

Human Dynamics in Smart Cities
Series Editors: Shih-Lung Shaw · Daniel Sui

Shih-Lung Shaw
Daniel Sui *Editors*

Mapping COVID-19 in Space and Time

Understanding the Spatial and Temporal
Dynamics of a Global Pandemic



Human Dynamics in Smart Cities

Series Editors

Shih-Lung Shaw, Department of Geography, University of Tennessee, Knoxville,
TN, USA

Daniel Sui, Department of Geography, Ohio State University, Columbus, OH, USA

This series covers advances in information and communication technology (ICT), mobile technology, and location-aware technology and ways in which they have fundamentally changed how social, political, economic and transportation systems work in today's globally connected world. These changes have raised many exciting research questions related to human dynamics at both disaggregate and aggregate levels that have attracted attentions of researchers from a wide range of disciplines. This book series aims to capture this emerging dynamic interdisciplinary field of research as a one-stop depository of our cumulative knowledge on this topic that will have profound implications for future human life in general and urban life in particular. Covering topics from theoretical perspectives, space-time analytics, modeling human dynamics, urban analytics, social media and big data, travel dynamics, to privacy issues, development of smart cities, and problems and prospects of human dynamics research. This will include contributions from the participants of the past and future Symposium on Human Dynamics Research held at the American Association of Geographers annual meeting as well as other researchers with research interests related to human dynamics via open submissions. The series invites contributions of theoretical, technical, or application aspects of human dynamics research from a global and interdisciplinary audience.

More information about this series at <http://www.springer.com/series/15897>

Shih-Lung Shaw · Daniel Sui
Editors

Mapping COVID-19 in Space and Time

Understanding the Spatial and Temporal
Dynamics of a Global Pandemic



Springer

Editors

Shih-Lung Shaw
Department of Geography
University of Tennessee
Knoxville, TN, USA

Daniel Sui
Geography and Public & International
Affairs
Virginia Polytechnic Institute and State
University
Blacksburg, VA, USA

ISSN 2523-7780

ISSN 2523-7799 (electronic)

Human Dynamics in Smart Cities

ISBN 978-3-030-72807-6

ISBN 978-3-030-72808-3 (eBook)

<https://doi.org/10.1007/978-3-030-72808-3>

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2021

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Foreword

Humans are social animals, given to meeting and assembling in groups and to the development of networks of friends. While much remains unknown about the transmission of the virus that causes COVID-19, it is clear that it depends on humans being in close proximity and thrives on people's social motivations. Terms such as *super-spreader* describe the ability of one infected individual to pass the virus to others whom they meet, share a drink or meal with, or ride together in an airplane. The result is a dynamic and evolving pattern of disease in space and time.

In recent years our ability to collect, analyze, and make inferences from space-time patterns has evolved very rapidly. We can make and publish maps, employ the sophisticated tools known as geographic information systems (GIS), track the locations of individuals through the GPS functions of smartphones, aggregate patterns into areal statistics, and plan the rollout of testing and vaccination sites. Our tools and databases allow us to look at the world in unprecedented detail, and in close to real time. Moreover the costs of all this have fallen dramatically: data can be shared on the Internet at virtually no cost, many of the necessary databases are freely available, today's smartphone computes with greater power than the supercomputers of thirty years ago and at a minute fraction of the cost, and much GIS software is open-source. No wonder, then, that the advent of COVID-19 has stimulated an outpouring of space-time research, as evidenced by the chapters of this book.

Yet beneath this rosy picture lie numerous issues that are central to the discipline of geographic information science (GIScience). Not the least among these, but central to a space-time perspective, is the question of positional accuracy. A GPS determines position with respect to the Earth's geodetic frame by, in effect, measuring the distances between the phone's current position and the Equator (to obtain latitude) and the Prime Meridian (to obtain longitude). Many factors impact the accuracy of these measurements, including tall buildings and overhanging trees, and as a result it is rarely possible to detect with certainty when two individuals and their smartphones come within 2 m of each other. Bluetooth allows direct communication

between devices and can sense the distance between them, but not reliably enough to avoid large numbers of false positives and false negatives. It may appear easy to compile statistics by county, by counting numbers of positive tests, for example, using the capability of GIS to assign points to polygons, but which location should be used, given that people move around in their daily activities and life histories: the county of residence or workplace, the county where the individual was infected, if known, or the county of the hospital where the individual was treated? This question may be moot for large counties, such as San Bernardino County in California, with an area of over 50,000 km², but is of critical importance for small counties, such as Manassas Park in Virginia, with an area of only 6.5 km².

While much of the research on COVID-19 using aggregate counts has relied on data at the county level, the populations and areas of the counties of the conterminous US both vary by four to five orders of magnitude—nothing about US counties is particularly “level.” Moreover we know from extensive research in GIScience over the past four decades that the spatial units used in this kind of research inevitably influence its outcomes; this Modifiable Areal Unit Problem (MAUP) creates significant issues for COVID-19 research and is often overlooked by researchers without the relevant GIScience awareness. Moreover, today’s tools invite their users to assemble data from diverse sources, often with little concern for the accuracy and provenance of the data. One of the more significant achievements of this book lies in the attention its chapters pay to issues like these.

The advent of COVID-19 has led to an outpouring of research, and this book is one of many that are in process or have already appeared. The work in microbiology that enabled the rapid development of effective vaccines has already led to Nobel prizes, and a flood of peer-reviewed papers has drawn inferences from the accumulating and readily accessible data about infections. As we move into a new phase of the pandemic, new issues appear that deserve the attention of the space–time research community. How should the logistic problems of vaccine distribution be addressed, to ensure equity as well as efficiency? What insights can a space–time perspective bring to the problems of vaccine resistance among individuals and social groups? What lessons have been learned from COVID-19 that will help us to respond more rapidly and effectively to the next major pandemic? What will be the long-term impacts of this pandemic on the spatial and temporal aspects of human behavior: commuting, residential choice, shopping, recreational travel, industrial location, in-person conferences, and social service provision? What new public policies will try to address the age-related impacts of COVID-19, including homes for the elderly?

As always, the best research solves important problems but also leads to a set of new questions. This pandemic is unique in occurring in an age when a large research community, armed with abundant tools and data, was able to respond quickly and to draw useful and important inferences. Yet we are left with a deep and abiding sense of awe that an organism measuring a hundred or so nanometers in diameter can disrupt our complacency in so many ways and inflict such damage on human civilization.

Michael F. Goodchild
Emeritus Professor of Geography
University of California
Santa Barbara, CA, USA
good@geog.ucsb.edu

Contents

1	Understanding the Spatial and Temporal Dynamics of a Global Pandemic	1
	Shih-Lung Shaw and Daniel Sui	
Part I Conceptual and Theoretical Perspectives		
2	A Power-Law-Based Approach to Mapping COVID-19 Cases in the United States	13
	Bin Jiang and Chris de Rijke	
3	Individual, Context, and Space: Using Spatial Approaches for Understanding Unequal Social and Psychological Fallout of COVID-19	25
	Grant Drawve, Casey T. Harris, and Kevin M. Fitzpatrick	
4	A Permanent Virtual Memorial for a Whistleblower of the COVID-19 Pandemic: A Case Study of Crypto Place on the Blockchain	47
	Xu Huang and Bo Zhao	
5	Emotional Responses Through COVID-19 in Singapore	61
	Yingwei Yan, Wei Chien Benny Chin, Chan-Hoong Leong, Yi-Chen Wang, and Chen-Chieh Feng	
6	A Socio-Ecological Perspective on COVID-19 Spatiotemporal Integrated Vulnerability in Singapore	81
	Chan-Hoong Leong, Wei Chien Benny Chin, Chen-Chieh Feng, and Yi-Chen Wang	

Part II Data Perspectives

- 7 Assessing Connections and Tradeoffs Between Geospatial Data Ethics, Privacy, and the Effectiveness of Digital Contact Tracing Technologies** 115
Peter Kedron and Andrew B. Trgovac
- 8 Challenges and Limitations of Geospatial Data and Analyses in the Context of COVID-19** 137
Sean G. Young, Jyotishka Datta, Bandana Kar, Xiao Huang, Malcolm D. Williamson, Jason A. Tullis, and Jackson Cothren
- 9 Multi-level Inter-regional Migrant Population Estimation Using Multi-source Spatiotemporal Big Data: A Case Study of Migrants in Hubei Province During the Outbreak of COVID-19 in Wuhan** 169
Jiale Qian, Zhang Liu, Yunyan Du, Nan Wang, Jiawei Yi, Yeran Sun, Ting Ma, Tao Pei, and Chenghu Zhou

Part III Analysis and Modeling Perspectives

- 10 Identifying and Characterising Active Travel Corridors for London in Response to Covid-19 Using Shortest Path and Streetspace Analysis** 191
Nicolas Palominos, Duncan A. Smith, and Sam Griffiths
- 11 Multi-scale CyberGIS Analytics for Detecting Spatiotemporal Patterns of COVID-19** 217
Fangzheng Lyu, Jeon-Young Kang, Shaohua Wang, Su Yeon Han, Zhiyu Li, and Shaowen Wang
- 12 Dynamic Spreading of COVID-19 Versus Community Mobility in Regions of England** 233
Tao Cheng, Xincheng Zhong, Yunzhe Liu, Yang Zhang, and Guangsheng Dong
- 13 Exploring Store Visit Changes During the COVID-19 Pandemic Using Mobile Phone Location Data** 253
Yunlei Liang, Kyle W. McNair, Song Gao, and Aslıgül Göçmen

Part IV Application and Policy Perspectives

- 14 Citizen Mobility and the Growth of Infections During the COVID-19 Pandemic with the Effects of Government Restrictions in Western Europe** 279
Mohd Sarim, Qunshan Zhao, and Nick Bailey

15 A Mathematical Model for Evaluating the Medical Resource Availability of COVID-19 in Time and Space	295
Fei-Ying Kuo and Tzai-Hung Wen	
16 Health Resilience Among European Countries in the Face of Pandemic: Reflections on European countries' preparedness for COVID-19	309
Yijing Li	
17 Improving Public Transportation Safety in COVID-19 Era Through Crowdsourcing Technique	325
Qisheng Pan, Zhonghua Jin, and Tao Tao	
18 Outlook and Next Steps: Understanding Human Dynamics in a Post-pandemic World—Beyond Mapping COVID-19 in Space and Time	347
Daniel Sui and Shih-Lung Shaw	

Chapter 1

Understanding the Spatial and Temporal Dynamics of a Global Pandemic



Shih-Lung Shaw and Daniel Sui

1.1 Introduction

Coronavirus disease 2019 (COVID-19) has caused unprecedented large-scale changes to human dynamics from local to global scale since early 2020. We have seen schools switching to online teaching, offices switching to teleworking, people with significantly fewer trips, store closures, job losses, healthcare challenges, and large number of deaths around many parts of the world due to COVID-19. On the one hand, we are trying to better understand the effects of COVID-19 and develop effective strategies of dealing with this pandemic. On the other hand, we are learning how vulnerable our societies are and re-thinking how we could re-design our social, economic, and healthcare systems to be more resilient to future pandemic events like the COVID-19.

Human societies are organized under various systems to meet our biological, economic, social, cultural, transportation, and other needs. These systems shape and are shaped by the spatial patterns and temporal rhythms of human dynamics. For example, we go to work or attend school at certain locations and during particular hours in a week. These spatiotemporal activities are closely related to the transportation systems, offices, schools, daycare centers, stores, and other facilities and services that operate in a coordinated manner to support human dynamics at various locations and with different schedules in our daily lives. When major events such as the COVID-19 pandemic disrupt the operation of these systems and the regular rhythm of human activities, human societies respond with policy and behavioral changes to address the challenges which often lead to different spatial and temporal dynamics. If a disruption lasts for a short time period, impacts on the existing systems and human

S.-L. Shaw (✉)
University of Tennessee, Knoxville, TN, USA
e-mail: sshaw@utk.edu

D. Sui
Geography and Public & International Affairs, Burruss Hall, Suite 340, 800 Drillfield Drive
Virginia Tech, Blacksburg, VA 24061, USA

societies usually are minor and the systems return back to normal shortly after a disruption. However, when an event disrupts normal operation of the societies for an extended period of time, various systems in human societies may or may not return back to the same operations and the same rhythms before the event.

By the time of this book's publication, the COVID-19 pandemic has been around for over one year with no clear sign when it will be under control or end. In fact, some countries were facing record-high daily confirmed cases and deaths that prompted tightened controls of human dynamics. In the meantime, many countries around the world struggled with major impacts of this pandemic such as store closings, job losses, healthcare crises, and COVID-related deaths. Researchers in a wide range of fields responded quickly to conduct studies that could help us better understand and gain insights into the causes and spread of this pandemic, medical treatments of this virus, strategies of adjusting various systems in our societies to respond to the pandemic, and implications of this pandemic to the future development of human societies. A quick search in Google Scholar returns a long list of studies covering different aspects of the COVID-19 pandemic. This book aims at contributing to the literature by focusing on the spatial and temporal dynamics of the global COVID-19 pandemic with a set of research papers written by scholars in Asia, Europe, and North America. This collection of papers also is closely aligned with the human dynamics theme of this *Human Dynamics in Smart Cities* book series.

1.1.1 Spatial and Temporal Dynamics of the COVID-19 Pandemic

As of January 14, 2021, COVID-19 had spread around the world with a total number of global cases exceeding 93 million and a total number of global deaths over 1.98 million (<https://coronavirus.jhu.edu/map.html>). The numbers of confirmed cases and deaths reached new highs in some countries by early 2021 as new variants of COVID-19 virus were identified. It is evident there exist uneven spatial patterns of COVID-19 cases across different countries and even among different regions within a country. In addition, these uneven spatial patterns have evolved over time with different countries or regions experiencing higher confirmed cases and/or deaths than other countries or regions. We also have observed various approaches and strategies implemented by the national and local governments around the world to control the pandemic. “Flatten the curve” has been one common strategy adopted by many governments that includes policies such as wearing a mask, social distancing, travel restriction, self-isolation, quarantine, lockdown, and border closure (e.g., Thunström et al. 2020, Matrajt and Leung 2020, CDC 2021, USA Today 2021). Although these “flatten the curve” policies often have positive effects on mitigating the spread of COVID-19, they also have caused many major changes of human behavior that in turn led to major impacts on our daily lives and the societies. Teleworking, virtual education, travel reduction, store closure, job loss, and low usage of office and other spaces

are just some examples of the changing dynamics in the societies around the world. With the varying policies and mixed outcomes found across the world, one major issue has been finding a balance between restricting human dynamics to control the spread of COVID-19 and sustaining the economy and normal lives.

Bonaccorsi et al. (2020) use human mobility data provided by Facebook to analyze how lockdown strategies affect economic conditions of individuals and local governments in Italy. They report two major effects of lockdown measures that mobility impact is higher in municipalities with a higher fiscal capacity and mobility contraction is stronger in municipalities with high inequality and lower per-capita income. Nicola et al. (2020) provide a review of the socioeconomic implications of COVID-19 and summarize the effects of COVID-19 on various industries in the primary, secondary, and tertiary sectors. They conclude that “It is prudent that governments and financial institutions constantly re-assess and re-evaluate the state of play and ensure that the ‘*whatever it takes*’ promise is truly delivered. (p. 190)” This reflects the dynamic nature of the COVID-19 pandemic. McKibbin and Fernando (2020) present seven scenarios for the global macroeconomic impacts of COVID-19. These seven scenarios vary by attack rate, mortality rate, and the countries experiencing the epidemiological shocks. The first three scenarios assume the epidemiological shocks are isolated to China with spillover effects to other nations through trade, capital flows and changes in risk premia in global financial markets. The next three scenarios assume the epidemiological shocks occur in all countries at different degrees. The last scenario assumes a mild pandemic reoccurs each year for the indefinite future. Their analysis results indicate that the global economy in the short run could be significantly impacted even with a contained COVID-19. As of early 2021, it is evident that the epidemiological shocks had occurred in many countries with a possibility of recurring each year, which correspond to the last four scenarios suggested by McKibbin and Fernando.

On the one hand, research is needed to address medical challenges such as understanding the COVID-19 virus, treatment of COVID-19 patients, and development of vaccines to overcome this pandemic. On the other hand, it is equally important to gain insights into the non-medical aspects of the COVID-19 pandemic. Rose-Redwood et al (2020) present an introductory article in a special issue of 42 commentaries on the uneven geographies of COVID-19 based on different theoretical, methodological, and thematic approaches. The uneven geographies of COVID-19 are not static spatial patterns. These spatial patterns and their associated spatial processes evolve over time. The chapters in this book are non-medical research with a focus on the spatial and temporal dynamics of the COVID-19 pandemic that are organized around four specific perspectives. The first one is *conceptual/theoretical perspectives*. Human dynamics have been studied in many different fields based on different conceptual and theoretical approaches (Shaw and Sui 2018a, b). For example, Barabási (2005) published an influential paper that led to many studies of human dynamics based on the concepts of statistical physics and data-driven approaches. In the meantime, time geography developed by Hägerstrand (1970) also has been widely used as the conceptual/theoretical foundation of studying human dynamics. Furthermore, as

human activities and interactions have increasingly moved into virtual space especially due to the restrictions on physical activities during the COVID-19 pandemic, new conceptual and theoretical approaches are needed to gain insights into the spatial and temporal human dynamics taking place under the pandemic. The first section of this book consists of papers related to the conceptual/theoretical perspectives.

Data are the prerequisite of conducting most research. There are many data issues and concerns when pursuing COVID-19 related research. The second section of this book therefore is centered around the *data perspectives*. COVID-19 research often involves sensitive data in support of studies such as contact tracing. One critical challenge is to address the conflicts between privacy protection and the effectiveness of digital contact tracing using geospatial technologies with respect to the principles of efficacy, privacy, equity, transparency, and accountability. In addition, the National Academy of Sciences, Engineering and Medicine (2019) published a report of *Reproducibility and Replicability in Science* as a result of a mandate from the U.S. Congress. Reproducibility is defined in this report as “obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis. (p. 6)” Replicability is defined as “obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data. (p. 6)” Chapters in this section cover issues related to reproducibility and replicability of COVID-19 research as well as other data-related challenges.

COVID-19 research has been pursued with a wide range of analysis methods and models. The section of *analysis and modeling perspectives* in this book focuses on spatiotemporal analysis and modeling of COVID-19 dynamics. Time is a critical dimension in spatiotemporal analysis and modeling of dynamic phenomena. With the changing dynamics of COVID-19, it is important to examine the dynamic changes across space and over time, investigate potential spatial lag and temporal lag effects of the policies and strategies implemented to control the pandemic, and design better systems to serve changing human needs in a space-time context. These spatiotemporal analysis methods and models in support of real-time or near-real-time applications may need special designs and infrastructures to support them due to their specific computational requirements. Chapters in this section present some approaches and challenges associated with analysis and modeling of spatiotemporal dynamics of the COVID-19 pandemic.

The final section of this book consists of chapters related to the *application and policy perspectives*. With this unprecedented pandemic, most governments and individuals were uncertain what would be good and effective ways of dealing with the pandemic in many fronts. For example, should schools stay with in-person teaching? What would be effective policies of flattening the curve without having major impacts on the economy? Would the proposed policies cause major negative impacts on particular population groups due to the nature of varying spatiotemporal activity needs among different population groups? There exist many potential application and policy studies related to the COVID-19 pandemic and there is a limit to what one book could possibly cover. This book focuses on the spatial and temporal dynamics of the COVID-19 pandemic to complement the existing publications.

1.2 Overview of the Chapters in this Edited Volume

This book is organized into four sections with a total of eighteen chapters. Chapter 1 is written by the editors as an introduction to this book. Other substantive chapters are organized around the four sections of conceptual and theoretical perspectives, data perspectives, analysis and modeling perspectives, and application and policy perspectives. This book ends with another chapter by the editors with a summary and suggestions of future research directions.

The section of *conceptual and theoretical perspectives* consists of five papers. Jiang and Rijke take a statistical physics approach to examining the spatial and temporal distributions of COVID-19 cases against the underlying population distribution in the U.S. They indicate that the distribution of COVID-19 cases becomes a power-law distribution as time passes, which is largely shaped by the underlying U.S. population distribution. The chapter by Drawve, Harris and Fitzpatrick presents the value of a geographic lens for understanding the multidimensional social, economic, and psychological impacts of COVID-19. They use survey data of over 10,000 individuals collected in March 2020 to describe how spatial processes are a critical piece of understanding the current public health crisis in the social sciences. Chapter 4 by Huang and Zhao introduces the concept of crypto place using the blockchain technique as an encrypted monument in the COVID-19 pandemic that consists of three dimensions: decentralized location, immutable locale, and transaction-based sense of place. They suggest that, although people are confined to the secular/physical world, they can transcend the space-time limitations in the real world and achieve psychological freedom of subjective existence. This concept is especially relevant in the hybrid physical-virtual world enabled by modern technologies. Yan et al. focus on the emotional impacts on people in Singapore during the COVID-19 pandemic using Twitter data and sentiment analysis. This study brings the emotional dimension of human dynamics to COVID-19 research. They report that people in Singapore generally had a high level of trust and positive attitude towards each other during the pandemic, which could be attributed to smart crisis management plan initiated by the government. In the meantime, they indicate that people in different native language groups exhibited differences in their emotional profiles within this multi-ethnic and multi-lingual nation. The last chapter in this section by Leong et al. develops a socio-ecological framework of epidemiology to gain insights into local vulnerability by considering changes of risks and human mobility across space and over time. This exploratory study integrates spatial and temporal urban public transportation data to develop a measure on epidemic vulnerability. The findings suggest spatiotemporal vulnerability could be shaped by multiple forces such as urban social and built environments, local demographic characteristics, and the composite spatial lag effects from contiguous subzones. The pandemic disruptions are likely to encourage a decentralization trend and a new routine that augment the importance of mobility distance and spatial lag effects.

The next section of *data perspectives* includes three chapters covering different aspects of data challenges. Contact tracing is a useful tool for mitigation and suppression of the COVID-19 pandemic. However, digital contact tracing using geospatial tracking technologies brings up a number of concerns and issues. Kedron and Trgovac introduce a framework for assessing conflicts between privacy and effectiveness by examining linkages between the potential capabilities of geospatial tracking technologies and ethical and privacy principles of geospatial data handling. They present efficacy, privacy, equity, transparency, and accountability as five key principles that address the potential privacy, ethical, and social impacts of these technologies. This chapter highlights a need of bringing researchers with varied backgrounds to critically examine the implications of digital contact tracing technologies through their unique lens. Young et al., on the other hand, discuss the challenges and limitations of geospatial data and analysis for COVID-19. This chapter presents how human mobility data can augment traditional data sources to explore relationships between human behaviors and spatiotemporal spread of COVID-19. It also discusses the importance of considering the scale of analysis when modeling COVID-19 spreads and some uncertainty issues associated with geospatial disease modeling. Discussions in this chapter are framed in the context of reproducibility and replicability of COVID-19 research. Qian et al. examine population movements in Hubei province during the COVID-19 outbreak in Wuhan, China. They use location data from online apps and other relevant data to develop a method of estimating population flows between Wuhan and other cities and counties in Hubei province around the Chinese New Year in 2020. Such population movement estimates derived from location tracking data of online apps can be useful in assessing potential risk levels of infectious disease spread associated with a major holiday.

There are four chapters included in the section of *analysis and modeling perspectives*. Many different analysis methods and models have been used to gain insights into various aspects of the COVID-19 pandemic. One common characteristic of the chapters included in this section is they all place an emphasis on the spatial and temporal aspects of COVID-19. Palominos, Smith and Griffiths conduct a study of developing new cycling and walking schemes and designing place-based streets in London, United Kingdom to mitigate virus spread. They investigate potential streetspace reallocations needed to create a micro-mobility network that prioritizes space for active travel and public transport under the current demands for healthy transport modes due to the COVID-19 pandemic. The proposed method is based on network analysis and considers a completely connected active travel network in which direct and fast routes are prioritized for active travel modes at the expense of modes that are less efficient and have a greater environmental impact. The chapter by Lyu et al. argues there is an urgent need of dynamic and interactive mapping informed by rigorous geospatial analysis and modeling for understanding how COVID-19 spreads at various spatial and temporal scales to support decision making of mitigating the spread and minimizing negative impacts. Due to the data-intensive and computation-intensive nature of geospatial analysis and modeling, they develop a multi-scale cyberGIS approach to detect spatiotemporal clusters of COVID-19 death cases and explore potential factors contributing to the spatiotemporal clusters and patterns. Cheng et al.

evaluate the effectiveness of non-pharmaceutical interventions (NPIs) such as lockdown in reducing the mobility and the spreading of the COVID-19 virus in both the first and the second waves of outbreaks in the nine regions of England, United Kingdom. They employ Susceptible-Exposed-Infectious-Recovered (SEIR) model with Bayesian analysis and the maximum information coefficient to investigate the spatial-temporal relationships between community mobility and the reproduction rate based on Google's mobility reports. The last chapter by Liang et al. examines the dynamics of customer visits to five popular chain-store brands in pre-lockdown and after lockdown periods in three large cities of the United States. They use a time-aware dynamic Huff model to analyze store visiting patterns at the start of the pandemic and after the lockdowns based on mobile phone location data. It is evident from the four chapters in this section that various methods and models are available for spatiotemporal analysis of pandemic data. However, it is important to assess these methods and models in relation to the discussions in the section of conceptual and theoretical perspectives and the section of data perspectives. For example, to what extent are the analysis findings reported in this section reproducible and/or replicable? What are the connections between the conceptual/theoretical foundation and the choice of methods/models for spatiotemporal analysis?

The final section of *application and policy perspectives* includes four chapters that provide policy implications. Sarim, Zhao and Bailey focus on the mobility restrictions imposed by governments to reduce the spread of COVID-19 cases. They use the Oxford COVID-19 Government Response Tracker to measure the severity of restrictions on mobility in eight Western European countries and the data from Google mobility reports to assess the impacts of mobility restrictions on different types of human activity. This study groups the eight countries into three lockdown categories (strict, intermediate, and flexible). The findings suggest that strict restrictions of mobility such as working from home order is important to limit the spread of COVID-19 cases. However, there is a time lag effect of the mobility restrictions due to an incubation period of the COVID-19 virus. It takes at least 2-3 weeks of lockdown to be effective of controlling the virus spread. In the meantime, this study indicates that essential shopping and exercise in parks with social distancing and face mask do not appear to increase the infection growth rate. Kuo and Wen argue in the next chapter that, due to the spatiotemporal spread of COVID-19 cases, it is important to incorporate the spatial-temporal dynamics of epidemic-induced demand into the assessment of medical resource availability at various stages of the pandemic spread. They propose a model that integrates epidemic dynamic process and resource accessibility to assess the changing ration levels of medical resources in a space-time context based on the data collected in Taipei, Taiwan as a case study. This study illustrates that the resource demand for testing reagents and the ration level at different locations vary across space and over time. The proposed model could be employed to support health authorities in making decisions on medical resource allocation plans to confront the spread of a pandemic. The chapter by Li uses a Pandemic Resilience of Place (PROP) model to compare relative resilience against the COVID-19 pandemic among 35 European nations. Three groups are identified among the 35 European nations based on their time-series trend of accumulated confirmed COVID-19 cases

and deaths, which are (1) nations with the rate of confirmed cases increased faster than deaths, (2) nations with the rate of deaths increased faster than confirmed cases, and (3) nations with similar rates of confirmed cases and deaths. This study further uses seven indicators to measure national preparedness for the COVID-19 pandemic and makes recommendations on the actions that should be taken at the national and provincial scales to enhance the resilience level. In the final chapter of this section, Pan et al. propose a conceptual framework based on crowdsourcing methods to collect timely epidemiological information from transit riders and operators to monitor and respond to the infectious disease in public transit systems. This framework is designed to support public transit management on implementing rules and regulations that can reduce the risk of spreading COVID-19 and increase public transportation safety.

There have been many publications addressing various critical research questions related to the COVID-19 pandemic. The collection of chapters in this book places an emphasis on the spatial and temporal dynamics of this pandemic. Authors of these chapters are from geography, planning, sociology, engineering, mathematics, urban design, geographic information science, and other fields in Asia, Europe, and North America. They offer various perspectives on the theories/concepts, data, analysis/modeling, and application/policy for studies of the spatial and temporal dynamics of the COVID-19 pandemic, which add a set of useful papers to the literature. This book also offers a unique addition to the *Human Dynamics in Smart Cities* book series by examining important changes to human dynamics due to a major pandemic.

References

- Barabási, A.-L. (2005). The origin of bursts and heavy tails in human dynamics. *Nature*, 435(7039), 207–211.
- Centers for Diseases Control and Prevention (CDC). (2021). *COVID-19 travel recommendations by destinations*. Retrieved January 12, 2021, from <https://www.cdc.gov/coronavirus/2019-ncov/travelers/map-and-travel-notices.html>.
- Giovanni Bonacorsia, G., Pierrib, F., Cinellic, M., Floria, A., Galeazzid, A., Porcellie, F., et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*, 117(27), 15530–15535.
- Hägerstrand, T. (1970). What about people in regional science? *Papers of the Regional Science Association*, 24(1), 7–21.
- Matrajt, L., & Leung, T. (2020). Evaluating the effectiveness of social distancing interventions to delay or flatten the epidemic curve of coronavirus disease. *Emerging Infectious Diseases*, 26(8), 1740–1748.
- McKibbin, W., & Fernando, R. (2020). The global macroeconomic impacts of COVID-19: Seven scenarios. Working Paper 19/2000, Center for Applied Macroeconomic Analysis, Crawford School of Public Policy, Australian National University. Retrieved from <https://cama.cra.wf.ard.edu.au/publication/cama-working-paper-series/16221/global-macroeconomic-impacts-covid-19-seven-scenarios>.
- National Academies of Sciences, Engineering and Medicine. (2019). *Reproducibility and replicability in science*. Washington DC: The National Academies Press.

- Nicola, M., Alsaifib, Z., Sohrabic, C., Kerwand, A., Al-Jabird, A., Iosifidisc, C., et al. (2020). The socio-economic implications of the coronavirus pandemic (COVID-19): A review. *International Journal of Surgery*, 78, 185–193.
- Rose-Redwood, R., Kitchin, R., Apostolopoulou, E., Rickards, L., Blackman, T., Crampton, J., et al. (2020). Geographies of the COVID-19 pandemic. *Dialogues in Human Geography*, 10(2), 97–106.
- Shaw, S.-L., & Sui, D. (2018a). GIScience for human dynamics research in a changing world. *Transactions in GIS*, 22(4), 891–899.
- Shaw, S.-L., & Sui, D. (Eds.). (2018b). *Human Dynamics Research in Smart and Connected Communities*. Cham, Switzerland: Springer International Publishing AG.
- Thunström, L., Newbold, S. C., Finnoff, D., Ashworth, M., & Shogren, J. F. (2020). The benefits and costs of using social distancing to flatten the curve for COVID-19. *Journal of Benefit-Cost Analysis*, 11(2), 179–195.
- USA Today. (2021). COVID-19 restrictions: Map of COVID-19 case trends, restrictions and mobility. Retrieved January 13, 2021, from <https://www.usatoday.com/storytelling/coronavirus-reopening-america-map/>.

Part I

Conceptual and Theoretical Perspectives

Chapter 2

A Power-Law-Based Approach to Mapping COVID-19 Cases in the United States



Bin Jiang and Chris de Rijke

2.1 Introduction

The novel coronavirus COVID-19 has rapidly spread around the world and triggered an unprecedented pandemic in the few months since January 2020. At the time of writing this paper, over 34.2 million people globally had been infected, with over 1 million deaths, and the situation is still developing. How to better understand the spread mechanisms of the coronavirus in space and in time across different levels of scale concerns many scientists such as geographers, cartographers, and epidemiologists. Many previous studies have already examined the spatial distribution of COVID-19 cases using conventional geographic information systems (GIS) and mapping methods such as hotspot and time series analyses (ESRI 2020). These methods are developed essentially under Gaussian statistics (Jiang 2015) with the assumption that data varies around a characteristic mean (e.g., 1.75 meters as the characteristic mean for human height). A common problem of these methods is that the resulting spatial patterns are sensitive to human subjective decisions like parameter settings. For example, either the number of classes or class intervals has to be decided subjectively. In contrast, we adopt a power-law-based approach under Pareto statistics for examining spatial and temporal distribution of COVID-19 cases in the United States.

We examine the spatiotemporal distribution of all COVID-19 cases in the US across multiple scales of space and time. In space, there are two levels of scale—state and county—whereas in time there are three levels: monthly, weekly, and daily. We detected a power law distribution ($y = kx^{-a} + m$, where a is called power

B. Jiang · C. de Rijke

Faculty of Engineering and Sustainable Development, Division of GIScience, University of Gävle, 801 76 Gävle, Sweden
e-mail: bin.jiang@hig.se

C. de Rijke
e-mail: chris.de.rijke@hig.se

law exponent between 1 and 3, and k and m are two constants.) for each of three parameters: population, infection, and death. All these three parameters demonstrate power laws with cut-off, despite of some fluctuations for both infection and death. The power law indicates that these three parameters bear an inherent hierarchy or spatial heterogeneity, with far more small events than high ones. To derive this hierarchy, we used head/tail breaks (Jiang 2013) so that each state or county is assigned a ht-index for each of these three parameters to indicate its hierarchical level. The derived hierarchical levels provide new insights into the development of the pandemic for individual states and counties relative to their populations. For example, the pandemic is largely shaped by the underlying population with the R-square value between infection and population up to 0.82. The power-law-based approach enables us to see spatiotemporal patterns that the conventional methods are unable to discover. The approach has a profound implication on power-law-related research in terms of whether data exhibits a power law or any other similar distribution. That is, from a dynamic view, power law is usually observed when a complex system is fully developed, before which the system is likely to exhibit other less-power-law distributions such as lognormal and exponential. For example, there is little doubt that a tree as a complex biological system demonstrates a power law distribution for its trunk, branches, and leaves, because there are far more leaves than branches, and far more branches than trunk. However, the tree is unlikely to hold a power law at the stage of the germination of the seed. We will further discuss this implication before the conclusion.

The remainder of this paper is structured as follows. Section 2.2 introduces the data source initially collected by Johns Hopkins University, and the head/tail breaks illustrated by a simple example of the 10 numbers. The power law detection is based on the maximum likelihood method (Clauset et al. 2009), arguably the most robust statistical test. Section 2.3 presents our results and discussion, as well as an animation map (<http://lifegis.hig.se/COVID19/>). Section 2.4 highlights the implication we briefly mentioned above. Finally, Sect. 2.5 draws a conclusion of this paper and points to possible future work.

2.2 Data Source and Methodology

Over three million people have been infected and 208,000 people have died from COVID-19 in the US from January to June of 2020. Johns Hopkins University (2020) has gathered this data and published it on the GitHub website. This data is compared with the country's population at both state and county levels. In general, the two parameters—*infection* and *death*—are highly related to population. Like the population in the US, the numbers of *infection* and *death* are highly concentrated in a few well-populated states and counties. In this study, we intend to compare COVID-19 cases against the underlying population, in order to develop new insights into spatiotemporal patterns of the pandemic from the multiple scales of space and time.

Like all countries, the US population is not evenly distributed, and it has a very high degree of concentration in a few cities, states, or counties, the so-called inherent hierarchy or spatial heterogeneity. At the city level, this kind of spatial distribution is usually characterized by a power law distribution, such as Zipf's law (1949). Zipf's law states that in terms of population the first largest city is twice as big as the second largest, three times as big as the third largest and so on. At the county level, the top 20% counties accommodate 80% of the population—the so-called 80/20 principle (Koch 1998) that is credited to the Italian economist and polymath Vilfredo Pareto (1848–1923). What is behind Zipf's law and the 80/20 principle—or the power law in general—is the inherent hierarchy or spatial heterogeneity, which can be illustrated through the head/tail breaks classification scheme (Jiang 2013). This is a recursive function that can be used to derive the inherent hierarchy of data with a heavy-tailed distribution. The derived hierarchical levels or classes reflect the recurrence of far more small numbers than large ones, or spatial heterogeneity, characterized by the ht-index (Jiang and Yin 2014).

Unlike conventional classification methods, with which the number of classes or class intervals are subjectively determined, head/tail breaks adopts the wisdom of crowds thinking (Surowiecki 2004), through which both the number of classes and class intervals are objectively determined by the data; in other words, the data speaks for itself. Head/tail breaks is recursive function, through which a dataset is conceived as the head of the head of head and so on, and all the tails and the last head constitute the derived classes or inherent hierarchical levels.

To further illustrate the recursive function, let us use a simple example of the 10 numbers $[1, 1/2, 1/3, \dots, 1/10]$, which follow exactly a rank-size distribution in the so called rank-size plot in which the x-axis is rank, while the y-axis is size. Strictly speaking, these 10 numbers cannot be said to be distributed according to Zipf's law, for it is a statistical regularity. Instead these 10 numbers $[1 + e_1, 1/2 + e_2, 1/3 + e_3, \dots, 1/10 + e_{10}]$ (where e_i is a very small value epsilon) are said to fit Zipf's law. Back to the first 10 numbers (Fig. 2.1, Jiang and Slocum 2020), its average is 0.29, which partitions the 10 numbers into two groups: those greater than the average $[1, 1/2, 1/3]$

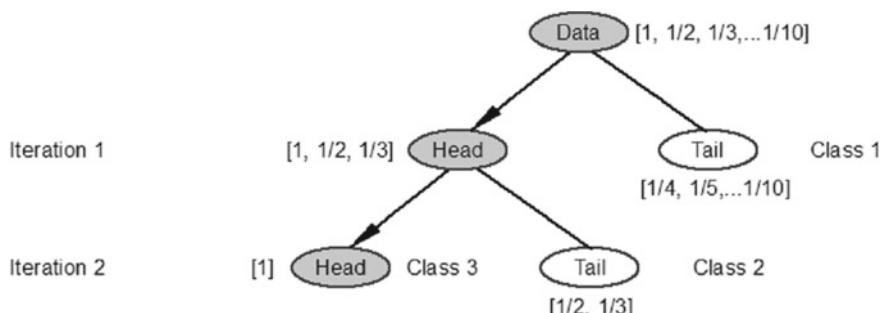


Fig. 2.1 Illustration of head/tail breaks classification with a simple example of the 10 numbers. Note The 10 numbers $[1, 1/2, 1/3, \dots, 1/10]$ are classified into three classes: $[1/4, 1/5, \dots, 1/10]$, $[1/2, 1/3]$, and $[1]$, which can be said to have three inherent hierarchical levels

called the head, and those less than the average [1/4, 1/5, 1/6, ... 1/10] called the tail. For those in the head [1, 1/2, 1/3], their average is 0.61, which further partitions the three into two groups: the one greater than the average [1] called the head and those less the average [1/2, 1/3] called the tail. The number of iterations or the notion of far more smalls than larges occurs twice, so the ht-index is three, indicating three inherent hierarchical levels. As shown in this example, the head percentage is far less than the preset 40%. The 40% is a very loose condition for something to be a minority to meet the notion far more smalls than larges.

In this study, we detect power laws using the robust maximum likelihood method (Clauset et al. 2009), calculate the ht-index for all COVID-19 cases in the US, and compare these calculated parameters with those of the population at both state and county levels along the time dimension. This type of comparison provides new insights into the spatiotemporal patterns of the pandemic. Before getting into the results, we would like to make one point explicitly clear about power law exponent a . It is a good indicator for heterogeneity of data: the higher the exponent, the more heterogeneous the data. For example, given two power laws, $y = x^{-2}$, and $y = x^{-3}$, the one with the exponent 3 is more heterogeneous than the one with the exponent 2. Throughout our study, we will show that the ht-index is a better indicator than the power law exponent for characterizing the data heterogeneity.

2.3 Results and Discussion

The population of the US looks like a power law distribution at both state and county levels, as shown in Fig. 2.2, but they are power laws with an exponential cut-off, strictly speaking. It is pretty the same for both infection and death gradually developing towards power laws with cut-off as time passes (shown as light blue and light red to dark blue and dark red). In March or before early April, both infection and death exhibit moderate power laws, while after April they are power laws with cut-off. As an example, Table 2.1 shows that how death at the county level demonstrate a power law or a power law with cut-off with detailed statistics.

The likelihood ratio (LR) (Clauset et al. 2009) can be used to determine power laws or power laws with cut-off in comparison to its alternative heavy tailed distributions such as lognormal and exponential. As a rule, a positive LR favors the power law fit, while a negative LR says the alternative fit. On the other hand, the LR is trustworthy if the statistical fluctuation of LR is relatively small. Therefore, an additional p-value is defined to the LR is trustworthy statistically (Clauset et al., 2009); if $p < 0.1$, then the LR is trustable.

At the state level, the LR is not statistically significant as the p-values are too high, therefore we cannot be certain that either of the distributions is a better fit. This is likely due to the small sample size ($n = 51$). At the county level with the large sample ($n = 3262$), the alternative lognormal distribution is more likely than the power law distribution for most of the weeks. However, the support for power laws with cut-off

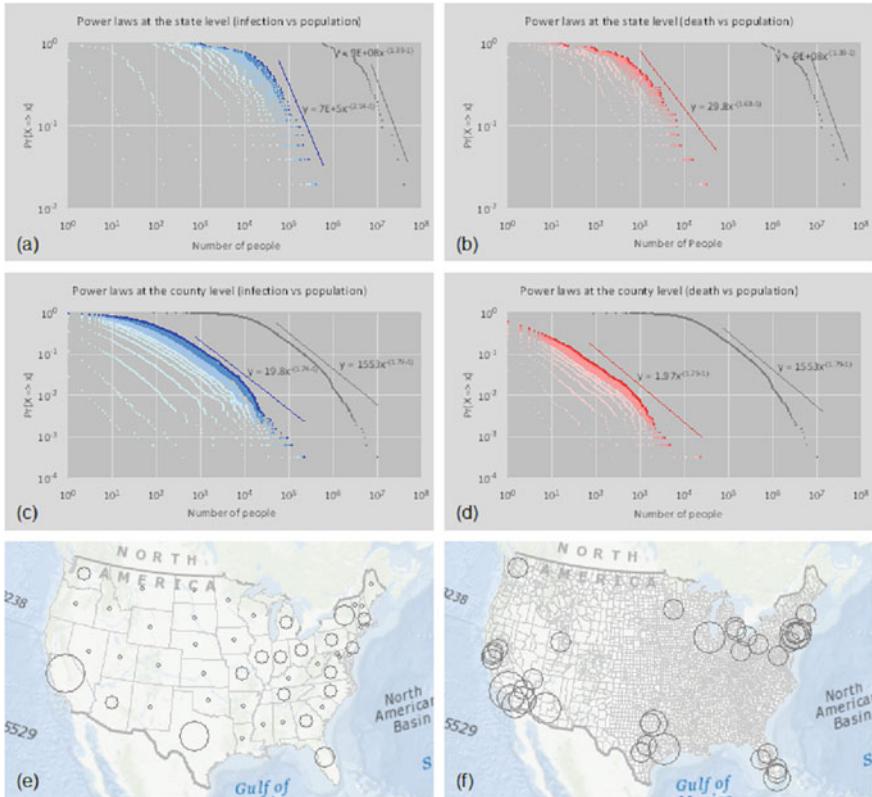


Fig. 2.2 (Color online) Power laws of infection (blue), death (red), and population (gray). Note At the state level (**a, b**), because of the large areal unit, the power laws with cut-off are not so striking, whereas at the county level (**c, d**), because of the small areal unit, the power law with cut-off are very striking. The hierarchy of population is mapped at both (**e**) state and (**f**) county levels, indicating far more less-populated states than well-populated ones, or far more less-populated counties than well-populated ones. Interestingly, the notion of far more smaller than large recurs four and six times at the state and county levels, respectively, thus with the ht-indexes being 5 and 7, indicating the inherent hierarchical levels. All the five levels are shown in panel (**e**), whereas only the top four levels are shown in panel (**f**) for the sake of legibility

are even more likely than the lognormal distribution. In the end, power laws with cut-off are more likely than the lognormal and exponential distributions.

The log-log plots in Fig. 2.2 indicate that the overall spatial distributions of infection and death are very much shaped by the underlying population. That is, those populated states and counties tend to have far more cases of infection or death. This is of course not out of our expectation, since the more the population, the more likely the infection or death. Given the power law distribution of the population and by applying the head/tail breaks, we derive ht-indexes of 5 and 7 at the state and county levels, respectively. In other words, the population is automatically classified into 5 and 7 classes, as shown in Fig. 2.2 (e, f). These two patterns regarding population at

Table 2.1 Support of a moderate power law or a power law with cut-off in comparison to alternatives

	Lognormal		Exponential		PL with cutoff		Support
	LR	P	LR	P	LR	P	
Week 12	-4.74	0.1	122.24	0.02	-4.46	0	Moderate PL
Week 13	-0.29	0.82	229.23	0.03	-0.2	0.52	Moderate PL
Week 14	-0.03	0.9	182.22	0.03	-0.02	0.86	Moderate PL
Week 15	-0.3	0.68	175.6	0.03	-0.18	0.55	Moderate PL
Week 16	-4.04	0.01	514.44	0	-2.35	0.03	PL with cutoff
Week 17	-1.68	0.27	340.5	0	-1.62	0.07	PL with cutoff
Week 18	-1.51	0.29	353.76	0	-1.6	0.07	PL with cutoff
...							
Week 27	-2.21	0.18	308.44	0	-3.48	0.01	PL with cutoff
Week 28	-2.66	0.14	328.65	0	-3.78	0.01	PL with cutoff

Note LR = likelihood ratio, PL = power law, P = p-value as defined in Clauset et al. (2009)

the state and county levels reflect the patterns of COVID-19 cases fairly well. That is, the states and counties on the West and East Coasts tend to have higher numbers of COVID-19 cases than those inland, which will be examined in the following. These two patterns at the state and county levels constitute the basic patterns to which COVID-19 cases can be compared in order to develop new insights into the pandemic in terms of its spatial and temporal patterns.

It is clear from Fig. 2.2 that the power law distributions have different exponents. The different power law exponents indicate the different degree of heterogeneity or hierarchy; that is, the higher the exponents, the more heterogeneous the data. In this connection, the ht-index is a better indicator than the power exponent as it better reflects the inherent hierarchy. As shown in Fig. 2.3a, b, the ht-indexes of both infection and death increase towards that of population. There is little wonder that the ht-index of the population remains unchanged—that is, 5 at the state level and 7 at the county level—indicating that the population is more heterogeneous at the county level than that in the state level (Fig. 2.3a, b). This is because the population in the large areal unit of states tends to be more homogenized than that in the small areal unit of counties. According to this logic, the population in the small areal unit of cities tends to be more heterogenized than that in the large areal unit of counties. This is indeed true, as shown in the literature (e.g., Newman 2005). What is interesting for infection and death is that they have a very low ht-index of 0 or 1 at the very beginning and increase rapidly towards 5 and 7, with some fluctuation in the course of development of the pandemic. This means that lockdown policies or social distancing measures are definitely effective at containing and combating the spread of the virus; otherwise, the situation would be far more devastating than it is currently. The result shows that COVID-19 cases are largely shaped by the underlying population, seen through the increasing correlations between infection and population and between death and population (Fig. 2.3c, d). In other words, two patterns shown in Fig. 2.2 (e,

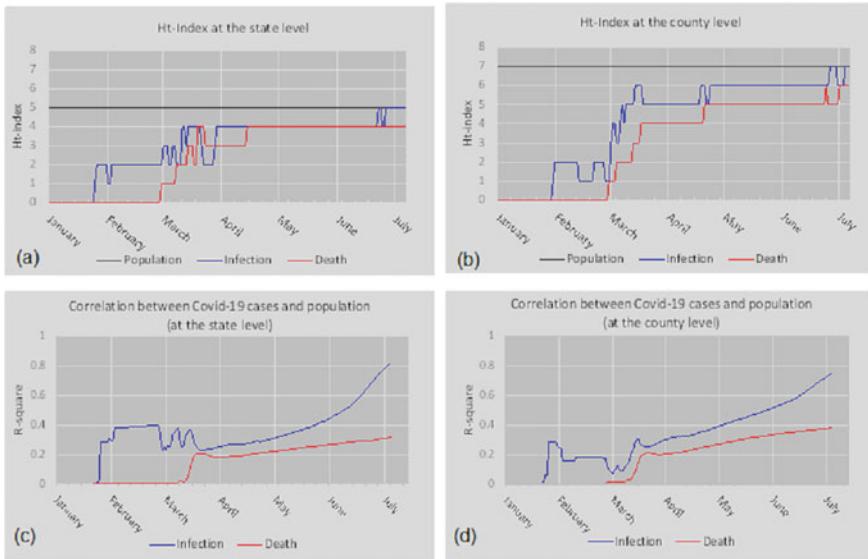


Fig. 2.3 (Color online) Relationship between COVID-19 cases and populations. Note The ht-indexes for the population are 5 and 7 respectively at the state (**a**) and county (**b**) levels, while the ht-indexes for the infection and death increase from mild to wild status despite some slight fluctuations. The correlations between infection and population, and between death and population increase at both the state (**c**) and county (**d**) levels

f) largely reflect those of infection and death; that is, populated states and counties tend to have more far more COVID-19 cases.

As elaborated above, the ht-indexes of infection and death at both the state and country levels are persistently approaching to that of population, and the correlations between infection and population, and between death and population increase also as time goes (Fig. 2.3). This is the overall picture. On the other hand, the hierarchical levels for these three parameters (population, infection, and death) provide a much more complex and interesting picture about the pandemic (Fig. 2.4). By examining the ht-indexes of the three parameters (population, infection and death) for individual states and counties, we can see how the pandemic hits individual states and counties differently relative to their total populations. For example, New York and its nearby states are hit most hard as reflected by the larger red circles, whereas California and Texas are affected far less, as shown by larger gray circles (Fig. 2.4d). It is important to assess how this latest situation evolved from a dynamic point of view. For example, the situation in January and February was very mild; only five states had a relatively high degree of infection, with Washington State having the highest. The situation took a drastic turn into very wild in March, when there were suddenly six states with larger red circles, indicating that hierarchical levels of death were larger than those of population and infection. This was a dangerous sign. From March to April, and from May to June, the situation got worsened, with a few exceptions such as Washington

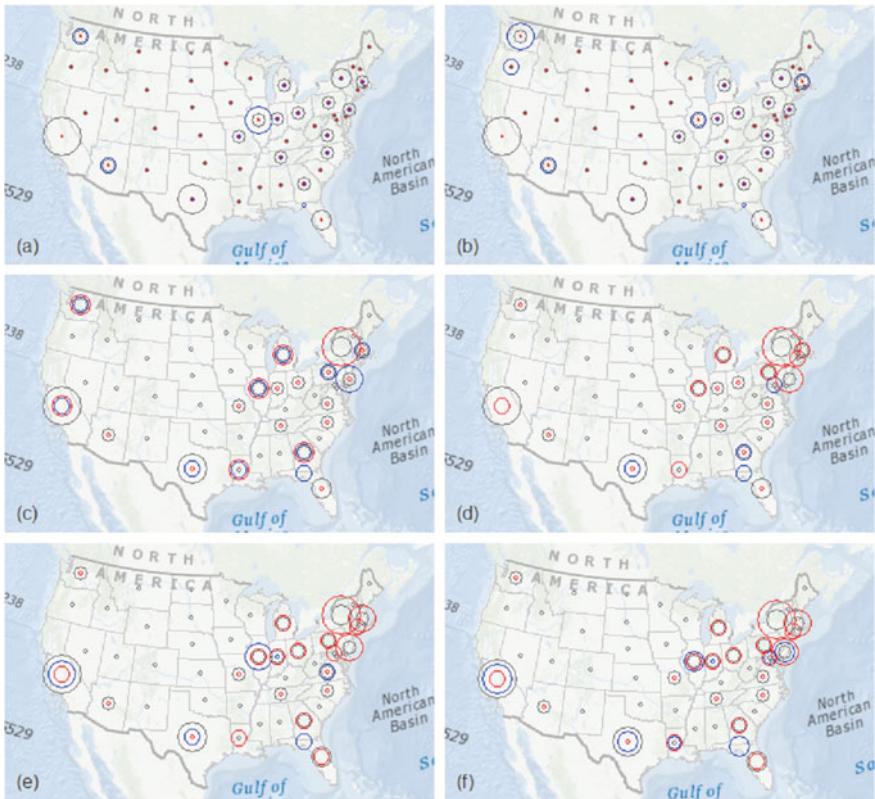


Fig. 2.4 (Color online) The hierarchy of COVID-19 cases compared with that of population. Note Each state has three circles: gray for population, blue for infection, and red for death. Panels (a)–(f) show the status of the pandemic in January, February, March, April, May, and June respectively at the state level. For the county level, please refer to the animation map at <http://lifegis.hig.se/COVID19/>

State. These are the new insights that are developed from the state level. The same insights can be seen at the county level, and the reader can refer to or further explore the animation map as cited in the note of Fig. 2.4.

2.4 Implication

This study has an important implication for power-law-related studies. The distributions of many natural and societal phenomena follow a power law over a wide range of magnitude, which has been extensively studied in a variety of scientific fields, such as physics, biology, economics, geography, demography, and social sciences (e.g., Bak 1996, Newman 2005). Surrounding a power law distribution and its variants

such as lognormal and exponential, an increasing number of research works have been made to illustrate what is the appropriate distribution for a real-world data. The first author of this paper has long developed an argument that a power law is an idealist status, when a complex system becomes mature or well-developed (Jiang and Yin 2014). Before the idealized status, the system is likely to show some deviation from a power law, thus a less-power-law distribution such as lognormal or a power law with an exponential cutoff. In this regard, it is better to use the ht-index to characterize the dynamic process or evolution of the system. This study proves that the ht-index is a good indicator, apparently a better one than the power law exponent, for characterizing the inherent hierarchy or heterogeneity of a complex system from a dynamic point of view.

2.5 Conclusion

In this paper, we have found that COVID-19 cases in the United States have developed over time from a less heterogeneous state to a more heterogeneous one, or equivalently from a very flat hierarchy to a very steep hierarchy, persistently approaching that status of the populations. Thus, the COVID-19 spatiotemporal patterns are largely shaped by the underlying population patterns, i.e., well-populated states or counties tend to have more people affected or died. While this finding may seem obvious, deviations from this overall trend help us see the particularities of the COVID-19 patterns at local scales. On the one hand, the spatial distribution of COVID-19 cases is persistently approaching a power law with cut-off, despite the implemented lockdown and social distance measures, indicating enormous spatial heterogeneity in terms of the distribution of COVID-19 cases. On the other hand, the observation that the ht-index of COVID-19 cases does not exceed that of population implies that lockdown and social distance measures do indeed have some effect; otherwise, the situation would become far more devastating than it is now. The power-law-based approach enables us to uncover these interesting patterns of COVID-19 cases, so opens a new way of mapping geographic phenomena. Our future work points to this direction.

2.6 Data Availability Statement

The data used and generated in this study are available at <https://doi.org/10.6084/m9.figshare.13295540>. The covid-19 data is based on the GitHub repository maintained by Johns Hopkins University (2020). It can be found at: https://github.com/CSSEGI/SandData/COVID-19/tree/master/csse_covid_19_data. Powerlaw calculations have been done with Aaron Clauset's MatLab code found at: <http://tuvalu.santafe.edu/~aaronc/powerlaws/>. The software used to spatially process and visualize the data is ArcGIS 10.7 by ESRI. Additionally head/tail breaks have been calculated with a

head/tail breaks calculator which can be found at: <https://github.com/ChrisdeRijke/HeadTailBreaksCalculator>

Acknowledgements This chapter is a reprint of the journal paper (Jiang and de Rijke 2021) with permission of the publisher Taylor & Francis. We would like to thank the anonymous referees for their valuable comments and Dr. Aaron Clauset for his insightful discussion. This paper was partially supported by the Swedish Research Council FORMAS through the ALEXANDER project with grant number FR-2017/0009.

References

- Bak, P. (1996). *How nature works: The science of self-organized criticality*. New York: Springer.
- Clauset, A., Shalizi, C. R., & Newman, M. E. J. (2009). Power-law distributions in empirical data. *SIAM Review*, 51, 661–703.
- ESRI (2020). *COVID-19 GIS Hub*, <https://coronavirus-disasterresponse.hub.arcgis.com/> https://github.com/CSSEGISandData/COVID-19/tree/master/csse_covid_19_data
- Jiang, B. (2013). Head/tail breaks: A new classification scheme for data with a heavy-tailed distribution. *The Professional Geographer*, 65(3), 482–494.
- Jiang, B. (2017). Geospatial analysis requires a different way of thinking: The problem of spatial heterogeneity. In *Trends in spatial analysis and modelling: Decision-support and planning strategies* (pp. 23–40). Berlin: Springer.
- Jiang, B., & de Rijke C. (2021). A power-law-based approach to mapping COVID-19 cases in the United States. *Geo-spatial Information Science*. <https://doi.org/10.1080/10095020.2020.1871306>
- Jiang, B., & Slocum, T. (2020). A map is a living structure with the recurring notion of far more smalls than larges. *ISPRS International Journal of Geo-Information*, 9(6), 388. Reprinted as the cover story in the magazine *Coordinates*, August issue, 6–17, 2020.
- Jiang, B., & Yin, J. (2014). Ht-index for quantifying the fractal or scaling structure of geographic features. *Annals of the Association of American Geographers*, 104(3), 530–541.
- Johns Hopkins University. (2020). *COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University*,
- Koch, R. (1998). *The 80/20 Principle: The secret of achieving more with less*. New York: DOUBLEDAY.
- Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, 46(5), 323–351.
- Surowiecki, J. (2004). *The wisdom of crowds: Why the many are smarter than the few*. London: ABACUS.
- Zipf, G. K. (1949). *Human behavior and the principles of least effort*. Cambridge, MA: Addison Wesley.

Dr. Bin Jiang is professor of computational geography at Faculty of Engineering and Sustainable Development (Division of GIScience) of the University of Gävle, Sweden. His research interests center on geospatial analysis of urban structure and dynamics, e.g., topological analysis, and scaling hierarchy applied to buildings, streets, and cities, or geospatial big data in general. Inspired by Christopher Alexander's work, he developed a mathematical model of beauty—beautimeter, which helps address not only why a structure is beautiful, but also how beauty the structure is.

Mr. Chris de Rijke is a research assistant at Faculty of Engineering and Sustainable Development (Division of GIScience) of the University of Gävle, Sweden. He holds bachelor and master's degrees in earth sciences and economics, and recently a master's degree in GIS. He has been researching on living structure and topological analysis supported by the novel concepts of natural cities and natural streets using big data such as OpenStreetMap data, nighttime imagery, and social media data.

Chapter 3

Individual, Context, and Space: Using Spatial Approaches for Understanding Unequal Social and Psychological Fallout of COVID-19



Grant Drawve, Casey T. Harris, and Kevin M. Fitzpatrick

3.1 Introduction

The tracing of the novel coronavirus (SARS-CoV2 or COVID19) started in late December of 2019 in Wuhan, China with the first confirmed case emerging in the United States on January 20, 2020 (Hauck et al. 2020; John Hopkins University 2020). While early media and public discourse described the country as low risk for an outbreak, the Trump administration declared COVID-19 a public health emergency in the United States on January 31, 2020 following a similar announcement from the World Health Organization the day before. Within the first few days of March, the United States had surpassed 100 cases and over 1,000 confirmed cases by the second week of March (Dong et al. 2020).

Unsurprisingly, fear of COVID-19 spread quickly across the United States, tracking closely with the rise in confirmed cases. Newspaper headlines began issuing stark warnings to “prepare for coronavirus fears” (Dawson and Puente 2020). Because little was known about its effects on short- and long-term health or the overall susceptibility of specific individuals at the time, fear of the virus during the earliest months centered on its health consequences. This fear quickly expanded to include generalized fear and concern about the social, economic, and psychological antecedents of the pandemic. Responding to the growing scope of the public health crisis, social scientists looked to untangle how the pandemic impacted not only physical health, but mental health, institutional engagement, family wellbeing, food systems and food insecurity, and economic conditions of the people and places touched by the virus’s spread.

Like the diffusion of the virus itself, the social impact of COVID-19 has rippled across United States communities and population groups in very uneven ways. Stay

G. Drawve (✉) · C. T. Harris · K. M. Fitzpatrick
Department of Sociology & Criminology, University of Arkansas, Fayetteville, AR, USA
e-mail: drawve@uark.edu

at home orders, physical distancing mandates, and the closures of schools, businesses, and places of worship impacted communities on the East and West coasts early (Mervosh et al. 2020). Other communities throughout the heartland and South remained largely unrestricted through mid-April, despite the continued spread of the virus (Kaiser Family Foundation 2020). Regardless of context, minority groups quickly emerged as high risk for serious infection and death (Yancy 2020), while pandemic-induced job loss impacted Black and Hispanic females more acutely than males or even White females (Gezici and Ozay 2020). Like natural disaster crises throughout history—including hurricanes, famines, and floods—the COVID-19 pandemic has entrenched inequalities that social scientists are now looking to empirically describe, explain, and mitigate.

3.1.1 *Silos in COVID-19 Social Science Research*

The explosion of social scientific research has generally fallen into three groups, each marking a methodological lens through which to view the impact of the pandemic: individual, contextual, and spatial. Yet, these trajectories remain largely “siloled” with only limited integration across each. First, a large number of studies have sprung up to explore differences across *individuals* at risk of COVID-19 infection (Chou et al. 2020; Xu et al. 2020; Zheng et al. 2020), as well as an enhanced vulnerability to the social and psychological outcomes that have developed as the pandemic has unfolded. Many of these early studies reveal that individuals have heightened levels of distress to the virus’s spread (Dawson and Golijani-Moghaddam 2020), increased depression (Fitzpatrick et al. 2020c), elevated suicidality (Fitzpatrick et al. 2020b), heightened fear (Fitzpatrick et al. 2020a; Presti et al. 2020), and growing food insecurity (Wolfson and Leung 2020). These conditions appear more acutely among the most socially vulnerable, including minorities, females, families with children present, and those who are unemployed or have been laid off (Fitzpatrick et al. 2020c). Yet, missing from many of these studies is consideration for the types of places and communities in which vulnerable individuals live. Indeed, those who are already food insecure or under-employed may experience different outcomes during the coronavirus pandemic when they live in communities with stronger/weaker food supply chains or local economies, respectively.

Building on such observations, a second area of social science research has taken a *contextual* approach. Rather than emphasizing individuals and their personal vulnerabilities and resources, such research focuses instead on the milieu effects of social and health vulnerabilities across places. Though notably smaller than the individual-level studies, this line of inquiry finds that some communities are more susceptible to the spread of the virus itself, as well as to the pandemic’s social repercussions (e.g., economic upheaval, mental health problems, hunger, housing turnover) than others. For example, more densely populated, aging, and poverty-stricken communities appear to be at greater risk for coronavirus infection (Zhang and Schwartz 2020), while places with greater poverty, minority population size, and segregation

have higher death rates, net of underlying population health risks (Kim and Bostwick 2020). This contextual approach is perhaps best illustrated by the tailoring of the Centers for Disease Control and Prevention (CDC) Social Vulnerability Index (SVI) into a COVID-19 Community Vulnerability Index (CCVI), which quantifies the vulnerability of counties for COVID-19 based upon socio-demographic and healthcare system factors (Surgo Foundation 2020). Still, other studies find that the secondary effects of the pandemic, including elevated rates of suicide (Riblet et al. 2020) and increased risk of domestic violence (Bullinger et al. 2020), are determined at least in part by the underlying structural resources, population and housing composition, and demographic characteristics.

Finally, third, a smaller number of social scientists have used a *spatial* lens, looking not only at individuals or the types of communities in which they live, but also how those people and places are connected across geographic space over time. Building on the convergence of people in time and space necessary for the transmission of the virus itself (Smith and Mennis 2020), a recent review notes the importance of spatial-temporal analysis, health and social geography, environmental, and web-based mapping for understanding the social fallout (Franch-Pardo et al. 2020). For example, communities more socially connected to early pandemic hotspots have seen greater case counts even when controlling for physical distance between regions and demographic characteristics. Additionally, using GPS location data, COVID-19 data, and population characteristics at the county-level, researchers have found social interaction and personal mobility declines with rising local infection rates and government-issued stay-at-home order. However, counties with greater elderly populations, a lower share of Republicans, and those more densely populated, appear to be more responsive to control efforts (Engle et al. 2020). Even the adoption of coronavirus mitigation policies appears to be driven in part by considerations of what is happening in neighboring geographic areas, both in the U.S. and abroad (Sebhaut et al. 2020).

3.1.2 *Spatial Approaches in Practice*

Critically, the lack of methodological integration that has dominated social science research so far marks a prominent shortcoming for policymakers and other stakeholders seeking to alleviate the burden of COVID-19 on individuals and communities throughout the United States. The true manifestation of the pandemic on the social lives of the general population involves a complex web of micro- and macro-processes. Individuals and their communities exist not as islands but are woven together across geographic space both physically through transportation networks and supply chains, as well as socially through kin and friendship networks, the sharing of social capital, information and general assistance sharing, institutions (e.g., schools, businesses, religious and civic organizations), and local and regional policies. Addressing who is most vulnerable to the economic, health, familial, or social-psychological fallout from the coronavirus pandemic requires the integration

of spatial approaches. Early research has already leveraged spatial tools, including those developed specifically for the detection of COVID-19 hotspots (Desjardins et al. 2020; Hohl et al. 2020; Sun et al. 2020) and/or the allocation of response resources (Kang et al. 2020; Lakhani 2020). However, spatial approaches gain even more value when employed alongside the more traditional individual and contextual data that have dominated sociological research thus far.

As an illustration of spatial approaches and their value, we draw from our own study of the diffusion of fear and its corollary social and psychological outcomes conducted in March 2020. Supported by the National Science Foundation (SES#2027148) and using a national opt-in Qualtrics panel, we collected a complete-data sample of 10,368 respondents that measure responses to the pandemic, including overall and specific COVID-19 fear and worry, as well as depression, anxiety, suicide ideation, trust, religiosity, social ties, food insecurity, and a host of other social and behavioral outcomes. This project sought to understand the response of individuals within their respective communities, the clustering of social problems across geographic space, and the response of individuals to the geographic distribution of COVID-19 itself. To do so, we geo-located all individuals to their respective residential counties at the time of response. Overall, this post-stratification weighted sample of 10,368 respondents for whom we have data, reside in 1,644 different counties (representing 52% of all U.S. counties).

We acknowledge that the context of the United States remains unique compared to other countries experiencing the COVID-19 pandemic. The United States is comprised of 50 states with 3,142 county subunits. Our own focus is on the 48 continental states, excluding Alaska and Hawaii. The United States has a long history of spatial disparities on socio-demographic characteristics. Figures 3.1, 3.2, and 3.3 highlight a small number of those disparities as obtained through Social Explorer from the American Community Survey 5-year estimates (2014-2018). Each state and county has a unique ID allowing for multiple data joins, including spatial. Each survey respondents reported their zip code, which was used with the U.S Department

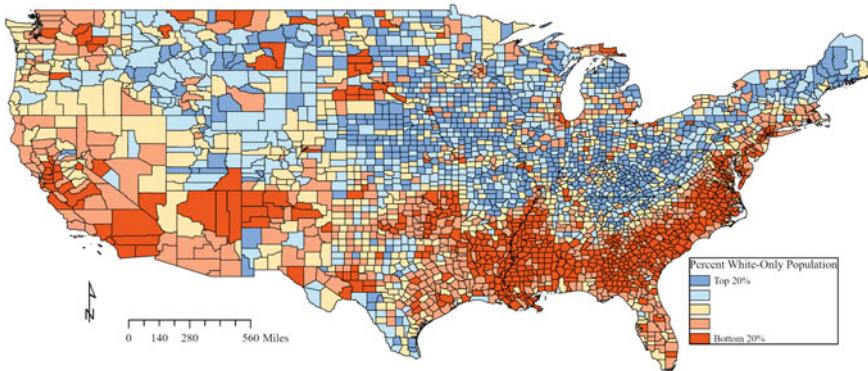


Fig. 3.1 Percent White-Only

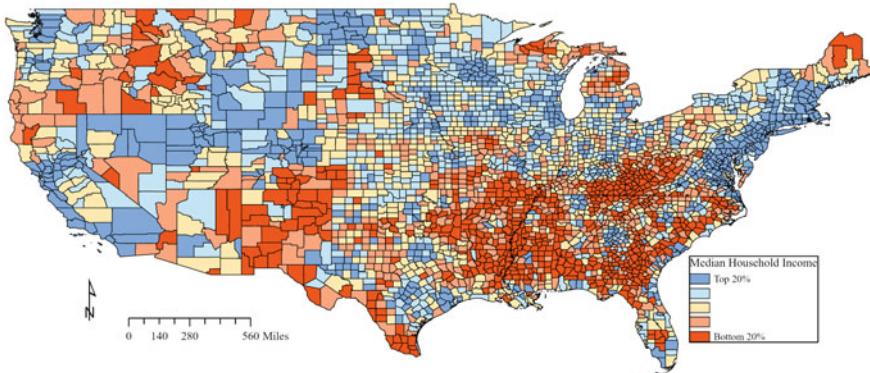


Fig. 3.2 Median Household Income

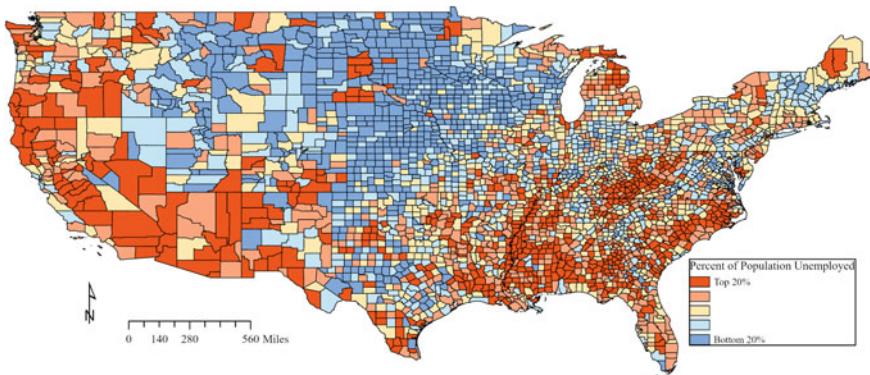


Fig. 3.3 Percent Unemployed

of Housing and Urban Development crosswalk file to assign respondent zip codes to a specific county (see Wilson and Alexander 2018).

As seen in Fig. 3.1, the east coast and southern and western states have an identifiably higher non-White population concentration as compared to the Midwest and Northeast, while Fig. 3.2 shows considerable spatial variation in median household income with the northeastern coastal region containing those counties in the uppermost quintile. Meanwhile, many of the southern counties remain both more diverse and plagued by lower median household income, while Fig. 3.3 shows that these same counties and states (along with those in the West region) contain higher levels of civilian unemployment. To further underscore the importance of how place matters, we also map the Surgo Foundation adapted the CDC Social Vulnerability Index for COVID-19, known as the COVID-19 Community Vulnerability Index (CCVI), in Fig. 3.4. As is clear from this map, some counties are at greater risk than others for COVID-19 than others: counties throughout the South, West, and parts of the

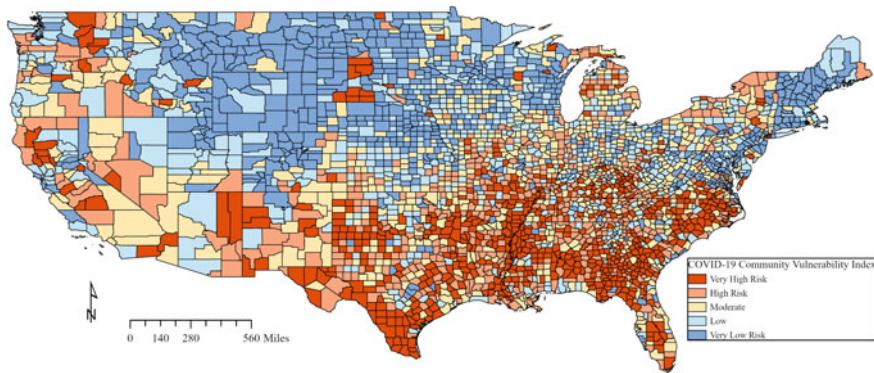


Fig. 3.4 COVID-19 Community Vulnerability Index

Midwest are at “very high” risk compared to those in the Northeast or Great Plains regions. This tailored metric includes a number of measures related to socioeconomic status, household composition and disability, minority status and language, housing type and transportation, epidemiologic factors, and healthcare system factors.

Turning to our own data, we note that the mid-to-late March study period remains particularly consequential for a number of reasons. Namely, the social responses of individuals within their respective communities and across space impacted everything from the adoption of personal mitigation strategies to support for public health policies that would ultimately determine the spread of the virus itself. Likewise, how individuals fared socially, economically, and psychologically during the earliest stages of the pandemic has downstream impacts on social problems that manifest themselves over longer periods of time. By the time of funding award and data collection, the number of confirmed virus cases had climbed from 42,469 to 155,947, an increase of over 350% in less than one month. Yet, the timing of the survey early in the pandemic marks an important phase in which the country had not yet become saturated by the virus or fully implemented lockdown restrictions in a near-universal manner. In turn, the responses from the survey reveal more variation than might be observed at the time of this writing, where confirmed cases exceed 11.2 million and nearly 250,000 deaths (Obtained November 17, 2020). Figure 3.5 below provides a visual illustration of the counties with confirmed COVID-19 cases (Dong et al. 2020) and the location of survey respondents from our data collection.

As is clear from the figure, there are well-defined clusters of confirmed COVID-19 cases across the nation, while Fig. 3.5 also shows that our survey respondents live varying distances from key outbreak sites (e.g., New York City, Washington state, Louisiana). In turn, though not shown here, local indicators of spatial autocorrelation (LISA) constructed using ArcGIS Pro show statistically significant spatial association for coronavirus cases, as would be expected given the diffusion of the virus early in the pandemic from select cities and metropolitan areas. LISA is an approach to identify spatial clusters with similar values of bordering neighbors (queen contiguity for

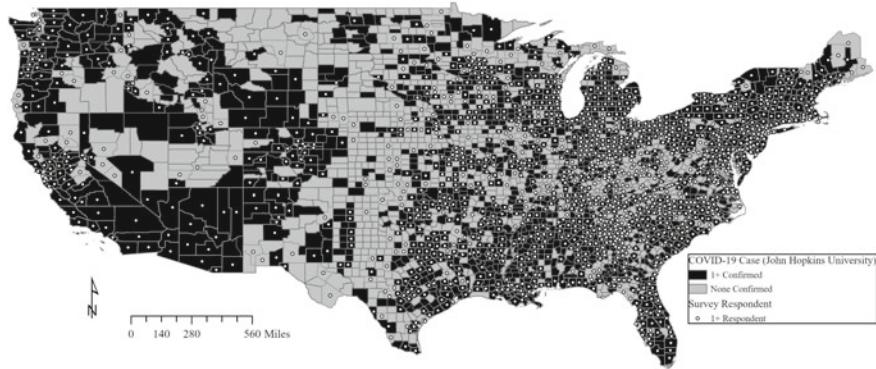


Fig. 3.5 Survey Respondents & Confirmed COVID-19 Cases

confirmed cases; distance method for survey data) or outliers based on neighboring values. In this case, an analysis of fear reveals counties with surrounding high-count neighbors (high-high clusters) throughout the New York metropolitan area and along the West Coast, as well as low case-count areas (low-low clusters) throughout the central United States and Great Plains regions. Again, our survey data contain critical variation in the spatial proximity or exposure to COVID-19 at the time of our data collection, allowing us to examine how differential exposure to the virus impacts social behavior in places with more or less viral prevalence.

Given the spatial clustering of COVID-19 cases within the United States at this point in time, a series of subsequent questions follow. What are the social implications of the novel coronavirus pandemic? Do the social problems most associated with this public health crisis cluster in space, too? How do individual, contextual, and spatial vulnerabilities combine to shape the social lives of U.S. residents during the pandemic? We turn now to an examination of several social outcomes in order to illustrate how spatial approaches provide for more meaningful responses to these types of questions. In doing so, we explicitly recognize (a) the interrelationship of the communities in which respondents reside, as well as (b) sociological and psychological responses and how they differ as a function of geo-spatial processes (e.g., proximity to confirmed cases and deaths, local density of deaths). We bring to bear a number of spatial tools to illustrate how GIS can be implemented within social science research, primarily through programs like ArcGIS Pro, GeoDa (Anselin et al. 2006), and Stata as described in specific analyses below. Importantly, while we illustrate the value of several specific spatial tools (e.g., LISA maps), nearly every aspect of the current project utilizes a range of geo-spatial techniques for advancing social science inquiry, including spatial joins, boundary clipping, attribute joins, boundary crosswalk (Zip to County), neighbor connectivity, and many other tools that are employed in data preparation and cleaning.

3.2 A Conceptual Framework

3.2.1 Fear and Concern for COVID-19

Fear underlies many social behaviors during a crisis, particularly one in which the boundaries of risk are ill-defined or unknown as has been the case early for the COVID-19 pandemic. As an example, during a natural disaster like a hurricane or flood, fear of damage to one's property or the risk to one's life affects whether someone seeks shelter, evacuates, or takes necessary steps to mitigate risk (Espinola et al. 2016). Thus, assessing threat and expressing concern for an event—including the current health crisis—remains an important predictor of social action. Fear and concern hold implications for downstream socio-emotional outcomes. Those who have more fear or concern may experience greater anxiety or other mental health problems that have both immediate and long-term implications (Laurance 2003). Heightened fear and concern may even drive individuals to change their perception of community institutions as they grapple with how to manage the resources those institutions provide for mitigating risk (Sweeney et al. 2015).

The application of spatial statistics in our own data reveal that individuals' fear clusters in space. While there are a number of strategies used to assess generalized fear and anxiety in individuals (Kogan and Edelstein 2004), our interest has been in giving as little guidance as possible to the respondent as to how they should think about or frame COVID-19 fear. To that end, we asked respondents to numerically rate on a sliding scale of zero to ten (0–10): how would you currently rate your fear of COVID-19? The local indicators of spatial autocorrelation (LISA) map for responses in the across our respondent counties are presented in Fig. 3.6 below as a visualization of key geographic patterns for overall coronavirus fear. The LISA analysis employed a spatial weight based on distance (200 miles; 999 permutations). The distance weight was used rather than the contiguity weight since we do not have

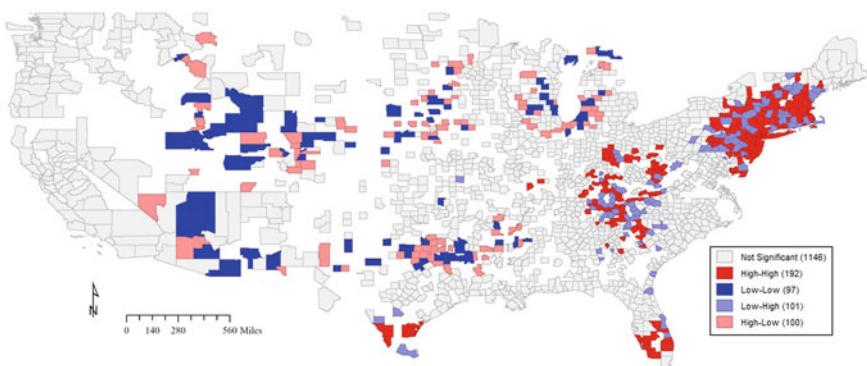


Fig. 3.6 LISA Fear of COVID-19

respondents from all US counties, leaving gaps and county islands with no adjoining neighbors.

Like the virus itself, *survey respondents have more comparable levels of fear to residents of neighboring communities than those further away*. For example, Fig. 3.6 illustrates that fear clusters positively (high-high) in Manhattan and the surrounding boroughs of New York City, while radiating outward into New Jersey, Pennsylvania, and other parts of the upper Atlantic region. In contrast, COVID-19 fear is not clustered in space to the same degree in the areas just beyond, where individuals living in Western Pennsylvania, Maine, or Maryland express less geographically consistent levels of worry or concern for COVID-19 on our overall scale. Additionally, as seen across Fig. 3.6, many communities were not significant. Interesting though, Florida and Texas have High-High counties for Fear of COVID-19 that could reflect tourist destinations still operating during our survey in Florida and fear along the southern US—Mexico border based on border crossings. Indeed, weeks after administering our survey (April 22), President Trump issued a proclamation suspending and limiting entry of immigrants into the United States, which could help contextualize the fear clusters in southern Texas.

Not surprisingly, individual, subjective feelings of fear appear to cluster in the same locations as the virus itself. To further explore that, we estimated hierarchical linear regression models predicting individual fear as a function of basic individual characteristics (gender, race/ethnicity, age), contextual features (county rates of poverty and unemployment, multi-group measure of racial/ethnic diversity), and the average coronavirus case rate of each respondent's immediate county contiguous neighbors. The latter were constructed in GeoDa and then used in Stata as part of the mixed-effects modeling package. GeoDa is useful for identifying the connectivity of each county and the average overall neighbors. The purpose here is to address the persistent gap in prior social scientific research on coronavirus by exploring the unique and simultaneous effects of individual, contextual, and spatial vulnerabilities on social phenomenon like fear.

The results of these models reveal that *individuals living in counties surrounded by more per capita COVID-19 cases or more coronavirus deaths express statistically significantly higher levels of fear*. Even adding in additional contextual characteristics of the communities that people live in (i.e., poverty, unemployment, and racial/ethnic diversity), the same result is produced. In short, proximity to COVID-19 cases and fatalities elevates individual fear net of other fear-inducing vulnerabilities among individuals and their communities.

3.2.2 Mental Health: Depression and Suicide Ideation

Fear is only one social outcome of COVID-19, however. For instance, Abel and McQueen (2020) note in their editorial that spatial distance remains distinct from social distance, despite the fact that the two phenomena are conflated in the rhetoric of the general public and policymakers alike. Since much of the public has been

forced to abruptly change their routines and regular interactions, the combination of both forms of distancing produces mental health problems, including depression and anxiety. Google search term trends bear out the criticality of distancing as a potential precursor to mental health symptoms at the time of our survey especially. Searching for the specific terms “social distancing” peaked in late March through early April, particularly between the beginning and end of our survey period.

As COVID-19 has become more prevalent throughout the United States, greater attention continues to be drawn to the mental health consequences of the virus. For many, the sudden shift in individuals’ routines has resulted in feelings of social isolation, anxiety, and depression. To add to the burden, most schools, churches, and businesses have closed or reduced interaction spaces, leaving many Americans struggling to make sense of the new normal or with opportunities to utilize mental health coping mechanisms (see Holmes et al. 2020). Reger and colleagues (2020), for example, describe the COVID-19 pandemic as the perfect storm for mental health, ripe for elevated risk of suicide (Reger et al. 2020).

Our own data bear out these more general anecdotes, revealing that the current public health crisis extends to socio-emotional outcomes, as well. In particular, we find evidence of heightened depressive symptomatology, anxiety (Fitzpatrick et al. 2020a), and suicidality (Fitzpatrick et al. 2020b). Specifically, the average depressive symptom score in our sample of 10,368 respondents is above the clinical caseness level and more than 25% of respondents report moderate to severe anxiety levels, far more prevalence than observed during other recent periods of time. Likewise, nearly 15% of sampled respondents are categorized as high risk for suicide. Such findings are consistent with a systematic review and meta-analysis by Salari and colleagues (2020) that examines a wide range of stress, anxiety, and depression outcomes related to COVID-19 (Salari et al. 2020).

Critically, such mental health outcomes are driven not only by individual vulnerabilities, like marital status, race, gender, and socioeconomic status (e.g., income), but by community contextual factors and the geo-spatial relationships between those communities, as well (Fitzpatrick and LaGroy 2011). Below, we illustrate this in our own data. Figure 3.7 displays a LISA map for depressive symptomatology using the Center for Epidemiological Studies for Depression (CES-D) scale. Though not shown, similar maps for anxiety symptoms measured using the General Anxiety Disorder-7 (GAD-7) scale and suicide behavior measured through the Suicide Behaviors Questionnaire Revised (SBQ-R) reveal similar patterns.

Even early in the pandemic, before the virus had impacted most communities or entrenched itself for a long period of time, *individuals with elevated levels of depression, anxiety, and suicidality clustered in space*. For example, there is evidence of positive spatial autocorrelation with depressive symptoms among individuals living within mid-Atlantic and Appalachian counties, as well as throughout Oklahoma, both clusters near early outbreaks in the poultry and meatpacking industries. In contrast, negative spatial autocorrelation (Low-Low) is observed in Florida and the upper Midwest. Similar clustering patterns are observed for both anxiety symptoms and suicide ideation.

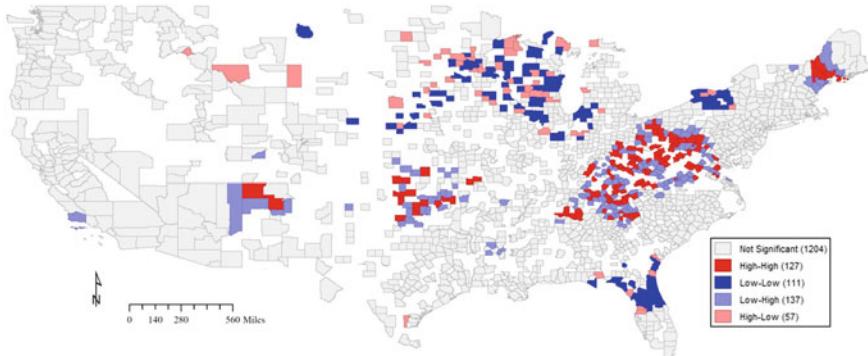


Fig. 3.7 LISA CES-Depression Scale

Not until leveraging spatial approaches does the impact of geography on mental health really emerge. As with our analysis of fear, we estimate a series of hierarchical linear and logistic regression models predicting individual depression, anxiety, and suicide as a function of individual characteristics (gender, race/ethnicity, age), county-level covariates (county rates of poverty and unemployment, multi-group measure of racial/ethnic diversity), and key spatial features (average coronavirus case and death rates among contiguous neighbors, distance to nearest case). Critically, and like our analysis of fear, *proximity to confirmed cases of the virus or neighboring death rates is linked to more elevated levels of depression, anxiety, and suicide ideation among respondents*. That is, living close to known cases and deaths results in greater individual mental health risk. This holds net of other individual risk factors and county-level characteristics that might also impact mental health, including poverty, unemployment, and racial/ethnic diversity.

3.2.3 Food Insecurity

Of course, the coronavirus pandemic has touched more than the socio-emotional lives of individuals living within the United States. As the virus spread across the United States, millions of people lost jobs as a result of lockdown orders and the closure of churches, schools, and other large institutions, further exacerbating mental health problems in a population that was increasingly under- and unemployed. As evidence, the Bureau of Labor Statistics notes that this was most prominent during the time of our own study with a spike in unemployment from March (4.4%) to April (14.7%). In turn, the rapid decline of the economy left many individuals and families vulnerable to additional health-related fallout, including food insecurity. Google search trends indicate searches for food banks and food pantries dramatically increased through March and April, suggesting the pandemic led to rapid and large-scale inability to secure basic nutritional needs.

Food insecurity plagues American communities even without a widespread health crisis. More disadvantaged and rural places tend to have reduced access to food as compared to their more affluent and urban counterparts (Hinrichs and Lyson 2007; Fitzpatrick and Willis 2015). In turn, food insecurity links to poorer health outcomes among those more personally vulnerable, especially if they are living in communities with fewer avenues for securing essential food (Leonard et al. 2018). The COVID-19 pandemic has exacerbated this inequality and, in turn, some public health scholars have noted that a domestic and international “hunger pandemic” may be on the immediate horizon (OXFAM International 2020). Where pre-pandemic estimates placed about 11–12% of US households as food insecure (Coleman-Jensen et al. 2019), substantial increases in food insecurity have already emerged (e.g., a 33% increase in Vermont between the end of March and early April) (Niles et al. 2020).

Because of the centrality of food systems and nutritional access in the daily lives of the United States population, our own study sought to understand how individual and community vulnerabilities impacted food insecurity during the pandemic (e.g., see Fitzpatrick et al. 2020). In doing so, we operate from the assumption that food insecurity remains concentrated in some locations more so than others. To measure local food shortages during the early pandemic, we asked respondents to identify (yes/no) whether they had recently observed food-related conditions in their local communities, including shortages of food at local shopping locations (e.g., grocery stores, large retailers, convenience stores, etc.).

Figure 3.8 shows the spatial distribution of persons observing food shortages at shopping locales during the mid-to-late March period during which our survey was deployed. As is clear from the figure, pockets of food shortages at local shopping establishments emerged early and unevenly. For example, shortages are positively correlated in space among individuals living in counties throughout the lower Appalachian and Southern regions, whereas individuals living in the upper Midwest (e.g., Minnesota, Wisconsin, Iowa) reported fewer shortages in their own counties just as others did in neighboring counties.

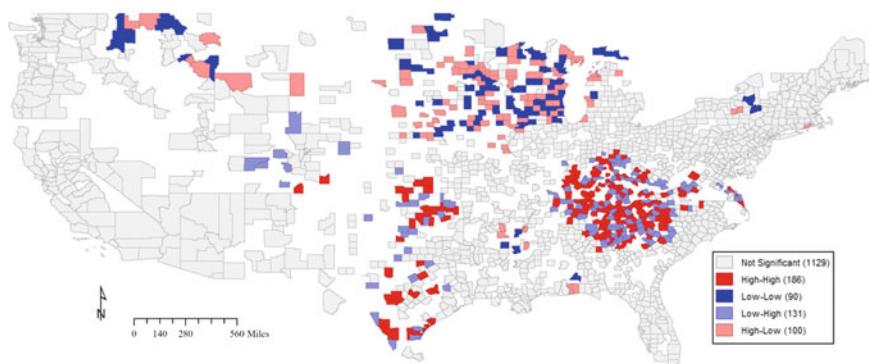


Fig. 3.8 Food Shortage LISA

The implications of such shortages on individual food insecurity are equally clear. The most vulnerable persons—those already at-risk for food insecurity—may be more acutely impacted if they are living within geographic areas of more disrupted food supply chains. Critically, this carries implications for social scientists needing to leverage spatial approaches because food shortages that cluster in space make solving the food insecurity puzzle increasingly more difficult: those who are in need of food must travel further to secure it, resulting in greater levels of food insecurity individually.

To that end, we map below food insecurity using the full 10-item USDA scale. Broadly, food insecurity clusters in the same kinds of places as those who have experienced local shortages of food and supplies, as described above. Taken together, two key social outcomes during the early COVID-19 pandemic were *spatially concentrated shortages of nutritional goods and personal difficulties securing food for oneself and family*. This comes as little surprise as the USDA has found that the areas our own data identify as having the most food-insecure families are among the highest regions for the percentage of food-unstable households (Coleman-Jensen et al. 2020) (Fig. 3.9).

Interestingly, there are communities that do not report being food insecure but experience food shortages. For instance, looking at Oklahoma, there were no statistically significant local clusters of food insecurity, though there were communities experiencing food shortages. In contrast, southern Alabama has high-high clusters of food insecurity but not clusters for experiencing food shortages. These patterns could reflect different buying behaviors but also the ability, or lack thereof, to purchase food during a pandemic. Given that our survey was administered early in the pandemic in the United States, future attention should focus on these community differences as the pandemic unfolded into winter months and supply chains (and buying behaviors) change.

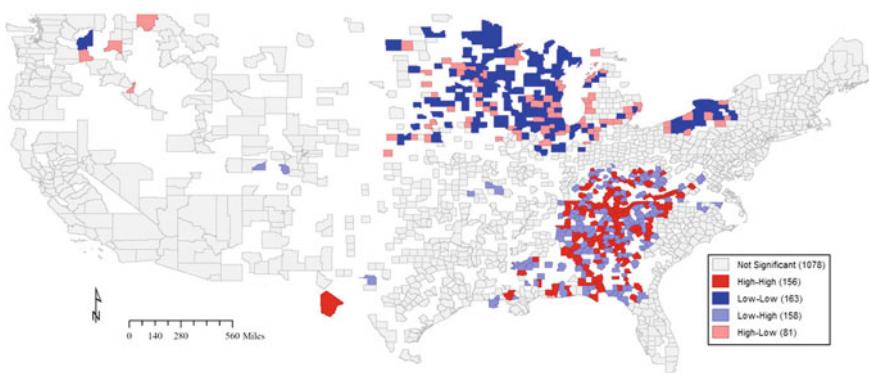


Fig. 3.9 Food Insecure LISA

3.2.4 Early Adoption of Personal Mitigation Strategies

Faced with a growing pandemic, many nations around the world have relied upon centralized public health policies to combat the spread of coronavirus (e.g., Pan et al. 2020). In contrast, the United States' response has instead devolved into a mosaic of state-level policies, many of which rely heavily on the adoption of personal mitigation strategies. Most notably, the Center for Disease Control (CDC) has outlined requirements to reopen America again (CDC 2020), emphasizing testing and contact tracing, handwashing, avoiding public transit and crowded events/areas, and utilizing remote interaction where possible.

Like other social and behavioral outcomes that require individual participation, the adoption of such mitigation strategies has been uneven. Some population groups and communities have embraced personal responsibility to a greater degree than others. Emerging research of survey data reveals disparities in the adoption of personal mitigation strategies by political identification, education, marital status, gender, and the presence of children (Atchison et al. 2020; Haischer et al. 2020; Pew Research Center 2020; Ritter and Brenan 2020). Others find that the political, socio-economic, and demographic composition of a community predicts the rate of mitigation adoption, as well (e.g., Henry-Nickie and Hudak 2020).

Our own survey of early COVID-19 responses also highlights disparities across individuals as a function of personal characteristics, community contextual factors, and spatial processes. To simplify for the current analysis, we focus on the binary (yes/no) responses to scenarios in which respondents indicated whether they were personally and currently (a) avoiding social events and gatherings (i.e., social distancing) or (b) staying home from work. The spatial distribution of responses is provided in Fig. 3.10 below.

Two patterns emerge clearly from Fig. 3.10. First, the adoption of social distancing was more likely to cluster together in counties throughout the New York-Pennsylvania area, whereas respondents living in other parts of the United States showed more spatial variability in their respective likelihood of avoiding public gatherings or crowds. Second, there is some similarity for the spatial concentration of staying home from work (e.g., throughout the New York area), but that important spatial clusters of increased likelihood exist throughout Ohio, where Governor Mike Dewine adopted aggressive stay-at-home orders early in the pandemic. The spatial concentration of low likelihood of staying home from work tends to occur in counties throughout Iowa and rural Indiana, as well as Texas and the Carolinas. As with the social outcomes described above, *the adoption of personal mitigation strategies are not random across counties in the United States* with respondents in some communities much more likely to follow health guidelines than those living in others.

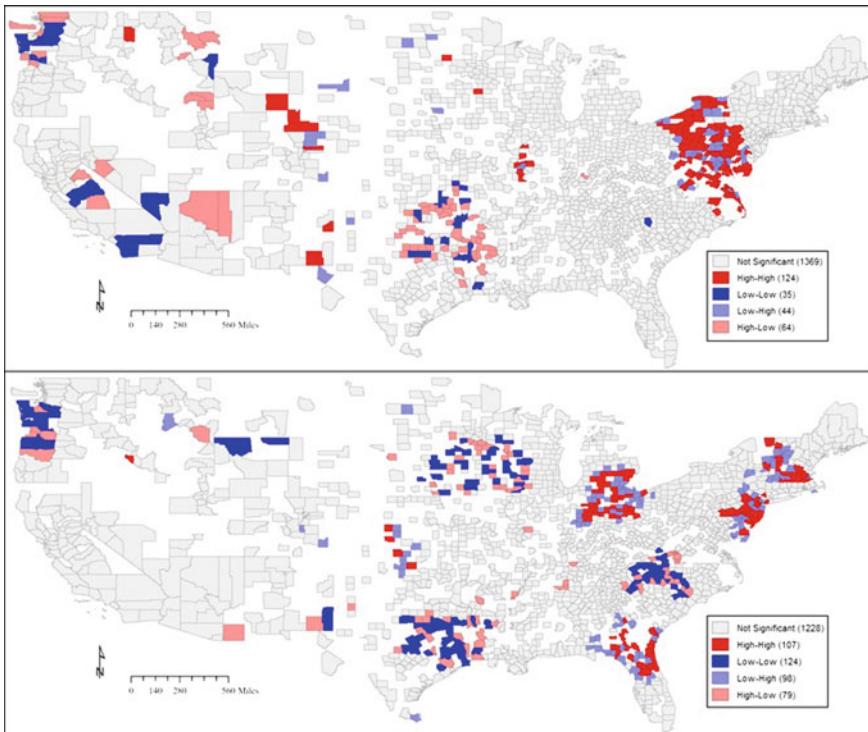


Fig. 3.10 Avoid Social Events & Staying at Home from Work LISA

3.3 Discussion

The spread of COVID-19 has and will continue to impact individuals and their communities for years to come. As with natural disasters and other public health crises in the past, the physical, emotional, social, and economic toll of the pandemic will fall disproportionately on some people and some locales more than others. Social scientists have already begun to describe such inequality, though the contributions of existing scholarship remain more limited than they need be. Much of the prior research emphasizes the personal or individual vulnerabilities of the most marginalized in our society—minorities, the lower class, those with young children, the marginally employed—and how COVID-19 has generated greater levels of fear, depression, or economic uncertainty for them. Others have pointed to the unequal impact of the coronavirus pandemic on specific communities, highlighting the ways that impoverished neighborhoods or aging populations create fertile ground for social upheaval amidst the uncertainty those places now face. Both areas of research constitute the bulk of extant empiricism.

Crucially, other methodological frameworks that emphasize *geo-spatial processes* remain under-developed, particularly as they work alongside and in conjunction

with individual and contextual mechanisms. That individuals and their communities are interconnected through the exact structures highlighted by social scientists (e.g., human and social capital networks, infrastructure and transportation, organizations and institutions) remains perhaps the most important contribution of the spatial approaches to social scientific research. Our goal here has been to illustrate the value of thinking about space and capturing, however simply, the role played by geographic processes (e.g., proximity, local density).

Table 3.1 provides a summary of social and psychological outcomes which our own NSF-funded study explored amidst the COVID-19 pandemic. Critically, our analyses nests 10,368 individuals within their respective counties, predicting a handful of social outcomes—fear, depression, anxiety, and suicidality—as a function of individual, contextual, and spatial characteristics. Regarding the latter, we focus especially on the role played by proximity to and density of COVID-19 cases and deaths, with some additional exploration of spatial distance and the descriptive spatial patterning of each social outcome itself. Generally, Table 3.1 reinforces that when individuals reside in counties surrounded by a greater average number of cases and/or deaths, those individuals report significantly elevated levels of fear, depression, anxiety, and suicide ideation.

Of note (though not shown), several of our covariates also had interesting relationships with the outcomes examined in Table 3.1. For example, we found consistent evidence of racial/ethnic disparities at the individual level, while female respondents were more likely to have elevated levels of fear and mental health symptomology than males. Additionally, county socio-economic characteristics and diversity appear to be less consequential for fear, but very impactful for mental health outcomes (Table 3.1).

Table 3.1 Summary of Coefficients from Hierarchical Linear and Logistic Regression Models Illustrating GIS-Generated Measures to Predict Individual-Level Social Outcomes

Social Outcome	Avg. Per Capita Cases (Neighbors)	Avg. Per Capita Deaths (Neighbors)
Subjective Fear (0–10)	+	+
Depression (CES-D)	+	NS
Anxiety (GAD-7)	+	+
Suicide Ideation (0–1)	NS	+

Note For parsimony, all models include controls for individual age in years, race (Black, Hispanic, Asian, Native American, Other), gender (female); county-level poverty, unemployment, and racial/ethnic diversity (multi-group entropy) (see Reardon and Firebaugh 2002)

+ indicates statistically significant positive relationship ($p < 0.05$)

– indicates statistically significant negative relationship ($p < 0.05$)

NS indicates non-significant relationship ($p < 0.05$)

Clearly, place matters, but how individuals respond to their local context is more complicated than considering only the characteristics of the places in which they live. Individuals living near each other and, critically, nearer to the virus itself, share more similarities than those living further away. As Tobler's first law of geography aptly notes, "everything is related to everything else, but near things are more related than distant things" (Tobler 1970, p. 236). This applies to the social responses to the COVID-19 pandemic as it does to many other physical and social phenomena.

At the same time, our analyses reveal important variation in the way that spatial processes work with regards to human behavior. Perhaps the best illustration of this relationship was the socio-emotional measures being strongly related to the spatial concentration of COVID-19 cases, whereas suicide ideation was not. Instead, the average neighboring death rate was more important. One possible reason for this might be that extreme mental health problems resulting in the contemplation of lethal self-harm only increases when the loss of life hits close to home—literally, in neighboring communities. This points to the need to more fully parse out how individuals perceive their own vulnerabilities, as well as those of their communities and neighboring locales. There remain fruitful avenues as spatial approaches are more fully integrated into the mainstream of social scientific work on the COVID-19 pandemic.

At the same time, we must reiterate the importance of our own study timeline. The survey from which we draw our data was administered in mid-to-late March when the virus was emerging unevenly across the United States and, in turn, differentially impacting the social lives of the population. COVID-19 is still very much a pandemic six months later. Interesting to note though, when measuring many of our survey outcomes, the sheer volume of news-media attention COVID-19 garnered could also be influencing how spatial relationships emerged. This provides an important snapshot into early social, mental, and behavioral outcomes, but our intent is not to take away from our current discussion or emphasis on the geography of the pandemic today. Broadly, our goal has been to argue that social outcomes—like the physical phenomena that form the bedrock of much geo-spatial analysis—that space and time are critical components of public health models seeking to bolster social support and improve individual and community resiliency.

3.3.1 Implications and Future Research

While our primary aim has been to illustrate the value of a geographic lens to understanding the sociological repercussions of the COVID-19 pandemic, this same spatial lens also points to unique policy and practice as we look toward recovery and resilience. In particular, attention must focus on the geography of inequality. For example, Sun and colleagues (2020) highlight the spatial clustering of coronavirus cases and deaths along coastlines and the historic, Southern "Black Belt," where communities were already vulnerable to poor health outcomes. Similarly, Novacek et al. (2020) provides a series of recommendations for mental health approaches to

vulnerable minorities, many of whom live in highly segregated and disadvantaged spatial contexts that look to offset systemic inequalities in mental health. Thus, it will be important moving forward, for service providers to be aware of such disparities when developing outreach and service provision plans throughout the United States. Even emerging clinical trials and subsequent plans for the distribution of a vaccine reflect concern for including the most individually, contextually, and spatially vulnerable people and communities (Johnson 2020).

Access to resources and the utilization of service will become increasingly important for responding to COVID-19 related social outcomes. The physical locations of different service providers and institutions are available through the Department of Homeland Security's data portal Homeland Infrastructure Foundation-Level Data, including for hospitals, urgent care facilities, universities, local law enforcement, and nursing homes. Taking into consideration the proximity to and density of care providers may have important implications for COVID-19 fear, mental health consequences, and even food access above and beyond the kinds of measures we were able to bring to bear in the current analysis.

Additionally, more research will be needed that further leverages the GIS toolkit to understand the social ramifications of COVID-19 that builds directly off of the work we have highlighted here. For instance, early research using the Center for Disease Control and Prevention's Social Vulnerability Index (SVI) has identified communities (counties) with greater underlying social vulnerability and their relative risk for confirmed cases and deaths. How do healthcare infrastructures affect long-term physical and mental health and, more importantly, how does the geographic concentration of weak infrastructure matter? How does the connectivity of communities across linguistic and cultural lines affect recovery from the current pandemic? The tailoring of the CDC SVI into a new COVID-19 community vulnerability index (CCVI) may help in this regard as social scientists continue to explore additional coronavirus-specific contextual and spatial characteristics alongside the individual vulnerabilities faced by specific sub-groups of the population. There may even be duration effects that more fully integrate time and space together: perhaps those individuals, families, and communities that are exposed to high case and deaths rates for longer periods of time and in greater spatial concentrations will be most at-risk for social and economic problems in the future (or recover more slowly once vaccination and prevention measures are in place).

3.4 Conclusion

Social scientists have made important strides in detailing the consequences of the COVID-19 pandemic above and beyond the physical illnesses and deaths that have grown exponentially since the start of the year. As Holmes and colleagues (2020) argue, more multidisciplinary research is greatly needed to fully understand the true impact of COVID-19. We believe that the utilization of geographical information systems and spatial analysis by its' nature emphasizes such synergy, avoiding the

more narrowly-focused “silos” that hamper efforts to fully understand the social problems that have come to define the middle and later stages of the COVID-19 pandemic.

References

- Abel, T., & McQueen, D. (2020). The COVID-19 pandemic calls for spatial distancing and social closeness: Not for social distancing! *International Journal of Public Health*, 65(3), 231. <https://doi.org/10.1007/s00038-020-01366-7>.
- Anselin, L., Syabri, I., & Kho, Y. (2006). GeoDa: An introduction to spatial data analysis. *Geographical Analysis*, 38(1), 5–22.
- Atchison, C. J., Bowman, L., Vrinten, C., Redd, R., Pristera, P., Eaton, J.W., & Ward, H. (2020). *Perceptions and behavioural responses of the general public during the COVID-19 pandemic: A cross-sectional survey of UK adults*. doi.org/<https://doi.org/10.1101/2020.04.01.20050039>
- Bullinger, L. R., Carr, J. B., & Packham, A. (2020). *COVID-19 and crime: Effects of stay-at-home orders on domestic violence*. The National Bureau of Economic Research: Doi.org/. <https://doi.org/10.3386/w27667>.
- Bureau of Labor Statistics. (2020). *Graphics for Economic News Releases: Civilian Unemployment Rate*. <https://www.bls.gov/charts/employment-situation/civilian-unemployment-rate.htm>.
- CDC. (2020, May). *CDC activities and initiatives supporting the COVID-19 response and the president's plan for Opening America up again*. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention.
- Chou, R., Dana, T., Buckley, D. I., Selph, S., & Fu, R. (2020). Epidemiology of and risk factors for coronavirus infection in health care workers: A living rapid review. *Annals of Internal Medicine*. <https://doi.org/10.7326/M20-1632>.
- Coleman-Jensen, A., Rabbit, M. P., Gregory, C. A., & Singh, A. (2019). *Household Food Security in the United States in 2018* (p. 270). Economic Research Service Report Number: United States Department of Agriculture.
- Dawson, D. L., & Golijani-Moghaddam, N. (2020). COVID-19: Psychological flexibility, coping, mental health, and wellbeing in the UK during the pandemic. *Journal of Contextual Behavioral Science*, 17, 126–134. <https://doi.org/10.1016/j.jcbs.2020.07.010>.
- Dawson, A., & Puente, M. (2020, March 6). *Tampa Bay's first-responders plan, prepare for coronavirus*. Tampa Bay Times. <https://www.tampabay.com/news/health/2020/03/07/tampa-bays-first-responders-plan-prepare-for-coronavirus-fears/>
- Department of Homeland Security. (2020). Homeland Infrastructure Foundation-Level Data (HIFLD). <https://hifld-geoplatform.opendata.arcgis.com/>
- Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, 118,. <https://doi.org/10.1016/j.apgeog.2020.102202>.
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet: Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Engle, S., Stromme, J., & Zhou, A. (2020). Staying at home: Mobility effects of COVID-19. Retrieved from SSRN: <https://ssrn.com/abstract=3565703>.
- Espinola, M., Shultz, J. M., Espinel, Z., Althouse, B. M., Cooper, J. L., Baingana, F., et al. (2016). Fear-related behaviors in situations of mass threat. *Disaster Health*, 3(4), 102–111. <https://doi.org/10.1080/21665044.2016.1263141>.
- Fitzpatrick, K. M., & LaGory, M. E. (2011). *Unhealthy cities: Poverty, race, and place in America*. New York: Routledge.
- Fitzpatrick, K. M., & Willis, D. (2015). *A place-based perspective of food in society*. New York, NY: Palgrave Macmillan.

- Fitzpatrick, K. M., Harris, C., & Drawve, G. (2020a). Fear of COVID-19 and the mental health consequences in America. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(S1), S17–S21. <https://doi.org/10.1037/tra0000924>.
- Fitzpatrick, K. M., Harris, C., & Drawve, G. (2020b). How bad is it? Suicidality in the middle of the COVID-19 pandemic. *Suicide and Life-Threatening Behavior*. <https://doi.org/10.1111/sltb.12655>.
- Fitzpatrick, K. M., Harris, C., & Drawve, G. (2020c). Living in the midst of fear: Depressive symptomatology among US adults during the COVID-19 pandemic. *Depression and Anxiety*. <https://doi.org/10.1002/da.23080>.
- Fitzpatrick, K. M., Harris, C., Drawve, G., & Willis, D. E. (2020d). Assessing food insecurity among US adults during the COVID-19 Pandemic. *Journal of Hunger & Environmental Nutrition*. <https://doi.org/10.1080/19320248.2020.1830221>.
- Franch-Pardo, I., Nappelano, B. M., Rosete-Verges, F., & Billa L. (2020). Spatial analysis and GIS in the study of COVID-19. A review. *Science of the Total Environment*, 739(15), 140033. doi.org/<https://doi.org/10.1016/j.scitotenv.2020.140033>
- Gezici, A., & Ozay, O. (2020). How race and gender shape COVID-19 unemployment probability. Gezici, Armagan and Ozay, Ozge, How Race and Gender Shape COVID-19 Unemployment Probability (July 16, 2020). Retrieved from SSRN: <https://ssrn.com/abstract=3675022>.
- Haischer, M. H., Beilfuss, R., Hart, M. R., Opielinski, L., Wrucke, D., Zirgaitis, G., Uhrich, T. D., & Hunter, S. K. (2020). *Who is wearing a mask? Gender, age-, and location-related differences during the COVID-19 pandemic*. medrxiv.org/content/<https://doi.org/10.1101/2020.07.13.20152736v3>
- Hauck, G., Gelles, K., Bravo, V., & Thorson, M. (2020, June 23). *Five months in: A timeline of how COVID-19 has unfolded in the US*. USA TODAY. <https://www.usatoday.com/in-depth/news/nation/2020/04/21/coronavirus-updates-how-covid-19-unfolded-u-s-timeline/2990956001/>
- Henry-Nickie, M., & Hudak, J. (2020, May 19). *Social distancing in black and white neighborhoods in Detroit: A data-drive look at vulnerable communities*. Brookings. <https://www.brookings.edu/blog/fixgov/2020/05/19/social-distancing-in-black-and-white-neighborhoods-in-detroit-a-data-driven-look-at-vulnerable-communities/>.
- Hinrichs, C. C., & Lyson, T. A. (2007). *Remaking the North American food system: Strategies for sustainability*. Lincoln: University of Nebraska Press.
- Hohl, A., Delmelle, E. M., Desjardins, M. R., & Lan, Y. (2020). Daily surveillance of COVID-19 using the prospective space-time scan statistic in the United States. *Spatial and Spatio-temporal Epidemiology*, 34.. <https://doi.org/10.1016/j.sste.2020.100354>.
- Homles, E. A., O'Connor, R., Perry, V. H., Tracey, I., Wessely, S., Arseneault, L., et al. (2020). Multi-disciplinary research priorities for the COVID-19 pandemic: A call of action for mental health science. *The Lancet Psychiatry*, 7(6), 547–560. [https://doi.org/10.1016/S2215-0366\(20\)30168-1](https://doi.org/10.1016/S2215-0366(20)30168-1).
- John Hopkins University. (2020, September). *Hubei Timeline*. <https://coronavirus.jhu.edu/data/hubei-timeline>.
- Johnson, C. Y. (2020, July 26). *A trial for coronavirus vaccine researchers: Making sure Black and Hispanic communities are included in studies*. The Washington Post. <https://www.washingtonpost.com/health/2020/07/26/trial-coronavirus-vaccine-researchers-making-sure-black-hispanic-communities-are-included-studies/>.
- Kaiser Family Foundation. (2020, April 9). *When state stay-at-home orders due to coronavirus went into effect*. <https://www.kff.org/other/slide/when-state-stay-at-home-orders-due-to-coronavirus-went-into-effect/>.
- Kang, J.-Y., Michels, A., Lyu, F., Wang, S., Agbodo, N., Freeman, V. L., et al. (2020). Rapidly measuring spatial accessibility of COVID-19 healthcare resources: A case study of Illinois. *USA. medRxiv preprint*. <https://doi.org/10.1101/2020.05.06.20093534>.
- Kim, S. J., & Bostwick, W. (2020). Social vulnerability and racial inequality in COVID-19 deaths in Chicago. *Health Education & Behavior*, 47(4), 509–513.

- Kogan, J. N., & Edelstein, B. A. (2004). Modification and psychometric examination of a self-report measure of fear in older adults. *Journal of Anxiety Disorders*, 18(3), 397–409. [https://doi.org/10.1016/S0887-6185\(02\)00260-8](https://doi.org/10.1016/S0887-6185(02)00260-8).
- Kuchler, T., Russel, D., & Stroebel, J. (2020). *The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook*. arXiv:2004.03055v2
- Lakhani, A. (2020). Which Melbourne metropolitan areas are vulnerable to COVID-19 based on age, disability, and access to health services? Using spatial analysis to identify service gaps and inform delivery. *Journal of Pain and Symptom Management*, 60(1), e41–e44. <https://doi.org/10.1016/j.jpainsympman.2020.03.041>.
- Laurance, J. (2003). *Pure Madness: How fear drives the mental health system*. New York, NY: Routledge.
- Leonard, T., Hughes, A. E., Donegan, C., Santillan, A., & Pruitt, S. L. (2018). Overlapping geographic clusters of food security and health: Where do social determinants and health outcomes converge in the U.S.? *SSM—Population Health*, 5, 160–170. doi.org/<https://doi.org/10.1016/j.ssmph.2018.06.006>
- Mervosh, S., Lu, D., & Swales, V. (2020, April 20). *See which states and cities have told residents to stay at home*. The New York Times. <https://www.nytimes.com/interactive/2020/us/coronavirus-stay-at-home-order.html>.
- Niles, M. T., Bertmann, F., Belarmino, E. H., Wentworth, T., Biehl, E., & Neff, R. A. (2020). The early food insecurity impacts of COVID-19. doi.org/<https://doi.org/10.1101/2020.05.09.20096412>
- Novacek, D. M., Hampton-Anderson, J. N., Ebor, M. T., Loeb, T. B., & Wyatt, G. E. (2020). Mental health ramifications of the COVID-19 pandemic for Black Americans: Clinical and research recommendations. *Psychological Trauma: Theory, Research, Practice, and Policy*, 12(5), 449–451.
- OXFAM International. (2020). *World on the brink of a ‘hunger pandemic’: Coronavirus threatens to push million into starvation*. <https://www.oxfam.org/en/world-brink-hunger-pandemic-corona-virus-threatens-push-millions-starvation>.
- Pan, A., Liu, L., Wang, C., Guo, H., Hao, X., Wang, Q., et al. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA Network*, 323(19), 1915–1923. <https://doi.org/10.1001/jama.2020.6130>.
- Pew Research Center. (2020, June 25). *Republicans, democrats move even further apart in Coronavirus concerns*. <https://www.pewresearch.org/politics/2020/06/25/republicans-democrats-move-even-further-apart-in-coronavirus-concerns/>.
- Reardon, S. F., & Firebaugh, G. (2002). Measures of multigroup segregation. *Sociological Methodology*, 32(1): doi.org/<https://doi.org/10.1111/1467-9531.00110>
- Reger, M. A., Stanley, I. H., & Joiner, T. E. (2020). Suicide mortality and coronavirus disease 2019—A perfect storm? *JAMA Psychiatry*. <https://doi.org/10.1001/jamapsychiatry.2020.1060>.
- Riblet, N. B., Stevens, S. S., Watts, B. V., & Shiner, B. (2020). A pandemic of body, mind, and spirit: The burden of “social distancing” in rural communities during an era of heightened suicide risk. *The Journal of Rural Health*. <https://doi.org/10.1111/jrh.12456>.
- Ritter, Z., & Brenan, M. (2020, May 13). *New April guidelines boost perceived efficacy of face masks*. GALLUP. <https://news.gallup.com/poll/310400/new-april-guidelines-boost-perceived-efficacy-face-masks.aspx>.
- Salari, N., Hosseiniyan-Far, A., Jalali, R., Vaisi-Raygani, A., Rasoulopoor, S., Mohammadi, M., Rasoulopoor, S., & Khaledi-Paveh, B. (2020). Prevalence of stress, anxiety, depression among the general population during the COVID-19 pandemic: A systematic review and meta-analysis. *Globalization and Health*, 16(57). doi.org/<https://doi.org/10.1186/s12992-020-00589-w>
- Sebhatau, A., Wennberg, K., Arora-Jonsson, S., & Lindberg, S. I. (2020). Explaining the homogeneous diffusion of COVID-19 nonpharmaceutical interventions across heterogeneous countries. *Proceedings of the National Academy of Sciences PNAS*. <https://doi.org/10.1073/pnas.2010625117>.

- Smith, C. D., & Mennis, J. (2020). Incorporating geographic information science and technology in response to the COVID-19 pandemic. *Preventing Chronic Disease, 17*, E58. <https://doi.org/10.5888/pcd17.200246>.
- Sun, F., Matthews, S. A., Yang, T-C., & Hu, M-H. (2020). A spatial analysis of the COVID-19 period prevalence in the U.S. counties through June 28, 2020: Where geography matters? *Annals of Epidemiology*. doi.org/<https://doi.org/10.1016/j.anepidem.2020.07.014>
- Surgo Foundation. (2020). *The COVID-19 Community Vulnerability Index (CCVI)*. <https://precisionforcovid.org/ccvi>.
- Sweeney, A., Gillard, S., Wykes, T., & Rose, D. (2015). The role of fear in mental health service users' experiences: A qualitative exploration. *Social Psychiatry and Psychiatric Epidemiology, 50*, 1079–1087.
- Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography, 46*, 234–240.
- Wilson, R., & Din, A. (2018). Understanding and Enhancing the U.S. Department of Housing and Urban Development's ZIP Code Crosswalk Files. *Cityscape: A Journal of Policy Development and Research, 20* (2): 277–294. <https://www.huduser.gov/portal/periodicals/cityscape/vol20num2/ch16.pdf>.
- Xu, L., Mao, Y., & Chen, G. (2020). Risk factors for 2019 novel coronavirus disease (COVID-19) patients progressing to critical illness: A systematic review and meta-analysis. *Aging, 12*(12), 12410–12421. <https://doi.org/10.18632/aging.103383>.
- Yancy, C. W. (2020). COVID-19 and African Americans. *JAMA, 323*(19), 1891–1892. <https://doi.org/10.1001/jama.2020.6548>.
- Zhang, C. H., & Schwartz, G. G. (2020). Spatial disparities in coronavirus incidence and mortality in the United States: An ecological analysis as of May 2020. *The Journal of Rural Health, 36*, 433–445. <https://doi.org/10.1111/jrh.12476>.
- Zheng, Z., Peng, F., Xu, B., Zhao, J., Liu, H., Peng, J., et al. (2020). Risk factors of critical & mortal COVID-19 cases: A systematic literature review and meta-analysis. *Journal of Infection, 81*(2), e16–e25.

Chapter 4

A Permanent Virtual Memorial for a Whistleblower of the COVID-19 Pandemic: A Case Study of Crypto Place on the Blockchain



Xu Huang and Bo Zhao

4.1 Introduction

Dr. Li Wenliang, one of the whistleblowers who shared information of the initial COVID-19 outbreak in China with the world, passed away at the age of 34 from the coronavirus in the early morning of February 7, 2020. Upon hearing of his death, a few mourners placed flowers at the front gate of Wuhan Central Hospital (Feng and Cheng 2020). In addition to this physical, impromptu and fleeting commemoration, a variety of online and offline memorials have been generated. Among them, a virtual memorial was created by an anonymous mourner as a smart contract on the Ethereum platform. By decrypting the input data of this newly created smart contract to utf-8 text, we can see an ASCII-code-formed image of memorial (see Fig. 4.1).¹ Unlike a physical memorial, this encrypted monument did not require any complicated approval process by authorities. The anonymous founder chose this virtual setting possibly in defiance of the perceived censorship of China's Internet, as early online news about Dr. Li was widely censored (Zhao 2020).

Such a monument would not only be accessible to the public on the Ethereum blockchain but be unaffected by natural disasters or authoritarian censorship. From this regard, it represents the emergence of a new type of place—“crypto place,” which is briefly described in our recently published paper (Zhao and Huang 2020). This chapter includes a detailed discussion of the initial concerns and analyses regarding

¹ Any Internet user can visit an instance of this smart contact on Etherscan. The URL for the described smart contact is <https://etherscan.io/tx/0xb16c93d6f51ee7a68d626be22add24455db0a4644c68da69af6705981c77820e>. Once this URL is opened, a user can browse the ASCII code formed tombstone image by pressing the “View Input as UTF-8” button.

X. Huang

School of Geography, Nanjing Normal University, Nanjing 210046, China

B. Zhao (✉)

Department of Geography, University of Washington, Seattle, WA 98195, USA

e-mail: zhaobo@uw.edu

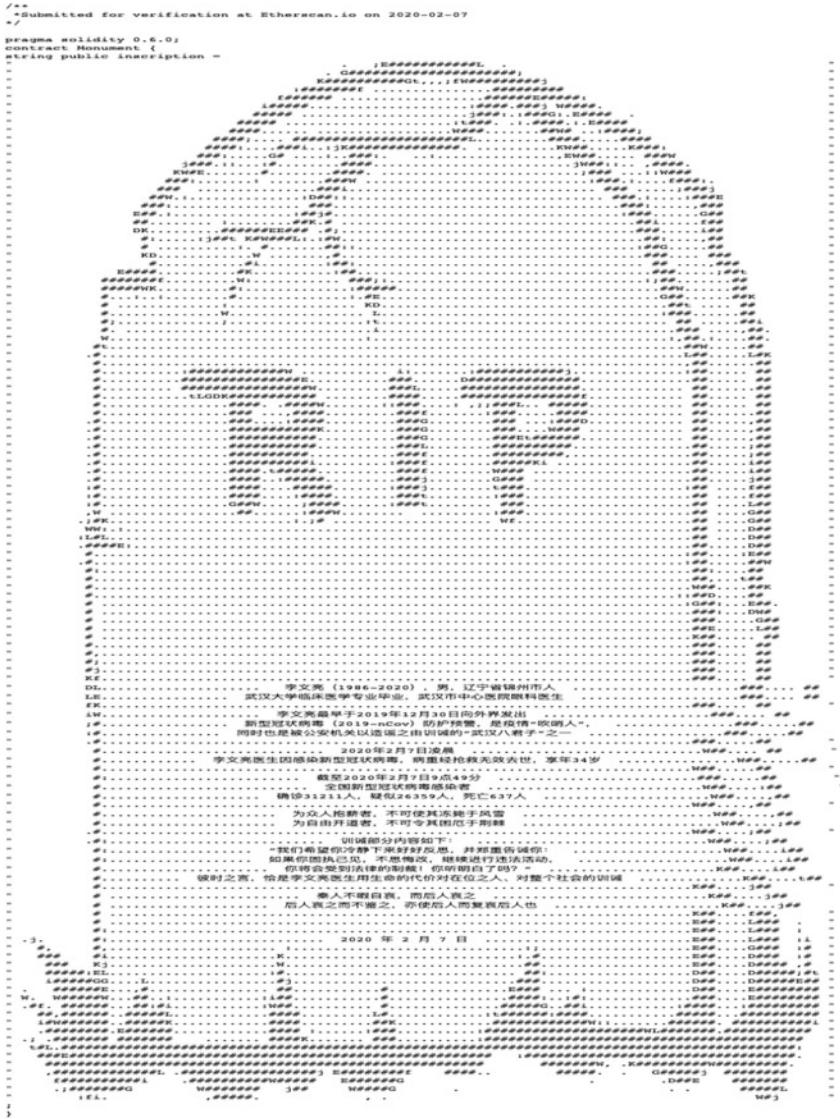


Fig. 4.1 The decrypted virtual memorial honoring Dr. Wenliang on the Ethereum blockchain

this emerging technology. The purpose of this discussion is not to make any political judgments about this encrypted monument, but rather to focus primarily on the features of the encrypted place and its social implications.

In the remainder of this chapter, we review the literature on monuments and their relevance to place. In addition, we discuss the promise of blockchain and analyze the concept of crypto place through a case study of the abovementioned encrypted

virtual memorial for Dr. Li. This chapter will make unique theoretical and practical contributions by examining the increasing significance of blockchain to geography and GIScience studies.

4.2 Monumental Landscapes and Place Memory

A monument is an artifactual structure that is created to commemorate an individual or event (Johnson 1995), such as a memorial, a statue, an archaeological site, or a gravestone. In addition to traditional physical memorials, our analysis will include virtual monuments on social media or the Internet, including an augmented reality monument entitled “Stonewall Forever” to celebrate the fight for equal rights for the LGBTQ community (Fitzgerald 2019) and another online memorial, the “Global Cemetery” that enables remote visitors to place virtual flowers on the virtual “graves” of their loved ones. Like their physical counterparts, these virtual monuments provide a constant visual message for visitors (Özbek 2018) and allow them to commemorate the past and to envision the future (Harvey 1979; Stangl 2008).

Once a monument is built, it becomes an integral part of its location or even a destination in and of itself, where new memories and meanings will be linked to this place (Auster 1997). Moreover, one’s experiences at a monument should be viewed from both the perspective of time and place. As Arendt argued (1958), “Monuments generate sweet dreams of immortality in all of us.”

While monuments are created to incorporate a sense of perpetuity for the place, most physical monuments, ironically, have been destroyed by natural processes, vandalism, or even political disputes (e.g., the removal of the U.S. monument for the Confederate soldiers), excluding some rare indestructible examples like Mount Rushmore or the Great Wall of China. The meanings of monuments are not only continuously negotiated between their founders and visitors but also are enriched by the present and the past (Auster 1997), the death of the honored individual and the passage of time. In most cases, descendants assign new meanings to a place (Muzaini 2015), while older memories are very likely to fade away as time passes (Özbek 2018).

A monument invites debates about conflicting attitudes, interests or ideologies (Boym 2009), even though political officials often have absolute power to decide what specific meaning a monument should carry. Even virtual monuments on web pages or social media platforms cannot escape the control of authorities (Zhao 2020). Thus, a place, whether physical or virtual, can hardly contain any permanent meanings. However, thanks to the emergence of blockchain technology, a crypto place can escape the control of authorities and the influence of political debates. Thus, in the next section, we will describe how the blockchain serves as a landscape within this new type of place.

4.3 The Landscape of Ethereum Blockchain

Blockchain is widely recognized as the theoretical foundation for cryptocurrencies such as Bitcoin that enables its users to trade digital assets without the involvement of a central bank or of any sovereign body (Crosby 2016). Ethereum has transcended Bitcoin through the smart contract function which provides an effective and secure means of payment for e-commerce (Frey et al. 2016; Subramanian 2017) and location-based services (Jia 2018; Yang et al. 2019). A blockchain provides freedom from the control of governments or institutions; it allows individuals to perform transactions safely and securely by shifting (1) from government-issued currencies to a cryptocurrency (e.g., Bitcoin, Ethereum); (2) from trusted third-parties to automated and community-driven counter-party risk mitigation and (3) from party-level ledger keeping to shared/crowd ledger management (Lacity 2020).

Ethereum blockchain’s dispersed peer-to-peer network can be described as a “landscape,” in which there are two types of nodes: the Ethereum Virtual Machines (EVMs) and miners. The former implements the smart contracts to initiate the transference of cryptocurrency for a variety of services. Unlike the EVM nodes, a miner creates new blocks and then attaches them in a global ledger. In other words, a blockchain maintains a ledger, or a continuously growing list of transaction records.

There are two types of transactions that can be initiated within the Ethereum blockchain: a) transferring ether to another user, or b) eliciting a trade within a smart contract. A user must pay a small fee of “ether”,² also known as the “gas fee,” to initiate the transaction. Then, the transaction must be digitally signed with the sender’s private key to ensure legitimacy and verify his or her identity so that the transaction can be authenticated and the balances of the Ethereum users can be changed. Once complete, a receipt will be generated and stored in a block. Each transaction receipt contains information about the sender and recipient, the value of the transferred assets, the transaction fee, and a piece of encrypted data that contains all transaction-related information, such as the balances (Zhang et al. 2018). It must be noted that the blockchain secures transactions and concomitant data through cryptographic algorithms. All private information on the blockchain (e.g., a user’s account address, transaction data, and smart contract address, etc.) is encrypted through the hash function. Accordingly, only people with access to the corresponding key are able to decrypt and view the information. Thus, the hashing encryption allows the blockchain to transfer private information securely.

4.4 The Encrypted Memorial as a Crypto Place

A physical place is comprised of a location as well as a sense of place that is developed by the people who have interacted with it (Agnew 1987). In this light, we argue that a

²Ether is the basic denomination unit of Ethereum, as the dollar is the basic denomination unit of the U.S. currency system.

smart contract can create a new type of place—a “crypto place”—on the blockchain landscape. As previously stated, once a smart contract is initiated, a unique contract address will be generated and then encrypted in hash code as a way to index this smart contract. Simultaneously, an initial transaction record is created to store the relevant initial parameters of this smart contract. Then, it is ready for Ethereum users to transfer ether to it. Thus, the monument for Dr. Li can be regarded as a crypto place because its location can be specified by the unique transaction hash address—0x6e46d3ab7335fffb0d14927e0b418cc08fe60505. It was created by the Ethereum user with the following address 0x16a51077e0548cb265b223d180844beafb76e03b on the block height of 9432824 in the Ethereum blockchain on February 7th, 2020. The creation of this crypto place cost a gas fee of 0.067675152 that is equal to \$17.05 based on the ether to U.S. dollar exchange rate at the time it was created. The profile of this crypto place is stored in the initial transaction record (address: 0xb16c93d6f51ee7a68d626be22add24455db0a4644c68da69af6705981c77820e). Soon after its creation, six users transferred ether to this contract to show their support to Dr. Li.

4.4.1 Decentralized Location

The location of a place refers to a geographical area in a particular environment. A Encrypted place is located by its contractual address. For example, Dr. Li's encrypted monument can be found via the following unique address 0x6e46d3ab7335fffb0d14927e0b418cc08fe60505. This crpto place can be easily accessed using any Ethereum mirror source (e.g., Ethereum's Etherscan or Google BigQuery datasets). It is important to note that this unique contract address can direct the visitors to a copy of the crytpo place (a transaction to a smart contract); multiple copies spread out in different Ethereum nodes. In other words, the contract address and its referred crypto place follows an one-to-many rather than an one-to-one relationship between a location and its reference site. This type of decentralized reference system is able to function efficiently primarily because each transaction is anchored to a unique block, and each block is synchronized to several other Ethereum nodes, creating a streamlined network. Thus, a decentralized location does not exist in a single Ethereum node but across multiple nodes. Since it is not owned or overseen by any single user or organization (Di Masso and Dixon 2015), this lowers the risk that a crypto place will be censored by authorities (Curtis 2003) and also creates a secure place for blockchain users to express extreme sentiments or criticism towards that crypto place.

4.4.2 *Immutable Locales*

The locale of an encrypted site contains both material and immaterial components. The hard disk that stores the record of a transaction is the material component, while the immaterial components include its address, transaction status, location block, timestamp, the addresses of the sender (from) and receiver (to), its transfer value, fuel charge and input data, as illustrated by the transcation information for the encrypted memorial in Table 4.1. The input parameters allow the creator of the crypto place to record any type of encrypted multimedia information such as text, images, audio or video files (Wang et al. 2018).

Unlike conventional monuments that can be damaged or even destroyed (cf., Auster 1997; Özbek 2018), the locale of a crypto place is nearly immutable. This is because any change in an existing transaction must take place in all the blocks and simultaneously within all the Ethereum nodes, which is extremely difficult for any hacker to accomplish.

Therefore, an Ethereum user must be very careful when creating a smart contract due to the immutable feature of the encrypted locale. Any smart contract will exist perpetually and almost impossible to remove from the blockchain. Such a feature might engender serious social implications. For example, if any private information was published as part of a locale, it might lead to the long-misuse of private data.

Table 4.1 The smart contract profile of the encrypted memorial for Dr. Li

Field name	Value
Hash	0xb16c93d6f51ee7a68d626be22add24455db0a4644c68da69af6705981c77820e
Transaction Index	22
Block Hash	0x7704b4e8188249acc6f41469040332468830debacd5895949d5498819683f5bf
Block Number	9432824
From	0x16A51077E0548Cb265b223D180844BeAfb76E03b
To	None
Nonce	303
R	0x5a25e1f33829f6f7cc8b1115ad3e7094aa0243be762f5fb51feb026bb5786545
S	0x28b70c72428881e80740e82cbd00d78270f2bcd36e967ded85d0288bf61b96a5
V	38
Input	0x60806040526040518061328001604052 // Length: 27,824 digits
Value	0
Gas	8459394
Gas Price	8,000,000,000

4.4.3 Transaction-Based Sense of Place

Ethereum users can share their experiences with and memories about a crypto place with others through transactions. To do so, a user must follow the crypto place's smart contract to initiate a new transaction. Then, the initiator can encrypt a memory into the input data parameter of the newly-created transaction. Others will be able to read the recorded memory only by decrypting the input data into UTF-8 text. Usually, a new transaction process costs some gas fee. Therefore, the value-laden nature of blockchain transactions somewhat reflects the extent of the sender's support and concern about a particular crypto place. Also, a sender can express further support by adding "a written note" to this input data. For example, Table 4.2 lists all the transactions associated with Dr. Li's encrypted memorial. Each transaction includes its hash identifier, the height of the locating block, an initiated time stamp, transfer value, and gas fee. It is clear that many Ethereum users have initiated transactions and have transferred ether as a way of expressing their respects to Dr. Li.

It must be noted that most transactions have no recorded content within the input parameter. Only one user ($\text{id} = 4$) expressed his or her regard for this monument in both Chinese and English as below.

I born and live in Hubei, China. I suffered great pain this spring, I lost a relative that I loved very much. And I witnessed many other painful people like me. Please give me some eths, I want to feel be loved and blessed, and I want to help people. I will use all the eths to help painful people here I see and know for real. I have been a good people all my life till now, but now I'm feeling that I'm turning into a bad person inside out, I am longing for love. I have some talents and have worked hard, but I am nothing but mere mortal. I am so incompetent.

This user described the miserable suffering during the COVID-19 pandemic, in order to raise ether. By showing this user's purpose, we argue that the presence of input data in smart contract gives each Ethereum user an opportunity to publicly express his or her sense of place.

As shown above, the sense of place expressed on the blockchain is transaction-based and therefore value-laden. This economic dimension has not been fully unveiled by any previous research on sense of place, such as the classical model proposed by Scannell and Gifford (2010). The transfer of digital assets and use of gas implies that the sender respects the encrypted place enough to pay for this input, which can be used to share postings about his or her sense of place. In other words, since the size of the input parameters is positively and exponentially correlated with the cost, each user must carefully weigh the cost/benefit of initiating a transaction by asking themselves whether it is worthwhile to post these feelings. In that sense, it is much more difficult or at least more expensive to spread news or rumors on Ethereum than on most free social media platforms (e.g., Twitter, Facebook or TikTok).

Table 4.2 Transactions to the smart contract

Id	Address	Block #	Timestamp	From	Value In	Gas Fee
1	0xed132d7889693211b2d571ed6f186bb2f15d651cd82717e222d8f088284fb55	9433273	27/2/2020 3:48 GMT	0x1ef63d4eabe1f641656d282f75hb5b513755ec615	0.00100000	0.000105000
2	0x02c37ebe3998c007e6290d102460a7e7a06595765aaa34c48366cb8334f4bd54	9433396	27/2/2020 4:17 GMT	0x851af22f2c5f18eda43b70b895254ed8c307e320	0.00521000	0.000063000
3	0x14cc0198cdad9970d27ff3cd8899e3158115dd466e46aa1581e3bb003caeadb1ae	9434837	27/2/2020 9:47 GMT	0xta18807d7423ae833f24785c125e9e00772853b	0.00750000	0.000105000
4	0x5591c1c0ff7814847e6a7aa121777e7b03be813a2279253033c6f5821f2c4aaa	9436538	27/2/2020 16:07 GMT	0xfd146c56791b47e39480e80dc2343364375ae537	0.00100000	0.001324846
5	0x88ccb61e0053305fef176e7375a5b139fd76baifaacb18ab9a2c184d7e2081af	9460898	2/11/2020 9:38 GMT	0xe420831bcbd7c707539fdb4092bd61c52a91984d2	0.00202027	0.000063000
6	0xb18435654d8f4168a6d0803dd02f150aa3fd0bb672c4675124b34dcda93f3c96	9499435	2/17/2020 7:59 GMT	0xb07eb598e6at35e774c595bcb213f08d95e5e19	0.00100000	0.000021000

4.5 Discussion

After investigating the three features of a crypto place, we will further discuss how this new type of place enhances the authenticity of human experience and expands the concept of place.

4.5.1 *The Authenticity of the Experience*

Place is an essential defining concept in the field of human geography (Tuan 1974; Relph 1976; Massey 1984; Harvey 1986; Hetherington 1997; McDowell 1997; Keith and Pile 2004; Cresswell 2008; Malpas 2018). While there are many connotations of this concept, the majority of people tend to seek an authentic experience from a place (Zhao et al. 2019) no matter it is in a crypto or a physical area.

As Edward Casey (2000) argues in *Remembering: A Phenomenological Study*, “The memory of place, or human memory more generally, is always associated with human bodily experience. Traditional monuments are often large and imposing, and always provide a strong visual stimulus,” a sense of awe and a deep connection with tradition and past generations; however, it is also linked to the contemporary world. Any experience is already intersubjective, just as it is always already bodily and implanted. When one visits the Lincoln Memorial, one must, like so many others have over the decades, climb many steps to see beyond the columns, read the inscriptions of his speeches, and gaze at the larger-than-life statue of Lincoln sitting in his chair.

For these reasons, an encrypted monument seems to provide less of an enthralling experience. While the landscape is attractive, it is far less authentic due to its inability to stimulate the senses. It may even be treated as a temporary commemoration, a kind of spur-of-the-moment tribute. The display may be immediate, but it lacks any permanence or ongoing attention. It may not provide a lasting sense of connection to the physical place, especially if the encrypted monument requires expensive maintenance. Impromptu memorials do provide necessary reminders of death and the fragility of human beings, but in fact they are only temporary.

Thus, a relevant question must be raised of how an encrypted monument can provide a more realistic and all-encompassing experience through virtual reality. There are two passages that highlight the potential of this technology. Brooks declares (summarizing Sutherland [1965]), “Do not think of that thing as a screen, think of it as a window, a window that lets you look into the virtual world.” The challenge for computer graphic artists is to make the virtual world look real, sound real, allow the user to move and interact in real time, and have the experience even feel real (1999, p. 16). Moreover, according to Negroponte (1995):

Digital life will include less and less reliance on being in a particular place at a particular time, and the delivery of the place itself will begin to be possible. If I can actually look out of the electronic window of my living room in Boston and see the Alps, hear the cowbells, and smell the (digital) manure of summer, in a way I am very much in Switzerland (p. 165).

These possibilities will enrich our understanding about sense of place, which can be defined as the lived experience of being in the world (Heidegger 2008). In this regard, Ludwig Binswanger's concept "Being-beyond-the-world" eloquently illustrates the experience of encrypted place, which emphasizes the potential for human beings to transcend the limitations of everyday life in the mundane world (Binswanger 1941). Indeed, as blockchain technology enables us to claim new territories and encrypt new types of places, we are able to give new meaning to these new spaces. This process demonstrates that while we are confined to the mundane physical world, we can still break the limits of space and time.

4.5.2 Creating a Sense of Place at the Encrypted Memorial

When we take a closer look at the experience of encrypted memorial we may begin to question the role that they can play in the preservation of memory, the establishment of community and the creation of place. In many ways, they appear to serve these functions; however, they fail to establish the kind of embodiment necessary for traditional transmission and preservation of memory. In addition, due to the fact that they are not entirely accessible to the public, they become a very personal coping mechanism, sometimes serving commercial interests as well.

Despite their inability to provide a total experience, encrypted memorials can still promote the interaction with the visitors. The ability to attract new visitors is not dependent on spatial distance or common friends. The online comment function associated with most virtual memorials allows people from around the world to communicate electronically with others who share a similar feeling, thus forming a virtual community.

The creation of a place, whether fictional or real, physical or virtual, requires some degree of creativity. It is impossible to describe every detail, so the creator must select essential elements of a place and rely on a combination of his or her own expression and the imagination of the audience to fill in the gaps and create an image of the place as a whole. Of course, many digital virtual places are interactive and can be explored from different perspectives, which requires active imaginations on the part of the visitors. The original designer of a virtual place must, to some extent, anticipate how participation might change the place and provide appropriate inspiration for it. Although in a fully interactive virtual place, as in a real one, the imaginative involvement of participants will lead to some unforeseen changes. There are no rules or guidelines for this, and the most compelling virtual places can be perceived as ever-changing works of art, reflecting the combined imaginations of the creators and the participants.

Similar to physical places, such imaginative involvement helps the sense of the virtual place to develop. Not only will it involve many senses and emotions but because it is electronically mediated, the experience of it will vary from individual to individual. It will also have an active community surrounding it in that the sense of place is shared with millions of participants who are connected by the virtual worlds

they choose to explore. However, for those who choose to explore different virtual worlds, this unique sense of place can be considered a variant and a complement to the current dispersed sense associated with a real place. It simultaneously recognizes geographic diversity and seeks ways to give places compelling identities.

4.6 Conclusion

In this chapter, we introduced the emerging phenomenon of crypto place on the blockchain during the COVID-19 pandemic. It represents a newly emerged territory that consists of three dimensions: decentralized location, an immutable locale, and a transaction-based sense of place. These three unique characteristics have greatly expanded the geographical concept of place.

Blockchain technology gives us the ability to create a permanent place in an encrypted form, which allows us to maintain our sense of place in a more secure way, especially when faced with restrictions from public officials or online censorship. However, we should not neglect some of the negative effects of this new technology. For example, this emerging form of crypto place can also be used to spread fake news, which leads to a misinformed public. To cope with the complex social implications of blockchain technology, we hope to call on scholars in various fields to examine its role in today's decentralized and data-intensive world. For future research, we hope to utilize blockchain technology in geographic studies to create crypto places that preserve the collective memory of humanity and prevent the vanishing of cultural landscapes.

References

- Adams, P. (1998). Network topologies and virtual place. *Annals of the Association of American Geographers*, 88(1), 88–106.
- Agnew, J. (1987). *The United States in the world-economy: A regional geography*. Cambridge, UK: Cambridge University Press.
- Arendt, H. (1958). *The Human Condition*. Chicago: University of Chicago Press.
- Auster, M. (1997). Monument in a landscape: The question of “meaning”. *Australian Geographer*, 28, 221–230.
- Binswanger, L. (1941). On the relationship between Husserl’s phenomenology and psychological insight. *Philosophy and Phenomenological Research*, 2(2), 199–210.
- Boym, C. (2009). If there were a monument. *Newsweek*, 153(24), 43.
- Cresswell, T. (2008). Place: encountering geography as philosophy. *Geography*, 93(3), 132–139.
- Crosby, M., Pradan, P., Sanjeev, V., & Vignesh, K. (2016). Blockchain technology: Beyond bitcoin. *Applied Innovation*, 2(6–10), 71.
- Curtis, S. (2003). Policy and place: General practice in the UK. *Transactions of the Institute of British Geographers*, 28, 28–40.
- Dannen, C. (2017). *Introducing Ethereum and Solidity* (Vol. 1). London: Berkeley, A Press.
- Di Masso, A., & Dixon, J. (2015). More than words: Place, discourse and struggle over public space in Barcelona. *Qualitative Research in Psychology*, 12(1), 45–60.

- Feng, E., & Cheng, A. (2020). *Critics say china has suppressed and censored information in coronavirus outbreak.*
- Fitzgerald, M. (2020). Google Celebrates 50 Years of LGBTQ Pride with ‘Living Monument’ to the Stonewall Riots. *Times*. <https://time.com/5600553/google-stonewall-living-monument/>.
- Frey, R., Dominic, W., & Alexander, I. (2016). *Collaborative filtering on the blockchain: A secure recommender system for e-commerce*. <https://aisel.aisnet.org/amcis2016/ISSec/Presentations/36/>
- Harvey, D. (1979). Monument and myth. *Annals of the Association of American Geographers*, 69(3), 362–381.
- Harvey, D. (1989). *The condition of postmodernity* (Vol. 14). Oxford: Blackwell.
- Heidegger, M. (2008). *Building Dwelling Thinking*. London: HarperPerennial, Basic Writings.
- Hetherington, K. (1998). In place of geometry: the materiality of place. *The Sociological Review*, 45(1_suppl), 183–199.
- Huang, C., Wang, Y., Li, X., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet*. [https://www.thelancet.com/journals/lancet/article/PII_S0140-6736\(20\)30183-5](https://www.thelancet.com/journals/lancet/article/PII_S0140-6736(20)30183-5).
- Janet, D., & Books, L. (2015). Remembering places: a phenomenological study of the relationship between memory and place. *Journal of Cultural Geography*, 33(1), 184–185.
- Jia, B., Zhou, T., Li, W., et al. (2018). A blockchain-based location privacy protection incentive mechanism in crowd sensing networks. *Sensors*, 18(11), 3894.
- Johnson, N. C. (1995). Cast in stone: Monuments, geography, nationalism. *Environment and Planning D*, 13, 51–65.
- Keith, M., & Steve, P. (2004). *Place and the Politics of Identity*. London: Routledge.
- Kavaratzis, M. (2005). Place branding: A review of trends and conceptual models. *The Marketing Review*, 5(4), 329–342. NPR. <https://www.npr.org/sections/goatsandsoda/2020/02/08/803766743/critics-say-china-has-suppressed-and-censored-information-in-coronavirus-outbreak>.
- Kotler, P., & Gertner, D. (2002). Country as brand, product, and beyond: A place marketing and brand management perspective. *Journal of brand management*, 9(4), 249–261.
- Lacity, M.C. (2020). *Blockchain Foundations: For the Internet of Value*. Epic Books.
- Malpas, J., (2018). *Place and experience: A philosophical topography*. Routledge.
- McDowell, L. (1997). *Undoing place? A geographical reader*. Arnold.
- Muzaini, H. (2015). On the matter of forgetting and ‘memory returns’. *Transactions of the Institute of British Geographers*, 40, 102–112.
- Owley, J. (2018). Understanding the complicated landscape of civil war monuments. *Indiana Law Journal*, 93, 15.
- Özbek, E. (2018). The Destruction of the Monument to Humanity: Historical Conflict and Monumentalization. *International Public History*, 1(2). <https://doi.org/10.1515/iph-2018-0011>
- Pallot, J. (2005). Russia’s penal peripheries: space, place and penalty in soviet and post-Soviet Russia. *Transactions of the Institute of British Geographers*, 30, 98–112.
- Pierce, J., Martin, D. G., & Murphy, J. T. (2011). Relational place-making: the networked politics of place. *Transactions of the Institute of British Geographers*, 36(1), 54–70.
- Rakodi, C. (2006). Relationships of power and place: the social construction of African cities. *GeoForum*, 37(3), 312–317.
- Relph, E. (1976). *Place and Placelessness*. London: Pion.
- Scannell, L., & Gifford, R. (2010). Defining Place Attachment: A tripartite organizing framework. *Journal of Environmental Psychology*, 30, 1–10.
- Shafagh, H., Burkhalter, L., Hithnawi, A., et al. (2017). Towards blockchain-based auditable storage and sharing of IoT data. In *Proceedings of 2017 on Cloud Computing Security Workshop*, 45–50.
- Stangl, P. (2008). The vernacular and the monumental: memory and landscape in post-war Berlin. *GeoJournal*, 73(3), 245–253.
- Subramanian, H. (2017). Decentralized blockchain-based electronic marketplaces. *Communications of the ACM*, 61(1), 78–84.
- Wang, S., Zhang, Y., & Zhang, Y. (2018). A blockchain-based framework for data sharing with fine-grained access control in decentralized storage systems. *IEEE Access* 6, 38437–38450.

- Wellman, B. (2001). Physical place and cyberplace: The rise of personalized networking. *International Journal of Urban and Regional Research*, 25(2), 227–252.
- Wieczner, J. (2017). The 21St-Century Bank Robbery. *Fortune*, 176 (3), 34–41.
- Xu, X., Zhang, L., & Wong, T. (2010). Structure-based ASCII art. In *ACM SIGGRAPH 2010* papers, 1–10.
- Yang, M., Zhu, T., Liang, K., et al. (2019). A blockchain-based location privacy-preserving crowdsensing system. Future *Generation Computer Systems*, 94, 408–418.
- Zhao, B., & Huang, X. (2020). Encrypted monument: The birth of crypto place on the block-chain. *Geoforum*, 116, 149–152.
- Zhang, Y., Kasahara, S., Shen, Y., et al. (2018). Smart contract-based access control for the internet of things. *IEEE Internet of Things Journal*, 6(2), 1594–1605.
- Zhao, W. (2020). China's Coronavirus Whistleblower Is Now Memorialized on Ethereum, Coindesk. <https://www.coindesk.com/chinas-coronavirus-whistleblower-is-now-memorialized-on-ethereum>

Xu Huang is an Associate Professor in the School of Geography at Nanjing Normal University, Nanjing, 210046, China. His research interests include the sense of place, residential mobility and place making.

Bo Zhao is an Assistant Professor in the Department of Geography at the University of Washington, Seattle, WA 98195. His research interests include GIScience, digital geographies, location spoofing, and humanistic geography.

Chapter 5

Emotional Responses Through COVID-19 in Singapore



**Yingwei Yan, Wei Chien Benny Chin, Chan-Hoong Leong, Yi-Chen Wang,
and Chen-Chieh Feng**

5.1 Introduction

Since the first case of COVID-19 in Wuhan, China, was announced in early 2020, the virus has spread worldwide, resulting a global pandemic declared by the World Health Organization (WHO) on 11 March 2020. As a popular destination for visitors from China, Singapore was one of the first countries to be affected by the novel coronavirus with the first case of infection confirmed on 23 January 2020. While many countries adopted a wait-and-see approach to the less understood disease, Singapore was proactive in setting up measures to screen possible cases given the SARS experience in 2003. Singapore was hailed as a role model for the world in fighting against COVID-19 as it had successfully contained infection cases in the early stages while keeping everyday business as usual. Unfortunately, the success did not shield Singapore from the subsequent waves of COVID-19 infections, arising mainly from returning Singaporeans taking refuge from the pandemic outbreaks outside of China. This led Singapore into a lockdown, or what local authority called

Y. Yan

Key Lab of Guangdong for Utilization of Remote Sensing and Geographical Information System,
Guangdong Open Laboratory of Geospatial Information Technology and Application, Guangzhou
Institute of Geography, Guangzhou, China
e-mail: yanyingwei@u.nus.edu

Y. Yan · Y.-C. Wang · C.-C. Feng
National University of Singapore, Singapore, Singapore

W. C. B. Chin (✉)
Singapore University of Technology and Design, Singapore, Singapore
e-mail: benny_chin@sutd.edu.sg

C.-H. Leong
Singapore University of Social Sciences, Singapore, Singapore
e-mail: chanhoong@suss.edu.sg

the “circuit breaker” (CB), starting from 7 April 2020. During the CB period, non-essential services were restricted to work-from-home or complete shutdown, and all schools were moved to full online learning. While this mobility restriction evidently changed the usual activities, routines or livelihoods of the residents, the impact of CB on their emotions is less known.

Emotion (e.g., feeling) is an integral aspect of human experiences, the variation of which can be a direct consequence of crises and this in turn leads to further changes in the physical and mental dimensions of human dynamics (Seagal and Horne 2020; Shaw et al. 2016). Emotional traumas in societies as a result of a major calamity may lead to massive stress-related disorders (Ćosić et al. 2020). For COVID-19 in particular, a WHO technical guidance notes the main psychological impact to date being “elevated rates of stress or anxiety” (WHO 2020). Unlike other crisis, such as terrorist attacks and hurricanes, COVID-19 is intangible and can cause long-term uncertainties that lead to fear, loneliness, distress reactions, and mental health disorders (Aslam et al. 2020; Ćosić et al. 2020).

Recently, sentiment analysis of social media data in the context of COVID-19 pandemic has attracted attention from the research community (Barkur et al. 2020; Li et al. 2020; Lwin et al. 2020). Indeed, social media data can be harnessed to understand the responses of residents to crisis events because the data are both cost-effective to collect, and their richness in volume and spatiotemporal coverage is unrivaled against traditional data sources (Wang and Ye 2018; Yan et al. 2017; Yan et al. 2020b). With respect to human dynamics, compliance to evacuation procedures and communication behaviors of residents during disaster and crisis events have been explored using social media (Kim et al. 2017; Martín et al. 2017; Takahashi et al. 2015). These studies tend to focus on the physical (e.g., evacuation actions) and mental dimension (e.g., perspectives) of human dynamics (Seagal and Horne 2020). In terms of the emotional dimension of human dynamics, sentiment analysis of Tweets posted in crisis events has received considerable attention in recent years. These research have contributed to a better understanding of the online community dynamics in response to terrorist attack (Shaikh et al. 2017), the dynamics of social networking (Neppalli et al. 2017), the sentiments towards the refugee crisis (Öztürk and Ayvaz 2018), and the post-earthquake sentiment variations in tourism destinations (Yan et al. 2020a).

Despite the value of social media sentiment analysis in understanding emotional responses to crisis, existing studies have seldom investigated the changes of sentiments, which have the potential to shed lights on how people react when transitioning from normal to crisis situations. This study performed a sentiment analysis of Tweets posted in Singapore and aimed to reveal the temporal dynamics of the online communities’ emotions in the context of COVID-19 global pandemic. In addition, Singapore is a country with a high density of population (7866 km^{-2}) and multi-ethnics and multi-lingual environment (DSS 2020). Singapore resident population composed of ethnic Chinese (74.4%), Malays (13.4%), Indian (9.0%), and other minority races (3.2%). Furthermore, the citizens and permanent residents proportion composed about 70.6% of the total population, leaving 29.4% of the population as foreign employees, dependents, or students with long-term residential passes. As such, it is

crucial to examine the content of all Tweets using a multicultural lens. Therefore, this research further analyzed the emotional differences of Twitter users in different ethnicities (speaking different languages) of Singapore during the pandemic.

5.2 Material and Methods

The workflow of this study is illustrated in Fig. 5.1. It started with Tweets collection, followed by six Tweets pre-processing steps, including the language detection of the retrieved Tweets, translation of those non-English Tweets into English, tokenization, stop words removal, and lemmatization of the Tweets, and removal of the Tweets from news media. Lastly, sentiment analyses were performed to reveal both the overall sentiment dynamics and the language-based sentiment dynamics. Details are introduced in Sects. 5.2.1, 5.2.2 and 5.2.3.

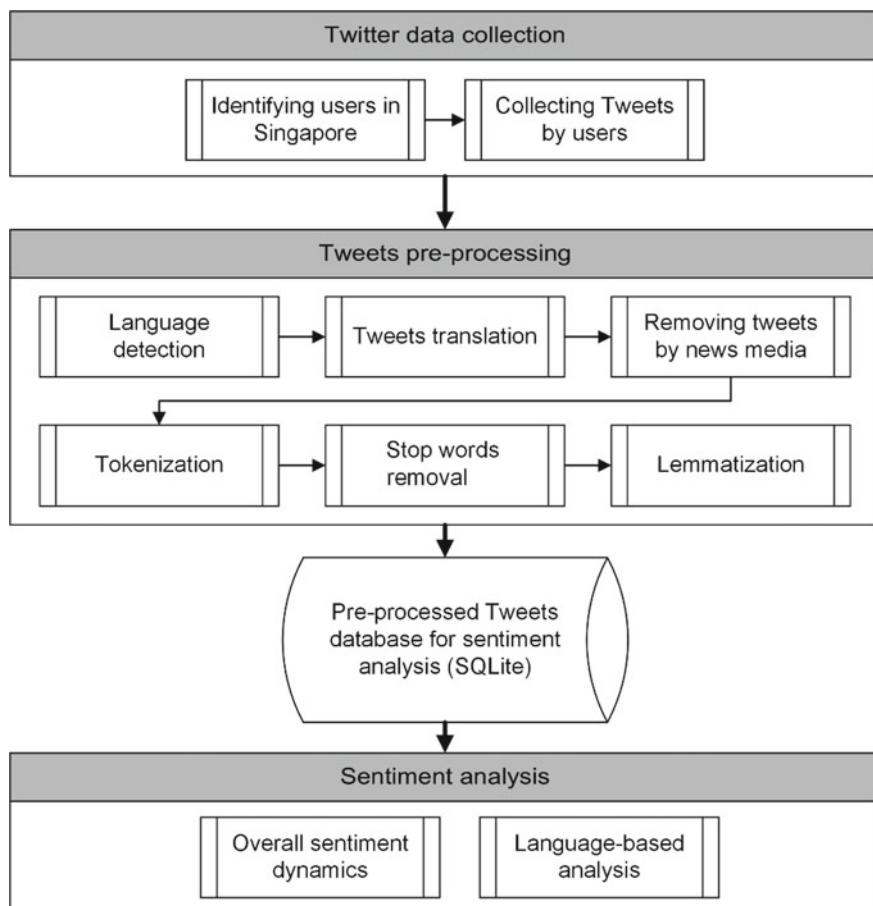


Fig. 5.1 Workflow of this study

5.2.1 Twitter Data Collection

Tweets related to COVID-19 were collected via TAGS (<https://tags.hawksey.info/>), a free Google Sheets template allowing us to set up and run automated collection of search results from Twitter. The geocode parameter of the Standard Search Application Programming Interface (API) of Twitter was used to retrieve Tweets that were located within Singapore. The Standard Search API takes location information preferentially from the Geotagging API, but will fall back to users' Twitter profile.

Real-time Tweets with their texts and hashtags containing the keywords of “COVID” or “coronavirus” (i.e., COVID-19-related Tweets) were collected between 16 April 2020 and 31 May 2020. After removing the non-individual (e.g. news media) accounts, a total of 12,067 users were found. Because COVID-19 pandemic started from January 2020 and Singapore has imposed semi-lockdown measures during the period of data collection, we assumed that these users resided in Singapore from January to May of 2020. After that, all the Tweets (regardless whether a Tweet was related to COVID-19 and regardless whether a Tweet was geotagged) posted by the same group of Twitter users who posted the COVID-19-related Tweets were further collected through Twitter’s official API based on the user IDs. A total of 18,535,620 Tweets were collected. We retrieved the Tweets from the users posted in the first five months of 2020 for the main analysis, including (a) January and February: the nascent stage of COVID-19 outbreak when the disease was first imported to Singapore and transmitted locally (1,622,470 Tweets), (b) March: the stage where the COVID-19 outbreak escalated in Singapore (1,227,370 Tweets), (c) April and May: when the “Circuit Breaker” semi-lockdown measures were imposed (3,572,745 Tweets). We also collected the Tweets of the same months posted by the users in the preceding year (2019) to serve as the baseline for comparison (1,368,162 Tweets). All these Tweets were composed by text in different languages. Tweets composed by non-words objects (e.g., static images, GIFs, only URL links, only emoji characters or special characters) were excluded.

5.2.2 Tweets Pre-processing

After the retrieval of the Tweets, Google Sheets were used for the language detection of the Tweets and for the Tweets translation. Following Yan et al. (2020b), Tweets were further pre-processed by tokenization (cohesive strings from the Tweets were split up into single words or “tokens”), stop words removal (frequently occurring short-function words without valuable content such as “of” and “to” were removed to reduce noise), and lemmatization (the words were converted to their root form); these pre-processing steps reduced the semantic dimension of the raw Tweets for creating word vectors. All pre-processed Tweets were stored in a SQLite (3.22.0) database file for the subsequent sentiment analysis.

5.2.3 *Sentiment Analysis*

The sentiment analysis of this study was performed based on Plutchik (2001), which postulates eight primary bipolar emotions: joy (feeling happy) versus sadness (feeling sad); anger (feeling angry) versus fear (feeling of being afraid); trust (stronger admiration and weaker acceptance) versus disgust (feeling something is wrong or nasty); and surprise (being unprepared for something) versus anticipation (looking forward positively to something). This approach has been adopted extensively. Particularly, the National Research Council (NRC) Canada presented a large Word-Emotion Association Lexicon created through a massive online annotation project based on Plutchik's eight basic emotions (Mohammad and Turney 2013).

The NRC Word-Emotion Association Lexicon has been implemented in the Syuzhet package of R (<https://www.rdocumentation.org/packages/syuzhet/versions/1.0.4>). The Syuzhet package enabled us to generate emotional valence value for each Tweet. It also enabled us to identify the words in each Tweet that indicated eight basic types of human emotions (i.e., anger, fear, anticipation, trust, surprise, sadness, joy, and disgust) and the words in each Tweet that indicated positive and negative emotions. Both the overall sentiment analysis (without separating the Tweets by languages) and sentiment analysis by languages (the Tweets were separated by languages) were performed. We grouped the Tweets by days and analyzed the daily emotional valence trends in the following sections. In addition, Augmented Dickey-Fuller test (Kwiatkowski et al. 1992) were used to check if each of the emotional valence daily trends was stationary.

Hierarchical clustering was performed to investigate the clustering and hierarchical relationships among the emotional responses of the Tweets in different languages and emotion types. The tool used to perform the hierarchical clustering was a function (AgglomerativeClustering) from the Scikit-learn package of Python. To understand the similarity and differences between languages and emotions, the daily emotional dynamic for each emotion of a language was used as a data point, with each day as a vector. In data mining and statistics fields, agglomerative hierarchical clustering is a method to group similar objects (in this study, language) based on the observation patterns (i.e., the trends of the eight emotions) using a bottom-up approach. It starts with each object as a cluster, followed by calculating the differences between each pair of clusters, then groups the most similar clusters (i.e., with the smallest differences); it repeats the steps of difference calculation and cluster grouping. In this study, the trend of each emotion was used as the patterns for comparing between languages. In other words, if two languages were grouped as a cluster, it indicated that the two languages had the same trend of a specific emotion during the five-month period.

5.3 Results and Discussion

5.3.1 Overall Sentiment During COVID-19

The emotional valence in the first five months of 2019 and 2020 (Fig. 5.2) shows similar percentage of Tweets exhibiting zero emotional valence. The proportion of Tweets with negative emotional valence is slightly higher in 2020 (35%) than in 2019 (29.4%). In both 2019 and 2020, more than half of the Tweets have a positive emotional valence. Aggregating the emotional valence leads to the daily average of emotional valence of 2019 and of 2020 (Fig. 5.2) both in the positive value range, suggesting a weak positive emotion throughout the study period. In addition, the emotional valence trends of the two years were stationary (Table 5.1). Nonetheless, due to the proportion of negative Tweets in 2020 is larger than in 2019, the overall trends in 2020 daily average values are lower than in 2019. To confirm that the overall emotional valence values of the two years are indeed different, an ANOVA test was conducted. The test result ($F(1,301) = 506.609, p < 0.001$) confirmed that the drop of overall emotional valence during COVID-19 period was statistically significant.

Narrowing in on the changes of emotional valence over the five-month period in 2019 and in 2020 (Fig. 5.3), the two-time series show strikingly high emotional valence on the 1 January (Day 1). By gleaning through the Tweets of 1 January, it was confirmed that the high emotional valence was reasonable as a large number of “new year greetings” were tweeted. After the new year of 2020, the overall emotional valence dipped twice, one towards the end of January (26 January, Day 26) and the other right after the start of February (8 February, Day 39). It became more positive in the subsequent month before dipping again near the end of March 2020 (29 March, Day 89). Conversely, the emotion valence did not drop but rather improved somewhat in early April at the onset of CB and remained steady with minor fluctuation around

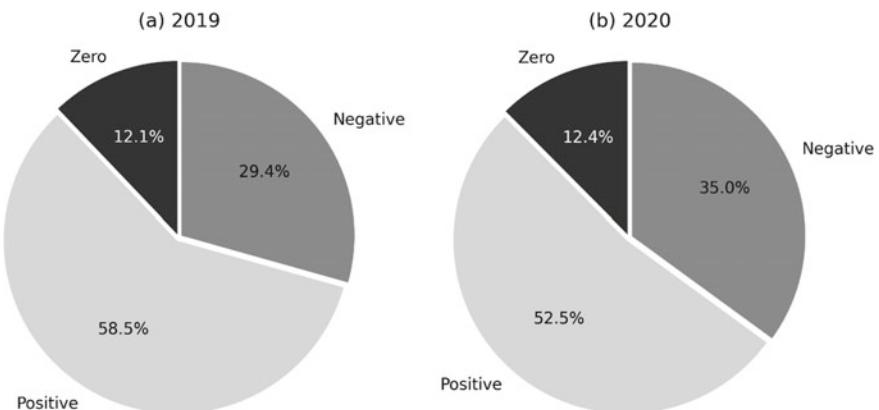


Fig. 5.2 Proportion of positive, negative, and zero emotional valence for **a** 2019 and **b** 2020

Table 5.1 Stationarity test of the emotional valence and the eight emotions for 2019 and 2020. Both trends and daily differences (differencing) were tested

Year	2019		2020	
Variable	Trend	Daily changes	Trend	Daily changes
EV	-4.72*	-7.69*	-4.99*	-11.57*
Trust	-2.26	-5.59*	-2.67	-6.54*
Anticipation	-8.38*	-8.17*	-8.66*	-12.51*
Joy	-4.21*	-5.21*	-1.68	-4.34*
Fear	-8.03*	-6.89*	-3.77*	-11.32*
Sadness	-7.48*	-6.27*	-6.81*	-7.80*
Surprise	-10.08*	-7.62*	-6.76*	-9.62*
Anger	-7.86*	-7.48*	-4.01*	-11.00*
Disgust	-3.99*	-6.05*	-7.50*	-10.61*

* $p < 0.01$; EV: emotional valence

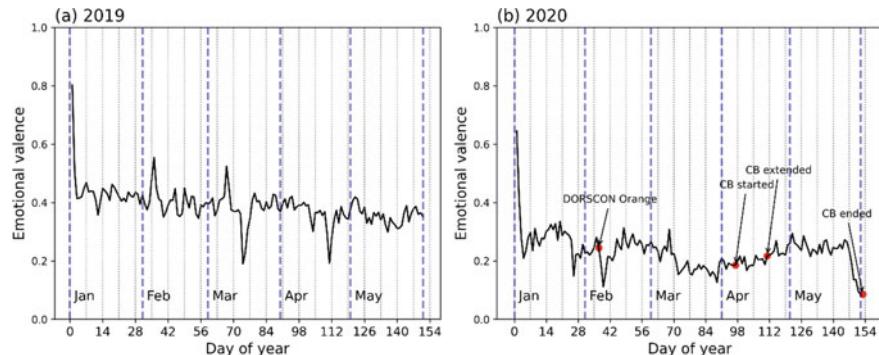


Fig. 5.3 Mean emotional valence by day-of-year of **a** 2019 and **b** 2020

0.25 until the end of CB (31 May, Day 152), when it dipped to the lowest point of the five months studied.

Breaking down the overall sentiment according to the eight basic types of emotions in both years for the five-month period exhibit similar temporal patterns (Fig. 5.4). Except for the 2019 trend for trust and the 2020 trends for trust and joy, all the other trends were stationary at 1% significance level (Table 5.1). Additionally, all the daily differences trends (i.e., differencing) were stationary (Table 5.1). The stationarity test results imply that, for the different types of emotions, although there existed certain emotional unstableness over time, there existed no significant emotional fluctuation day to day. There is a higher proportion of positive emotion types in both years, these include trust, anticipation, and joy. Overall, there is proportionately less negative emotion types, with more messages on fear and sadness than anger and disgust being observed in both years. In terms of the ordering by emotion types, there is no clear

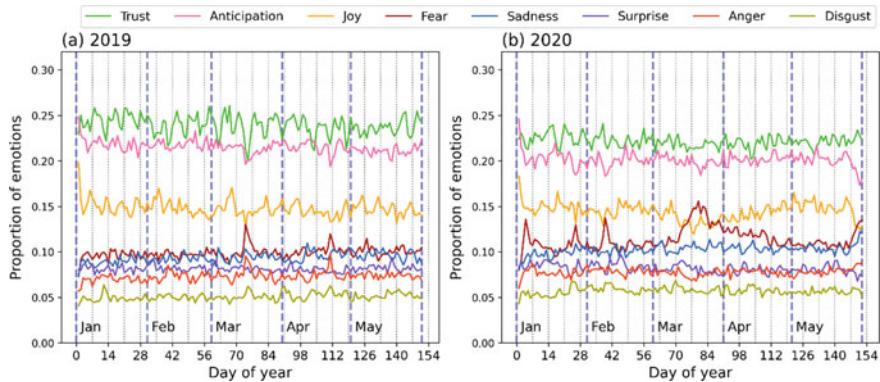


Fig. 5.4 Emotion proportion by types during COVID-19 of **a** 2019 and **b** 2020

distinction between the temporal patterns of 2019 and 2020 except that the negative types have a slightly higher proportion in 2020 as suggested by the overall sentiment analysis above. One peculiar deviation is the visible rise on the proportion of fear for about 10 days in mid-to-late-March, 2020, which may be related to the rising confirmed cases that prompted the implementation of CB in early April. In addition, it is observed that towards the end of CB, there was a visible drop of the proportion of anticipation and joy, along with a visible rise of the proportion of fear and sadness. Beside these increasing and decreasing trends, the proportion of fear had three visible one-day peaks during January (4 January and 26 January, Day 4 and Day 26) and February (8 February, Day 39) 2020, each had a corresponding drop of proportions in trust, anticipation, and joy.

5.3.2 Emotional Responses to COVID-19 in Singapore

In general, the results show that Singapore had a high level of trust and positive attitude during the 2020 pandemic (slightly less positive compared with the figures for 2019) (Figs. 5.2, 5.3 and 5.4). Singapore government's response to the crisis was swift. Within days of the Wuhan lockdown on 23 January (Day 23, the fear emotion increased after the Wuhan lockdown, see Fig. 5.4), free face masks were provided, social distancing was required, temperature checks were implemented in schools and work places, and hand sanitizers were provided where necessary. In the meanwhile, a network of about 900 designated clinics called Public Health Preparedness Clinics (PHPCs) was activated to help the authorities better detect and manage COVID-19 outbreak as the first line of treatment (ST 2020). Citizens and permanent residents would pay an affordable flat rate fee for consultation and treatment at PHPCs. This encouraged people to consult doctors and allowed the government to have enhanced surveillance systems and extensive testing operations to identify cases at their early

diagnoses. Anyone infected was sent to hospitals or community isolation facilities. People entering Singapore must serve a quarantine during which their health statuses were monitored; breaking quarantine will be a criminal offense.

Additionally, the government was actively identifying clusters and linkages of COVID-19 infections, tracking where every patient had been and identify everyone they had interacted with since becoming infected in order to break the chain of transmission. Furthermore, residents were encouraged to download a tracing mobile application for the authorities to check the places they had visited and who they had come in contact. With these measures in place, people in Singapore would know where the exposed risks were, so that they could adapt their behaviors and feel confident that the virus was under control. As a result of these measures, most of the workplaces, schools, resultants, cafes, and bars stayed open until the start of the CB.

Following several cases of infections without any links to previous cases or travel history to mainland China, Singapore raised its Disease Outbreak Response System Condition (DORSCON) level to Orange from Yellow on 7 February, 2020 (Day 38) (CNA 2020). This resulted in the fear emotion of the general public (Fig. 5.4) and panic buying and stock piling were reported. However, the government soon urged calm and assured that there was “no risk of us running a shortage of essential food or household items” (BT 2020).

Facing the wave of outbreak before the CB in early April, the proportion of fear emotion had increased and the proportion of joy emotion decreased (Fig. 5.4). The panic lasted only a short time, as the emotional valence curve of fear descended and the emotional valence curve of joy ascended very soon after. The level of community transmission (COVID-19 spreads in such a way that the source of origin of the infection is unknown) was still low during the CB. The dramatic increase of the infections mainly occurred within the foreign workers dormitories, but they had little interactions with the general public (Today 2020). The mortality rate in Singapore remained low (Worldometer 2020).

5.3.3 Sentiment Analysis by the Top 10 Languages

The multi-lingual nature of Singapore was evidently captured in Twitter messages during the study period, with a total of 105 languages found in the Tweets being analyzed in this study. The major ethnic groups in Singapore residents are Chinese, Malay, and Indian (DSS 2020) and their corresponding languages are, respectively, Chinese language, Malay language, and Tamil or Hindi languages. Using 18,000 Tweets as the threshold, the 10 most frequently used languages were identified (Table 5.2). The four languages of the major ethnic groups were contained in the top 10 languages. English was most frequently used, followed by Malay. The frequencies corresponded well to the fact that English and Malay are the main language and the national language of Singapore, respectively. The next language used is not Chinese even though ethnic Chinese accounts for 74% of Singaporeans as of 2019 (DSS 2020). Rather, the number of Tweets in Chinese is in the seventh place, suggesting the

Table 5.2 The top 10 languages used in COVID-19 related Tweets in Singapore

Rank	Language	Number of Tweets	Percentage (%)
1	English	5,433,782	84.51
2	Malay	259,485	4.04
3	Japanese	170,802	2.66
4	Indonesian	151,604	2.36
5	Korean	84,960	1.32
6	Filipino	61,019	0.95
7	Chinese	48,855	0.76
8	Tamil	40,946	0.64
9	Thai	30,894	0.48
10	Hindi	18,578	0.29
	Other languages	128,524	1.99
	Total	6,429,449	100.00

possibility that majority of ethnic Chinese in Singapore tweeted in English. Coupled with the fact that the use of Chinese trails Japanese, Indonesian, Korean, and Filipino, it is plausible that non-English Tweets corresponded to non-local communities. The Tweets distribution of the top 10 languages may reflect the penetration rates of Twitter in different ethnicities.

Focusing on the top 10 languages used in the Tweets during the COVID-19 period, Fig. 5.5 shows the daily average emotional valence over the study period. All except Japanese and Tamil languages were stationary at 1% significance level (Table 5.3); the daily differences trends were all stationary (Supplementary Table 5.1). The stationarity test results suggested that, for the Twitter users in different language groups, despite certain emotional unstableness over time, there existed no significant emotional fluctuation day to day. Because a large percentage of the Tweets were in English, the trend of emotional valence values for English (Fig. 5.5b) was similar to the overall trend (Fig. 5.5a), both of which had relatively low uncertainty. The fluctuation patterns of the other nine languages (Fig. 5.5c–k) were different from each other, suggesting that the people of different ethnicities were experiencing different sentiments during the first five months of COVID-19. Alternatively, sentiments reflected in English, Japanese, and Chinese Tweets, were relatively stable compared to the sentiments reflected in the other seven languages, which not only fluctuated over time but also exhibited different patterns in the five months studied. Both Japanese and Chinese had a moderate fluctuation in January, and the fluctuations stabilized from February to the end of May. The sentiments in Malay and Indonesian, the two languages that were highly cognate (Feng and Mark 2017), had moderate fluctuations but the former had a larger fluctuation in the first two months and a smaller fluctuation starting from March, whereas the latter had a larger fluctuation in the first two months, following by a small fluctuation in March, and a moderate to large fluctuation in April and May. This further reinforced the idea that the users of the two languages were likely from separate, local and non-local communities.

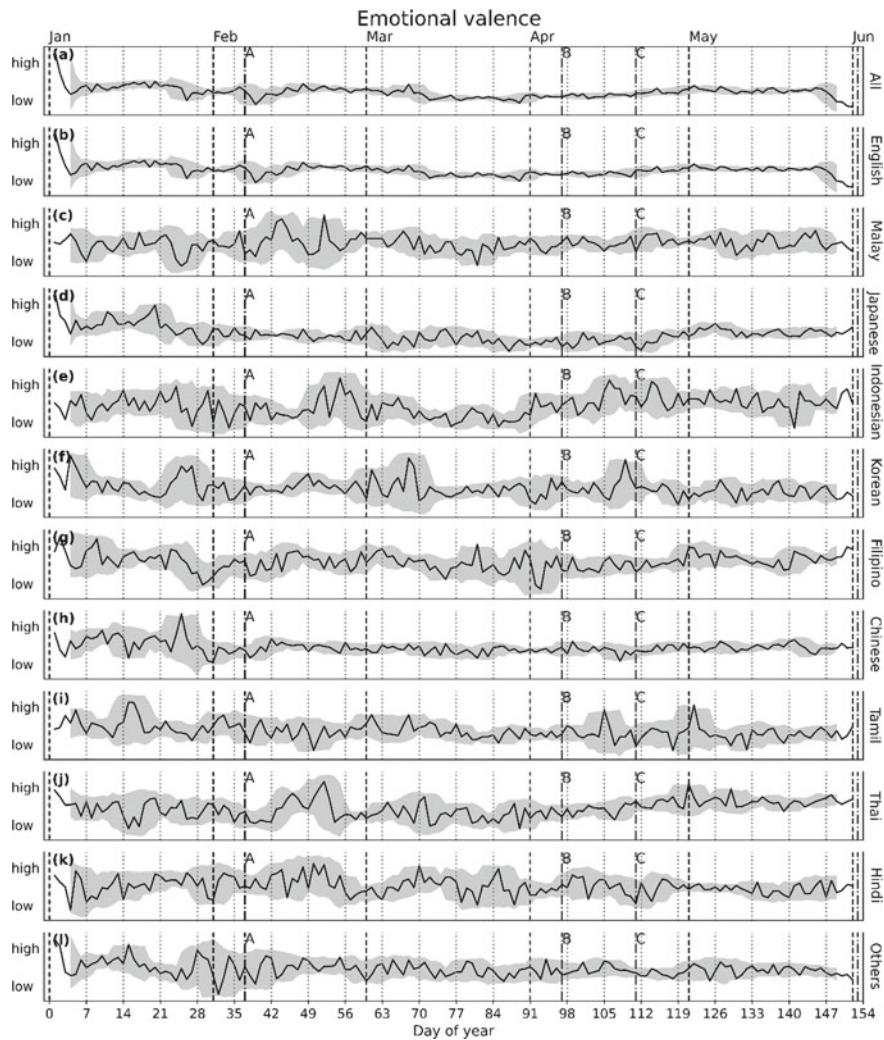


Fig. 5.5 Daily average of the emotional valence values for 2020 by the top 10 languages used. The black lines indicate the daily average value, and the gray area shows the trends of ± 3 days moving average with ± 2 standard deviation. The vertical dash-dot lines indicate the three key dates: **a** DORSCON level to Orange, **b** CB measures started, **c** the announcement of CB being extended to 1 June

The remaining five languages showed persistent fluctuations throughout the study period. In general, most languages exhibited a significant drop of emotional valence values in the last two weeks of January, some earlier, including Japanese (decrease starting from Day 20 to Day 29) and Thai (dropped to a valley on Day 15 to Day 17), and some later, including Indonesian, Korean, Filipino, Chinese (these four dropped

Table 5.3 The stationarity test for the trends of emotional valence and the eight emotions. The languages were labelled with the corresponding language code (ISO 639–3)

Lang	EV	Trust	Anticip	Joy	Fear	Sad	Surpr	Anger	Disgust
ENG	-5.12*	-2.76	-8.50*	-4.86*	-3.88*	-7.06*	-7.38*	-3.88*	-7.44*
MSA	-8.23*	-8.63*	-4.72*	-9.12*	-8.20*	-8.59*	-8.01*	-8.95*	-7.47*
JPN	-2.13	-1.73	-2.25	-1.83	-2.26	-2.05	-6.45*	-10.75*	-4.64*
IND	-3.48*	-5.47*	-7.38*	-2.45	-7.56*	-9.77*	-4.79*	-8.56*	-9.26*
KOR	-8.20*	-10.06*	-11.63*	-9.05*	-7.02*	-9.11*	-10.16*	-6.06*	-9.47*
FIL	-5.08*	-11.83*	-8.15*	-11.23*	-7.79*	-11.48*	-11.60*	-10.00*	-7.03*
ZHO	-8.18*	-9.12*	-10.87*	-2.29	-2.57	-9.32*	-9.31*	-5.77*	-7.96*
TAM	-2.52	-10.50*	-11.02*	-10.43*	-4.12*	-4.30*	-10.54*	-12.14*	-6.13*
THA	-7.10*	-12.24*	-7.95*	-6.07*	-1.13	-7.43*	-3.11	-10.88*	-3.11
HIN	-7.01*	-12.82*	-5.83*	-11.26*	-4.94*	-11.35*	-12.20*	-11.70*	-10.48*

* $p < 0.01$; EV: emotional valence; Anticip.: Anticipation; Sad.: Sadness; Surpr.: Surprise

Supplementary Table 5.1 Supplementary stationarity test for the daily changes of the language-based emotional valences

Lang	EV	Trust	Anticip	Joy	Fear	Sad	Surpr	Anger	Disgust
ENG	-11.19*	-6.44*	-10.89*	-7.73*	-11.60*	-7.48*	-9.96*	-11.27*	-7.81*
MSA	-7.85*	-6.42*	-11.27*	-5.36*	-7.46*	-6.13*	-7.03*	-7.67*	-7.17*
JPN	-7.59*	-8.80*	-8.57*	-8.60*	-7.71*	-8.80*	-6.28*	-7.85*	-8.95*
IND	-5.34*	-7.40*	-7.29*	-4.87*	-9.17*	-10.38*	-6.49*	-7.28*	-7.19*
KOR	-8.61*	-7.33*	-6.53*	-7.16*	-6.99*	-6.97*	-7.84*	-9.72*	-7.12*
FIL	-7.31*	-8.04*	-7.04*	-7.63*	-7.56*	-9.18*	-7.83*	-9.21*	-7.95*
ZHO	-6.96*	-6.97*	-4.62*	-5.24*	-9.25*	-6.02*	-7.93*	-5.37*	-8.27*
TAM	-6.70*	-9.29*	-9.57*	-7.75*	-8.66*	-12.76*	-9.27*	-6.15*	-7.43*
THA	-9.24*	-10.22*	-8.27*	-6.39*	-6.06*	-10.93*	-9.95*	-6.94*	-9.96*
HIN	-9.77*	-8.47*	-10.81*	-6.83*	-6.36*	-6.40*	-6.49*	-8.70*	-6.73*

* $p < 0.01$; Anticip.: Anticipation; Sad.: Sadness; Surpr.: Surprise

after the peak of Chinese New Year between Day 21 and Day 28), Malay (which had a valley on Day 25), and Hindi (dropped to a valley on Day 30 and Day 31).

The overall emotion pattern (Fig. 5.6b) reveals that more positive emotion types are more dominant given that they account for higher proportions of the Tweets, and the Tweets related to trust in every language is generally above 20%. Nevertheless, for the proportion of emotion by types, different languages have exhibited different patterns (Fig. 5.6). Japanese is unique in that the anticipation is always slightly higher than trust, while in other languages either trust is mostly higher than anticipation or their proportions are indistinguishable (Fig. 5.6). Tweets in five languages show clear “gaps” (generally 5–10%) between one or more positive emotion types and other emotion types. For the Tweets in English and Japanese, trust and anticipation are the

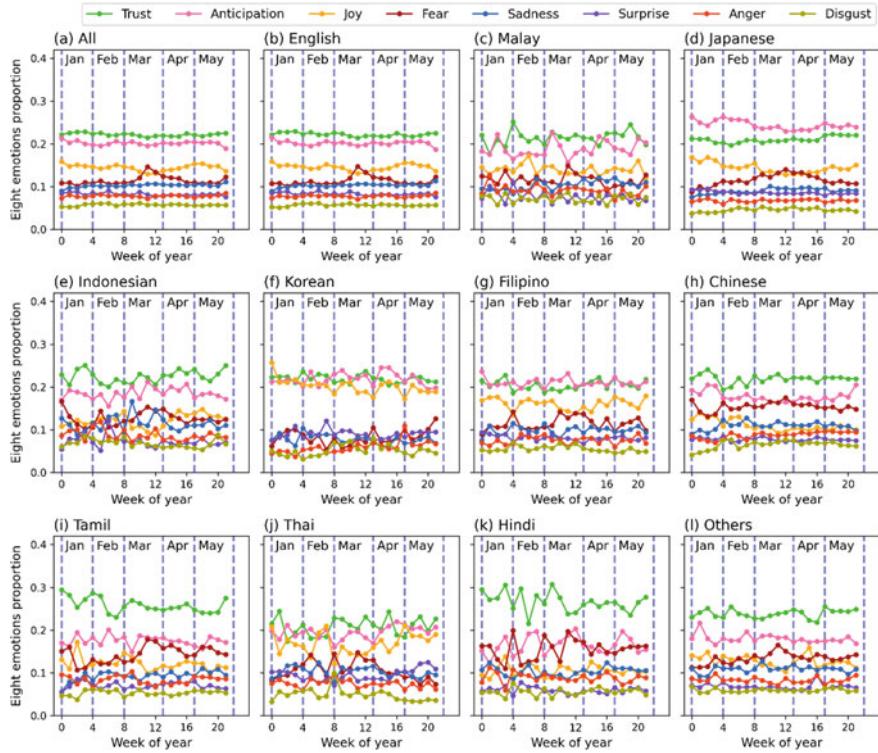


Fig. 5.6 Patterns of the weekly average proportion of the eight emotions by **a** all languages; **b–k** the top 10 languages analyzed; and **l** all other languages

two emotion types separated from the rest emotion types. For the Tweets in Korean, trust, anticipation, and joy are clearly more dominant than the rest emotion types. For Tamil and Hindi, trust stands out from the rest emotion types. The remaining languages exhibit mixtures of emotions although they may at some point exhibit separations between some positive emotion types with other types (Fig. 5.6). The most intriguing temporal pattern of individual emotion types across all languages is the rise of the proportions of Tweets related to fear around mid-March, except for Korean. The proportion of Tweets in Korean did exhibit a rise in March, but it is near the end of March. Most of the trends in Fig. 5.6 were stationary at 1% significance level (Table 5.3). All of the corresponding daily differences were also tested, and were stationary at 1% significance level (Supplementary Table 5.1). The stationarity test results suggest that, for the Twitter users in different language groups and for the different types of emotions, although there existed certain emotional unstableness over time, there existed no significant emotional fluctuation day to day.

To explore the similarities and differences of emotional responses among languages, each of the eight daily average emotion proportions was analyzed using hierarchical clustering. The results are shown in dendrograms (Fig. 5.7). Some

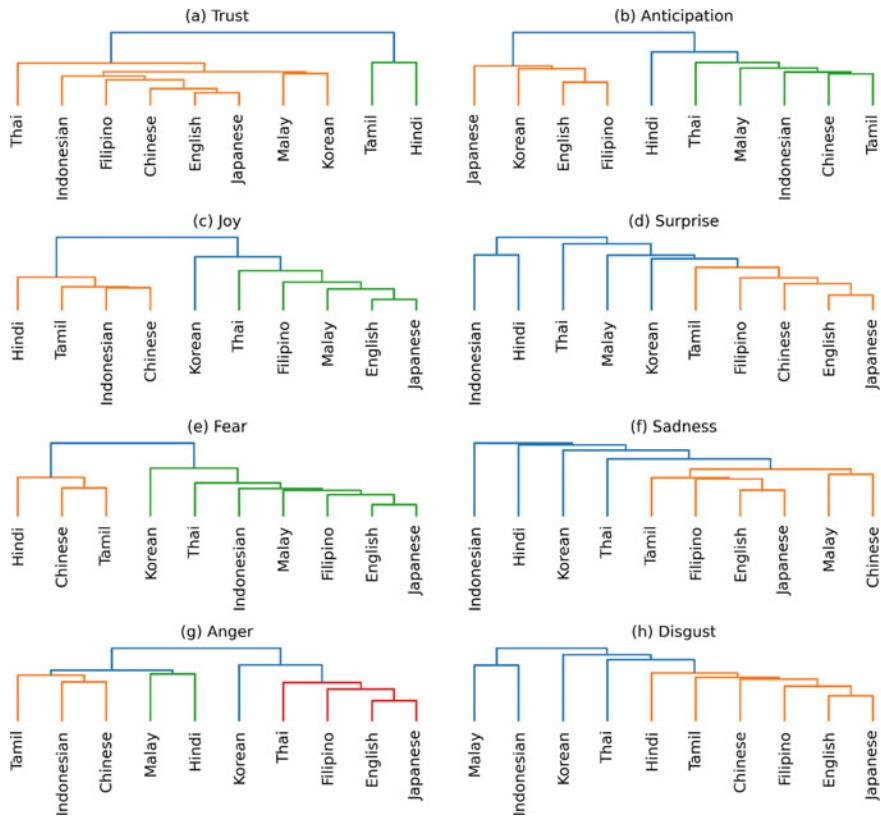


Fig. 5.7 The hierarchical clustering dendrograms for the daily average of the eight emotion proportions in Tweets

emotions exhibited similar clustering patterns such as (1) joy and fear as a group, and (2) surprise, sadness and disgust as another group, demonstrating the intra-group language similarity. The main difference between joy and fear is on Indonesian; it belongs to the left cluster (orange) in joy, but to the right cluster (green) in fear. Additionally, surprise, sadness, and disgust have a similar clustering pattern. For these three emotions, Tamil, Filipino, English, and Japanese have similar trends; they are among the lowest levels in all three emotions. Indonesian joins the tree in the last level for most of the emotions, indicating that it is more different from the other nine languages. Hindi also differs from the other eight languages for surprise and sadness emotions. English and Japanese usually appear at the lowest level, indicating their similarity in the trends in most emotions (except for anticipation, in which they appear in third lowest level).

5.3.4 COVID-19 Outbreaks in a Multi-ethnics and Multi-lingual Environment

Singapore has a high population density and many diverse ethnic groups. It is a norm among Singapore residents to live next to others who speak different languages. This cultural plurality is also evident in the virtual community of social media users (e.g., Twitter). From our data—the Tweets from people who live in Singapore—the Tweets in English accounted for 84.5%, the Tweets in nine other major languages represent 14%, and the Tweets in all other languages 2% (Table 5.2). Despite the high population density, the people in different language groups reported different emotional dynamics over the first five months in 2020 (Figs. 5.5 and 5.6), meaning that the people of different ethnics experienced the impacts of COVID-19 differently, even though they stayed next to each other. There are several possible reasons for the differences. First, the people of different ethnics received information from different types of media. Different sources of news may have different time-lag effects, causing the temporal differences in the emotional variations. Second, people from different ethnic community may embrace different attitudes in coping with the pandemic, hence experienced the outbreak differently and expressed their feelings using varying emotions. Different languages may express the same event but emphasises different sentiments. Some may describe more terms in “fear”, while others may use “angry”. Third, the non-resident population may be concerned about their families or relatives residing in their home and other countries that encountered different COVID-19 challenges compared to the situation in Singapore. Fourth, despite the highly diverse residential environment—the result of the social mixing programs in Singapore—people in the same ethnic group tend to gather around certain locations, and the clusters might appear in different locations in different time, leading to the differences in emotional dynamics.

The hierarchical clustering results (Fig. 5.7) showed the similarities and differences of languages in different emotion types. Some languages demonstrated similarities across multiple emotions, such as English and Japanese. Some emotions have similar language clustering patterns, such as joy and fear as a group, surprise, sadness, and disgust as another group, implying that the intra-group dynamics of the emotions were related. The inter-group differences indicated the underlying factors that stimulated the emotion dynamics were different. Those factors could be related to the four causes mentioned above or to the nature of the languages, which require further in-depth investigations.

While some similar clustering patterns exist among languages and emotions, the results showed more differences than similarities among the languages’ dynamic patterns in terms of the emotions. As such, the multi-lingual and multi-ethnic diversity perspective should be considered when investigating the impacts of the disease outbreak on the people in Singapore. Granted, existing efforts in fighting against

COVID-19 have considered this perspective by, for example, delivering useful information in multiple languages and dialects. The sentiment analysis results, nonetheless, showed that considering strategies amenable to the needs of different communities to minimize negative sentiments is still needed. One strategy may be to ramp up counselling, social, and other related services in a way that is sensitive to the difficulties facing specific communities so as to reduce the chances of further worsening of their mental wellbeing. The results and the methodological framework presented here demonstrates a means that enables the identification of the sentiment changes by communities or groups in a crisis event, such as the COVID-19 pandemic, and forms the basis from which resource provisions can be more precisely rendered and more adaptive to the needs on the ground.

Beyond informing policies, the sentiment analysis presented here suggested the importance the need to consider language as a factor in the analytical framework of crisis management, especially in regions with diverse ethnicities and possible diverse spoken languages, such as Singapore. As the diversity of ethnics in population is usually neglected in previous studies because of the difficulties in data collections and data processing, it is challenging if not possible to compare the expression of same emotion from two people from different ethnicities. Using Singapore as a case study and the languages of Tweets as a surrogate to represent different ethnic groups, it is demonstrated that considering ethnicity is indispensable in such analyses. Practically, this study also showcased one possible way to incorporate the notion of ethnicity in the understanding of the dynamic responses to a pandemic event.

5.4 Conclusions and Future Works

Social media platforms offer residents venues to express freely their feelings and opinions. In a crisis situation, mining the reactions of people on social media could be an efficient way to evaluate the effectiveness of the intervention measures and the decisions made by the related authorities. In this study, sentiment analysis was performed with Tweets posted in Singapore in 2019 and 2020, revealing the variations in the emotion dimension of human dynamics in the context of COVID-19. The study found that people in Singapore generally had a high level of trust and positive attitude facing the crisis, which could be a result of the government's smart crisis response. The study also revealed the differences in the emotions of the people in different language groups, highlighting the uniqueness in the emotional reactions of Singapore as a multi-ethnic and multi-lingual nation.

The significance of this study is at least threefold. First, understanding the social emotional impacts of the first and largest global pandemic event in the twenty-first century allows us to better understand how people cope with such events and how it affects people's well-being and social resilience. Second, the insights generated in this study may also facilitate the authorities' evaluation of the impact of their anti-epidemic measures on people's mental states. Third, this study contributes to

the literature on the policies that assist people to address possible emotional reaction to a pandemic.

For future works, first, this study focused on only the Tweets' text contents and the Tweets' visual contents (i.e., pictures and memes) were not included in the analysis. It would be interesting to take the visual contents into consideration in future works. Second, topic modeling such as Latent Dirichlet Allocation can be coupled with the sentiment analysis to discover more hidden details and provide more contexts for better explaining the sentimental patterns. Third, emotion is a complex and mixed product of human feelings, but this study models the sentiment of Tweets using only the eight basic emotions. Future studies are suggested to explore more on the emotions, e.g., the stronger or weaker emotions and the integrated feelings. In addition, as a data source, the sampling issue and bias of social media related to demographics, culture, user behavior, and even its own API (at most one percent sample of all Tweets are retrievable) should be addressed in future research.

Acknowledgements We thank the support from the GDAS' Project of Science and Technology Development (grant number: 2020GDASYL-20200103005), the National Natural Science Foundation of China (grant number: 41901330), and the Singapore University of Technology and Design (grant number: Cities Sector: PIE-SGP-CTRS-1803). The funding sources have no involvement in study design, collection, analysis and interpretation of data, writing of the report, and the outlet of the manuscript.

References

- Aslam, F., Awan, T. M., Syed, J. H., Kashif, A., & Parveen, M. (2020). Sentiments and emotions evoked by news headlines of coronavirus disease (COVID-19) outbreak. *Humanities and Social Sciences Communications*, 7, 23. <https://doi.org/10.1057/s41599-020-0523-3>.
- Barkur, G., & Vibha, Kamath, G. B. (2020). Sentiment analysis of nationwide lockdown due to COVID 19 outbreak: Evidence from India. *Asian Journal of Psychiatry*, 51, 102089. <https://doi.org/10.1016/j.ajp.2020.102089>.
- BT. (2020). Panic buying hits Singapore after virus alert raised. Retrieved August 13, 2020, from <https://www.businesstimes.com.sg/government-economy/panic-buying-hits-singapore-after-virus-alert-raised>.
- CAN. (2020). Coronavirus outbreak: Singapore raises DORSCON level to Orange; schools to suspend inter-school, external activities. Retrieved August 7, 2020, from <https://www.channelnewsasia.com/news/singapore/wuhan-coronavirus-dorscon-orange-singapore-risk-assessment-12405180>.
- Ćosić, K., Popović, S., Šarlija, M., & Kesedžić, I. (2020). Impact of human disasters and Covid-19 pandemic on mental health: Potential of digital psychiatry. *Psychiatria Danubina*, 32, 25–31. <https://doi.org/10.24869/psyd.2020.25>.
- DSS. (2020). Population and population structure: Singapore residents by age group, ethnic group and sex, end June, annual. Retrieved August 20, 2020, from <https://www.singstat.gov.sg/find-data/search-by-theme/population/population-and-population-structure/latest-data>.
- Feng, C.-C., & Mark, D. M. (2017). Cross-linguistic research on landscape categories using GEOnet names server data: A case study for Indonesia and Malaysia. *The Professional Geographer*, 69, 567–578. <https://doi.org/10.1080/00330124.2017.1288575>.

- Kim, I.-H., Feng, C.-C., Wang, Y.-C., Spitzberg, B. H., & Tsou, M.-H. (2017). Exploratory spatiotemporal analysis in risk communication during the MERS outbreak in South Korea. *The Professional Geographer*, 69, 629–643. <https://doi.org/10.1080/00330124.2017.1288577>.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., & Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54, 159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y).
- Li, S., Wang, Y., Xue, J., Zhao, N., & Zhu, T. (2020). The impact of COVID-19 epidemic declaration on psychological consequences: A study on active Weibo users. *International Journal of Environmental Research and Public Health*, 17, 2032. <https://doi.org/10.3390/ijerph17062032>.
- Lwin, M. O., Lu, J., Sheldenkar, A., Schulz, P. J., Shin, W., Gupta, R., & Yang, Y. (2020). Global sentiments surrounding the COVID-19 pandemic on Twitter: Analysis of Twitter trends. *JMIR Public Health and Surveillance*, 6, e19447. <https://doi.org/10.2196/19447>.
- Martín, Y., Li, Z., & Cutter, S. L. (2017). Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. *PLoS ONE*, 12, e0181701. <https://doi.org/10.1371/journal.pone.0181701>.
- Mohammad, S. M., & Turney, P. D. (2013). NRC emotion lexicon. National Research Council Canada, Canada.
- Neppalli, V. K., Caragea, C., Squicciarini, A., Tapia, A., & Stehle, S. (2017). Sentiment analysis during Hurricane Sandy in emergency response. *International Journal of Disaster Risk Reduction*, 21, 213–222. <https://doi.org/10.1016/j.ijdrr.2016.12.011>.
- Öztürk, N., & Ayvaz, S. (2018). Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis. *Telematics and Informatics*, 35, 136–147. <https://doi.org/10.1016/j.tele.2017.10.006>.
- Plutchik, R. (2001). The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89, 344–350.
- Seagal, S., & Horne, D. (2020). Human dynamics for the 21st century. *Systems Thinker*. Retrieved July 17, 2020, from <https://thesystemsthinker.com/human-dynamics-for-the-21st-century/>.
- Shaikh, S., Feldman, L. B., Barach, E., & Marzouki, Y. (2017). Tweet sentiment analysis with pronoun choice reveals online community dynamics in response to crisis events. *Advances in cross-cultural decision making* (pp. 345–356). Cham: Springer International Publishing.
- Shaw, S.-L., Tsou, M.-H., & Ye, X. (2016). Editorial: Human dynamics in the mobile and big data era. *International Journal of Geographical Information Science*, 30, 1687–1693. <https://doi.org/10.1080/13658816.2016.1164317>.
- ST. (2020). Coronavirus: Designated clinics being activated from today. Retrieved August 7, 2020, from <https://www.straitstimes.com/singapore/health/designated-clinics-being-activated-from-today>.
- Takahashi, B., Tandoc, E. C., & Carmichael, C. (2015). Communicating on Twitter during a disaster: An analysis of tweets during Typhoon Haiyan in the Philippines. *Computers in Human Behavior*, 50, 392–398. <https://doi.org/10.1016/j.chb.2015.04.020>.
- Today. (2020). 89% of migrant workers at dorms cleared of Covid-19: MOM. Retrieved August 7, 2020, from <https://www.todayonline.com/singapore/89-per-cent-migrant-workers-at-dormitories-cleared-of-covid-19-mom>.
- Wang, Z., & Ye, X. (2018). Social media analytics for natural disaster management. *International Journal of Geographical Information Science*, 32, 49–72. <https://doi.org/10.1080/13658816.2017.1367003>.
- WHO. (2020). World Health Organization: Mental health and COVID-19. Retrieved July 18, 2020, from <https://www.euro.who.int/en/health-topics/health-emergencies/coronavirus-covid-19/technical-guidance/mental-health-and-covid-19>.
- Worldometer. (2020). Coronavirus cases and deaths. Retrieved August 28, 2020, from <https://www.worldometers.info/coronavirus/country/singapore/>.

- Yan, Y., Chen, J., & Wang, Z. (2020a). Mining public sentiments and perspectives from geotagged social media data for appraising the post-earthquake recovery of tourism destinations. *Applied Geography*, 123, 102306. <https://doi.org/10.1016/j.apgeog.2020.102306>.
- Yan, Y., Eckle, M., Kuo, C.-L., Herfort, B., Fan, H., & Zipf, A. (2017). Monitoring and assessing post-disaster tourism recovery using geotagged social media data. *ISPRS International Journal of Geo-Information*, 6, 144. <https://doi.org/10.3390/ijgi6050144>.
- Yan, Y., Feng, C.-C., Huang, W., Fan, H., Wang, Y.-C., & Zipf, A. (2020b). Volunteered geographic information research in the first decade: A narrative review of selected journal articles in GIScience. *International Journal of Geographical Information Science*, 34, 1765–1791. <https://doi.org/10.1080/13658816.2020.1730848>.

Chapter 6

A Socio-Ecological Perspective on COVID-19 Spatiotemporal Integrated Vulnerability in Singapore



Chan-Hoong Leong, Wei Chien Benny Chin, Chen-Chieh Feng,
and Yi-Chen Wang

6.1 Introduction

The global COVID-19 pandemic has impacted public health, economics, and political confidence in profound and complex ways. Singapore reported its first COVID-19 case on January 23, a visitor from Wuhan, as the source of the outbreak. The first Singaporean infected case was reported on February 4, involving a customer service assistant who worked at a retail shop visited by Chinese tourists who were carriers of the disease. By early March, local transmission clusters were reported at places of worship, entertainment outlets, and clubs. On April 7, the city-state was placed on a partial lockdown, locally coined as a “Circuit Breaker” (CB), to contain the viral transmission. Other than those in essential services such as healthcare, public sanitation, and energy providers, all workers must stop work or do so online. Eldercare and Senior Activity Centres were closed, and school lessons transited to home-based learning. The movement of people was curtailed and strict social distancing rules were enforced to discourage interactions. Recreational facilities were closed, and restaurants could only offer take-away food.

At the start of the CB, a total of 1,481 COVID-19 cases, or a daily average of 50 to 100 cases, was reported in Singapore ([Ministry of Health \(n.d.\)](#)). The number of daily infections peaked at 1,426 new cases on April 20, with the majority from migrant workers’ dormitories. By early June, the CB had seemingly broken the chains of infection as the number of daily infections declined, averaging less than 10 community cases a day, and about 200 to 300 cases a day in the dormitories. On June

C.-H. Leong (✉)
Singapore University of Social Sciences, Singapore, Singapore
e-mail: chleong@suss.edu.sg

W. C. B. Chin
Singapore University of Technology and Design, Singapore, Singapore

C.-C. Feng · Y.-C. Wang
National University of Singapore, Singapore, Singapore

2, schools were gradually re-opened (Phase 1 opening), and by June 19, the CB was partially lifted, where restricted forms of shopping, restaurant dine-in, office work, and recreational activities were permitted (Phase 2 opening). There were a total of 41,615 cases of COVID-19 infections as of June 19, 2020.

The evidence so far suggests that the disease spreads directly and indirectly from human-to-human contact. An infected person may pass on the virus to the next individual through the surface of an object that the person come in contact with, and more likely so if the latter touches the nose, mouth or the eyes. In contrast to past global epidemic like SARS, or MERS, COVID-19 is an airborne and highly virulent respiratory disease with long incubation period (Morawska and Milton 2020). Indoor venues, places with poor ventilation, and where people linger around longer are more prone to transmission (Lewis 2020). The stealthy and exceedingly contagious respiratory virus makes this disease a potential deadly strain.

Scientific knowledge in this pandemic is still limited, particularly how the types of social and built environmental attributes may contribute to epidemiological risk. The goal of this chapter is to develop a measure of spatiotemporal vulnerability in Singapore. Specifically, we use the data on movement restrictions from the COVID-19 outbreak to highlight how an epidemic may evolve over space and time due to the inherent risks to certain subgroups and the type of amenities that attract human congregation and transmission. The chapter will first give an overview on the socio-built environmental characteristics that are linked to community resilience and pathogenic infections, follow by a review on how temporal fluctuations in movement of people may shape local vulnerability. This study coalesces different data sources using demographic profiles, human mobility, and urban spatial functions to produce a new perspective and measure on spatiotemporal vulnerability.

6.2 Literature Review on Environment and Vulnerabilities

There is a historical curiosity to understand how a location is linked to human behaviours, motivations, and wellbeing, and by extension, the place's vulnerability to devastations from hazards. Policymakers, urban planners, disaster response management, and public health officials have a strategic interest to find out what are the social or built environmental features that can shape adaptation as this knowledge can provide the conduit to assess and predict disaster, and provide the foundation for prevention and recovery (Adger 2006; Cutter 1996; Cutter et al. 2003, 2008). Built features such as pedestrian friendly walkways, aesthetically designed corridors, and access to facilities are known to encourage increased mobility, social interactions, and community participation (e.g., Leslie et al. 2007); places with abundance of nature, vegetation, and sunlight are found to improve healing and rejuvenation from psychological distress (e.g., Bratman et al. 2012; McMahan and Estesb 2015).

Alternatively, social features such as a neighborhood's demographic profile on age, income, density, and racial composite, give policymakers insights on potential areas that are susceptible to socio-economic disruptions, experience higher risk to

mental and physical health, or face greater impediments to post-disaster recovery and humanitarian aid. The Social Vulnerability Index (SVI) developed by the US Centers for Disease Control and Prevention (CDC) demonstrates how information about the social environment is harnessed to identify vulnerable locations and where the deployment of emergency shelters are needed most (e.g., Flanagan et al. 2018). SVI composes 15 US census tract indicators, covering four demographic and geographical attributes, namely socio-economic status, household composition, minority status, housing and transportation. Overall, locations with a higher concentration of elderly, minority races, and lower income households are predisposed to greater vulnerability.

Recent evidence from the US and Europe on the COVID-19 outbreak corroborated this observation. Places that are densely populated, with more senior citizens, young children, non-whites, working class residents, and people with pre-existing medical conditions reported a higher infection rate and mortality compared to middle-upper class neighborhoods (e.g., Karaye and Horney 2020; Mueller et al. 2020; Nayak et al. 2020; Snyder and Parks 2020; Yonker et al. 2020). The findings reflect the conflation between race, class, and occupational hazards, as more minorities live in disadvantaged estates, employed in lower-paying occupations, lack access to healthcare, and medical insurance. Social vulnerabilities are a product of structural socio-economic inequalities, and their disabilities are amplified during a crisis, with the least resources to cope with the ensuing disruptions (Cutter et al. 2003).

Last but not the least, the conceptual framework on spatial vulnerability to pandemics has helped informed locale emergency management, including heat-related injuries (Lehnert et al. 2020), floods (Rufat et al. 2015), and earthquakes (Schmidlein et al. 2011). Notwithstanding these contributions, many of the spatial vulnerability indicators are static and time-invariant. It does not capture the dynamics of human mobility, which has a temporal and class dimension attached. In other words, the risk changes across space and time. The COVID-19 outbreak is also different from previous epidemics in a few significant ways. First, there is a greater recognition and practice of health preventive measures such as safe distancing, work from home, and travel restrictions. Specifically, young children, elderly, and those with medical conditions are confined at home during the early stages of lockdown to minimize interactions.

Second, COVID-19 is considerably more infectious than the previous known diseases such as SARS and MERS (Petrosillo et al. 2020). The lessons learned from the SARS pandemic, such as isolation of suspected and confirmed cases, are considerably less effective this time. Exposure to the infectious disease, specifically in places or Point of Interest (POI) where people congregate remain the crucial determinant that underscores transmissions. These include supermarkets, shopping malls, and public transportation nodes. The centrality (i.e., intensity of linkages to other locations/nodes) and transitivity of locations (i.e., clustering with other locations that are also directly linked up between them) predicted the severity of infections. This pattern of transmission is also reflected at the regional level. Studies on the Black Death plague for instance, found that cities that are major intersection networks are significantly more vulnerable (Gómez and Verdú 2017). Proximity to spatially

vulnerable POI and networks of densely connected nodes increases the probabilities of local infections (Bogich et al. 2013; Tatem et al. 2006).

Third, there is a pungent public discourse on socio-economic inequality in the COVID-19 pandemic (Schwartz and Cook 2020). Most of the higher income earners, the professionals, managers, and executives, can perform their job remotely, i.e., work from home. Conversely, those employed in provision of essential services, such as public transport, postal and food delivery, building maintenance, energy providers, law enforcement, and water and waste management services, still have to commute to work daily and exposing themselves to the risk of infection. This narrative is not unique to Singapore. Similar debates can be heard in other infected countries. Taken together, the evidence suggests that a lockdown may have dissimilar spatiotemporal impact to people from different socio-economic background. In addition, due to the compact nature of city-states like Singapore, we expect the overall risk to be affected by both local and proximity effects. Specifically, neighboring areas that are considered vulnerable will have a spill over impact on local residential risk.

In summary, it is important to develop an Integrated Vulnerability (IV) model to address geographic variation on risk assessment and disruption according to the CB timeline (i.e., lockdown) in Singapore. This variability is a consequence of resource inequality across space that is dynamic but highly influential in light of the virulent nature of COVID-19. Residential areas with disproportionately more vulnerable residents (i.e., young and elderly residents), low income families (using housing prices as surrogate), concentration of POIs/amenities and built features known to attract human congregation and disease transmissions (e.g., public transports, markets) will have a higher spatial vulnerability. In line with expectations from network theory where higher concentrations of centrality and transitivity accelerated pandemic outbreaks, neighborhoods that are contiguous to other estates with similar spatial vulnerabilities will likely produce a higher risk than remote, dis-contiguous areas with limited exposure to the hazardous POIs (i.e., spatial lag effect). Spatial vulnerabilities however will change in line as the rules governing CB change over time, as more business, work, and school activities gradually resume prior to the lifting of the lockdown.

6.3 Data and Method

The measure of IV to COVID-19 is segmented by monthly data and using subzone as the basic enumeration unit. Singapore is a compact island city-state with an area of 720 km², and a population of five million people, including long-term and transient residents. A subzone is the smallest zonal unit used in urban planning. It is commonly centered around a focal point like a neighborhood center¹. There are 323 subzones with varying sizes, population, land use, and demographic profiles. Demographic

¹About 10 subzones make up a planning area, a broader division of regional towns with about 70,000 to 100,000 residents.

data for a subzone can be obtained from Singapore Department of Statistics (www.singstat.gov.sg). The data includes the configuration of housing types, and population segmented by gender, race, and age. The map representing all subzone divisions and its demography showing population density and the proportions of young, working age adults, and elderly residents can be found in Fig. 6.1.

Monthly data on public transport utilization and housing resale prices is used (Table 6.1). The former shows travel pattern among commuters of buses and trains and is obtained from Ministry of Transport, Singapore. Majority public transport commuters rely on a store-value card when using the transport system. The points of entry and exit in each train station or bus stop are registered in a national transport database, and the consolidated data reflects the intensity of human activities in the various subzones. This movement elevates the risk of COVID-19 transmission as the virus is known to spread from human to human in close proximity.

Housing resale prices are obtained from the Ministry of National Development. This data includes all property transactions over a three-year period between 2017 to 2019. More than 9 in 10 properties in Singapore are apartments co-located in the same block sharing the same postal code. Less than 10% of all properties are

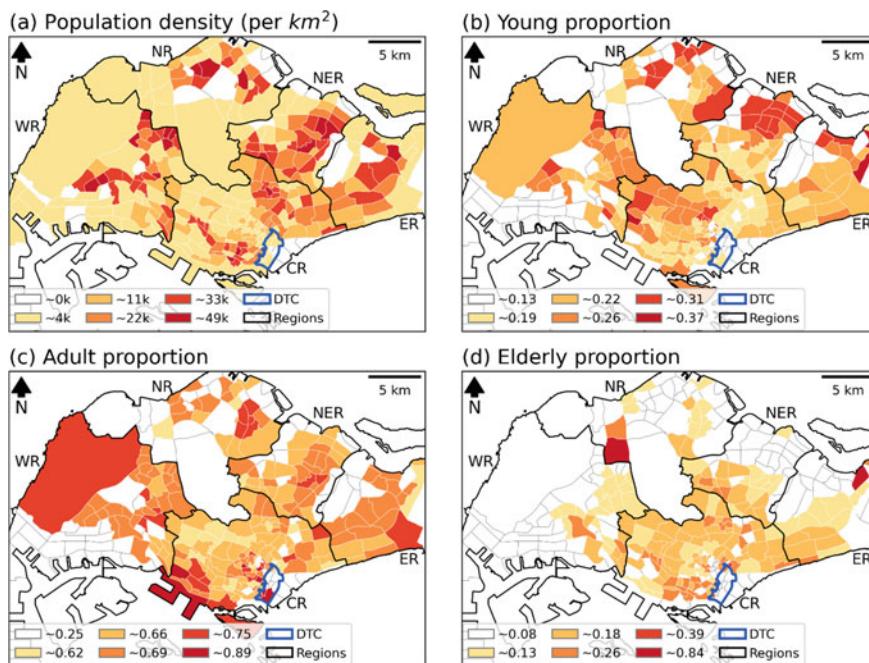


Fig. 6.1 Map of Singapore representing **a** population density, **b** proportion of young residents (age 19 and below), **c** proportion of adult residents (between age 20 and 64), and **d** elderly residents (age 65 and above). The polygon in blue shows the Downtown Core, which constitutes the bulk of the Central Business District. WR, NR, NER, ER, and CR, represent Western region, Northern region, Northeast region, Eastern region, and Central region, respectively

Table 6.1 Spatial and temporal information used in this study

Category	Details	Static/dynamic
Demographic	Population in each subzone, segmented by 5-age-interval	Static
Public transport flow	Number of passengers from one bus-stop/train-station to another	Dynamic
Residential	Resale housing prices by building blocks	Static
Facilities POI	12 types of POI, classified under four amenity categories: – commercial: shopping malls (199), supermarkets (324), wet markets/hawker centers (114); – transportation: train stations (183), bus interchanges (48), bus stops (5,045); – facilities: sport facilities (182), medical facilities (54), community clubs (112) – other: worship places (654), coffee shops (1,136), community in bloom parks (1,024)	Static
Urban planning	9 categories: residential, business/commercial, agricultural, transportation, education, health/medical, religious, public open space, and other	Static

classified as a landed unit (i.e., own a plot of land), some of which have its own postal code or sharing one with a handful of other dwellings. The average transaction price for each postal code is computed. This data serves as surrogate on socio-economic status, where a higher average transaction price reflects middle-upper class, and vice versa.

To evaluate the land use mixture of each subzone, the data on land use categories was extracted from the Singapore Master Plan 2019 provided by Urban Redevelopment Authority of Singapore. The original data came in Keyhole Markup Language (kml) format, which contains detail polygons of building blocks, street profile, etc., and a column of land use description for each polygon. A total of 33 land use descriptions were provided, which contained subdivisions of some major land use categories or a mixture of them (e.g., business, residential and residential with commercial on first storey).

6.3.1 Data Pre-processing

The obtained data was converted to subzone-based attributes and unit interval values for the calculation of local risks. There are six measures in each subzone, including population density (PD_i), elderly population density (ED_i), property resale transactions (PT_i), accessibility of subzone (Acc_i), land use diversity (LE_i), and population activity intensity (IF_i).

6.3.1.1 Demographic and Socio-Economic (PD_i , ED_i , PT_i) Data

Two demographic aspects were included in the measurement of local risk—population density (PD_i) and elderly density (ED_i). The total subzone resident population and elderly population (age 65 and above) were divided by the area size to obtain the two density values for each subzone. The two population density values were converted to unit interval using their minimum and maximum values. We used the average housing resale price to capture the socio-economic aspect of each subzone. For those subzones with no housing resale data, a simple estimation based on 10 nearest neighbors and weighted by the spatial distance was performed. The property transaction value for each subzone (PT_i) was calculated as the natural logarithmic of the (estimated) average resale price, with a conversion to unit interval using the minimum and maximum values.

6.3.1.2 Measurement of Accessibility (Acc_i)

This study used potential accessibility to some commonly utilized facilities to capture the possibility of viral exposure. In total, 12 types of POI were included, classified under four amenity categories (Table 6.1). The number of POIs for each type varies considerably. There are nine facility types with less than 1,000 POIs each, these include medical facilities (54), train stations (Mass Rapid Transit (MRT)/Light Rail Transit (LRT), 183), bus interchanges (48), sport facilities (182), supermarkets (324), wet markets and hawker centers (114), shopping malls (199), places of worship (churches, Chinese temples, mosques, and Hindu temples, a total of 654), and community clubs (112). There are three facility types with more than 1,000 POIs each, namely bus stops (5,045), coffee shops (National Environment Agency, 1,136), and community in bloom parks (1,024).²

For the first group (i.e., less than 1,000 POIs in each type), because of limited accessibility, residents must go to the nearest facility to access the specific function, such as healthcare or transportation. Thus, for each of these facility types, we calculated the accessibility (Acc_i^k) as the natural logarithmic of the inverse square distance from each subzone centroid (i) to the nearest facility of the specific type (k), or in mathematical form: $\ln((dist_i^k)^{-2})$. For the second group, these facilities are more accessible island wide; residents have options if they need to use the specific utility. Thus, for each of these facility types, we count their number within a subzone and compute the natural logarithm of the count as a measure of accessibility, or in mathematical form: $\ln(count_i^k)$. For each subzone, the 12 accessibility values are scaled to unit interval using the minimum and maximum values of each facility type, and the arithmetic mean of the twelve scaled values is used as the integrated accessibility (Acc_i).

²A “coffee shop” is colloquial term that refers to a place that offers dine-in food, drinks and a place where people gather to interact.

6.3.1.3 Land Use Diversity (LE_i)

Land use diversity measures the mixture of activities in each subzone. The data on 33 land use descriptions was categorized into nine major land use types (including residential, business/commercial, agricultural, transportation, education, health/medical, religious, public open space, and other) for analysis. For each subzone, the area for each of the nine land use types was calculated and integrated into a diversity measurement (LE_i) based on the Shannon's entropy equation (Shannon 1948), and normalized using natural logarithmic of nine. The normalized entropy values range within unit interval.

6.3.1.4 Population Activity Intensity (IF_i)

The subzone demographic data reflects the residential population and where they live, but not where they work or study. To capture the daily activity of the population, we included a measure of public transport utilization, specifically the incoming flow of bus and train users as a surrogate index of the population activity intensity. Figure 6.2 shows the daily average number of trips for trains (including MRT and LRT) and buses from December 2019 to July 2020, as a reference to understanding the changes in commuting patterns in Singapore during the COVID-19 epidemic. At the start of the global outbreak in February and March, there is a significant drop in weekend trip counts as residents reduce non-essential outdoor activity to minimize social interactions. The CB was implemented on April 7 and lasted till June 2. Public commuting activity for both weekdays and weekends dropped significantly in the

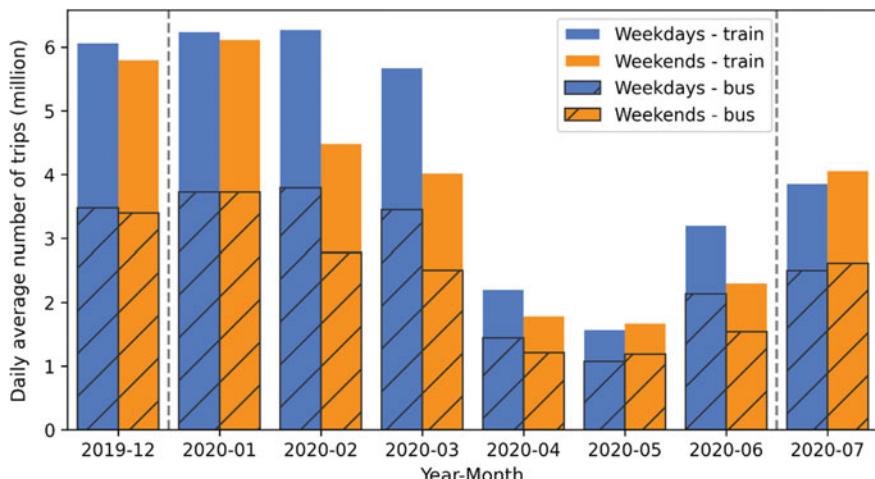


Fig. 6.2 The daily average number of trips for trains (MRT and LRT) and buses from December 2019 to July 2020

months of April and May. The first two phases of reopening took place on June 2 and 19, respectively. By the end of July, the number of trips was still considerably lower than usual compared to January data. We used the weekdays' incoming flow to a subzone as a surrogate variable to measure the population activity intensity in each subzone (IF_i). The incoming flows were converted into unit interval using the minimum and maximum values throughout the six months so that the resulting values were comparable between months.

A local vulnerability measure (LR) is first defined based on a non-weighted aggregate of local attributes in each subzone, e.g., demographic profile, socio-economic status, accessibility to facilities, land use, and population activity intensity. Population activity intensity assesses the number of people coming into each subzone. LR also incorporates POIs known to have high infections, such as shopping malls, supermarkets, and popular wet markets. Demographic profiles, socio-economic status, accessibility to facilities, and land use indicators are static and time invariant, activity intensity however fluctuates over the period when the lockdown was initiated, to the different phases of reopening.

$$LR_i = \frac{1}{6} (PD_i + ED_i + Acc_i + LE_i + PT_i + IF_i)$$

6.3.2 Metric of Spatial Lag Vulnerability

Spatial lag effects of vulnerability comprised two distinct components. First, the transition of the vulnerable effects due to human mobility. As the pandemic is transmitted primarily through human contact, the movement of people across space has implications to vulnerability, and proximity to active subzones and places with higher local risk would contribute to greater IV. Second, the spatial lag effects may differ for different groups of people. Working age adults in particular, are more mobile and subzones with a higher proportion of working age adults would have a higher transition of vulnerability.

6.3.2.1 Transition of Vulnerable Effects

The spatial lag effects are observed as the transition of elements from one place to another as a result of population flow (Fig. 6.3). In normal conditions (Fig. 6.3a), where there is no restriction of movement, population flow follows the gravity or radiation model of human movement patterns, both leading to an intense flow of people between neighboring subzones. Subzones with intense interactions would have greater likelihood of cross infections, and thus the condition of one location may be transferred to its neighboring subzones. During the CB period, the movement of people and interactions between places that were further apart were restricted

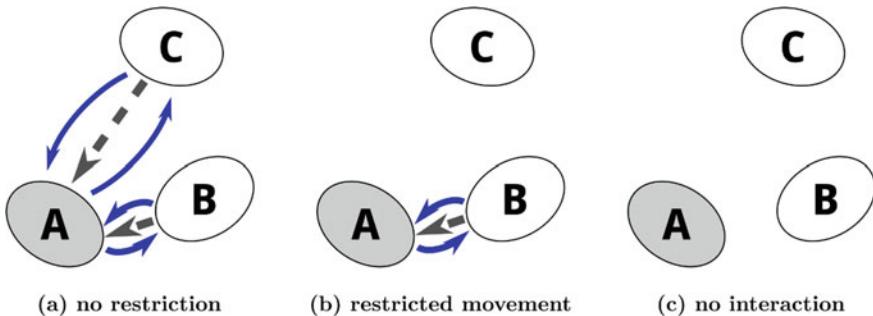


Fig. 6.3 An illustration on the transition of vulnerability between neighborhoods A, B, and C, with A as the target location: **a** no restriction, thus strong transition (dashed arrows) occurred with intense human exchanges (blue arrows); **b** when long distance movement is restricted, only residents in nearby neighborhoods have interactions with A, hence the shorter distance which contributes to the transition of vulnerability from B to A; and **c** when the CB is placed, no one is allowed to move across regions, hence zero transition

(Fig. 6.3b) or completely cut off across zonal boundaries (Fig. 6.3c). The analyses thus aimed to capture the transition of vulnerability, incorporating the spatial lag effects due to proximity and zonal contiguity.

6.3.2.2 Movements of Different Group of People

During CB, only providers of essential services are permitted to travel. This analysis examines the spatial lag effects of different age groups according to their IV exposure. The resident population was categorized into three age groups: young (19 and below), working age adults (20 to 64), and elderly (65 and above). The frequency distributions of the subzones by age groups are shown in Fig. 6.4. Most of the subzones have about 66.1% of working age adults, 20.5% of dependent young residents, and 13.5% of elderly residents. Since the working-age adults are expected to be economically active and spatially more mobile than both young and old residents, the potential transition of vulnerability caused by adult people would be greater than the other two age groups.

Figure 6.5 illustrated the different spatial lag effects of varying movement restrictions for different age groups in hypothetical scenarios. Under normal circumstances, everyone is free to move around, but most people tend to move within a distance range for their daily routines. In Fig. 6.5b, we assume everyone can move freely without restrictions and for any distance. During the early stage of the outbreak (Fig. 6.5c), the elderly is considered the most vulnerable group because their higher mortality rate; they were advised not to travel long distances and avoid crowded places. When the schools were closed (Fig. 6.5d), the younger residents were confined to home-based learning and hence zero distance travelled. Under a CB (partial lockdown) (Fig. 6.5e), most people are not allowed to leave their home except for essential workers. Under

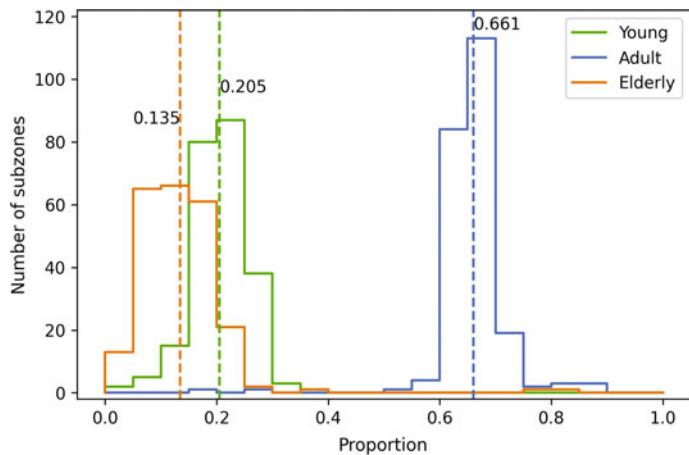


Fig. 6.4 The proportion of the three age groups in subzones. The values adjacent to the three colored dashed lines show the average proportion of the corresponding age group

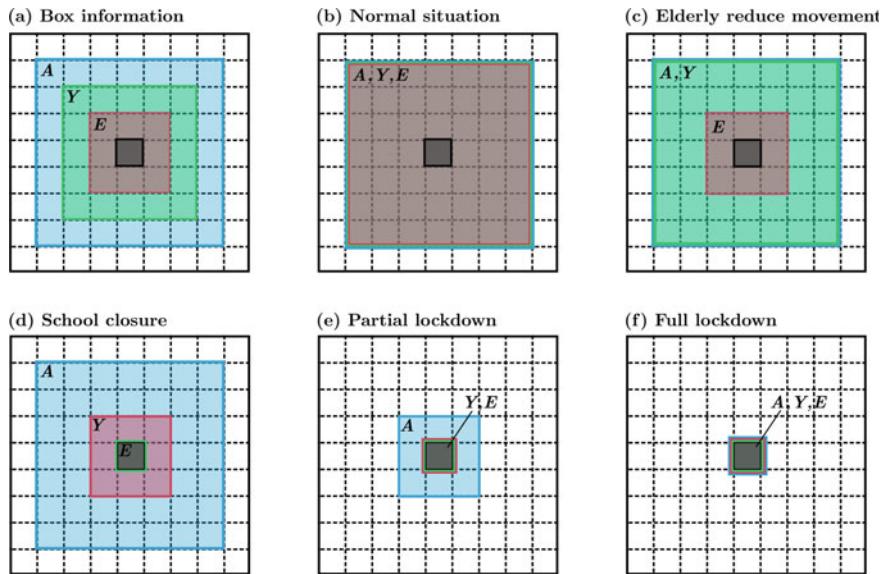


Fig. 6.5 An illustration of the transition effects by multiple age-group as a result of different restrictions and mobility between age groups: **a** The target location (dark gray), shows the moving distance for elderly (red, labelled with E), for working adults (blue, labelled with A), and for young students (green, labelled with Y); **b** to **f** indicates the transition effects of the three age groups from the varying neighboring distance to the target location

a full lockdown (Fig. 6.5f), no one is allowed to cross zonal boundaries, and the effects of neighboring areas would be zero.

For each age group (G), the spatial lag of vulnerable for a subzone i (Lag_i^G) is calculated with the following equation, which multiply the proportion of the respective population to total population ($prop_i^G$) with the average local risk of its neighboring subzones for the age group (using the neighboring weight $W_{i,k}^G$, which is 1.0 if the distance between two subzones is less than or equal to the moving distance for the age group, otherwise zero). Note that the spatial lag would have different neighboring weights ($W_{i,k}^G$) for different age group (young, adult, elderly).

$$Lag_i^G = prop_i^G \times \frac{\sum_k (W_{i,k}^G \times LR_k)}{\sum_k W_{i,k}^G}$$

where subscript i , subscript k , and superscript G represent a subzone, the contiguous subzones, and age group, respectively. Following this, we integrate the spatial lag effect in the computation of an IV index. For each subzone, the IV index:

$$IV_i = p(LR_i) + (1 - p)(Lag_i^{elderly} + Lag_i^{adult} + Lag_i^{young})$$

The IV in the equation reflects the vulnerability of a location; it is a function of its local risks (LR_i), and the neighboring vulnerability conditions. Local and neighborhood conditions however pose different impact on overall vulnerability, and this depends on the strength of their effects. There is a p parameter, representing the strength of effects from local environment, and $(1-p)$ represents the strength of effects from the neighboring subzones. Thus, a p -local value of 1.0 demonstrates complete local effect, no spatial lag; a p -local value of 0 demonstrates complete neighborhood effect, i.e., vulnerability is affected by characteristics of contiguous subzones.

6.3.3 Study Design

To explore the IV of disease outbreaks in Singapore and the effects of the parameters (moving distances for varying groups and the local effect), the monthly scenario analysis and sensitivity analysis were performed.

6.3.3.1 Monthly Scenario Analysis

In this section, we analyzed the travel patterns of Singapore residents from January to June of 2020, and provided an outline on the model assumptions for LR, spatial lag models, and IV. Table 6.2 shows the parameter configurations. In January, when the first few imported cases were identified, all moving distances were set to 25 km,

Table 6.2 The parameters configuration for the monthly scenario analysis

Month	p-local	Distance-young (km)	Distance-adult (km)	Distance-elderly (km)
January	0.5	25	25	25
February	0.5	20	25	10
March	0.5	10	25	10
April	0.25	0.5	25	0.5
May	0.2	0.5	25	0.5
June	0.35	10	25	10

a large enough radius that covers most of the subzones, signifying no restriction of movement. In February, and in light of many unlinked community cases (i.e., unknown sources of transmission), the government issued advisories against unnecessary outdoor events and traveling, and thus the reduction in the moving distance of the elderly and young residents.

The rapid increased of cases in late March led to the CB in April and May, and this restricted the mobility of young and elderly people. Only a short commute to the neighborhood grocery stores and food centers were allowed, and thus the travel distance for the two groups was set as 0.5 km. The travel distance for working age adults was set to 25 km (non-restricted) as some of them are essential workers who provide critical services to the country. Although a significant reduction of travelling was observed, there were approximately 2 million trips recorded during the CB period, and hence the effects of spatial lag for this category of residents (i.e., working age adults) was assumed to be similar to January for April and May.

The partial lifting of CB began in early June, where some were allowed to go back to work and school, although the businesses and individuals were urged to restrict their outdoor activities and social gatherings, and to work from home as far as possible. For this period, the moving distance of young and elderly people was increased to 10 km. Prior to the CB, the p-local effect is set to 0.5, i.e., half of the vulnerability of a subzone depended on its own local risk conditions, and the other half from the neighboring subzones. During CB, the p-local was reduced to 0.25 (April) and 0.2 (May), as people were not supposed to leave their homes. The p-local increased to 0.35 for June, as Singapore gradually reopens. As aforementioned, the local risk for each subzone was assessed using six variables, and where five of them were static. Only one of them, the population activity intensity, was based on the monthly public transport flow data.

6.3.3.2 Sensitivity Analysis

The following sensitivity analysis tested the different distance parameters and adjusted the local effects (p) incrementally to understand how the IV index changes. We used only the January public transport flow data as the population activity intensity variable, because the flow patterns for January was more stable compared to

the other months. Since the population proportion of young and elderly are lower than the working adults, the effect of the changing distances for young and elderly is expected to be low. As such, we combined the two age groups to share a moving distance³ parameter in the sensitivity analysis. The range of the parameters' settings is as follows:

- effect from local (p-local): 0.0, 0.25, 0.5, 0.75, 1.0.
- moving distance for adult: 50, 100, 500, 1,000, and 3,000 m.
- moving distance for young/elderly: 50, 100, 500, 1,000, and 3,000 m.

In a pre-analysis, longer distances for young, adult, or elderly (from 3 to 30 km) showed similar patterns as the 3 km, indicating less sensitive effects for a further distance. Therefore, we present the results for emphasizing the fluctuation in the smaller distance range.

6.4 Results

6.4.1 Spatial Distribution of Local Risk

Local vulnerability (i.e., no impact from contiguous or proximal subzones) for each subzone is assessed based on local demographic and socio-economic profiles, spatial accessibility to essential facilities, and monthly incoming flows as a surrogate index of population activity intensity. Figure 6.6 shows the spatial distribution of local vulnerability. The subzones were categorized into six groups using the same set of break values derived from Jenk's natural breaks method, using the local vulnerability values of all six months. The frequency distribution plots for the local vulnerability of all six months and for each month were presented in Supplementary Figure S1. Since most of the variables were static throughout the six months and only the local population activity intensity changed over time, the overall spatial distribution was similar between months. The distribution mainly reflects the distribution of residential population density. Thus, the southern and central part of Singapore where the Central Business District (CBD) is located had slightly lower values. The Central region has a low population and density. Previous study (Chin and Bouffanais 2020) which attempted to identify super-spreader and super-susceptible locations in Singapore also obtained similar and counter-intuitive results—the Central Region (the CBD of Singapore) contained less than expected super-spreader and super-susceptible locations. These results highlighted that while the CBD might be important in the disease control, the residential areas could also have high risk of disease spreading thus should not be neglected (Huang et al. 2019).

³Moving distance parameter for sensitivity analysis is different from Table 6.2 as it was meant to measure a more active radius of mobility and interactions.

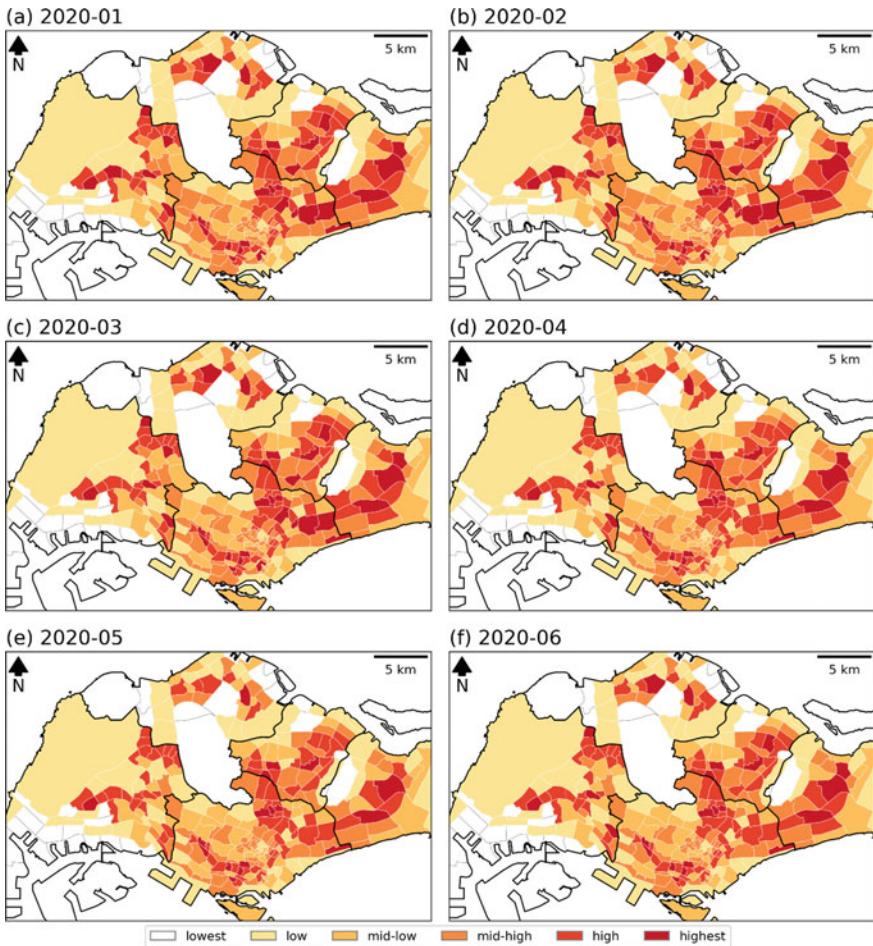


Fig. 6.6 Spatial distribution of local risk for the first six months. The local vulnerability values were categorized into six groups using a set of common break values (generated using Jenk's natural breaks method with all values stacked). Subzones with a darker shade are locally more vulnerable

6.4.2 Spatiotemporal Distribution of Vulnerability with Spatial Lag Effects

Using the aforementioned parameter settings, the effects of neighboring areas were included in the IV index for the six months (Fig. 6.7). Similar to Fig. 6.6, the subzones were categorized into six groups using the same set of break values derived from Jenk's natural breaks method and using the IV measures of all six months. The frequency distribution of the IV were presented in Supplementary Figure S2. The IV in January showed similar patterns with the local spatial vulnerability distribution (Fig. 6.6a). Most subzones in the southern part of Singapore, including the CBD,

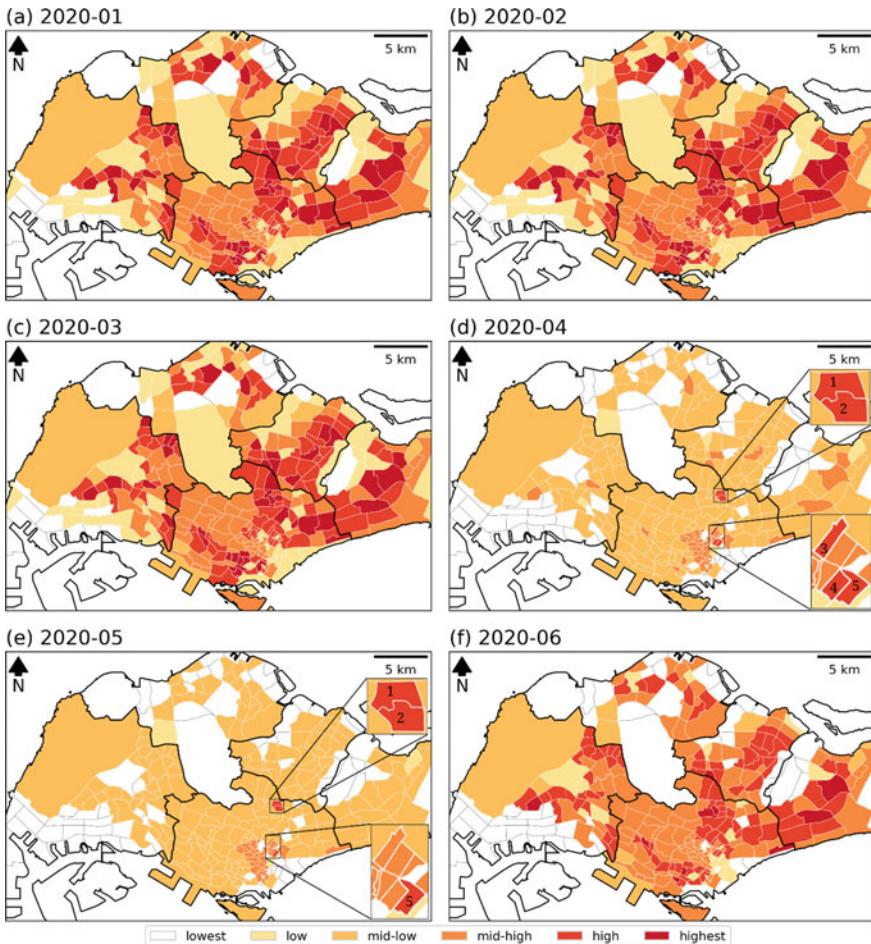


Fig. 6.7 Spatiotemporal distribution of the IV (local and spatial lagged) for the six months. The IV values were categorized into six groups using a set of common break values derived from Jenk's natural breaks method with all values stacked. A darker shade represents increased vulnerability. The highlighted subzones in (d) and (e) are: (1) Lorong 8 Toa Payoh, (2) Pei Chun, (3) Farrer Park, (4) Victoria, and (5) Bugis, respectively

where many of the businesses are located, have increment values from “lowest”-to-“mid-low”, “mid-low”- to-“high”. The relatively lower IV is partly due to the smaller population size and reduced density in this region. With the changes of spatial lag distances for elderly and young in February and March (reduced from 25 to 10 km), the spatial distribution patterns did not change. This may be attributed to the smaller representation of the two age groups (20.5% for young and 13.5% for elderly) in the population compared to the adult proportion (66.1%), leading to weak effects of reducing movement distance for the former.

During April and May, when the CB measures were implemented and most working adults were confined at home except essential workers, the IV of most subzones was reduced to level “lowest”-to-“low”, with only a few subzones remained on level “mid-low”-to-“mid-high” (Fig. 6.7). Two subzones (Lorong 8 Toa Payoh and Pei Chun) remained at level “high” in April and May; Lorong 8 Toa Payoh’s IV was level “high” for all six months, whereas Pei Chun’s was on level “highest” on January to March, dropped to level “high” from April and June. Three subzones’ IV ((3) Farrer Park, (4) Victoria, and (5) Bugis) also remained at level “high” for five months (from January to June except May); in May, only Bugis remained at level “high”, while Farrer Park and Victoria dropped to level “mid-high”. In other words, the measure to reduce moving distance for the five subzones was ineffective.

Figure 6.8a shows the changes in IV for each subzone over a six-month period. A clustering analysis of the trends based on k-means clustering with the Scikit-learn Python package (using silhouette score to determine the number of cluster-parameter k) was performed to identify the changes in IV for each subzone. Four distinct clusters were identified (Fig. 6.8b): In the first cluster, the IV values were similar for the first three months. After which, most of the scores declined in April, some of which accelerated down south to a trough in May (green, cluster-C). The second cluster demonstrated similar trajectory, albeit at a slower rate in May (blue, cluster-D). In June, the vulnerability values of most subzones began to pick up as the CB was partially lifted. The third cluster however, represents subzones that did not register a downward trend in April and May (red, cluster-A), i.e., consistently high. For the fourth and final cluster, the reduction rate was slower than the other subzones (orange, cluster-B). This indicates that the same moving distance restriction measures may have varying impact for different subzones.

Figure 6.9 shows the spatial distribution of the four clusters. Most subzones belong to cluster-C or D (blue or green), covering most areas of Singapore. All subzones of cluster-A (red) and cluster-B (orange) were in Central Region (Fig. 6.9b). Two subzones were classified as cluster-A (red cluster), including (1) Lorong 8 Toa Payoh and (2) People’s Park. The IV slightly increased during April and May, which may

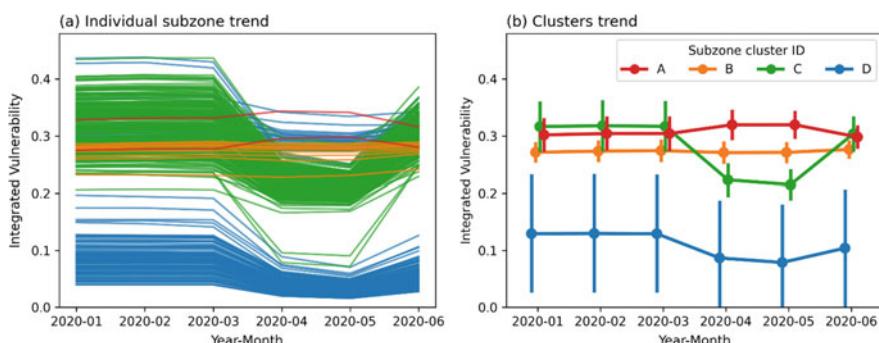
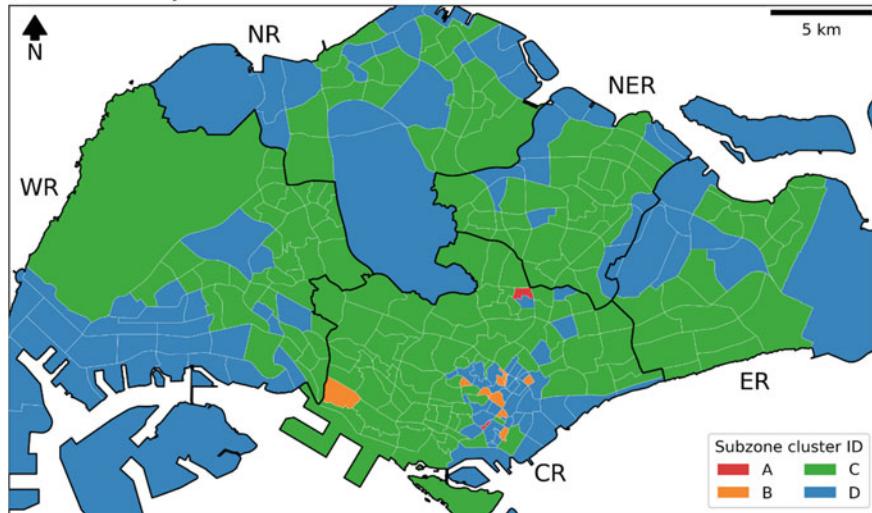


Fig. 6.8 Clustering results of the IV trend for subzones: **a** each line representing a subzone’s IV, and **b** four clusters of IV for clustering of (a)

(a) Whole study area



(b) Part of Central Region

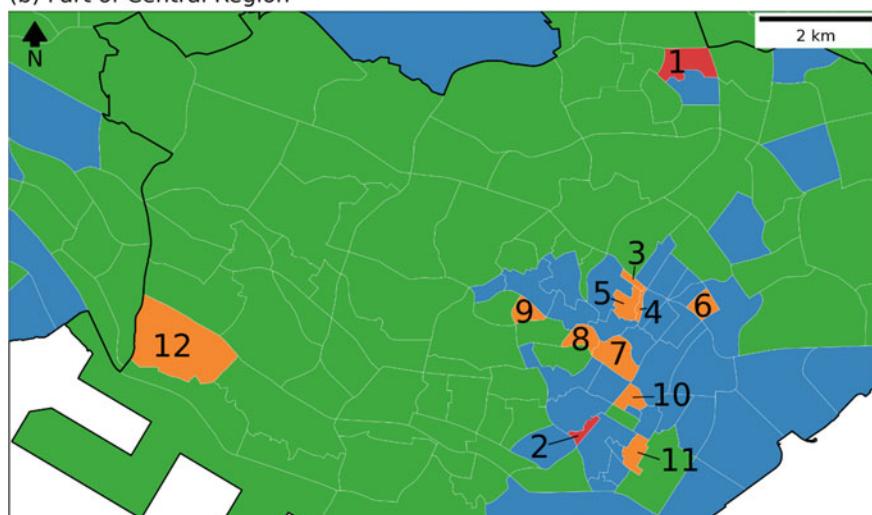


Fig. 6.9 Spatial distribution of the four clusters. The five regions of Singapore were labelled in (a): CR: Central Region, ER: East Region, NER: North East Region, NR: North Region, and WR: West Region. The subzones highlighted in the Central Region (b) are: (1) Lorong 8 Toa Payoh, (2) People's Park, (3) Mackenzie, (4) Selegie, (5) Mount Emily, (6) Kampong Glam, (7) Fort Canning, (8) Oxley, (9) Paterson, (10) Boat Quay, (11) Cecil, and (12) National University of Singapore

be a result of reduction in local effect (the parameter p), and increased neighboring effect ($1-p$), as the two subzones were surrounded by other subzones with higher local vulnerability. Ten subzones ((3) Mackenzie, (4) Selegie, (5) Mount Emily, (6) Kampong Glam, (7) Fort Canning, (8) Oxley, (9) Paterson, (10) Boat Quay, (11) Cecil, and (12) National University of Singapore) were classified as cluster-B (orange cluster). Almost all areas in the National University of Singapore subzone are used for education purposes, thus it contains low residential population, which implies level-low in local risk (Fig. 6.6). Alternatively, because of the low residential population, its IV was less sensitive to spatial lag effects, which was mainly based on the population proportions and distance settings, hence there was no reduction in April and May. The other nine subzones were located near the core area of Central Region. Each of the nine subzones has a small area size, and surrounded by other small subzones. In comparison to other areas, these subzones have a larger number of neighboring subzones in a given radius, and this leads to a relatively stable patterns and is less sensitive to the changes of moving distances parameters.

6.4.3 Sensitivity Analysis for Understanding the Effect of Restriction Distances and Local Proportion

Figure 6.10 shows the aggregated results for the combination of three variables, the effect from local (in the five subplots), the mobility distance for working adult (legend showing the lines with different colors), and distance for young/elderly (in horizontal axis). When local effect is set to 1.0 (Fig. 6.10e), i.e., effect from neighboring region is zero, the IV is fixed on the same value, and the distance for both adult and young/elderly would not have any effect. When p decreased to 0.75 (Fig. 6.10d), the IV started to decrease, with a gap appeared between the two lines with distance for adult set to 500 m and 1 km. The distance for young/elderly did not have significant effect when p set to 0.75. When p is set to 0.5 (Fig. 6.10c), which indicate half of the effect comes from the local, and half from its neighbors through the spatial lag effect, the IV will be significantly higher if the range of mobility is above 1 km across all age group.

In Fig. 6.10a, b, where the p parameter for the local effects were set to zero and 0.25, respectively (i.e., strong neighboring effect), the gaps in IV for the lines diverged, and the increment trends by the distance for elderly/young were most significant; in short, greater mobility range predicted higher IV. It should be noted that the standard deviation for the two figures is larger for the adult and elderly/young groups when the distance is set above 1 km and with a low value of local effect, meaning that the distributions were flatten and uncertainty is larger if the local effect is zero, i.e., IV is solely depending on its spatial neighbors.

These results highlighted the overall effects of the three parameters. Specifically, the proportion of local effect had a significant influence to the overall vulnerability; the impact from adult mobility distances appeared between 500 m to 1 km; and the

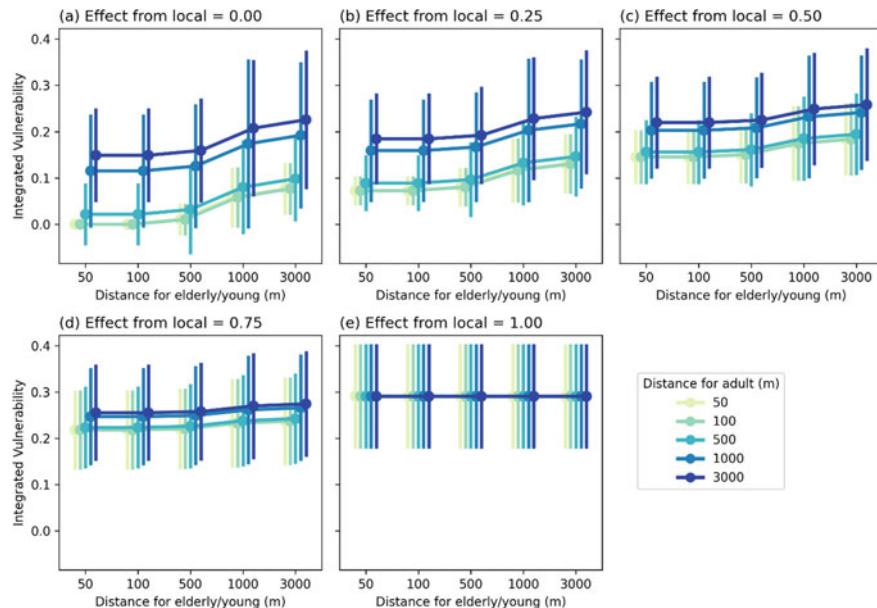


Fig. 6.10 Sensitivity analysis of the three parameters (local effects (p), distance for adults, and distance for young and elderly). The circle marker and error bars indicate the mean and standard deviation of all subzones for a given set of parameters

IV increases significantly if the distance for both adult or young/elderly is increased beyond 1 km. The details of the distributions were displayed in Supplementary Figure S3.

To find out the sensitivity of subzones' IV on adult mobility distances, we ran a k-means clustering with the lines of trends for each subzone by distance for young/elderly, segmented by the distance for adult (in individual subplot). The subzones with zero changes were filtered out from the clustering analysis (shown as cluster-F in Fig. 6.11). Figure 6.11 shows the clusters of trends by distance for young/elderly, which contained five clusters for each distance for adult parameters. Each cluster displays a unique trend by distance for young/elderly, which mainly occurs on the point of significant jump. For instance, where adult mobility is restricted to 50 m, cluster-C had an IV jump between 1 and 3 km; cluster-B increased between 500 m to 1 km; and cluster-A is seemingly most sensitive to travel distance of young/elderly, as it increased significantly between 100 and 500 m. The other two clusters (cluster-D and cluster-E) had low fluctuations, and the cluster-A is consistently higher than cluster-C. This means that the effect of distance for different subzone is also varied. In all other mobility radius (100, 500, 1,000, and 3,000 m), cluster-A has consistently demonstrated a significant leap wherever travel distance of young/elderly increased from 100 to 500 m.

The spatial distribution of the six clusters from Fig. 6.11 is illustrated in Fig. 6.12. Most subzones were either cluster F, cluster-C, or cluster-B. Cluster-A—seemingly

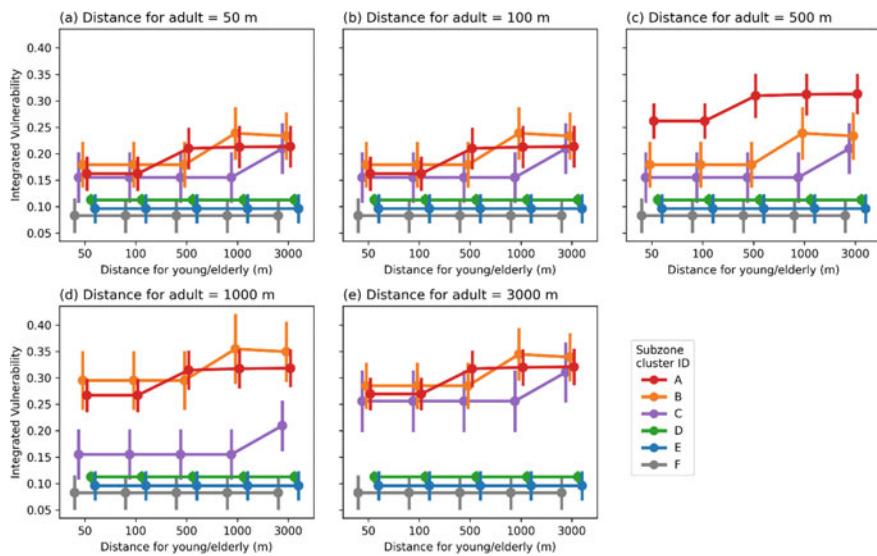


Fig. 6.11 Comparison of the effects of distance restrictions for adult versus young and elderly, with local effect (p) fixed at 0.5. Each color indicates different trend cluster

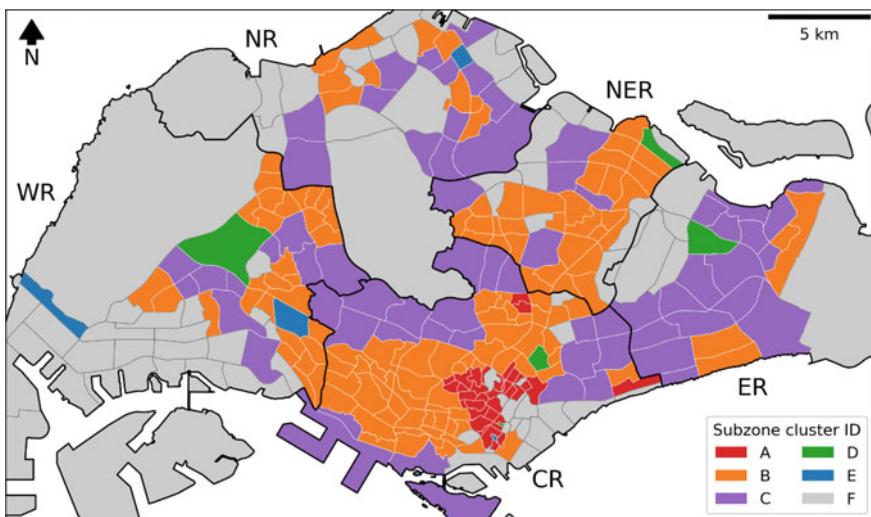


Fig. 6.12 Spatial distribution of the clusters. The five regions were labelled as: CR: Central Region, ER: East Region, NER: North East Region, NR: North Region, and WR: West Region

the most sensitive cluster, is concentrated in the Central Region, which contains the CBD. These are places where movement should be restricted to less than 100 m to achieve the reduced IV. This is followed by subzones in cluster-B, where movement should be restricted to no more than 500 m, and lastly for subzones in cluster-C indicated that the IV would increase unless movement distance is restricted to less than 1 km.

6.5 Discussion

This study examines spatiotemporal vulnerability through a socio-ecological lens, considering a subzone's population density (i.e., densely populated districts are considered more vulnerable), residents' age profiles (i.e., higher proportion of young and elderly residents are more vulnerable), socio-economic conditions (i.e., higher housing resale prices as surrogate measure of higher status), proximity to POIs linked to COVID-19 infections (e.g., supermarkets, wet markets, shopping malls, and train and bus stations), land use diversity (i.e., more diverse land use associated with greater social interactions), and population activity intensity (i.e., number of commuters passing through each subzone, where more intense activities are linked to greater vulnerability).

While a static perspective to assess spatial vulnerability is well documented in the empirical literature (e.g., CDC's Social Vulnerability Index), the impact from the temporal dynamics is less established. The COVID-19 epidemic provided the opportunity to study population movement in a period of disruptions, and the factors that underscore vulnerability to disease transmissions. Specifically, the CB (i.e., lockdown) that was implemented as part of the health preventive measures to contain the outbreak created the perfect chance to scrutinize geographical regions (i.e., subzones) that are differentially exposed to the risk factors as a function of time and space.

6.5.1 Spatiotemporal Distribution of IV

Spatiotemporal vulnerability incorporates impact of local attributes (e.g., proportion of young and elderly residents who are most at risk, proximity to POI's known for infections) and the spatial lag effects. Locale vulnerability is not only affected by local risks, but also its neighboring's risk and vulnerability, including the population distributions of various age groups in the region.

In line with observations in Fig. 6.6, static, local risk factors do not change over time. However, as a result of the CB in April and May, the reduction in movement has transformed spatiotemporal vulnerability (Fig. 6.7)—all but a handful of locations have demonstrated reduced vulnerability. The subzones that reported consistent IV over the CB period include Lorong 8 Toa Payoh, Pei Chun, Farrer Park, Victoria, and Bugis. The last three subzones are part of or closest to the CBD, and thus the

result was unsurprising. However, the comparatively higher vulnerability reported in Lorong 8 Toa Payoh and Pei Chun were unexpected. The two subzones are home to plots of public houses, vehicle depots for buses and taxis, and industrial parks. A closer examination on the two subzones reveals that they are bounded by two busy highways on the North and the East (Fig. 6.13), and this greatly limits the accessibility, mobility, and exposure of residents in the two neighborhoods. The two highways are barricaded, and the flow in and out of the two subzones are restricted to the South and the West ends. As such, the elevated IV in these two areas could have been a reflection on urban spatial bottleneck. This conjecture evidently warrants further studies.

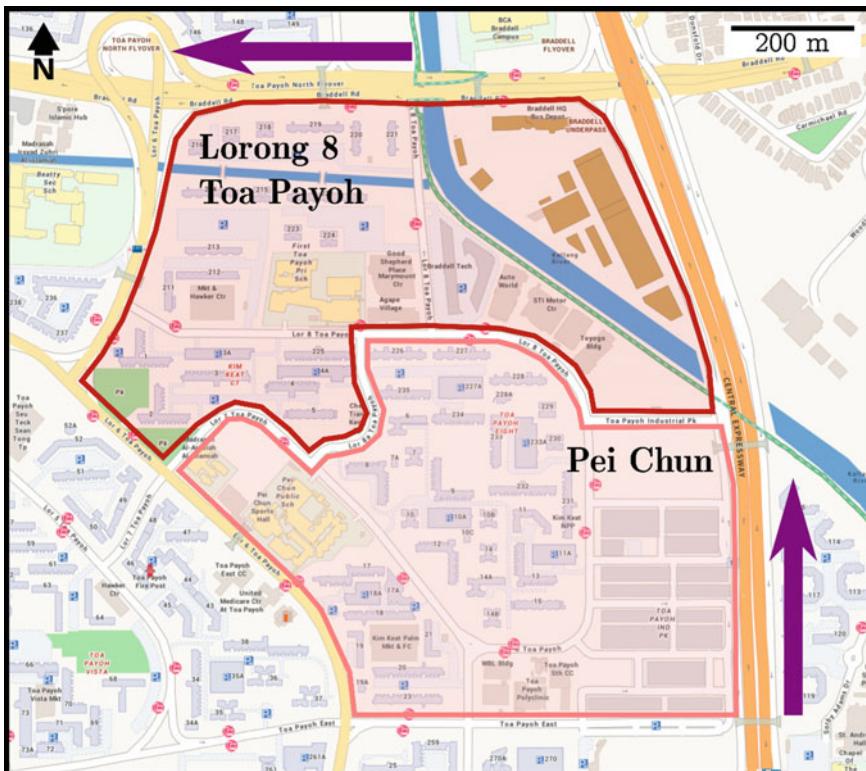


Fig. 6.13 Spatial layout of Lorong 8 Toa Payoh (boundary in red) and Pei Chun (boundary in pink) subzones, extracted from OneMap (location: 1.339 N, 103.858E, accessed: December 21, 2020, <https://www.onemap.sg/main/v2/>). The purple arrows show the two highways that locked in the two subzones to a corner. The blue line is a canal, with industrial buildings located on both sides

6.5.2 Differential Impact of Mobility Distance and Circuit Breaker

Using sensitivity analyses (Fig. 6.8), we identified distinct clusters of subzones according to their changes on the IV over a six-month period. Some subzones in the Central Region remain on the top of the IV alert list throughout the months of April and May (Fig. 6.9). In other words, reducing movement distance may not be as effective for some subzones. The results suggest that these areas could receive more epidemiological attention during the CB.

Subzones in cluster A (Fig. 6.12) show increase in IV within a shorter distance range (100–500 m) compared to subzones that are further away from the CBD and Central Region. These areas have higher spatiotemporal vulnerability, ostensibly due to the locale centrality, locations of various key administrative buildings (e.g., Immigration and Checkpoint Authority), retail and business districts (e.g., Orchard Road shopping malls, Clarke Quay and Robertson Quay), recreational facilities (e.g., East Coast Park), and transportation nodes (e.g., vehicle depots and maintenance hub for buses and taxis). Safe distancing and other health preventive measures can be strengthened and actively enforced in these areas.

Lastly, and in line with the geographic distributions of IV (Fig. 6.7), two subzones consistently show elevated risk, namely Lorong 8 Toa Payoh and Pei Chun (Fig. 6.12). The two estates are home to an estimate population of 7,160 and 10,360, respectively ([Singapore Department of Statistics \(n.d.\)](#)). Both neighborhoods offer a wide range of amenities and industrial land use. The IV index on Lorong 8 Toa Payoh has also remained stable overtime in-spite of the CB in April and May (Fig. 6.9). This further underscores the need for policy calibration and intervention at the local level. It should be noted that the two subzones are not the only estates with diverse, industrial land use. Whether higher IV is a result of the unusual landlock illustrated in Fig. 6.13 needs to be verified.

6.5.3 Differential Impact of Spatial Lag Effects

As the young and elderly residents were confined at home during CB, their distance mobility range was curtailed. However, this is not so for the working adults, some of whom were still required to travel to work. Consequently, the distance mobility for the latter group would expose the former (i.e., young, elderly) to greater epidemiological risk, at both the local and neighborhood level. The increased in IV on the young and elderly groups became obvious when the local effect was set as 0.5 (i.e., similar impact from local and regional demographic attributes) and where the distance mobility range for working adults was at least 500 m or more (Fig. 6.10).

Additionally, sensitivity analyses based on restriction of distance movement and proportion of various age demography found that as spatial lag effect increases (i.e., p reduces to 0, neighboring regions become more influential on vulnerability), the

mobility range of young and elderly people would have an impact on IV but this increment is seemingly less than the overall impact of local effects (Fig. 6.10). The result may be attributed to the smaller proportion of young and elderly residents compared to working age adults. It may also be because the differences in local risk between age groups (e.g., the accessibility to key amenities may be less important to working age adult) were not incorporated. When local effect is set to 1.0 (Fig. 6.10e), i.e., effect from neighboring region is zero and vulnerability is entirely local, the IV is fixed on the same value, and the distance for both adult and young/elderly would not have any effect.

6.6 Conclusions and Future Work

This is an exploratory study to integrate spatial and temporal information using urban public transportation data to inform and develop a measure on epidemic vulnerability. To the best of our knowledge, this study is the first of its kind and will bridge an important conceptual vacuum in Singapore, one of the most densely populated city-states in the world. Notwithstanding, there are a few methodological limitations that warrant further investigation. First, we combined and analysed six months of static human mobility data across 323 subzones. The period could be extended before and after the CB for at least a year. The additional data would allow us to simulate spatial and temporal models simultaneously, and help identify potential interactions between the two components. Comparisons between pre- and post-CB will give further insights on the efficacy of the movement restrictions.

Second, the IV model is based on various assumptions about the demographic profiles of subzones and the characteristics believed to be vulnerable to COVID-19 transmission, e.g., elderly residents as a more vulnerable group. The model's predictive validity however needs to be established and benchmarked against actual infection cases in each subzone, similar to the CDC's Social Vulnerability Index. The framework to include varying age groups for different spatial lag effects can be extended to other types of population categories, e.g., ethnicity, gender, socio-economic/education status because different categories of people might have contributed to different spatial lag effects. Future studies can explore the effect of these categories.

Taking an overarching view, the results suggest that spatiotemporal vulnerability could be shaped by multiple forces, which include the urban social and built environment, local demographic characteristics, and the composite spatial lag effects from contiguous subzones. The lockdown period in April and May provided a natural "treatment" manipulation, where we observe changes in actual mobility that result from movement restrictions. Using GIS, sensitivity tests, subgroup and cluster analyses, the findings gave us an invaluable opportunity to study this naturalistic change on spatiotemporal risk. This study has set the stage and inspired researchers to experiment with more robust and bolder methods in future endeavours.

With no end in sight for the COVID-19 pandemic, the health advisories and movement restrictions are likely to remain status quo for another year, and in turn, reinforce the work-from-home arrangement, and curate a new consumption pattern and transport utilization. The disruptions and the macro-economic forces will catalyst a decentralization trend, where residents make fewer trips to the CBD area, but more activities and movement within their neighborhood town networks (Zhong et al. 2014). This would herald a new routine and augment the importance of mobility distance and spatial lag effects.

Acknowledgements The preparation of this manuscript by the second author was supported by Singapore University of Technology and Design (Cities Sector: PIE-SGP-CTRS-1803).

Supplementary Figures

See Fig. S1.

See Fig. S2.

See Fig. S3.

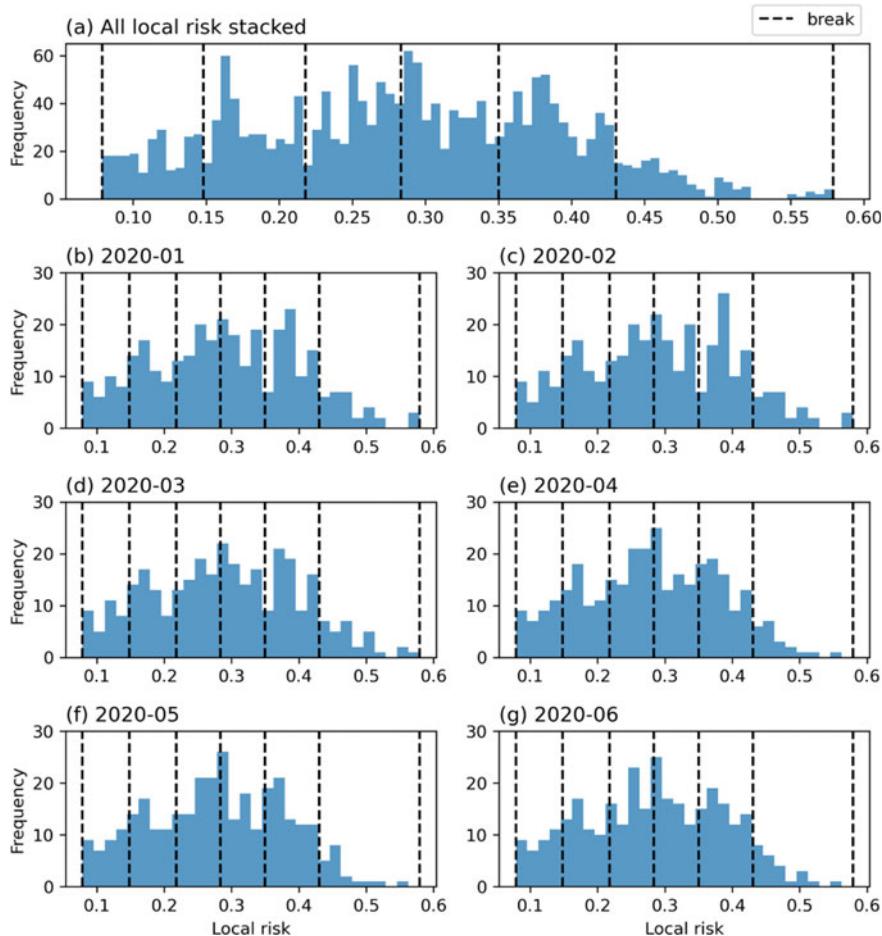


Figure S1 The histograms of: **a** all local vulnerability values, **b–g** the local vulnerability values for each month from January to June. All seven sub-plots shared the same set of Jenk's natural breaks that was calculated based on all local vulnerability values. Jenk's natural breaks method would generate breaks that aims to minimize the variance within group and maximize the variance between groups. As a result, all the five breaks (excluding the minimum and maximum values) located at the lower point (valley) in the histogram presented in **(a)**. We generated Jenk's natural breaks on all local vulnerability values so that the six months shared the same set of break values, and thus the results in different months can be compared

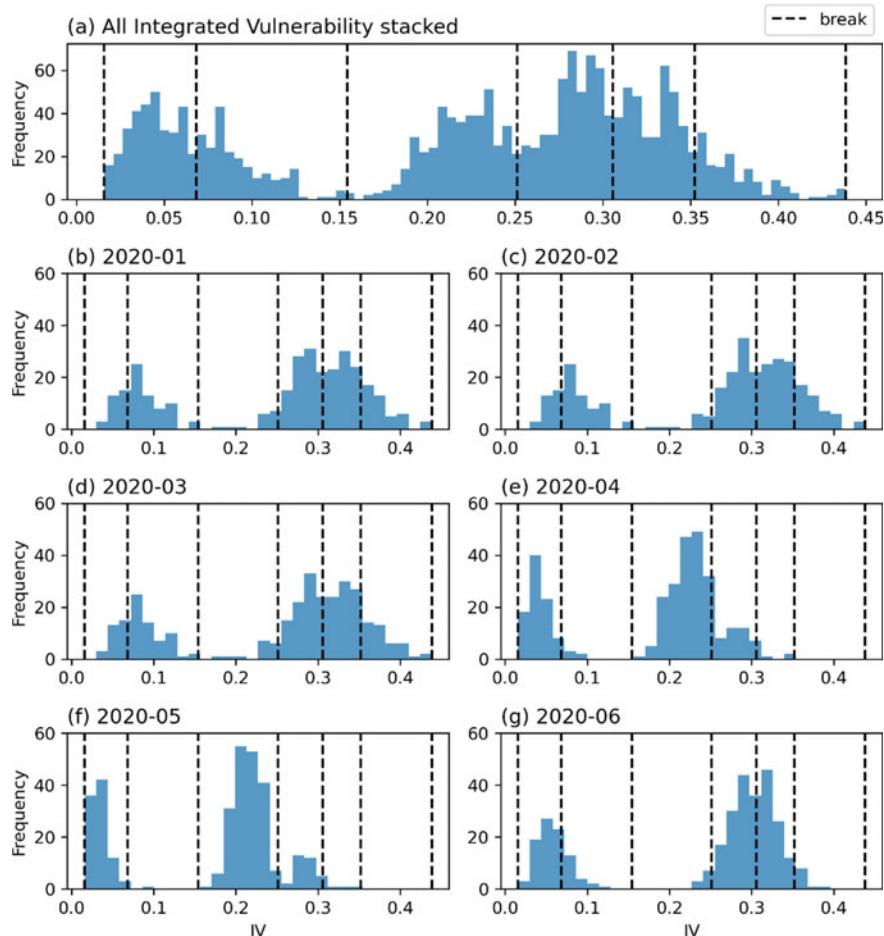


Figure S2 The histograms of: **a** all IV, **b–g** the IV for each month from January to June. All seven sub-plots shared the same set of Jenk's natural breaks that was calculated based on all IV. Similar to Figure S1, the Jenk's natural break values located at the lower points in **(a)**. We generated Jenk's natural breaks on all IV so that the six months shared the same set of break values, and thus the results in different months can be compared

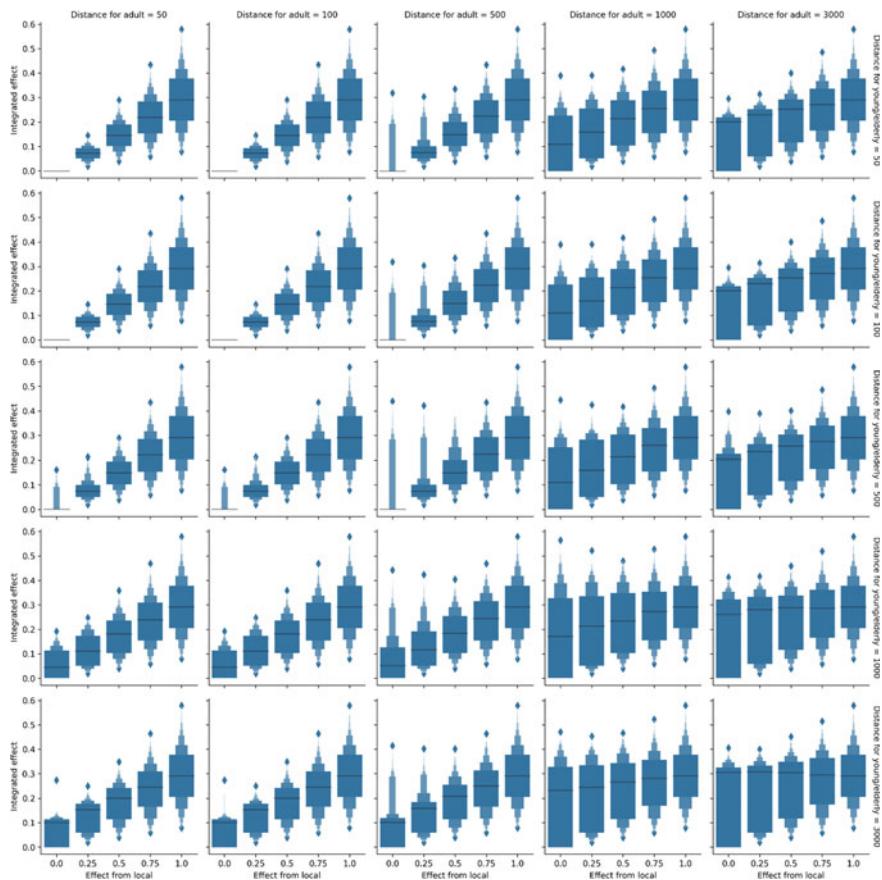


Figure S3 The boxplot of the integrated vulnerability for each parameter combination set

References

- Adger, W. N. (2006). Vulnerability. *Global Environmental Change*, 16, 268–281. <https://doi.org/10.1016/j.gloenvcha.2006.02.006>.
- Bogich, T. L., Funk, S., Malcolm, T. R., Chhun, N., Epstein, J. H., Chmura, A. A., et al. (2013). Using network theory to identify the causes of disease outbreaks of unknown origin. *Journal of the Royal Society, Interface*, 10, 20120904.
- Bratman, G. N., Hamilton, J. P., & Daily, G. C. (2012). The impacts of nature experience on human cognitive function and mental health. *Annals of the New York Academy of Sciences*, 1249, 118–136. <https://doi.org/10.1111/j.1749-6632.2011.06400.x>.
- Chin, W. C. B., & Bouffanais, R. (2020). Spatial super-spreaders and super-susceptibles in human movement networks. *Scientific Reports*, 10, 18642. <https://doi.org/10.1038/s41598-020-75697-z>.
- Cutter, S. L. (1996). Vulnerability to environmental hazards. *Progress in Human Geography*, 20(4), 529–539. <https://doi.org/10.1177/2F030913259602000407>.
- Cutter, S. L., Boruff, B. J., & Shirley, W. L. (2003). Social vulnerability to environmental hazards. *Social Science Quarterly*, 84(2), 242–261. <https://doi.org/10.1111/1540-6237.8402002>.

- Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., et al. (2008). A place-based model for understanding community resilience. *Global Environmental Change*, 18, 598–606. <https://doi.org/10.1016/j.gloenvcha.2008.07.013>.
- Flanagan, B., Hallisey, E., Adams, E., & Lavery, A. (2018). Measuring community vulnerability to natural and anthropogenic hazards: The centers for disease control and prevention's social vulnerability index. *Journal of Environmental Health*, 80, 34–36.
- Gómez, J., & Verdú, M. (2017). Network theory may explain the vulnerability of medieval human settlements to the Black Death pandemic. *Scientific Reports*, 7, 43467. <https://doi.org/10.1038/srep43467>.
- Huang, C. Y., Chin, W. C. B., Wen, T. H., Fu, Y. H., & Tsai, Y. S. (2019). Epirank: Modeling bidirectional disease spread in asymmetric commuting networks. *Scientific reports*, 9(1), 1–15. <https://doi.org/10.1038/s41598-019-41719-8>.
- Karaye, I. M., & Horney, J. A. (2020). The Impact of Social Vulnerability on COVID-19 in the U.S.: An analysis of spatially varying relationships. *American Journal of Preventive Medicine*, 59(3), 317–325. <https://doi.org/10.1016/j.amepre.2020.06.006>.
- Lehnert, E. A., Wilt, G., Flanagan, B., & Hallisey, E. (2020). Spatial exploration of the CDC's social vulnerability index and heat-related health outcomes in Georgia. *International Journal of Disaster Risk Reduction*, 46,. <https://doi.org/10.1016/j.ijdrr.2020.101517>.
- Leslie, E., Coffee, N., Frank, L., Owen, N., Bauman, A., & Hugo, G. (2007). Walkability of local communities: Using geographic information systems to objectively assess relevant environmental attributes. *Health and Place*, 13(1), 111–122. <https://doi.org/10.1016/j.healthplace.2005.11.001>.
- Lewis, D. (2020). Coronavirus in the air. *Nature*, 583, 510–513.
- McMahana, E. A., & Estesb, D. (2015). The effect of contact with natural environments on positive and negative affect: A meta-analysis. *Journal of Positive Psychology*, 10(6), 507–519. <https://doi.org/10.1080/17439760.2014.994224>.
- Ministry of Health (n.d.). Updates on COVID-19 (Coronavirus disease 2019) local situation. Retrieved July 31, 2020 at <https://www.moh.gov.sg/covid-19>.
- Morawska, L., & Milton, D. K. (2020). It is time to address airborne transmission of COVID-19. *Clinical Infectious Diseases*, ciaa939. <https://doi.org/10.1093/cid/ciaa939>.
- Mueller, A. L., McNamara, M. S., & Sinclair, D. A. (2020). Why does COVID-19 disproportionately affect older people? *Aging*, 12(10), 9959–9981. <https://doi.org/10.1863/aging.103344>.
- Nayak, A., Islam, S. J., Mehta, A., Ko, Y. A., Patel, S. A., Goyal, A., et al. (2020). Impact of social vulnerability on COVID-19 incidence and outcomes in the United States. *MedRxiv*. <https://doi.org/10.1101/2020.04.10.20060962>.
- Petrosillo, N., Viceconte, G., Ergonul, O., Ippolito, G., & Petersen, E. (2020). COVID-19, SARS and MERS: Are they closely related? *Clinical Microbiology & Infection*, 26(6), 729–734. <https://doi.org/10.1016/j.cmi.2020.03.026>.
- Rufat, S., Tate, E., Burton, C. G., & Maroof, A. S. (2015). Social vulnerability to floods: Review of case studies and implications for measurement. *International Journal of Disaster Risk Reduction*, 14(4), 470–486. <https://doi.org/10.1016/j.ijdrr.2015.09.013>.
- Schmidlein, M. C., Shafer, J. M., Berry, M., & Cutter, S. L. (2011). Modeled earthquake losses and social vulnerability in Charleston, South Carolina. *Applied Geography*, 31(1), 269–281. <https://doi.org/10.1016/j.apgeog.2010.06.001>.
- Schwartz, M., & Cook, L. R. (2020). These N.Y.C. neighborhoods have the highest rates of virus death. *The New York Times*. Retrieved May 19, 2020, at <https://www.nytimes.com/2020/05/18/nyregion/coronavirus-deaths-nyc.html>.
- Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27(4), 623–656.
- Singapore Department of Statistics (n.d.). Geographic distribution. Retrieved Nov 25, 2020, at <https://www.singstat.gov.sg/find-data/search-by-theme/population/geographic-distribution/latest-data>.
- Snyder, B., & Parks, V. (2020). Spatial variation in socio-ecological vulnerability to COVID-19 in the contiguous United States. Retrieved May 19, 2020 <http://doi.org/10.2139/ssrn.3587713>.

- Tatem, A. J., Rogers, D. J., & Hay, S. I. (2006). Global transport networks and infectious disease spread. *Advances in Parasitology*, 62, 293–343. [https://doi.org/10.1016/S0065-308X\(05\)62009-X](https://doi.org/10.1016/S0065-308X(05)62009-X).
- Yonker, L. M., Neilan, A. M., Bartsch, Y., Patel, A. B., Regan, J., Arya, P., et al. (2020). A.Pediatric severe acute respiratory syndrome coronavirus 2 (SARS-COV-2): Clinical presentation, infectivity, and immune responses. *Journal of Pediatrics*, 227, 45–52. <https://doi.org/10.1016/j.jpeds.2020.08.037>.
- Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, C. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science*, 28(11), 2178–2199. <https://doi.org/10.1080/13658816.2014.914521>.

Part II

Data Perspectives

Chapter 7

Assessing Connections and Tradeoffs Between Geospatial Data Ethics, Privacy, and the Effectiveness of Digital Contact Tracing Technologies



Peter Kedron and Andrew B. Trgovac

7.1 Introduction

Contact tracing is a set of activities employed by public health professionals to assist with the abatement of infectious diseases. A part of contact tracing, the identification of individuals who may have come into contact with a person known to be infected with a disease, has been identified as key to the mitigation and suppression of COVID-19 (CDC 2020; Kahn 2020). Effective contact tracing allows public health authorities to sever chains of transmission and shift policy to case-based interventions such as selective individual quarantines rather than population-wide interventions such as social distancing. While public health authorities have the ability to conduct manual contact tracing, many do not have the capacity to identify and trace infected individuals at the scale or speed needed to respond to the COVID-19 pandemic. To improve the reach and effectiveness of contact tracing, many academics and policymakers have proposed a shift to technology-assisted contact tracing (TACT) systems that use the geospatial technologies (e.g., GPS, WiFi, Bluetooth) embedded in mobile devices to gather, store, transfer, and share the location and contact histories of individuals (see Ferretti et al. 2020; Kahn 2020).

However, the movement to develop TACT and introduce digital contact tracing technologies (DCTT) into pandemic management has raised a number of practical, ethical, and privacy concerns. As a practical matter, the accuracy and reliability of DCTT are limited by both the geospatial technologies these systems are built upon and the ways in which the general public will, or will not, use them. As an ethical matter, differential access to key technologies and spatial variation in the efficacy of DCTT mean that some segments of society may not benefit from DCTT as much as

P. Kedron (✉) · A. B. Trgovac

School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ, USA
e-mail: Peter.Kedron@asu.edu

P. Kedron

Spatial Analysis Research Center (SPARC), Arizona State University, Tempe, AZ, USA

others. At the same time, the large scale collection of detailed data about the location and social networks of individuals raises concerns about privacy and possible social stigmatization. Understanding and addressing each of these issues is fundamentally a geographic challenge because each issue is shaped by the capabilities of geospatial technologies and how those technologies are used to suppress COVID-19. Meeting these challenges requires identifying and evaluating the tradeoffs that exist between technology uses that may improve the efficacy of DCTT at the expense of selected groups or individual liberties.

This chapter presents a framework that can be used to gather the data needed to assess the potential tradeoffs between privacy and effectiveness that may arise as DCTT are developed and deployed to address COVID-19. How well DCTT support TACT efforts, and whether those efforts adhere to ethical principles presented in the public health and location privacy literatures depends largely on the purpose and boundaries of data collection, the entities responsible for data collection, the entities granted access to data, the definition of allowable data use, and how and when the data will be disposed of. To expand on each of these topics, the remainder of the chapter is organized as follows. First, we briefly introduce traditional forms of contact tracing and identify how the transmission characteristics of COVID-19 have strained this public system. We then present TACT systems and the DCTT they are built upon. We explain how these systems are intended to function and identify potential benefits and limitations of DCTT using selected examples from the COVID-19 pandemic. In the fourth section of the chapter, we present the ethical and privacy principles at the center of the DCTT debate. We develop an assessment framework around those central principles and the characteristics of TACT presented earlier in the chapter and discuss key data that will be needed to operationalize the framework. Finally, we conclude the chapter by identifying key contextualizing factors that have emerged during the first six months of the pandemic and reflect on potential contributions geospatial researchers can make going forward.

7.2 Contact Tracing, Digital Contact Tracing Technologies, and Technology Assisted Contact Tracing

In this chapter, we draw the following distinctions, which we emphasize here for clarity. When we use the term *contact tracing*, we are referring to the traditional process of using in-person or phone interviews to identify individuals that may have been exposed to a disease by coming into contact with an individual carrying that disease, and to the set of related activities (e.g., follow-up calls, connecting individuals to support resources) used to limit disease spread. In the United States, contact tracing is conducted by, or under the jurisdiction of, state or local public health departments that are bound by the Public Health Code of Ethics (APHA 2019) and related law (see Gostin and Wiley 2016). We use *digital contact tracing technologies* (DCTT) to refer to a specific set of technologies (e.g., GPS, Bluetooth, contact

tracing apps) tied to mobile phones that can be used together to gather data about the location of individuals and/or their proximity to others. The related term *technology-assisted contact tracing* (TACT) refers to the use of DCTT to augment traditional contact tracing by recording the location histories of DCTT users, notifying users of potential exposure, or otherwise intervening in the interest of public health.

7.2.1 Contact Tracing

The goal of contact tracing is to identify every individual a person with an infectious disease has come into contact with, so those individuals who were exposed and potentially infected with the disease can be quarantined away from the general population, thus breaking the chain of disease transmission. While we primarily focus on the identification and notification of contacts in this chapter, contact tracing involves a wider set of activities that includes connecting infected individuals and their contacts with needed services, conducting follow-up activities to encourage and ensure that infected individuals take appropriate disease mitigating actions, and generally providing the support services needed to limit disease transmission.

Contact tracing is a cornerstone of infectious disease control. The eradication of smallpox relied on extensive contact tracing and subsequent patient isolation and community immunization (Eames and Keeling 2003). Public health professionals have also used contact tracing to control diseases that spread through sexual networks, such as syphilis and HIV (Cates et al. 1996; Judson and Vernon 1988). Most recently, contact tracing and subsequent control measures such as quarantine were critical in limiting outbreaks of SARS in 2003 and Ebola in 2014 (Riley et al. 2003; Saurabh and Prateek 2017).

For COVID-19, contact tracing begins with contact identification and notification, but also includes the immediate quarantining of individuals exposed to the virus or isolating of individuals infected with the virus. To reduce the burden these actions place on individuals and improve their compliance, contact tracers also share information about essential services such as child care and elder care and conduct regular follow-ups with affected individuals. Follow-up contact is another key component of contact tracing efforts as it helps individuals gain and maintain access to the resources they need to sustain isolation or quarantine until the risk of transmission has ended.

Given the wide set of activities involved, numerous factors can impact the effectiveness of contact tracing. First, contact tracing is most effective when testing is widespread, accessible, and has produced rapid results. Abundant testing speeds identification of disease transmission, which facilitates and improves the response of public health departments. Second, individuals must be willing to be tested for a disease, respond to contact tracers, and disclose infection for contact tracing to have the desired effect. Third, even when individuals are tested and respond to contact tracers, contact tracing will only impact disease transmission if those affected by the disease have access to the resources they need to successfully isolate or quarantine. Accessing resources begins with knowing that support services such as testing,

therapeutics, mental health support, childcare, and grocery delivery exist and how to access them. Contact tracers provide this knowledge during interviews. Moreover, effective use of these services depends on repeated contact to ensure individuals isolating or quarantining have continued access to resources that meet their needs. As such, the number of contact tracers relative to the size of the infected population can severely limit the effectiveness of contact tracing—because contact tracers cannot notify and follow-up sufficiently with those infected or exposed, and because service providers may be overwhelmed. Finally, contact tracing is shaped by the dynamics of disease transmission. Contact tracing is most effective when a disease spreads through prolonged direct physical contact with an infected person (direct transmission) after the onset of symptoms (symptomatic transmission) and when infected.

As described, contact tracing that relies on voluntary interviews with infected individuals that can be both time consuming and costly. Using contact tracing to mitigate and control COVID-19 has been challenging because the underlying virus (SARS-CoV-2) can be spread directly through physical contact or the exchange of respiratory droplets (Stadnytskyi et al. 2020) and indirectly through contact with contaminated surfaces¹ (Kampf et al. 2020). At the same time, SARS-CoV-2 can be spread during the incubation period of the virus before the onset of symptoms (pre-symptomatic transmission) or in cases where an infected individual never develops symptoms (asymptomatic transmission) (Bai et al. 2020; Mizumoto et al. 2020). Pre-symptomatic and Asymptomatic transmission disrupts contract tracing efforts for at least two reasons. First, infected individuals who are unaware that they are carrying and spreading the disease cannot identify themselves to contact tracers. Second, even in cases where individuals do develop symptoms and then volunteer for contact tracing, those same individuals may have already spread the disease preceding the onset of symptoms while they were still pre-symptomatic.² Moreover, contact tracing is most effective when conducted soon after an infected individual is identified. For SARS-CoV-2, local health departments seek to initiate contact tracing within 24 h of a confirmed positive test. In situations where transmission is widespread, departments may simply lack the resources to complete contract tracing in a timely manner. Due to this combination of factors, identifying personal contacts and delineating social networks may not alone be able to capture the underlying transmission mechanisms of SARS-CoV-2.

Contact tracing must also reconstruct the location history of individuals carrying SARS-CoV-2 to capture possible indirect, asymptomatic transmission. Examining

¹Although the risk of transmission of SARS-CoV-2 by contaminated surfaces is likely less than initially believed (Goldman 2020; Kanamori 2020).

²At the time of writing it remains unclear how prevalent pre-symptomatic transmission is and over what time period it extends. Research indicates that while infectiousness can start 12 days before the onset of symptoms, only a small percentage of transmission occurs before three days prior to symptom onset (a model-based estimate by He et al. (2020) suggests 9%). Epidemiological studies conducted in a variety of transmission contexts suggest actual rates may be lower than those predicted by transmission models (see Slifka and Gao 2020 for a summary). If presymptomatic transmission rates are low they become less of a barrier to traditional contact tracing.

the analogous case of the 2003 outbreak of SARS in Taiwan, Chen et al. (2007) demonstrate that introducing geographic contacts (when people share a location but not a direct interaction) into the construction of the disease contact network dramatically expands the size and connectivity of the network and by extension the number of possible chains of transmission. Studying tuberculosis, Klovadahl et al. (2001) and McElroy et al. (2003) similarly found that adding places as nodes in their social network analyses led to the identification of otherwise unrecognized contacts between patients and improved understanding of disease transmission. These findings collectively imply that a failure to account for geographic contacts and location in diseases with indirect, pre-symptomatic or asymptomatic spread is likely to dramatically misrepresent disease transmission.

Six months into the COVID-19 pandemic, contact tracing efforts have had mixed success—constrained by the dynamics of the virus, the lack of testing availability and the rapid return of test results, the ability of individuals to recall their location and contact history, and the magnitude of the COVID-19 pandemic overwhelming the number of available contact tracers (Steinhauer and Goodnough 2020). To improve this aspect of pandemic response, many have proposed a shift to TACT systems that use DCTT to record the location and contact history of users and notify them of potential exposures to SARS-CoV-2.

7.2.2 Digital Contact Tracing Technologies Used in Technology-Assisted Contact Tracing

Through the use of TACT, public health departments seek to improve the response to COVID-19 by using a digital record of the absolute location (where someone is in space) and/or relative location (where someone is in relation to someone or something else) of an individual to accelerate the contact tracing process. TACT may take several forms. One approach is to use DCTT to continuously record the location history of an individual that has downloaded a contact tracing application onto their mobile device. In the event that the individual tests positive for COVID-19, this information would then be used by contact tracers to retrace where that person had been, and to help identify who they came in contact with using the location history to prompt the individual's memory during an interview.

Individual location histories and the population level mobility patterns that can be constructed from those histories have been shown to be valuable resources during the management of Ebola (Wesolowski et al. 2014a, b), influenza (Farrahi et al. 2014; Dong et al. 2019), Cholera (Finger et al. 2016). In some countries (e.g., South Korea), a user's location data is combined with other forms of personal data (e.g., purchasing histories) to add depth to the location histories. However, this approach is not common in the United States, although some systems do allow users to link personal data (e.g., age, sex) to their location histories and give those users the opportunity to release the information to health departments if they wish.

An alternative approach is to focus on the relative location of individuals. In this approach, anonymized identification codes are exchanged through Bluetooth low energy (BLE) signals between the mobile devices of users that have downloaded an application to create a contact list for each individual (Ferretti et al. 2020). If an individual tests positive for COVID-19, the system would then send a notification to that user's contacts signaling that they should be tested for the virus or take other preventative actions (e.g., quarantine). To preserve privacy, the location or contact histories of the individual can be stored on a user individual mobile device and contacts could be anonymized to reduce the risk of identification. This decentralized, proximity-centered approach to recording potential contacts is the foundation of a collaborative Apple-Google exposure notification system (Apple-Google 2020), which several states have adopted as the framework for their own digital contact tracing efforts.

Irrespective of approach, to be effective at mitigating the spread of SARS-CoV-2, the DCTT used in a TACT system must be capable of identifying epidemiologically meaningful contacts. At a minimum, this requirement means that DCTT must be accurate enough to detect when one individual comes into close contact (within 2 m) with another individual or identify when an individual enters an area (e.g., a restaurant) with a risk of virus transmission. Clinical evidence (Bourouiba 2020) and retrospective studies of the secondary attack rate of COVID-19 in different environmental settings (see Cheng et al. 2020; Rosenberg et al. 2020) suggest transmission risks rise with time of exposure, which has led public health agencies to use temporal thresholds around 15 min when evaluating transmission risk. For this reason, DCTT should also be able to provide an accurate measure of either absolute or relative location for a similar period of time. If location data and/or proximity data are accurately recorded, securely stored, and accessible to public health officials, these agencies could use individual or aggregated data to improve contact tracing, model the course of the pandemic, and coordinate testing and response resources. To adhere to the ethical guidelines for public health data (see Lee and Gostin 2009), it is critical that these TACT have a clearly stated purpose, that that data use is limited to the that purpose, and that data access is limited to public health personnel. For example, if the purpose of a TACT system is only to identify contact and facilitate tracer interviews limiting data collection to an individual's location and proximity history may be sufficient. However, if a TACT system is to send automated follow-up, notify individuals of changes in support services, or monitor the use of those services the system may need a broader set of data and may need to be accessible by a wider set of public health personnel.

7.3 Technological, Operational, and Environmental Constraints on Technology Assisted Contact Tracing

Whether a TACT system is able to identify epidemiologically meaningful contacts while also remaining aligned with public values, such as privacy, depends on at least three factors—(1) what *technologies* the system uses, (2) how the system is *designed* to use those technologies, and (3) the *environment* the system is used in. It is useful to distinguish between these factors because each plays a different role in determining the impacts and effectiveness of a particular TACT system.

Technology: The specific hardware and software a TACT system uses determine the capabilities of that system by setting limits on the collection, storage, and transfer of location and contact data. At present, TACT uses four geospatial technologies to automate the collection of user location data: Cellular, GNSS, WiFi, and BLE. Each of these technologies offered a different level of spatial accuracy and temporal resolution. Generally, location information derived from cellular towers that has a spatial accuracy between 1 and 5 km is not precise enough to identify meaningful contacts, particularly in rural areas with few tower locations. Locations derived using the global navigation satellite systems (GNSS), such as GPS, are typically accurate to 5–20 m (Lee et al. 2016), which means that those systems can potentially be used to identify co-located contacts. However, GNSS alone is likely not accurate enough to identify all high-risk interactions between individuals. WiFi network access points and BLE signals can be used to infer individual locations both indoors and outdoors from device scans (Kwet 2019). BLE signals extend approximately 10–100 m around a device depending on the hardware being used (Bluetooth 2020). By combining WiFi and BLE information with location data from a GNSS, a TACT system could improve location accuracy to <1 m, but the exact level of accuracy will depend on obstructions (e.g., walls in buildings) and network density.

How accurately each of these technologies record the location and proximity histories of users depends on the hardware (e.g., antenna, chipset) and software (e.g., operating system, application) of each individual user's mobile device. For proximity tracking systems that rely on BLE exchange, the Bluetooth chipset of a device determines the strength of the signals sent by the device and how a device interprets incoming signals. Antenna position also affects how well signals are received. Chipsets and antenna orientation are customized by manufacturers (Bluetooth 2020), which makes it possible that the same signal will be interpreted differently by different devices. Software configurations can similarly impact the spatial accuracy of BLE by altering the transmission power, broadcasting interval, and duration of signals. In a review of 20 contact tracing applications, Zhao et al. (2020) show that these factors vary considerably across applications and that the minimal device-specific tuning that has been completed to date for these applications raises questions about their effectiveness. Positional accuracy similarly varies based on the model and configuration of each individual mobile device using GNSS or WiFi (see Menard et al. 2011). In addition users may disable location services, which restricts signal use and lowers positional accuracy.

Operational Design: While the technologies used in TACT set limits on how a system might function, how location and contact data are actually gathered, stored, transferred, and shared in practice are choices made by the system developers. Design decisions should be made by the public health agencies that commission or are responsible for the development of the central application and are ultimately responsible for the contact tracing effort. These agencies have the knowledge, experience, and responsibility to weigh the need for the type and quantity of information about an individual need to respond to COVID-19 against the ethical responsibilities of the public health agency and the privacy tolerance of the public. When location and proximity data are gathered it may be useful to also collect and link related user data (e.g., activities) that can later be used to prompt user memory during contact tracing interviews. At the same time, public health officials must collaborate with developers to make decisions about the frequency at which DCTT will gather location data. For example, a developer creating a BLE-based application can modulate not only how often a mobile device emits signals but how strong those signals are.

Once gathered, another key design decision public health agencies must make is how and where user data will be stored. TACT can be centralized or decentralized. Centralized systems store user location and contact data on servers. When a user tests positive and notifies the server, the system will analyze its overall database to determine exposure risk and notify those potentially exposed. While there are privacy-preserving protocols for this type of system (e.g., Inria 2020), decisions remain as to how potential exposure will be determined and if/when notification will be given and by whom, or whether key notification functions will be automated. Decentralized systems attempt to avoid the privacy concerns associated with a central server by hosting user data on their local devices and only occasionally synchronizing that data with a larger database. Nonetheless, decisions must be made about how long to store data, how frequently to synchronize, and how to provide notifications.

As data is stored and shared in either type of system, public health agencies must make policies related to the circumstances under which data will be released to with third parties (e.g., government agencies, employers), even in aggregated form, and whether data will be combined with other available data (e.g., purchasing histories). Public health agencies under ethical and legal obligations to collect the least amount of data possible, store and use the data safely, and dispose of the data once the public health effort is complete would not share data unless compelled to do so. However, agencies but could face circumstances, such as subpoenas, where other seek to compel that information from the agency. Similarly, agencies must balance the potential benefits of linking location or proximity data with other forms of data that could benefit contact tracing efforts and analysis of COVID-19 etiology, but will also likely raise risks of user identification or loss of public trust. Public trust in TACT systems and the public health departments operating them is essential for public use of the system and therefore their ability to gather needed information and support mitigation efforts.

Environment: The performance of DCTT can vary dramatically with the environment in which they are used. In an open field with an unobstructed view of the sky,

a GNSS receiver can produce location data accurate to within a meter. In obstructed environments (e.g., canopy cover, buildings) the accuracy of GNSS can be in the tens of meters, limiting their usefulness in certain forms of contact tracing. To assess the potential effectiveness of TACT, we need to also consider the environments in which location and contact data will be collected.

For example, location data collected using a GNSS receiver will typically have greater spatial accuracy than location data collected using a cellular receiver. However, this relationship may be reversed in dense urban environments. Signal interference and multipath errors produced by tall buildings can lower the accuracy of location data gathered using a GNSS receiver (Lachapell et al. 2018), but the greater number of cellular towers in urban areas can raise the accuracy of location data collected using a cellular receiver. GNSS accuracy in multipath urban environments can be improved if combined with WiFi, as demonstrated by Merry and Bettinger (2019). However, the horizontal position accuracy of GNSS and WiFi-enabled mobile devices is between 5 and 15 m in multipath urban environments and can vary with time of day and configuration of the surrounding urban environment (Zandbergen 2009; Menard et al. 2011; Garnett and Stewart 2015). Szot et al. (2019) suggest that static vertical position accuracy for devices using a combination of location technologies is in the 3–4 m range. With this level of accuracy, we can use DCTT to create location histories for individuals but may find it difficult to pinpoint their location to the level of a room in a building. Vertical accuracy may be particularly problematic in this regard as an error of 4 m could place an individual on two to three different floors of a building.

Proximity detecting technologies face similar challenges. BLE signals can penetrate some forms of building materials such as glass and wood. This feature creates the possibility that individuals located in different rooms or different floors within a building may be identified as being in contact when in reality, they were never in a situation in which SAR-CoV-2 could be transmitted. False-positive or false-negative contacts may also be generated if surrounding WiFi signals use channels that overlap with BLE signals. This overlap can also create interference between signals, which can impact the accuracy of distance measurements (Wen et al. 2020).

7.4 Assessing the Ethical and Privacy Implications of Technology-Assisted Contact Tracing

The ethical and privacy implications location tracking technologies raise are well-documented in the literature (Armstrong and Ruggles 2005; Curtis et al. 2006; Krumm 2009; Johnson and Sieber 2013), and the related study of the ethical and privacy implications of DCTT and TACT is now beginning (see Chan et al. 2020; Hekmati et al. 2020; Wen et al. 2020). The literature on the ethics and privacy of location tracking technologies revolves around issues of (i) notice-and-consent to

location data collection and (ii) the risk of revealing the identity of an individual through their location history.

While the application of location tracking technologies to contact tracing is relatively new, public health agencies have long grappled with the implications of collecting, storing, handling and using confidential information (Fairchild et al. 2007; Lee and Gostin 2009). Public health ethical practices enforce the principal that all confidential information that could bring harm to either community or the individual must be protected by public health institutions (Thomas et al. 2002). Further, any framework for designing public health programs must consider threats to privacy, particularly disease surveillance or contact tracing, prior to implementation (Kass 2001). Even in extreme public health situations, such as global HIV/AIDS epidemic, personal information is guarded by a structure of administrative and regulatory protections within the public health administration (Bayer and Fairchild 2002).

Providing notice to an individual that their location data will be collected and receiving their affirmative consent to do so are at the center of U.S. protections of location privacy (Boshell 2019; Rothenberg 2020). As early as 1994, Onsrud et al. (1994) introduced eight principles related to privacy and the handling of geospatial data, including—(1) limiting the collection of personal information, (2) only collecting relevant, accurate, and up-to-date data, (3) clearly stating the purpose of the data being collected and limiting use to those purposes, (4) not allowing secondary uses of personal information without individual consent or authorized by law, (5) protecting personal data, (6) open policies surrounding the use of personal data, (7) individual ability to inspect and correct their personal data files, and (8) data controllers should be held accountable for complying with the guidelines. These principles are similarly reflected in both the Public Health Code of Ethics (APHA 2019) and the GIS Code of Ethics (GIS Certification Institute 2003). The Public Health Code of Ethics establishes that for policies and practices to be ethical, special attention must be given to “protecting the privacy and confidentiality of individuals when gather data” and stresses the removal of “personal identifying information from the data set as soon as it is no longer needed.” Further, the data collected on individuals or communities should be limited to “only data elements and specimens necessary for disease control or protection.”

The GIS Code of Ethics states that it is a GIS professional’s duty to “allow individuals to withhold consent from being added to a database, correct information about themselves in a database, and remove themselves from a database.” More recently, these principles can be seen in rights-based approaches to privacy like the signal code of Greenwood et al. (2017), which focuses on crisis situations and argues that data privacy and security, data agency, and regress and rectification are fundamental rights that also facilitate crisis response. Empirical evidence suggests these principles are not often being applied to much of the location data being regularly collected from many mobile devices. See for example the cases of LocationSmart (Krebs 2018; Oremus 2018), or the illegal sale of augmented GPS data (Brodkin 2019).

Even in situations where the majority of the principles related to notice and consent to collect location data are maintained, the risk of identifying individuals remains. To preserve privacy, location data are de-identified and are often aggregated in time and

space. However, location data is difficult to fully anonymize. If individual identifying information (e.g., name, age) is removed from the data, but spatial information is unaltered, it is often easy to identify an individual. Research suggests that a small number of spatial-temporal locations is needed to identify a large portion of the U.S. population (De Montjoye et al. 2013). Even when data are spatially aggregated, it is often possible to identify individuals. Golle and Partridge (2009) suggest that when location data are aggregated to census tracts, but work and home locations can be inferred, half of the population can be identified as one of ten individuals. Incorporating demographic information can narrow this set to one.

7.5 Ethical and Privacy Principles Linked to Digital Contact Tracing Technologies

As DCTT emerge as a tool for pandemic response, an accompanying literature is developing around how these technologies can and should be used to mitigate and suppress COVID-19 and quell future outbreaks of communicable disease. To address this cross-cutting issue, contributing scholars are drawing central ideas from the public health literature (CDC 2020; Kahn 2020), privacy law (Gilmor 2020), digital data governance (Raskar et al. 2020), human rights frameworks (Morley et al. 2020), and moral philosophy (Morley et al. 2020; Gasser et al. 2020; Hart et al. 2020). While terminology varies slightly across authors, this emerging body of literature has coalesced around a set of five principles that address the potential privacy, ethical, and social impacts of these technologies. Here, we briefly present each principle along with an accompanying set of characteristics that define the principle, and a short description.

1. **Efficacy** (necessary, proportionate, scientifically valid, impactful): The best available scientific evidence should show that TACT will improve pandemic response efforts and that the positive effects DCTT create will outweigh the negative effects of these technologies. Those responsible for these systems should also monitor their performance and provide some measures of their ongoing positive and negative impacts. Compatibility, across-platform functionality, and backward compatibility are also key to effectiveness. It should also be recognized that effectiveness may be difficult to quantify and assign to different parts of the larger contact tracing effort, as contact tracing may also involve the use of auxiliary data collected during tracing interviews, outside data sources, or user inputs when at the time of application download.
2. **Privacy** (voluntary, consent, limited, anonymous, editable, secure, temporary): Participation in TACT should be voluntary and non-participation should not incur any punitive measures. Any data collected from users should be limited to the purposes of COVID-19 prevention, be securely stored, and destroyed when no longer relevant for this purpose. Users should be given clear notice about the types of data that will be collected and how that data will be stored and

transferred. User data collected for COVID-19 prevention should not be share with third parties. Users should also have the ability to opt out of participation at any time and have as much control over their data as possible.

3. **Equity** (equally available, equally accessible): DCTT should be free and available to any user that wishes to use them. These technologies should also be accessible to people of different backgrounds, levels of experience, incomes, and other differentiating characteristics. Some form of oversight should be designed to identify if DCTT are creating inequitable public health outcomes.
4. **Transparency** (open source, accessible, customizable): As an extension of the privacy and equity discussions above, the rules governing the collection and management of user data should be understandable and publicly accessible. Ideally, TACT systems would use open architectures and standards, so others can audit and amend the systems. Evaluating and monitoring TACT for data misuse, privacy, and other concerns is far easier if the underlying platform of the system is openly available.
5. **Accountability** (auditable, amendable): Assessing and ensuring the efficacy, equity, and privacy of DCTT and the TACT systems they are used in pandemic response relies on the auditability and amendability of these systems. TACT systems should undergo regular, independent assessments organized around the above principles. Those assessments should also be publicly accessible.

7.6 Connecting Ethical and Privacy Principles to Technological, Operational, and Environmental Constraints

Each of the five principles presented above must be linked with measurable characteristics of TACT and the DCTT that support their operation to be useful as assessment criteria. Table 7.1 contains a listing of key data about a TACT that can be collected and used to assess the ethical standing of the system. While we link each piece of data to the principle we believe it is most directly related to, we recognize that much of this data can be linked to multiple principles. For example, the spatial accuracy of the DCTT will clearly impact the efficacy of the contact tracing system. However, spatial accuracy also has clear and direct implications for privacy and equity. Higher spatial accuracy makes it easier to identify individuals, which can by extension lead to the differential treatment of individuals.

To explore the connections between each principle and the data that can be used by that principle, we frame the following discussion around the technical, operational, and environmental characteristics that may shape adherence to each principle in practice.

Efficacy: To monitor the efficacy of DCTT, it will be critical to gather data on the accuracy and precision of the proximity and location information these technologies collect. Understanding when and where accuracy and precision degrade will be key to

Table 7.1 Key data needed to assess the ethical and privacy implications of technology-assisted contact**Efficacy**

- How is contact defined as proximity in space and duration in time?
- How frequently is location and/or proximity data collected?
- What technologies are used in location and/or proximity data collection?
- What is the locational accuracy and precision of these technologies?
- What is the consistency of accuracy and precision of the location/proximity data across mobile devices?
- What is the consistency of accuracy and precision of the location/proximity data across environments?

Privacy

- Is the technology optional to install and use?
- Is there an articulated plan for the collection, storage, and use of data that is consistent with public health ethics principles and guidance for best practices (e.g., public health code of ethics)?
- Is there a mechanism for providing notice of private data collection, storage, and transfer?
- Is there a mechanism for gathering consent to collect and use private data?
- What is the timing of consent (e.g., is consent gathered prior to data collection, data release)?
- What are the limitations on the use of data collected (e.g., limited to contact tracing effort)?
- What measures are taken to preserve anonymity (e.g., will data be stored on user devices)?
- How will the data be protected from unauthorized use (e.g., security protocols)?
- What are the third party sharing and data transfer policies (e.g., shared with other government agencies)?
- Are there any exceptional situations in which the data will be shared for non-public-health purposes (e.g., subpoena)?
- What are the user rights/abilities to edit or delete their data?
- What is the plan for data deletion (e.g., will data be deleted when it is no longer useful for contact tracing)?
- Has a privacy officer been appointed to ensure effective privacy/ethical standards?

Equity

- What mobile devices support the contact tracing system?
- Is there a cost to users? If so, what is it?
- Are there features for users with disabilities (e.g., design for visually or hearing impaired)?
- Are interfaces designed for all users?
- What are the options for individuals without access to the preferred TACT platform?

Transparency

- Is the underlying data collection code available and editable?
- Is data storage plan available?
- Is the privacy officer available to relate how ethical standards were upheld?
- Was the public included in the development and implementation of the TACT?

Accountability

- What criteria are used to assess the impact and efficacy of the TACT?
- Is there an independent third party auditor of the system?
- Will impact assessments and efficacy audits be made public?
- Is there a stated process to update the TACT as new information becomes available?
- Is there a plan to phase out the system as the pandemic recedes?

adjusting or augmenting these systems to increase effectiveness. As presented above, variations in the environment and the spatial distribution of key infrastructures can alter the effectiveness of these technologies. However,

DCTT should also be examined for device-specific performance variations. Early research in this area by Wen et al. (2020) suggests that different devices receiving the same signal may not produce the same proximity measurements. Adjusting and correcting for how different devices process BLE of GNSS signals to ensure accuracy will be critical to improving overall system efficacy. Linking those adjustments to our evolving understanding of the dynamics of SAR-CoV-2 transmission will also be a key to maintaining efficacy. To do so we will need to know how contact is defined by different systems and whether that definition is updating as our knowledge progresses.

In addition to these technical limitations, efficacy will also be determined by adoption. Hinch et al. (2020) suggest that TACT can begin to produce protective effects when adoption rates are as low as 10% of the population and that benefits rise with greater adoption. A finding reinforced by a systematic review by Braithwaite et al. (2020). However, adoption rates have varied by country. Countries that have mandated adoption, such as Qatar, have seen adoption rates over 90%, but most countries with official DCTT have adoption rates below 20% (Rivero 2020). To assess efficacy, the performance of any particular DCTT and related TACT needs to be placed in this social context. A perfectly accurate and highly precise DCTT is unlikely to be effective at producing population-level benefits if it has a small number of non-representative users. Plans to include users and affected communities in TACT development are one way to raise the chances of voluntary use of these systems. While community involvement may present practical challenges, increased use may offset such potential costs.

Privacy: DCTT adoption appears to be limited in large part by the privacy concerns of potential users. Academics and civil liberties groups (see Gilmor 2020) acknowledge the productive role DCTT can play in pandemic response, but also caution about the potential for misuse of user data or possible expanded use of user data beyond a system's initial intention. These concerns are grounded in the misuse and sale of user data observed in other forms of application-based data collection and the economic incentives to collect and sell information about users (Zuboff 2019).

Whether a TACT system violates privacy is largely a function of the operational decisions of those responsible for the system. To understand whether DCTT are being used by TACT developers in ways that challenge privacy, we therefore need to gather information about how data are managed during all phases of the contact tracing effort. The need for this information arises not because public health agencies are conducting contract tracing, but because new entities (e.g., private companies) are involved in the data collection effort. Public health agencies have a well-established record of performing contact tracing and a clear set of ethical principles and related policies and laws that guide their behavior (Thomas et al. 2002, Lee and Gostin 2009, APHA 2019, Gostin and Wiley 2016). To preserve privacy, data must be gathered with clear notice to and consent from the user and notice should

be given in a way that is understandable by a non-specialist. Information about how user data is deidentified, transferred, and stored is also needed to ensure the system has adequate measures in place to preserve user anonymity. Decentralized systems focused on proximity measurement that do not collect location data have the advantage of reducing opportunities for later reidentification.

However, decentralized TACT systems still need to be evaluated for editability and the length of data storage. Editability can be assessed by gathering information about how and when systems prompt users to retain their data, share their data with contact tracers, and which specific aspects of their data users can change or remove. Systems may for instance, allow users to decide if their location and proximity data are shared with a central authority but not allow users to edit those same histories. Similarly, the Center for Disease Control and Prevention (2020) notes that TACT following privacy best practices will only retain user data as long as they are useful for disease prevention. The ACLU suggests that this condition means that the location and proximity histories of users should be deleted after 14 days as this is the typical length of the contagious period for COVID-19. This practice has the added benefit of removing incentives to aggregate and sell user data to third parties for unrelated purposes. Finally, TACT systems should share information about the editing and data deletion process in user agreements so users can make informed consent decisions.

Equity: Actions that preserve privacy may have the additional benefit of enhancing equity. In its simplest form, equity can be measured by access to DCTT. If a particular DCTT is not available to all users it is unlikely to achieve the broadest and most equitable outcomes. Access can be measured by cost of use. For example, does a user have to pay to use the system? However, cost of use can also be measured in effort and accessibility. For example, TACT systems should have intuitive user interfaces that can be used by all members of society (e.g., those with visual impairment). Access is also a function of the platforms a TACT supports. Most TACT work across mobile device platforms, but some socially underprivileged groups may not have access to these platforms further placing an undue burden on them. Additionally, socially privileged groups may be allowed to opt out of the program adding additional pressure to a group that are already stressed. Lastly, these systems should also be assessed for their interoperability with the different contact tracing systems of different public health agencies. If health professionals cannot access data approved for release by users these systems may be less useful to mitigation efforts. The Association of Public Health Laboratories is now developing a national server and systems to facilitate state-to-state interoperability in the United States.

Transparency/Accountability: Ensuring the equity of TACT systems is closely related to maintaining their transparency and accountability. Understanding if a TACT is producing equitable outcomes will require systematic reviews of who the users of a system are, how those users are using the DCTT, and whether those technologies are underperforming expectations in particular environments. To date, we have limited evidence in this area. While some of the technical details for DCTT are generally available, the operational details of TACT are less accessible and often can only be gathered with some effort. At the same time, many of these systems do not

have clear auditing plans or procedures outlining how amendments will be made and what triggers those amendments. This situation is somewhat understandable given that many of these systems are new. However, Clear operational rules and guidelines need to be created and matched with enforcement mechanisms to ensure these systems enact espoused principles.

The framework needed for a clear chain of responsibility and enforcement is already in place under the auspices of the state public health authority. To ensure efficacy, these systems should also be customized and amended as new evidence becomes available about disease transmission and system effectiveness. For example, there is increasing evidence that SAR-CoV-2 may be transmissible through the air and not just via respiratory droplets, which has prompted revision of CDC guidelines for the disease (CDC 2020). If this proves to be the case, TACT systems will need to be amended to reflect the science and remain effective at capturing transmission risk. For instance, if SAR-CoV-2 can remain suspended in the air for a period after an individual leaves a location, capturing movement by other into and out of that location increases in importance.

7.7 Technology-Assisted Contact Tracing in the Context of the COVID-19 Pandemic

At the time of writing, TACT has played different roles in the response to COVID-19 across the globe (see Kahn 2020; Lee et al. 2020; Lin et al. 2020). In the United States, DCTT and TACT are only beginning to emerge as part of the toolkit for pandemic response. However, even as these technologies emerge their effectiveness and impact is constrained by a number of factors. First, TACT is most effective in environments with widespread, easily accessible, and regularized testing for COVID-19. Recording location and proximity histories of users is only useful for disease prevention if users exchanging information are tested and then share that information with public health authorities or other users in privacy-preserving ways. Without knowledge of infection, chains of transmission cannot be identified and broken. Negative results are equally useful because they provide health professionals with a clearer picture of disease prevalence in the population and allow for more accurate estimation of key indicators like case fatality rates. Using the framework presented here in an environment without adequate testing resources may skew evaluation of DCTT, as the TACT they support may simply not be receiving enough information to be effective.

Second, the availability of DCTT does not necessarily mean that these technologies will be adopted and widely used. In the United States, it remains unclear whether DCTT will be widely adopted by the public. However, the benefits of these technologies are subject to strong network effects—the more people that adopt a DCTT the more benefit it produces. To facilitate adoption DCTT developers could focus their release and adoption efforts on sub-segments of the population (e.g., in a hotspot city) to ensure broad localized use. Farranato et al. (2020) identify that this strategy

is commonly used to successfully launch mobile applications of all types and has the virtues of building a critical mass of users, which allows for the gradual scaling of adoption. If this approach to adoption is used, the evaluation criteria outlined here will need to account for the localized nature of the deployment strategy. In the simplest case, efficacy would need to be measured in the adoption rate within the target sub-population and not as a raw count of users. Similarly, additional considerations for privacy may be needed. If the target population is small enough it may be more difficult to preserve anonymity. This selected release strategy could also run counter to the equality principles outlined here. Nonetheless, targeted release and widespread adoption within an urban center that is a hotspot of SAR-CoV-2 transmission could have clear benefits. Another approach to expanding adoption could be to focus effort during TACT development on gathering future user feedback and involving affected communities in meaningful ways in system development. Such an approach would not only match the ethical principles of the public health profession, it would likely raise adoption rates because it would increase community trust in TACT.

Third, to some extent, an undirected version of the targeted release strategy is becoming the default for the use of TACT in many locations. In the United States, for example, state governments have made different decisions about whether to use TACT in the response efforts. At the same time, DCTT are being deployed at schools, universities, and corporations. The use of these technologies in these settings raises a host of additional questions that extend the assessment criteria presented in this chapter beyond the technical, operational, and environmental considerations presented here. One key difference are the incentives of the parties responsible for the data collection and management. As Zuboff (2019) demonstrates, companies have clear incentives to collect large quantities of diverse data on users as this information creates economies of scale and scope in secondary data markets. These incentives therefore encourage behavior in direct opposition to the limited, purpose-specific, and collaborative forms of data collection outlined in the public health code of ethics. More directly, detailed location and proximity data could be used by companies to monitor behavior or map social interactions. For firms, this type of information could be used to increase the efficiency of production through the reorganization of labor, but at the expense of worker autonomy and self-determination. Another key difference is the responsibilities some non-governmental organizations have for their members. For example, nursing care facilities may carry a duty of care to all residents that the institution believes would be fulfilled by digital contact tracing but may simultaneously conflict with the wishes of individual residents. Primary and secondary schools may have similar *in loco parentis* responsibilities that might come into conflict with the rights or preferences of their individual students.

Fourth, how effective DCTT are rests on how they are integrated into the wider set of activities that make up a contact tracing effort. While this chapter, and much of the emerging literature on DCTT and TACT, focuses on whether and how well technologies can record location and proximity histories, it is as important to consider how these technologies may be used to ethically facilitate follow-up activities and service provision. For example, the location data gathered by a TACT system could

be used to automatically identify which service providers are located closest to an individual isolating after potential exposure to COVID-19. This may be ethical and helpful if the individual volunteered their address information to the system, but becomes questionable if the address is inferred from the data in the system. As another example, in a situation where contact tracers are overwhelmed by the scale of the pandemic, data gathered by DCTT could be used to create prioritization schemes to identify individuals which individuals should be scheduled for interviews in what order. Simple summary statistics of the number of people an individual that has received a positive test has come into contact with, or the number and diversity of locations they have visited could be calculated from proximity and location histories to support such an effort. Creating such a scheme would force us to reengage the principles identified above and outlined in the GIS and public health codes of ethics.

The need to understand these and the many other questions that will arise as DCTT evolve during and after the COVID-19 pandemic is an opportunity for geographic research. Addressing these questions will provide a foundation for public health geographers, legal geographers, economic geographers, critical geographers, as well as others from the field's varied sub-disciplines to critically examine the implications of DCTT through their unique lens. Further, it demonstrates the importance of having spatial scientists and scholars involved in the evaluation and analysis of technologies that potentially have such far-reaching ethical implications.

References

- American Public Health Association. (2019) Public health code of ethics. Retrieved November 16, 2020, from https://www.apha.org/-/media/files/pdf/membersgroups/ethics/code_of_ethics.ashx.
- Apple-Google. (2020) Privacy-preserving contact tracing. <https://covid19.apple.com/contacttracing>.
- Armstrong, M. P., & Ruggles, A. J. (2005) Geographic information technologies and personal privacy. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 40(4), 63–73.
- Bai, Y., Yao, L., Wei, T., et al. (2020). Presumed asymptomatic carrier transmission of COVID-19. *JAMA*, 323(14), 1406–1407.
- Bayer, R., & Fairchild, A. (2002). The limits of privacy: Surveillance and the control of the disease. *Health Care Analysis*, 10, 19–35.
- Bluetooth. (2020). What is the range of Bluetooth? Retrieved April 24, 2020, from <https://www.bluetooth.com/learn-about-bluetooth/bluetooth-technology/range/>.
- Boshell, P. (2019) *The power of place: Geolocation tracking and privacy*. American Bar Association.
- Bourouiba, L. (2020). Turbulent gas clouds and respiratory pathogen emissions: Potential implications for reducing transmission of COVID-19. *JAMA*, 323(18), 1837–1838.
- Braithwaite, I., Calendar, T., Bullock, M., et al. (2020). Automated and partly automated contact tracing: A systematic review to inform the control of COVID-19. *The Lancet Digital Health*. [https://doi.org/10.1016/S2589-7500\(20\)30184-9](https://doi.org/10.1016/S2589-7500(20)30184-9).
- Brodkin, K. (2019) Selling 911 location data is illegal—U.S. carriers reportedly did it anyway. *Ars Technica*. Retrieved October 02, 2020, from <https://arstechnica.com/tech-policy/2019/02/att-t-mobile-sprint-reportedly-broke-us-law-by-selling-911-location-data/>.

- Cates, W., Rothenberg, R., & Blount, J. (1996). Syphilis control: The historic context and epidemiologic basis for interrupting sexual transmission of *Treponema pallidum*. *Sexually Transmitted Diseases*, 23(1), 68–75.
- Center for Disease Control and Prevention. (2020). Preliminary criteria for the evaluation of digital contact tracing tools for COVID-19 – version 1.2 (U.S. Center for Disease Control and Prevention, 2020)
- Chan, J., Gollakota, S., Horvitz, E., et al. (2020). PACT: Privacy sensitive protocols and mechanisms for mobile contact tracing. arXiv preprint [arXiv:2004.03544](https://arxiv.org/abs/2004.03544).
- Chen, Y.-D., Tseng, C., King, C.-C., et al. (2007). Incorporating geographical contacts into social network analysis for contact tracing in epidemiology: A study of Taiwan SARS data. In D. Zeng, et al. (Eds.), *Intelligence and security: Biosurveillance*. Lecture Notes in Computer Science (Vol. 4506). Berlin: Springer.
- Cheng, H. Y., Jian, S. W., Ng, T. C., et al. (2020). Taiwan COVID-19 Outbreak Investigation Team. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Internal Medicine*. <https://doi.org/10.1001/jamaintmed.2020.2020>.
- Curtis, A., Mills, J. W., & Leitner, M. (2006). Keeping an eye on privacy issues with geospatial data. *Nature*, 441(7090), 150–150.
- De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., et al. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3, 1376.
- Dong, W., Guan, T., Lepri, B., et al. (2019). PocketCare: Tracking the flu with mobile phones using partial observations of proximity and symptoms. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2), 1–23.
- Eames, K. T., & Keeling, M. J. (2003). Contact tracing in disease control. *Proceedings of the Biological Sciences*, 270, 2565–2571.
- Fairchild, A., Gable, L., Gostin, L., et al. (2007). Public goods, private data: HIV and the history, ethics, and uses of identifiable public health information. *Public Health Reports*, 122, 7–15.
- Farrahi, K., Emonet, R., & Cebrian, M. (2014). Epidemic contact tracing via communication traces. *PLoS One*, 9(5). <https://doi.org/10.1371/journal.pone.0095133>
- Farranato, C., Iansiti, M., Bartosiak, M., et al. (2020). How to get people to actually use contact-tracing apps. *Harvard Business Review*. Retrieved October 03, 2020, from <https://hbr.org/2020/07/how-to-get-people-to-actually-use-contact-tracing-apps>.
- Ferretti, L., Wymant, C., Kendall, M., et al. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science*, 368(6491), eabb6936.
- Finger, F., Genolet, T., Mari, L., et al. (2016). Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. *Proceedings of the National Academy of Sciences of the United States of America*, 113(23), 6421–6426.
- Garnett, R., & Stewart, R. (2015). Comparison of GPS units and mobile Apple GPS capabilities in an urban landscape. *Cartography and Geographic Information Science*, 42, 1–8.
- Gasser, U., Ienca, M., Scheibner, J., et al. (2020). Digital tools against COVID-19: Taxonomy, ethical challenges, and navigation aid. *The Lancet Digital Health*. [https://doi.org/10.1016/s2589-7500\(20\)30137-0](https://doi.org/10.1016/s2589-7500(20)30137-0).
- Gilmor, D. (2020). Principles for technology-assisted contact-tracing. *American Civil Liberties Union*.
- GIS Certification Institute. (2003). *GIS code of ethics*. Retrieved October 10, 2020, from <https://www.gisci.org/Ethics/CodeofEthics.aspx>.
- Golle, P., & Partridge, K. (2009). On the anonymity of home/work location pairs. In *International Conference on Pervasive Computing* (pp. 390–397). Berlin, Heidelberg: Springer.
- Goldman, E. (2020). Exaggerated risk of transmission of COVID-19 by fomites. *The Lancet Infectious Diseases*, 20(8), 892–893.
- Gostin, L., & Wiley, L. (2016). *Public health law: Power, duty, restraint*. Oakland, CA: University of California Press.

- Greenwood, F., Howarth, C., Poole, D., et al. (2017). The signal code: a human rights approach to information during crisis. *Signal human security and technology—Standards and ethics series* 02. Harvard, MA.
- Hart, V., Siddarth, D., Cantrell, B., et al. (2020). Outpacing the virus: Digital response to containing the spread of COVID-19 while mitigating privacy risks, COVID-19 rapid response impact initiative—white paper 5 (Harvard University, Edmond J. Safre Center for Ethics, 2020).
- He, X., Lau, E. H. Y., Wu, P., et al. (2020). (2020) Author correction: Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*, 26, 1491–1493.
- Hekmati, A., Ramachandran, G., & Krishnamachari, B. (2020). CONTAIN: Privacy-oriented contact tracing protocols for epidemics. arXiv preprint. [arXiv:2004.05251](https://arxiv.org/abs/2004.05251).
- Hinch, R., Probert, W., Nurtay, A., et al. (2020). Effective configurations of digital contact tracing app: a report to NHSX. https://github.com/BDI-pathogens/covid-19_instant_tracing.
- Inria. (2020). ROBERT: ROBust and privacy-presERving proximity tracing protocol. Retrieved October 05, 2020, from <https://github.com/ROBERT-proximity-tracing>.
- Johnson, P. A., & Sieber, R. E. (2013). Situating the adoption of VGI by government. In D. Sui, S. Elwood, & M. Goodchild (Eds.), *Crowdsourcing geographic knowledge* (pp. 65–81). Dordrecht: Springer.
- Judson, F., & Vernon, T. (1988). The Impact of AIDS on state and local health departments: Issues and a few answers. *American Journal of Public Health*, 78, 387–393. <https://doi.org/10.2105/AJPH.78.4.387>.
- Kahn, J. (Ed.). (2020). *Digital contact tracing for pandemic response*. Johns Hopkins University Press. <https://doi.org/10.1353/book.75831>.
- Kampf, G., Todt, D., Pfaender, S., et al. (2020). Persistence of coronaviruses in inanimate surfaces and their inactivation with biocidal agents. *Journal of Hospital Infection*, 104, 246–251.
- Kanamori, H. (2020). Rethinking environmental contamination and fomite transmission of SARS-CoV-2 in the healthcare. *Journal of Infection*. <https://doi.org/10.1016/j.jinf.2020.08.041>.
- Kass, N. (2001). An ethics framework for public health. *American Journal of Public Health*, 91, 1776–1782. <https://doi.org/10.2105/AJPH.91.11.1776>.
- Klov Dahl, A. S., Graviss, E. A., Yaganehdoost, A., et al. (2001). Networks and tuberculosis: And undetected community outbreak involving public places. *Social Science and Medicine*, 52(5), 681–694.
- Krebs, B. (2018). Tracking firm LocationSmart leaded location data for customers of all Major US mobile carriers without consent in real time via its website. *Krebs Security*. Retrieved October 02, 2020, from <https://krebsonsecurity.com/2018/05/tracking-firm-locationsmart-leaked-location-data-for-customers-of-all-major-u-s-mobile-carriers-in-real-time-via-its-web-site/>.
- Krumm, J. (2009). A survey of computational location privacy. *Personal and Ubiquitous Computing*, 13(6), 391–399.
- Kwet, M. (2019). In stores, secret Bluetooth surveillance tracks your every move. *The New York Times*. Retrieved April 24, 2020, from <https://www.nytimes.com/interactive/2019/06/14/opinion/bluetooth-wireless-tracking-privacy.html>.
- Lachapell, G., Gratton, P., Horrelt, J., et al. (2018). Evaluation of a low cost hand held unit with GNSS raw data capability and comparison with an Android smartphone. *Sensors*, 18(2), 4185.
- Lee, L., & Gostin, L. (2009). Ethical collection, storage, and use of public health data: A proposal for a national privacy protection. *JAMA*, 302, 82–84.
- Lee, L., Jones, M., Ridenour, G. S., et al. (2016). Comparison of accuracy and precision of GPS-enabled mobile devices. In 2016 IEEE International Conference on Computer and Information Technology (pp. 73–82).
- Lee, V. J., Chiew, C. J., & Khong, W. X. (2020) Interrupting transmission of COVID-19: Lessons from containment efforts in Singapore. *Journal of Travel Medicine*, 27, taaa039.
- Lin, C., Braund, W. E., Auerbach, J., et al. (2020). Policy decisions and use of information technology to fight COVID-19, Taiwan. *Emerging Infectious Diseases*, 6, 1506–1512.

- McElroy, P. D., Rothenberg, R. B., Varghese, R., et al. (2003). A network-informed approach to investigating a tuberculosis outbreak: Implications for enhancing contact investigations. *The International Journal of Tuberculosis and Lung Disease*, 7, S486–S493.
- Menard, T., Miller, J., Mowak, M., et al. (2011). Comparing the GPS capabilities of the Samsung Galaxy S, Motorola Droid X, and the Apple iPhone for vehicle tracking using FreeSim_Mobile. In *14th International IEEE Conference on Intelligent Transportation Systems*, Washington, DC.
- Merry, K., & Bettinger, P. (2019). Smartphone GPS accuracy study in an urban environment. *PLoS ONE*, 14(7). <https://doi.org/10.1371/journal.pone.0219890>.
- Morley, J., Cowls, J., Taddeo, M., et al. (2020). (2020) Ethical guidelines for COVID-19 tracing apps. *Nature*, 582, 29–31.
- Mizumoto, K., Kagaya, K., Zarebski, A., & Chowell, G. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*, 25(10), 2000180.
- Onsrud, H. J., Johnson, J. P., & Lopez, X. (1994). Protecting personal privacy in using geographic information systems. *Photogrammetric Engineering and Remote Sensing*, 60(9), 1083–1095.
- Oremus, W. (2018). The privacy scandal that should be bigger than Cambridge Analytica. *Slate*. Retrieved October 02, 2020, from <https://slate.com/technology/2018/05/the-locationsmart-scandal-is-bigger-than-cambridge-analytica-heres-why-no-one-is-talking-about-it.html>.
- Raskar, R., Schunemann, I., Barbar, R., et al. (2020). Apps gone rouge: Maintaining personal privacy in an epidemic. [arXiv:2003.08567v1](https://arxiv.org/abs/2003.08567v1).
- Riley, S., Fraser, C., Donnelly, C. A., et al. (2003). Transmission dynamics of the etiological agent of SARS in Hong Kong: Impact of public health interventions. *Science*, 300(5627), 1961–1966.
- Rivero, N. (2020). Global contact tracing app downloads lag behind effective levels. *Quartz*. Retrieved October 04, 2020, from <https://qz.com/1880457/global-contact-tracing-app-downloads-lag-behind-effective-levels/>.
- Rosenberg, E. S., Dufort, E. M., Blog, D. S., et al. (2020). New York State Coronavirus 2019 Response Team. COVID-19 testing, epidemic features, hospital outcomes, and household prevalence, New York State-March 2020. *Clinical Infectious Diseases*. <https://doi.org/10.1093/cid/ciaa549>.
- Rothenberg, M. (2020). *The privacy law sourcebook 2020*. Electronic Privacy Information Center.
- Thomas, J., Sage, M., Dillenber, J., & Guillory, V. J. (2002). A code of ethics for public health. *American Journal of Public Health*, 92, 1057–1059. <https://doi.org/10.2105/AJPH.92.7.1057>.
- Saurabh, S., & Prateek, S. (2017). Role of contact tracing in containing the 2014 Ebola outbreak: A review. *African Health Sciences*, 17(1), 225–236.
- Slifka, M., & Gao, L. (2020). Is presymptomatic spread a major contributor to COVID-19 transmission? *Nature Medicine*, 26, 1531–1533.
- Stadnytskyi, V., Bax, C. E., Bax, A., et al. (2020). The airborne lifetime of small speech droplets and their potential importance in SARS-CoV-2 transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 117(22), 11875–11877.
- Steinhauer, J., & Goodnough, A. (2020) Contact tracing is failing in many states. Here's why. *New York Times*. Retrieved September 17, 2020, from <https://www.nytimes.com/2020/07/31/health/covid-contact-tracing-tests.html>.
- Szot, T., Specht, C., Specht, M., et al. (2019). Comparative analysis of positioning accuracy of Samsung Galaxy smartphones in stationary measurements. *PLoS ONE*, 14(4), e0215562.
- Wen, H., Zhao, Q., Lin, Z., et al. (2020) A study of the privacy of covid-19 contact tracing apps. In *International Conference on Security and Privacy in Communication Networks*.
- Wesolowski, A., Buckee, C., Bengtsson, L., et al. (2014a). Commentary: Containing the Ebola outbreak—The potential and challenge of mobile network data. *PLOS Currents*, 6. <https://doi.org/10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e>.
- Wesolowski, A., Stresman, G., Eagle, N., et al. (2014b). Quantifying travel behaviour for infectious disease research: A comparison of data from surveys and mobile phones. *Scientific Reports UK*, 4, 5678.

- Zandbergen, P. A. (2009). Accuracy of iPhone locations: A comparison of assisted GPS WiFi and Cellular Positioning. *Transactions in GIS*, 13, 5–26.
- Zhao, Q., Wen, H., Lin, Z., et al. (2020). On the accuracy of measured proximity of Bluetooth-based contact tracing Apps.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. Profile Books.

Chapter 8

Challenges and Limitations of Geospatial Data and Analyses in the Context of COVID-19



Sean G. Young, Jyotishka Datta, Bandana Kar, Xiao Huang,
Malcolm D. Williamson, Jason A. Tullis, and Jackson Cothren

8.1 Reproducibility, Replicability, and Provenance

Connections between human health and geographic location have been recognized since at least the time of the ancient Greek physician Hippocrates (Hippocrates 1849). While certainly not a new concept, efforts to study human health and disease using geospatial data and analyses have only recently begun to attract widespread attention outside of a few specialized disciplines (Barrett 2000). In particular, the COVID-19 pandemic has brought this relatively small but growing research field into the public spotlight and generated unprecedented demand for maps and analyses, with the result that more researchers and data scientists than ever before are using geospatial data. The associated high profile geospatial analyses of COVID-19 data have amplified questions of reproducibility and replicability (R&R) and related issues such as trust, privacy, and quality across the information technology (IT)-intensive sciences. Many complex geospatial R&R issues can be simplified or better understood by a focus on who has (or should have) access to associated geospatial provenance information (Tullis and Kar 2020) which is defined below.

S. G. Young (✉)

Department of Environmental and Occupational Health, University of Arkansas for Medical Sciences, Little Rock, AR, USA

e-mail: sgyoung@uams.edu

J. Datta

Department of Statistics, Virginia Tech, Blacksburg, VA, USA

B. Kar

Oak Ridge National Laboratory, Oak Ridge, TN, USA

X. Huang · J. A. Tullis · J. Cothren

Department of Geosciences and Center for Advanced Spatial Technologies, University of Arkansas, Fayetteville, AR, USA

M. D. Williamson

Center for Advanced Spatial Technologies, University of Arkansas, Fayetteville, AR, USA

Shortly before the pandemic, an expert panel chaired by Harvey V. Fineberg published a *consensus study report* on R&R representing the official view of the National Academies of Sciences, Engineering, and Medicine (National Academies 2019). As adopted in the report, “*reproducibility* is obtaining consistent results using the same input data; computational steps, methods, and code; and conditions of analysis.... [and *replicability*] is obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data (National Academies 2019).” While these definitions are clearly distinct, they are also linked. For example, if a COVID-19 dashboard cannot be *reproduced* due to lack of contextual metadata, comparative interpretation of differences in a *replicated* dashboard becomes problematic.

Provenance records are well understood to be critical for authenticity in the art domain where they “[entail] an artifact’s complete ownership history, but ideally will also include artistic, social, and political influences upon the work from its creation to the present day (Tullis et al. 2015).” The Worldwide Web Consortium’s (W3C’s) Provenance Incubator Group defines provenance of a resource (e.g., a COVID-19 map of per-capita deaths) as a contextual metadata “record that describes entities and processes involved in producing and delivering or otherwise influencing that resource (W3C Provenance Incubator Group. 2010).” Careful recordkeeping with detailed provenance information is required to ensure reproducibility (and therefore trust), geospatial or otherwise (Tullis and Kar 2020; National Academies 2019; Tullis et al. 2015; W3C Provenance Incubator Group 2010). Indeed, data provenance has been included in the preliminary Research Data Framework (RDaF) from the National Institute of Standards and Technology (NIST) as a category within the Process/Analyze Function, indicating its close ties with programmatic needs (National Institute of Standards and Technology 2020).

By its very nature, geographic analyses incorporate spatial variations and complexities that strain replicability across space (Kedron et al. 2019). That strain, coupled with the rapid pace and evolution of COVID-19-related geospatial applications, only increases the need for accurate provenance information that is accessible to the geospatial consumer and other stakeholders. In this chapter we will explore some of the current challenges and limitations, but also opportunities, associated with geospatial data and analyses for COVID-19, including mobility data, spatial and temporal scales, and sources of model uncertainty.

8.2 Using Social Media to Monitor Human Mobility Dynamics

Reports indicate the considerable mobility reduction following the implementation of social distancing measures is responsible for the reduction of the effective reproduction number R_0 , consequently reducing the transmission rate of SARS-CoV-2 (Kraemer et al. 2020; Qiu et al. 2020). Since the initial outbreak of COVID-19,

numerous efforts have been made to monitor human mobility dynamics. Google and Apple have released their aggregated and anonymized community mobility reports based on data collected from their services (i.e., Google Location Services and Apple Maps). Mobile network operators started to collaborate with authorities to assist in the estimation of the effectiveness of control measures. However, concerns have been raised on whether sharing customer data is appropriate, even in a time of crisis. The rapid global spread of COVID-19 highlights the need for a more harmonized, aggregation flexible, privacy-respecting, and easily accessible approach to monitoring human mobility [19].

Fortunately, the rise of social media platforms offers a new solution to closely monitoring human mobility dynamics, as they provide valuable and timely information regarding whether people are actively reducing their exposure to COVID-19 by reducing distances traveled and by how much (Hawelka et al. 2014; Jurdak et al. 2015). Twitter, a popular social media platform, has millions of active users worldwide. The mobility dynamics derived from Twitter data can indicate how people respond to the pandemic geographically and whether policies are implemented effectively (Xu et al. 2020). In this section, we aim to present how collaborative efforts to reduce human mobility are reflected from massive Twitter data in selected countries.

8.2.1 Twitter Data

Twitter is a massive social networking site that allows users to send and receive short posts called tweets. Tweets contain metadata that describes when the message is posted, and some of them contain geolocation information indicating where a message is posted. Depending on the Twitter user's specific configuration, such spatial information is either in the format of place names (e.g., Fayetteville, AR) or latitude/longitude coordinates (Li et al. 2018). The Twitter data in this study were collected using the Twitter Streaming Application Programming Interface (API) by the Geoinformation and Big Data Research Lab (GIBD) at the University of South Carolina. The dataset contains a total of 587 million geotagged tweets collected from 10 million Twitter users with temporal coverage from January 1, 2020, to May 31, 2020. Geotagged tweets with spatial resolution coarser than the city level were removed to increase the credibility of the mobility pattern. Following Martín et al. (2020), we excluded non-human tweets (e.g., automated weather reports and job offers) by investigating the tweet source. For example, tweets automatically posted for job offers from sources like CareerArc are removed. Since the computation of mobility requires tweets posted from at least two locations, only users who posted tweets on at least two consecutive days are included in the geodesic distance calculation (details in Sect. 2.2.1).

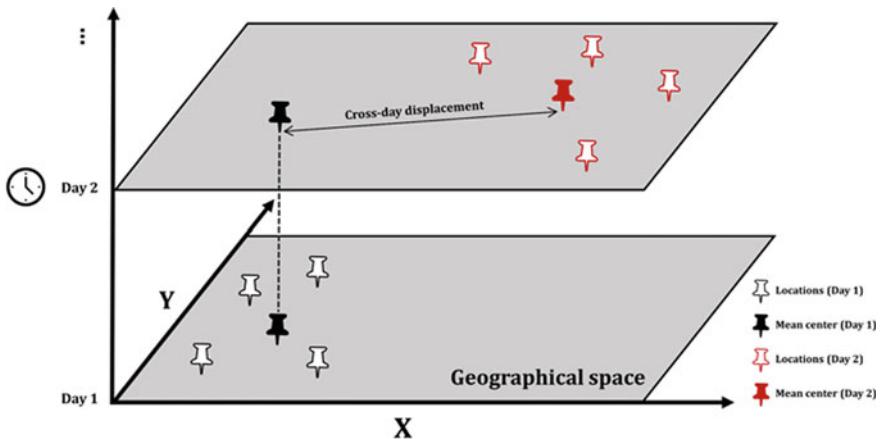


Fig. 8.1 Concept of cross-day displacement in a space-time cube

8.2.2 Mobility Methodology

8.2.2.1 Travel Distance Calculation

The travel distance is the cross-day displacement that quantifies the mean center shift of users between two consecutive days. For a selected Twitter user i , the collection of locations of user i in day j can be represented by $\{L_{i,j}^1, L_{i,j}^2, \dots, L_{i,j}^n\}$. Within the total of n locations in day j , $L_{i,j}^n$ always precedes $L_{i,j}^{n+1}$ in time while $L_{i,j}^1$ denotes the initial location of user i in day j . To compute the cross-day displacement, a mean center ($L_{i,j}^1$) is calculated by respectively averaging the x and y coordinates of locations in $\{L_{i,j}^1, L_{i,j}^2, \dots, L_{i,j}^n\}$, i.e., $P_{i,j}^1 = \mu\{L_{i,j}^1, L_{i,j}^2, \dots, L_{i,j}^n\}$, where μ denotes the mean center operator. The concept of cross-day displacement is illustrated in a space-time cube (Fig. 8.1).

8.2.2.2 Mobility Normalization

Given that a week has been widely recognized as an independent cycle in mobility (Leguay et al. 2006), the baselines for cross-day displacement were set for each corresponding day of a week. As some countries already imposed strict mobility-reducing policies as early as late-February, the baseline period was set to span from January 13 (to exclude abnormal mobility patterns due to the New Year holiday season) to February 15. To ensure that sufficient Twitter posts are available to generate stable time series, we selected the top 32 countries with the most Twitter samples in the dataset. Let R represent the geographical space of a certain country and D_j^R represent the cross-day displacement of R in day j . D_j^R is defined as the mean value

of all $D_{i,j}^R$, i.e., $D_{i,j}^R = \frac{\sum_i D_{i,j}^R}{N}$, where N denotes the total number of selected users in day j within \mathbb{R} . The normalized mobility of country R in day j is the ratio of $D_{i,j}^R$ to its baseline value of a corresponding day in a week (red dashed line in Fig. 8.2).

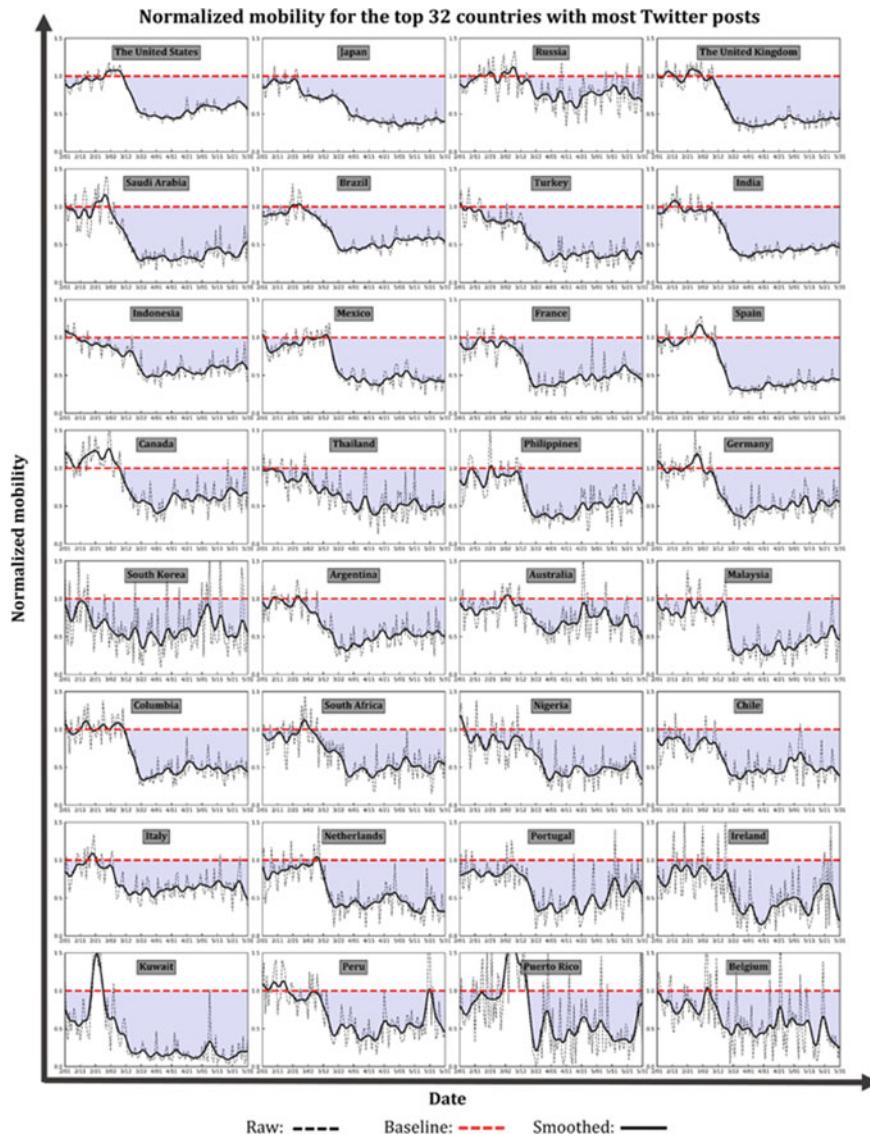


Fig. 8.2 Normalized mobility for the top 32 countries with the most Twitter posts in the dataset

8.2.3 Differential Mobility Impacts by Country

In general, the impact of the COVID-19 pandemic on mobility is obvious. The triggers of mobility changes correspond well with the announcements of mitigation measures, proving that Twitter-based mobility may be useful to reflect the effectiveness of those measures. The mobility for the 32 selected countries experienced a remarkable reduction in March, April, and May (Fig. 8.2), suggesting the effectiveness of mobility-reducing measures implemented in these countries. However, the impacts on mobility varied substantially among countries in different epidemic phases. The mobility in the U.S. started to drop in mid-March, following the declaration of COVID-19 as a pandemic on March 11 and as a U.S. National Emergency on March 13. The mobility remained low in April but started to bounce back in May, presumably due to the loosened measures from the Opening Up America Again policies and the nationwide “Black Lives Matter” protests at the end of May. In Japan, the early drop of mobility in late-February followed the announcement to close all Japanese schools by Prime Minister Shinzo Abe (Kuniya 2020). In addition, Japan extended the declaration of emergency on April 16, which explains the second drop and the low mobility in April (Booker 2020). However, Japan ended the state of emergency on March 25. Due to limited temporal coverage of the collected Twitter posts, the impact of the loosening measures remains unknown. Malaysia experienced a sharp mobility reduction on March 18, coinciding with the date when the Movement Control Order (MCO) from the federal government took effect (Jun 2020). Countries like India, Mexico, Spain, and France also exhibited sudden mobility drops following national mitigation measures. The sudden reductions in those countries demonstrate that mobility-reducing measures were effectively and efficiently implemented. In South Korea, a country that, along with China, suffered from the initial spread of the epidemic in its early stage, low mobility was observed as early as February, due to the early imposition of strict measures. In light of the gradually easing situation (Fisher and Wilder-Smith 2020), the social distancing measures in South Korea started to be lifted, leading to mobility increases in April.

8.2.4 Limitations of Social Media Data for Mobility Research

There are a number of limitations in gauging human mobility dynamics using social media data. First and foremost is their representativeness, which relies on the demographics of the individual users in relation to the demographics of the local population (Huang et al. 2020a). The representativeness shortcoming, however, is also shared with other mobility data sources like cellphone applications and traditional questionnaire surveys. The former has to face the criticisms that people who don’t usually use them or are without digital devices are significantly underrepresented, and the respondents of the latter are often older, more educated, and include fewer minority respondents (Martín et al. 2017; Sloan and Morgan 2015). Future studies

should delve into the demographics of Twitter users for a better understanding of their representativeness.

We must also pay attention to issues regarding data privacy and data ethics within the Research Data Framework (RDaF). While many researchers assume communications through social media are in the public domain and therefore not entitled to the scrutiny of ethics panels, surveys suggest a majority of social media users had opposing views on the fair use of their posts in publications and their rights as research subjects (Williams et al. 2017). Codes of ethical conduct are still needed to guide data collection, analysis, representation, and reproduction. Proper data aggregation and anonymization of Twitter-derived mobility patterns are needed before those patterns are presented in public venues.

Although social media data are less privacy-concerning compared with many passive data collecting approaches, this active nature limits the available trajectory of data spatiotemporally, and potentially causes skewed origin-destination information. Thus, the mobility dynamics from social media data only reflect the mobility patterns that users are willing to share and should not be generalized to represent the complete mobility patterns of the population without further investigation. A study by Huang et al. (2020b) compared Twitter mobility with the Apple mobility report, the Google mobility report, and mobility data from Descartes Labs during the COVID-19 pandemic. Future work is still needed to investigate the similarity and dissimilarity between Twitter mobility records and other mobility reports. To reflect the multifaceted nature of human mobility and to capture a large spectrum of the population, the possibility of deriving an integrated mobility index from multiple sources deserves exploration.

Despite the massive timely responses contributed by millions of Twitter users, we find that Twitter data are still insufficient to derive stable mobility time series even in regions where Twitter is among the most popular social media platforms. This is mainly due to the limited capability of Twitter's free API that returns 1% of the content and the fact that only around 1% of the tweets can be precisely geotagged (Morstatter et al. 2013). Future studies should explore the potential of generating more reliable mobility dynamics using the Twitter Decahose and Firehose streams, with the capability of returning 10 and 100% of Twitter data respectively. The rapid development of geoparsing techniques (Gelernter and Mushegian 2011; Wang et al. 2020a) (the process of extracting place-names from text) and the visual-textual fusion scheme (Huang et al. 2020c) potentially provide new solutions to the utilization of ungeotagged Twitter posts.

Access to the provenance of geotagged Twitter-derived analyses is complex because (a) the information is digitally fragmented, (b) some questions may be difficult to answer such as the motivations behind the use of location services by Twitter users, and (c) geoprocessing decisions may not be clearly communicated. While reproduction of the results is generally strong when versioned source control is published in conjunction with geospatial products, additional work is needed on the replicability of Twitter-derived analyses. This is largely due to the complexity of and access to Twitter's full provenance information.

8.3 Scales of Analysis

During this global pandemic, a significant number of models, dashboards, and geo-visualization tools have been developed and deployed to understand the spread of COVID-19 infections, mortality rates, and access to health care facilities based on mobility, network connectivity, and socio-economic conditions. For instance, dashboards, such as those available from Johns Hopkins University (<https://coronavirus.jhu.edu/map.html>) and the New York Times (<https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html>), are widely used by media and the public to understand the severity of the disease spread and deaths. Disease spread models such as the θ -SEIHRD, exponential growth, and susceptible–infected–removed (SIR) compartmental models (Ivorra et al. 2020; Bertozzi et al. 2020) are used to inform policies about reopening economies and resource allocation planning. Despite their wide usage, these models and tools rarely provide information about the spatial and temporal scales of analysis at which data are available, models are implemented and analyzed, and results are reported. This raises the question of why the scale of analysis should be considered while modeling a pandemic.

Epidemics are both geographical and temporal in nature. Demographic characteristics tend to influence the spread of disease across space and time; the socio-economic data associated with such diseases tend to be non-uniform. Mapping of non-uniformly distributed data has always presented problems, particularly in selecting an appropriate mapping unit. The widely adopted practice in the U.S. during the COVID-19 pandemic of mapping data to counties may provide false impressions to policy makers and the public of the locations at greatest and least risk. Methods that better compare like-populated areas and account for normal mobility may provide more accurate information and better decision-making support. Because of the dynamic nature of the disease, temporal scale is also crucial to understand the implications of models, specifically from a policy implementation and adaptation perspective.

COVID-19 impacts appear to be more pronounced among ethnic minorities and economically disadvantaged populations (Webb Hooper et al. 2020). As an airborne virus spread through close contact, densely populated areas like New York City, Los Angeles, and Atlanta have become epicenters of the disease. The inequality in access to healthcare facilities for certain vulnerable population groups has also exacerbated disease impacts (Wang and Tang 2020; Karaye and Horney 2020). Thus, reporting the number of COVID-19 cases and deaths at the county level rarely provides precise information about the locations where the disease spread is higher. County-level reports are also ineffective at predicting where there is a potential for spread due to high population density or targeting areas with high concentrations of vulnerable populations that may need more testing and/or treatment resources. Addressing these questions requires understanding the impact of spatial and temporal scales of analysis on analytical outcomes.

What do we mean by scale of analysis? Scale can be spatial, temporal, and social. Although there are different connotations for spatial scale (Kar 2008; Goodchild and Quattrochi 1997), in this context, spatial scale refers to the *operational scale of*

analysis or the spatial extent at which the process is studied or reported (Lam and Quattrochi 1992). Temporal scale, on the other hand, corresponds to the duration over which the process has been studied (e.g., two months versus 1 year), while temporal resolution refers to the granularity used for the analysis (e.g., a week versus 24 h) (Quijas and Balvanera 2013). Due primarily to privacy concerns, COVID-19 cases and deaths in the United States are usually aggregated and reported at the county level. The aggregate nature of this data also requires us to understand the concept of the Modifiable Areal Unit Problem (MAUP). The MAUP represents the error associated with statistical analysis due to the use of aggregate data sources (Openshaw 1984). MAUP results from *scale* (or *aggregation*) and *zoning* (or *grouping*) (Openshaw 1984). The *scale*-related uncertainties in MAUP stem from the aggregation of results and data occurring in one areal unit into other larger or smaller units (e.g., blocks into block group or county). The *zoning* related errors result from grouping of areal units at the same scale of analysis into different zones (e.g., political redistricting and gerrymandering) (Openshaw 1984; Atkinson and Tate 2000). The reporting of COVID-19 cases and deaths at the county level results in the scale-related MAUP, as it does not account for the smaller units with higher numbers of cases nor the underlying population density, which is a contributor to disease spread, and therefore provides a biased view of the areas with the most cases.

Due to the availability of large volumes of data from different sources about COVID-19, models and tools have been developed to understand and visualize the spread of the pandemic and its socio-economic impacts. Because these products rarely discuss the underlying scale of the data or the results, it is difficult to compare the results, and particularly challenging to reproduce and replicate them. This variation in results has added to the confusion about COVID-19 and its potential impacts, thereby reducing the efficacy of the information provided by public health professionals to reduce the pandemic impact and spread.

Using Arkansas and Tennessee as two case study states, we examined the *impact of spatiotemporal scales of analysis on reported COVID-19 cases*. Unlike most other states, Arkansas did not have a policy in place for partial or complete shutdown of non-essential businesses until early April, 2020, and started reopening by early May. Tennessee initially had shutdown policies in place, but by June 2020, the policies were relaxed, and non-essential businesses were allowed to function. These two states were selected for this study because of their relaxed approach to shut down policies in the spring and early summer of 2020. Because the study focused on understanding and examining the implications of scales of analysis used to report COVID-19 cases (non-uniformly distributed data), this study could be replicated elsewhere.

We examined the spatial distribution of reported cases at the county level using population and population density as normalizing factors. We also explored the variation in cases based on population density by accounting for areas within a county that are unoccupied. To understand the impact of temporal resolution, we examined the reported number of cases within a week vs over the weekend. A discussion of the methodology implemented is presented below.

8.3.1 Impacts of Measurement Scale

The number of reported daily cases by county were obtained from the Arkansas Department of Health and the Tennessee Department of Health for the week of July 16th through July 22nd, 2020 (Thursday through Wednesday, so that the weekend was centered). The average new daily cases were calculated for each state for the seven-day period using data for the five weekdays (Thursday, Friday, Monday, Tuesday, Wednesday) and the two days of the weekend (Saturday and Sunday). The 7-day mean daily new case data were normalized by population, where each county's mean new case value was divided by the 2017 population for that county, and then multiplied by the state's mean county population value to scale the range of normalized values to be comparable to the mean new case values.

Figure 8.3 depicts the 7-day average of daily new cases reported at the county level in Tennessee (TN). According to Fig. 8.3a, the north-central and eastern counties of

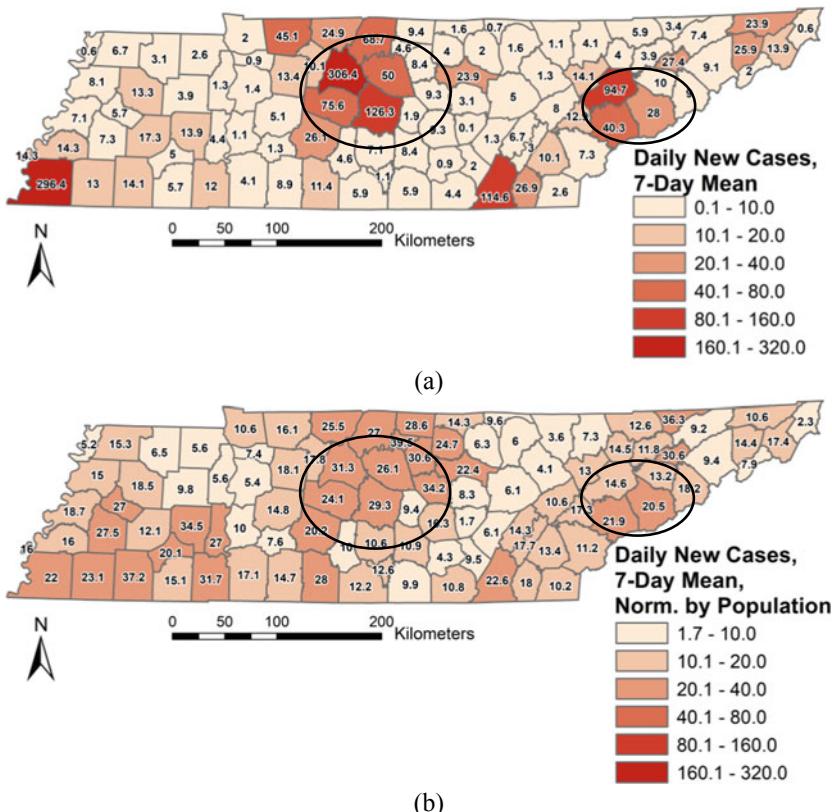


Fig. 8.3 For the week of July 16-22, 2020, **a** 7-day average of daily new cases at the county level in Tennessee, and **b** 7-day average normalized by county population

the state (identified by black ovals) had a higher average of daily new cases with some counties experiencing more than 200 new cases per day. After normalizing by population (Fig. 8.3b), it is evident that the same counties in the north-central and eastern part of TN experienced more moderately elevated daily new cases. In contrast, several rural counties throughout the state had relatively higher daily new cases after normalizing by population.

A similar trend can be seen in Arkansas (AR) where the 7-day average daily new cases was higher in the north-western, central and east-central part of the state (Fig. 8.4a—highlighted by black ovals). Although the same counties appeared to experience elevated average daily new cases based on the 7-day average normalized by 2017 population (Fig. 8.4b), a set of counties in the south-eastern and south-western part of AR also had a notable rise in daily new COVID-19 cases. The counties in the southern part of Arkansas (Fig. 8.4b) had a higher number of cases per capita in comparison to the counties in the central and north-western and north-eastern part of the state (Fig. 8.4a).

8.3.2 Impacts of Temporal Resolution

Separating the average new cases by weekdays and the weekend for the one-week period in July (Fig. 8.5a, b), we observe high averages clustered in the same counties in north-central and eastern Tennessee seen in Fig. 8.3a. While several of these highest case counties saw increased averages on weekends, in 62/95 counties (~65%), the average number of cases were higher during weekdays as opposed to the weekends.

Compared to Tennessee, a much higher proportion of counties in Arkansas saw average daily new cases increase during the weekend compared to weekdays (69/75 counties; 92% compared to Tennessee's 35%; see Fig. 8.6). The counties in the central, north-west, south and north-east part of Arkansas appeared to have experienced a substantial rise in the average number of cases during the weekend (18th and 19⁺ of July) as opposed to the rest of the week (July 16th–22nd).

8.3.3 Impacts of Spatial Scale

Though publicly available COVID-19 data about number of positive cases and deaths is rarely reported at a granularity finer than counties, there has been plenty of evidence that outbreaks follow the concentration of people as well as connectivity resulting from socio-economic activities (Hamidi et al. 2020). Other than heavily urbanized counties, most counties in the U.S tend to have limited concentrations of people in cities and towns, surrounded by areas of low to no habitation. To investigate the reality of population distribution in moderately urbanized counties, Knox County, TN was selected. The 2010 census block boundary and associated population data were obtained, and the population density was calculated as number of people per square

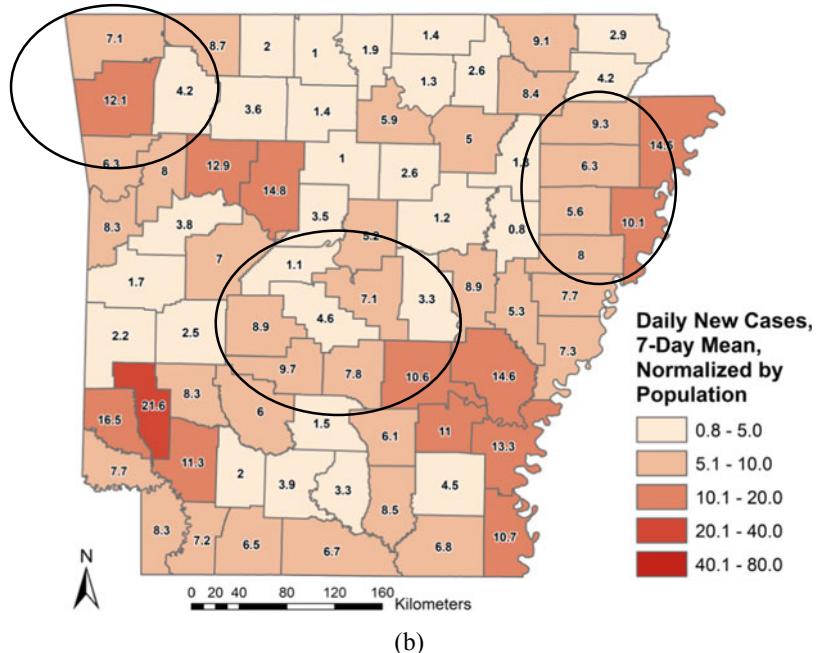
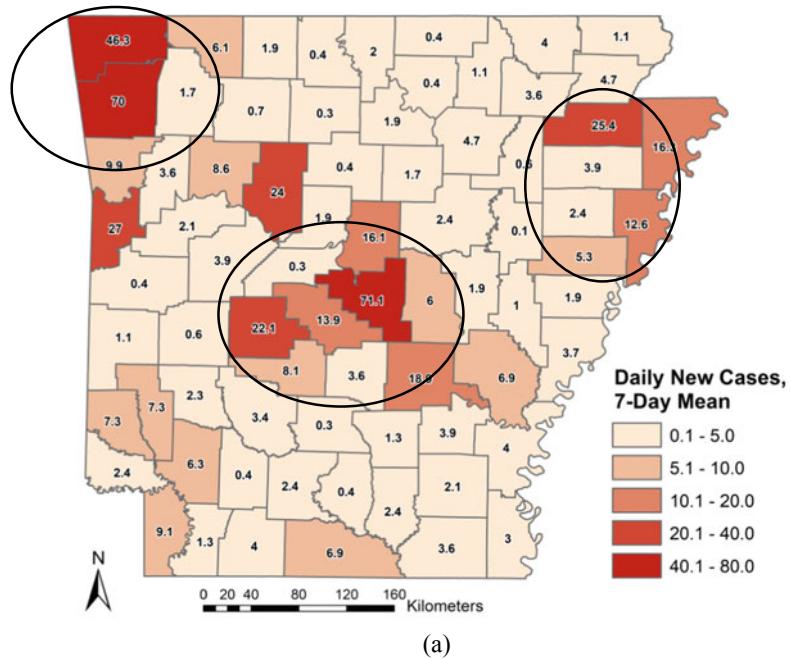


Fig. 8.4 For the week of July 16-22, 2020, **a** 7-day average of daily new cases at the county level in Arkansas, and **b** 7-day average normalized by county population

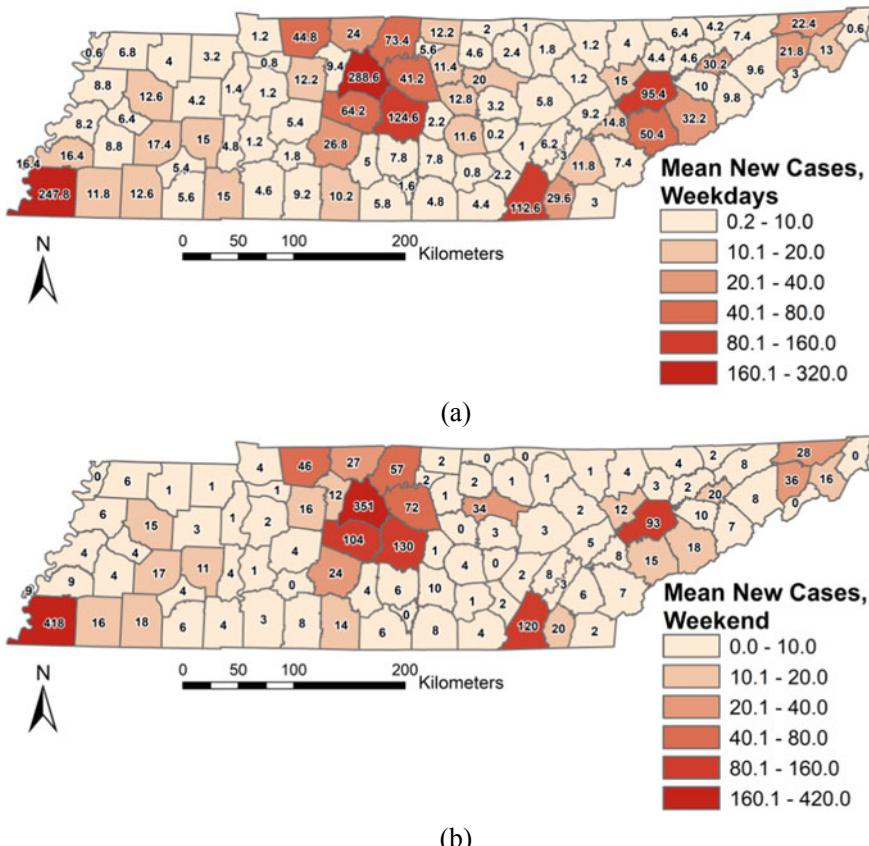
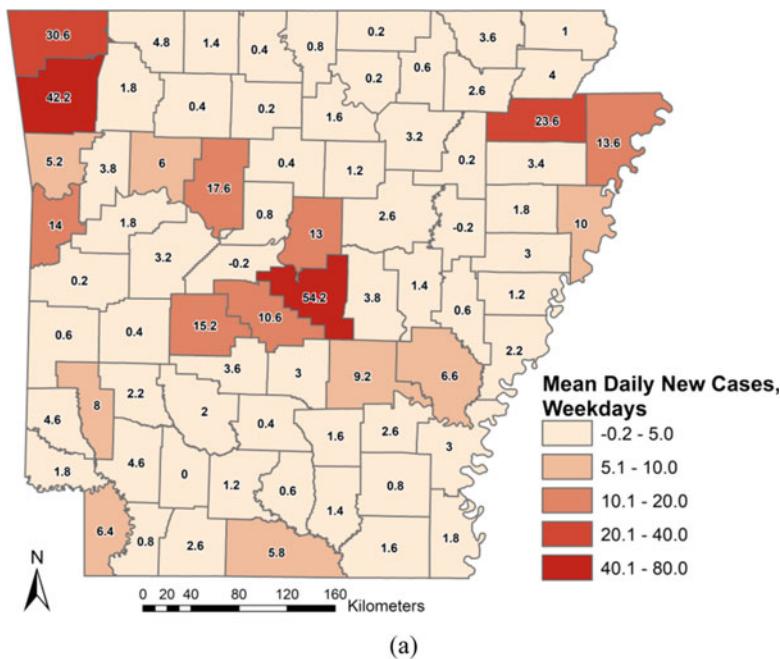


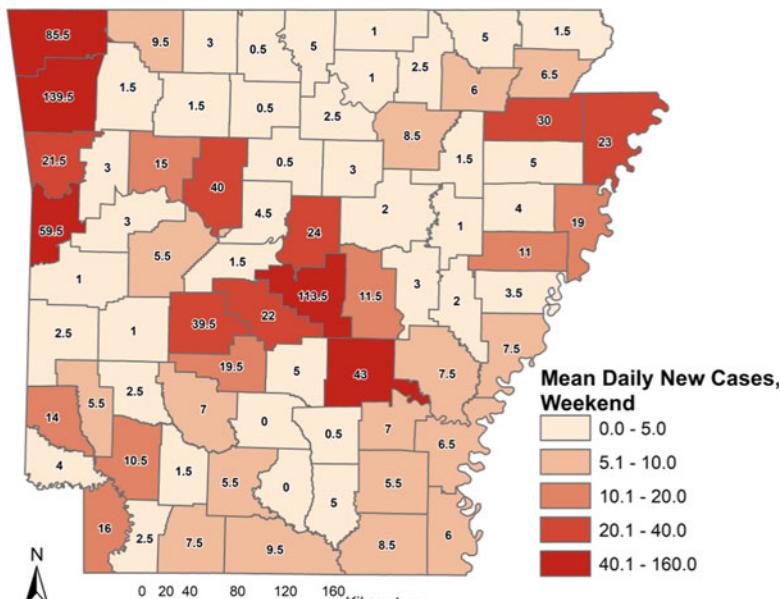
Fig. 8.5 Average number of new COVID-19 cases at the county level in Tennessee during the week of July 16–22, 2020 on **a** weekdays (Monday–Friday), and **b** the weekend (Saturday and Sunday)

kilometer. After temporarily removing blocks with zero population, the remaining blocks were classified into four quartiles based upon population density. The lowest quartile was recombined with the zero-population blocks; these represent areas with such a low population density that they would be very unlikely to sustain an outbreak. The other three quartiles represent increasing likelihood of supporting COVID-19 outbreaks due to increasing population density.

Census blocks and block groups differ in more than just size. Blocks serve as a minimum mapping unit for Census demographics data and are solely defined by the intersection of geographic boundaries: roads, water bodies, railroads, political boundaries, and major power lines. Block groups are used as the minimum unit for publishing Census Bureau sample data, i.e., data that are only collected from a sample of households. In order to protect individual privacy, block groups are required to have a minimum population, with most having between 600 and 3000 residents. Thus, a block group will never have a population of zero, while there are



(a)



(b)

Fig. 8.6 Average number of new COVID-19 cases at the county level in Arkansas during the week of July 16–22, 2020 on **a** weekdays (Monday–Friday), and **b** the weekend (Saturday and Sunday)

many Census blocks with a population of zero, both in rural and urban locations. To evaluate the impact of spatial scale, and to generalize these “likelihood” maps a bit more, the census blocks were converted to census block groups by dissolving (removing) the boundaries between blocks that belong to the same block group. Values of the individual blocks for population and area were summed for the newly formed block groups. After calculating the population per square kilometer for each of the block groups, they were classified with the same break points as the census block map, producing a map that provides a more generalized view of the same approach.

As can be seen from Fig. 8.7a, there are several blocks that have zero population in Knox County, Tennessee. While the block level density distribution identifies the pockets with zero population or sparsely populated blocks, those blocks are not identifiable at the block group level (Fig. 8.7b). In fact, the area identified by the black circle in Fig. 8.7a includes sparsely populated blocks or blocks without any population. The same area in Fig. 8.7b depicts moderate density areas in the county. Thus, using the block group for normalizing reported COVID-19 cases can present misleading information as it will incorrectly suggest unoccupied blocks have cases. From a resource planning perspective, using this information could involve providing assistance to locations where it is not needed, which may also involve removing resources from another location where they are needed.

In late summer 2020, the Arkansas Center for Health Improvement (ACHI) began providing COVID-19 case data, both cumulative and active, by ZIP Code. In situations where there are less than 10 cases per ZIP Code, no numbers are reported. Mapping the case data at this level of granularity greatly changes the visual interpretation of where outbreaks are occurring. To illustrate this, one can compare two maps of active case data for the same day (August 17th, 2020), one mapped by ZIP Codes (Fig. 8.8b) and one mapped by counties (Fig. 8.8a), using a similar classification scheme. These maps demonstrate how the generalization of data to county mapping units tends to obscure more local patterns evident at the greater granularity of ZIP Code boundaries.

8.4 What Can Be Done About Challenges with Scale?

Despite agreement in the spatial science community about the implications of scales of analysis, this is generally not addressed and accounted for in many disciplines including public health. As can be seen from the Figs. 8.3 and 8.4, not surprisingly, using a 7-day average versus 7-day average normalized by population provides different information and highlights a different set of counties that are experiencing high average daily new cases. The number of reported cases during weekdays vs weekends (Figs. 8.5 and 8.6) is also different as counties with high numbers of cases during the weekend are not always the same ones that experience high numbers during weekdays. While county level data about cases, death rates, and mobility

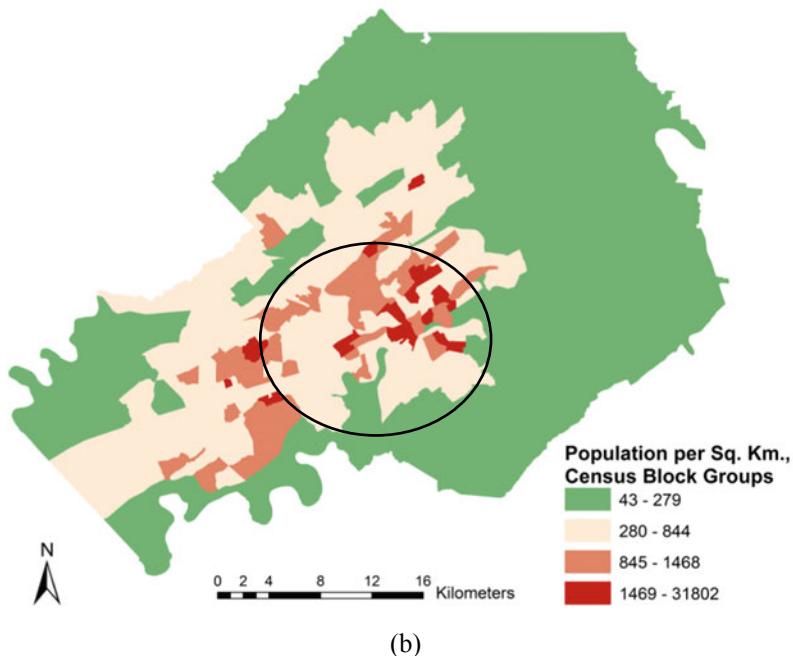
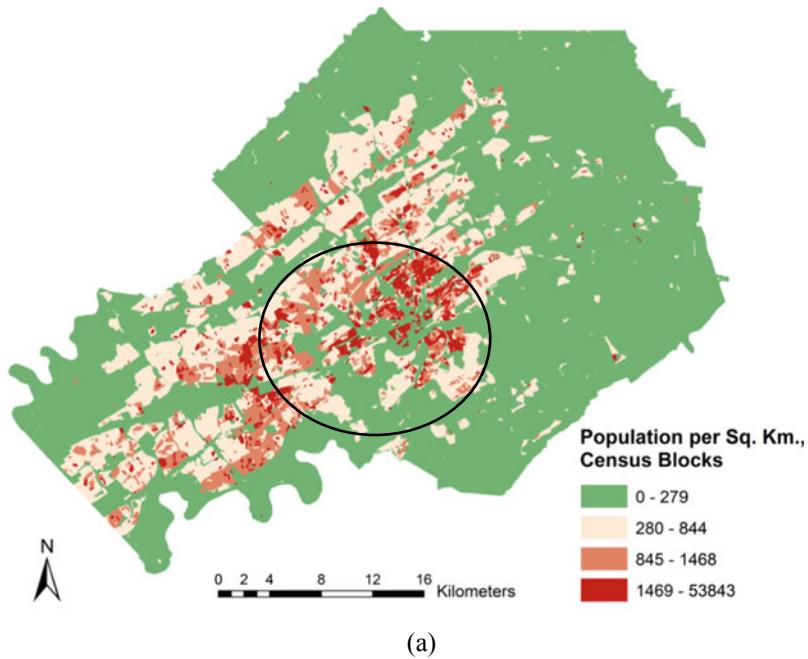


Fig. 8.7 Comparison of population density in Knox County, Tennessee at **a** census block scale, and **b** census block group scale

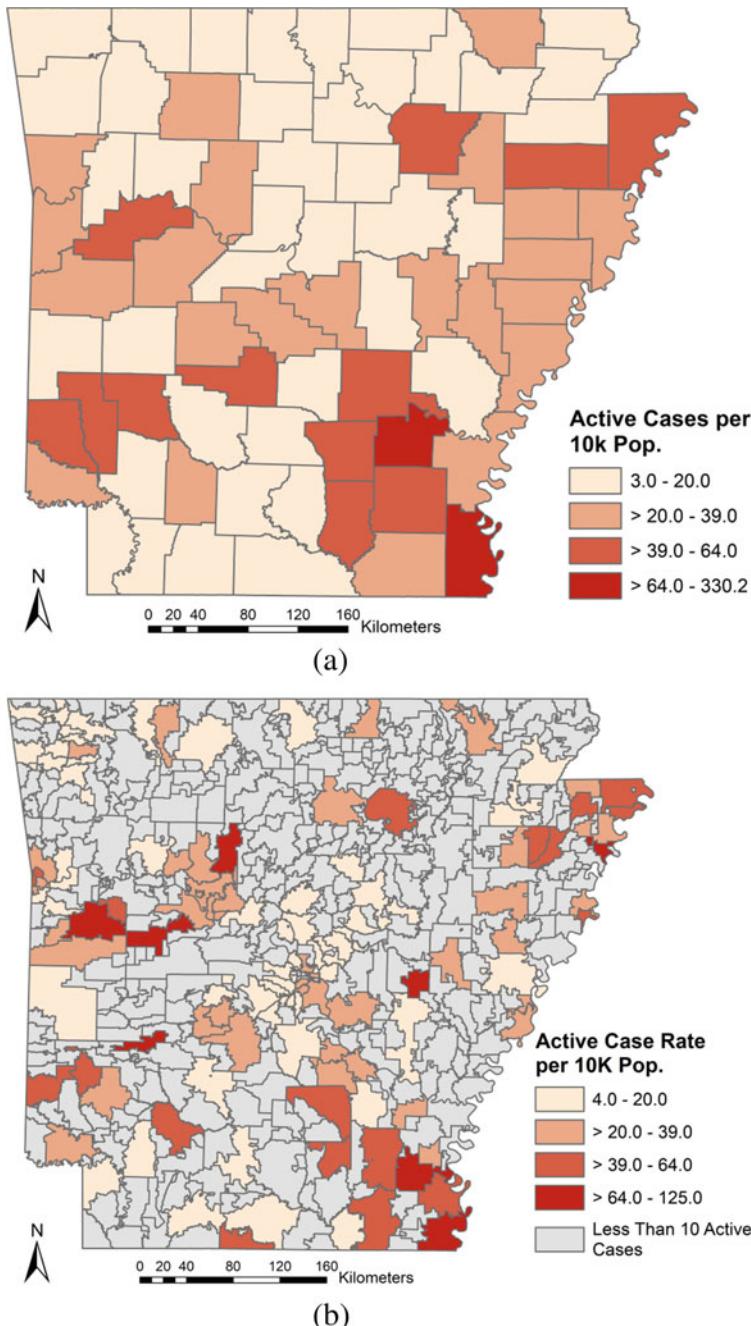


Fig. 8.8 Comparison of active COVID-19 cases per 10,000 people in Arkansas on August 17, 2020 at **a** county scale, and **b** ZIP Code scale with numbers below 10 suppressed

eliminates most privacy concerns, Figs. 8.7 and 8.8 demonstrate the tendency for increasingly aggregated data to obscure the true spatial patterns of high case rates.

A major takeaway of these preliminary analyses is that representing and quantifying non-uniform data such as the number of COVID-19 cases and subsequent deaths at coarser spatial scales of analysis (e.g., county) tends to undermine not only the spatial distribution of the cases, but also the number of cases. Likewise, using different temporal resolutions (e.g., 24 h vs. 7-day average) also does not reflect the true number of infected cases at a specific location. Because COVID-19 spread is influenced by population density and the spatial distribution of certain vulnerable groups (Roy and Kar 2020), normalizing the number of cases with population density at different spatial scales of analysis (county vs census tract vs. census block) provides a different perspective with regard to the intensity of the spread. The findings resonate with previous studies that have demonstrated that using coarser spatial scales and temporal resolutions tend to generalize results and obscure spatial distributions, concentrations, and patterns of non-uniform data like COVID-19 cases, which is crucial for response and resource planning need of communities.

The disparity in information can be confusing and misleading to stakeholders unless explanations about the spatial and temporal scales of analysis are provided. Given the urgency of this pandemic, stakeholders and policy makers are interested in getting reliable information to help with resource planning. Because the majority of the tools and dashboards are visualizing the infected cases and death rates, and rarely discuss the underlying data that were used to create those outputs, it is not surprising to see a variety of policies and responses have been deployed to contain this pandemic.

Provenance information associated with scale of analysis (and associated R&R within a given geography) is highly accessible to geospatial experts who are familiar with the associated workflow(s) including code and data. Unfortunately, many COVID-19 visualizations, dashboards, statistics, etc. are “published” without clear linkages to team or personal workflow environments. Many questions can be quickly answered by expert reviewers when such linkages are made available. Making outputs based on different scales of analyses available to the decision-makers will also minimize confusion, allowing stakeholders to utilize multiple outputs to identify hotspots.

8.5 Understanding Uncertainty in Projection Models

Beyond describing and mapping mobility patterns and the ongoing spread of COVID-19, there has also been remarkably high interest in predictive modeling, both geospatial and temporal. A key issue in predictive epidemiological models for infectious diseases is the characterization and, when possible, quantification of uncertainty. We here elaborate on several key sources of uncertainty in the projection models and forecasts adopted by different agencies in analyzing the COVID-19 data. We will first delineate the two most common approaches for modeling the rate and spread of

an infectious disease. Then we will shift our attention to other sources of uncertainties, viz. assumptions on model parameters, model estimation, and both systematic and random uncertainties from input data, including temporal lags in reporting. Since these forecasts have an immense impact on public health policy and individual risk-based decision making, it is paramount to understand and convey the sources and nature of uncertainties.

8.5.1 Model Choice

For the sake of brevity, we will focus on two of the most popular models for COVID-19 spread: (1) the epidemiological compartmental models such as eSIR or SEIR or their extensions (Song et al. 2020; Osthust et al. 2017; Wang et al. 2020b), and (2) the IHME model (Murray et al. 2020) that estimates the logarithm of the cumulative death rates using a curve-fitting approach and information on the social distancing practiced. This is critical to understand for both practitioners and scientists, as there is little in common between these models and thus, model choice adds to the amount of overall uncertainties in the predicted numbers.

8.5.1.1 Epidemiological Compartmental Models: SIR

The most commonly used tool for predicting the theoretical number of individuals in different stages of an infectious disease is the Susceptible-Infected-Removed (SIR) model (Kermack and McKendrick 1927). The standard SIR model captures the infection dynamics using a set of differential Eq. (8.1) connecting each compartment. The basic SIR model takes as inputs two time series of proportions of infected and removed cases, denoted by Y_t^I and Y_t^R at time t , respectively, where the removed part of the population is a sum of the recovered cases and deaths at time t . The dynamic model is represented below:

$$\frac{d\theta_t^S}{dt} = -\beta\theta_t^S\theta_t^I, \quad \frac{d\theta_t^I}{dt} = \beta\theta_t^S\theta_t^I - \gamma\theta_t^I, \quad \text{and} \quad \frac{d\theta_t^R}{dt} = \gamma\theta_t^I \quad (8.1)$$

Here, θ_t^S is the true proportion of susceptible individuals at time t and θ_t^I and θ_t^R are the respective prevalence of infection and removal at time t . The basic SIR model assumes that the entire population falls into one of these three compartments, i.e., at any time t , $\theta_t^S + \theta_t^I + \theta_t^R = 1$. The parameters $\beta > 0$ and $\gamma > 0$ denote the disease transmission and removal rates respectively. It can be shown that the basic reproduction number R_0 is given by $R_0 = \beta/\gamma$ for the SIR model.

However, the basic SIR model is restrictive in terms of incorporating uncertainties and measurement errors and it assumes a constant rate for both transmission and removal, which does not reflect reality where these rates are affected by government-imposed isolation measures. Several extensions of this model have been

used by epidemiologists for modeling different infectious diseases, such as SEIR (Susceptible-Exposed-Infected-Removed) (Lekone and Finkenstädt 2006), network-SIR (Boccara and Cheong 1992), and extended SIR or eSIR (Song et al. 2020; Wang et al. 2020b). For this section, we will illustrate one of the most recent models, eSIR, and show the effect of parameter choices on the outcomes of eSIR. The eSIR model makes the following improvements:

1. The observed proportions of susceptible and infected follows a two-dimensional time series with the underlying prevalence parameters θ_t^I and θ_t^R following a Markov model, each epoch depending on the last one.
2. It allows a time-varying probability $\pi(t)$ that a susceptible person meets an infected person or vice versa: $\pi(t)$ is called a transmission modifier as it modifies the probability of a susceptible person meeting an infected person.

For the Bayesian modeling framework in eSIR, the prevalence of different compartments at time t is modeled via a Beta-Dirichlet state-space model (Osthus et al. 2017), where the states at time $t + 1$ depend on the states at time t via a Markovian structure. The proportions at time t follow a Dirichlet distribution centered at the solution to the system of dynamic ODE (1), or its approximation, e.g., the fourth order Runge-Kutta approximation as used by the eSIR package. The vector $\eta = (\beta, \gamma, \theta_0, \lambda, \kappa)$ is the set of all parameters, with $\lambda I, \lambda^R, \kappa$ being used as scale parameters. Note that the parameters of the Beta distributions are chosen to ensure that the expected proportions are equal to the prevalence of infection. The authors of the eSIR package used log-normal priors on both R_0 and γ , with parameters calibrated to obtain a prior mean of R_0 around 2 and a prior mean of γ , mean infectious period, of 12 days. The remaining scale parameters are given relatively ‘flat’ or weakly informative Gamma priors with high variance. As noted before, the eSIR extends the basic SIR model by incorporating $\pi(t)$, a transmission modifier input by the user, that can be used to tune the extent of social distancing. For a real data analysis, $\pi(t)$, the transmission rate modifier needs to be specified according to isolation measures taken in a given region. Obviously, different values of the prior parameters as well as different values of $\pi(t)$ leads to different dynamic behaviors; we will discuss these in detail in Sect. 8.5.2.

8.5.1.2 Curve-Fitting Models: IHME Model

The popular Institute for Health Metrics and Evaluation (IHME) (Murray et al. 2020) model uses worldwide COVID-19 mortality data and data on government intervention and extent of social distancing measures, aggregated from various local and global sources, to forecast expected deaths using a curve-fitting approach. As Jewell et al. (2020) pointed out, there were updates and major revisions to the IHME model (Jewell et al. 2020), and hence it is prudent to describe the latest curve-fitting model that was publicly available at the time of writing this report. It is also worth mentioning that the IHME website (<http://www.healthdata.org/covid/publications>) in July 2020 listed several different modeling approaches used by the modeling

team after the initial curve-fitting approach, viz. IHME-CF (curve fitting), IHME-CF-SEIR (hybrid curve fitting and compartmental) and IHME-MS-SEIR (mortality spline and compartmental) model. Please see Friedman et al. (2020) for a comparison of predictive accuracy between these models (Friedman et al. 2020). We describe the IHME-CF model below.

According to the IHME model, the cumulative death rate at each location is modeled as a nonlinear mixed model with a Gaussian error distribution after suitably transforming by location and scale.

$$D(t, \alpha, \beta, p) = \frac{p}{2} \Psi(\alpha(t - \beta)), \quad (8.2)$$

where $\Psi(z) = 2/\sqrt{\pi} \int_0^z \exp(-t^2) dt$ is the Gaussian error function and the parameters p , α and β control the location specific maximum cumulative death rate, growth parameter and inflection point and t denotes the epoch for death rate crossing 0.31 per million. Information on social distancing implementation were used to estimate the location shift parameter β from locations that were considered to have peaked in a training window of observation.

Despite its popularity and widespread use by many policymakers in the US, the IHME model has several shortcomings, namely, underestimation of uncertainty and poor coverage probability. For example, a thorough analysis of the different stages of the IHME model by Marchant et al. (2020) showed that for the initial model (April) almost 76% of states had actual death counts outside the 95% confidence intervals and the updated model (May) improved this at the cost of poorer coverage probability (Marchant et al. 2020). This has raised a lot of concerns in the public health and epidemiology community about the IHME curve-fitting model (see Jewell et al. 2020; Begley 2020). We would like to emphasize that the IHME model (5) makes several assumptions, such as assumption of Gaussian shape of epidemic curves or continued conformity to social distancing guidelines/mandates and their effectiveness, among others. It has also been noted that the IHME modeling approach has been through several ‘major modeling updates,’ which has been cited as a sign of lack of reliability (Begley 2020). As Jewell et al. (2020) points out, “the IHME projections are based not on transmission dynamics but on a statistical model with no epidemiologic basis,” and several assumptions used by the model are unlikely to be satisfied in practice (Jewell et al. 2020). We point the readers to Jewell et al. (2020) for a detailed discussion of these issues and to the website <http://www.covid-projections.com/> which shows how the IHME estimates have been revised over time.

8.5.2 *Uncertainty from Statistical Inference*

Here we focus on uncertainty from parameter choices, which can severely impact forecasts from epidemiological models. As the operational mechanisms of the IHME

curve-fitting model are not publicly available to test different assumptions independently, we focus on the standard epidemiological model eSIR (Song et al. 2020) introduced before. We consider two key hyper-parameters that are usually supplied by the user depending on external factors:

1. The reproduction number R_0 , defined as $R_0 = \beta/\gamma$ where β and γ are parameters from Eq. (8.1), that reflects the speed of transmission of an infectious disease.
2. The transmission modifier $\pi(t)$, that reflects the strength of isolation or quarantine measures.

8.5.2.1 Effect of R_0

The R_0 parameter, called the basic reproduction number, is often interpreted as the expected number of cases generated by one infected individual in a completely susceptible population. It is considered the most important quantity to be estimated for infectious disease modeling and provides information for understanding the extent of spread of an emerging disease and designing interventions (Diekmann et al. 2010). In general, R_0 can be estimated by the dominant eigenvalue of the *next generation matrix*, that connects the newly infected individual states in consecutive generations or time points (Diekmann et al. 2010). For the SIR model given in Eq. (8.1), the parameters β and γ are average transmission rate and removed rate and R_0 is given by the ratio of β/γ .

In a Bayesian approach for a SIR state-space model, the practitioner supplies a prior guess for the parameters related to spread of infection, namely R_0 , which depends on the parameters β and γ , and the parameters are updated conditional on the observed data. In practice, one can either specify these prior choices directly into the model if expert knowledge is available or use these as tuning parameters to choose a value that yields best prediction performance. For example, Song et al. (2020) and the subsequent application Ray et al. (2020) specify the default values of these parameters based on the first month SARS-CoV-2 outbreak (Ray et al. 2020). To find a suitable prior value of R_0 , Ray et al. (2020) compared two policy arrangements in India (no versus moderate intervention) and choose β and γ in a way that the values of R_0 of 2.0 and 1.5 were obtained corresponding to the two scenarios (Ray et al. 2020).

8.5.2.2 Effect of Policy Arrangement, or Changing $\pi(T)$

As discussed earlier, the addition of $\pi(t)$, the transmission modifier, is a key element in incorporating policy effects into our model in real time. To elaborate on what goes into $\pi(t)$, suppose at a time t , $q^S(t)$ is the chance of an at-risk person being in home isolation, and $q^R(t)$ is the chance of an infected person being in hospital quarantine, then, we will have:

$$\pi(t) = (1 - q^S(t))(1 - q^R(t)) \in (0,1).$$

Note that the above extended SIR model assumes primarily that “both population-level chance of being susceptible and population-level chance of being infected remain the same, but the chance of a susceptible person meeting with an infected person is reduced by $\pi(t)$ ” (Song et al. 2020).

We show in Fig. 8.9 the effect of $\pi(t)$ in an early stage of COVID-19 evolution, based on the reported data in the state of Arkansas up to April 8, 2020. We compare the effect of policy arrangement via the peak of the prevalence of *infected* or ‘I’ compartment in our eSIR model over the projection time-period, as well as the shape

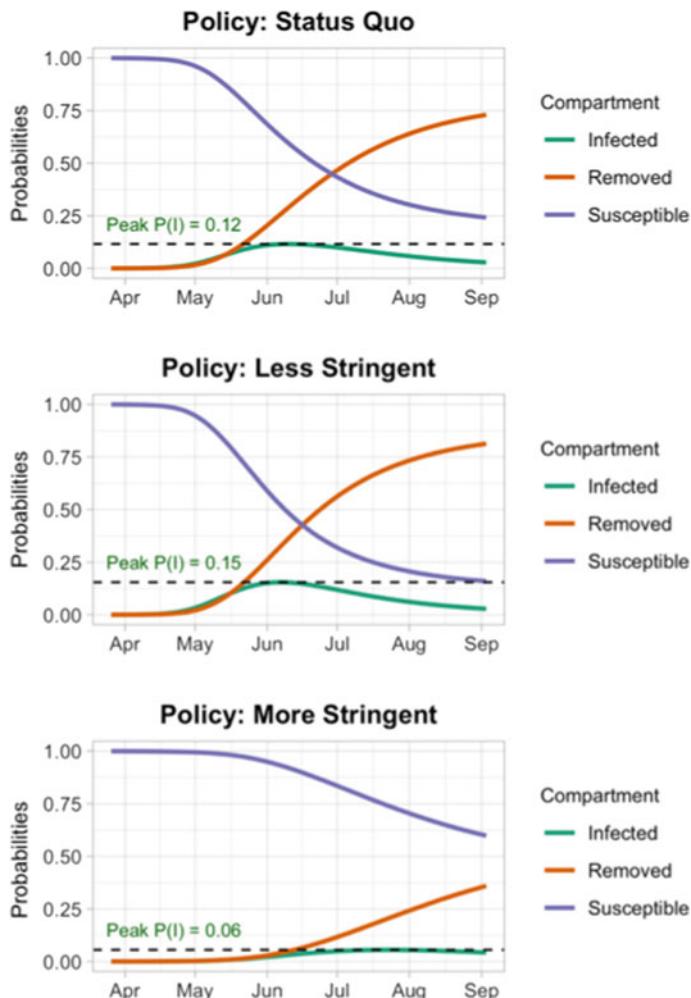


Fig. 8.9 Infection forecasts under three policy arrangements

of the individual trajectories of all three compartments, viz. Susceptible, Infected and Removed. Intuitively, a stronger and earlier non-pharmaceutical intervention should result in a smaller peak as well as a delayed attainment of the peak infection prevalence. For this analysis, we assume “status quo” represents a moderate level of stringency in state policies dealing with COVID-19 (e.g., emergency declaration, school closures, bar/restaurant limits, ban on large gatherings). Under these circumstances, $\pi(t)$, the probability of contact between a susceptible and an infected individual is relatively moderate (0.7) and the peak prevalence of infection ($P(I)$) will be at a moderate level (0.12). If we make these policies less stringent and lift current mitigation measures, the chance of related individual contact will be relatively high (0.9) and the peak prevalence of infection ($P(I)$) will increase to a relatively high level (0.15). For the state of Arkansas with a population of over 3 million, this policy change could translate to an additional 90,000 infections at the peak of the epidemic. By contrast, if we add more stringent policy measures to the status quo (e.g., stay-at-home order, non-essential business closures, state border control) at this point, the chance of related individual contact will be low (0.5) and the peak prevalence of infection ($P(I)$) will drop to a relatively low level (0.06), half of the status quo peak. Critically, there is little theoretical guidance on how these $\pi(t)$ values should be selected.

8.5.3 *Provenance-Related Uncertainty*

The uncertainties from state-wise or county-wise data are inherently tied to the provenance, accuracy, and variability of the estimated quantities. Infection or death data tend to be either lagged or under-reported both on weekends and when facilities are overrun. In addition, some hospitals and clinics analyze their own tests, while others send them to outside labs, further delaying results. For example, consider the daily cases in Arkansas during July 2020, comparing the date each case was reported to the Arkansas Department of Health with the date the actual test was performed (Fig. 8.10a). Figure 8.10b shows the percentage difference between these two counts of cases each day, ranging from 2.5 to 96.5% difference. Of all the tests performed on a given day, only an unknown proportion will be analyzed and reported on that same day. Over subsequent days and sometimes weeks, test results from that single day will continue to be reported to central authorities. Furthermore, reporting lag is not a consistent period of time (e.g., always a 3-day lag), and may also vary geographically with test results from more isolated rural areas generally taking longer to be analyzed and reported. As this lagged reporting occurs, datasets with access to the actual test dates can become more accurate over time, while datasets relying on reporting dates, such as the widely used New York Times dataset, grow increasingly outdated as more test results come in.

We believe many if not most users engaged in descriptive analyses or predictive modeling of COVID-19 are not aware of this discrepancy between test date and reporting date, or the uncertainty it introduces. One possible adjustment which may

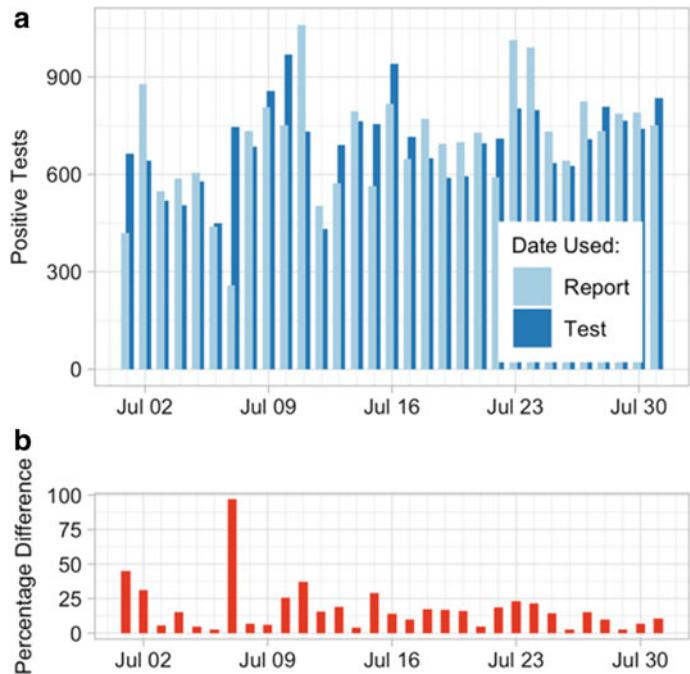


Fig. 8.10 a Daily positive test results in Arkansas during July 2020, separated by date of reporting and date of testing; and b the percentage difference between these two daily counts

help alleviate this challenge is the use of smoothed data, such as 3-, 5-, or 7-day moving averages (see Fig. 8.11). Not only are smoothed estimates closer to consistent

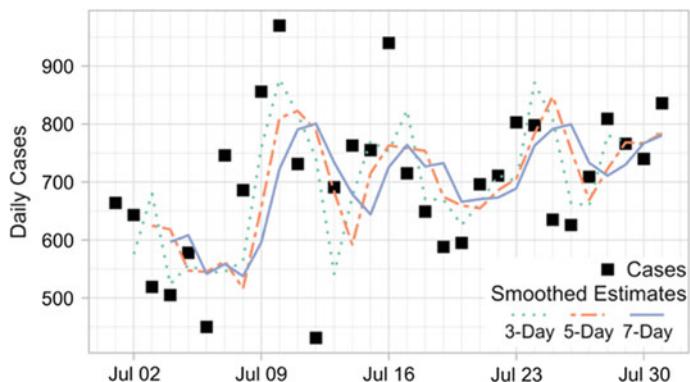


Fig. 8.11 Different smoothed estimates of daily COVID-19 cases, using 3-, 5-, or 7-day rolling averages

between test date and reporting date, but they have the added advantage of reducing the impact of “weekend effects” in both testing and reporting.

8.5.4 *Other Challenges with COVID-19 Projection Modeling*

As we have discussed in this section, uncertainty in COVID-19 projection can come from many different sources, including but not limited to model choice, statistical inference procedure, and source of data. However, there are also other factors that might influence the projections significantly. First, medical resources, interventions, and even definitions and guidelines issued by health agencies evolve over time, affecting the rates that are beyond the scope of simple mathematical models described above. Second, the mathematical models ignore the heterogeneous nature of the population, e.g., gender, age, and other demographic indicators, but in reality, the prevalence in these groups is far from homogeneous. Some communities with disadvantaged population are at a higher risk of COVID-19 infection. Given these inherent restrictions, it would be prudent to leverage information from other sources, such as seroprevalence studies or mobility data (see Sect. 2), to confirm or validate projections and estimated rates.

8.6 Conclusion

Not since John Snow’s 1854 cholera map of London has a public health crisis brought so much attention to geospatial data and analyses, and their importance to understanding, predicting, and preventing the spread of disease. In this chapter we have attempted to highlight some of the most relevant and pressing examples of challenges and limitations associated with these data and analyses, including the troubling issues regarding reproducibility and replicability. These and other challenges and limitations not discussed demonstrate the need for better adherence to research data standards and procedures, like those outlined in the RDaF currently under development (National Institute of Standards and Technology 2020). Despite the laundry list of concerns, we remain optimistic that the benefits and unique insights afforded by geospatial approaches to COVID-19 (and other public health crises) far outweigh the limitations.

For example, since the initial outbreak of COVID-19, the emerging concepts of “Web 2.0” (O’Reilly 2005), “Big Data” (McAfee and Brynjolfsson 2012) and “Citizen as Sensors” (Goodchild 2007), have been widely embraced to obtain timely information regarding human mobility, which is closely related to the transmission of SARS-CoV-2. Social media data has been particularly useful. First, the popularity of social media allows mobility dynamics from multiple regions to be cross-compared. As a more harmonized source compared with cellphone records from certain providers, the massive social media data provide a new venue where citizens

can form an organic sensor network, benefiting epidemic management in all phases. Second, there are fewer privacy concerns with social media data, as users must grant permission to share their data and determine the locational accuracy of the posts from their personal settings. Third, social media are cheap and easily accessible, although computational resources and data storage are needed to handle the large amount of data. Twitter, for example, provides a free API that returns 1% of all the content. Finally, social media data offer timely and spatially explicit geo-information, which transcends traditional surveying approaches that usually lack immediacy.

With more traditional data sources, knowing the exact location of individuals who have been infected by COVID-19 is crucial to allocate resources and evaluate access to health facilities. From a policy perspective, this accounting at individual level is also beneficial to address the economic burden caused by the pandemic. Because of privacy concerns, the alternative is to provide information at a coarse scale of analysis like county. This generalization of reporting prevents effective resource planning, such as supply of ventilators and masks to certain hospitals that are in close proximity to hotspots of infected cases. This generalization can also be misleading when normalized by population (as seen from Fig. 8.3) as it reduces the apparent concentration of cases in certain counties. Observational scales of analysis influence analytical outcomes; specifically, descriptive powers of relationships increase at coarser scales of analysis, thereby reducing the variability that is generally evident at finer scales of analysis as well as increasing the reliability of the results (Mandelbrot 1967; Kar and Hodgson 2012; Estes et al. 2018). Policy makers and stakeholders should be aware of the implication of using coarse vs fine scales of analysis. COVID-19 stakeholders working with geospatial data and tools should ensure that sufficient provenance information is available for expert review before making scale-dependent policy decisions. It is the responsibility of geospatial researchers to work with public health professionals to ensure the intricacies of spatial and temporal scales of analysis is conveyed to policy makers so that effective decisions can be taken without creating erroneous and misleading outputs demonstrated here.

Sources of uncertainty associated with predictive models of COVID-19 are varied and pervasive, from basic model choice and parameterization, to provenance and data quality issues such as reporting lag. At the same time, access to and affordability of advanced computational resources, such as high-performance computing, continues to improve, allowing ever more sophisticated models to be developed. The recent improvements to traditional SIR modeling, such as the inclusion of a transmission modifier implemented in the eSIR package, is just one example. As geospatial epidemiological models improve and become more widely accessible, the need for R&R will only increase. Future epidemiological models will undoubtedly find ways to incorporate increasingly realistic approximations of spatiotemporal dynamics, as well as address many of the uncertainties outlined above. Accurate capture and appropriate sharing of provenance information will be critical to the success of such endeavors.

With the increased scrutiny of public health research, especially in the context of COVID-19, it is essential for stakeholders (practitioners and researchers alike) to protect and manage their datasets to ensure (i) reproduction and replication of these

research is possible in the future and (ii) the generation of trustworthy outcomes for policy purposes. The NIST's Research Data Framework (RDaF) provides guidelines and best practices for managing and protecting different data sets to not only reduce the risks of data mismanagement, but also benefit the stakeholder community by promoting and maintaining trustworthy datasets with appropriate provenance information. This chapter has highlighted (i) the value of mobility datasets in exploring the spatio-temporal spread of COVID-19 disease, (ii) the importance of scales of analyses in examining and modeling disease spread, (iii) inherent issues of uncertainties associated with geospatial disease modeling, and (iv) the need for reproducibility and replicability of research to address the aforementioned points. Acknowledging and addressing these challenges and limitations of geospatial datasets and analyses is essential in order to allow the invaluable potential of geospatial tools and technologies to become fully engaged in COVID-19 research, and public health at large.

Acknowledgements This manuscript has been authored in part by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE). The US government retains and the publisher, by accepting the article for publication, acknowledges that the US government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this manuscript, or allow others to do so, for US government purposes. The manuscript contents are solely the opinions of the authors and do not constitute a statement of policy, decision, or position on behalf of Oak Ridge National Laboratory, UT-Battelle, the Department of Energy, or the US Government.

References

- Atkinson, P. M., & Tate, N. J. (2000). Spatial scale problems and geostatistical solutions: A review. *The Professional Geographer*, 52, 607–623. <https://doi.org/10.1111/0033-0124.00250>.
- Barrett, F. A. (2000). *Disease & geography: The history of an idea*. Atkinson College, Dept. of Geography, Toronto.
- Begley, S. (2020). Influential COVID-19 models shouldn't guide U.S. policies, critics say. In: Statnews.com. <https://www.statnews.com/2020/04/17/influential-covid-19-model-uses-flawed-methods-shouldnt-guide-policies-critics-say/>. Accessed 20 Aug 2020.
- Bertozzi, A. L., Franco, E., Mohler, G., Short, M. B., & Sledge, D. (2020). The challenges of modeling and forecasting the spread of COVID-19. ArXiv200404741 Q-Bio.
- Boccara, N., & Cheong, K. (1992). Automata network SIR models for the spread of infectious diseases in populations of moving individuals. *Journal of Physics A: Mathematical and General*, 25, 2447–2461. <https://doi.org/10.1088/0305-4470/25/9/018>.
- Booker, B. (2020). Japan to declare nationwide state of emergency as virus spreads. In Japan Times. <https://www.japantimes.co.jp/news/2020/04/16/national/japan-nationwide-state-of-emergency-coronavirus/>. Accessed 20 Aug 2020.
- Diekmann, O., Heesterbeek, J. A. P., & Roberts, M. G. (2010). The construction of next-generation matrices for compartmental epidemic models. *Journal of the Royal Society, Interface*, 7, 873–885. <https://doi.org/10.1098/rsif.2009.0386>.
- Estes, L., Elsen, P. R., Treuer, T., Ahmed, L., Caylor, K., Chang, J., et al. (2018). The spatial and temporal domains of modern ecology. *Nature Ecology and Evolution*, 2, 819–826. <https://doi.org/10.1038/s41559-018-0524-4>.

- Fisher, D., & Wilder-Smith, A. (2020). The global community needs to swiftly ramp up the response to contain COVID-19. *The Lancet*, 395, 1109–1110. [https://doi.org/10.1016/S0140-6736\(20\)30679-6](https://doi.org/10.1016/S0140-6736(20)30679-6).
- Friedman, J., Liu, P., Gakidou, E., & IHME COVID-19 Model Comparison Team. (2020). Predictive performance of international COVID-19 mortality forecasting models. *Epidemiology*.
- Gelernter, J., & Mushegian, N. (2011). Geo-parsing messages from microtext. *Transactions in GIS*, 15, 753–773. <https://doi.org/10.1111/j.1467-9671.2011.01294.x>.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *GeoJournal*, 69, 211–221. <https://doi.org/10.1007/s10708-007-9111-y>.
- Goodchild, M. F., & Quattrochi, D. A. (1997). Scale, multiscaling, remote sensing, and GIS. In M. F. Goodchild & D. A. Quattrochi (Eds.), *Scale in remote sensing and GIS* (pp. 1–12). Florida, USA: CRC Press.
- Hamidi, S., Sabouri, S., & Ewing, R. (2020). Does density aggravate the COVID-19 pandemic?: Early findings and lessons for planners. *Journal of the American Planning Association*, 86, 495–509. <https://doi.org/10.1080/01944363.2020.1777891>.
- Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., & Ratti, C. (2014). Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Systems*, 41, 260–271. <https://doi.org/10.1080/15230406.2014.890072>.
- Hippocrates. (1849). *The genuine works of Hippocrates*. New York: Printed for the Sydenham society.
- Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020a). Twitter reveals human mobility dynamics during the COVID-19 pandemic. *PLoS ONE*, 15,. <https://doi.org/10.1371/journal.pone.0241957>.
- Huang, X., Li, Z., Jiang, Y., Ye, X., Deng, C., Zhang, J., & Li, X. (2020b). The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S. during the COVID-19 pandemic. medRxiv 2020.07.31.20143016. <https://doi.org/10.1101/2020.07.31.20143016>.
- Huang, X., Li, Z., Wang, C., & Ning, H. (2020c). Identifying disaster related social media for rapid response: A visual-textual fused CNN architecture. *International Journal of Digital Earth*, 13, 1017–1039. <https://doi.org/10.1080/17538947.2019.1633425>.
- Ivorra, B., Ferrández, M. R., Vela-Pérez, M., & Ramos, A. M. (2020). Mathematical modeling of the spread of the coronavirus disease 2019 (COVID-19) taking into account the undetected infections. The case of China. *Communications in Nonlinear Science and Numerical Simulation*, 88, 105303. <https://doi.org/10.1016/j.cnsns.2020.105303>.
- Jewell, N. P., Lewnard, J. A., & Jewell, B. L. (2020). Caution warranted: Using the Institute for Health Metrics and Evaluation model for predicting the course of the COVID-19 pandemic. *Annals of Internal Medicine*, 173, 226–227. <https://doi.org/10.7326/M20-1565>.
- Jun, S. W. (2020). Movement control order not a lockdown, says former health minister. In: Malay Mail. <https://www.malaymail.com/news/malaysia/2020/03/17/movement-control-order-not-a-lockdown-says-former-health-minister/1847232>. Accessed 20 Aug 2020.
- Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015). Understanding human mobility from Twitter. *PLoS ONE*, 10,. <https://doi.org/10.1371/journal.pone.0131469>.
- Kar, B. (2008). *Scaling modeled potential residential loss from a storm surge*. Ph.D., University of South Carolina.
- Kar, B., & Hodgson, M. E. (2012). Observational scale and modeled potential residential loss from a storm surge. *GIScience & Remote Sensing*, 49, 202–227. <https://doi.org/10.2747/1548-1603.49.2.202>.
- Karaye, I. M., & Horney, J. A. (2020). The impact of social vulnerability on COVID-19 in the U.S.: An analysis of spatially varying relationships. *American Journal of Preventive Medicine*, 59, 317–325. <https://doi.org/10.1016/j.amepre.2020.06.006>.
- Kedron, P., Frazier, A. E., Trgovac, A. B., Nelson, T., & Fotheringham, A. S. (2019). Reproducibility and replicability in geographical analysis. *Geographical Analysis*. <https://doi.org/10.1111/gean.12221>.

- Kermack, W. O., & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society London*, 115, 700–721.
- Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Covid, O., Hanage, W. P., Brownstein, J. S., Layan, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. G., & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. 6.
- Kuniya, T. (2020). Prediction of the epidemic peak of coronavirus disease in Japan, 2020. *Journal of Clinical Medicine*, 9, 789. <https://doi.org/10.3390/jcm9030789>.
- Lam, N. S.-N., & Quattrochi, D. A. (1992). On the issues of scale, resolution, and fractal analysis in the mapping sciences*. *The Professional Geographer*, 44, 88–98. <https://doi.org/10.1111/j.0033-0124.1992.00088.x>.
- Leguay, J., Friedman, T., & Conan, V. (2006). Evaluating mobility pattern space routing for DTNs. In *Proceedings of the IEEE INFOCOM 2006 25th IEEE international conference on computer communications* (pp. 1–10). <https://doi.org/10.1109/INFOCOM.2006.299>.
- Lekone, P. E., & Finkenstädt, B. F. (2006). Statistical inference in a stochastic epidemic SEIR model with control intervention: Ebola as a case study. *Biometrics*, 62, 1170–1177. <https://doi.org/10.1111/j.1541-0420.2006.00609.x>.
- Li, Z., Wang, C., Emrich, C. T., & Guo, D. (2018). A novel approach to leveraging social media for rapid flood mapping: a case study of the 2015 South Carolina floods. *Cartography and Geographic Information Systems*, 45, 97–110. <https://doi.org/10.1080/15230406.2016.1271356>.
- Mandelbrot, B. (1967). How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science*, 156, 636–638. <https://doi.org/10.1126/science.156.3775.636>.
- Marchant, R., Samia, N. I., Rosen, O., Tanner, M. A., & Cripps, S. (2020). Learning as we go: An examination of the statistical accuracy of COVID19 daily death count predictions. ArXiv200404734 Quantitative Biology Status.
- Martín, Y., Li, Z., & Cutter, S. L. (2017). Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. *PLoS ONE*, 12,. <https://doi.org/10.1371/journal.pone.0181701>.
- Martín, Y., Cutter, S. L., & Li, Z. (2020). Bridging twitter and survey data for evacuation assessment of Hurricane Matthew and Hurricane Irma. *Natural Hazards Review*, 21, 04020003. [https://doi.org/10.1061/\(ASCE\)NH.1527-6996.0000354](https://doi.org/10.1061/(ASCE)NH.1527-6996.0000354).
- McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review* 60–68.
- Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013). Is the sample good enough? Comparing data from Twitter's Streaming API with Twitter's Firehose. ArXiv13065204 Physics.
- Murray CJ, IHME COVID-19 health service utilization forecasting team. (2020). Forecasting the impact of the first wave of the COVID-19 pandemic on hospital demand and deaths for the USA and European Economic Area countries. *Infectious Diseases (except HIV/AIDS)*.
- National Academies. (2019). *Reproducibility and replicability in science, of Sciences, Engineering, and Medicine*. Washington, D.C.: National Academies Press.
- National Institute of Standards and Technology. (2020). *Research Data Framework (RDaF): Motivation, development, and a preliminary framework core*.
- O'Reilly, T. (2005). What is Web 2.0? Design patterns and business models for the next generation of software. In: O'Reilly.com. <https://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>. Accessed 20 Aug 2020.
- Openshaw, S. (1984). The modifiable areal unit problem. *Concepts and techniques in modern geography* (pp. 1–41). Norwich: Geo Books.
- Osthus, D., Hickmann, K. S., Caragea, P. C., Higdon, D., & Del Valle, S. Y. (2017). Forecasting seasonal influenza with a state-space SIR model. *The Annals of Applied Statistics*, 11, 202–224. <https://doi.org/10.1214/16-AOAS1000>.
- Qiu, J., Shen, B., Zhao, M., Wang, Z., Xie, B., & Xu, Y. (2020). A nationwide survey of psychological distress among Chinese people in the COVID-19 epidemic: Implications and policy recommendations. *Gen Psychiatry*, 33,. <https://doi.org/10.1136/gpsych-2020-100213>.

- Quijas, S., Balvanera, P. (2013). Biodiversity and ecosystem services. In: *Encyclopedia of biodiversity* (pp 341–356). Elsevier.
- Ray, D., Salvatore, M., Bhattacharyya, R., Wang, L., Du, J., Mohammed, S., et al. (2020). Predictions, role of interventions and effects of a historic national lockdown in India's response to the COVID-19 pandemic: Data science call to arms. *Harvard Data Science Review*. <https://doi.org/10.1162/99608f92.60e08ed5>.
- Roy, A., Kar, B. (2020). Characterizing the spread of COVID-19 from human mobility patterns and SocioDemographic indicators. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on advances in resilient and intelligent cities* (pp 39–48). ACM, Seattle Washington.
- Sloan, L., & Morgan, J. (2015). Who Tweets with their location? Understanding the relationship between demographic characteristics and the use of geoservices and geotagging on Twitter. *PLoS ONE*, 10,. <https://doi.org/10.1371/journal.pone.0142209>.
- Song, P. X., Wang, L., Zhou, Y., He, J., Zhu, B., Wang, F., Tang, L., & Eisenberg, M. (2020). An epidemiological forecast model and software assessing interventions on COVID-19 epidemic in China. *Infectious Diseases (except HIV/AIDS)*.
- Tullis, J. A., & Kar, B. (2020). Where is the provenance? Ethical replicability and reproducibility in giscience and its critical applications. *American Association of Geographers Annals*. <https://doi.org/10.1080/24694452.2020.1806029>.
- Tullis, J. A., Alsumaiti, T. S., Cothren, J. D., Lanter, D. P., Limp, W. F., Linck, R. F., Shi, X., & Young, S. G. (2015). Geoprocessing, workflows, and provenance. In P. S. Thenkabail (Ed.) *Remote sensing handbook. Remotely sensed data characterization, classification, and accuracies* (1st edn, vol. 1, p. 678). Taylor & Francis.
- W3C Provenance Incubator Group. (2010). *Provenance XG final report*. W3C.
- Wang, Z., & Tang, K. (2020). Combating COVID-19: Health equity matters. *Nature Medicine*, 26, 458. <https://doi.org/10.1038/s41591-020-0823-6>.
- Wang, J., Hu, Y., & Joseph, K. (2020a). NeuroTPR: A neuro-net toponym recognition model for extracting locations from social media messages. *Transactions in GIS*, 24, 719–735. <https://doi.org/10.1111/tgis.12627>.
- Wang, C., Liu, L., Hao, X., Guo, H., Wang, Q., Huang, J., He, N., Yu, H., Lin, X., Pan, A., Wei, S., & Wu, T. (2020b). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. *Epidemiology*.
- Webb Hooper, M., Nápoles, A. M., & Pérez-Stable, E. J. (2020). COVID-19 and racial/ethnic disparities. *JAMA*, 323, 2466. <https://doi.org/10.1001/jama.2020.8598>.
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Towards an ethical framework for publishing Twitter data in social research: Taking into account users' views, online context and algorithmic estimation. *Sociology*, 51, 1149–1168. <https://doi.org/10.1177/0038038517708140>.
- Xu, P., Dredze, M., Broniatowski, D. A. (2020). The twitter social mobility index: Measuring social distancing practices from geolocated tweets. ArXiv200402397 Cs.

Chapter 9

Multi-level Inter-regional Migrant Population Estimation Using Multi-source Spatiotemporal Big Data: A Case Study of Migrants in Hubei Province During the Outbreak of COVID-19 in Wuhan



Jiale Qian, Zhang Liu, Yunyan Du, Nan Wang, Jiawei Yi, Yeran Sun, Ting Ma, Tao Pei, and Chenghu Zhou

9.1 Introduction

By the end of January 2020, the prevention, control, and tracking of the novel coronavirus (COVID-19) had entered a new stage. Resolutely preventing the spread of the epidemic has become the top priority for China. Since Wuhan announced the “city closure” on January 23, 2020, many other cities in Hubei Province also adopted the “city closure” strategy to block the flow of people. However, due to the Spring Festival travel before the Chinese New Year (January 25, 2020), a large number of students and migrant workers returned home before the holiday, resulting in a large-scale population movement across the country. According to the information released by the Hubei provincial government on the evening of January 26, 2020 (People’s Daily), more than 5 million people had left Wuhan before “city closure,” accounting for 35.7% of the outflow population. The whereabouts of these people became one of the focuses of epidemic prevention and control. Baidu’s migration big data showed that more than 70% of the outflow population from Wuhan went to other

J. Qian · Z. Liu · Y. Du (✉) · N. Wang · J. Yi · Y. Sun · T. Pei · C. Zhou

State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

e-mail: duyy@lreis.ac.cn

University of Chinese Academy of Sciences, Beijing 100049, China

T. Ma

School of Geography and Planning, Sun Yat-Sen University, Guangzhou 510275, China

Department of Geography, Swansea University, Swansea S28PP, UK

areas in Hubei Province. Therefore, the rapid and accurate estimation of the population number and its spatial distribution characteristics from Wuhan to other areas can provide a scientific basis for making epidemic prevention decisions. The key to solve this problem is to use multi-source spatiotemporal big data, combined with a holiday population migration network and dynamic population estimation method, to estimate the number of inter-regional population migration and its dynamic distribution in different spatial structures within the region.

In recent years, with the rapid development of location-aware big data, it is possible to perceive the fine-scale individual trajectories and collective human activity dynamics at a large scale, which provides an opportunity for observing large-scale migration events and dynamic population estimation during holidays. At present, the research on population migration during holidays mainly focuses on the analysis of temporal and spatial patterns and network structure characteristics of population migration (Wang et al. 2019, 2014; Hu 2019; Wei et al. 2018; Li et al. 2016; Zhu et al. 2018; Hu et al. 2017; Xu et al. 2017). For example, Wang et al. (2014) used Baidu migration data to analyze the spatial and temporal characteristics of population migration during the Spring Festival, and the results showed that the destinations and sources of the migrant population had obvious geographical proximity; Li et al. (2016) used multi-source tourism big data to study the temporal and spatial characteristics of Spring Festival tourism, and the results showed that the tourism network had significant geographic agglomeration and multi-center characteristics; Hu et al. (2017) used the Weibo data to study the pattern of personnel movement during the Spring Festival, the results showed that most people only move between a small number of cities, and work and study are the main driving force of population movement; Xu et al. (2017) used Tencent migration big data during the Spring Festival, and the network analysis method to assess the unbalanced migration of population between cities and the spatial differences in urban development. The existing studies can better reveal the spatial-temporal patterns and network interaction characteristics of population flow at different scales during holidays. However, intercity migration is often measured by the relative proportion of inflows and outflows, and less attention is paid to the estimation of the real population number of intercity migration, which makes it an urgent research problem to estimate the number of people from Wuhan to other cities of Hubei Province combined with the big data of open-source migration.

The core idea of dynamic population estimation is to use the quantitative relationship between location-aware big data and the ground real population to estimate the fine-scale dynamic population distribution. The existing data sources used in dynamic population estimation studies mainly include mobile phone signaling data, social media big data, subway card data, etc. (Leyk et al. 2019; Wardrop et al. 2018; Yao et al. 2017; Patel et al. 2017; Kontokosta and Johnson 2017). There are three kinds of methods: (1) Weighted interpolation method. Based on the data or auxiliary variables, a simple weighted model is constructed to obtain the population distribution weight, and then the location perception big data is transformed into the ground population (Kubíček et al. 2019; Ma et al. 2017; Tsou et al. 2018). For example, Kubíček et al. (2019) used a comprehensive weighting method of building floor area and function type to realize dynamic population estimation based on mobile phone

data. (2) Statistical modeling method. Statistical regression models are used to model the functional relationship between location-aware big data and census statistics and estimate dynamic population (Deville et al. 2014; Liu et al. 2018; Khodabandeh et al. 2018, 2016; Feng et al. 2018). For example, Deville et al. (2014) used a logarithmic linear model to construct the functional relationship between the street-scale nighttime mobile phone user density and the census population density and then estimated the population dynamics at the fine spatio-temporal scale. (3) Artificial intelligence modeling method. Based on machine learning or deep learning algorithm, the spatiotemporal patterns and influencing factors of population distribution are learned to estimate the dynamic population (Zong et al. 2019; Chen et al. 2018). For example, Chen et al. (2018) used a neural network to model the temporal dependence and spatial correlation of the grid-scale population flow based on the mobile phone CDRs data and then made a near real-time prediction of the urban population. In general, the existing dynamic population estimation methods can effectively use location-aware big data to estimate the dynamic population at a fine scale. However, few studies have focused on the quickly dynamic estimation of the number of inter-regional population migration and its distribution in different spatial structures within the region from the perspective of population migration.

In view of the above problems, a multi-level spatial distribution dynamic estimation model of inter-regional migration population, which integrates multi-source geographic spatiotemporal big data, such as population migration big data, aggregated location perception big data, and land cover data, was proposed to quickly estimate the number of people flowing from Wuhan to cities, counties, and rural areas of Hubei Province before the new year's eve of 2020 (January 24, 2020) and analyze the spatial distribution characteristics of the migrant population in different scales. Firstly, the dynamic estimation method of multi-level spatial distribution of inter-regional migration population was described. Secondly, the population change characteristics at city scale during the Spring Festival in 2018, the population distribution characteristics of all cities from Wuhan in 2020, and the spatial pattern of population change at district and county level were analyzed. Finally, based on the population flow pattern of cities and counties in Hubei Province during the Spring Festival, the future development of epidemic situation was judged and discussed. For the sake of simplicity and convenience, we specifically defined: (1) Normal time: from 46 to 26 d before the first day of the New Year. Normal time in 2018: January 1–January 21, 2018; (2) Spring Festival: the first to the fourth day of the Lunar New Year. Spring Festival in 2018: February 16–February 19, 2018. Spring Festival in 2020: January 25–January 28, 2020. (3) The 12 days before New Year's Eve represents January 11 to January 22, 2020.

9.2 Data Sources and Methods

9.2.1 Data Sources

The data source that we used in this study contained Tencent location request data (TLR), Baidu migration data, Urban area data, 2018 statistical yearbooks data, and a cumulative number of confirmed covid-19 infection.

- (1) Tencent location request data was used to estimate the daily dynamic population of cities and counties in Hubei Province during the normal period and the Spring Festival in 2018. The dataset comes from Tencent platform (<https://heat.qq.com/>), which records the location of users when using location-based services (LBS) provided by Tencent series products. These software include Wechat, QQ, Jingdong, Didi travel, Wangzhe glory, etc. The service content covers social networking, games, shopping, travel, communication, and other aspects. It receives as many as 60 billion positioning calls every day, covering up to 1 billion smartphone users. Therefore, TLR has a typical representative in dynamic population estimation research (Chen et al. 2019). The spatial resolution of the dataset is about $1 \text{ km} \times 1 \text{ km}$, and the temporal resolution is 1d.

Because the TLR cannot be obtained during the Spring Festival transportation in 2020 and needs regression modeling based on the year-end permanent population statistics data at the prefecture-level city before estimating the dynamic population number of local cities. At present, only the prefecture-level city-scale permanent population statistics data at the end of 2018 can be collected completely (the official data of 2019 has not been officially released).

Besides, as shown in Fig. 9.1, the crowd began to flow around 25 days before the New Year's Day in 2018 (January 22–February 15) and reached the highest or the lowest value on the New Year's Eve (February 15), which means that the population inflow or outflow is basically completed before the New Year's Eve, and will remain stable for a short period from the first to the fourth day, and then the flow of people increases rapidly. Wuhan adopted the “closed city” measure at 10 am on January 23, 2020. At this time, the population has nearly completed the migration before the New Year's Eve. Therefore, this article believes that the use of TLR during the Spring Festival in 2018, can better characterize 2020.

Based on the above two factors, the TLR from January 1 to March 18, 2018 is used for the analysis. The Spring Festival transportation in 2018 starts from February 1, 2018, to March 12, 2018, and February 15, 2018, is the New Year's Eve.

- (2) Baidu migration data, which is from Baidu map insight platform (<https://qia.nxi.baidu.com/>), is used to estimate the number of people flowing from Wuhan to cities in Hubei Province before the Spring Festival in 2020. The dataset provides the ratio of the population number who move in or out of other cities

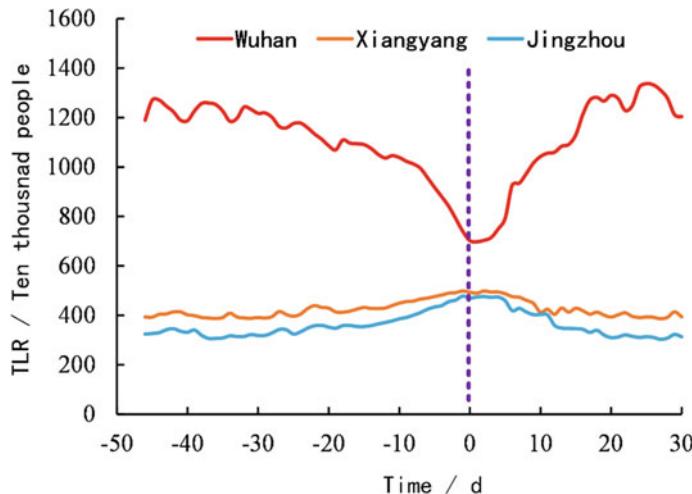


Fig. 9.1 Trends of Tencent location requests in prefecture-level cities during the 2018 Spring Festival. The x-coordinate is the number of days from January 16, 2018 (New Year's Day), and the y-coordinate is the total number of Tencent location requests (units: 10,000)

every day to the total number of people in or out of the city. In this paper, we obtain the proportion of people flowing from Wuhan in the 12 days before New Year's Eve in Hubei Province.

Combined with Baidu's migration data, it can be found (Fig. 9.2) that in the week before the Spring Festival in 2020 (January 18–January 24), the top 14 cities in the number of people emigrating from Wuhan are all cities in Hubei Province, accounting for more than 70% of the total emigrating population. Among them, the proportion of people emigrating to Huanggang, Xiaogan, Jingzhou, and Xianning is the highest. In Hubei Province, Wuhan is basically

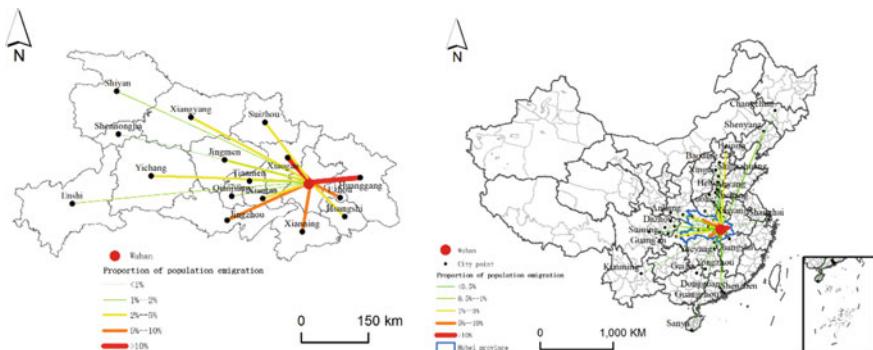


Fig. 9.2 The proportion of population movement from Wuhan to Hubei province and nationwide in the 12 days before New Year's Eve 2020

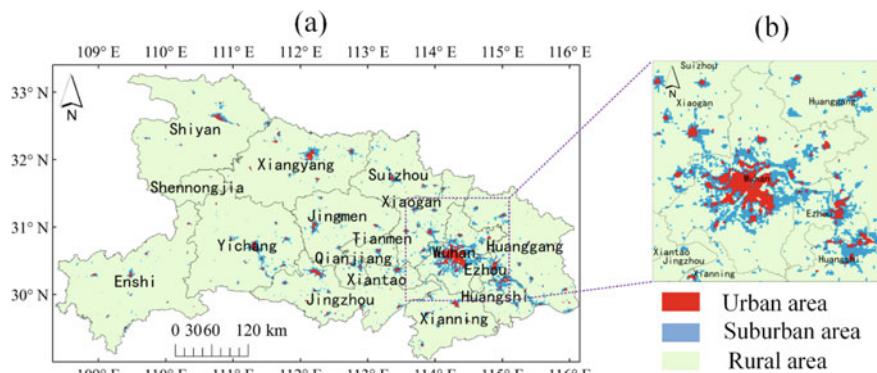


Fig. 9.3 Urban, suburban, and rural areas in Hubei Province in 2016

the source of migration. Therefore, our basic judgment is that a large number of people from Wuhan will flow into other prefecture-level cities in Hubei Province before the Spring Festival.

- (3) Urban area: Hubei Province has 17 cities and 103 counties. The resident population of the province is 58.51 million. Among them, Wuhan is the capital city of Hubei Province, which has the largest population (11.08 million). In order to obtain the extent of the urban, suburban, and rural areas of each city, we use two data sources. The urban area in 2016 was extracted by Song et al. (2018) according to POI and road network. The total scope of urban and suburban areas in 2017 was defined by Gong et al. (2019) based on the impervious surface extracted from multi-source remote sensing data such as Landsat, sentinel-1 radar data, and night light remote sensing data, with a spatial resolution of 30 m. The data does not distinguish between the urban and suburban areas. Through the combination of the two data sources, the extent of the urban, suburban, and rural areas in Hubei Province is obtained (Fig. 9.3), with a spatial resolution of approximately 1 km × 1 km, which is consistent with Tencent location request data.
- (4) At the end of 2018, the permanent population statistics of prefecture-level cities in China are derived from the statistical yearbooks published by provincial statistical bureaus, which are used to model and calculate the real population. Considering the synchronization and integrity between Tencent positioning request big data and prefecture-level city-scale year-end permanent population statistics, this paper adopts the National prefecture-level city permanent population statistics at the end of 2018.
- (5) As of January 29, 2020, the cumulative number of confirmed Covid-19 infection in all prefecture-level cities in Hubei Province was obtained from the website of Hubei Provincial Health Committee (<https://wjw.hubei.gov.cn/>). The published data were used to analyze the correlation between population change and cumulative infection at prefecture level.

9.2.2 Research Methods

9.2.2.1 Technical Route

The research methods in this paper are mainly divided into three parts (Fig. 9.2). Firstly, based on the TLR at grid scale, the total number of TLR per day in each county and its urban, suburban, and rural areas was calculated based on the grid-scale TLR, and the dynamic population number of each city and county and its urban, suburban and rural areas was estimated by combining with the demographic data. Then, based on Baidu's migration data 12 days before the New Year's Eve of 2020, we calculated the population number of cities and counties in Hubei Province and their internal regions from Wuhan. Finally, we analyzed the characteristics of population changes during the Spring Festival in each city and county.

9.2.2.2 Estimation Method

(1) Dynamic population estimation in cities and counties

First of all, the average value of the total number of TLR in various cities in the normal period of 2018 was calculated. Then, a logarithmic linear regression model of the total number of TLR and the resident population at the prefecture scale was established (Eq. (9.1)).

$$\log(\rho_i) = a \times \log(\sigma_i) + b \quad (9.1)$$

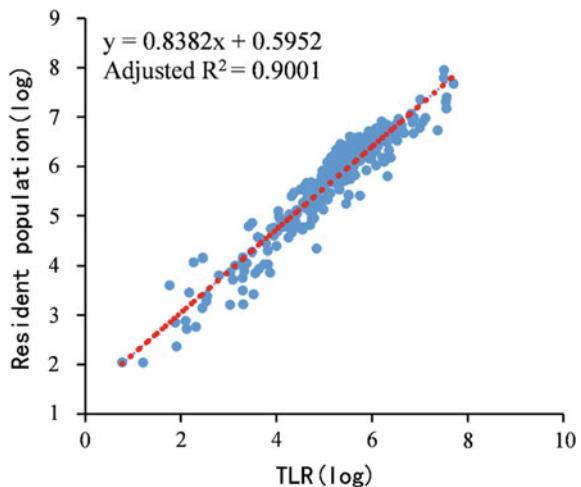
where σ is the total number of TLR at the prefecture-level city scale; ρ is the number of resident population at the end of the prefecture scale; i represents a single prefecture-level city; a is the super-linear impact of resident population on TLR; b is the scale ratio.

It can be seen that the relationship between the TLR and resident population is very stable (Fig. 9.4). The correlation is as high as 0.90.

At the same time, we calculate the total number of TLR in each city and county and its urban, suburban, and rural areas for the normal period and the Spring Festival period in 2018, and then use a logarithmic linear regression model to calculate the number of daily dynamic populations within the county and its urban, suburban, and rural areas each city, and the estimated value of the total urban population fine-tunes the estimated value of the population in the urban, suburban, and rural areas, the formula is as follows:

$$\rho_{region(i)}' = \frac{P}{\widehat{P}} \rho_{region(i)} \quad (9.2)$$

Fig. 9.4 Regression relationship between Tencent location requests and resident population



Formulas $\rho_{region(i)}$ and $\rho_{region(i)}'$ represent the estimated population of urban, suburban, or rural areas of city and county i and the estimated population after fine-tuning; \hat{P} is the estimated population of urban, suburban, and rural areas in city and county i ; Sum of values; P is the estimated population of city and county i .

(2) Estimated from the population of Wuhan who moved to each level city

In order to estimate the number of people who flowed from Wuhan to other cities in Hubei Province before the “closing of the city,” we first use the population average pop_{spring} during the Spring Festival in 2018 minus the average population $pop_{background}$ during the normal time to get the population change during the Spring Festival in each city and county. Then, using Baidu’s migration big data, we calculate the average proportion $p_{migration}$ of people who migrated from Wuhan to the city’s total population 12d before the New Year’s Eve in 2020, and then multiply it by the number of population changes in each city during the Spring Festival to get the number of people who moved in from Wuhan before the New Year’s Eve. The specific formula is as follows:

$$pop_{(i)} = (pop_{spring(i)} - pop_{background(i)}) \times p_{migration(i)} \quad (9.3)$$

where i represents a city.

Then, we estimate the upper limit $pop_{upper(i)}$ and lower limit $pop_{lower(i)}$ of the number of people flowing from Wuhan into rural areas of each city in Hubei Province. The inference of the upper limit and lower limit will be elaborated in Sect. 3.2. The specific calculation process is as follows: ① Calculate the ratio of the rural population to the total population in each city during the Spring Festival in 2018 and set it as $p_{sum(i)}$. It is assumed that the probability of migrants from Wuhan flowing into different regions of the city is equal, Then according to formula (9.4), the lower limit

$\text{pop}_{lower(i)}$ of the number of people flowing from Wuhan into rural areas of each city in Hubei Province can be obtained; ② Calculate the proportion of the rural increasing population during the Spring Festival in 2018 compared to the normal period to the total increase in the urban population in each city. The proportion of the population increase in each region is set as $\text{pop}_{change(i)}$, then the upper limit $\text{pop}_{upper(i)}$ of the people flowing from Wuhan into rural areas of each level city in Hubei Province can be obtained from Eq. (9.5).

$$\text{pop}_{lower(i)} = p_{sum(i)} * \text{pop}_{(i)} \quad (9.4)$$

$$\text{pop}_{upper(i)} = \text{pop}_{change(i)} * \text{pop}_{(i)} \quad (9.5)$$

We then further analyzed the characteristics of the total population changes in each city during the Spring Festival, the distribution characteristics of people who migrated from Wuhan to each city, and the spatial pattern of population changes in districts and counties.

9.3 Results and Analysis

9.3.1 Characteristics of Population Change at City Scale During the Spring Festival

The average number of people in each city during the Spring Festival in 2018 is shown in Table 9.1. It can be seen that: ① The cities with the largest urban population during the Spring Festival were Wuhan (7.047 million), Huanggang (6.027 million), Xiangyang (4.944 million) and Jingzhou (4.734 million); ② In terms of urban population ratio, except for Wuhan which reached 67.2%, all other cities were less than 50%, with an average ratio of 31%; ③ The cities with the largest population in rural areas are Huanggang (3.514 million), Jingzhou (2.792 million), and Xiangyang (2.410 million) and Enshi (2.315 million); ④ In terms of the proportion of the rural population, except for Wuhan (17.4%) and Ezhou (27.6%) with lower proportions, the other cities all reach or exceed 40%, with an average proportion of 51.7%.

Table 9.2 shows the population change during the Spring Festival in 2018 in each city compared to the normal period.

- (1) The outflow of people during the Spring Festival in Wuhan reached 5.154 million. The top three cities with the largest population increase during the Spring Festival in Hubei Province were Huanggang (2.513 million), Jingzhou (1.479 million), and Xiaogan (1.119 million). Combined with Baidu migration big data, it is found that the top three cities in Wuhan's migration destinations

Table 9.1 Population of prefecture-level cities in Hubei Province during the 2018 Spring Festival
(unit: the first four variables' unit are all ten thousand)

City	City pop change	Urban pop change	Suburban pop change	Rural pop change	Urban pop change (%)	Suburban pop change (%)	Rural pop change (%)
Wuhan	704.7	473.8	108.3	122.5	67.2%	15.4%	17.4%
Huanggang	602.7	148.8	102.5	351.4	24.7%	17.0%	58.3%
Jingzhou	494.4	178.0	75.4	241.0	36.0%	15.3%	48.8%
Xiaogan	473.4	124.8	69.3	279.2	26.4%	14.6%	59.0%
Enshi	420.2	134.0	90.6	195.6	31.9%	21.6%	46.5%
Xiangyang	369.5	86.0	52.0	231.5	23.3%	14.1%	62.7%
Xianning	332.3	115.1	55.3	161.9	34.6%	16.6%	48.7%
Tianmen	326.6	135.0	50.6	141.0	41.3%	15.5%	43.2%
Huangshi	290.0	103.3	71.5	115.1	35.6%	24.7%	39.7%
Suizhou	282.5	104.8	46.5	131.1	37.1%	16.5%	46.4%
Shiyan	237.3	74.5	36.8	126.0	31.4%	15.5%	53.1%
Xiantao	214.9	64.9	52.1	98.0	30.2%	24.2%	45.6%
Jingmen	150.0	48.1	33.2	68.7	32.1%	22.1%	45.8%
Qianjiang	143.5	27.1	22.0	94.5	18.9%	15.3%	65.8%
Ezhou	138.0	57.2	42.6	38.1	41.5%	30.9%	27.6%
Yichang	110.3	28.4	26.8	55.1	25.8%	24.3%	49.9%
Shennongjia	11.8	3.0	1.6	7.3	25.1%	13.5%	61.4%

are Xiaogan, Huanggang, and Jingzhou. The results from the two data sources are consistent.

- (2) The amount of population change in different regions within the same city is quite different. During the Spring Festival, the number of people in Wuhan's urban and suburban areas decreased by 5.663 million people, while the number of people in rural areas increased by 509,000. In other cities, except Huanggang (107,000 people), Tianmen (47,000 people), and Suizhou (43,000 people) showing a slight increase in the number of urban residents, the number of urban residents in other cities all decreased during the Spring Festival. The cities with the largest decrease were Xiangyang (271,000), Jingzhou (244,000), and Enshi (230,000). At the same time, the number of people in most suburbs has increased slightly. The number of people in rural areas has increased, with the largest increase in Huanggang (2.21 million), Jingzhou (1.599 million), Enshi (1.213 million), Xiangyang (1.171 million), and Xiaogan (1.056 million).
- (3) From the perspective of the proportion of the increase in the number of people in the rural areas of each city in the total change of the urban population, the average proportion of the cities except Wuhan is 124.7%. Combining the

Table 9.2 Population changes of prefecture-level cities in Hubei Province during the 2018 Spring Festival. The highest to lowest value in each column is represented by a gradient color from red to blue (unit: the first four variables' unit are all ten thousand)

City	City pop change	Urban pop change	Suburban pop change	Rural pop change	Urban pop change (%)	Suburban pop change (%)	Rural pop change (%)
Wuhan	-515.4	-495.9	-70.4	50.9	96.2%	13.7%	-9.9%
Huanggang	251.3	10.7	22.8	217.8	4.2%	9.1%	86.7%
Jingzhou	147.9	-24.4	12.4	159.9	-16.5%	8.4%	108.1%
Xiaogan	111.9	-5.0	11.3	105.6	-4.5%	10.1%	94.3%
Enshi	96.6	-23.0	-1.6	121.3	-23.8%	-1.7%	125.5%
Xiangyang	96.3	-27.1	6.3	117.1	-28.2%	6.5%	121.6%
Xianning	75.1	-1.7	4.0	72.7	-2.3%	5.4%	96.9%
Tianmen	68.9	4.7	5.1	59.1	6.8%	7.4%	85.7%
Huangshi	68.0	-6.9	8.2	66.7	-10.1%	12.1%	98.1%
Suizhou	65.3	4.3	11.6	49.5	6.5%	17.7%	75.8%
Shiyan	56.2	-18.7	5.3	69.6	-33.4%	9.4%	123.9%
Xiantao	46.2	-0.6	6.1	40.7	-1.4%	13.2%	88.1%
Jingmen	41.5	-19.4	0.9	60.0	-46.8%	2.1%	144.6%
Qianjiang	27.6	-4.1	4.7	27.1	-14.8%	16.9%	97.9%
Ezhou	22.2	-1.9	5.5	18.6	-8.6%	24.6%	84.0%
Yichang	21.4	-41.5	-8.3	71.1	-194.2%	-38.7%	333.0%
Shennongjia	0.8	-0.7	-0.3	1.8	-94.9%	-36.5%	231.3%

characteristics of the change in the number of urban, suburban, and rural populations during the Spring Festival period. We infer that for each city, not only a large number of people from other cities flow into the countryside, but also the people who move out of the city's own urban area will also flow into the countryside, leading to a substantial population increase in rural areas. Among them, the cities with the highest proportion are Yichang (333.0%), Shennongjia Forest Area (231.3%), Jingmen (144.6%), Enshi (125.5%), Shiyan (123.9%), and Xiangyang (121.6%).

In general, during the Spring Festival, except for Wuhan, other cities are dominated by population inflows as a whole. But there are large differences in population flows in different regions within cities. Urban areas are dominated by small population outflows, and rural areas are dominated by large population inflows..

9.3.2 Distribution Characteristics of People Who Migrated from Wuhan During the Spring Festival

Table 9.3 shows the estimation results of the number of people flowing from Wuhan to rural areas in Hubei Province before the Spring Festival (January 25, 2020). The following three points can be learned.

- (1) Before the Spring Festival in 2020, the cities with the highest average proportion of migrants from Wuhan to the total urban migrants were Xiaogan (56.8%), Ezhou (46.2%), and Xiantao (37.4%). The cities with the largest number of immigrants from Wuhan before the Spring Festival are Huanggang (893,800), Xiaogan (635,200), and Jingzhou (338,300).
- (2) As can be seen from the proportion of the population flowing into rural areas of each city in Wuhan, the average lower limit ratio is 50% and the average upper limit ratio is 90%. It can also be seen from the upper and lower limits of the population inflows from Wuhan to urban and rural areas that the sum of

Table 9.3 Population migrates from Wuhan to rural areas in prefecture-level cities before New Year's Eve, January 24, 2020 (unit: inflow number, lower limit to rural, upper limit to rural' unit are ten thousand. Confirmed cases' unit is per person)

City	Inflow percentage (%)	Inflow number	Lower limit to rural (%)	Upper limit to rural (%)	Lower limit to rural	Upper limit to rural	confirmed cases
Huanggang	35.6%	89.4	58.3%	86.7%	52.1	77.5	324
Xiaogan	56.8%	63.5	46.5%	90.3%	29.6	57.4	274
Jingzhou	22.9%	33.8	59.0%	92.8%	20.0	31.4	101
Xianning	35.7%	26.8	46.4%	94.7%	12.4	25.4	112
Tianmen	25.4%	17.5	65.8%	85.7%	11.5	15.0	34
Suizhou	32.7%	21.4	45.6%	75.8%	9.7	16.2	116
Xiangyang	20.5%	19.7	48.8%	94.9%	9.6	18.7	131
Xiantao	37.4%	17.2	45.8%	87.0%	7.9	15.0	32
Huangshi	27.8%	18.9	39.7%	89.0%	7.5	16.8	86
Enshi	11.5%	11.1	62.7%	100.0%	6.9	11.1	51
Jingmen	28.6%	11.9	53.1%	98.6%	6.3	11.7	142
Shiyan	19.2%	10.8	43.2%	92.9%	4.7	10.0	88
Qianjiang	23.5%	6.5	49.9%	85.3%	3.2	5.5	8
Ezhou	46.2%	10.2	27.6%	77.3%	2.8	7.9	84
Yichang	19.1%	4.1	48.7%	100.0%	2.0	4.1	63
Shennongjia	14.8%	0.1	61.4%	100.0%	0.1	0.1	3

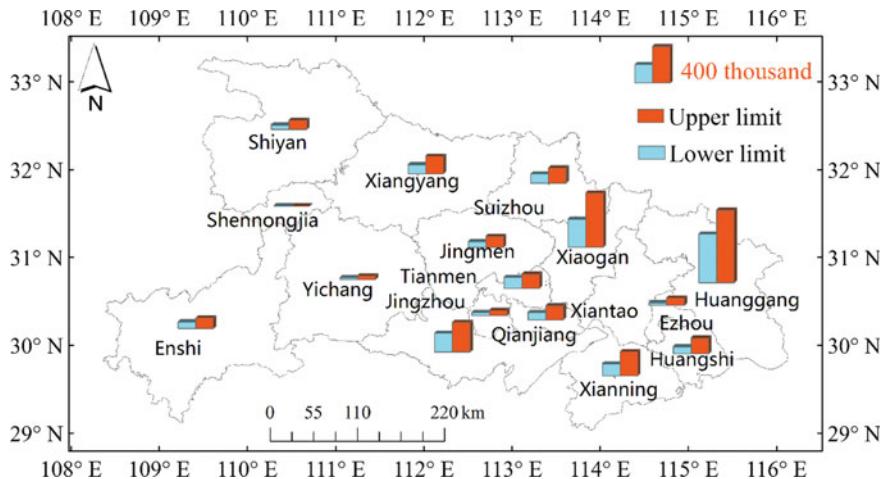


Fig. 9.5 Spatial distribution of the upper and lower limits of the population who immigrate from Wuhan to rural areas in prefecture-level cities before New Year's Eve, January 24, 2020

the population inflows from Wuhan to urban and rural areas ranges from 1.864 million to 3.238 million, of which Huanggang (521,000), Xiaogan (296,000 people), and Jingzhou (200,000 people) have a lower limit of more than 200,000. In terms of spatial distribution (Fig. 9.5), these three cities all directly border Wuhan. On the whole, the pressure on epidemic prevention and control in these three cities may be greater, and the epidemic prevention situation in rural areas should be focused on.

- (3) In addition, the Pearson correlation coefficients between the upper limit and lower limit of the number of people flowing from Wuhan into rural areas and the cumulative number of infections in each city are 0.91 and 0.87 (p -value ≤ 0.001), while the Pearson correlation coefficient between the population change and the cumulative number of infections is relatively low (Pearson's $r = 0.77$, p -value ≤ 0.001), indicating that to a certain extent the pressure of epidemic prevention in rural areas needs to be taken seriously.

9.3.3 Characteristics of Population Changes in Districts and Counties During the Spring Festival

The spatial distribution of the total population change during the Spring Festival period in the district and county scale of Hubei Province in 2018 presents three circle structures compared with the usual period (Fig. 9.6a). ① The first circle is the central area of the epidemic, including Wuhan City and its surrounding districts and counties, with population outflow mainly. ② The second circle is the focus area, including Huanggang, Huangshi, Xiantao, Tianmen, Qianjiang, Suizhou, Xiangyang,

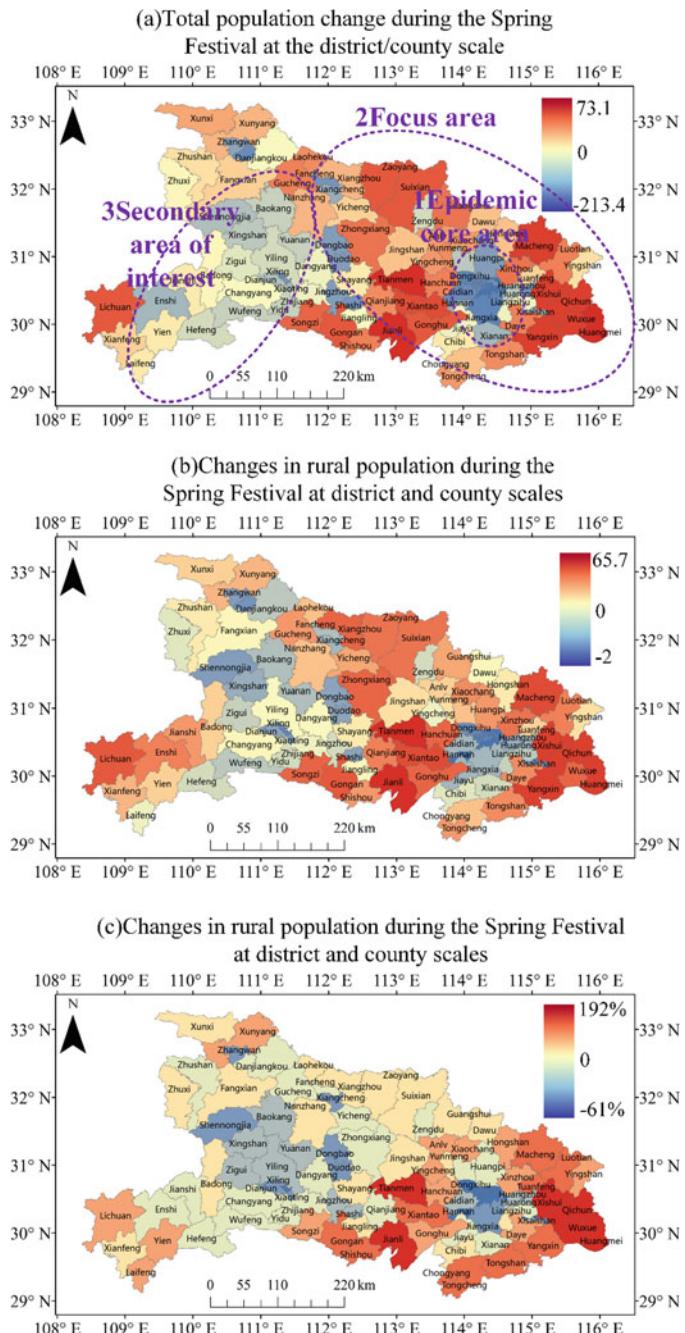


Fig. 9.6 The spatial distribution of county-level population change during the 2018 Spring Festival in Hubei Province

and parts of Xiaogan, Jingmen, Jingzhou, and Xianning. Combined with the cumulative confirmed case data (as of January 29, 2020), Huanggang (324 cases), Xiaogan (274 cases), Jingmen (142 cases), Xiangyang (131 cases), Suizhou (116 cases), The number of confirmed cases in Xianning (112 cases) and Jingzhou (101 cases) has exceeded 100, showing a rapid growth trend. This shows that the severity of the epidemic in each city and county is highly positively correlated with the influx of people in Wuhan, and it is in the rapid spread of the epidemic. ③ The third circle is a secondary concern area, including Yichang, Enshi, Shennongjia forest areas in western Hubei and parts of Jingmen, which is weaker than the second circle. At the same time, the population change in rural areas during the Spring Festival at the county level shows a spatial structure that is highly similar to that of the total population change (Fig. 9.6b), and the spatial distribution of the ratio of the population change in rural areas to the total population change during the Spring Festival at the county level (Fig. 9.6c) shows that in most districts and counties in cities around Wuhan, such as Huanggang, Huangshi, Xianning, Jingzhou, Xiantao, Tianmen, and Xiaogan, the change in population in rural areas and the change in total population during the Spring Festival The proportions are all over 100%, reaching an average of 140%, indicating that a large number of people in these areas have flowed into rural areas during the Spring Festival. It can be inferred from this that the pressure on epidemic prevention and control in the second circle may be greater. Attention should be paid to strengthening epidemic prevention and control measures in the second circle around Wuhan, and particular attention should be paid to population control in rural areas in the second circle.

Specific to the population changes during the Spring Festival of a single district and county, it is found that the districts and counties in Hubei Province with the largest population inflow during the Spring Festival include Jianli County in Jingzhou (730,700), Tianmen (680,900), and Huangmei in Huanggang (657,000), Yangxin, Huangshi (618,300). Especially in Huanggang, there are five counties with population changes over 500,000 during the Spring Festival, namely Huangmei (657,000), Qichun (591,800), Xishui (547,100), Macheng (525,600) People), Wuxue (522,300 people). After calculation, the counties with the largest population changes in rural areas include Jianli County in Jingzhou (656,800), Tianmen (590,500), Yangxin of Huangshi City (507,400), and Huanggang has four counties with population during the Spring Festival. The amount of change exceeded 400,000, and they were Huangmei (498,500), Qichun (473,300), Xishui (493,900), and Macheng (448,100). It can be seen that epidemic prevention and control measures in these districts, counties, and rural areas should be highly valued.

9.4 Discussion

The uncertainty of the results of this study is mainly caused by the following four aspects:

Firstly, since TLR data cannot distinguish the number of individual users, using the relative ratio of the TLR with the number of background populations to estimate the number of returnees will bring certain deviations. And due to data limitations, this study did not consider the spatial heterogeneity of background population distribution and flow characteristics. This will also bring certain deviations to the population estimation results.

Secondly, taking into account the age group and coverage limitations of users of Tencent's social media platform, as well as the differences in the frequency of use of location request services by people in different regions, the estimation result may, on the one hand, underestimate the number of younger and older returnees; on the other hand, lead to differences in population estimation errors in different spatial regions, such as the overestimation of urban populations and the underestimation of rural populations. As people enter rural areas from urban areas, the demand for location services may decrease due to fewer usage scenarios, which may result in an underestimation of population changes in rural areas during the Spring Festival. In addition, it should be noted that this article focuses on population changes in cities and counties during the Spring Festival rather than absolute population numbers. Therefore, the large number of left-behind elderly and children in rural areas will not have a significant impact on the analysis results of this article.

Thirdly, the resident population data used for correction did not consider the number of short-term migrants. At the same time, this article uses the 2018 TLR to participate in inferring the population movement during the 2020 Spring Festival, and does not take the change into account the 2 years population change (2018–2020). Both factors will affect the estimation results. However, we infer that the results of this article can still reflect the spatial pattern of population changes during the 2020 Spring Festival in Hubei Province because the scale and spatial pattern of population migration between cities each year are relatively stable and will not produce very large changes in the short term. Wuhan officially adopted the measures of “closing the city” at 10 am on January 23, 2020 (the day before New Year’s Eve, 2020). At this time, most of the population has completed migration before the Spring Festival. Therefore, the spatial distribution pattern of the population during the Spring Festival in 2020 should be relatively consistent with the Spring Festival in 2018.

Finally, limited by the obtaining difficulty of data sources, the accuracy of population estimation results needs further verification. At present, most dynamic population estimation is limited to the internal scale of cities, and there are very few researches on population dynamic estimation at the provincial or even national scale. This is mainly due to privacy protection and cross-regional data interoperability, which makes it extremely difficult to obtain large-scale mobile phone call detailed data and other data sources. And there is no large-scale detailed migration survey statistics for individuals during the Spring Festival in China. These factors make it impossible to directly verify the estimation results of population changes in cities and counties during the Spring Festival. However, Chen et al. (2019) used Tencent location big data to estimate the grid-scale population distribution in Nanjing and compared it with the international mainstream population products Worldpop and Landscan. They found that the population estimation based on Tencent location big

data can achieve better estimation accuracy at the grid scale. To a certain extent, this provides support for the accuracy verification of the estimation results in this paper.

Although the various restrictions mentioned above will produce a certain degree of uncertainty in the estimation results of this study. But considering the real-time and spatial clarity of TLR data, related studies (Ma et al. 2018) also show that similar calculations are useful for short-term populations. The estimation of the relative spatial distribution of migration is credible, which provides a reliable support for the estimation of the relative proportion of the spatial distribution of returnees in this study.

9.5 Conclusion

Due to the severe situation of the COVID-19 epidemic, Wuhan began to implement the “closed city” at 10:00 on January 23, 2020. However, due to the Spring Festival travel before the Chinese New Year (January 25, 2020), a large number of students and migrant workers have returned home before the holiday, resulting in a large-scale population movement across the country. More than 5 million people living or staying in Wuhan had left Wuhan. Baidu Migration Data shows that more than 70% of the outflow population from Wuhan went to other cities in Hubei Province. Quickly and accurately estimating the number of people flowing from Wuhan to other cities and counties in Hubei Province, and analyzing their spatial distribution characteristics can provide a scientific basis for preventing the spread of the epidemic and making epidemic prevention decisions. Based on crowd migration big data, we aggregate Tencent located request data, land cover data, and demographic data, and propose a dynamic estimation model of multi-level spatial distribution of inter-regional migrants that integrates multi-source geographic and temporal-spatial big data to dynamically estimate inter-regional population migration and its distribution in different spatial structures within the region.

The main conclusions of this article are as follows:

- (1) During the Spring Festival in 2018, the average proportion of the rural population in each city at the level of the urban population was 51.7%. With the exception of Wuhan, other prefecture-level cities were mainly population inflows. In most cities, the number of people in urban areas decreased slightly, and the number of people in suburbs increased slightly, while the number of people in rural areas increased significantly. The increase in the number of people in the rural areas of the city accounted for an average of 124.7% of the total changes in the urban population. This indicates that the population flow in different regions within the same city during the Spring Festival in 2018 was quite different, and the net population inflows in Hubei Province before the New Year's Eve on February 15, 2018, is highly inclined to enter rural areas.
- (2) The outflow of the Wuhan population during the Spring Festival in 2020 is about 5.154 million, accounting for 42% of the entire urban population, of

which 71% flowed to cities and counties in Hubei Province. At least 51.3% of the people who flowed from Wuhan to cities in Hubei Province flowed into rural areas. Among the prefecture-level cities, the most people who have flowed from Wuhan to rural areas are Huanggang (more than 296,000), Xiaogan (more than 296,000), and Jingzhou (more than 200,000). At the same time, the total inflow of people from Wuhan to these three cities is also the most.

- (3) According to the spatial pattern of the total population change during the Spring Festival in 2018 at the district and county scale, three epidemic prevention and control circles are delineated: ① The first circle is the core area of the epidemic, including Wuhan and its surrounding areas, with crowd outflows as the main area. ② The second circle is the focus area, including Huanggang, Huangshi, Xiantao, Tianmen, Qianjiang, Suizhou, Xiangyang, and parts of Xiaogan, Jingmen, Jingzhou, and Xianning. The total population and the large increase in population in rural areas are significantly higher than that of the districts and counties located in the third circle; ③ The third circle is a secondary concern area, including parts of Yichang, Enshi, Shennongjia, and Jingmen in western Hubei, with small population inflows.
- (4) During the Spring Festival in 2018, the districts and counties with the largest influx of people in administrative regions and rural areas include Jianli, Tianmen, Yangxin, Huanggang, and Huangmei, Qichun, Xishui, Macheng, and Wuxue. In these districts and counties, the number of people during the Spring Festival has increased by more than 500,000 compared to normal times, and the number of people in rural areas has increased by more than 400,000.

In general, we propose a dynamic estimation model for the multi-level spatial distribution of interregional migrants that integrates multi-source geographic space-time big data and realizes the flow from Wuhan to cities in Hubei Province before the Spring Festival of 2020 (January 24, 2020) Rapid estimation of population numbers in counties and their rural areas and analysis of spatial characteristics, and two suggestions for epidemic prevention and control in Hubei Province: (1) It is recommended to increase the prevention and control of the epidemic in the second circle around Wuhan. Population control in rural areas in the second circle (Huanggang, Huangshi, Xiantao, Tianmen, Qianjiang, Suizhou, Xiaogan, Xiangyang, Jingmen and parts of Jingzhou). (2) Huanggang, Xiaogan, and Jingzhou among the prefecture-level cities in Hubei Province should pay special attention to rural areas in terms of epidemic prevention and control, and increase regional medical facilities and equipment to improve epidemic prevention capabilities.

Multi-source location-aware big data can also play a greater role in improving the ability and efficiency of epidemic prevention and control, such as the estimation of the number and flow of returnees in Hubei Province after the Spring Festival holiday, the scientific deployment of medical personnel and materials based on dynamic population distribution characteristics, Epidemic spread prediction that integrates dynamic crowd flow information, and fine-scale intra-regional epidemic risk estimation, etc. By combining expertise in different fields, such as the internal structure and functions of cities and towns, urban and rural and regional characteristics, infectious

disease models, etc., location-aware big data will be able to provide more in-depth and accurate epidemic analysis and prediction results.

References

- Chen, J., Pei, T., Shaw, S. L., et al. (2018). Fine-grained prediction of urban population using mobile phone location data. *International Journal of Geographical Information Science*, 32(9), 1770–1786.
- Chen, Y. H., Zhang, R. J., Ge, Y., et al. (2019). Downscaling census data for gridded population mapping with geographically weighted area-to-point regression kriging. *IEEE Access*, 7, 149132–149141.
- Deville, P., Linard, C., Martin, S., et al. (2014). Dynamic population mapping using mobile phone data. *Proceedings of the National Academy of Sciences*, 111(45), 15888–15893.
- Feng, J., Li, Y., Xu, F. L., et al. (2018). A Bimodal model to estimate dynamic metropolitan population by mobile phone data. *Sensors*, 18(10), 3431.
- Gong, P., Li, X. C., & Zhang, W. (2019). 40-Year (1978–2017) human settlement changes in China reflected by impervious surfaces from satellite remote sensing. *Science Bulletin*, 64(11), 756–763.
- Hu, M. (2019). Visualizing the largest annual human migration during the Spring Festival travel season in China. *Environment and Planning A: Economy and Space*, 51(8), 1618–1621.
- Hu, X. Q., Li, H., Bao, X. G. (2017). Urban population mobility patterns in Spring Festival transportation: Insights from Weibo data. In *2017 International Conference on Service Systems and Service Management* (pp. 1–6). IEEE.
- Khodabandehlou, G., Gauthier, V., El-Yacoubi, M., et al. (2016). Population estimation from mobile network traffic metadata. In *2016 IEEE 17th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WOWMOM)* (pp. 1–9). IEEE.
- Khodabandehlou, G., Gauthier, V., Fiore, M., et al. (2018). Estimation of static and dynamic urban populations with mobile network metadata. *IEEE Transactions on Mobile Computing*, 18(9), 2034–2047.
- Kontokosta, C. E., & Johnson, N. (2017). Urban phenology: Toward a real-time census of the city using Wi-Fi data. *Computers, Environment and Urban Systems*, 64, 144–153.
- Kubiček, P., Konečný, M., Stachoň, Z., et al. (2019). Population distribution modelling at fine spatio-temporal scale based on mobile phone data. *International Journal of Digital Earth*, 12(11), 1319–1340.
- Leyk, S., Gaughan, A. E., Adamo, S. B., et al. (2019). The spatial allocation of population: A review of large-scale gridded population data products and their fitness for use. *Earth System Science Data*, 11, 3.
- Li, J. W., Ye, Q. Q., Deng, X. K., et al. (2016). Spatial-temporal analysis on Spring Festival travel rush in China based on multisource big data. *Sustainability*, 8(11), 1184.
- Liu, Z., Ma, T., Du, Y., et al. (2018). Mapping hourly dynamics of urban population using trajectories reconstructed from mobile phone records. *Transactions in GIS*, 22(2), 494–513.
- Ma, Y. J., Xu, W., Zhao, X. J., et al. (2017). Modeling the hourly distribution of population at a high spatiotemporal resolution using subway smart card data: A case study in the central area of Beijing. *ISPRS International Journal of Geo-Information*, 6(5), 128.
- Ma, T., Lu, R., Zhao, N., et al. (2018). An estimate of rural exodus in China using location-aware data. *PLoS One*, 13(7).
- Patel, N. N., Stevens, F. R., Huang, Z. J., et al. (2017). Improving large area population mapping using geotweet densities. *Transactions in GIS*, 21(2), 317–331.
- People's Daily. *Mayor of Wuhan: About 5 million people left Wuhan*.

- Song, Y. Z., Long, Y., Wu, P., et al. (2018). Are all cities with similar urban form or not? Redefining cities with ubiquitous points of interest and evaluating them with indicators at city and block levels in China. *International Journal of Geographical Information Science*, 32(12), 2447–2476.
- Tsou, M. H., Zhang, H., Nara, A., et al. (2018). *Estimating hourly population distribution change at high spatiotemporal resolution in urban areas using geo-tagged tweets, land use data, and dasymetric maps*. arXiv:1810.06554.
- Wang, X. W., Liu, C., Mao, W. L., et al. (2014). *Tracing the largest seasonal migration on earth*. arXiv:1411.0983.
- Wang, Y. X., Dong, L., Liu, Y., et al. (2019). Migration patterns in China extracted from mobile positioning data. *Habitat International*, 86, 71–80.
- Wardrop, N. A., Jochem, W. C., Bird, T. J., et al. (2018). Spatially disaggregated population estimates in the absence of national population and housing census data. *Proceedings of the National Academy of Sciences*, 115(14), 3529–3537.
- Wei, Y., Song, W., Xiu, C. L., et al. (2018). The rich-club phenomenon of China's population flow network during the country's spring festival. *Applied Geography*, 96, 77–85.
- Xu, J., Li, A. Y., Li, D., et al. (2017). Difference of urban development in China from the perspective of passenger transport around Spring Festival. *Applied Geography*, 87, 85–96.
- Yao, Y., Liu, X. P., Li, X., et al. (2017). Mapping fine-scale population distributions at the building level by integrating multisource geospatial big data. *International Journal of Geographical Information Science*, 31(6), 1220–1244.
- Zhu, D., Huang, Z., Shi, L., et al. (2018). Inferring spatial interaction patterns from sequential snapshots of spatial distributions. *International Journal of Geographical Information Science*, 32(4), 783–805.
- Zong, Z., Feng, J., Liu, K., Shi, H., Li, Y. (2019). *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 33, pp. 1294–1301).

Part III

Analysis and Modeling Perspectives

Chapter 10

Identifying and Characterising Active Travel Corridors for London in Response to Covid-19 Using Shortest Path and Streetspace Analysis



Nicolas Palominos, Duncan A. Smith, and Sam Griffiths

10.1 Introduction

Worldwide, city authorities and transport agencies are implementing fast emergency streetspace reorganisation strategies in response to the Covid-19 pandemic (NACTO Covid-19 Transportation Response Center 2020). The Department for Transport in the UK launched a statutory guidance for network management in response to Covid-19 (Traffic Management Act 2004), which includes recommendations for reallocating road space to people walking and cycling and other measures to discourage the use of cars, such as e-scooter trials and a ‘fix your bike’ voucher scheme (Cycling and walking—GOV.UK 2020). The transport and public space management has at least two specific new requirements to prevent virus propagation. First, the movement of people has social-distancing restrictions resulting in more street area consumption, and second it is desirable to provide alternatives to mass public transport to avoid the general overcrowded environment of such systems. Both have triggered urban interventions resulting in the reallocation of streetspace for walking and cycling (Batty 2020).

Despite the efforts by city authorities, a problem they face is the limited comprehensive understanding of how streetspace is allocated citywide, how streetspace provision relates to streetspace demand and how these relate to variations in travel behaviour, weakening the understanding of the effectiveness of the streetspace safedistancing measures. In addition, because actual streetspace and street networks

N. Palominos (✉) · D. A. Smith

Centre for Advanced Spatial Analysis, UCL, First Floor 90 Tottenham Court Road, W1T 4TJ London, England

e-mail: n.palominos.16@ucl.ac.uk

D. A. Smith

e-mail: duncan.a.smith@ucl.ac.uk

S. Griffiths

The Bartlett School of Architecture, UCL, 3.18 22 Gordon Street, WC1H 0QB London, England

were mostly designed to satisfy the demands of car traffic, the alternative solutions proposed should be sufficiently attractive with equal levels of convenient connectivity and high environmental quality to effectively counterbalance car travel choice and produce a mode shift towards active travel. The picture gets further complicated by the effects of the extra streetspace demand at the neighbourhood scale for which streetspace reorganisation has also appeared as a solution (e.g. High Streets pavement-widening and low-traffic neighbourhoods). If street interventions fail it is likely that car usage will increase, accentuating well-known health impacts associated with air and noise pollution, traffic injuries and sedentary lifestyles.

Streetspace allocation is a key street design parameter especially relevant in circumstances where the available space is scarce and demand is high as it is expected when rail transport operates with less capacity (e.g. an estimated 15% for London). The analysis of streetspace can be informative in several ways; fine-grain and enhanced urban analytics for monitoring and understanding the impacts of streetspace interventions, and for the proposition of alternative scenarios that guide future interventions for sustainable healthy cities.

From multiple perspectives, there are significant arguments to reclaim streets from private cars and prioritise people in the design and planning of streets. Moreover, it is clear that travel patterns can substantially change through the changes in road capacity and streetspace allocation. Given the current demands for healthy transport modes, in this chapter, we investigate the potential streetspace re-allocations needed to create a micro-mobility network which prioritises space for active travel and public transport. Forthcoming sustainable transport policies provide the setting for the analysis. These are the expansion of the Ultra Low Emission Zone in London for 2021 (ULEZ 2021, see Fig. 10.1), and the target to have 80% of trips done by foot, cycle or public transport by 2041 (TfL Transport for London 2018).

Given the study area, we assume the intensity of usage of places from transport data with railway and underground stations conceptualised as 'activity nodes' (Alexander 1977). Then, we identify the critical pathways of connections between these places using shortest-path network analysis. Finally, we present a descriptive analysis of two optimal network scenarios applying the street metrics developed in previous work on quantifying streetspace in London [19].

The analysis is conducted by creating a pathways model connecting all railway and underground stations at the street level. The street segments contained by the pathways represent 30% of the total street length, and show a big variance of carrying load (or transport 'flow' as defined by Hollander (2016)). Moreover, a selection of 8 pathways has 38% of aggregated carrying load, although these correspond to just 2% of the total street length within the ULEZ 2021. Because the shortest-path calculations are a factor of network centrality analysis, some associations can be made with regards to the type of streets that constitute the network of pathways.

At a higher spatial resolution, the analysis of two reallocation of streetspace alternatives show that although the streetspace has considerable variations along pathways, there is sufficient space for vehicular and pedestrian uses to coexist. Also, the impacts of narrowing carriageways are relatively more beneficial to footway space than disadvantageous for carriageway space.

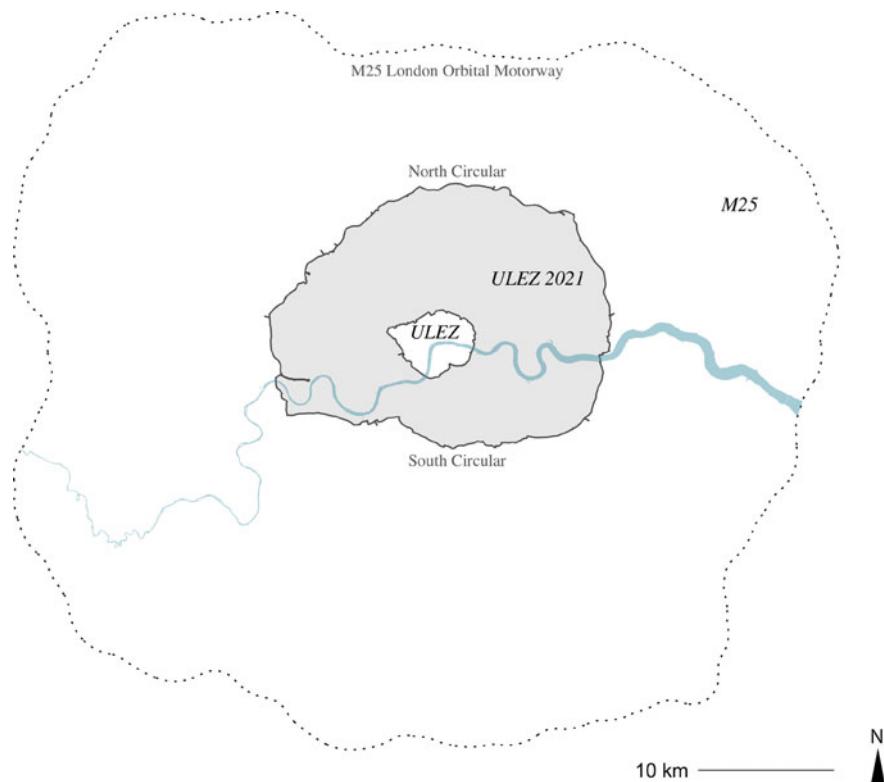


Fig. 10.1 Boundaries of Ultra Low Emission Zones and M25 zone organised as concentric rings

The following sections begin with an overview of relevant indicators related to streets and street usage. Then follows a brief descriptive analysis of the street network in the study area and the nodes definition. Next, we present the methodology for generating a micro-mobility network and analyse the results. At last, we conclude with a summary and a discussion of the key findings.

10.2 Street Use and Transport General Trends and Facts

This section includes a review of key indicators related to the street ecosystem and general transport indicators for London. Commonly these indicators are presented in reports prepared by the metropolitan transport authority Transport for London (TfL) and other think tanks specialised in urban issues. The indicators cover a wide range of domains from the built environment to air pollution and are presented without any particular organisation as most of them relate to two or more domains (transport, health, economy, etc.).

While Covid-19-related street measures are noteworthy in number and extent, the design and planning of streets have already been shifting from car-oriented to people-oriented towards more sustainable cities. Some initiatives are based on counter balancing the negative impacts of private car use and promoting a modal split change (Gössling 2020), others go beyond the transport focus arguing that streets are multifunctional urban entities (Institute for Architecture and Urban Studies 1978; Marshall et al. 2018) and suggest that streets should be considered as drivers of urban prosperity (Mboup 2013).

In general, street network studies reduce the complexity of the space of the street by using a linear representation to facilitate network-based structural investigations of the street systems (Marshall et al. 2018). For example, spatial configuration analysis is a well-known approach in urban morphology for the study of street patterns using road centre line street representations. Findings of this approach include important associations between configurational metrics and street social and economic activity (Porta et al. 2009, 2012, Hillier and Iida 2005), among others. Nevertheless, the analysis of physical metrics that are fundamental attributes impacting the way a street functions, such as the footway and carriageway widths, are often overlooked. The focus on streetspace allocation in combination with street level connectivity presented here is a concrete contribution not only for expanding street network studies and the insights these can bring into street planning and design but also for other realms of sustainable urban design noted as relevant at multiple scales from local initiatives to the country-level industrial strategy. Moreover, the methods presented here can offer valuable analytical capacity for developing new cycling and walking schemes and designing place-based streets that are more appropriate to control virus propagation and promote a green postpandemic recovery. A key aspect for promoting active travel is to make this kind of trips more convenient, attractive and less costly. We suggest that identifying the most direct and shortest routes for active travel prioritisation can bring important benefits for urban transport without disrupting bus services but on the contrary creating exclusive circulation corridors.

Shortest-paths analysis has been studied elsewhere to highlight the tree-like structure of transport modes (Allen 2018) and to model route choice behaviour of ride-hailing services (Manley et al. 2015). Nevertheless, the analysis presented here has the purpose to examine and prescribe new street morphologies for future urban mobility. This considers assuming a street network in which active travel modes are prioritised with direct and fast routes at the expense of modes that are less efficient and have a greater environmental impact such as the car. This proposal of promoting walking, cycling and micro-mobility aims to absorb the trips of public transport services that are operating with less capacity due to social distancing restrictions, while at the same time controls an increase in car travel.

The streets of London carry the majority of the daily trips which are mainly done by active and sustainable modes (walking, cycling and public transport). In 2018 this accounted for 63% of trips. This modal share would require a 0.7 annual increase approximately to reach the Mayor's 80% target by 2041. In more detail, the period between 2000 and 2018 shows a decline of private transport from 48 to 37%, a small increase of 1% in walking starting at 24% and a bigger increase of 9% in public

transport starting at 27%. The 36% of public transport share is composed by 22 and 14% of the trips done by rail/underground and bus respectively (TfL Transport for London 2018). This is relevant from a street environment perspective as all public transport trips typically include a short walk at the beginning or end of trips. As an illustration, for the calculation of the Public Transport Access Level (PTAL) a value of 12-minutes walk is used by TfL. However, the streets of London are not only crucial for transport purposes but also have an important role as the main public space of the city occupying an estimated 80% of the total surface of public space (TfL Transport for London 2017).

From the economic perspective, a number of reports for London suggest greater value of pedestrian-oriented streets. For example, it has been demonstrated that pedestrians spend 65% more than drivers on average per month. Moreover, improvements on the street environment, including pavement widening, add significant value to private property (Sadik-Khan and Solomonow 2017, NewLondon Architecture 2016). This can be explained by the negative impacts of motorised traffic on the environmental quality of streets. For example, in London, motorised road transport is a relevant contributor to air pollution generating emissions that are harmful to human health (14% of nitrogen oxides and 56% of particulate matter less than 2.5 microns in diameter) (GLA Greater London Authority 2018).

As an illustration for modal shift towards sustainable transport, policy guidelines enumerate several negative impacts of high levels of heavy traffic; air pollution, loss of urban public space, accidents, severance, noise and vibration and economic inefficiency and loss of competitiveness of central areas, among others (European Commission. Environment Directorate-General 2004; Barrett et al. 2017). In contrast, it has been reported that place-based street improvements provide considerable value not only to street users but also to surrounding businesses (Carmona et al. 2018, Sadik-Khan and Solomonow 2017). Notably, associations have been found between the quality of walking amenities and the performance of innovation districts, suggesting that face-to-face contacts enables innovation (Zandiatashbar and Hamidi 2018). Additionally, it has been implied that, despite the trivial they might be, the sum of many little contacts between pedestrians form the trust of a city (Jacobs 2016) and that pedestrian streets can provide the place for people to rub shoulders which is an “essential social ‘glue’ in society” (Alexander 1977). In the United Kingdom, one of the key areas of the Government’s Industrial Strategy, Future Mobility, was reformulated with emphasis on the role of urban design and planning and the need to develop new street design standards to optimise sustainable and low environmental impact travel systems (Mazzucato and Willetts 2019).

Overall, it is possible to observe in London a trend of car use decline and greater awareness of the social, economic and environmental benefits of pedestrian-oriented street environments. This situation has already been recognised by public authorities and research groups in London (Barrett et al. 2017, GLA Greater London Authority 2018; New London Architecture 2019; Council 2015; City of London Corporation 2019).

10.3 Streets and Nodes of the Ultra Low Emission Zones

In this section we analyse the streetspace designation metrics of the current 2020 and proposed Ultra Low Emission Zones (ULEZ), and introduce a definition of ‘activity nodes’ within the ULEZ2021. Both ultra low emission zones are graphically represented in Fig. 10.1. The current 2020 ULEZ corresponds with the Congestion Charge zone (Euston Rd., City Rd., Tower Bridge Rd., Kennington Ln., Vauxhall Bridge Rd. and Park Ln.), and the 2021 expansion is defined by the North and South Circular roads. The charts in Fig. 10.2 show the central tendency of streetspace allocation measures for the M25, ULEZ 2021 and ULEZ 2020 zones, which are quite revealing in several ways. First, it can be seen a very regular pattern in the relation between zones for all streetspace designation measures. Second, that the ULEZ has the highest values for all streetspace metrics and the M25 zone has the lowest with a striking exception for footways where the ULEZ 2021 has the lowest values. Finally, the chart shows a decline trend of total streetspace from centre to periphery and the overall predominance of carriageway streetspace over footway streetspace across all zones, which is consistent with the description presented in Palominos and Smith (2019), that was conducted using a different approach.

For the purpose of defining ‘activity nodes’ within the ULEZ 2021, we make the assumption that the surroundings of railway and underground stations have the potential of concentrating public life, activities and community facilities that mutually support each other (‘activity nodes’ is proposed as a pattern by Alexander (1977)). The surroundings of stations located in the inner city already have plenty of amenities and attract an important number of people. In like manner, stations with less demand have the potential to do so by the strategic densification around stations that accommodates city growth with a sustainable approach (TfL Transport for London 2019). As some authors suggest this approach has been successfully applied worldwide and

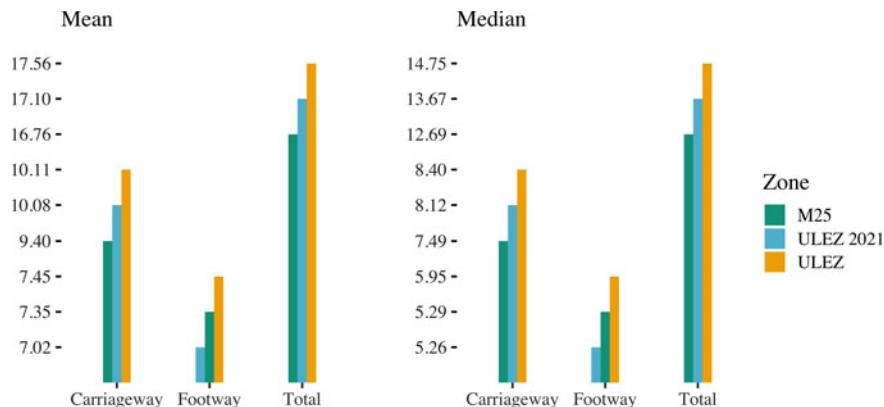


Fig. 10.2 Central tendency comparison of streetspace metrics between the zones within the M25 Orbital, the current 2020 ULEZ (Congestion Charge Zone) and the planned ULEZ 2021

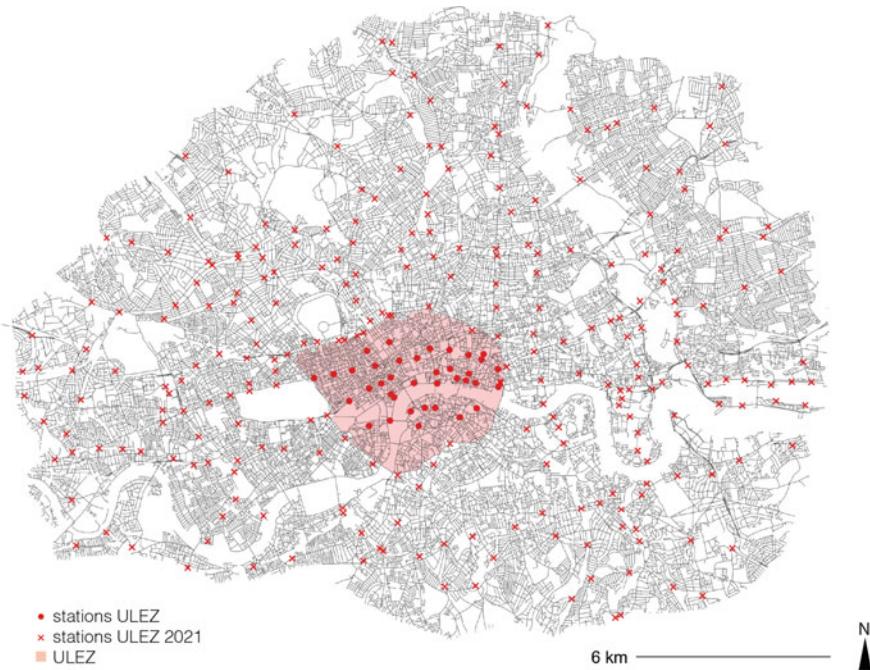


Fig. 10.3 Railway and underground stations in the ULEZ 2021

is referred to as Transit Oriented Development (TOD) (Ibraeva et al. 2020). With this in mind the ULEZ 2021 has plenty of potential 'activity nodes'. Figure 10.3 shows the dispersion of stations in the current 2020 and proposed ULEZ, which number increases with the ULEZ extension in 266 stations from 37 to 303.

10.4 A Street Level Micro-Mobility Network

Having defined what is meant by 'activity nodes', we will now move on to present the methodology for generating a street level micro-mobility network. Micro-mobility comprises a smaller kind of urban mobility in two forms: the size of the vehicle and the trip range. This type of mobility has disseminated in many cities worldwide supported by platform technologies that allow a convenient 'as-needed' flexible transport solution, usually for short and medium distance trips. The most common types of vehicles are e-scooters, dockless bikes and station-based bikes (such as the cycle hire scheme operating in London since 2010), which have a small physical footprint and weight, although they have a limited passenger capacity. Still, micromobility has the potential for both increasing the access and adding options to public transport (as

a first mile/last mile solution), and for replacing short-distance car trips (Tice 2019; National League of Cities 2019).

Before proceeding to define the micro-mobility network, it is important to discuss the implications of prioritising streetspace for active travel and public transport framed under a people-oriented street design approach. In the context of scarce physical streetspace, the incorporation of a new type of vehicle intensifies the existing competing demands for urban space. For example, micro-mobility vehicles occupy extra streetspace for both parking and circulating. Parking has been the focus of attention of public authorities to solve the additional streetspace clutter that these vehicles generate when parked inappropriately (e.g. the designation of parking areas for dockless bikes). Circulating is not yet fully admitted (e.g. electric scooters in the UK), nevertheless micro-mobility vehicles have greater competitive advantages because they are more space-efficient than private cars (see Table 10.1). In addition, it could be argued that with an adequate management, micro-mobility vehicles can allow greater social contact and community connections.

As previously stated, micro-mobility, active travel and public transport are different types of transport solutions that have similar objectives. In addition, we have defined activity nodes that are actual and potential attractive destinations. Assuming that the intensity and diversity of activities of the nodes is complex to define and a matter in constant evolution and adaptation, the micro-mobility network is created by connecting all nodes through the shortest-paths, in order to create the conditions for an integrated public transport system that optimises travel distance using the actual street infrastructure (see Fig. 10.4). Essentially, this network provides the convenient and desirable conditions for short, medium and long trips. Also, it can potentially enable multi-modal integration and maximise the efficiency of streetspace in concordance with sustainable urban goals. It would be expected that the paths which are most intensively used could gradually turn into ‘promenades’ of mixed-use activity such that the remaining in-between areas are at short distance from lively and vibrant streets and centres (‘promenade’ is proposed as a pattern by Alexander (1977)).

From the design perspective it is important to highlight that the shortest-path type of structure for transportation purposes is far from being the optimal for construction

Table 10.1 Per-person travel space requirements for different modes

unit	Standing/Parked		Speed		Travelling	
	sqm	sqf	kph	mph	sqm	sqf
Pedestrian	0.5	5	5	3	2	20
Micro-vehicle	1	11	25	16	7	70
Bicycle	2	20	16	10	5	50
Bus Passenger	2	20	48	30	7	75
Automobile	37	400	48	30	139	139

Source National League of Cities (2019), Tice (2019)

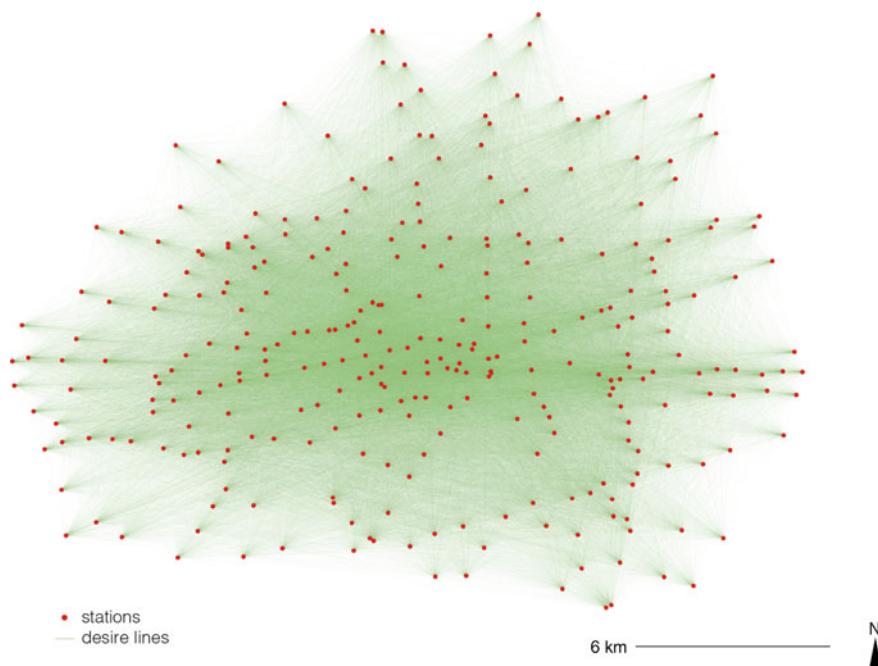


Fig. 10.4 Complete graph connecting stations: node-to-node shortest-paths ($n = 45,753$)

(see discussion from a network perspective in Barthélemy (2011)). Yet, the alternative proposed here adopts such structure to prioritise the convenience of users by optimising travel distance and at the same time optimises construction by utilising pre-existing infrastructure.

The map in Fig. 10.5 shows the structure of the micro-mobility network highlighting the street segments that concentrate the greater number of through-routes and the travel pathways these form. It is possible to observe a cellular/radial spatial pattern that coincides with some the actual cycling infrastructure. Fig. 10.6i shows the current cycle lanes which are fully or partially segregated, on-carriageways or shared lanes (e.g. bus lanes). Some of these cycle lanes are also part of the designated cycle routes, which are part of executed, ongoing and future investments. Overall, however, this visual comparison shows that the connectivity and complexity of the modelled network are much greater than the observed reality of the cycling network.

Having discussed the methodology to construct the network, the next sections address the descriptive analysis of the network and travel pathways and the potential streetspace re-allocations needed to create a micro-mobility network.

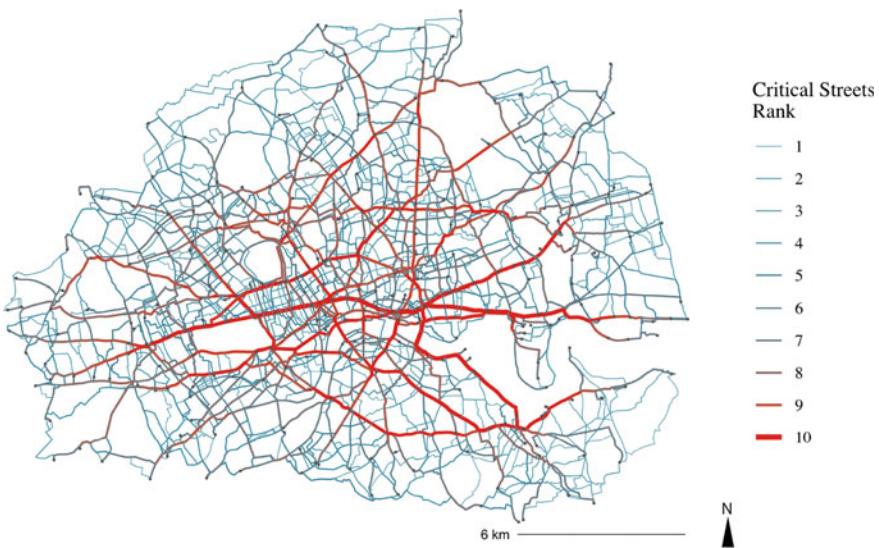


Fig. 10.5 Critical streets ranked according to shortest-paths through-routes

10.4.1 Cycling Infrastructure and Missing Links

The growth of cycling trips in recent years has been accompanied by investment in new and upgraded infrastructure. In this section we analyse with more detail the cycling infrastructure illustrated in Fig. 10.6*i*, and we compare it with the core critical streets network in Fig. 10.5. As can be seen, there is an overlap between the critical streets with higher rank and the provision of cycling infrastructure (e.g. some of the radials like Kingsland Rd, Edgware Rd and the Victoria Embankment). To identify the core critical streets we selected the segments at the highest 20% of through routes, which corresponds to 4.513 street segments with values from 490 up to 7031 traversing shortest-paths (see Fig. 10.6*ii*).

The 5 cycle lane types in Fig. 10.6*i* provide a general description of the quality of the cycle lanes that occupy part of the street. The cycle lane types were generated from the cycling infrastructure documentation published by TfL (see database schema at

Table 10.2 describe the attributes of 3 cycle lane types and Fig. 10.7 shows real-world examples. The 3 cycle lane types information was joined to the streetspace road centre line representation (RCL) resulting in street segments with mixed cycle lane types (2 additional). Because physically segregated lanes could be considered of the highest standard (type *a*), the mixed cycle lane types were defined with reference to these.

From Fig. 10.6*i* and *ii* and the breakdown in Table 10.3 it is possible to observe the main characteristics of the existing cycling lanes. Near 64% of the cycling street network only has indicative cycling infrastructure (Signings and markings (*b*) and Shared on carriageway with no priority (*c*)). Although, the spatial pattern described

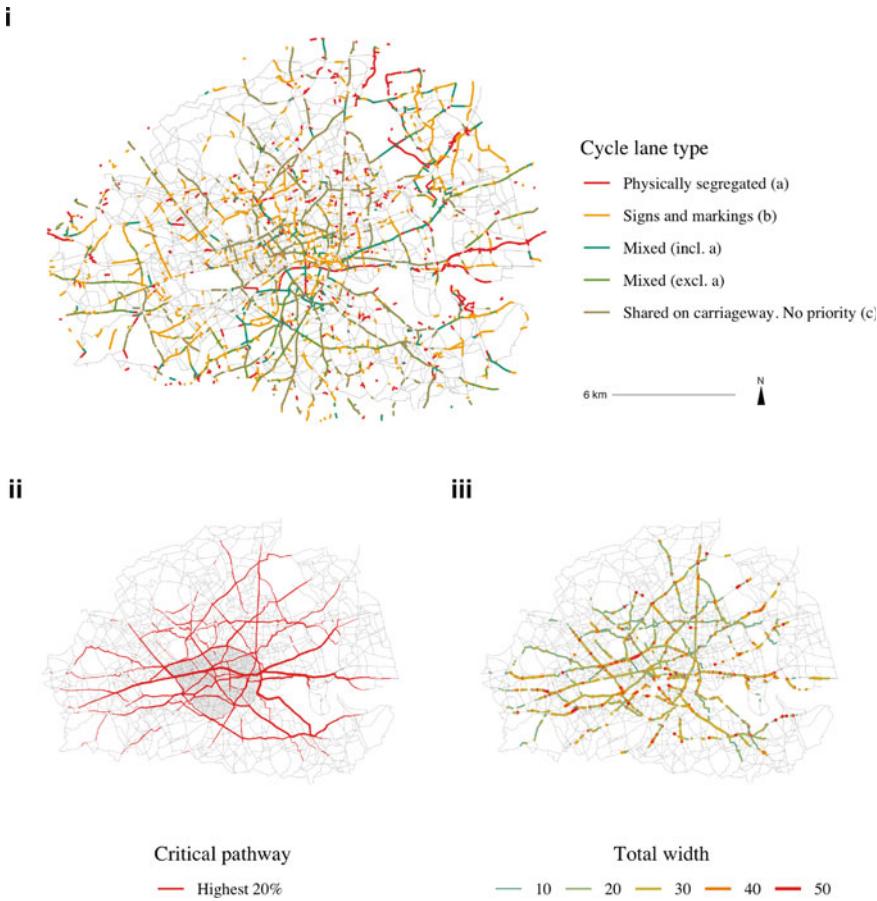


Fig. 10.6 Cycling infrastructure and missing links: (i) Cycle lane types, (ii) Core of critical streets, (iii) Difference between *ii* and *i* classified by total street width (*i* is excluding *c*). See figures in Table 10.3. Data source (i) <https://cycling.data.tfl.gov.uk>

Table 10.2 Description of cycle lane types from TfL survey

Cycle lane type	Physically segregated (a)	Signs and markings (b)	Shared on carriageway. No priority (c)
TRUE	Segregated Partially Seg Stepped	Mandatory Advisory Priority	Shared On carriageway
FALSE	On park On waterside		Priority

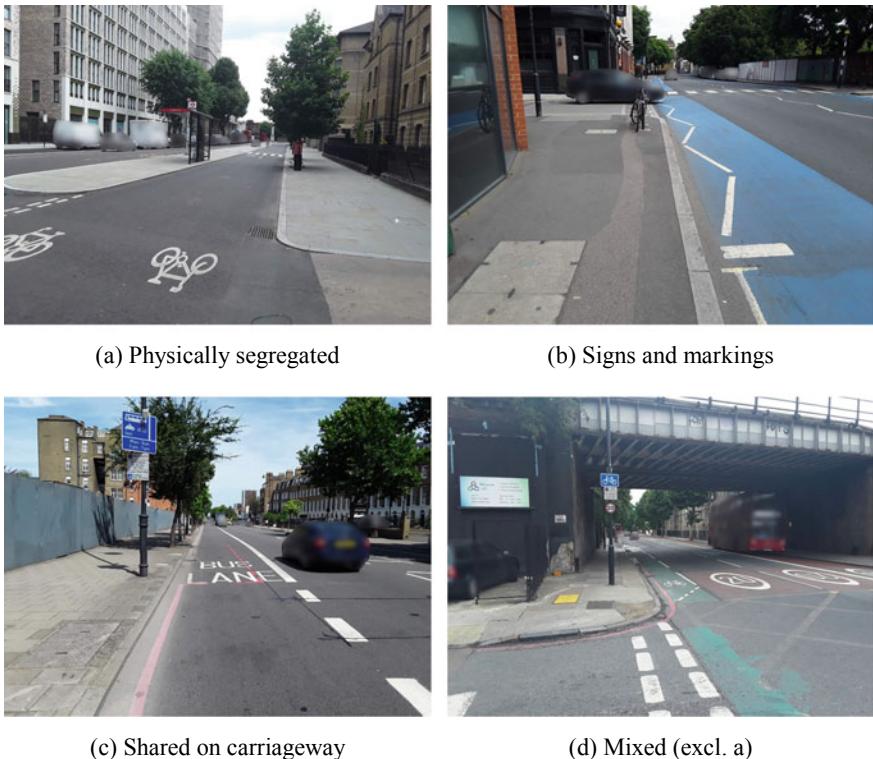


Fig. 10.7 Examples of cycle lane types. *Images source* <https://cycling.data.tfl.gov.uk/>

Table 10.3 Summary of Fig. 10.6

	Physically segregated (a)	Signs and markings (b)	Mixed (incl. a) (excl. a)	Mixed No priority (c)	Shared on carriageway. length	Total length
Cycle lanes (km)	85.60	177.70	47.10	33.30	117.20	460.90
<i>pc</i>	0.19	0.39	0.10	0.07	0.25	1.00

$$\frac{\text{Total length}}{\text{Critical pathway (km)}} = \frac{460.90}{261.3} = 1.77$$

total width breaks (m)	(0,10]	(10,20]	(20,30]	(30,40]	(40,50]	Total length
Missing connections (km)	2.5	76.3	68.0	31.5	5.7	184.0
<i>pc</i>	0.01	0.41	0.37	0.17	0.03	1.00

by them is of continuous lines similar to some of the critical pathways in Fig. 10.6ii. The same could be observed from the physically segregated cycle lanes (type *a*), however, these are both shorter (29%) and scattered with some continuity in Central London (Victoria Embankment) and the radial connections towards the East, plus isolated radials in the north-eastern part of the study area.

The missing links illustrated in Fig. 10.6iii are obtained from the difference between the core critical pathways and the existing cycling lanes (excluding Shared on carriageway with no priority (*c*)). It is possible to observe that an important part of the critical pathways network would need to be built or enabled to create the continuous paths. The visualisation shows the total street width of the missing connections to estimate the easiness to fit in formally designated road space for cycling or micro-mobility. Additionally, it is possible to observe the stretch of critical streets without any cycle lane infrastructure but that have the potential to create a continuous route. The general piecemeal pattern might be the result of street retrofitting investment strategies, which despite the existence of a general plan of designated cycle routes, lacks the adequate infrastructure and continuity of a purposefully constructed cycling/micro-mobility network. Nevertheless, the construction of a continuous network has the constraints imposed by the scarcity of streetspace represented by the total street width available. The bottom of Table 10.3 shows the breakdown according to total street width illustrating that most streets segments needed to complete the critical core are relatively narrow (10–20 m), reflecting the challenges of streetspace reallocation.

10.4.2 General Patterns of the Street-Level Travel Pathways

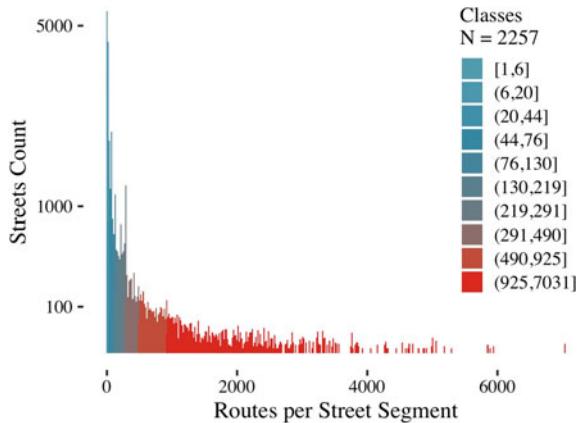
We have argued that the design approach for the construction of the network optimises the utilisation of the existing streets. The figures in Table 10.4 show that the total street length of the ULEZ 2021 is around 4,784 km., and that 30% of the total street length is needed to create the network. However, the street segments have a varying carrying load represented by the total number of through-routes along them (see Fig. 10.5). This can be observed in the distribution of routes per street segment which shows a left skewed distribution. This means that there are many streets that are part of few travel routes and a few critical street segments that are traversed by many travel routes (see Fig. 10.8).

The carrying load metric for each street segment is useful to measure the relative importance of the different pathways. To compare pathways we define a rate of critical pathway importance P by adding the carrying load of all street segments in

Table 10.4 ULEZ 2021 and Network total street length

	Km	Perc
Total ULEZ 2021	4784	100
Network	1434	30

Fig. 10.8 Street segments-pathways distribution



the pathway S and dividing it by the total number of street segments or edges E in the pathway to control by pathway length.

$$P_{ij} = \sum \frac{S_{ij}}{E_{ij}} \quad (10.1)$$

Table 10.5 presents the results obtained from the descriptive statistical analysis of the pathways. Because of the location and spread of the activity nodes, the pathway length and total number of segments (E) show a considerable range with maximum values more than 200 bigger than minimum values. Similarly, the carrying load (S) has maximum values 8000 times bigger than minimum values, which is useful to identify important travel routes as this measure could be assimilated with a measure of aggregated betweenness centrality. Also, this can favour strategic approaches for selecting significant pathways for intervention, which would have a greater impact for the whole network. The 10.7 km mean value of pathway length reveals that under appropriate street-infrastructure conditions a one-way average commute could take 25 min. on an electric micro-vehicle and 35 min. riding a bike.

The critical pathway metric is better understood looking at the spatial pattern. Fig. 10.9b show the pathway with maximum critical pathway value P , at the city centre (Cornhill, Leadenhall St, Aldgate High St.), representing the thoroughfare with the highest density of traversing routes. The pathways in colour red on Fig. 10.9d

Table 10.5 Pathways metrics summary

Variable	Min	Q1	Median	Q3	Max	Mean
Carrying Load (S)	145	73308	181291.5	371886.8	1213439	255723.3
Number of segments	3	113	177	249	524	185.2
Critical Pathway (P)	5.6	613.9	1046.2	1680.4	4771.5	1226.5
Pathway length in km	0.1	6.6	10.2	14.3	28.7	107

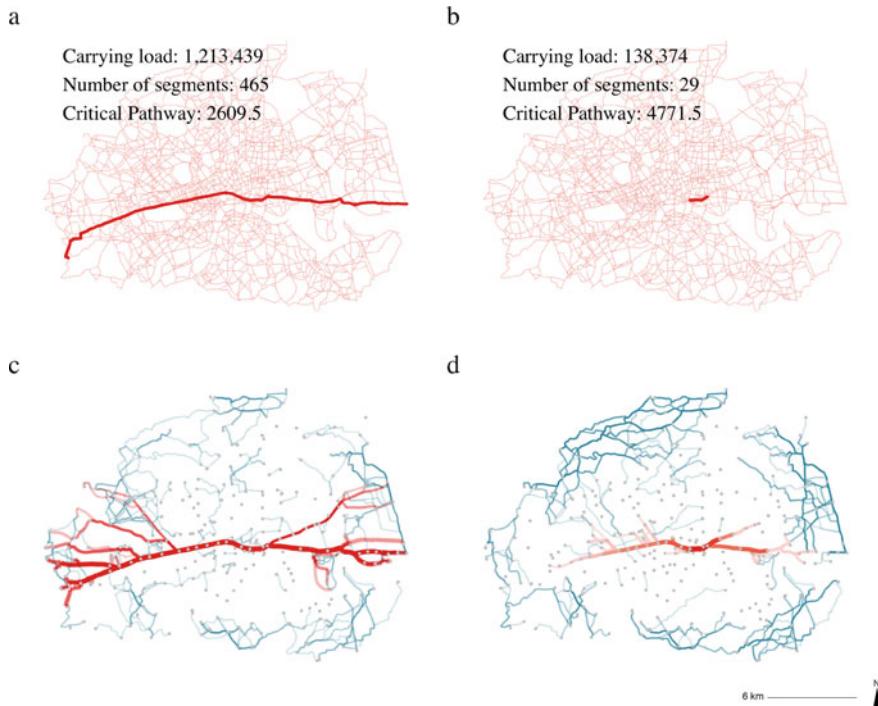


Fig. 10.9 Spatial pattern of pathways carrying load S and critical pathway P . Panels a and b are the maximum values of S and P , and panels c and d show the top (red) and bottom (blue) 1% values of S and P . Stations represented in light grey

Correspond to the highest 1% values of P . Interestingly, along the West extension there appear a series of branches towards the North which correspond with relatively short pathways adjacent to the main East-West thoroughfare, which get high P because of their proximity to the pathway with highest P (Southampton Row, Gower St., Tottenham Court Rd and Cleveland St.). Fig. 10.9c shows a pattern of longer branches in red representing the highest 1% values of C , which can be defined as long and high-density pathways connected to the centre. For Fig. 10.9c and d, the pathways in blue are the lowest 1% values, where it is possible to identify peripheral and few central pathways. Overall, it stands out the total lack of important pathways South of the river. This result is somewhat counter-intuitive, because some of the pathways in the South also traverse the main thoroughfare (see Fig. 10.5), yet longer routes are needed and also the concentration of nodes is much bigger in the North (see for example the sequence of stations in light grey in Fig. 10.9c), therefore the overall carrying load is greater (e.g. S value of all top 1% pathways in Fig. 10.9c is over 1 million).

10.4.3 Selection of Critical Pathways that Conform the Micro-Mobility Network

A number of strategies could be adopted to select the pathways out of the more than 40 thousands. For example, a balance between Northern and Southern areas of the city could be desirable, or a focus on areas with higher potential urban growth. It is clear from the pathway analysis that the intervention on the East-West pathway in Fig. 10.9 would represent an impact for an important number of pathways. However, the geographical balance between different areas within the ULEZ 2021 is missing from this analysis.

In order to identify pathways that could complement the East-West main thoroughfare, we decomposed the critical streets map in Fig. 10.5 into 20 groups classified by the rank of their streets segments frequency and presented cumulatively (see Fig. 10.10).

This sequence resembles a pattern of urban growth where by comparative analysis it is possible to identify the formation of critical pathways. For the reasons we discussed above the first pathway to be generated is the East-West thoroughfare. Then, a branch to the South-East follows and a bifurcation of the main thoroughfare in diagonal in a North-East direction (from panel 2 onward). Also, on the first panels it is possible to observe the formation of a pathway in diagonal towards the North-West (represented with more clarity in panel 4; for reference this corresponds to Edgware Road). In panel 16 a dozen of pathways form a network with an extensive geographic coverage. From these, for the sake of simplicity, we selected the 8 pathways represented in Fig. 10.11 which contain the pathways with the maximum S and P , and most of the pathways highlighted in the rank visualisation which are connected together and form a network.

The criteria for pathways selection is to avoid overlapping between paths while at the same time to connect stations sufficiently separated so that the 8 pathways network has a considerable geographic coverage of the ULEZ 2021. Table 10.6 shows a summary of the pathway metrics for the 8 pathways network. The metrics can be compared with the summary of the whole network in Table 10.5. As can be seen, most of the P values are close to the whole network mean (1226,5), although there are 2 notable exceptions which correspond to pathways with a South-North direction and a trajectory that crosses rather than overlaps with the centre (Westferry-South Tottenham and Clapham Common-White Hart Lane). Importantly, the S values are mostly above the general median (181291.5) except for Westferry-South Tottenham. The total street length of the 8 pathways network is 96 km approximately, many times smaller than the whole network (1434 km), yet it concentrates near 38% of the aggregated carrying load.

A closer inspection of the pathways characteristics can be done by looking at the streetspace designation metrics. Table 10.7 presents an overview of total street, footway and carriageway widths. Overall, figures show that on average the selected pathways have considerable designated footways and carriageway streetspace. However, because of the piecemeal street improvements across the

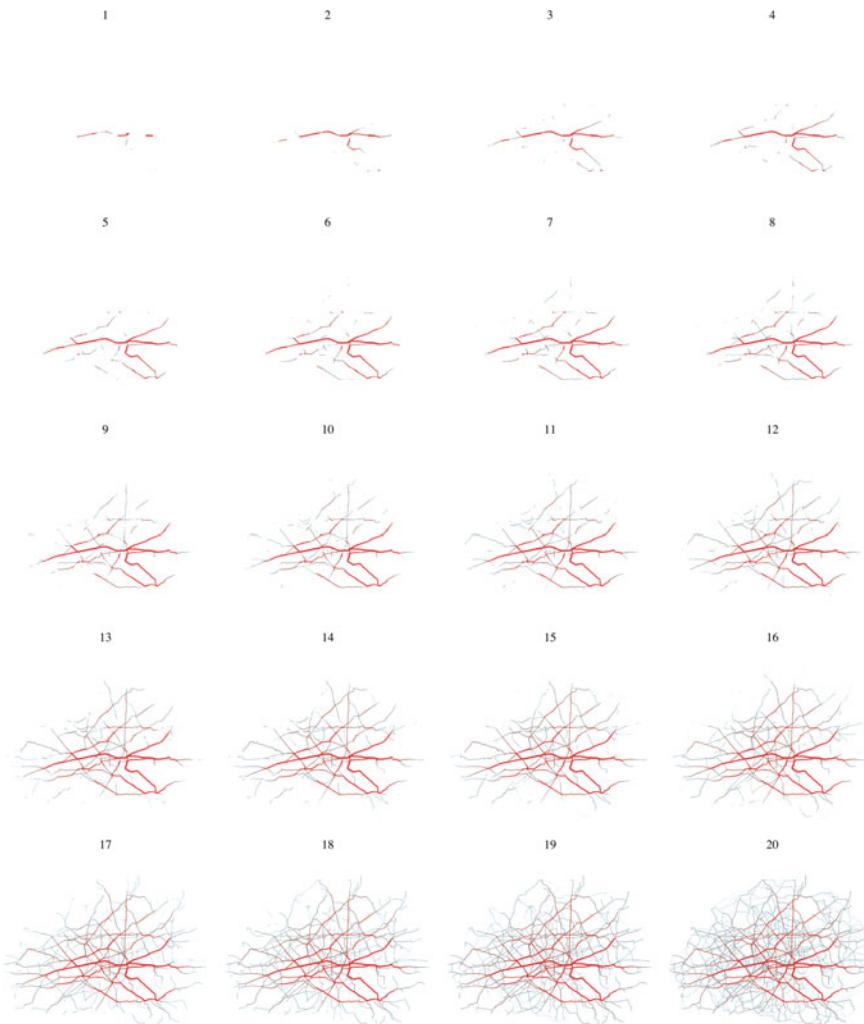


Fig. 10.10 Cumulative critical streets sequence according to street carrying load

history of London, it is possible to anticipate that there will be important variations in the streetspace designations metrics along the pathways.

Before the quantitative examination of possible scenarios to re-allocate street space, it is worth visualising what do we mean by this. We assumed that the prioritisation of active travel and public transport implies reducing carriageway space to a minimum width that allow the circulation of buses. The carriageway width is assigned for two scenarios: the demarcation of one and two lanes (3.5 and 7 m width respectively).

Fig. 10.12 illustrates the variations of the streetspace designation metrics for the two scenarios in a sample of 60 street segments from the Aldgate East-Bow Church

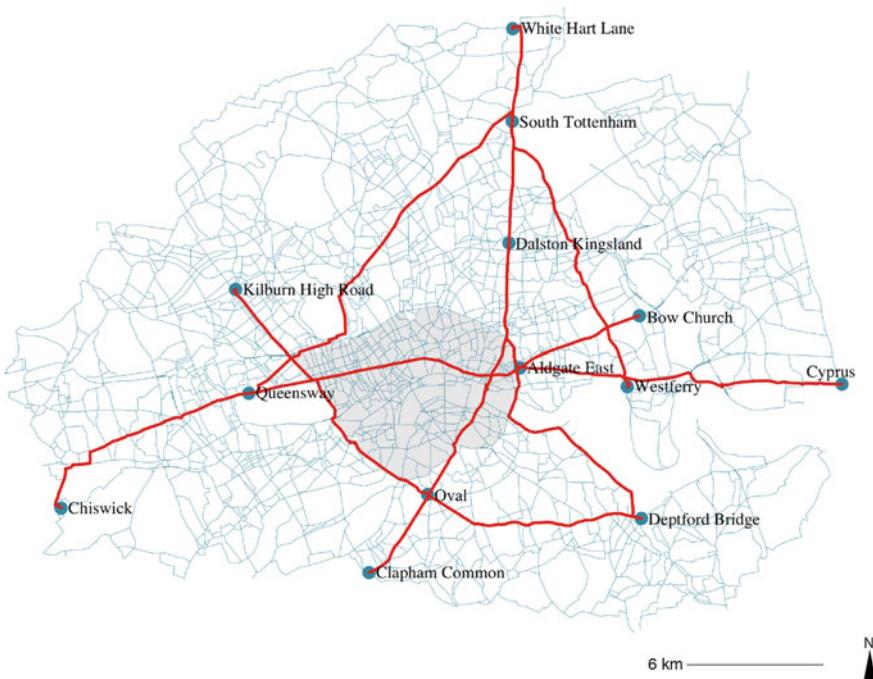


Fig. 10.11 Network of 8 selected pathways

Table 10.6 Eight pathways network metrics summary

From	To	Total length (m)	S	E	P
Chiswick	Cyprus	25,696	1,214,066	467	2600
Deptford Bridge	Dalston Kingsland	10,892	435,321	195	2232
Clapham Common	White Hart Lane	17,934	330,986	350	946
Queensway	South Tottenham	12,801	239,540	203	1180
Kilburn High Road	Oval	8890	225,359	161	1400
Aldgate East	Bow Church	3854	211,983	77	2753
Oval	Deptford Bridge	6695	195,528	119	1643
Westferry	South Tottenham	9181	122,143	176	694

pathway. From the chart, it can be seen that by designating a fixed carriageway it is possible to uniform the otherwise chaotic sequence of street cross-sections. Accordingly, this allows to identify places along the pathway with greater potential or challenge for active travel prioritisation. While the indication of designated town centres provides clues to estimate streetspace demand, the addition of other variables that reflect the complexity of street usage would enrich this analysis (e.g. street markets, bike stations, bus stops, etc). Nevertheless, the multiple dissection of the

Table 10.7 Eight pathways streetspace designation metrics summary

From to	Total street width mean	Footway width mean	Carriageway width mean
Aldgate East Bow Church	30.2	14.4	15.7
Clapham Common White Hart Lane	26.8	11.0	15.8
Oval Deptford Bridge	24.8	9.6	15.1
Chiswick Cypress	27.8	11.2	16.6
Kilburn High Road Oval	29.3	11.1	18.3
QueenswaySouth Tottenham	25.0	10.4	14.6
WestferrySouth Tottenham	24.7	9.6	15.1
Deptford Bridge Dalston Kingsland	25.4	10.1	15.3

pathway serves as a baseline that presents key geometrical information of the streets environments that compose the pathway.

The 8 pathways network scenarios are summarised in Table 10.8. To measure the variance between the actual, two lanes and one lane scenarios, we calculated a total approximate area by multiplying actual and proposed street widths by the street length. What is interesting about the data in this table is that the carriageway variance for both proposed scenarios are very similar, reflecting the modular nature of street design with regards to vehicular space. The opposite is true for pedestrian space which is the 'left-over' space after the carriageway has been determined. Also, since the positive and negative values can be seen as gains and losses of space, the figures reflect that the proposed scenarios entail a gain of 73% and 109% footway space on average for the two lanes and one lane scenario respectively, and a loss of 51% and 75% of carriageway space. These last measures are consistent with the figures in Table 10.7 that show a carriageway total mean of 15 m approximately which could fit 4 lanes.

Overall, because footways widths are smaller than carriageway widths, on the whole the relative gains surpass the losses. In other words, on streets such as the ones forming the selected pathways, changes of the streetspace designation metrics like the studied here, can have a relatively greater impact for the footway space than for the carriageway space.

10.5 Summary and discussion

In this chapter we have presented a method to analyse the street network of central London and proposed alternative scenarios for the conformation of a micro-mobility network that prioritises active travel and public transport. The method covers from the definition of a structure of routes to a fine-grain characterisation of street segments, and is inspired by well-known strategies of rail infrastructure optimisation (e.g. use oriented and construction-oriented design). This considers proposing a completely

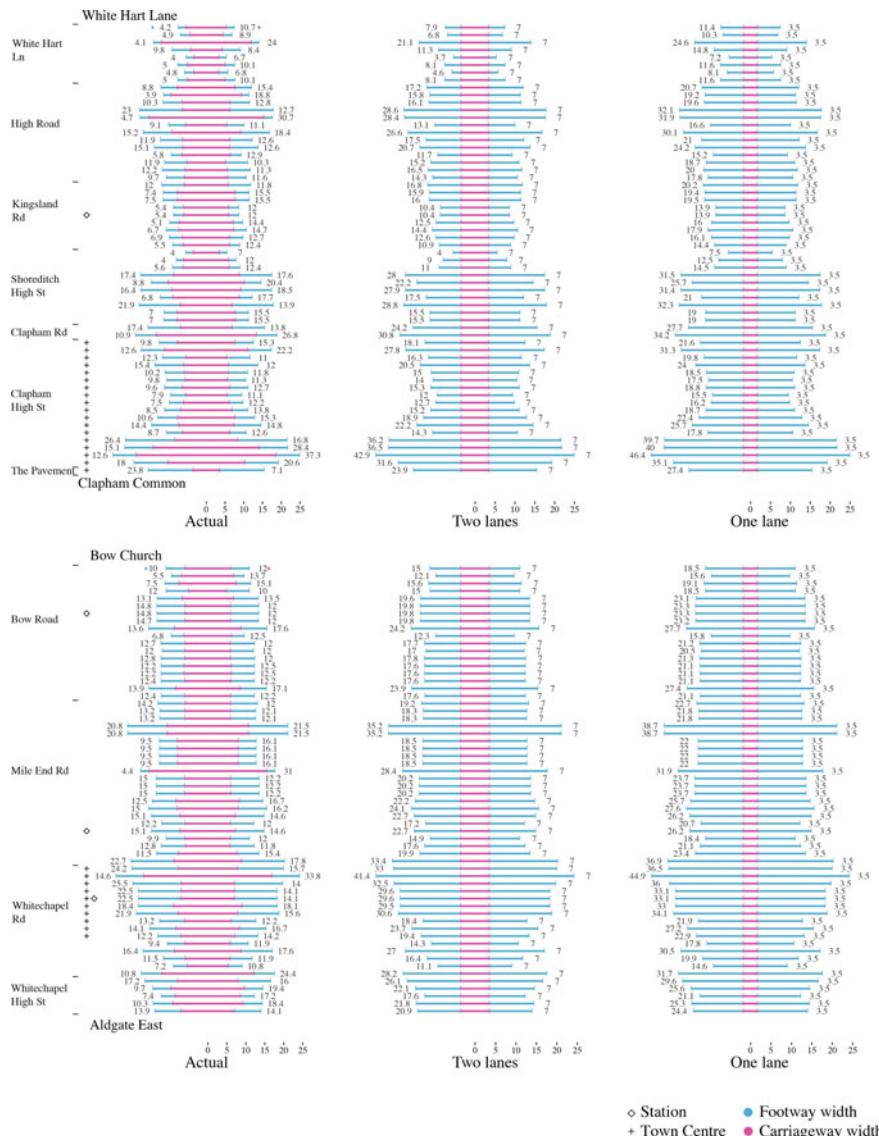


Fig. 10.12 Pathway anatomy: Multiple dissection of selected pathways and micromobility scenarios (sample of 20 segments at the start, middle and end of paths)

connected active travel network in which direct and fast routes are prioritised for active travel modes at the expense of modes that are less efficient and have a greater environmental impact. However, because the construction of a network from an ordinary number of points results in a high number of connections (over 40 thousands in this case), the problem of pathways selection arises. The analysis could be

Table 10.8 Summary of pathways two lane and one lane scenarios

Pathway	Actual total app. area (sqm)			Two lanes			One lane			
	Footway (fo)	Carriageway (ca)	fo	var	ca	var	fo	var	ca	var
Ald_Bow	55,160	55,829	84,008	0.52	26,981	-0.52	97,498	0.77	13,490	-0.76
Cla_Whi	184,071	266,856	325,389	0.77	125,538	-0.53	388,158	1.11	62,769	-0.76
Ova_Dep	58,183	89,514	100,829	0.73	46,868	-0.48	124,263	1.14	23,434	-0.74
Chi_Cyp	264,531	366,850	451,512	0.71	179,868	-0.51	541,447	1.05	89,934	-0.75
Kil_Ova	89,563	143,317	170,651	0.91	62,228	-0.57	201,766	1.25	31,114	-0.78
Que_Sou	116,806	171,064	198,263	0.7	89,607	-0.48	243,066	1.08	44,804	-0.74
Wes_Sou	77,411	125,812	138,954	0.8	64,269	-0.49	171,088	1.21	32,134	-0.74
Dep_Dal	99,214	145,390	168,361	0.7	76,244	-0.48	206,483	1.08	38,122	-0.74
Mean	118,117	170,579	204,746	0.73	83,950	-0.51	246,721	1.09	41,975	-0.75

further refined with the purposeful selection of certain nodes, to reduce the number of pathways or to focus on certain areas. For example, it would have been possible to identify key amenities of public interest such as schools or hospitals. In fact, this is not far from the solutions that some cities have implemented in the context of the Covid-19 pandemic. This include from pavement widening and the delimitation of car-free zones to setting up temporary cycle lanes to control public transport over-crowding and ease the compliance of social-distancing recommendations (NACTO Covid-19 Transportation Response Center 2020). In London the Covid-19 crisis accelerated ongoing policies of streetspace reclamation that are more adequate for promoting sustainable transport travel (TfL Transport for London 2020). Notably, all these solutions correspond to urban planning schemes that revolve around the idea of reclaiming streetspace from private cars.

The shortest-path analysis of the central area of London, showed that only 30% of the actual street length used to connect all railway and underground stations through the shortest route. The pathway length frequency distribution is normal and pathways have an average length of 10.7 km. Although, the number of routes traversing street segments (carrying load) has a left skewed distribution with maximum values 8000 times bigger than minimum values. The method combining shortest-path with streetspace analysis can be useful for both planning active travel prioritisation strategies and Covid-19 pandemic street management measures to keep cities functioning. In the meantime, street use will continue evolving in response to the pandemic and transition to long-term recovery while mass transit operates with less capacity.

The visual comparison of transport networks illustrates both the endurance of historic networks but at the same time the difficulties of constructing connected and complex networks, despite the use of existing streets (e.g. dispersion and disconnectedness of the current cycle network). Additionally, the spatial visualisation of the street segments carrying load distribution in small multiples, resembles a pattern of urban growth that allow to identify the formation of critical pathways.

Moreover, the multiple dissection of pathways reveals the variations of streetspace designation metrics along a route at high spatial resolution. Street reallocation interventions visualised in this manner allow to understand the positive impacts of narrowing carriageways for increasing the public value of streets (i.e. healthier street environments), while still keeping space for motorised-traffic like buses. This, in addition to the fact that only some streets are needed to connect all stations, reinforces the idea that small interventions on few streets at the city and street levels can have a big impact.

Other known approaches to traffic management consist of the definition of a neighbourhood unit that group together minor streets surrounded by major streets. This same concept was developed by Buchanan as the 'environmental areas' from which through-traffic was excluded and instead it was channelled through the perimeter streets forming the city corridors (Appleyard et al. 1981). Such strategies have had real-world applications in the area of Barnsbury in London and in some areas of Barcelona under the name of 'superblocks' (Rueda 2019). However, even though these approaches favour the creation of good quality pedestrian environments, they

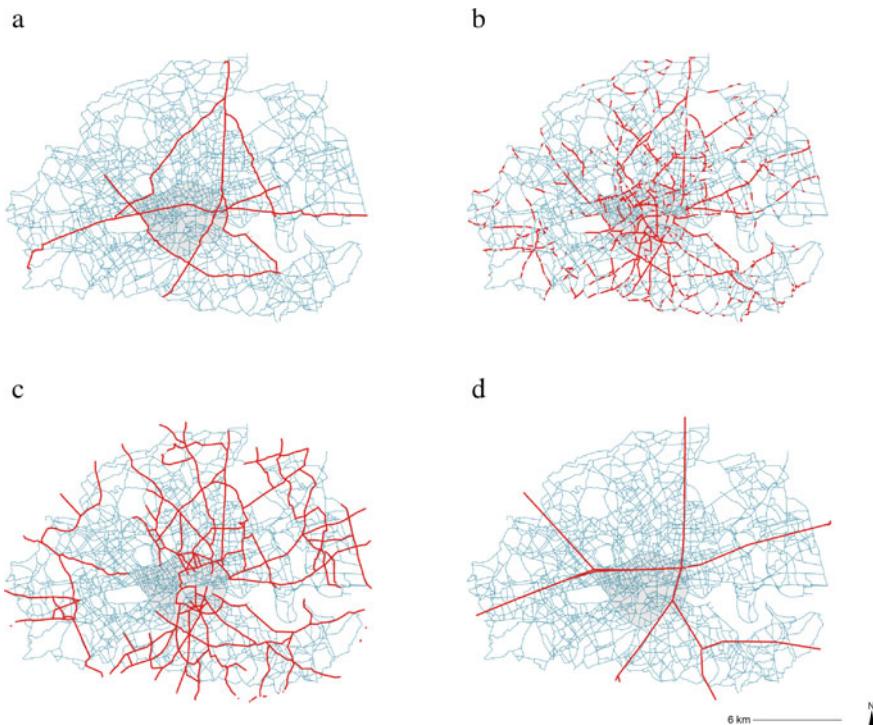


Fig. 10.13 Transport network comparison. **a** is the 8 pathways network, **b** cycle lanes (segregated and shared with bus lanes), **c** former tramways (mid 1900s) and **d** Roman roads diagram. Data sources b: <https://cycling.data.tfl.gov.uk/>, c: https://sharemap.org/public/Trams_in_London, d: <https://darmc.harvard.edu/dataavailability>

operate in an inward-like manner as opposed to the strategy presented here, were the street enhancements are done on the main corridors of the city.

Because the analysis presented here is based in shortest-path and streetspace analysis, some associations were found with the analysis of previous research (see Palominos and Smith (2019), Palominos and Ballal (2018)). First, it was possible to identify the central streets of the system, which in this case has a clear East-West pattern at the North side of the river. Second, a hierarchy of streets with a central-periphery pattern which correspond to 30% of the total street length inside the ULEZ 2021. Third, a left skewed long tail frequency distribution of routes per street segment, similar to centrality distribution. Finally, from urban planning perspective, the simultaneous examination of the strategic and design scales of the street system allows for a more comprehensive analysis and overview of interventions.

Similarly, some associations could be drawn with pre-existing and existing transport networks. For example, the correspondence of the 8 pathways network with the former tramways network, the current cycling infrastructure and bus route network and the Roman roads. Fig. 10.13 demonstrate the slow rate of change of some path-

ways over time. That is the case, for example, of the Clapham Common-White Hart Lane pathway, starting at Clapham Road in the South a continuing along Kingsland Road in the North, following the same trajectory of part of a tramway route and a Roman road (to Chichester in the South and to York in the North). Similar juxtapositions can be found for the rest of the pathways in the 8 pathways network (e.g. Seven Sisters Road, Edgware Road, Whitechapel Road, etc). Certainly, the 8 pathway structure is partially contained by the cycling routes, yet that is only because the majority of the cycling routes represented in Fig. 10.13 share the streetspace with bus lanes. In addition to this, the piecemeal-like pattern of the actual cycling routes reflect the difficulties of transforming the space of the street from car priority to bicycle priority.

In the last section of this chapter, we analysed the pathways in more detail looking at the streetspace designation metrics of a subset of the shortest-paths network. It was not a surprise to find that the pathways are formed by relatively wide streets with more space assigned to the vehicular part. The pathways anatomy visualisation of multiple cross-sections offers an alternative more detailed perspective of the streetspace designation metrics variations along a route. Also, we studied two possible scenarios for prioritising active travel and public transport through the reallocation of carriageway space. These show that by reducing carriageway space to minimum functional standards can have a significant increase of streetspace for place-based street improvements and space-efficient modes of travel. Given the current social distancing restrictions and the decreased capacity of mass transit services due to the Covid-19 pandemic, these street interventions should have a positive effect for both controlling virus propagation and encouraging a sustainable and healthy recovery (Batty 2020).

Data Accessibility Data for the cycling infrastructure analysis are available online from TfL's Cycling Infrastructure Database. Metrics of streetspace designation metrics are available from Zenodo: <https://doi.org/10.5281/zenodo.3783807>

References

- Alexander, C. (1977). *A pattern language: Towns, buildings, construction*. Oxford university press.
- Allen, J. J. (2018). Comparative visualizations of transport networks in Calgary using shortestpath trees. *Environment and Planning A: Economy and Space*, 50(1), 9–13.
- Appleyard, D., Gerson, M. S., & Lintell, M. (1981). *Livable streets, protected neighborhoods*. University of California Press.
- Barrett, S., Colthorpe, T., & Wedderburn, M. (2017). *Street smarts: Report of the commission on the future of London's roads and streets*. London: Centre for London.
- Barthélémy, M. (2011). Spatial networks. *Physics Reports*, 499(1–3), 1–101.
- Batty, M. (2020). Social distancing at scale. *Environment and Planning B: Urban Analytics and City Science*, 47(9), 1533–1536.
- Carmona, M., Gabrieli, T., Hickman, R., Laopoulos, T., & Livingstone, N. (2018). Street appeal: The value of street improvements. *Progress in Planning*, 126, 1–51.

- City of London Corporation. (2019). City streets transport for a changing square mile. Retrieved September 2020, from <https://www.cityoflondon.gov.uk/assets/ServicesEnvironment/city-of-london-transport-strategy.pdf>.
- Claris, S., & Scopelliti, D. (2016). Cities alive: Towards a walking world. Retrieved September 2020, from <https://www.arup.com/perspectives/cities-alive-towards-a-walking-world>.
- Cycling and walking—GOV.UK. (2020). Retrieved December 2020, from <https://www.gov.uk/travel/cycling-and-walking>.
- European Commission. Environment Directorate-General. (2004). Reclaiming city streets for people: chaos or quality of life?. Office for Official Publications of the European Communities.
- GLA (Greater London Authority). (2018). Mayor's transport strategy. Retrieved September, 2020 from <https://www.london.gov.uk/sites/default/files/mayors-transport-strategy-2018.pdf>.
- Gössling, S. (2020). Why cities need to take road space from cars—and how this could be done. *Journal of Urban Design*, 1–6.
- Hackney Council. (2015). Hackney transport strategy 2015–2025. Retrieved September 2020, from <https://drive.google.com/file/d/173Yz8wZuSdjnXb15q0dSsIhWzEkRGkBo/view>.
- Hillier, B., & Iida, S. (2005). Network and psychological effects in urban movement. In *International Conference on Spatial Information Theory* (pp. 475–490). Springer, Berlin, Heidelberg.
- Hollander, Y. (2016). Transport modelling for a complete beginner. CTthink!.
- Institute for Architecture and Urban Studies. (1978). On streets. In S. Anderson (Ed.), (pp. 1–7). Cambridge, MA: MIT press.
- Ibraeva, A., de Almeida Correia, G. H., Silva, C., & Antunes, A. P. (2020). Transit-oriented development: A review of research achievements and challenges. *Transportation Research Part A: Policy and Practice*, 132, 110–130.
- Jacobs, J. (2016). *The death and life of great American cities*. New York: Random House.
- Manley, E. J., Addison, J. D., & Cheng, T. (2015). Shortest path or anchor-based route choice: A large-scale empirical analysis of minicab routing in London. *Journal of Transport Geography*, 43, 123–139.
- Marshall, S., Gil, J., Kropf, K., Tomko, M., & Figueiredo, L. (2018). Street network studies: From networks to models and their representations. *Networks and Spatial Economics*, 18(3), 735–749.
- Mboup, G. (2013). Streets as public spaces and drivers of urban prosperity. United Nations Human Settlements Programme (UN-Habitat).
- NACTO Covid-19 Transportation Response Center. (2020). Retrieved September 2020, from <https://nacto.org/program/covid19/>.
- National League of Cities (2019). Micromobility in cities a history and policy review. Retrieved September 2020, from <https://www.nlc.org/resource/micromobility-in-cities-a-history-and-policy-overview>.
- New London Architecture. (2019). Future streets. Retrieved September 2020, from <https://nlaproducts.s3.amazonaws.com/3051/NLA-Future-Streets-Publication.pdf>.
- Palominos, N., & Ballal, H. (2018). Web-based analytical tools to support decision-making in the context of tactical street design. *Disegnarecon*, 11(20).
- Palominos, N., & Smith, D. A. (2019). Quantifying and mapping streetspace: A geocomputational method for the citywide analysis of pedestrian and vehicular streetspace (p. 21). CASA Working Paper 212.
- Porta, S., Strano, E., Iacoviello, V., Messora, R., Latora, V., Cardillo, A., & Scellato, S. (2009). Street centrality and densities of retail and services in Bologna, Italy. *Environment and Planning B: Planning and Design*, 36(3), 450–465.
- Porta, S., Latora, V., Wang, F., Rueda, S., Strano, E., Scellato, S., & Latora, L. (2012). Street centrality and the location of economic activities in Barcelona. *Urban Studies*, 49(7), 1471–1488.
- Rueda, S. (2019). Superblocks for the design of new cities and renovation of existing ones: Barcelona's case. In *Integrating Human Health into Urban and Transport Planning* (pp. 135–153). Springer, Cham.
- Sadik-Khan, J., & Solomonow, S. (2017). Street fight: Handbook for an urban revolution. Penguin.

- TfL (Transport for London). (2017). Healthy streets for London. London: Transport for London. Retrieved September 2020, from <https://content.tfl.gov.uk/healthy-streets-for-london.pdf>.
- TfL (Transport for London). (2018). Travel in London report 12. Retrieved September 2020, from <https://content.tfl.gov.uk/travel-in-london-report-12.pdf>.
- TfL (Transport for London). (2019). Transport for London business plan 2019–2024. Retrieved September 2020, from <https://content.tfl.gov.uk/tfl-business-plan-2019-24.pdf>.
- TfL (Transport for London). (2020). Streetspace for London. Retrieved December 2020, from <https://www.tfl.gov.uk/travel-information/improvements-and-projects/streetspace-for-london>.
- Tice, P. C. (2019). Micromobility and the built environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 63, No. 1, pp. 929–932). Sage CA: Los Angeles, CA: SAGE Publications.
- Traffic Management Act. (2004). Network management in response to COVID-19. (2020). GOV.UK. Retrieved December 2020, from <https://www.gov.uk/government/publications/reallocating-road-space-in-response-tocovid-19-statutory-guidance-for-local-authorities/traffic-management-act-2004-networkmanagement-in-response-to-covid-19>.
- UCL Commission for Mission-Oriented Innovation and Industrial Strategy (MOIIS) cochaired by Mazzucato, M. and Willetts, D. (2019). A Mission-Oriented UK Industrial Strategy. UCL Institute for Innovation and Public Purpose, Policy Report, (IIPP WP 2019–04). <https://www.ucl.ac.uk/bartlett/public-purpose/wp2019-04>.
- Zandiataashbar, A., & Hamidi, S. (2018). Impacts of transit and walking amenities on robust local knowledge economy. *Cities*, 81, 161–171.

Chapter 11

Multi-scale CyberGIS Analytics for Detecting Spatiotemporal Patterns of COVID-19



Fangzheng Lyu, Jeon-Young Kang, Shaohua Wang, Su Yeon Han, Zhiyu Li,
and Shaowen Wang

11.1 Introduction

The COVID-19 pandemic has put the world into an extremely difficult situation (WHO 2020). In the United States alone, the total number of cases has gone over 7 million, and the cases of death reached approximately 200,000 as of September 16, 2020. Although a majority of COVID-19 patients experience mild symptoms, including dry cough, sore throat, and fever (Zhou et al. 2020; Sohrabi et al. 2020), the patients who died had to cope with severe pneumonia, septic shock, and respiratory failure (Backer et al. 2020; Linton et al. 2020; Chen et al. 2020). It is of importance to understand spatiotemporal patterns of COVID-19 for effective control and prevention of the spread of COVID-19 due to its highly transmissive nature. A timely response is critical to control and prevent further transmission (Bitzegeio et al. 2020).

Space-time clustering and geo-surveillance play a vital role in not only gaining an improved understanding of the nature of the disease spread, but also providing timely information for decision-making support. Exploring spatial epidemiology data containing a temporal signature is also one of the effective ways to advance the knowledge frontiers of infectious diseases (Aldstad and Arthur 2006; Delmelle et al. 2014; Rogerson and Yamada 2008). With the temporal aspects, the spatial scan statistics are useful for answering questions about whether spatial clusters have abnormally arisen (Desjardins et al. 2020; Kulldorff 1999, 2001; Naus and Sylvan 2006). Commonly

F. Lyu · S. Wang · S. Y. Han · Z. Li · S. Wang (✉)

CyberGIS Center for Advanced Digital and Spatial Studies, University of Illinois at Urbana-Champaign, Urbana, IL, USA

e-mail: shaowen@illinois.edu

Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA

J.-Y. Kang

Department of Geography Education, Kongju National University, Gongju-si,
Chungcheongnam-do, South Korea

used methods include space-time interaction modeling (Mantel 1967), spatiotemporal k -nearest neighbors test (Jacquez 1996), scan statistics (Shi and Pun-Cheng 2019; Glaz et al. 2001), space-time density-based clustering (ST-DBSCAN) (Kullendorff 2001; Kulldorff et al. 2005), space-time kernel density estimation (STKDE) (Brunsdon et al. 2007), and windowed nearest neighbor methods (Pei et al. 2010).

The detection of localized disease outbreaks can support policymakers and public health officials to control and prevent the disease spread as early as possible (Takahashi et al. 2008). The STKDE has been developed as an effective method to test disease intensities in space and time and visualize them in a three-dimensional (3D) framework (Delmelle et al. 2014; Hohl et al. 2016). The results from STKDE may provide clues on why COVID-19 cases have densely occurred around particular areas during specific times. A deeper investigation of STKDE results with socio-economic characteristics can help improve related intervention efforts (e.g., social distancing and quarantine).

Therefore, this study aims to explore spatiotemporal intensities of COVID-19 death cases in the US using STKDE. Specifically, the following research questions are addressed to examine the spatiotemporal distribution of COVID-19 deaths. First, where are the COVID-19 clusters of death cases located spatially and temporally? Second, what spatiotemporal patterns can be interpreted from the COVID-19 clusters and probability density of COVID-19 death cases with STKDE? Third, what are the socio-economic characteristics of such clusters at the neighborhood scale?

By examining the spatiotemporal patterns of COVID-19 deaths in the US, this study focuses on advancing multi-scale understanding of the disease spread and its impacts based on cyberGIS that enables computationally intensive STKDE. The rest of this paper is structured as follows. Section 11.2 introduces our study area and corresponding data for analysis. In Sect. 11.3, a cyberGIS-based STKDE workflow is described. Section 11.4 illustrates the results of our cyberGIS-enabled analysis. We conclude the chapter with a discussion on the findings and point out future research directions in Sect. 11.5.

11.2 Data and Study Area

During the past several months in 2020, the US has become an epicenter of the COVID-19 pandemic with the largest number of reported cases. As a result, unfortunately, a large number of individuals died as victims of this deadly disease. To analyze spatiotemporal patterns of COVID-19 in the US, this study focuses on all the death cases reported in the US from February 29, 2020 when the first death was reported in the US, until August 4th. As the pandemic intensifies across the globe, various web applications and portals have been developed to track COVID-19 cases in the US and world, including, for example, WHO (<https://who.sprinklr.com>), CDC (<https://www.cdc.gov/coronavirus/2019-ncov/cases-in-us.html>), ECDC (<https://www.ecdc.europa.eu/en/geographical-distribution-2019-ncov-cases>) and academic institutions including Johns

Hopkins University (<https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6>), University of Washington (<https://hgis.uw.edu/virus/>), and University of Virginia (<https://nssac.bii.virginia.edu/covid-19/dashboard>). Meanwhile, many organizations, such as WHO, CDC, New York Times, and some state agencies, provide open access to datasets that contain the information of confirmed COVID-19 cases and deaths.

In this study, our COVID-19 dataset is retrieved from the New York Times via its Github repository (<https://github.com/nytimes/covid-19-data/>). Compiled from the state and local governments and health departments, this dataset provides COVID-19 cases at the county level with daily updates. Figure 11.1 shows the number of death cases per 100,000 people in the US on August 4th, 2020 from the WhereCOVID-19 platform (WhereCOVID-19 2020) (<https://wherecovid19.cigi.illinois.edu/>). Across the entire US, most counties are impacted by COVID-19 regardless of their locations. However, some counties suffer more than others due to the uneven distribution of COVID-19 cases and deaths. Therefore, we are interested in exploring the spatiotemporal characteristics of COVID-19 distribution by analyzing the reported death cases. The primary reason for our focus on the death cases instead of the confirmed cases is that COVID-19 positive cases were not captured well for the entire US due to testing availability and quality (Abbasi 2020). However, the reporting of death cases is generally considered more accurate as it is subject to more rigorous review processes.

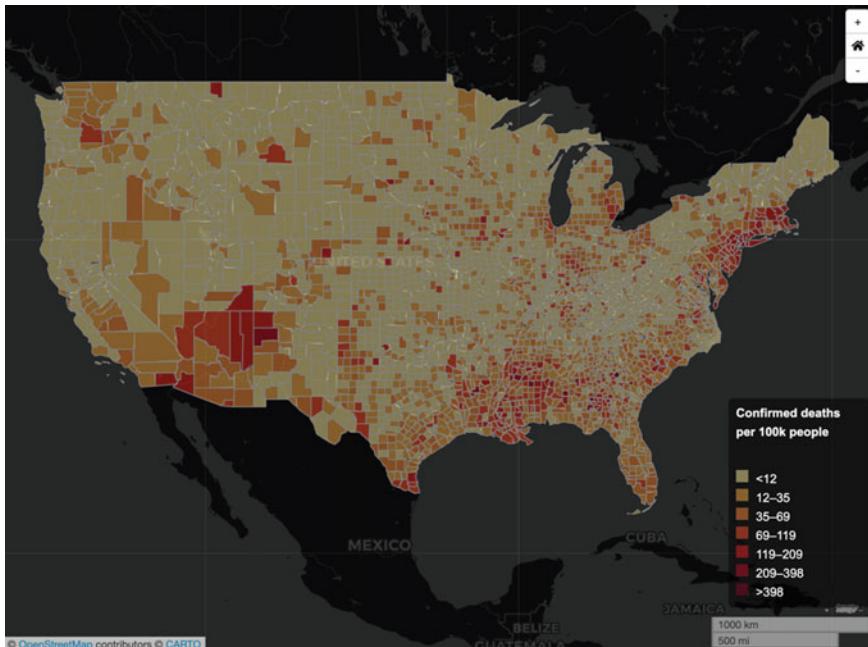


Fig. 11.1 COVID-19 death cases per 100,000 people in the US on August 4th, 2020

The dataset was processed as follows: (1) due to the fact that the original dataset lists the cumulative cases for each county, the dataset is reorganized into daily cases; (2) each case is mapped to its county's centroid; (3) when there are multiple COVID-19 death cases in one certain county within a single day, it is assumed that all the cases are temporally evenly distributed within the day; and (4) the World Geodetic System (WGS 1984) is replaced by Spherical Mercator in this study for more accurate measurement of the location of each county.

11.3 Method

11.3.1 Workflow

To enable data-, computation- and collaboration-intensive geospatial problem solving, cyberGIS—geospatial information science and systems (GIS) based on advanced computing and cyberinfrastructure (CI) (Wang 2010) has been applied to a variety of domains, such as bioenergy (Hu et al. 2017), public health (Padmanabhan et al. 2014; Kang et al. 2020a, b), hydrology (Lyu et al. 2019), and emergency management (Vandewalle et al. 2019). CyberGISX is a cutting-edge cyberGIS platform for supporting computationally reproducible geospatial analytics using CyberGIS-Jupyter notebooks (Yin et al. 2019). Figure 11.2 shows the architecture

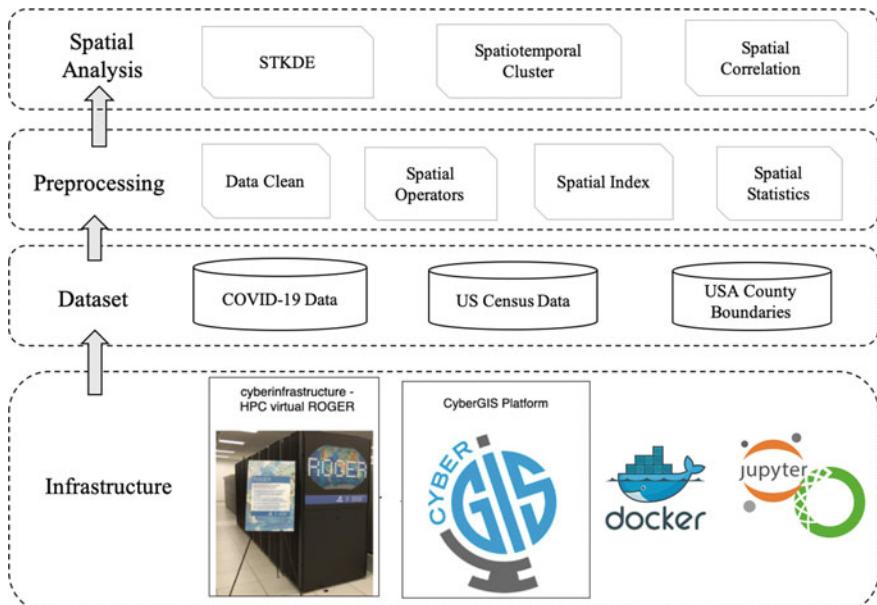


Fig. 11.2 CyberGIS architecture for detecting spatiotemporal patterns of COVID-19

for detecting spatiotemporal clustering of COVID-19 in the US. The architecture includes infrastructure, dataset, preprocessing, and spatial analysis tiers.

The infrastructure tier represents advanced cyberinfrastructure capabilities including such high-performance computing resources as Virtual ROGER (<https://cybergis.illinois.edu/infrastructures/>). The CyberGISX platform integrates geospatial data and analytics at scale (Wang et al. 2013). Using Docker facilitates fast and large-scale deployment of geospatial software and applications (Wang et al. 2019). The preprocessing tier is important to obtain data-intensive spatial analysis results with high efficiency. Data cleaning utilities help remove outliers in raw datasets. Spatial operators are used for joining the results of STKDE and the U.S. census data. The 3D spatiotemporal information revolving around COVID-19 death cases is stored in a *kd*-tree-based spatial index for computing spatial-temporal kernel density estimation (STKDE) efficiently. In the spatial analysis tier, STKDE is used to detect spatiotemporal patterns of COVID-19 in the US. Spatial correlation analytics is used for revealing the relationships between the results of STKDE and population characteristics.

11.3.2 Space-Time Kernel Density Estimation

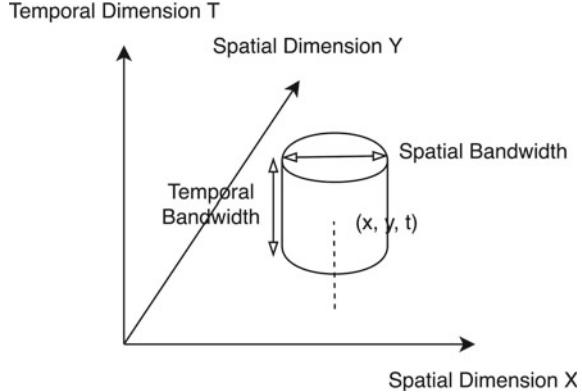
Kernel Density Estimation (KDE), devised for estimating a smooth empirical probability function (Silverman 1986), is a widely used spatial analysis method that can transform a set of geospatial points into a density surface (Nakaya and Yano 2010). Similar to a two-dimensional spatial kernel, a space-time kernel considers the temporal dimension as a dimension that is orthogonal to the spatial dimension. Brunsdon et al. (2007) proposed the following kernel for space-time kernel density estimation as specified by Eq. (11.1):

$$f(x, y, t) = \frac{1}{nh_1^2h_2} \sum_{i=1}^n k_s\left(\frac{x - x_i}{h_1}, \frac{y - y_i}{h_1}\right)k_t\left(\frac{t - t_i}{h_2}\right) \quad (11.1)$$

where $f(x, y, t)$ is the output of space-time kernel density estimation at the location of (x, y, t) , n is the number of events, which is the number of COVID-19 death cases in our study, h_1 is spatial bandwidth, h_2 is temporal bandwidth, and k_s and k_t are the probability functions defined across space and over time respectively. Figure 11.3 illustrates a space-time kernel for a given spatial-temporal point (x, y, t) . All the events within the cylinder are taken into consideration when calculating the probability density for (x, y, t) .

Meanwhile, the function k_s and k_t are defined using the Epanechnikov kernel (Epanechnikov 1969), which is a widely used kernel function for kernel density estimation (Eq. 11.2).

Fig. 11.3 Spatial-temporal kernel



$$k_s(u, v) = \begin{cases} \frac{2}{\pi}(1 - (u^2 + v^2)), & (u^2 + v^2) < 1 \\ 0, & \text{otherwise} \end{cases}$$

$$k_t(w) = \begin{cases} \frac{3}{4}(1 - w^2), & w^2 < 1 \\ 0, & \text{otherwise} \end{cases} \quad (11.2)$$

11.3.3 Bandwidth and Scale

Bandwidth selection is crucial in STKDE analysis. A small bandwidth can generate an output density surface with a strong emphasis on local features while often making the density surface not smooth. On the other hand, a large bandwidth can generate a smooth surface, even though it may filter out some major fine-scale details. In this study, a commonly used algorithm for determining optimal bandwidth proposed by Härdle (2004) and Nakaya and Yano (2010) is adopted with the Eq. (11.3) applied to determine the bandwidth for three-dimensional kernel density estimation.

$$h_k = 2.21n^{-\frac{1}{7}}\sigma_k \quad (11.3)$$

where h_k represents the bandwidth for the kernel at the k th dimension, n is the number of events, which is the total number of cases, while σ_k is the standard deviation of the k th dimension for all the cases. In this study, after applying the bandwidth formula for the dataset, the spatial bandwidth is set to 300 km while the temporal bandwidth is specified as 1.8 days.

In addition, the selection of the spatial and temporal resolutions is a question about balancing accuracy and computational intensity. A fine spatial and temporal resolution can generate more detailed results from STKDE, which tends to be computationally intensive. On the other hand, a coarse spatial and temporal resolution likely fails to provide a clear picture of the density surfaces. In this study, we aim to achieve

fine-scale STKDE results by looking into the distribution of COVID-19 death cases across the conterminous US. Therefore, a spatial resolution of 20 km and a temporal resolution of 1 day are chosen. Here, 1 day is believed to be the finest temporal resolution given that data is collected on a daily basis. Meanwhile, since we are handling county-level COVID-19 data, 20 km is a spatial resolution that can make sure that at least one spatial point lies within each county. Also, the computational intensity is resolvable for the computation at the scale of the entire conterminous US. As a result, we sampled STKDE results in a space-time cube that consists of about 500,000 spatial-temporal points. The fine spatial resolution selected in this study helps to achieve a better understanding of the spatial-temporal patterns of COVID-19 in a detailed fashion. Due to the high computational intensity incurred with the fine spatial resolution, we employ advanced cyberinfrastructure and develop a scalable STKDE algorithm to resolve the computational intensity within a reasonable time.

After finishing the STKDE modeling and normalization by rescaling the data to have a value between 0 and 1, we apply a 95% threshold to the existing non-zero values to find the spatiotemporal points where the intensity for COVID-19 death cases is significant. Based on the thresholding result, the spatiotemporal clusters are derived.

11.3.4 Neighborhood Analysis

To understand the association between COVID-19 clusters from STKDE results and the socio-economic characteristics of corresponding neighborhoods, we analyze the neighborhood characteristics of each cluster by calculating its z -score and assess related social-economic characteristics based on STKDE results and county-level socioeconomic data. Specifically, we analyze the following nine attributes to study the racial, age, and income factors of COVID-19 clusters: (1) the percentage of white population; (2) the percentage of black population; (3) the percentage of American Indian population; (4) the percentage of Asian population; (5) the percentage of Hawaiian population; (6) the percentage of the population from other race and ethnicity categories; (7) the percentage of the population over 60; (8) low-income population; and (9) median income.

Based on the cluster boundary derived from STKDE results and socio-economic data collected from the US census, we calculated the z -score of each county based on the characteristics of all counties that are located in the conterminous US. Then, based on the boundary of each cluster, we can find the counties that lie within each cluster. Compared with all the US counties except those in Alaska and Hawaii, the z -score for each county was calculated. After integrating county-level values, we derive the z -score and p -value for each cluster.

11.4 Results

11.4.1 STKDE

The probability density calculated with STKDE represents how severe a place on a certain day is affected by COVID-19 based on death data. Based on the probability distribution of standardized density from the output of STKDE (Fig. 11.4), about 99% of the spatial-temporal points have value of 0. This is likely caused by the fact that in the early stage of the disease spread, the number of death cases happening in the US was relatively small and there were many counties in the US that had 0 death at that time. However, the number of deaths increased significantly in March and April, resulting in the normalized density from STKDE to rise rapidly. As shown in Fig. 11.5, the cumulative standardized density of STKDE increases steadily as the standardized density of STKDE increases. That is to say, the number of spatial-temporal points that are significantly higher than the others is still relatively small. Moreover, the spatial-temporal distribution of COVID-19 death cases in the US is highly clustered. As a result, the distribution of COVID-19 cases is skewed spatially and temporally.

The STKDE results provide an improved understanding of the spatiotemporal distribution of COVID-19 death cases across the US. They identify the probability density of spatial-temporal points for COVID-19 daily cases in the US, which enables our follow-on correlation analysis. In addition, the results from STKDE help identify spatial-temporal clusters. Furthermore, these results can further enable a clear 3D visualization of the accumulative COVID-19 deaths in the conterminous US.

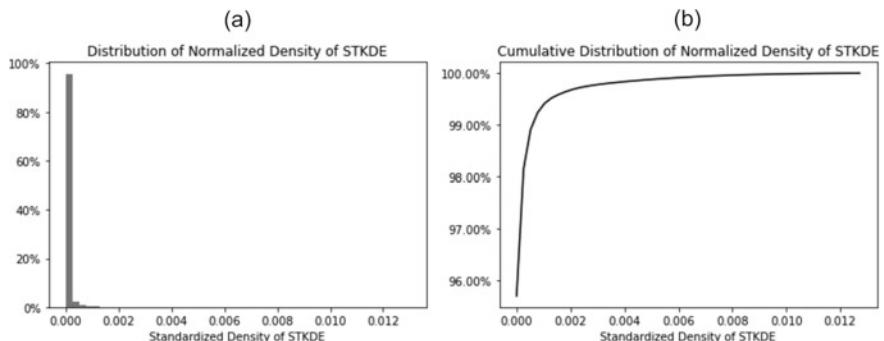


Fig. 11.4 Distribution of standardized density from the output of STKDE. **a** Probability distribution, **b** cumulative distribution

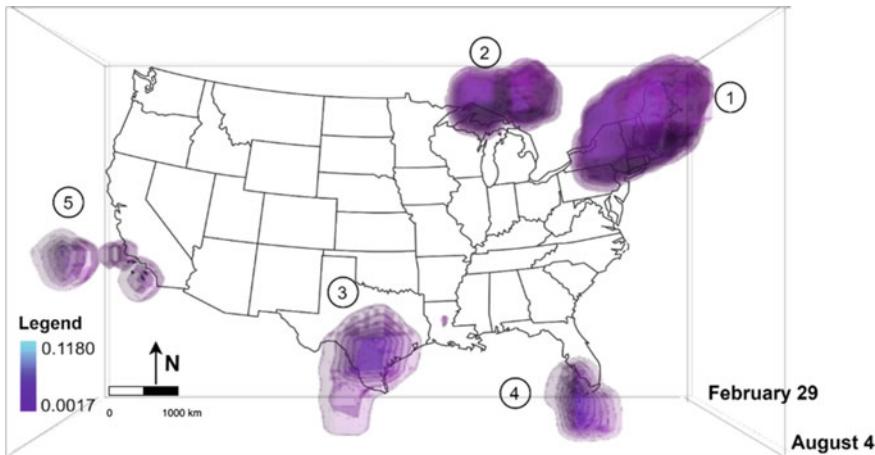


Fig. 11.5 Clustering output of COVID-19 death cases based on STKDE (above view). 1: New York City, 2: Chicago/Detroit/Indianapolis, 3: Houston, 4: Miami/Tampa, and 5: LA/San Diego

11.4.2 Clustering Analysis

The STKDE results can be used to detect spatial-temporal clusters to discover significant events and patterns. In particular, the COVID-19 clusters of abnormally high probability density can be detected in this study. The clustering results can help identify not only hotspots of COVID-19 death cases, but also provide temporal information regarding each domain. Valuable information can be interpreted from the clustering results to address the following questions: (1) where are the COVID-19 clusters spatially and temporally located? (2) for each COVID-19 cluster, what is its spatial span for different timestamps? (3) when did the outbreak first begin for different clusters and when did the clusters suddenly grow bigger? and (4) how to distinguish spatiotemporal ranges of the clusters for predicting the future trend of the COVID-19 spread at different locations? The STKDE output is used to detect COVID-19 clusters of death cases in the conterminous US. The COVID-19 death data collected from the first date of death reported in the US, which is February 29th, to August 4th, is used as input for our STKDE analysis.

Based on the clustering analysis in the US (Fig. 11.5), there are 5 major clusters of COVID-19 death cases, as follows: (1) the biggest cluster was located in the state of New York, especially in New York City, with the dark color indicating that it was the center of an outbreak; (2) the second largest cluster was located near Chicago, Illinois, and this cluster was extending northeast toward the state of Michigan and south to Indiana; (3) the third-largest cluster was located in the state of Texas, centered near the city of Houston; (4) the fourth largest cluster was located near south Florida; and (5) the fifth cluster was located in the state of California, near the city of Los Angeles. In the fifth cluster, the two epicenters (Los Angeles and San Diego) were not directly connected to each other, indicating that the COVID-19 death cases between

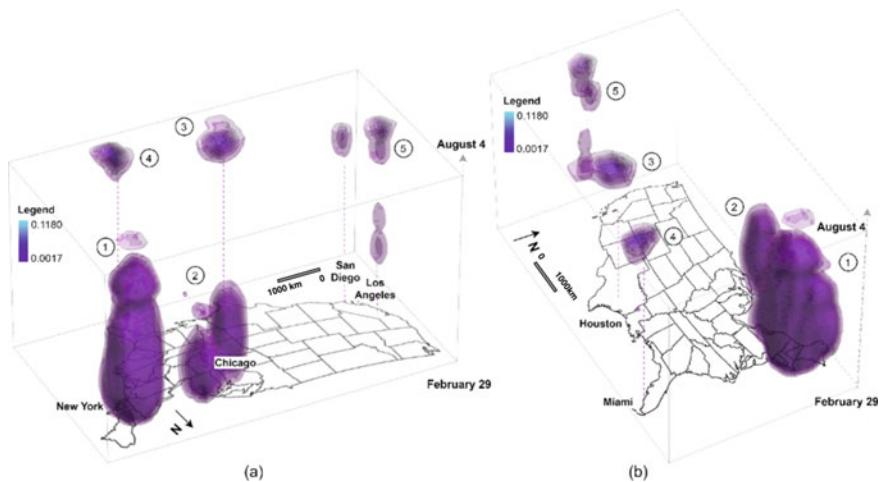


Fig. 11.6 Clustering output of COVID-19 death cases based on STKDE

two epicenters were not significant. From the perspective of the spatial distribution of COVID-19 death cases, each of the clusters was centered at least in one major urban area. The east coast suffered more from the pandemic compared with the west coast. Particularly, the State of New York and the area near Chicago are two regions that were most severely impacted. The clustering results suggest that regions with low population density are not likely to form any significant clusters.

From the perspective of the temporal dimension (Fig. 11.6), several fine-scale temporal patterns can be distinguished as follows. The cluster in the State of New York started in mid-March, and since then, this cluster had been steadily growing bigger spatially and larger in intensity and continues to be the largest COVID-19 cluster among all clusters. Until mid-April, this cluster reached its maxima in terms of intensity and spatial span. After that time, this cluster started shrinking with a small spike near the end of June.

The second largest cluster, which centers at Chicago, emerged around mid-April, slowly extending and connecting northeast toward another center for this cluster at Detroit. Until early June, the Chicago/Detroit cluster had stayed similar in terms of intensity and spatial span, and that was the time COVID-19 death cases reached their peak in this cluster. After that, the sub-cluster located near Detroit disappeared first with the sub-cluster centered at Chicago disappearing in around two months. On August 4th, this cluster did not exist anymore, implying that the death cases in Chicago/Detroit clusters had been well under control.

The clusters located in Miami and Houston share some similar characteristics. Both of the clusters started more recently around early July, and they took over the New York cluster to become the new epicenter of COVID-19 deaths in the US. During a 2-week span starting from early July, we can observe that both clusters were growing in terms of intensity and spatial span. However, a difference does exist

between these two clusters. In particular, the cluster located near Houston spanned more widely spatially compared to the cluster in Miami.

In the cluster located across Los Angeles and San Diego, there were three small outbreaks of COVID-19 death cases represented by different spatial-temporal points. The first outbreak of COVID-19 death cases was a small cluster happening around mid-April in Los Angeles. At that time, the intensity of the cluster was not very strong with a relatively small number of death cases in Los Angeles. Starting from early August, there were two small clusters of COVID-19 death cases emerging in Los Angeles and San Diego. Similar to the aforementioned Miami and Houston clusters, additional actions need to be taken to stop the potential growth of these small clusters in the future.

The clusters centered around New York City and Chicago started to shrink after strict and effective policies like stay-at-home orders and recommendation of putting on face masks in public places were applied. Among all of the clusters, the east coast ones suffered more from the pandemic compared with the west coast ones. The cluster centered around New York City was the largest cluster in the US in terms of intensity and spatial span, and all the other clusters were less significant compared to the New York City one. The west part of the Midwest US, together with the east part of the west US, had relatively fewer cases and lower probability density based on STKDE analysis compared to other regions, which was likely caused by low population density (Oster et al. 2020).

Figure 11.7a illustrates the distribution of normalized output from STKDE, which represents the COVID-19 death case intensity, for each cluster. The probability density of the New York City cluster is in general much higher compared with other clusters. The Houston cluster exceeded the Chicago cluster in terms of maximum probability intensity and 75 percentile to become the second-largest epicenter of COVID-19 death cases. Figure 11.7b shows the maximum normalized output from STKDE for each cluster daily since each cluster first emerged, which can help better understand the COVID-19 spread speed in the cluster center. In addition, Fig. 11.7 shows a weekly pattern detected for the Chicago cluster (Ricon-Becker et al. 2020). To summarize, the STKDE clustering results not only provide a clear visualization of the distribution of COVID-19 death cases, but also help to detect spatiotemporal patterns. In addition, the output from STKDE can be used to conduct correlation analysis with other factors such as population density, which we will discuss in the following section.

11.4.3 Neighborhood Characteristics

We investigated the neighborhood characteristics of each cluster by looking into the COVID-19 death cases on April 4th, 2020. The socio-economic characteristics of the five clusters discussed above from the STKDE results are summarized with the following five major findings identified (Fig. 11.8a):

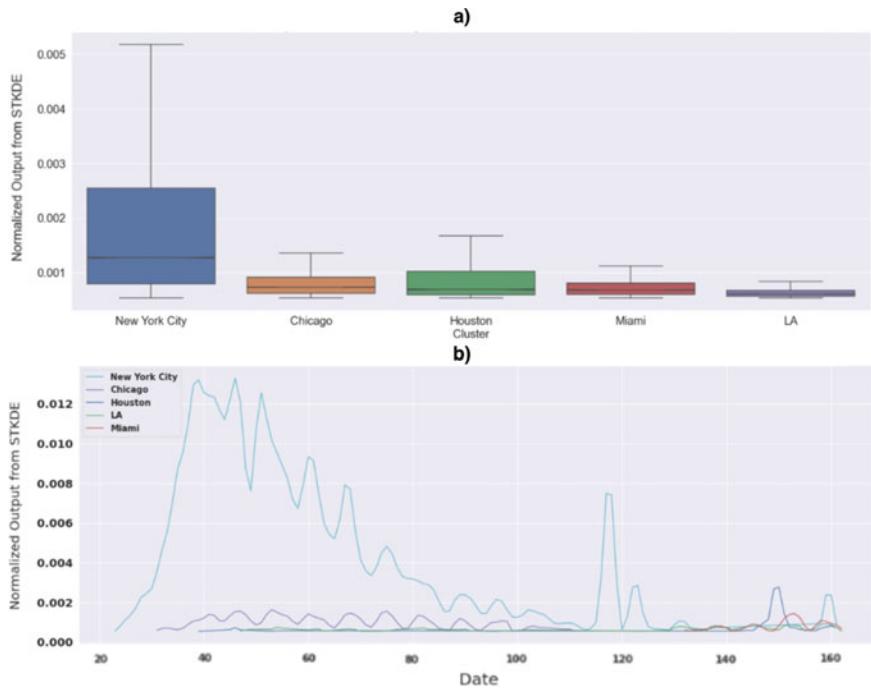


Fig. 11.7 Normalized output analysis from STKDE results: **a** boxplot for normalized output from STKDE for each cluster; **b** the maxima of normalized output from STKDE for each cluster on each day (the *x*-axis represents the number of days starting from February 29th on which the first death case was reported in the US)

- 4 out of 5 COVID-19 clusters have negative *z*-scores for white populations, indicating that no neighborhood of COVID-19 clusters is centered around white populations.
- In cluster 3, the *z*-score for the black population is over 1, indicating that in the neighborhoods of the Houston cluster, the density of the black population is high (Yang et al. 2020).
- For American Indian and Asian populations, the *z*-scores are all negative for all 5 clusters, implying that they are less vulnerable to COVID-19 compared with the other races.
- Counterintuitively, the *z*-score for the population over 60 is not very high in COVID-19 clusters. Though the older population are more likely to suffer from the pandemic, there is little sign that there is a clear relationship between COVID-19 clusters and places where the percentage of the older population is high.

Similarly, from Fig. 11.8b, it can be implied that (1) while we can find some correlation between racial factors and the STKDE results, there was no cluster being significant for any particular racial group given 10% confidence interval; and (2) among all socio-economic factors, the income level turns out to be most significant

	a) z-score								
	White	Black	Indian	Asian	Hawaiian	Other	Over60	Poor	Income
cluster1	-0.290985	0.264279	-0.212087	-0.212087	-0.201378	0.063760	-0.096341	-0.582043	1.078254
cluster2	0.436829	-0.325042	-0.215998	-0.215998	-0.208314	-0.196827	-0.293374	-0.405361	0.359915
cluster3	-0.901038	1.158796	-0.179583	-0.179583	-0.151405	-0.021866	-0.417035	0.656293	-0.383649
cluster4	-0.104948	0.129618	-0.184542	-0.184542	-0.058465	0.135201	1.108231	-0.008532	-0.081730
cluster5	-1.126193	-0.262529	-0.135905	-0.135905	1.213182	3.159590	-1.296976	-0.070190	1.248906

	b) p-value								
	White	Black	Indian	Asian	Hawaiian	Other	Over60	Poor	Income
cluster1	0.385531	0.395782	0.416019	0.416019	0.420202	0.474581	0.461625	0.280269	0.140460
cluster2	0.331118	0.372575	0.414495	0.414495	0.417492	0.421981	0.384618	0.342606	0.359455
cluster3	0.183784	0.123270	0.428740	0.428740	0.439828	0.491278	0.338326	0.255818	0.350619
cluster4	0.458209	0.448434	0.426794	0.426794	0.476689	0.446227	0.133881	0.496596	0.467431
cluster5	0.130042	0.396457	0.445948	0.445948	0.112530	0.000790	0.097320*	0.472021	0.105850

Fig. 11.8 Neighborhood characteristics of the 5 clusters: **a** z-score, with a darker box containing a larger z-score value; **b** p-value

with the lowest *p*-value, which might result from the fact the COVID-19 clusters were centered near cities where the average income level is high.

11.5 Concluding Discussion

The world continues to suffer greatly from the COVID-19 pandemic. For the control and prevention of COVID-19 spread, it is of importance to understand the dynamic nature of COVID-19 transmission. For example, timely identification of spatiotemporal clusters of COVID-19 can help to provide situational awareness and establish an etiologic linkage to the infections with socio-economic characteristics of local population. In this study, our fine-scale STKDE results detect COVID-19 daily intensities across space and over time. We found there were five clusters of COVID-19 death cases centered around New York City, Chicago/Detroit/Indianapolis, Houston, Miami/Tampa, and Los Angeles/San Diego.

Spatial indexing and computationally scalable geospatial algorithms were developed in the data processing and STKDE computing to make cyberGIS analytics feasible in this study. For example, using a time-slot-based method to join STKDE results and population data can significantly improve computational performance. The STKDE computational time is often sensitive to the resolutions of spatial and temporal units. A fine-tuning process was implemented to optimize STKDE parameters.

Further correlation analysis results show that a certain race/ethnic population density has a strong correlation coefficient. Based on the results, we addressed the health disparity issues existing in the US. A certain race/ethnic population is at a higher risk because of less access to healthcare support (Brown et al. 2016; Kang et al. 2020a, b). In addition to race/ethnic factors, socio-economic characteristics are also related. Low-income people might have to come to their workplace regularly even under the ‘stay at home’ order, which in turn, likely become more exposed to COVID-19 risk. Therefore, to better understand the disadvantaged population in the US, our next step is to comprehensively analyze the associations between the evolving COVID-19 risk and socioeconomic factors.

References

- Abbasi, K. (2020). COVID-19: Fail to prepare, prepare to fail. *Journal of the Royal Society of Medicine*, 113(4), 131–131. <https://doi.org/10.1177/0141076820918796>.
- Aldstad, J., & Arthur, G. (2006). Using AMOEBA to create a spatial weights matrix and identify spatial clusters. *Geographical Analysis*, 38(4), 327–343. <https://doi.org/10.1111/j.1538-4632.2006.00689.x>.
- Backer, J., Klinkenberg, D., & Wallinga, J. (2020). Incubation period of 2019 novel coronavirus (2019-nCoV) infections among travellers from Wuhan, China, 20–28 January 2020. *Eurosurveillance*, 25(5), <https://doi.org/10.2807/1560-7917.es.2020.25.5.2000062>.
- Bitzegeio, J., Britta, B., Marius, H., Dagmar, S., Lasse, R., Peter, A., et al. (2020). Two measles clusters in connection with short inner-European air travels indicating impediments to effective measles control: A cluster analysis. *Travel Medicine and Infectious Disease*, 33(January).
- Brown, G., & Jones, K. (2016). Mental health and medical health disparities in 5135 transgender veterans receiving healthcare in the veterans health administration: A case–control study. *LGBT Health*, 3(2), 122–131. <https://doi.org/10.1089/lgbt.2015.0058>.
- Brunsdon, C., Jonathan, C., & Gary, H. (2007). Visualising space and time in crime patterns: A comparison of methods. *Computers, Environment and Urban Systems*, 31(1), 52–75. <https://doi.org/10.1016/j.compenvurbsys.2005.07.009>.
- Chen, N., Zhou, M., Dong, X., Qu, J., Gong, F., Han, Y., et al. (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. *The Lancet*, 395(10223), 507–513. [https://doi.org/10.1016/s0140-6736\(20\)30211-7](https://doi.org/10.1016/s0140-6736(20)30211-7).
- Delmelle, E., Dony, C., Casas, I., Jia, M., & Tang, W. (2014). Visualizing the impact of space-time uncertainties on dengue fever patterns. *International Journal of Geographical Information Science*, 28(5), 1107–1127. <https://doi.org/10.1080/13658816.2013.871285>.
- Desjardins, M. R., Hohl, A., & Delmelle, E. M. (2020). Rapid surveillance of COVID-19 in the United States using a prospective space-time scan statistic: Detecting and evaluating emerging clusters. *Applied Geography*, 118. <https://doi.org/10.1016/j.apgeog.2020.102202>.
- Epanechnikov, V. A. (1969). Nonparametric estimation of a multivariate probability density. *Theory of Probability and Its Applications*, 14, 153–158.
- Glaz, J., Naus, J. I., Wallenstein, S., & Wallenstein, S. (2001). *Scan statistics*. Berlin/Heidelberg: Springer.
- Härdle, W., Müller, M., Sperlich, S., & Werwatz, A. (2004). *Nonparametric and semiparametric models*. Berlin: Springer.
- Hohl, A., Delmelle, E., Tang, W., & Casas, I. (2016). Accelerating the discovery of space-time patterns of infectious diseases using parallel computing. *Spatial and Spatio-Temporal Epidemiology*, 19(November), 10–20.

- Hu, H., Lin, T., Wang, S., & Rodriguez, L. F. (2017). A CyberGIS approach to uncertainty and sensitivity analysis in biomass supply chain optimization. *Applied Energy*, 203, 26–40.
- Jacquez, G. M. (1996). A k-nearest neighbor test for space-time interaction. *Statistics in Medicine*, 15, 1935–1949.
- Kang, J., Aldstadt, J., Vandewalle, R., Yin, D., & Wang, S. (2020a). A CyberGIS approach to spatiotemporally explicit uncertainty and global sensitivity analysis for agent-based modeling of vector-borne disease transmission. *Annals of the American Association of Geographers*, 110(6), 1855–1873. <https://doi.org/10.1080/24694452.2020.1723400>.
- Kang, J., Michels, A., Lyu, F., et al. (2020b). Rapidly measuring spatial accessibility of COVID-19 healthcare resources: A case study of Illinois, USA. *International Journal of Health Geographics*, 19, 36. <https://doi.org/10.1186/s12942-020-00229-x>.
- Kulldorff, M. (1999). Spatial scan statistics: Models, calculations, and applications. *Scan Statistics and Applications*. https://doi.org/10.1007/978-1-4612-1578-3_14.
- Kulldorff, M. (2001). Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 164(1), 61–72. <https://doi.org/10.1111/1467-985x.00186>.
- Kulldorff, M., Heffernan, R., Hartman, J., Assunçao, R., & Mostashari, F. (2005). A space–time permutation scan statistic for disease outbreak detection. *PLoS Medicine*, 2, e59. Scott, D. W. (2015). *Multivariate density estimation: Theory, practice, and visualization*. Hoboken, NJ: Wiley.
- Linton, N. M., Kobayashi, T., Yang, Y., Hayashi, K., Akhmetzhanov, A. R., Jung, S.-M., et al. (2020). Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: A Statistical analysis of publicly available case data. *Journal of Clinical Medicine*, 9(2), 538.
- Lyu, F., Yin, D., Padmanabhan, A., Choi, Y., Goodall, J., Castranova, A., et al. (2019). Reproducible hydrological modeling with CyberGIS-Jupyter: A case study on SUMMA. In *Proceedings of the Practice and Experience in Advanced Research Computing on Rise of the Machines (learning) (PEARC '19)*, Article 21 (pp. 1–6). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3332186.3333052>.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Nakaya, T., & Yano, K. (2010). visualising crime clusters in a space-time cube: An exploratory data-analysis approach using space-time kernel density estimation and scan statistics. *Transactions in GIS*, 14, 223–239. <https://doi.org/10.1111/j.1467-9671.2010.01194.x>.
- Naus, J., & Sylvan, W. (2006). Temporal surveillance using scan statistics. *Statistics in Medicine*, 25(2), 311–324. <https://doi.org/10.1002/sim.2209>.
- Oster, A. M., Kang, G. J., Cha, A. E., Beresovsky, V., Rose, C. E., Rainisch, G., et al. (2020). Trends in number and distribution of COVID-19 hotspot counties—United States, March 8–July 15, 2020. *MMWR. Morbidity and Mortality Weekly Report*, 69(33), 1127–1132. <https://doi.org/10.15585/mmwr.mm6933e2>.
- Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., et al. (2014). FluMapper: A CyberGIS application for interactive analysis of massive location-based social media. *Concurrency and Computation: Practice and Experience*, 26(13), 2253–2265.
- Pei, T., Zhou, C., Zhu, A. X., Li, B., & Qin, C. (2010). Windowed nearest neighbor method for mining spatio-temporal clusters in the presence of noise. *International Journal of Geographical Information Science*, 24, 925–948.
- Ricon-Becker, I., Tarrasch, R., Blinder, P., & Ben-Eliyahu, S. (2020). CITATION TOOLS A seven-day cycle in COVID-19 infection, hospitalization, and mortality rates: Do weekend social interactions kill susceptible people? medRxiv 2020.05.03.20089508. <https://doi.org/10.1101/2020.05.03.20089508>.
- Rogerson, P., & Yamada, I. (2008). Statistical detection and surveillance of geographic clusters. <https://doi.org/10.1201/9781584889366>.
- Shi, Z., & Pun-Cheng, L. S. (2019). Spatiotemporal data clustering: A survey of methods. *ISPRS International Journal of Geo-Information*, 8112.

- Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. London: Chapman & Hall.
- Sohrabi, C., Alsafi, Z., O'Neill, N., Khan, M., Kerwan, A., Al-Jabir, A., et al. (2020). World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19). *International Journal of Surgery*, 76, 71–76.
- Takahashi, K., Kulldorff, M., Tango, T., & Yih, K. (2008). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics*, 7(1), 1–14.
- Vandewalle, R., Kang, J., Yin, D., & Wang, S. (2019). Integrating CyberGIS-Jupyter and spatial agent-based modelling to evaluate emergency evacuation time. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on GeoSpatial Simulation—GeoSim '19*. <https://doi.org/10.1145/3356470.3365530>.
- Wang, S. (2010). A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Annals of the Association of American Geographers*, 100(3), 535–557. <https://doi.org/10.1080/00045601003791243>.
- Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M., Liu, Y., et al. (2013). CyberGIS software: A synthetic review and integration roadmap. *International Journal of Geographical Information Science*, 27(11), 2122–2145.
- Wang, S., Zhong, Y., & Wang, E. (2019). *An integrated GIS platform architecture for spatiotemporal big data*. Future Generations Computer Systems: FGCS. Retrieved from <https://www.sciencedirect.com/science/article/pii/S0167739X17319283>.
- WhereCOVID19. (2020). WhereCOVID-19 Platform. Retrieved from <https://wherencovid19.cigi.illinois.edu/>.
- WHO. (2020). *Novel coronavirus—China*. Retrieved from <https://www.who.int/csr/don/12-january-2020-novel-coronavirus-china/en/>.
- Yang, C., Sha, D., Liu, Q., et al. (2020). Taking the pulse of COVID-19: A spatiotemporal perspective. *International Journal of Digital Earth*, 13(10), 1186–1211. <https://doi.org/10.1080/17538947.2020.1809723>.
- Yin, D., Liu, Y., Hu, H., Terstriep, J., Hong, X., Padmanabhan, A., et al. (2019). CyberGIS-Jupyter for reproducible and scalable geospatial analytics. *Concurrency and Computation: Practice and Experience*, 31(11), e5040.
- Zhou, F., Yu, T., Du, R., Fan, G., Liu, Y., Liu, Z., et al. (2020). Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: A retrospective cohort study. *The Lancet*, 395(10229), 1054–1062. [https://doi.org/10.1016/s0140-6736\(20\)30566-3](https://doi.org/10.1016/s0140-6736(20)30566-3).

Chapter 12

Dynamic Spreading of COVID-19 Versus Community Mobility in Regions of England



Tao Cheng, Xinchen Zhong, Yunzhe Liu, Yang Zhang,
and Guangsheng Dong

12.1 Introduction

Since the first COVID-19 case was reported in December 2019 in Wuhan, China, the novel coronavirus (SARS-CoV-2) has been rapidly spreading to more than 190 countries or regions around the world (WHO 2020a). At the time of writing this article, this unprecedented global pandemic has infected over 54 million people and resulted in more than 1.3 million deaths (WHO 2020b); many other European countries represented by the UK are suffering from the second wave of this pandemic, resulting in the implementation of more strict restrictions including but not limited to the additional regional/national lockdowns (BBC 2020; Cabinet Office 2020).

The early evidence gathered in Wuhan has confirmed the person-to-person transmission of COVID-19 (Chan et al. 2020), meaning that this coronavirus can be transmitted through travels to various communities. As such, the lockdown has been implemented by governments around the world as a national strategy to contain the spreading of COVID-19 and to save life. However, lockdown also has substantial social and economic impacts. Therefore, it is crucial to evaluate the effectiveness of the NPIs in reducing the mobility and the spreading of the virus.

Mobility data provides information about people in different places at different times, which can be utilised to assist in the analysis of the spread of the pandemic (Bonacorsi et al. 2020; Gao et al. 2020). For instance, Bonacorsi et al. (2020) investigated how lockdown strategies affect the economic conditions of individuals and local governments based on a massive analysis of near-real-time Italian mobility data provided by Facebook. Gao et al. (2020) quantified the degree to which social distancing mandates have been followed in the US and its effect on the growth of COVID-19 cases based on daily travel distance and stay-at-home time derived from

T. Cheng (✉) · X. Zhong · Y. Liu · Y. Zhang · G. Dong
SpaceTimeLab, University College London, London WC16BT, UK
e-mail: tao.cheng@ucl.ac.uk

large-scale anonymous mobile phone location data provided by Descartes Labs and SafeGraph.

Furthermore, in order to investigate the impact of control measures in China, Kraemer et al. (2020) analysed the historic mobile data and detail cases with travel history in Wuhan, manifesting the effectiveness of the control measures taken by the Chinese government in alleviating the risk of COVID-19 spreading. Moreover, aggregated daily mobile data were used to generate social distance measures to evaluate its impact on new infection rates in 25 counties of the highest number of confirmed cases in the US (Badr et al. 2020). In addition, researchers have tried to use human mobility data to predict the number of confirmed cases of COVID-19 at the county level in the US (Kapoor et al. 2020; Wang and Yamamoto 2020). In general, most abovementioned studies of COVID-19 concentrate on investigating the relationship between human mobility and the number of confirmed cases.

The confirmed case data has been used to calculate the basic reproduction ratio (R_0), a measurement widely employed in epidemiology to quantify the potential transmission of a pathogen. R_0 is defined as the expected average number of secondary infected cases produced by a single infectious individual in an entirely susceptible population (Anderson and May 1992; Fraser et al. 2009; Nishiura 2010). If $R_0 > 1$, then the epidemic will start spreading in the population, e.g., if R_0 is 2, on average, each infected person infects two more people. When $R_0 < 1$, then the epidemic will shrink and die out, e.g., if R_0 is 0.5, meaning that on average for each two infected people, there will be only one new infection. If $R_0 = 1$, the epidemic will become endemic, i.e. every infected person will infect one other person, indicating that the total number of infections tends to be stable. In general, the larger the value of R_0 is, the harder it is to control the epidemic (Fine et al. 2011). The UK government has employed R_0 as an indicator to monitor the viral transmission in the UK and to guide its citizens to keep social distancing during the first nationwide lockdown implemented in late March to mid-June 2020 (Government Office for Science 2020).

However, the concept of R_0 is “deeply flawed” since “there are many diseases that can persist with $R_0 < 1$, while diseases with $R_0 > 1$ can die out” (Smith et al. 2011). Moreover, R_0 is a single static measure that does not adapt to changes in behaviour and restrictions (Systrom 2020). Given such problems with R_0 , the real-time effective reproduction Ratio (R_t) is accordingly employed as an alternative to R_0 in this analysis, which is the average number of secondary cases caused by an infected individual at time t (Bettencourt and Ribeiro 2008). R_t can be defined as “a time-dependent quantity that accounts for the population’s reduced susceptibility” (Cintron-Arias et al. 2009), which can change over time, and therefore can be utilised to dynamically monitor the ongoing epidemic. For instance, R_t value will decline if there is a reduction in viral transmission, while it will increase if the viral transmission rises. Similar to R_0 , when $R_t > 1$, the number of infected individuals will increase. If $R_t = 1$, the infection will become endemic. For instance, if R_t is greater than 1 for a long time, then lockdown should be continued. On the contrary, if $R_t < 1$ for a long time, people can return to normal life. The value of R_t is able to not only help us understand how effective the NPIs have been controlling an outbreak but also give

us vital information on whether we can increase or reduce restrictions for economic prosperity and human safety in real-time.

Within this context, our work focuses on exploring the dynamic change of mobility in relation to R_t of COVID-19 in nine regions of England. In particular, the UK has gone through the national-wide lockdown from 23/03/2020 to 02/07/2020, which has a significant impact on human mobility and reducing the spreading to a great extent. However, the second wave of the outbreak was on the way when this paper was written in September 2020. Moreover, the trends of the epidemics were different at regional and local levels. Therefore, why there are different trends of epidemic development? Is the lockdown effective at the local and regional level? How to assist different regions of England to develop targeted localised plans? There are crucial questions that we are aiming to answer in this work so that past experience in NPIs could be useful to guide the government policymaking to tackle the challenges of future outbreaks. The contributions of this work are presented as follows:

- We select R_t as an indicator to display the trend of the epidemic, which is a better way to show the growth or decline of the epidemic than the number of cases. In this way, we can evaluate the effectiveness of previous NPI policies by observing the change of R_t (e.g. whether it is less than 1).
- We reveal the spatial-temporal relationships between the epidemic situation and human mobility in nine regions, which can be implemented to assist each region to make localised NPIs.
- We compare correlation differences with different policy periods to understand the impact of the NPIs and the rebound of the outbreak in wave 2.

The remainder of this chapter proceeds as follows. Section 12.2 presents an overview of the study area and the data used in this study, which is followed by the methodology introduced in Sect. 12.3. The methodology consists of the calculation of R_t and its correlational analysis with the mobility data. The analysis results are presented in Sect. 12.4, which highlights the association of local mobility with the R_t . The difference between the first wave and second wave outbreaks is also discussed in Sect. 12.4, which leads to the summary and suggestions for local policy in Sect. 12.5.

12.2 Study Area and Data Description

The study area of this analysis covers all of the nine regions of England (Fig. 12.1), namely, South East, London, North West, East of England, West Midlands, South West, Yorkshire and the Humber, East Midlands, and North East. England is one of the four countries collectively making up the UK, which shares land borders with Wales to its west and Scotland to its north. According to ONS (2020), in 2019, over 56 million population (i.e. about 84% of the total population in the UK) resided in England.

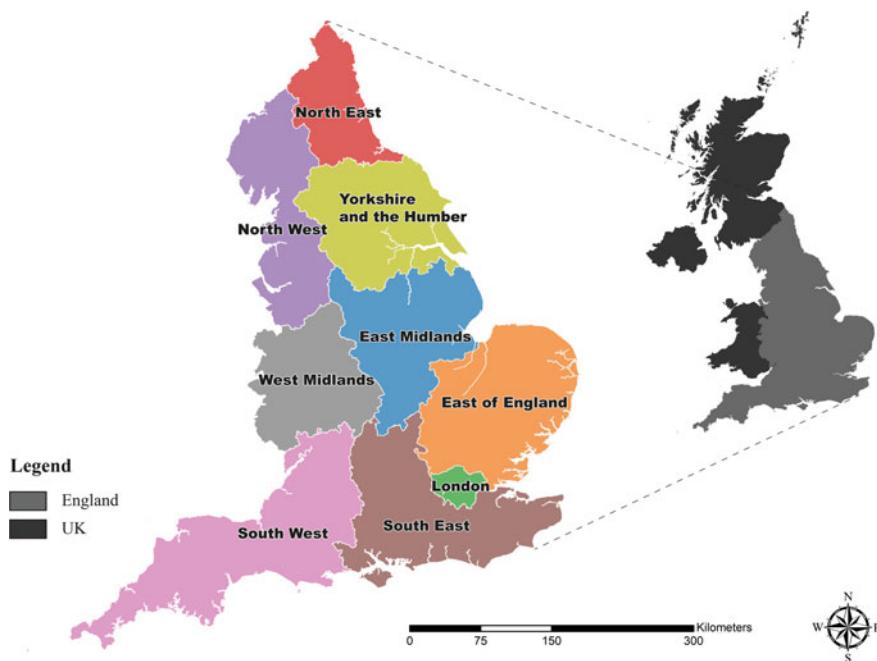


Fig. 12.1 Map of study area: nine regions in England

The daily confirmed COVID-19 data were extracted from the official coronavirus data website¹ published by the UK government. Figure 12.2 shows the number of daily cases for each region in England during the period from 27-02-2020 to 02-09-2020. According to the line diagram, generally, the number of cases in England increased rapidly from the end of February to the beginning of April, peaked in April, began to decline after May, and rebounded again from July to September. More specifically, from the end of February to June in 2020, the number of daily confirmed cases of all regions increased first and then decreased, with more cases confirmed in London, North West and South East. Moreover, London's peak number appeared the earliest (at the end of March), and this peak value was the largest (more than 1000). The second-largest peak value was from North West, which appeared at the end of April. From June to mid-July, the epidemic situation in all regions of England was relatively stable. However, there was a second upward trend in the number of cases in England from mid-July to early September, especially in North West and Yorkshire and The Humber.

In order to deepen the understanding of what has changed in answer to policies aimed at combating COVID-19, Google published the Community Mobility Reports (CMR)² to provide the movement trends over time by geography. These datasets show

¹<https://coronavirus.data.gov.uk/>.

²<https://www.google.com/covid19/mobility/>.

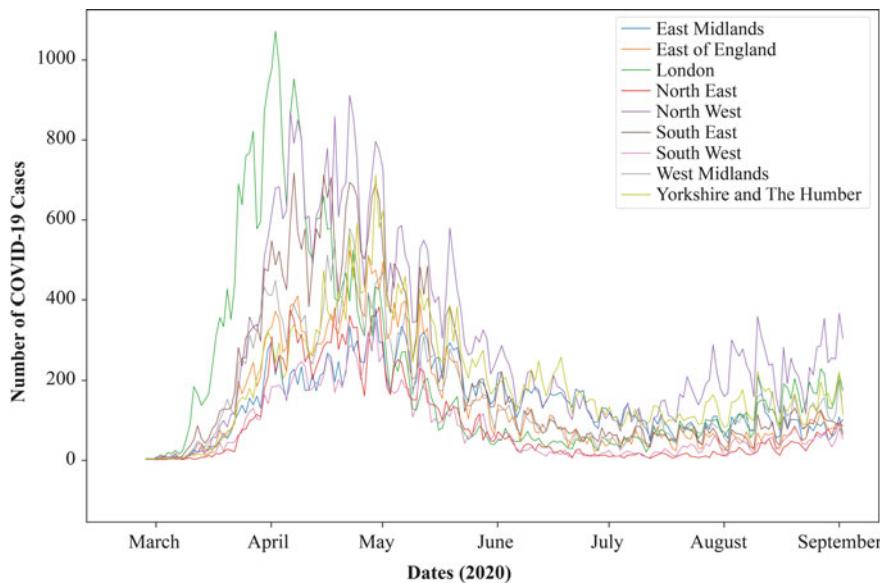


Fig. 12.2 Daily COVID-19 cases in nine regions (27/02/2020—02/09/2020)

how visits and length of stay at different places change for each day compared to a baseline value (baseline is the median value between 03/01/2020 and 06/02/2020). Table 12.1 provides some examples of community mobility data in London.

For the place categories, (1) *Grocery and Pharmacy* provide information on mobility trends for places like grocery markets, food warehouses, farmers markets, specialty food shops, drug stores, and pharmacies; (2) *Parks* provide information on mobility trends for places like local parks, national parks, public beaches, marinas, dog parks, plazas, and public gardens; (3) *Transit Stations* provide information on mobility trends for places like public transport hubs such as subway, bus, and train stations; (4) *Retail and Recreation* provide information on mobility trends for places like restaurants, cafes, shopping centres, theme parks, museums, libraries, and

Table 12.1 Example of Google community mobility data in London

Date	Retail and recreation	Grocery and pharmacy	Parks	Transit stations	Workplaces	Residential
2020-02-24	-4.706	-0.441	-7.118	-2.706	-0.794	1.029
2020-02-25	-1.059	3.147	3.324	-0.235	1.118	0.029
2020-02-26	-0.765	0.529	6.206	-2.676	1.382	0.088
...
2020-08-29	-34.588	-16.941	22.824	-38.676	-18.559	4.206
2020-08-30	-28.176	-17.882	61.5	-33.059	-14.559	1.059

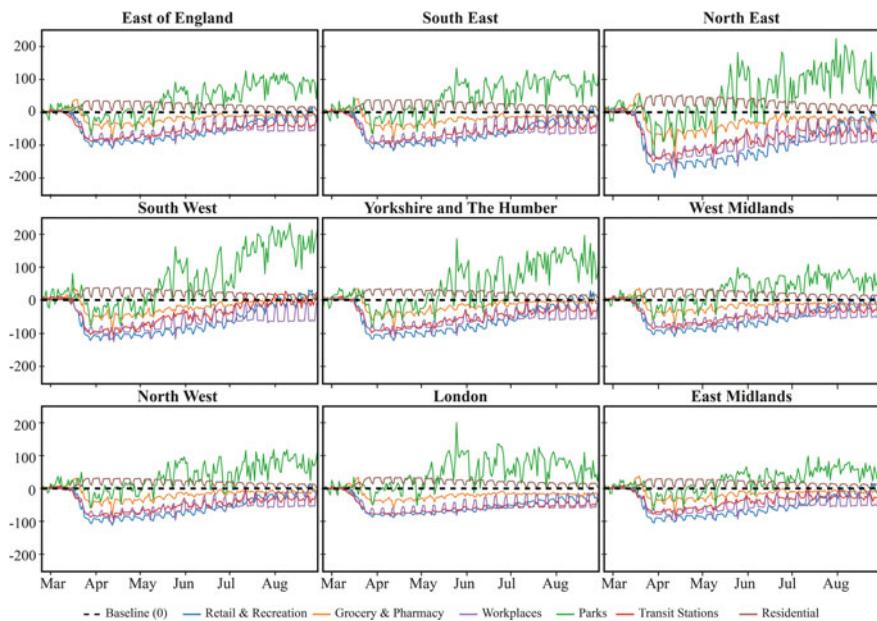


Fig. 12.3 Google mobility data for each region in England (27/02/2020—02/09/2020)

movie theatres; (5) *Residential* provide information on mobility trends for places of residence; (6) *Workplaces* provide information on mobility trends for places of work.

For the mobility data of England from Google, the temporal resolution is on a daily level, while the spatial resolution is at the county level. However, such spatial resolution does not in accord with the COVID-19 case data published by the UK government, which is available at the local authority level. In order to make the two data to be comparable, both of them were aggregated into the regional level based on their geographic location.

A series of line diagrams in Fig. 12.3 respectively illustrate the relative daily changes (%) of human mobility patterns derived from the Google Mobility Report data in nine regions of England from the middle of February to the end of August. Since the outbreak of COVID-19 in March, the residential data in all regions in England have stayed above baseline, indicating that people tend to stay at home to prevent themselves from being infected. The data of *retail and recreation*, *transit stations* and *workplaces* showed a downward trend and then an upward trend, then they restored to around the baseline from July to August. This phenomenon indicates that after the COVID-19 outbreak, people have reduced their visits to places of entertainment and the use of public transportation and are more likely to be working at home rather than go to workplaces. In July and August, with the policy of releasing lockdown, people have returned to work and entertainment, and have taken public transport, to promote economic recovery. The *grocery and pharmacy data* increased in the first ten days of March (panic shopping), then decreased (less frequent than

normal), and then returned to baseline in July and August. In addition, the *park data* also showed a downward trend and then an upward trend, but the volatility was sharper than others.

12.3 Methodology

12.3.1 Real-Time Effective Reproduction Ratio (R_t)

Generally, people within the epidemic scope of infectious diseases are divided into different categories with labels SEIR as follows (Kermack and McKendrick 1927):

- Susceptible (S), refers to those who do not have the disease, but lack immunity, and are easy to be infected after contact with the infected person.
- Exposed (E), refers to those who have been exposed to infection but are unable to infect others temporarily, which is applicable to infectious diseases with a long incubation period.
- Infectious (I), refers to the infected person, who can be transmitted to class S members to become class E or class I members.
- Recovered (R), refers to the person who is isolated or has immunity due to illness. If the immune period is limited, class R members can be changed into class S again.

Suppose that at each time step, an exposed individual would turn to be an infected individual with probability γ_1 , and an infected would be removed with probability γ_2 . The value of R_0 can be approximated as defined in Eq. 12.1 (Zhou et al. 2020):

$$R_0 = 1 + \lambda T_g + \rho(1 - \rho)(\lambda T_g)^2 \quad (12.1)$$

where $\lambda = \ln(Y_t)/t$ is the growth rate of the early exponential growth and Y_t is the number of infections with symptoms as of time t . $T_g = T_E + T_I$ is the generation time, where $T_E = \frac{1}{\gamma_1}$ indicates the exposure period and $T_I = \frac{1}{\gamma_2}$ indicates the infection period. $\rho = \frac{T_E}{T_g}$ is the proportion of exposure period to generation time.

The effective reproduction Ratio (R_t) is used for real-time analysis, which is the average number of new cases caused by an infected individual at time t . In order to estimate the R_t of COVID-19 in real-time, Systrom (2020) proposed a Bayesian approach, with Gaussian noise to estimate a time-varying R_t . The algorithm of real-time R_t is adopted here to assess the spatial-temporal difference of COVID-19 for England at a regional level.

In order to elaborate the algorithm in detail, the calculation of London's R_t will be taken as an example. First of all, we use the number of daily cases data downloaded from the government website of coronavirus in the UK as the input of this algorithm. The Gaussian filter is then applied to the time series to obtain the best view of the

'true' data as possible as we can. This is because the real-world process is not nearly as stochastic as the actual reporting (Systrom 2020).

After that, the posteriors are calculated by the following steps (Systrom 2020):

1. Calculate λ , the expected arrival rate of new cases per day

Given an average arrival rate of λ new cases per day, the probability of seeking k new cases is distributed according to the Poisson distribution (Systrom 2020):

$$P(K|\lambda) = \frac{\lambda^K e^{-\lambda}}{k!} \quad (12.2)$$

However, in order to look for $P(K|R_t)$ which is parameterised by R_t , it is necessary to seek the relationship between λ and R_t . According to the derivation by Bettencourt and Ribeiro (2008), the λ can be calculated as:

$$\lambda = k_{t-1} e^{\gamma(R_t - 1)} \quad (12.3)$$

where γ is the reciprocal of the serial interval, which is about seven days for COVID-19 from Centres for Disease Control and Prevention (Sanche et al. 2020); k_{t-1} is the number of new cases of the previous day.

2. Calculate each day's likelihood distribution of R_t

Therefore, the likelihood function can be reformulated as a Poisson parameterised by fixing k and varying R_t (Systrom 2020):

$$P(K|R_t) = \frac{\lambda^K e^{-\lambda}}{k!} \quad (12.4)$$

3. Get the posteriors and perform the Bayesian update

In order to get the posteriors, it is necessary to multiply the likelihood by the prior by using the cumulative product of each successive day. During this process, the Bayes' rule is applied iteratively to use information from both previous and current days and effectively averages it. For instance, based on the London cases shown in Fig. 12.2, the daily posteriors for R_t in London can be generated (Fig. 12.4). As shown in Fig. 12.4, the most likely value of real-time R_t for London is the index corresponding to the largest posterior.

12.3.2 Maximal Information Coefficient

Reshef et al. (2011) introduce the Maximal Information Coefficient (MIC), which is a new measure of dependence. At the same time, Reshef et al. (2011) introduce

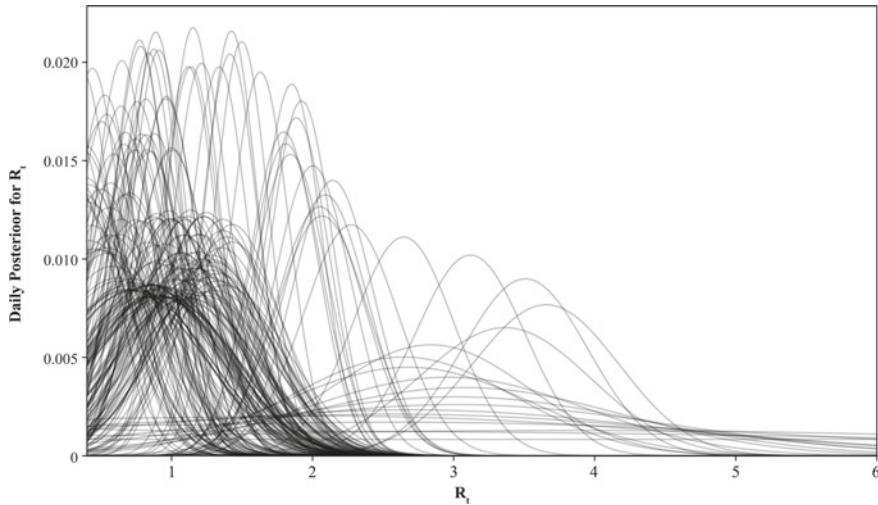


Fig. 12.4 Daily posterior for R_t in London

equitability: a measure of dependence is said to be equitable if it gives similar scores to relationships with similar noise levels, and show that MIC is highly equitable comparing with other current methods (Reshef et al. 2013). The Maximal Information Coefficient (MIC) is defined as in Eq. 12.5:

$$MIC(D) = \max_{xy < B(|D|)} \frac{I^*(D, x, y)}{\log_2 \min\{x, y\}} \quad (12.5)$$

where D is a set of ordered pairs. For a grid G , let $D|_G$ denote the probability distribution induced by the data D on the cells of G , and $I(\cdot)$ denote mutual information; $I^*(D, x, y) = \max_G I(D|_G)$, where the maximum is taken overall x -by- y grids G (possibly with empty rows/columns); B is a growing function satisfying $B(n) = o(n)$ and $B(n) = n^{0.6}$ is suggested as a default setting.

MIC is on the interval $[0, 1]$, where 0 indicates independence and 1 indicates a noiseless functional relationship. For nonlinear relationship or dependency, Reshef et al. (2013) proved that the MIC method performs better than other methods checking the correlation between variables, such as the Spearman rank correlation coefficient and Pearson correlation, which are not equitable and show a strong preference for some types of functions. Taking this point into consideration, the MIC method was selected to determine the nonlinear relationships between mobility data and R_t data.

Considering the incubation period of COVID-19 is around 7 days, and the isolation periods for most governments is two weeks, it is widely accepted that the impact of lockdown needs to be seen not immediately but over weeks, to be reflected in the reduced cases. To understand such delayed impact of lockdown, in particular the delayed impact of different types of mobility/activities, we have calculated the

correlation between each type of community mobility data and the real-time R_t value with different time delays. The time delay is set to 3 days, 5 days, 7 days, 9 days, 12 days, 15 days, 18 days and 21 days in order to understand the delayed impact of different mobility on R_t to what extent. Then, each type of mobility will have the most appropriate time delay according to the corresponding maximum correlation coefficient.

12.4 Results and Discussion

12.4.1 Spatial-Temporal Distribution of the Real-Time R_t in England

Figure 12.5 shows the results of the real-time R_t for each region in England from 27/02/2020 to 02/09/2020. In this figure, the points in red represent the $R_t > 1$, indicating the number of cases will increase, such as at the start of an epidemic; the grey points show the $R_t = 1$, indicating the disease is endemic; and the black points indicate the $R_t < 1$, implying that there will be a decline in the number of cases.

On the whole, the spatial-temporal variations of the R_t in regions of England have a certain similarity. That is to say, the epidemic broke out in March and April, tended to be stable and relatively safe in May and June, and rebounded in July and August. Generally speaking, this result is in line with government control measures. After

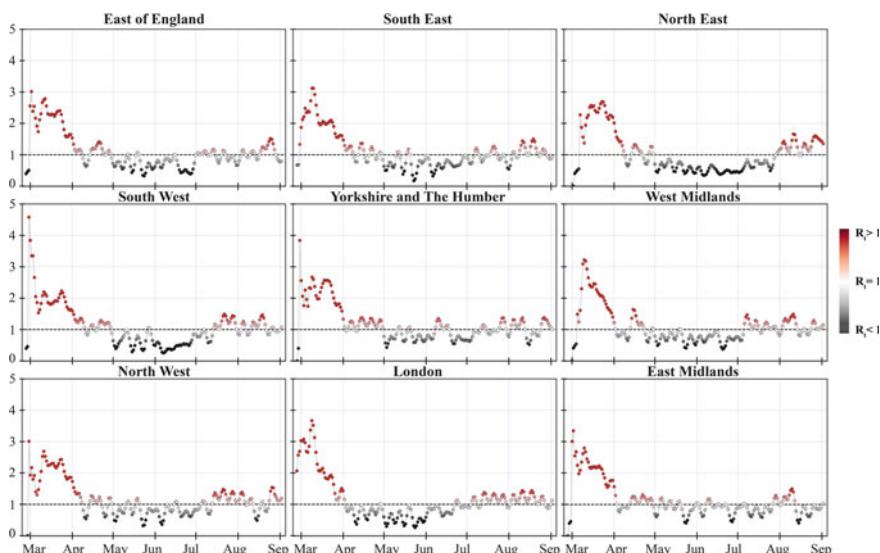


Fig. 12.5 Real-time R_t for nine regions in England

the outbreak of the epidemic of COVID-19 in March, the UK government adopted a series of lockdown measures during late March to mid-June, including but not limited to keeping social distance, closing restaurants, bars and other entertainment places, closing schools, requiring people to work at home and reducing going out. These measures promoted people to self-isolate at home and reduced contact with each other, so as to cut off the route of infection and improve the epidemic situation. However, with the lifting of lockdown in early July, the government launched a series of measures in order to promote economic recovery, such as the ‘Eat Out to Help’ Scheme in August (HM Revenue and Customs 2020), which could make the epidemic rebounds again from pubs, restaurants, factories and farms (The Guardian 2020). For example, if a customer of a restaurant is infected, then the epidemic will break out again in the range of people related to this restaurant.

However, Fig. 12.5 also reveals the differences in R_t between regions of England. For instance, from May to July, the epidemic situation in the North East region was well controlled, and the specific performance was that the R_t value was always less than 1 during this period. Additionally, the second rebound of the North East region occurred in August, slightly later than other regions. However, the R_t value of the East Midlands, North West, and Yorkshire and The Humber regions still fluctuated around 1 during this May–July, indicating that the epidemic situation in these regions has not been well controlled. Furthermore, for the first outbreak of COVID-19, London region is relatively good at controlling the epidemic, because the R_t value of London dropped quickly to less than 1 in April.

12.4.2 Relationships Between Mobility and R_t

Figure 12.6 shows the Maximal Information Coefficient (MIC) between the delayed mobility data and R_t of each region in England. It can be seen that R_t has a weak relationship with people going to parks in England, which may be due to parks are big enough for people to keep the social distance. For most regions except London and North East, R_t has moderate relationships with mobility data of retail and recreation, grocery and pharmacy, transit stations, workplaces and residential. As for London and North East, mobility of retail and recreation has the strongest relationship with R_t , followed by the mobility of transit stations and workplace.

In general, R_t is closely related to mobility subject to *retail and recreation*, *transit stations*, and *workplaces*, where it is difficult for people to maintain social distancing since these places are usually overcrowded while the space is small. Moreover, people were unlikely to wear masks in recreation places such as restaurants and bars, which exacerbated the epidemic of COVID-19. In addition, there is also a certain relationship between residential mobility and R_t since the self-isolation at home can reduce the risk of infection. Table 12.2 shows the maximum value of MIC for the correlations between R_t and the mobility types with the corresponding time delay in days. On average, dayled 12 days has strongest correlation between the R_t and mobility data which does reflect the delayed impact of lockdown. For all the mobility

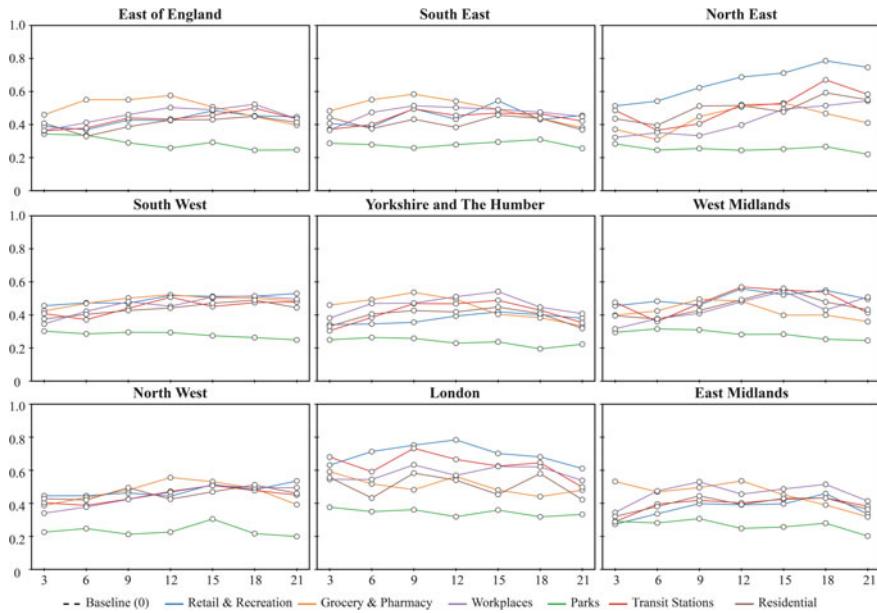


Fig. 12.6 MIC between delayed mobility data and R_t for each region in England

types, London reached the maximum correlation with relatively shorter time delays which means that the lockdown has better controlling impact in London than in other regions. This is also reflected in Fig. 12.5 that London's R_t drop to 1 quickest among all regions in early April, but South West, North East and Yorkshire and The Humber took longer to reach the maximum coefficient between the mobility and the R_t , which took longer to drop the R_t below 1 till early May in the first wave of the outbreak.

For the MIC value between the *retail and recreation* mobility and R_t , the maximum value is up to 0.8 in London and North East, indicating a very strong relationship; while MIC values of other regions are around 0.5–0.6, indication moderate relationships and similar as results of *grocery and pharmacy*, *workplaces* and *residential*.

In addition, there is a strong relationship between *transit stations* mobility and R_t in London with the MIC value greater than 0.7. This may be because people's daily travel in London is mainly dependent on public transportation. However, people were not required to wear masks on the tube in the early period, and it was difficult for people to keep social distance on the tube, especially in the morning and evening rush hours. These reasons led to a high correlation between *transit* mobility and R_t . This implies the importance of keeping social distance on public transit in London.

In general, the maximum MIC values between the retail and recreation mobility and R_t in London and North East are nearly 0.8, and the MIC values of each type of mobility and R_t in London are generally higher than those in other regions as well. We think the high correlation between mobility data and R_t in London is related to

Table 12.2 The maximum MIC with the corresponding time delays

	Retail and recreation	Grocery and pharmacy	Parks	Transit stations	Workplaces	Residential
	Max MIC	Time delay	Max MIC	Time delay	Max MIC	Time delay
East Midlands	0.459	18	0.535	12	0.307	9
East of England	0.484	15	0.576	12	0.343	3
London	0.784	12	0.592	3	0.377	3
North East	0.786	18	0.531	15	0.283	3
North West	0.535	21	0.556	12	0.305	15
South East	0.544	15	0.584	9	0.310	18
South West	0.530	21	0.522	12	0.302	3
West Midlands	0.559	12	0.495	9	0.316	6
Yorkshire and The Humber	0.419	15	0.547	9	0.264	6

the high population density in London to some extent. In other words, in areas with a larger population and more frequent activities, the correlation between real-time R_t and mobility data is more significant.

12.4.3 Comparison of the First and Second Waves

In order to compare the first outbreak and the second rebound of COVID-19, London and North East are selected as examples for analysis by splitting time. This is because, during the relatively mild period of the epidemic in England from the beginning of April to the end of June, R_t of London and North East remained stable below 1. In addition, we add some specific dates of lockdown policies for analysis:

- March 23rd 2020: Lockdown
- May 13th 2020: The first step for leaving lockdown:
 - (1) Encourage people who cannot work from home to return work;
 - (2) Remove the restriction that everyone goes out once a day;
 - (3) Keep 2 m when going out and abide by strict social isolation measure.
- June 1st 2020: The second step for leaving lockdown:
 - (1) Most sports events will be resumed;
 - (2) Restrictions on people's going out for exercise will be relaxed;
 - (3) Almost all primary schools will open.
- July 4th 2020: The third step for leaving lockdown:
 - (1) Reopen bars and restaurants;
 - (2) Shorten social distance and change 2 m to 1 m plus.

Figure 12.7 presents the time split of real-time R_t in London. The first outbreak occurred in London between February 27th and April 2nd with $R_t > 1$. The trend of the epidemic in London was relatively stable from April 3rd to June 22nd with $R_t < 1$. However, the epidemic in London had the second rebound since 23 June with $R_t \geq 1$ again.

As shown in Fig. 12.7, after lockdown, London's R_t value decreased and dropped below 1, indicating the lockdown policy was effective. After the first step of leaving lockdown, this value was still below 1. However, 23 days after the second step, it began to rebound to more than 1 and continued to be greater than 1 after the third step.

Table 12.3 shows the result of the Maximal Information Coefficient (MIC) between mobility data and R_t during three periods in London. The first outbreak of COVID-19 in London (the first period in Fig. 12.8 that R_t is greater than 1) is highly related to the mobility of *retail and recreation, grocery and pharmacy, transit stations, workplaces and residential*. The second rebound of the epidemic in London is related to mobility data of *retail and recreation, transit stations and workplace*,

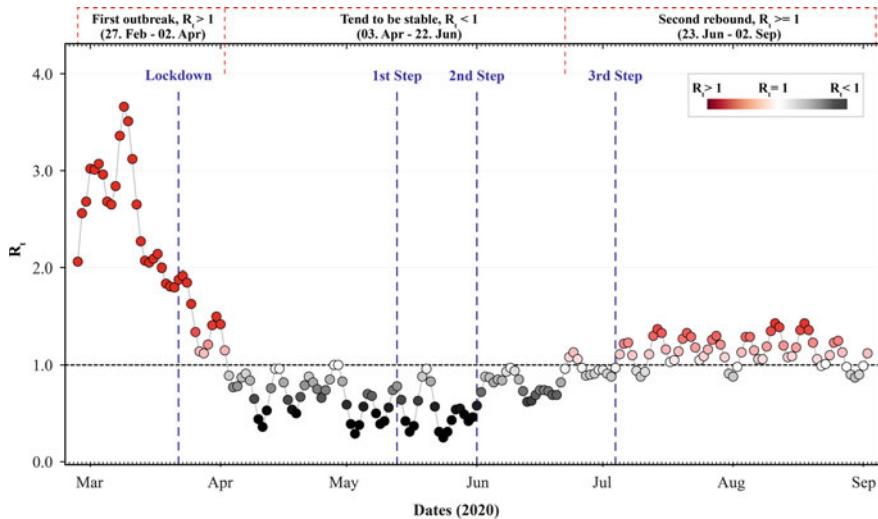


Fig. 12.7 Time split of R_t in London

Table 12.3 Maximal information coefficient during three periods in London

Period	2.27–4.02	4.3–6.22	6.23–9.02
Retail and recreation (Time delay: 12)	0.727	0.220	0.328
Grocery and pharmacy (Time delay: 3)	0.811	0.275	0.269
Parks (Time delay: 3)	0.306	0.248	0.201
Transit stations (Time delay: 9)	0.811	0.234	0.398
Workplaces (Time delay: 9)	0.811	0.232	0.315
Residential (Time delay: 9)	0.811	0.309	0.302

while the MIC values are lower than those of the first outbreak. From the perspective of decision-making, it is suggested that people in London should wear masks and keep the social distance in both public transportation and offices; and should reduce the frequency of going to retail and recreation places.

Figure 12.8 shows the time split of real-time R_t in North East. It can be seen that the first outbreak occurred between March 2nd and April 30th with $R_t > 1$. The trend of the epidemic in North East was relatively stable from May 1st to July 29th with $R_t < 1$. However, the epidemic had the second rebound since July 30th with $R_t \geq 1$ again.

As shown in Fig. 12.8, after lockdown, the R_t value in North East began to decrease and dropped below 1, indicating the lockdown policy was effective. After the first step and second step of leaving lockdown, this value was still below 1. However, nearly one month after the third step, it began to rebound to more than 1 and continued to be greater than 1.

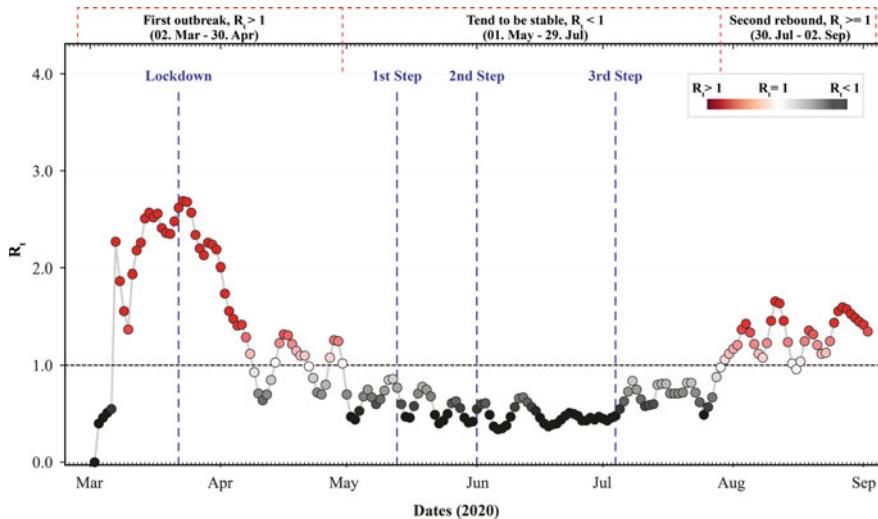


Fig. 12.8 Time split of R_t in North East

The result of the MIC values between mobility data and R_t during three periods in the North East is shown in Table 12.4. It can be seen that there are very strong relationships during the first outbreak of COVID-19 and mobility data of retail and recreation, grocery and pharmacy, transit stations, workplaces and residential, which is similar to the situation in London. However, the second rebound of the epidemic in the North East is mainly related to mobility for *workplaces*. That is to say, for the second rebound of COVID-19 in the North East, people in this region should pay special attention to the risk of infection in offices. From the perspective of decision-making, the government of the North East may be able to ask people to work from home as much as possible. In addition, people in this region need to do epidemic prevention and self-protection in entertainment places and public transport as well.

Table 12.4 Maximal information coefficient during three periods in the North East

Period	3.02–4.30	5.01–7.29	7.30–9.02
Retail and recreation (Time delay: 18)	0.846	0.480	0.356
Grocery and pharmacy (Time delay: 15)	0.935	0.248	0.299
Parks (Time delay: 3)	0.317	0.223	0.267
Transit stations (Time delay: 18)	0.905	0.295	0.359
Workplaces (Time delay: 21)	0.900	0.428	0.484
Residential (Time delay: 18)	0.966	0.340	0.424

12.5 Conclusions and Recommendations

With COVID-19 pandemic sweeping the world, it is necessary for countries and regions to plan effective NPIs to control the virus transmission. As human activities will spread the virus to a variety of places, it is meaningful to study the correlation between community mobility and epidemic situations. How to evaluate previous NPIs and plan further NPIs have become a priority issue for decision-making.

In this study, we aim to reveal the spatial-temporal relationships between the community mobility and R_t , which is an important monitoring index of epidemic situation, and provide suggestions for decision-making facing governments and travel plan facing people.

First of all, we computed the R_t of COVID-19 to show the trend of the pandemic and evaluate previous NPIs. For instance, if the value of R_t continues to decline or is less than 1 for a long time, the previous NPIs are proved to be effective. Then, we analysed the relationships between different types of mobility data and R_t by using the Maximal Information Coefficient (MIC). It shows that there is a weak correlation between the spread of the virus and parks, and people can choose parks as places for leisure and exercise during the outbreak. In addition, public transportation, recreation places, pharmacies and workplaces are relatively highly related to the spread of the epidemic, and people should pay attention to epidemic prevention in these places such as wearing masks and keeping the social distance.

We also compared the regional difference of R_t in both the first and second wave of the outbreak of the pandemic in London and North East. It shows that the lock-down policy during the first outbreak is effective in these areas; the second wave in London is related to public transit and recreation places, while that of North East is related to relatively high activity in the workplace. Such findings could provide useful guidelines to the local government which type of control measures they should take to contain the virus.

One of the major limitations of this study relates to the spatial resolution selected to conduct the analysis. The spatial resolution used in this work is at the regional level, which is relatively coarse and therefore might impose some problems subject to the Modifiable Areal Unit Problem (MAUP) (Openshaw and Taylor 1981) since the aggregation process. For instance, even when the R_t of a region is calculated below 1, some certain local areas in this region may have higher R_t values that are likely to exceed 1. Therefore, one direction of future work that would be favourable to the quality of value of the outcomes is to select the finer spatial resolution to conduct this analysis under then permission conditions of data availability. Nonetheless, despite such caveat, this paper has presented a promising attempt to explore the relationship between community mobility and R_t of COVID-19, which is implemented to reveal how NPIs influence the pandemic situation and human activities, and will likely be a useful exploration for applications within other contexts.

Acknowledgements This work was supported by the UKRI Medical Research Council: COVID-19 Rapid Response: Virus Watch: Understanding Community incidence, Symptom profiles, and Transmission of COVID-19 in relation to Population Movement and Behaviour (MC_PC_19070).

References

- Anderson, R., & May, R. (1992). *Infectious diseases of humans*. Oxford University Press.
- Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3).
- BBC. (2020). *Covid-19: Prime minister says UK “seeing a second wave”*. <https://www.bbc.co.uk/news/av/uk-54213129>.
- Bettencourt, L. M. A., & Ribeiro, R. M. (2008). Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE*, 3(5). <https://doi.org/10.1371/journal.pone.0002185>.
- Bonacorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.2007658117>.
- Cabinet Office. (2020). *New national restrictions from 5 November*. <https://www.gov.uk/guidance/new-national-restrictions-from-5-november>.
- Chan, J. F. W., Yuan, S., Kok, K. H., To, K. K. W., Chu, H., Yang, J., et al. (2020). A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: A study of a family cluster. *The Lancet*. [https://doi.org/10.1016/S0140-6736\(20\)30154-9](https://doi.org/10.1016/S0140-6736(20)30154-9).
- Cintrón-Arias, A., Castillo-Chavez, C., Bettencourt, L. M. A., Lloyd, A. L., & Banks, H. T. (2009). The estimation of the effective reproductive number from disease outbreak data. *Mathematical Biosciences and Engineering*. <https://doi.org/10.3934/mbe.2009.6.261>.
- Fine, P., Eames, K., & Heymann, D. L. (2011). “Herd immunity”: A rough guide. *Clinical Infectious Diseases*. <https://doi.org/10.1093/cid/cir007>.
- Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., et al. (2009). Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science*. <https://doi.org/10.1126/science.1176062>.
- Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Doepfer, D., Sethi, A. K., Reyes, J. F. M., Patz, J., & Yandell, B. S. (2020). Mobile phone location data reveal the effect and geographic variation of social distancing on the spread of the COVID-19 epidemic. arXiv.
- Government Office for Science. (2020). *The R number and growth rate in the UK*. <https://www.gov.uk/guidance/the-r-number-in-the-uk>.
- HM Revenue & Customs. (2020). *Guidance: Get a discount with the Eat Out to Help Out Scheme*. <https://www.gov.uk/guidance/get-a-discount-with-the-eat-out-to-help-out-scheme>.
- Kapoor, A., Ben, X., Liu, L., Perozzi, B., Barnes, M., Blais, M., & O’Banion, S. (2020). Examining COVID-19 forecasting using spatio-temporal graph neural networks. arXiv.
- Kermack, W., & McKendrick, A. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*. <https://doi.org/10.1098/rspa.1927.0118>.
- Kraemer, M. U. G., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., et al. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. <https://doi.org/10.1126/science.abb4218>.
- Nishiura, H. (2010). Correcting the actual reproduction number: A simple method to estimate R_0 from early epidemic growth data. *International Journal of Environmental Research and Public Health*. <https://doi.org/10.3390/ijerph7010291>.
- Office for National Statistics. (2020). *Population estimates for the UK, England and Wales, Scotland and Northern Ireland: mid-2019*. <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/bulletins/annualmidyearpopulationestimates/mid2019estimates>.
- Openshaw, S., & Taylor, P. J. (1981). The modifiable areal unit problem. *Quantitative Geography: A British View*. <https://doi.org/10.1002/9781118526729.ch3>.

- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., McVean, G., Turnbaugh, P. J., et al. (2011). Detecting novel associations in large data sets. *Science*. <https://doi.org/10.1126/science.1205438>.
- Reshef, D., Reshef, Y., Mitzenmacher, M., & Sabeti, P. (2013). *Equitability analysis of the maximal information coefficient, with comparisons*, 1–22. <https://arxiv.org/abs/1301.6314>.
- Sanche, S., Lin, Y., Xu, C., Romero-Severson, E., Hengartner, N., & Ke, R. (2020). High contagiousness and rapid spread of severe acute respiratory syndrome coronavirus 2. *Emerging Infectious Diseases*, 26(7), 1470–1477. <https://doi.org/10.3201/eid2607.200282>.
- Smith, R. J., Li, J., & Blakeley, D. (2011). The failure of R₀. *Computational and Mathematical Methods in Medicine*. <https://doi.org/10.1155/2011/527610>.
- Systrom, K. (2020). *Estimating COVID-19's R_t in real-time*. <https://github.com/k-sys/COVID-19/blob/master/RealtimeR0.ipynb>.
- The Guardian. (2020). “*Eat out to help out*” may have caused sixth of Covid clusters over summer. <https://www.theguardian.com/business/2020/oct/30/treasury-rejects-theory-eat-out-to-help-out-caused-rise-in-covid>.
- Wang, H., & Yamamoto, N. (2020). Using a partial differential equation with Google Mobility data to predict COVID-19 in Arizona. *Mathematical Biosciences and Engineering*. <https://doi.org/10.3934/mbe.2020266>.
- World Health Organization. (2020a). *Archived: WHO timeline—COVID-19*. <https://www.who.int/news/item/27-04-2020-who-timeline--covid-19>.
- World Health Organization. (2020b). *WHO coronavirus disease (COVID-19) dashboard*. <https://covid19.who.int/>.
- Zhou, T., Liu, Q., Yang, Z., Liao, J., Yang, K., Bai, W., et al. (2020). Preliminary prediction of the basic reproduction number of the Wuhan novel coronavirus 2019-nCoV. *Journal of Evidence-Based Medicine*. <https://doi.org/10.1111/jebm.12376>.

Chapter 13

Exploring Store Visit Changes During the COVID-19 Pandemic Using Mobile Phone Location Data



Yunlei Liang, Kyle W. McNair, Song Gao, and Ashlıgül Göçmen

13.1 Introduction

Day-to-day life has been dramatically changed by the novel coronavirus disease (COVID-19) pandemic. As of December 15th, 2020, there are over 70,000,000 confirmed cases of COVID-19 with over 1,600,000 deaths across the world (WHO 2020). With many researchers are working on finding effective treatments to control this disease, many federal governments and institutions have provided a wide-ranging set of non-pharmaceutical interventions and policies to slow down the spread of COVID-19; some of these policies have aimed at reducing travel flows and close contacts between individuals (Aloi et al. 2020; Lai et al. 2020; Gao et al. 2020; Painter and Qiu 2020).

In the United States, state and local governments have similarly enacted social distancing policies to limit contacts between individuals. These policies have included travel restrictions, closures of schools and nonessential businesses, bans

Y. Liang · K. W. McNair · S. Gao (✉)

GeoDS Lab, Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA

e-mail: song.gao@wisc.edu

Y. Liang

e-mail: yunlei.liang@wisc.edu

K. W. McNair

e-mail: kwmcnair@wisc.edu

A. Göçmen

Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA

Nelson Institute for Environmental Studies, University of Wisconsin-Madison, Madison, WI 53706, USA

A. Göçmen

e-mail: gocmen@wisc.edu

on large social gatherings, restrictions for indoor dining at restaurants, and statewide mask and stay-at-home orders (Courtemanche et al. 2020; Hong et al. 2020). People have changed their movements and store visit behaviors in response to those measures (Chang et al. 2020; Gao et al. 2020; Huang et al. 2020; Pan et al. 2020).

Google published the COVID-19 Community Mobility Reports to help the public understand how individuals' movement, as measured by visits to workplaces and popular destinations including parks, restaurants, and grocery stores, have changed in response to the lockdown policies (Aktay et al. 2020). Based on those reports, a few studies analyzed the impact of lockdown on community mobility and identified different trends in visits to different places during pre-lockdown and after lockdown periods (Mohler et al. 2020; Saha et al. 2020; Chetty et al. 2020). From those studies, it has been found that the behavioral changes are not uniform across places as the reduction in mobility is different for different types of places (Chang et al. 2020) and also in different countries (McKenzie and Adams 2020). Therefore, it is critical to further investigate the extent to which and how people's movements have changed. Some research studies have focused on discovering disparities in the movement patterns and related the change of mobility to different types of places people visit and different geographic areas or social groups people belong to. Li et al. (2020) used the origin-destination networks to understand how visits from Census Block Group to different types of urban hotspots have changed differently and how the reduction in movement varies across multiple U.S. cities. Another study by Hong et al. (2020) further combined localized socioeconomic, demographic and infrastructure information to discover whether the behavioral responses are related and affected by such characteristics. The researchers were able to identify distinct visit patterns in response to COVID-19 across neighborhoods and communities with different characteristics (Hong et al. 2020). In addition, Holtz et al. (2020) found that the mobility patterns of people in one area are greatly affected by the lockdown orders in its socially or geographically connected peer states. Those quantitative analyses provide insights into how lockdown policies affect mobility patterns from different perspectives. However, we still do not know much about how mobility patterns as measured by visits vary among particular stores and across different cities. Our work presented in this chapter explores this question about variability of mobility patterns among particular chain stores and across different cities in the U.S.

We use the definition of a brand as "a logo or branded store which has multiple locations all under the same logo or store banner" following the definition from our data provider (SafeGraph 2020a). The store is a specific location of a certain brand; in our work, the "store" can be considered as a retail establishment. In this study, we use large-scale mobile phone location data to analyze customer visits to five popular chain-store brands in pre-lockdown and after lockdown periods across three large U.S. cities to examine any variation among the visits to different places. We employ a time-aware dynamic Huff (T-Huff) model to estimate the visit probability from a particular Census Block Group to a chain-store brand. The traditional Huff model is typically used to estimate the trade area of a certain store (Huff 1963). A few studies have used the Huff model to delineate the trade area and calibrated it using data collected from user surveys or social media data (Suárez-Vega et al.

2015; Wang et al. 2016). For example, Wang et al. (2016) collected the social media posts to extract user samples and used that as the input to the Huff model for trade area delimitation. However, most of the Huff models are static, while the store visit patterns have temporal variations. The T-Huff model used in this study was proposed by Liang et al. (2020) in order to provide a second dimension of information (time) for trade area delineation besides the spatial dimension. By incorporating the temporal and spatial information in modeling people's visit patterns, the T-Huff model can capture the dynamics of visits and reflect the visit behaviors via the parameters of the model with higher accuracy than the traditional Huff model (Liang et al. 2020; McKenzie et al. 2015).

The contribution of this study is threefold: (1) We provide location business insights into the visit patterns by analyzing the median travel distance to each store and also by comparing the customer dwell time (the time spent in a store) distribution in different time periods. (2) We analyze the impacts of lockdown policies on store visits by comparing spatial distribution of visits using origin-destination mobility networks. (3) We are able to discover different visit patterns to different brands and identify regional variation in such visits using the T-Huff model.

The chapter is organized as follows. We first introduce the proposed T-Huff model in the methods Sect. 13.2, which is followed by the data and study area in Sect. 13.3. We then present comparative analyses, results and discussion in Sect. 13.4. Finally, we conclude the study and share some thoughts for future work in Sect. 13.5.

13.2 Methods

The original Huff model is primarily used to delineate the trade area of a store, which is an area containing potential customers (Huff 1963). It can estimate the visiting probability from one customer to a particular store with the assumption that this visiting probability is related to the attractiveness of the visited store and the travel cost (such as distance or travel time) between the customer and the store. The attractiveness here can be considered as what the store can offer to its customers, usually this variable can be represented by the store size, or the number of goods in the store.

In this study, we employ the above-mentioned time-aware dynamic Huff model (T-Huff), which was proposed based on the fact that people's visit preferences also have temporal variations (Liang et al. 2020). The authors provided evidence that the T-Huff model has higher estimation accuracy than the original Huff model in their experiments across ten U.S. cities and it is able to model the store visit patterns from both spatial and temporal perspectives. In this study, we apply the T-Huff model to study the store visit dynamics during the COVID-19 pandemic. We briefly introduce the model formula below; more technical details can be found in the original paper.

$$P_{ijt} = \frac{\frac{S_j^\alpha}{D_{ij}^\beta}}{\sum_{j=1}^n \frac{S_j^\alpha}{D_{ij}^\beta}} * P_{jt} \quad (13.1)$$

$$P_{jt} = \frac{V_{jt}}{\sum_{t=1}^m V_{jt}} \quad (13.2)$$

where P_{ijt} represents the visiting probability from customer i to the store j within time window t , S_j is the attractiveness of the store j and D_{ij} is the distance between customer i and the store j , n is the number of stores that customer i would visit. Equation 1 shows that the visiting probability is a result of comparison: the probability of visiting store j is calculated by comparing across all the stores that customer i will visit. P_{jt} is the probability of visiting store j within time window t and V_{jt} represents the number of total visits to store j within time window t . In this study, the time window t is selected as one hour, which is the finest resolution regarding data availability. So for one week, there are 168-dimensional visiting probabilities for one customer to one store.

Before using this model to predict the customer visiting probability, we need to calibrate it by adjusting two parameters: attraction exponent (α) and distance-decay coefficient (β) to make sure the model can reflect the reality. For each brand in each city, we use the Particle Swarm Optimization (PSO) to find the optimal sets of parameters that fit the data the best. The PSO method was proposed based on the movement of a bird flock (Eberhart and Kennedy 1995). It is selected for this study because it requires few assumptions and allows the design of different objective functions (Eberhart and Kennedy 1995; Xiao et al. 2013; Liang et al. 2020). Given each set of α and β , we can calculate the estimated visiting probability using the T-Huff model. We also computed the visiting probability using the actual visits each store received based on the collected mobile phone location tracking data (in the following Sect. 13.3) and considered it as the actual visiting probability. The cost function we used in the optimization function is the correlation between the estimated visiting probability and the actual visiting probability, and we try to maximize the correlation over the optimization process. For each brand in each city, we ran the optimization three times since there might be performance variations due to different starting positions. Based on the results from a few trials and the findings that PSO usually converges quickly (Shi and Eberhart 1999), we decided to have 50 iterations for each time with each iteration containing 10 candidate sets of values. The sets of α and β values that have the highest correlation will be selected as the optimal parameter fitting result.

Table 13.1 The chain-store brands and their categories

Brand	Category	NAICS code
Starbucks	Snack and nonalcoholic beverage bars	722515
McDonald's	Limited-service restaurants	722513
Target	Department stores	452210
Whole Foods Market	Grocery stores	445110
Trader Joe's	Grocery stores	445110

13.3 Data and Study Area

The Points of Interest (POIs) data for this study is provided by SafeGraph.¹ For each POI, the SafeGraph dataset contains its basic information including its location name, address, latitude/longitude, category, brand, etc. We selected four different types of POIs using the North American Industry Classification System (NAICS) code and conducted our analyses in three U.S. cities (New York, Los Angeles and Houston). For each category of POIs, we selected one or two top chain-store brands as the focused brand to study. The categories and specific brands are listed in Table 13.1. Those brands are selected because of two reasons: first, the five brands have the largest number of stores in their corresponding categories and are present in all three cities in our dataset; second, those brands are popular stores that are closely related to the daily life for many people living across the U.S. Analyzing the store visits to those brands may provide some representative and meaningful insights into human mobility pattern changes. Figure 13.1 shows the spatial distribution of the five chain-stores in our dataset in Los Angeles. McDonald's and Starbucks are present throughout the whole study area, whereas Target, Trader Joe's, and Whole Foods Market are primarily located in the northwestern Los Angeles according to the data provided by SafeGraph. Note that the POI database doesn't necessarily cover all the real-world stores for each brand.

In order to analyze the visit patterns in these brands, we use the “SafeGraph weekly patterns” dataset, which contains the place foot-traffic and aggregate information about how many visits are made to one POI, how often people visit each place, how long they stay, and where they come from.² For every POI, if a visit is accompanied by a mobile device, a trip from the home Census Block Group of the device owner to this POI is generated. The work of determining whether a device visits a POI is completed by the data provider using the GPS data (SafeGraph 2020b). The GPS points are first cleaned and then clustered if they are close in both space and time using a modified DBSCAN clustering algorithm. The clusters are then joined with a list of potential places which this cluster could have been referencing to. Then, a machine learning model called Learning-to-Rank is used to predict the most viable

¹<https://www.safegraph.com>.

²<https://docs.safegraph.com/docs/weekly-patterns>.

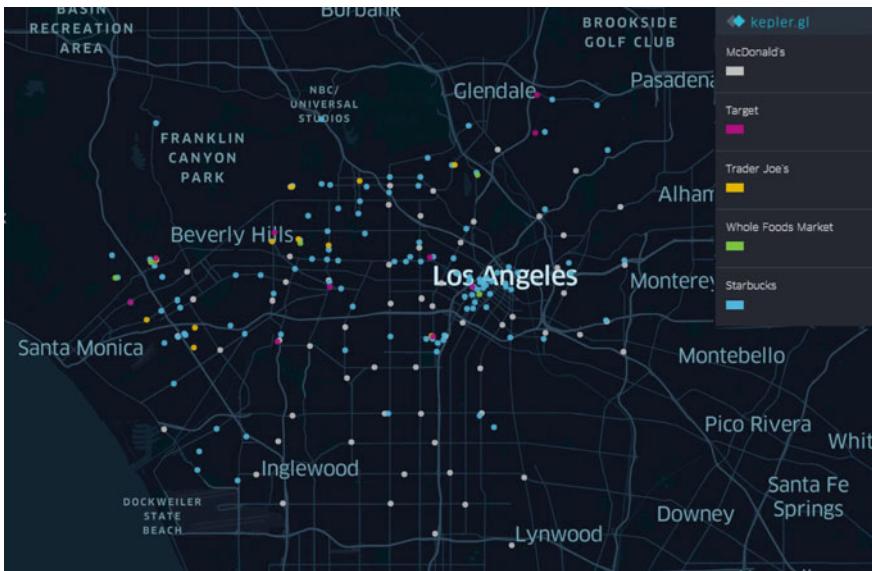


Fig. 13.1 The locations of five brands (McDonald’s, Target, Trader Joe’s, Whole Foods, and Starbucks) of stores in Los Angeles (in our database provided by SafeGraph)

place using features such as distance between clusters and places, the type of the places and the dwell time of the visit. Based on this data, we are able to map the flows from customers’ home Census Block Group to different stores. To discover the visit pattern changes under lockdowns during the COVID-19 pandemic, we selected two time periods for each city as the study time window. The first period is when COVID-19 started spreading and was declared a pandemic (March 11th, 2020). As the data is provided in a weekly format, we selected the two weeks March 2nd–8th and March 9th–15th to represent this period. For the second period, we focused on the weeks after statewide lockdown orders were issued. Our aim is to discover whether and how movement behaviors changed according to the lockdown orders. Based on the dates of the lockdown orders (New York, NY: March 22nd; Los Angeles, CA: March 19th; Houston, TX: April 2nd) (Wu et al. 2020), we picked the two weeks March 23rd–29th and March 30th–April 5th for New York and Los Angeles, and April 6th–12th and April 13th–19th for Houston as their post-lockdown periods, respectively.

Although the data used in this study come from a very large coverage—around 10% of mobile devices in the United States, it may still have sampling bias. The sampling bias has been examined in different geographic scales with three demographic factors: race, educational attainment and household income. The proportion of any given sub-group in the SafeGraph dataset should be the same as it in the US Census population if there’s no sampling bias. The data provider has provided evidence that their panel data is generally well sampled showing high correlations with the true population and low sum of absolute bias (Squire 2020), although a

small number of the Census Block Groups show some extreme outliers and this may be related to the process of linking a device to a specific Census Block Group.

This research got an IRB waiver from the University of Wisconsin-Madison for using aggregated mobile phone data.

13.4 Results and Discussion

13.4.1 *Store Foot-Traffic and Dwell Time Distribution Changes*

We first analyzed the specific foot-traffic change to each chain-store brand and discovered how the store visits have been impacted by the lockdown orders during the COVID-19 pandemic.

Tables 13.2, 13.3 and 13.4 show the foot-traffic statistics including the number of stores, the number of total visits, the average visits, and the reduced percentage of average visits over the two periods for the three cities, respectively. We find that the number of stores of the same brand (with active visits) varies across three cities but

Table 13.2 Store visit statistics in Los Angeles

	Early two weeks			Later two weeks			% of reduced visits
	Stores	Visits	Mean visits	Stores	Visits	Mean visits	
McDonald's	63	16358	259.7	63	8779	139.4	46.3
Starbucks	113	17866	158.1	105	7790	74.2	53.1
Target	10	6183	618.3	10	3459	345.9	44.1
Trader Joe's	11	2562	232.9	11	999	90.8	61.0
Whole Foods Market	5	615	123.0	5	274	54.8	55.4

Table 13.3 Store visit statistics in New York

	Early two weeks			Later two weeks			% of reduced visits
	Stores	Visits	Mean visits	Stores	Visits	Mean visits	
McDonald's	55	13188	239.8	53	2875	54.3	77.4
Starbucks	216	47703	220.9	212	8134	38.4	82.6
Target	8	4453	556.6	7	1400	200.0	64.1
Trader Joe's	7	591	84.4	6	152	25.3	70.0
Whole Foods Market	9	1432	159.1	8	502	62.8	60.6

Table 13.4 Store visit statistics in Houston

	Early two weeks			Later two weeks			% of reduced visits
	Stores	Visits	Mean visits	Stores	Visits	Mean visits	
McDonald's	109	41172	377.7	107	22227	207.7	45.0
Starbucks	114	43447	381.1	106	17643	166.4	56.3
Target	15	18374	1224.9	15	10622	708.1	42.2
Trader Joe's	3	1162	387.3	3	619	206.3	46.7
Whole Foods Market	9	2022	224.7	8	1023	127.9	43.1

the differences are generally not very large. The biggest difference is for Starbucks, New York City has almost twice the number of Starbucks compared to the other two cities. By comparing the average store visits over the two time periods, it shows that all stores in each city have much fewer average visits in the periods after the lockdown order was announced. Most of the stores lost more than 50% of their visits in the later period. Different cities show different trends in which stores have the greatest decrease in these two time periods. New York City had the largest decrease percentage of visits in all brands. This can be related to the fact that New York had the most COVID-19 cases in April and many people stayed at home to avoid the infection (The New York Times 2020). It is also possible that people in New York City did not drive as much and therefore did not use drive-thru opportunities in McDonald's and Starbucks.

In addition to the store visits, in each brand, we also analyzed the total count of visitors based on the bucketed dwell times. The bucketed dwell times represent five intervals for store visits: less than 5, 5–20, 21–60, 61–240 min, and greater than 240 min. Figures 13.2, 13.3 and 13.4 show a histogram of total visitor counts by each brand in the three cities over the two time periods.

The left side of Fig. 13.2 shows the dwell time patterns for McDonald's store visits. In all three cities prior to the lockdowns, the highest frequency of visits was between 5–20 min. Similarly, in all three cities after the lockdown, the highest frequency of visits was still between 5–20 min. However, a difference emerged between cities: the number of visitors in Houston and Los Angeles McDonald's stores were higher than visitors to New York McDonald's. Comparing the changes over the two weeks in three cities, New York City shows the greatest drops in all time intervals, corresponding to what we discovered from our analysis on the number of store visits.

Starbucks, a snack and non-alcoholic beverage bar category brand, shows a high volume of visitors in the 5–20 min dwell time interval in the right side of Fig. 13.2. Compared with McDonald's, more people spent longer time (more than 20 min) in Starbucks in the early two weeks. Though the early period indicates a higher count of visitors, we see a tremendous decrease during the latter two weeks. Similarly as we mentioned before, New York City shows the largest difference in visitors in the two time periods.

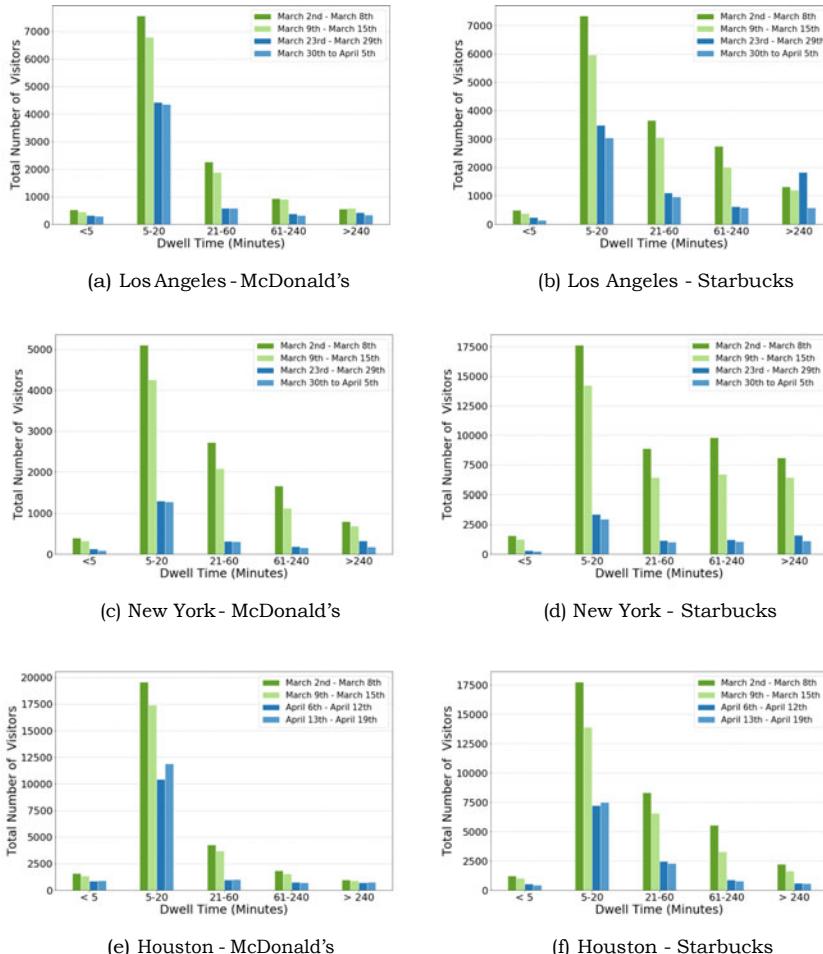


Fig. 13.2 The dwell time distribution for McDonald's and Starbucks

The dwell time distribution for Target, a department store, is shown in Fig. 13.3. In Los Angeles and Houston, the number of visits in the 5–20 and 21–60 min intervals was very high and visits longer than 60 min were much lower. These two cities show a similar trend in the later two-week periods. Although New York City Target stores had a high number of visitors for the 5–20 and 21–60 min intervals, they also had a high number of visitors that took place over 60 min in the early two weeks. During the later period, New York City Target stores had noticeably much lower visitor counts for all time intervals. Public was concerned about the sufficient supply of goods, therefore, they may have spent extra time shopping for additional items and storing them in the early period. In general, the number of Target store visits was decreased for all time periods in the lockdown periods for all three cities, likely reflecting the

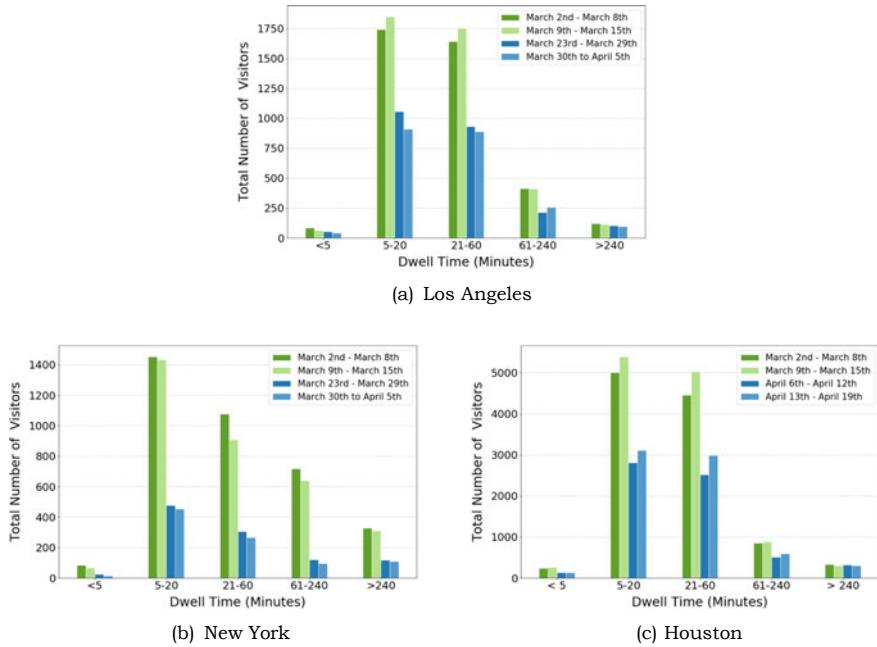


Fig. 13.3 The dwell time distribution for Target

preparedness with storage of home goods in the earlier period and the fast-pace of shopping consumers have adopted during the pandemic to limit their exposure to different parts of stores and different customers.

Finally, as shown in Fig. 13.4, we have a comparison of the grocery store brands Whole Foods Market and Trader Joe's. One finding is that in all three cities, there were more visits in the March 9th–March 15th week than the earlier week in the 21–60 min dwell time. This is also the week that COVID-19 was declared by WHO as a global pandemic (World Health Organization 2020). Therefore, the pandemic announcement, more specifically the willingness to stock up on groceries during a pandemic and longer lines may be an indicator to why people spent more time in grocery stores to purchase food among others.

13.4.2 Spatial Flow Distribution and Distance Decay

Using the collected mobile phone location data, we generated the origin-to-destination (OD) flow maps from the visitor's home Census Block Groups to each store location and compared the visit variations from the spatial point of view. Figure 13.5 displays maps for each of the three cities in this study showing visitor frequencies from home Census Block Groups to Whole Food Market stores in the two

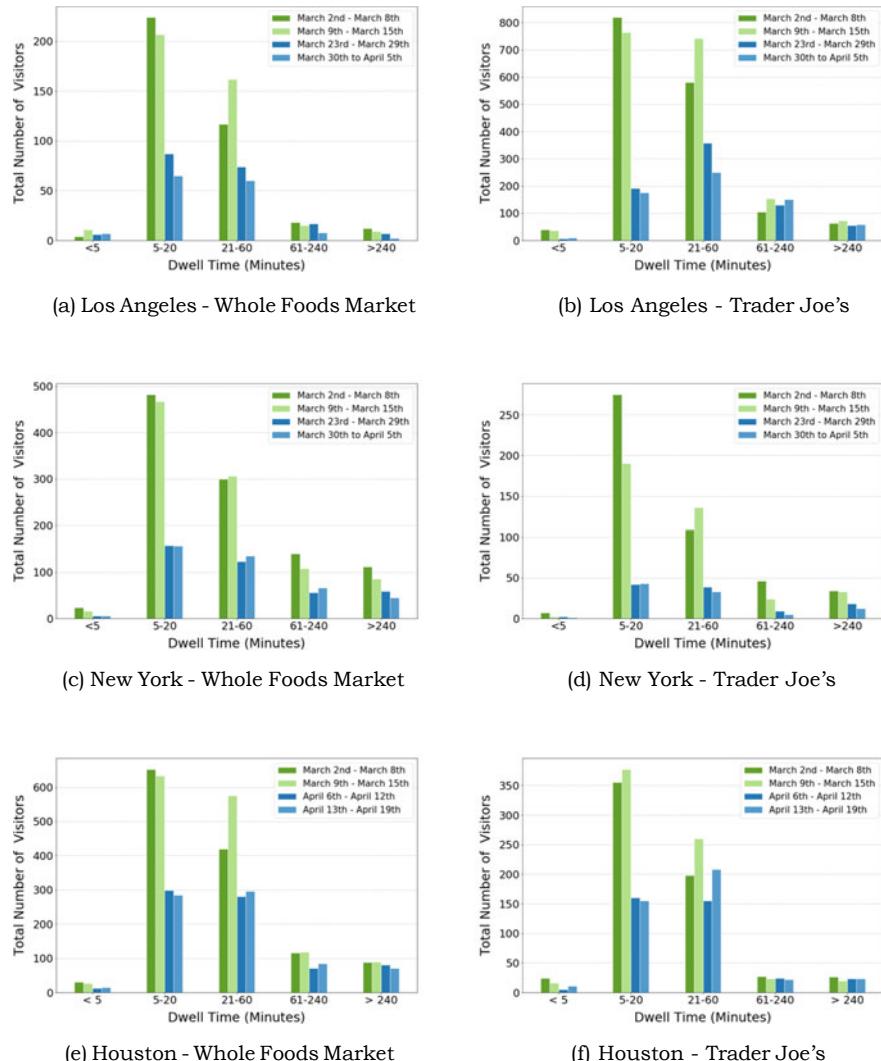


Fig. 13.4 The dwell time distribution for Grocery Store Brands

time periods. During the early two-week period, the flow maps reveal high frequency in various distances between Whole Foods Market store location and the visitor's home Census Block Group. There are also a few trips with frequency higher than 12 times in the early period. Most of the high-frequency trips (lighter colors) are within relatively close proximity to the store and there are also many long-distance visits with low frequency (darker colors) on the map.

For the latter two weeks, the maps reveal there is a lower frequency in visits. This is indicated by the overall colors on the map appear darker because some high-frequency



Fig. 13.5 Frequency of visits from home Census Block Groups to Whole Foods Markets

visits represented by the light colors disappeared. This could be an outcome of stocking up on groceries as well as limiting visits to stores, especially if an individual used public transportation. For instance, during the lockdown orders, cities across the country have advised the public to avoid using public transportation as much as possible (Tirachini and Cats 2020). For Houston, however, some long-distance trips have lighter colors in the later period (e.g., the two lines in the southeastern corner in Fig. 13.5f), meaning that some people still travelled long distances to

visit Whole Foods Market more often than in the earlier two weeks. This could be due to Whole Foods Market having specialty products that cannot be found in other grocery stores. For those remaining visits, visitors' home Census Block Groups were generally closer to Whole Foods Market compared with the early period. So with the lockdown orders, in addition to the reduction in the number of visits to Whole Foods Market stores, there is also a reduction in long-distance trips to visit such stores based on the flow maps. Similar findings have also been identified by other human mobility change research in the U.S. during the COVID-19 pandemic, providing evidence that the lockdown policies have made the short-distance trips gain a higher weight out of total trips than before (Gao et al. 2020; Aloi et al. 2020; The National Academies of Sciences, Engineering, and Medicine 2020).

Besides the flow maps, for each store of each brand, we computed the median travel distance from all the visitors to that store each week. Figure 13.6 shows the box plots of the median distance distribution for all the stores of each brand in the three cities. The red boxes are results for the early two weeks while the blue boxes are for the later weeks. In each boxplot, the interquartile ranges (IQR) of the data (from Q1: 25 percentile to Q3: 75 percentile) is represented in a box with an orange line at the median (Hunter et al. 2020). The two whiskers show the upper and lower range of the data, the upper whisker is defined as $Q3 + 1.5 \text{ IQR}$ and the lower whisker equals to $Q1 - 1.5 \text{ IQR}$. All data points beyond the whiskers are treated as outliers and excluded from our analysis.

For Los Angeles (Fig. 13.6a), in general, both the range between whiskers and the IQR of McDonald's, Target and Trader Joe's have a trend of moving down in the two time periods we examined, meaning that most of the travel distances became shorter in the latter lockdown period. Starbucks does not have a very obvious range change between the two time periods but it has a smaller median value in the later period, indicating that most trips to Starbucks have become shorter. Whole Foods Market actually shows a higher value for the bottom whisker in the later period, meaning that it has fewer short-distance trips. This result might be affected by the limited trip data for Whole Foods Market stores, as shown in Table 13.2, we only have 5 Whole Foods Market stores in Los Angeles in our dataset and fewer trips during lockdowns. So the change of travel distance to one store can easily affect the overall distribution of data.

For New York (Fig. 13.6b), all brands show a trend of moving down to smaller data ranges in the later weeks. Target has the most significant change: the range between two whiskers as well as the range of interquartile both shrank greatly in the later weeks. For Trader Joe's, its median travel distance dropped sharply, showing that though the range of the distance only changed slightly, the shorter-distance visits to this brand have a higher weight in the later weeks compared with that in the early period.

For Houston, most of the brands in Fig. 13.6c show the trend of having shorter travel distances as well as smaller data ranges in the later weeks compared with the early weeks. McDonald's and Starbucks both have fewer long travel distances as the upper whisker become much smaller in the later period. One noticeable difference

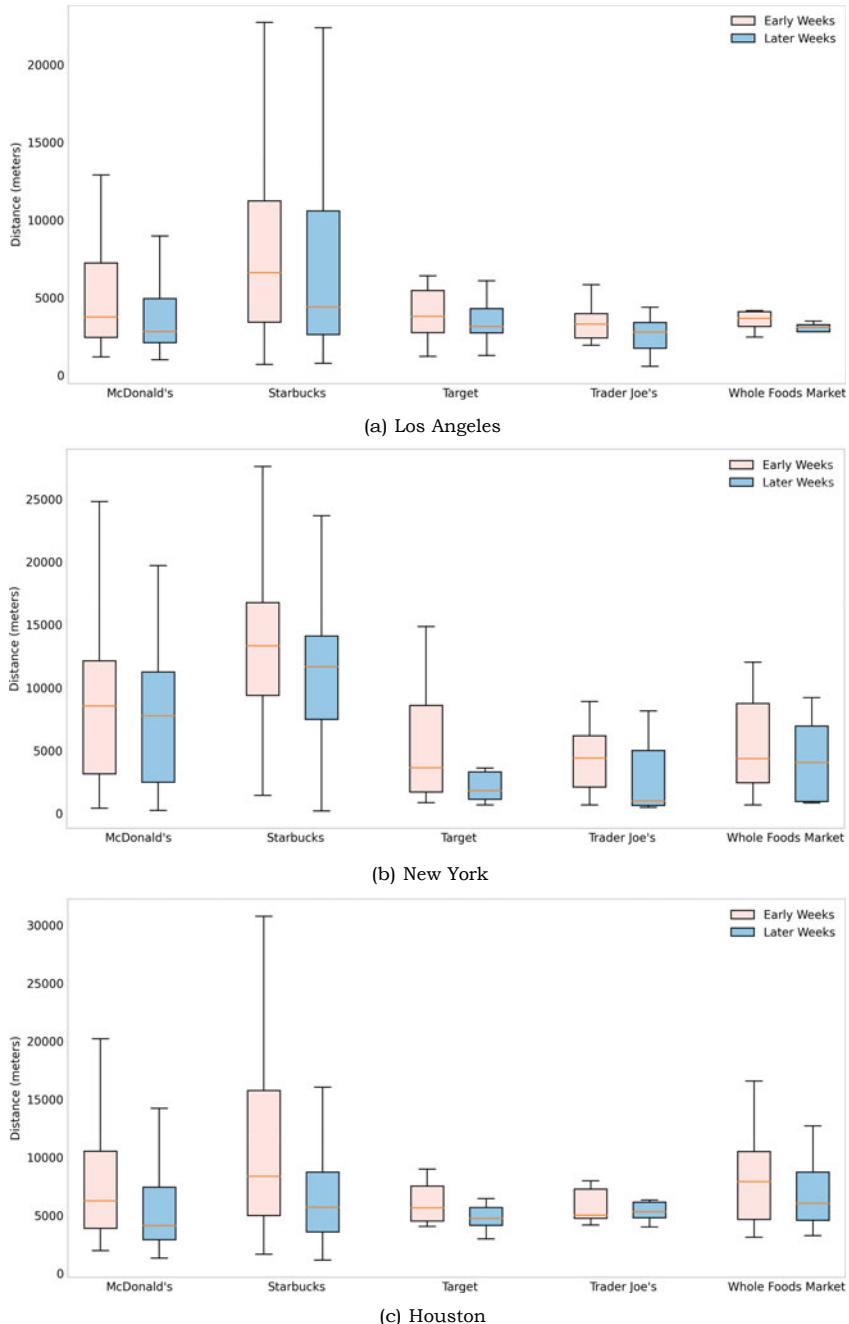


Fig. 13.6 The median travel distance (in meters) distributions for stores in Los Angeles, New York and Houston (outliers removed)

Table 13.5 Optimization using correlation for Log Angeles

Location name	Early two weeks			Later two weeks		
	Alpha	Beta	Correlation	Alpha	Beta	Correlation
McDonald's	0.0087	0.1746	0.9969	0.0304	0.0984	0.9939
Starbucks	0.0040	0.0375	0.9958	0.0001	0.0810	0.9958
Target	0.0053	0.0913	0.9955	0.0310	0.1110	0.9922
Trader Joe's	0.0110	0.0651	0.9992	0.0866	0.1017	0.9932
Whole Foods Market	0.0120	0.1277	0.9933	0.0007	0.0937	0.9302

is for Trader Joe's: it actually has a larger median value in the later weeks based on our data, which is the opposite of all other brands.

Given that the observed visit data are about 10% representative samples of the mobile devices in the U.S.,³ we further built the T-Huff model for each brand in each city during each time period to understand the overall spatial interaction patterns, which are important for location business insights as we cannot track all the customers across all the neighborhoods in different cities. A set of optimal α, β parameter values for the T-Huff model are obtained through the optimization. Table 13.5 shows the optimization results using the correlation between observed visit probability and fitted probability for Los Angeles. In general, the optimal correlations we obtained are very high and most of them are higher than 0.99, meaning that the predicted visiting probability is strongly related to the actual probability.

Comparing the results for each brand specifically, the parameter α controls the weight of store attractiveness in the model and the parameter β controls the impact of distance. A larger β means that distance plays a more important role when people take this trip. As the distance is the denominator in the visiting probability calculation (Eq. 13.1), a higher β indicates that people are less likely to visit a store that is far away from their home location (Liu et al. 2014; Liang et al. 2020). Therefore, when we compare the β values across brands, we may be able to discover people's visit preferences for long-distance trips to different brands in different time periods.

When we look row by row and compare the two β values for each brand in Table 13.5, Starbucks, Target and Trader Joe's has increased β values in the latter two weeks, meaning that Los Angeles residents seemed to become less willing to travel longer distances to visit those three brands after the lockdown. For McDonald's and Whole Food Market, we have lower β values in the later weeks. When we look back at Fig. 13.6a, for Whole Foods Market it indeed has a higher value for its lower whisker in the later periods, showing that we lost short travel distances for this brand. This makes the data have more longer-distance trips in the latter period and might lead to the decreased β value from the optimization result.

Table 13.6 shows the optimization results for New York. By comparing two β values, we see that McDonald's, Target and Whole Foods Market have increased β values, meaning that distance has a greater negative impact when people visited those

³<https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset>.

Table 13.6 Optimization using correlation for New York

Location name	Early two weeks			Later two weeks		
	Alpha	Beta	Correlation	Alpha	Beta	Correlation
McDonald's	0.0153	0.0840	0.9968	0.0360	0.0920	0.9877
Starbucks	0.0029	0.1336	0.9962	0.0259	0.0797	0.9929
Target	0.0331	0.0898	0.9963	0.0036	0.0976	0.9947
Trader Joe's	0.0010	0.0685	0.9970	0.1051	0.0177	0.9944
Whole Foods Market	0.0209	0.0549	0.9948	0.0239	0.0803	0.9960

Table 13.7 Optimization using correlation for Houston

Location name	Early two weeks			Later two weeks		
	Alpha	Beta	Correlation	Alpha	Beta	Correlation
McDonald's	0.0053	0.0103	0.9978	0.0077	0.0503	0.9985
Starbucks	0.0066	0.1891	0.9970	0.0088	0.0678	0.9958
Target	0.0025	0.2929	0.9975	0.0147	0.3333	0.9975
Trader Joe's	0.0008	0.2855	0.9987	0.0274	0.0351	0.9916
Whole Foods Market	0.0047	0.3047	0.9988	0.0052	0.0036	0.9931

brands. Starbucks and Trader Joe's have decreased β values in the latter two weeks. In Fig. 13.6b, Starbucks and Trader Joe's are the two brands with the slightest data range changes over the two periods. The visit pattern to those two brands in New York over the study periods may not have changed very significantly.

Table 13.7 provides the optimization results for Houston. McDonald's and Target both have higher β values for the latter weeks, showing that distance became more important to consider when visiting a store after the lockdown. For all other three brands, their β values dropped dramatically to less than 0.1 in the latter weeks. One reason for this might be the later period has fewer data and the optimization did not perform well when the data size is very small. In addition, as Los Angeles and New York had much more positive cases for COVID-19 over our study periods compared with Houston, it is also possible that people in Houston did not have consistent visit changes in response to COVID-19 in comparison with people in Los Angeles and New York; therefore, the results did not reveal a clear long-distance movement reduction for Houston.

13.4.3 Discussion

After we combined the results for all three cities, we found that Target, which is defined as a department store, has increased β values in the latter two weeks for all cities. This consistent result indicates that people are less likely to travel long

distances to visit this brand during the lockdown period. This may correspond to the lockdown orders that were issued to limit people's travels only to essential businesses. It is understandable that consumers would limit their purchases during a pandemic to essential items such as food and daily household goods such as cleaning supplies and toilet paper. Target brand carries household goods as well as groceries in some of its stores, and much more including clothing and travel-related items. Because much of those essential items (i.e., food and daily household goods) can be found in nearby other grocery stores, the need for individuals to travel long distances to Target may have been decreased. Moreover, many individuals go to Target occasionally and do not shop there on a daily or weekly basis. So it is possible that individuals may have shopped in greater quantity when the pandemic started in the early period not necessitating trips in the short-run. In this case, people may still visit nearby grocery stores as before, but they might show less interest in visiting department stores such as Target. Here, we use mobile phone location big data to show some evidence of this hypothesis.

McDonald's and Starbucks do not have a consistent trend of the β value changes over time, likely because there are usually more of these two brands in communities compared with the department stores and grocery stores. They do not require large spaces like department stores or large grocery stores and are designed to be located near people for easy access and convenience. So the visits to them may be less affected by distance as the visits are in relatively close proximity.

When we compared the overall β values across three cities, New York City has the lowest β value on average. If we only compare the β values for Target, Trader Joe's and Whole Foods Market (which are stores easier to be affected by distance than McDonald's and Starbucks) in the two periods, out of the six β values New York City has the lowest β value for four times. A lower β indicates that distance has a weaker negative impact on the travel behavior and people are less likely prevented to visit stores by the long distance between them and the stores (Liang et al. 2020; Liu et al. 2014).

New York has a much different urban make up than many other large cities in the U.S., including Los Angeles and Houston. New York City is considered the most automobile-independent large city due to its residential density, its urban fabric with mixed-use development, and its well-developed public transit system (Rundle et al. 2007; Newman and Kenworthy 2015). Similarly, New York is considered the most compact and connected metropolitan area in the U.S. based on a rating of 221 metropolitan areas for urban sprawl (Smart Growth America 2014). The Houston metropolitan area, on the other hand, appears as 182nd in this index; in other words, it is among the most sprawling and automobile-dependent metropolitan areas in this rating. Earlier studies have identified the sensitivity to travel distance also varies in different cities based on different transportation mode the city has (Pun-Cheng 2016; Liang et al. 2020). Though the general consensus is that the cost of a trip is proportional to the distance, it may not be the case when people are taking public transportation, where the cost can be similar for different lengths of trips (Pun-Cheng 2016). Therefore, residents in New York City may be more willing to travel a little further than the other two cities due to its mature public transportation system. This

might be one of the reasons why New York City always has lower distance-decay coefficient (β) values.

However, due to health concerns and the lockdown policy, people may have limited access to public transportation. One study in Spain found that public transport users have a greater decrease compared with private car users after the confinement measures imposed (Wang et al. 2020; Aloi et al. 2020). For New York City, the traffic was estimated to decrease by about 35% compared with the same period in 2019 (Aloi et al. 2020). However, based on the β values we found, people in New York City are still more likely to travel a longer distance to visit the same store than the other two cities, meaning that people's visit preferences or those essential workers at stores may remain the same despite the effects of the pandemic.

In addition to the T-Huff model, we also conducted a multiple linear regression (MLR) analysis to discover how the neighborhood characteristics and the POI attraction are associated with the visit changes under the lockdown policies during the COVID-19 pandemic. By including multiple demographic variables from US Census to predict the visit changes to McDonald's in Los Angeles, we found that population, median age and the ratio of people having a bachelor's degree or higher have a positive contribution to the drop in the visit. In other words, in areas with a higher population, higher median age and higher education level (larger ratio of people with a bachelor degree or higher), people tend to show a larger drop in the visit in the post-lockdown period. Similar findings have also been discovered by some studies showing that the education and income level may affect people's response to the lockdown policies (Brzezinski et al. 2020). Such analyses may help further understand the impacts of the lockdown policies in different communities and help adjust the policies to encourage less responsive neighborhoods to better follow the policies. A segregation effect was also found by Bonaccorsi et al. (2020) that the mobility contraction is stronger in areas where inequality is higher. A more comprehensive understanding of how individuals have been impacted, why they responded differently and who are the most fragile individuals under the lockdowns requires further research.

13.5 Conclusion and Future Work

In this study, we employed mobile phone location data to examine visit changes to different types of chain stores in two time periods (the period when COVID-19 became a pandemic and started spreading in U.S. and the period after a lockdown order was announced). We are able to discover the significant visit changes to five major chain-store brands in response to COVID-19 in three large U.S. cities (Los Angeles, New York and Houston). By applying a time-aware Huff model, we further examined different visit pattern changes to different types of stores. People in general reduced longer distance trips and became less willing to travel longer to some stores such as Whole Foods or Trader Joe's in different cities. Target, which is a department stores, is found to always have increased distance-decay coefficient (β) values in the

latter two weeks using the T-Huff model, meaning that longer distance trips had a greater negative impact when visiting this brand for customers after lockdown based on our study.

One limitation of this study is that the locations of the brands in our study may affect the accessibility to them by different transport modes. Some stores may not be easily accessed by public transit and this can affect people's willingness to visit those stores. Future work can further examine the distance from the stores to public transit routes and stops, and include this feature in the store attractiveness estimation as a factor related to the visit probability.

Some future research directions can be explored by expanding the scope of this study. For example, in addition to understanding the visits to specific brands, it is also possible to consider each business category as a whole and compare the visit patterns to different types of POIs such as restaurants or grocery stores. This may help provide a broader picture of the movement behaviors as many brands are local and cannot be compared across multiple cities. It is also worth studying how individuals' movements have changed with time. We only picked two time periods, one being at the beginning of the pandemic and the other soon after lockdowns were announced. However, it would be critical to investigate how mobility was adjusted under different policies as the states started reopening. Including longer investigation periods could help examine the influence of different policies as well as "behavioral fatigue", the phenomenon that people have become tired of staying at home and less self-regulating in the later periods of the pandemic (Sibony 2020; Brodeur et al. 2020; Zhang et al. 2020).

The results of this study may also provide suggestions for policy making related to public health. For example, the predicted temporal visit probabilities to different stores can be used to advise customers on finding a suitable time for shopping to avoid the crowds during the peak visit hours. Similarly, the temporal feature can also be applied to help businesses plan for reopening. For example, it can provide temporal information to stores that are considering opening part-time to select the most appropriate hours for reopening. In addition, analyzing how the store visit pattern has changed can help understand the effectiveness of the lockdown policies and provide insights for decision-making when governments are considering adjusting lockdown orders and preparing for reopening. Businesses in conjunction with local and state authorities could use findings from ours and other similar studies to advise store visit hours, with the assumption that there will be a long way to COVID-19 recovery even with the good news of vaccines and there might be similar outbreaks in future and that we might need to limit contact among individuals.

Beyond lockdowns, COVID-19 has also triggered different business and urban planning practices that have targeted safe business experiences for the health of individuals and the economy. These practices aim to reduce human interaction, especially in closed spaces. Many businesses have transferred their activities from indoors to outdoors: some restaurants began to provide outdoor dining options and some retail stores provided their wares on city sidewalks (Hendrickson 2020). While these practices are not new in many parts of the world, they are new in most U.S. cities as such

business practices are typically not allowed in cities due to existing zoning regulations. Post-COVID urban planning practice in the U.S. will likely see an increase in the relaxation of these codes and regulations. Moreover, the demand for places that individuals can be active outside as well as to experience the urban environment has risen significantly; prompting the expansion of sidewalks and conversion of streets to pedestrian-only streets across the globe (Wojahn 2020). Scholars, planners, and visionaries have been contemplating on what the post-pandemic city will look like and whether COVID-19 will be detrimental to the city itself (Couclelis 2020). We are optimistic, just like Couclelis is, that the city will survive, as it has done so in the past in the face of adversity as well as technological innovation, since the various cultural, economic, education, and social opportunities and the experiences of the built and natural environments cities offer are among the most important aspects of urban living. The planning practice and resulting outcomes will likely differ significantly from the practices and outcomes of the past several decades; indeed we imagine that post-pandemic cities may make our urban environments less automobile-dependent with the integration of land uses so that individuals can access services and goods more easily as well as with additional outdoor experience opportunities with increased greenery, active transportation options (e.g., bike lanes and paths), and dedication of outdoor space for public and businesses. Ultimately, these are strategies that sustainability-minded planners have been trying to bring to communities for several decades, and a pandemic might simply force us to do that.

Acknowledgements The authors would like to thank the funding support provided by the National Science Foundation (Award No. BCS-2027375). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Support for this research was provided by the University of Wisconsin—Madison Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation. This research was also supported by SafeGraph and The We Company (WeWork) through academic collaboration with the Geospatial Data Science (GeoDS) Lab at the University of Wisconsin—Madison.

References

- Aktay, A., Bavadekar, S., Cossoul, G., Davis, J., Desfontaines, D., Fabrikant, A., Gabrilovich, E., Gadepalli, K., Gipson, B., Guevara, M., et al. (2020). Google COVID-19 community mobility reports: Anonymization process description (version 1.0). *arXiv preprint arXiv:2004.04145*.
- Aloia, A., Alonso, B., Benavente, J., Cordera, R., Echaniz, E., Gonzalez, F., Ladisa, C., Lezama-Romanelli, R., Lopez-Parra, A., Mazzei, V., et al. (2020). Effects of the COVID-19 Lockdown on urban mobility: Empirical Evidence from the City of Santander (Spain). *Sustainability*, 12(9), 3870.
- Bonaccorsi, G., Pierri, F., Cinelli, M., Flori, A., Galeazzi, A., Porcelli, F., Schmidt, A. L., Valentisise, C. M., Scala, A., Quattrociocchi, W., et al. (2020). Economic and social consequences of human mobility restrictions under COVID-19. *Proceedings of the National Academy of Sciences*, 117(27), 15530–15535.
- Brodeur, A., Clark, A., Fleche, S., & Powdthavee, N. (2020). COVID-19, lockdowns and well-being: Evidence from Google trends.

- Brzezinski, A., Deiana, G., Kecht, V., & Van Dijcke, D. (2020). The COVID-19 pandemic: Government vs. community action across the united states. *Covid Economics: Vetted and Real-Time Papers*, 7, 115–156.
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., & Leskovec, J. (2020). Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, 1–8.
- Chetty, R., Friedman, J. N., Hendren, N., & Stepner, M. (2020). Real-time economics: A new platform to track the impacts of COVID-19 on people, businesses, and communities using private sector data. *NBER Working Paper*, 27431.
- Couclelis, H. (2020). There will be no post-COVID city. *Environment and Planning B: Urban Analytics and City Science*, 47(7), 1121–1123.
- Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). Strong social distancing measures in the United States reduced the COVID-19 growth rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. *Health Affairs*, 10–1377.
- Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *MHS'95 Proceedings of the Sixth International Symposium on Micro Machine and Human Science* (pp. 39–43). IEEE.
- Gao, S., Rao, J., Kang, Y., Liang, Y., Kruse, J., Dopfer, D., Sethi, A. K., Reyes, J. F. M., Yandell, B. S., & Patz, J. A. (2020). Association of mobile phone location data indications of travel and stay-at-home mandates with COVID-19 infection rates in the US. *JAMA Network Open*, 3(9), e2020485–e2020485.
- Hendrickson, V. (2020). Shopping al fresco: New York City will allow stores to operate outside. <https://www.marketwatch.com/story/shopping-al-fresco-new-york-city-will-allow-stores-to-operate-outside-11603923938>.
- Holtz, D., Zhao, M., Benzell, S. G., Cao, C. Y., Rahimian, M. A., Yang, J., Allen, J. N. L., Collis, A., Moehring, A. V., Sowrirajan, T., et al. (2020). Interdependence and the cost of uncoordinated responses to COVID-19.
- Hong, B., Bonczak, B., Gupta, A., Thorpe, L., & Kontokosta, C.E. (2020). Exposure density and neighborhood disparities in COVID-19 infection risk: Using large-scale geolocation data to understand burdens on vulnerable communities. *arXiv preprint arXiv:2008.01650*.
- Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020). Twitter reveals human mobility dynamics during the covid-19 pandemic. *PLoS ONE*, 15(11),
- Huff, D. L. (1963). A probabilistic analysis of shopping center trade areas. *Land Economics*, 39(1), 81–90.
- Hunter, J., Dale, D., Firing, E., & Droettboom, M., & The Matplotlib development team (2020). matplotlib.pyplot.boxplot. https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.pyplot.boxplot.html.
- Lai, S., Ruktanonchai, N. W., Zhou, L., Prosper, O., Luo, W., Floyd, J. A., Wesolowski, A., Zhang, C., Du, X., Yu, H., et al. (2020). Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*.
- Li, Q., Bessell, L., Xiao, X., Fan, C., Gao, X., & Mostafavi, A. (2020). Disparate Patterns of Movements and Visits to Points of Interests Located in Urban Hotspots across US Metropolitan Cities during COVID-19. *arXiv preprint arXiv:2006.14157*.
- Liang, Y., Gao, S., Cai, Y., Foutz, N. Z., & Wu, L. (2020). Calibrating the dynamic huff model for business analysis using location big data. *Transactions in GIS*, 24(3), 681–703.
- Liu, Y., Sui, Z., Kang, C., & Gao, Y. (2014). Uncovering patterns of inter-urban trip and spatial interaction from social media check-in data. *PloS One*, 9(1).
- McKenzie, G., & Adams, B. (2020). A country comparison of place-based activity response to covid-19 policies. *Applied Geography*, 125(12), 102363.
- McKenzie, G., Janowicz, K., Gao, S., Yang, J.-A., & Hu, Y. (2015). POI pulse: A multi-granular, semantic signature-based information observatory for the interactive visualization of big geosocial data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 50(2), 71–85.

- Mohler, G., Bertozzi, A. L., Carter, J., Short, M. B., Sledge, D., Tita, G. E., Uchida, C. D., & Brantingham, P. J. (2020). Impact of social distancing during COVID-19 pandemic on crime in Los Angeles and Indianapolis. *Journal of Criminal Justice*, 101692.
- Newman, P., & Kenworthy, J. (2015). The end of automobile dependence. In *The end of automobile dependence* (pp. 201–226). Springer.
- Painter, M., & Qiu, T. (2020). Political beliefs affect compliance with COVID-19 social distancing orders. Available at SSRN 3569098.
- Pan, Y., Darzi, A., Kabiri, A., Zhao, G., Luo, W., Xiong, C., et al. (2020). Quantifying human mobility behaviour changes during the COVID-19 outbreak in the united states. *Scientific Reports*, 10(1), 1–9.
- Pun-Cheng, L. S. (2016). Distance decay. In *International Encyclopedia of Geography: People, the Earth, Environment and Technology: People, the Earth, Environment and Technology* (pp. 1–5).
- Rundle, A., Roux, A. V. D., Freeman, L. M., Miller, D., Neckerman, K. M., & Weiss, C. C. (2007). The urban built environment and obesity in New York City: a multilevel analysis. *American Journal of Health Promotion*, 21(4 suppl), 326–334.
- SafeGraph. (2020a). Places schema. <https://docs.safegraph.com/docs>.
- SafeGraph. (2020b). Turn GPS data into store visit intelligence. <https://www.safegraph.com/visit-attribution>.
- Saha, J., Barman, B. B., & Chouhan, P. (2020). Lockdown for COVID-19 and its impact on pupil mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. *Children and Youth Services Review*, 105160.
- Shi, Y., & Eberhart, R. C. (1999). Empirical study of particle swarm optimization. In *Proceedings of the 1999 congress on evolutionary computation-CEC99 (Cat. No. 99TH8406)*, vol 3 (pp. 1945–1950). IEEE.
- Sibony, A.-L. (2020). The UK COVID-19 response: A behavioural irony? *European Journal of Risk Regulation*, 1–8.
- Smart Growth America. (2014). Measuring sprawl 2014. <https://www.smartgrowthamerica.org/app/legacy/documents/measuring-sprawl-2014.pdf>.
- Squire, R. (2020). What about bias in the safegraph dataset? <https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset>.
- Suarez-Vega, R., Gutierrez-Acuna, J. L., & Rodriguez-Diaz, M. (2015). Locating a supermarket using a locally calibrated Huff model. *International Journal of Geographical Information Science*, 29(2), 217–233.
- The National Academies of Sciences, Engineering, and Medicine. (2020). *TRB Webinar: How Much Will COVID-Affect Travel Behavior?* <https://www.nationalacademies.org/event/06-01-2020/trb-webinar-how-much-will-covid-19-affect-travel-behavior#sectionContact>.
- The New York Times. (2020). COVID in the U.S.: Latest Map and Case Count. <https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html#states>.
- Tirachini, A., & Cats, O. (2020). COVID-19 and public transportation: Current assessment, prospects, and research needs. *Journal of Public Transportation*, 22(1), 1.
- Wang, D., He, B.Y., Gao, J., Chow, J.Y., Ozbay, K., & Iyer, S. (2020). Impact of COVID-19 behavioral inertia on reopening strategies for New York City transit. *arXiv preprint arXiv:2006.13368*.
- Wang, Y., Jiang, W., Liu, S., Ye, X., & Wang, T. (2016). Evaluating trade areas using social media data with a calibrated huff model. *ISPRS International Journal of Geo-Information*, 5(7), 112.
- WHO. (2020). WHO coronavirus disease (COVID-19) dashboard. <https://covid19.who.int/>.
- Wojahn, P. (2020). Creating more space for safe active transportation and outdoor activity during COVID-19 and beyond. <https://www.nlc.org/article/2020/05/29/creating-more-space-for-safe-active-transportation-and-outdoor-activity-during-covid-19-and-beyond/>.
- World Health Organization. (2020). WHO director-general's opening remarks at the media briefing on COVID-19. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19--11-March-2020>.

- Wu, J., Smith, S., Khurana, M., Siemaszko, C., & DeJesus-Banos, B. (2020). Stay-at-home orders across the country. <https://www.nbcnews.com/health/health-news/heres-what-happened-with-stay-at-home-orders-across-country-n1168736>.
- Xiao, Y., Wang, F., Liu, Y., & Wang, J. (2013). Reconstructing gravitational attractions of major cities in China from air passenger flow data, 2001–2008: A particle swarm optimization approach. *The Professional Geographer*, 65(2), 265–282.
- Zhang, L., Ghader, S., Pack, M. L., Xiong, C., Darzi, A., Yang, M., Sun, Q., Kabiri, A., & Hu, S. (2020). An interactive COVID-19 mobility impact and social distancing analysis platform. *medRxiv*.

Part IV

Application and Policy Perspectives

Chapter 14

Citizen Mobility and the Growth of Infections During the COVID-19 Pandemic with the Effects of Government Restrictions in Western Europe



Mohd Sarim, Qunshan Zhao, and Nick Bailey

14.1 Introduction

The Coronavirus (COVID-19) pandemic has led to a massive disruption in the daily lives of people all around the world. The novel coronavirus (SARS-CoV-2) disease, also known as COVID-19 was first identified in December 2019 in the city of Wuhan, which is a part of the Hubei province in China. Due to a high frequency of travel and its highly infectious nature, it quickly spread across the globe. On January 30th, 2021, the World Health Organization officially declared the coronavirus outbreak as a public health emergency of international concern. It was later declared a global pandemic, owing to a rapid rise in the number of cases and deaths across multiple countries, including many in Western Europe which is the focus in this research. As of August 10th, 2020, over 19.7 million confirmed cases and more than 720 thousand deaths have been reported (W.H.O. 2020).

Since there are currently no effective vaccines for the disease at the moment, the primary strategies to mitigate its rapid spread have been to encourage greater personal hygiene and more social distancing. For the latter, several countries placed restrictions on the movement of people outside the home and restricted certain kinds of social or economic activity. These measures, also referred to as 'lockdowns' in the more extreme cases, have varied in terms of severity or stringency, dependent on the extent of the disease outbreak as well as the political choice of the governments. They have also varied in the details of which activities were permitted at any time: which kinds of social activity were permitted or whether people were required to work from home, for example. Governments have also been concerned to keep social distancing

M. Sarim · Q. Zhao (✉) · N. Bailey

Urban Big Data Centre, School of Social and Political Sciences, University of Glasgow, Glasgow G12 8RZ, UK

e-mail: Qunshan.Zhao@glasgow.ac.uk

N. Bailey

e-mail: Nick.Bailey@glasgow.ac.uk

measures to the minimum necessary, due to concerns about the social and economic impacts, and citizen compliance. To support compliance, it is important to study the effectiveness of these measures in limiting the transmission of the disease.

This research examines the mobility patterns across eight major Western European countries and their variation with the governmental responses. It seeks to identify different approaches to restrictions and to understand which kinds of restriction were most important for limiting the transmission of the disease. Mobility data is obtained from the COVID-19 Google Mobility Reports (Google 2020), whereas data on governmental responses and the number of cases is derived from the Oxford COVID-19 Government Response Tracker (Oxford: Blavatnik School of Government 2020).

14.2 Literature Review

Several studies have analyzed the mobility patterns after the start of the pandemic. Saha et al. (2020) used Google's Community Mobility Reports for India to analyze the impact of lockdown on community mobility. Exploratory analysis was used to plot state-wise changes in mobility at different categories of places. The data was divided into two timeframes: pre-lockdown and post-lockdown, and plots were generated showing the mobility changes for these timeframes. Additionally, spatial changes in mobility were shown for different times using maps. Warren and Skillman (2020) used anonymized data on mobile phone locations in the US to measure the changes in mobility. The location data for a designated time period before the lockdown was used to set up a baseline value for mobility. Then, the daily location data was used to calculate the percentage change in mobility for that day. These changes showed a dramatic drop in mobility across the US.

It is also important to understand the relationship between the transmission of COVID-19 and mobility to evaluate the impacts of the mobility restrictions. Most early studies in this regard focused on China. Kraemer et al. (2020) used real-time mobility data from Baidu Inc., and case data with travel history in order to determine the role of travel in the transmission of the virus. They used a generalized linear model to predict the daily case counts in other provinces, considering the mobility patterns in and outside Wuhan. The model predicted the number of cases with high accuracy. This work also helped to ascertain the effectiveness of the control measures in limiting the spread. A similar study was done by Zhao et al. (2020) that used correlation analysis to quantify the relationship between travel behavior and the number of transmissions to other areas. It found a positive correlation between the passenger traffic and the number of confirmed cases in 10 cities around Wuhan. Vinceti et al. (2020) used anonymized data from mobile phone movements to track citizen mobility in the most affected provinces of Italy. They modelled the daily trends of mobility and the number of cases using linear regression. The results showed a positive association of governmental interventions with the number of cases and the reduction in mobility.

Growth rate (cases per day) has been used to evaluate the impact of changes in mobility on the spread of the virus. Utsunomiya et al. (2020) created a framework to enable real-time analysis of the growth rate and growth acceleration using the moving regression technique and a hidden Markov model. The results showed that the growth acceleration started to decrease within a week of the restrictions being put in place. It also predicted a constant but small growth after 6 weeks. Badr et al. (2020) calculated the correlation between mobility and the growth rate of infections. Anonymized cell phone data was used to create a social distancing metric (mobility ratio), while the daily number of new cases were used to calculate the Growth Ratio. Correlations were then computed between these two metrics for different time lag intervals. The results showed that the correlation was highest for a time lag of 9–12 days, which is consistent with the incubation time of COVID-19.

Most studies have focused on a single country, with only a few (Seibold et al. 2020; Utsunomiya et al. 2020) expanding their analysis across multiple countries. However, only a few studies have used the Google Mobility Reports to compare the impacts of restrictions across various countries (Zhu et al. 2020), with no study focusing on different Western European countries with serious COVID-19 infection.

14.3 Data

14.3.1 *Mobility Levels*

The COVID-19 Community Mobility dataset (Google 2020) published by Google was used to assess the mobility levels. This data tracks the changes in mobility at six different categories of places using the location of mobile phones. Data is collected from devices that have ‘Location History’ setting turned on voluntarily by Google users. The mobility changes are compared to a baseline value for each day of the week. These baseline values are calculated by using the median mobility for the given day over a 5-weeks period over January and February 2020. Different days of the week have different baseline values, which means that comparison of day-to-day changes in mobility is not suitable. Another important point to note is the difference in the calculation of mobility changes at residences and all other places. The residential changes are measured as a change in the average duration of time (in hours) spent at home. Since a day only has 24 h, the percentage changes are limited in this case. On the other hand, the changes in all the other categories are measured in terms of the number of visitors, which means the changes are not bounded. The different categories of places are:

- Residential—areas earmarked as residential.
- Parks—areas officially designated as parks, such as National parks, castles and public gardens.
- Grocery and Pharmacy—places involving essential trips, like groceries and pharmacies.

- Retail and recreation—Shopping centers, restaurants, theatres and other places of recreational activities.
- Transit Stations—Bus, train and subway stations, as well as taxi stands and car rental agencies.
- Workplaces—locations marked as workplaces.

The values of transit and retail mobility were found to be highly correlated with the workplace mobility. Therefore, four categories of places are selected for this analysis: Residential, Parks, Grocery and Pharmacy, and Workplaces.

14.3.2 *Restrictions and Number of Cases*

The restrictions on mobility are assessed using the Oxford COVID-19 Government Response Tracker (Oxford: Blavatnik School of Government 2020). The governmental responses are measured through 17 indicators that correspond to various policy measures. Eight indicators identify the closure and containment policies, such as the closures of schools and workplaces. Five indicators measure healthcare responses such as testing policies and emergency investments, while the other four indicators correspond to the economic policies such as income support and fiscal measures. The Stringency Index, which is an average of eight closure and containment policy indicators and one healthcare measure, is used to quantify the responses of the governments to the pandemic (Hale et al. 2020). Using the Stringency Index has the advantage of having a graduated scale to model the restrictions, instead of a binary variable that denotes the presence of a lockdown. The Stringency Index ranges from 0 (no restrictions) to 100 (most stringent restrictions). In order to better visualize the mobility trends with introduction of restrictions, the Stringency Index is converted to the Freedom of Association Index, which is calculated as:

$$\text{Freedom of Association Index} = (100 - \text{Stringency Index})$$

The OxCGRT dataset also tracks the total number of confirmed cases and deaths due to COVID-19. These cumulative values are used to calculate the daily rise in the number of cases.

14.4 Lockdown Level Classification Based on Mobility Patterns

Plots are created for each country using the mobility changes at the selected categories of places (Residential, Parks, Grocery and Pharmacy, Workplaces) as well as the freedom of association index (All the analysis source codes are available at <https://github.com/mdsarim/COVID-19>). As the data captures weekly changes, the x-axis

is divided into weeks to better understand the patterns of mobility. For each country, a dramatic change in mobility is observed around the dates when there is a sharp decline in the freedom of association index ('lockdown date'). However, the freedom of association started to decline before the lockdown date. This was found to be the result of preventive measures taken before a full lockdown was put in place. These measures included closure of educational institutions, ban on gatherings and in some cases, restrictions on movement of people.

For all countries, some days witnessed unusual decreases in workplace mobility, and corresponding increases in residential mobility. Further examinations revealed those days to be public holidays. Each country witnessed a significant rise in residential mobility after the lockdown. There was a weekly pattern where the weekends saw lower changes than weekdays, which is understandable as workers tend to remain at home on weekends. The mobility changes for other categories, however, were not similar for each country. Based on the values for different categories, three distinct patterns were observed for the eight Western European countries.

14.4.1 Strict—France, Italy and Spain

The mobility patterns in France, Italy and Spain were found to be similar (Figs. 14.1, 14.2 and 14.3). Following the lockdown, there was a sharp decline in mobility at

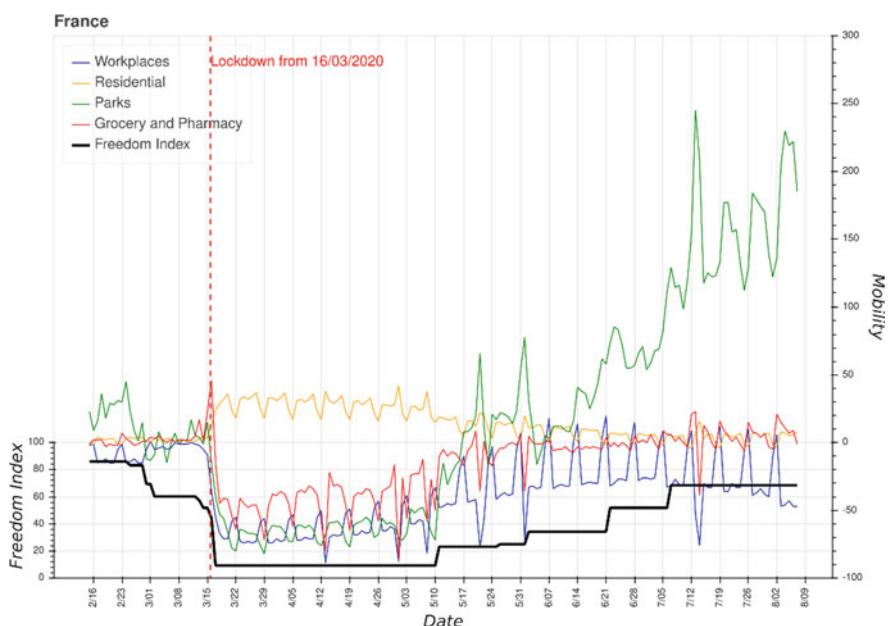


Fig. 14.1 Mobility and freedom of association in France

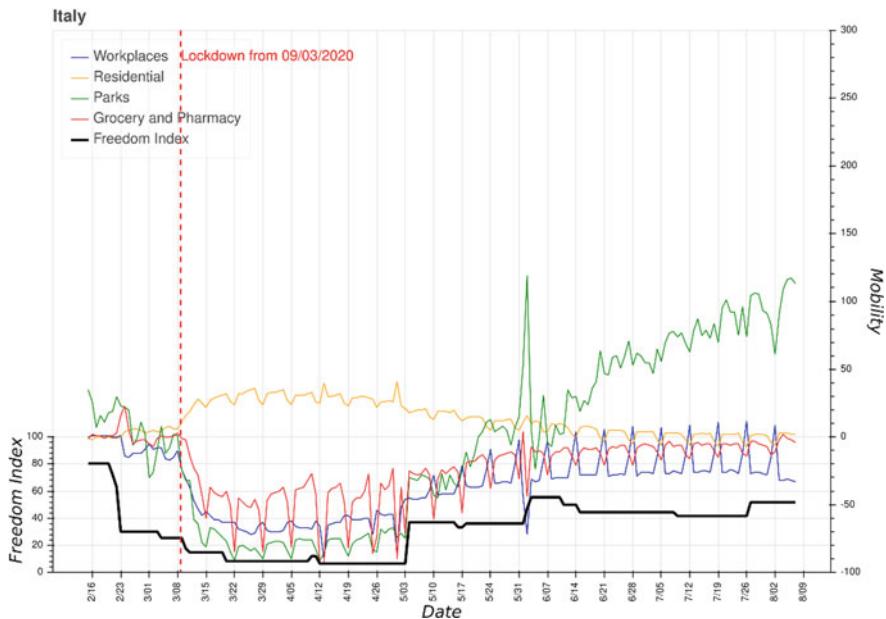


Fig. 14.2 Mobility and freedom of association in Italy

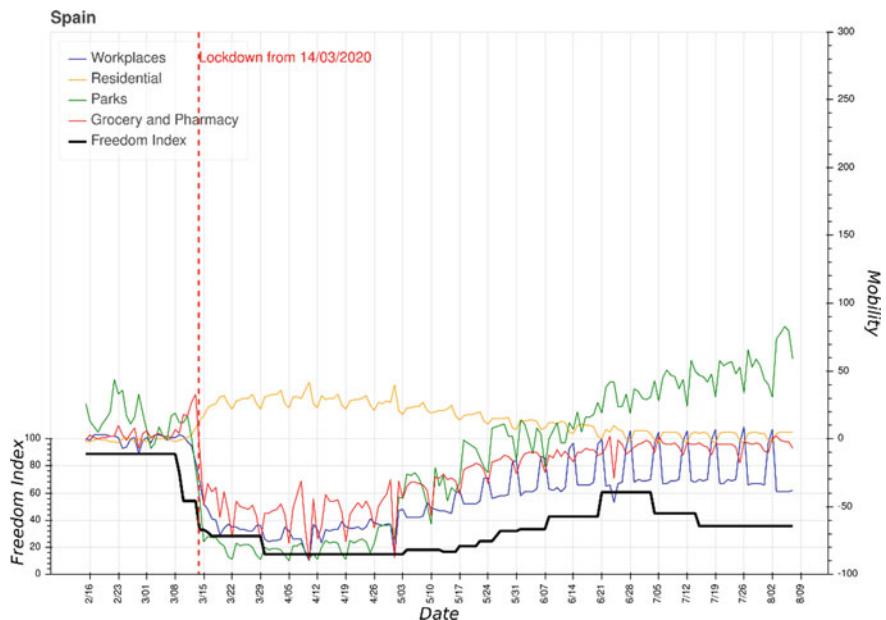


Fig. 14.3 Mobility and freedom of association in Spain

workplaces, parks and groceries in these countries. The freedom of association was the lowest (amongst the eight countries considered), suggesting the imposition of the strictest measures to curb the spread of the virus. This also pointed to a high compliance among the people. Such strict measures were deemed necessary as these were some of the worst-affected countries in terms of infections and deaths.

Mobility changes in Italy started when lockdown measures were imposed in the Italian provinces of Lodi and Padua on February 21st. Educational institutions and workplaces were closed, and public gatherings were banned ([Metro 2020](#)). France and Spain also initiated similar restrictions in early March. This led to a decline in the freedom of association even before the official lockdown date, as seen in the plots. Later, due to a continuous rise in the number of infections, country-wide lockdown measures were imposed in Italy (March 9th), Spain (March 14th) and France (March 16th). All non-essential businesses were ordered to close, and workers were encouraged to work from home if possible. Strict stay-at-home orders meant that people could go out only for essential supplies and in emergencies. This resulted in a decline of 60–70% in workplace mobility. Due to the closure of parks and public spaces, mobility for parks also saw a sharp decline ([Reuters 2020](#)), with the weekdays witnessing a fall of at least 60%. The weekends saw even more decline, suggesting a higher baseline value at parks during weekends. Mobility at groceries and pharmacies also decreased significantly post lockdown, with a change of 40% compared to the baseline levels until late April.

Lockdown measures were gradually eased since May, with workplaces and parks allowed to open ([Aljazeera 2020](#); [The Guardian 2020a](#)). This resulted in a significant increase in grocery and parks' mobility patterns, and a decrease in time spent at homes. While the mobility at groceries and homes hit pre-lockdown levels, more and more people were visiting parks. In case of workplaces, the mobility increased with the easing of restrictions, but did not reach pre-lockdown levels. Amid fears of a second wave of infections, Spain and Italy reintroduced some restrictions, but these did not have much impact on mobility.

14.4.2 Intermediate—Belgium and the UK

The freedom of association values for Belgium and the UK were close to those of the first group (scores around 20–22/100 as compared to 6–16/100, Figs. [14.4](#) and [14.5](#)). While they witnessed a steep decline in workplace mobility, they saw relatively small decrease in that of parks and groceries post-lockdown.

Belgium banned public gatherings and closed schools, cafes and other public places from March 12th. The UK government also advised against non-essential travel and public gatherings on March 16th, and ordered the closure of schools, restaurants and pubs from March 20th. These are also reflected in the freedom of association, which declined around the same time. Following a rise in cases, a country-wide lockdown was imposed from March 17th in Belgium, and from March 23rd in the UK. Stay-at-home orders were put in place, along with the closure of

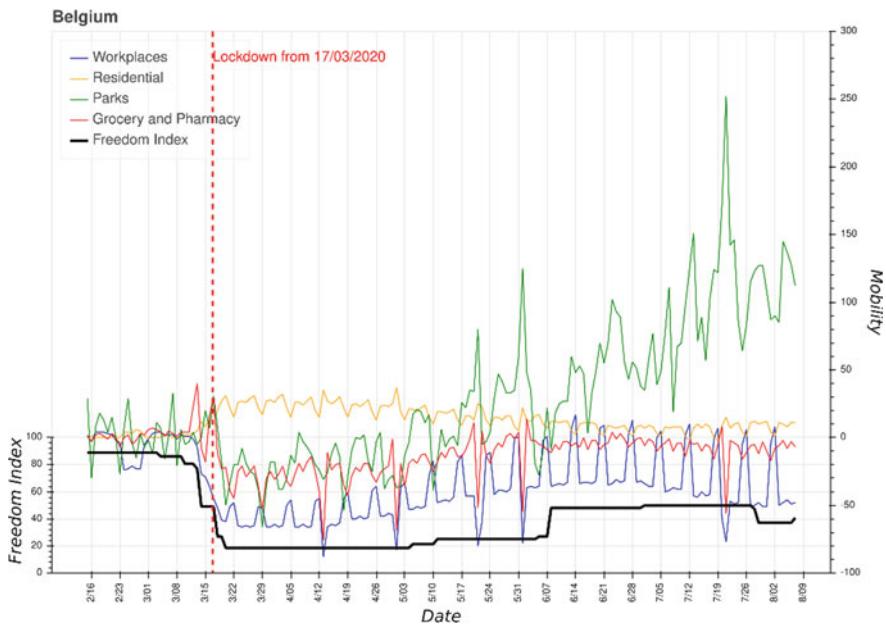


Fig. 14.4 Mobility and freedom of association in Belgium

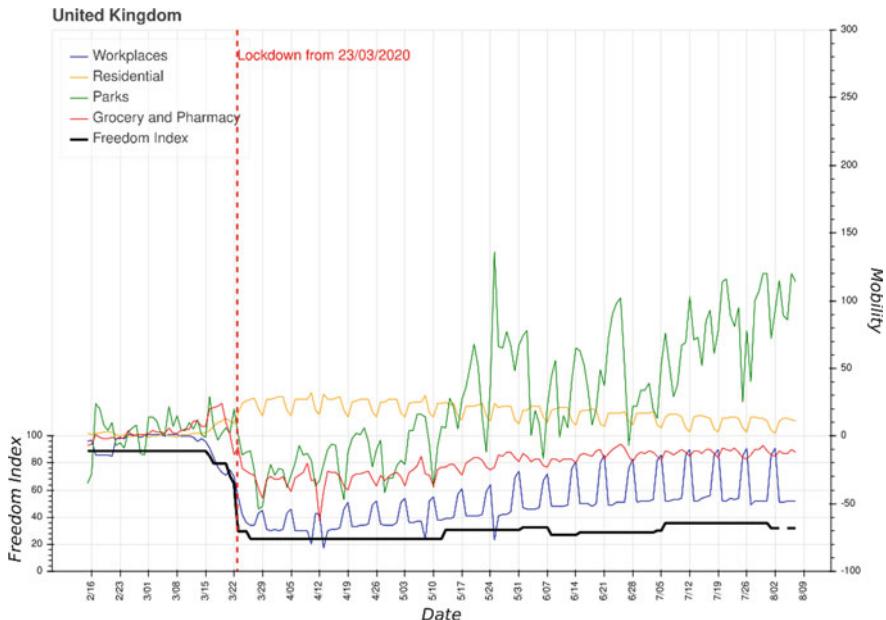


Fig. 14.5 Mobility and freedom of association in the UK

all non-essential businesses and shops (Belgian Federal Government 2020). The UK also introduced the ‘Coronavirus Act 2020’, which assigned emergency powers to the government for tackling the pandemic (Cabinet Office 2020; UK Legislations 2020). As a result, workplace mobility decreased by more than 60%. These countries did not close parks and allowed citizens to go out for physical activities such as running and cycling. However, such visits were limited by visiting frequency. As a result, the mobility at parks initially remained below pre-lockdown levels.

Both Belgium and the UK started easing the lockdown restrictions in early May (The Guardian 2020b). This resulted in a gradual increase in mobility. However, the changes were different for each category. While mobility at groceries and workplaces increased slowly, there was a huge rise in park visits. This could be due to the lifting of restrictions on the number of park visits, as well as good weather conditions. One important point to note is that despite easing of restrictions, the workplace mobility in Belgium peaked in June and started going down long before some restrictions were imposed again.

14.5 Flexible—Germany, Switzerland and the Netherlands

Germany, Switzerland, and the Netherlands witnessed similar values of freedom of association as Belgium and the UK (Figs. 14.6, 14.7, and 14.8). However, the

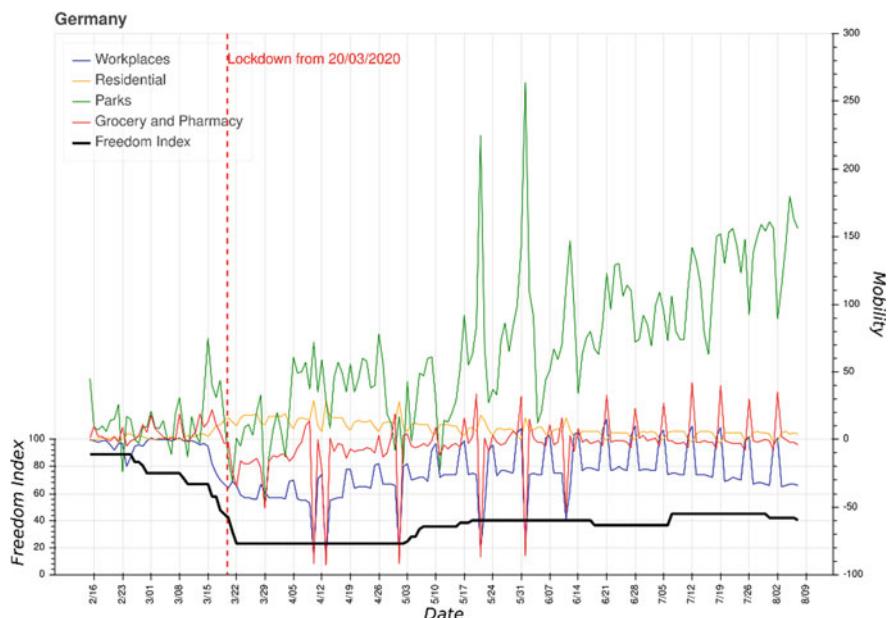


Fig. 14.6 Mobility and freedom of association in Germany

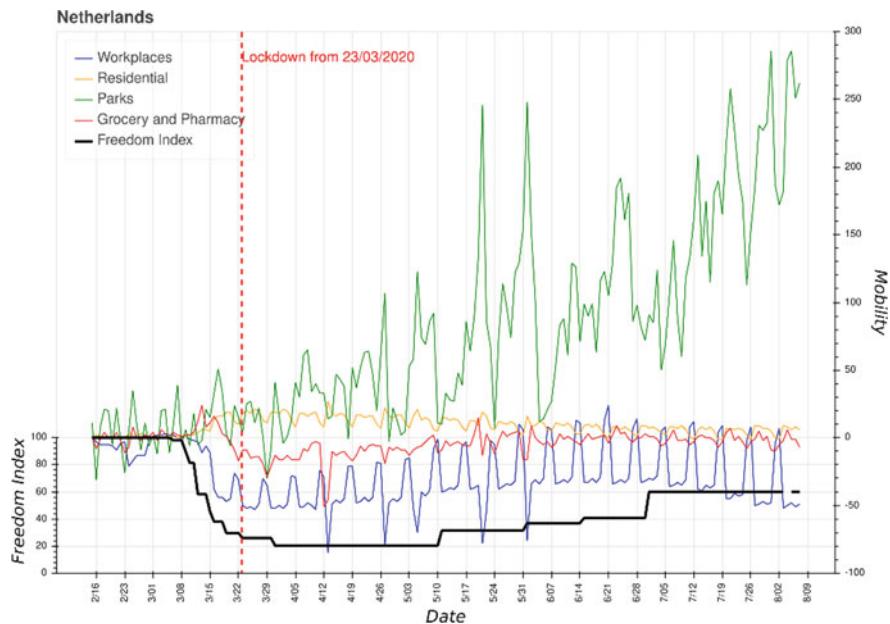


Fig. 14.7 Mobility and freedom of association in the Netherlands

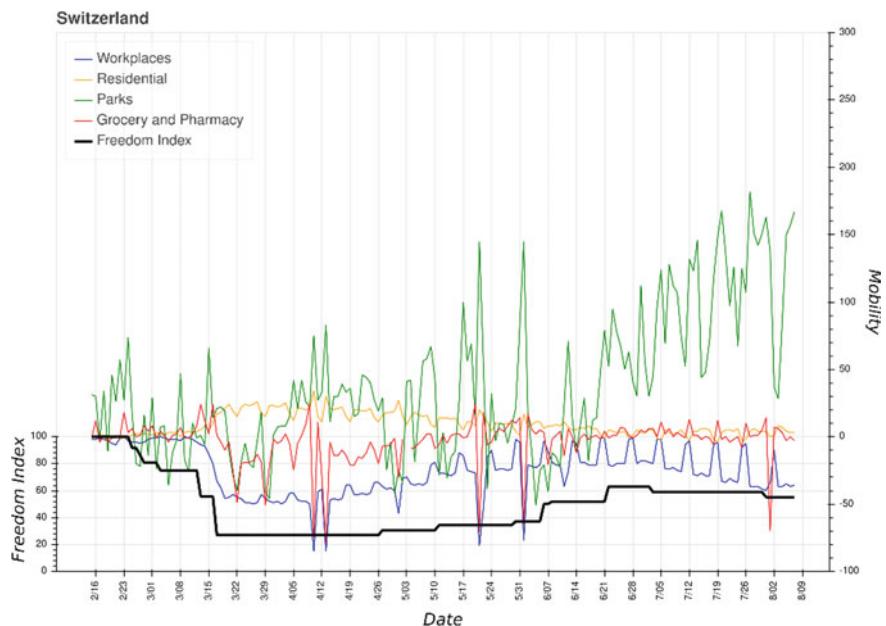


Fig. 14.8 Mobility and freedom of association in Switzerland

mobility changes were smaller at groceries and workplaces, with a huge rise in park visits.

Preventive measures were initiated in Germany and Switzerland towards the end of February, with closure of schools and cancellation of public events and gatherings (The Federal Council 2020). Similar steps were taken by the Netherlands in early March. Consequently, the freedom of association also started to decrease. The German states of Bavaria and Saarland were the first to introduce lockdown from March 20th (The Independent 2020), which was extended to other parts of the country on March 22nd. Although there was no official lockdown in Switzerland, schools and colleges were ordered to close from March 13th. Bars, restaurants and non-essential shops were also closed from March 16th. The Netherlands also enforced lockdown measures from March 23rd. As a result, mobility at workplaces decreased significantly, though it remained less than 50% (except public holidays). Mobility at groceries only decreased slightly (<20%) after lockdown, and increase in residential mobility was also small relative to the other categories.

A distinct feature of this category was the mobility at parks. These countries did not restrict visits to parks, which led to a general increase in mobility. Notably, the baseline values used were from January, when the weather is generally cold. Nevertheless, this continuous upward trend is not present for any of the other groups of countries, which suggests that lockdown did not reduce the number of visitors in parks. In Germany and Switzerland, there was a dip at the end of March and mid-May. On further examination, these days were found to have bad weather and low temperatures.

Since the end of April, these countries also started easing restrictions. This resulted in an increase in workplace mobility, and a corresponding decrease in time spent at home. The grocery mobility crept back to pre-lockdown levels, while there was a huge increase in park visits. Like Belgium, these countries also saw a peak in workplace mobility in June, and a downward trend afterwards.

14.6 Mobility Correlation Analysis

14.6.1 *Correlation Between Freedom of Association and Mobility*

To further understand the relationship between the freedom of association and the mobility at various places, correlation coefficients were calculated between the freedom of association index and each category (Table 14.1). These correlation coefficient values are consistent with the earlier classification. Residential mobility for each country shows a strong negative correlation with the freedom of association, since the lockdown required people to stay at home. France, Italy and Spain show strong correlation for all other categories as well, which means that restrictions were imposed and followed at all places. As mentioned earlier, parks in these countries

Table 14.1 Correlation coefficients between freedom of association and mobility

Country	Residential	Parks	Grocery	Workplaces
France	-0.829**	0.739**	0.741**	0.678**
Italy	-0.828**	0.771**	0.699**	0.769**
Spain	-0.777**	0.571**	0.714**	0.802**
Belgium	-0.803**	0.185	0.593**	0.680**
United Kingdom	-0.783**	-0.116	0.790**	0.821**
Germany	-0.636**	-0.146	0.287**	0.558**
Netherlands	-0.754**	-0.015	0.411**	0.507**
Switzerland	-0.793**	0.171*	0.328**	0.716**

* $p < 0.05$, ** $p < 0.01$

were closed during the lockdown, which caused a large decrease in mobility. In case of Belgium and the UK, all categories except parks were found to be highly correlated with the freedom of association. The mobility at groceries was highly correlated with the freedom of association except in the third group of countries (Germany, Switzerland and the Netherlands).

14.6.2 Correlation Between Growth Rate Ratio and Mobility Change

The OxCGRT dataset tracks the cumulative number of confirmed cases for each country. The daily number of cases on a particular day was calculated by subtracting the number of cases on the previous day from the number of cases on that day. To quantify the growth of infections, a measure called the Growth Rate Ratio (GRR) was used. The GRR, defined by Badr et al. (2020) in their study, is the ratio of the logarithmic change of cases over previous 3 days to that over the previous week. The daily cases were used to calculate the GRR using the following equation:

$$GR_i^t = \frac{\log\left(\sum_{i=2}^t \frac{C_i^t}{C_{i-1}^t}\right)}{\log\left(\sum_{i=6}^t \frac{C_i^t}{C_{i-7}^t}\right)} \quad (1)$$

where GR_i^t is the GRR and C_i^t is the number of cases in country i on day t . A GRR greater than 1 suggests a relative increase in growth rate compared to the previous week, with a decrease denoted by a value of less than 1.

Several studies have found the median incubation time of COVID-19 to be around 5 days, with 95% people showing symptoms within 12.5 days (Lauer et al. 2020; Li et al. 2020). As the incubation time differs in each case, accounting for the delay in reporting, 14 days has been considered the maximum incubation time. This has

also been used by countries as the quarantine period for anyone who has been at risk of infection. A time lag is expected between the change in mobility and the subsequent change in growth rate, since mobility changes are not expected to affect the growth rate immediately. Correlation coefficients were calculated between the mobility changes at different places and the 7 days rolling average of GRR values using different time lags. The statistically significant results for all countries are shown in Fig. 14.9 ($p < 0.05$).

The mobility values at each category of places were used to calculate correlation coefficients with the GRR. For Spain and Italy, the correlations peaked between a lag of 14 and 18 days, with all values above 0.6. Grocery mobility showed the weakest correlation with GRR, with all other places showing values above 0.9 during the peak. After 20 days the correlations declined sharply, which could be due to a decrease in the number of reported cases while mobility remaining stable. In case of France, the correlations started peaking around a lag of 18 days, and then started declining after 22 days. The GRR in Germany, Switzerland and the Netherlands showed high correlation with the workplace and residential mobility, and only a weak correlation with mobility at grocery and pharmacy. The mobility at parks showed a negative correlation for most of the statistically significant correlation values. The

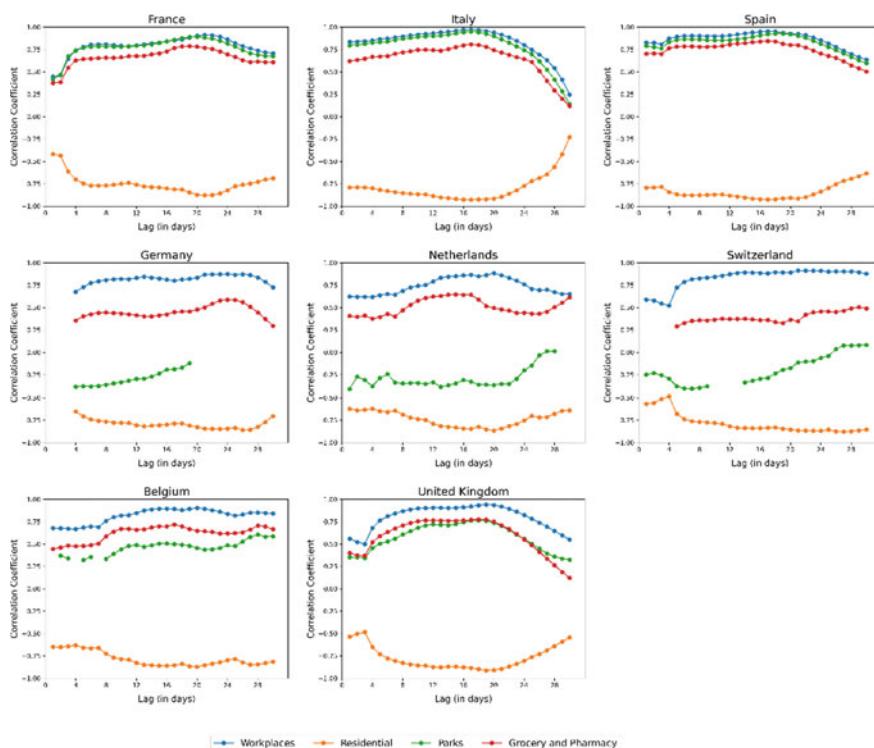


Fig. 14.9 Correlation Coefficients between GRR and mobility change for different time lags

coefficients in case of Belgium and the UK also followed largely the same patterns, with the notable difference being with parks, where the correlation was stronger than Germany, Netherlands, and Switzerland but weaker than Italy, France, and Spain. A notable observation is that there is a different time lag value for which the country achieves a peak value of correlation. This suggests that there are some-country-specific factors in play that have influenced the rate of growth of new infections. Nevertheless, there is satisfactory evidence to support the role of social distancing in reducing the spread of the virus.

14.7 Discussion and Conclusions

This research used aggregated mobility data provided by Google to monitor the changes in mobility in a number of major Western European countries. The mobility data was compared to the government responses, which were captured by the Oxford COVID-19 Government Response Tracker (OxCGRT). On the basis of mobility patterns, the eight countries were divided into three lockdown categories (strict, intermediate, and flexible). Furthermore, the daily number of cases was used to calculate the GRR for each country. Correlation coefficients were then calculated between GRR and the mobility changes, with different time lags. The results show that the GRR is highly correlated with the decrease in workplace and residential mobility for a time lag window, which suggests that the mobility restrictions could have been effective in decreasing the number of infections.

Based on the analysis of human mobility, freedom of association index, and COVID-19 growth rate ratio, the obvious conclusion is that the strict restriction of mobility such as working from home order is extremely important to limit the spread of COVID-19 infectious. Further, the limitation of mobility will not result in an immediate COVID-19 growth rate decline due to the virus incubation time. Thus, it is very important to impose enough time of lockdown (at least 2 to 3 weeks) to see the control of the virus from mobility restriction. Specifically, doing essential shopping and exercising in the parks with carefully social distancing and usage of face masks is not necessary to increase the growth rate of the virus.

There are a few limitations of this research that need to be discussed. Firstly, discussing the growth rates for different countries presents various challenges. For instance, the testing policies are not the same for each country, which could have effects on the number of reported cases. The level of infection and community spread before the restrictions also varies between countries. Secondly, the relationship between cases and mobility could work both ways. An increase in the number of cases could also result in a decrease in mobility, as people refrain from going out due to rising infections. Finally, there may be other factors such as preventive measures (handwashing and wearing masks) that could also affect the growth rate. The effect of such factors has not been discussed in this analysis.

The use of Google's Community Mobility data as a measure of population mobility also has some limitation. Firstly, the data collection methodology is not clearly

explained by Google, and it is not known whether any modelling techniques were used while collecting the data. Secondly, this type of data collection may only be suitable for developed countries. It may not work well for developing countries, where there may be a significantly lower percentage of people who own and use location-enabled smartphones, even with the access to the internet. The results in this analysis would be biased towards people who use smartphones and could be significantly different from the real-world scenario. Lastly, due to the reliance on the ‘Location History’ feature in the user’s phone being turned on, the data might not represent the actual mobility patterns. Therefore, caution should be exercised while using the results derived from this type of data. Additional methods and data validation should be used to further confirm the findings from this analysis.

Acknowledgements The work was made possible by the ESRC’s on-going support for the Urban Big Data Centre (UBDC) [ES/L011921/1 and ES/S007105/1]. The authors would like to thank the anonymous reviewers for their insightful comments and suggestions on an earlier version of this manuscript.

References

- Aljazeera. (2020). Italy leads Europe in easing coronavirus lockdown measures [www Document]. URL <https://www.aljazeera.com/news/2020/05/italy-leads-europe-easing-coronavirus-lockdown-measures-200504101525868.html>. Accessed 10 Aug 20.
- Badr, H. S., Du, H., Marshall, M., Dong, E., Squire, M. M., & Gardner, L. M. (2020). Association between mobility patterns and COVID-19 transmission in the USA: A mathematical modelling study. *The Lancet Infectious Diseases*. [https://doi.org/10.1016/S1473-3099\(20\)30553-3](https://doi.org/10.1016/S1473-3099(20)30553-3).
- Belgian Federal Government. (2020). Coronavirus: Reinforced measures [WWW Document]. URL https://www.belgium.be/en/news/2020/coronavirus_reinforced_measures. Accessed 10 Aug 20.
- Cabinet Office. (2020). Staying at home and away from others (social distancing) [www Document]. URL <https://www.gov.uk/government/publications/full-guidance-on-staying-at-home-and-away-from-others/full-guidance-on-staying-at-home-and-away-from-others>. Accessed 10 Aug 20.
- Google. (2020). Community mobility reports [WWW Document]. URL <https://www.google.com/covid19/mobility/>. Accessed 10 Aug 20.
- Hale, T., Petherick, A., Phillips, T., & Webster, S. (2020). Variation in government responses to COVID-19. *Blavatnik school of government working paper*.
- Kraemer, M. U. G., Yang, C.-H., Gutierrez, B., Wu, C.-H., Klein, B., Pigott, D. M., Covid, O., Hanage, W. P., Brownstein, J. S., Layman, M., Vespignani, A., Tian, H., Dye, C., Pybus, O. G., & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China 6.
- Lauer, S. A., Grantz, K. H., Bi, Q., Jones, F. K., Zheng, Q., Meredith, H. R., et al. (2020). The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: Estimation and application. *Annals of Internal Medicine*, 172, 577–582. <https://doi.org/10.7326/M20-0504>.
- Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., et al. (2020). Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. *New England Journal of Medicine*, 382, 1199–1207. <https://doi.org/10.1056/NEJMoa2001316>.

- Metro. (2020). What towns in Italy are on lockdown because of coronavirus? [WWW Document]. URL <https://metro.co.uk/2020/02/25/towns-italy-lockdown-coronavirus-12298246/>. Accessed 10 Aug 20.
- Oxford: Blavatnik School of Government. (2020). Oxford COVID-19 Government Response Tracker, Blavatnik School of Government [WWW Document]. URL <https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker>. Accessed 10 Aug 20.
- Reuters. (2020). Italy closes all parks in further tightening of coronavirus lockdown [WWW Document]. URL <https://uk.reuters.com/article/uk-health-coronavirus-italy-crackdown/italy-closes-all-parks-in-further-tightening-of-coronavirus-lockdown-idUKKBN21739M>. Accessed 10 Aug 20.
- Saha, J., Barman, B., & Chouhan, P. (2020). Lockdown for COVID-19 and its impact on community mobility in India: An analysis of the COVID-19 Community Mobility Reports, 2020. *Children and Youth Services Review*, 116,. <https://doi.org/10.1016/j.childyouth.2020.105160>.
- Seibold, B., Vucetic, Z., & Vucetic, S. (2020). Quantitative Relationship between Population Mobility and COVID-19 Growth Rate based on 14 Countries. [arXiv:2006.02459](https://arxiv.org/abs/2006.02459) [physics, q-bio].
- The Federal Council. (2020). Coronavirus: Federal Council bans large-scale events [WWW Document]. URL <https://www.admin.ch/gov/en/start/documentation/media-releases.msg-id-78289.html>. Accessed 10 Aug 20.
- The Guardian. (2020a). France, Italy and Spain prepare to ease coronavirus lockdowns [WWW Document]. URL <https://www.theguardian.com/world/2020/apr/26/france-italy-spain-prepare-to-ease-coronavirus-lockdowns>. Accessed 10 Aug 20.
- The Guardian. (2020b). Point by point: The UK government's plan to leave lockdown [WWW Document]. URL <https://www.theguardian.com/politics/2020/may/11/what-is-covered-in-the-uk-governments-lockdown-easing-plan>. Accessed 10 Aug 20.
- The Independent. (2020). Coronavirus: Germany introduces first lockdown as Bavaria imposes two-week ban on going outside [WWW Document]. URL <https://www.independent.co.uk/news/world/europe/coronavirus-germany-update-lockdown-today-bavaria-latest-a9413746.html>. Accessed 10 Aug 20.
- UK Legislations. (2020). Coronavirus Act 2020 [WWW Document]. URL <https://www.legislation.gov.uk/ukpga/2020/7/contents/enacted/data.htm>. Accessed 10 Aug 20.
- Utsunomiya, Y. T., Utsunomiya, A. T. H., Torrecilha, R. B. P., de Paulan, S. C., Milanesi, M., & Garcia, J. F. (2020). Growth rate and acceleration analysis of the COVID-19 pandemic reveals the effect of public health measures in real time. *Frontiers in Medicine*, 7. <https://doi.org/10.3389/fmed.2020.00247>.
- Vinceti, M., Filippini, T., Rothman, K. J., Ferrari, F., Goffi, A., Maffeis, G., et al. (2020). Lockdown timing and efficacy in controlling COVID-19 using mobile phone tracking. *EClinicalMedicine*, 25, 100457. <https://doi.org/10.1016/j.eclim.2020.100457>.
- Warren, M. S., & Skillman, S. W. (2020). Mobility changes in response to COVID-19. [arXiv:2003.14228](https://arxiv.org/abs/2003.14228) [cs].
- W.H.O. (2020). Coronavirus disease situation report-203 [WWW Document]. URL <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>. Accessed 10 Aug 20.
- Zhao, S., Zhuang, Z., Cao, P., Ran, J., Gao, D., Lou, Y., et al. (2020). Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: A correlational analysis. *Journal of Travel Medicine*. <https://doi.org/10.1093/jtm/taaa022>.
- Zhu, D., Mishra, S. R., Han, X., Santo, K. (2020). Social distancing in Latin America during the COVID-19 pandemic: An analysis using the Stringency Index and Google Community Mobility Reports. *Journal of Travel Medicine*, 125. <https://doi.org/10.1093/jtm/taaa125>.

Chapter 15

A Mathematical Model for Evaluating the Medical Resource Availability of COVID-19 in Time and Space



Fei-Ying Kuo and Tzai-Hung Wen

15.1 Introduction

Global pandemic COVID-19 rapidly spreads to many countries, causing enormous economic loss and many deaths in 2020 (Remuzzi & Remuzzi, 2020). Low accessibility and availability of medical resources (including screening/testing, prevention, and treatment facilities) for the at-risk populations is an urgent, important issue in preventing and controlling the pandemic (Ji, Ma, Peppelenbosch, & Pan, 2020). Thus, evaluating the accessibility and availability of medical resources is an essential topic for pandemic preparedness and responses. Healthcare accessibility and availability are determined by healthcare facilities' demand and service capacity (Wang, 2012). Floating catchment analysis (FCA) is a widely used model framework for measuring geographical accessibility (McGrail & Humphreys, 2014), and McGrail (2012) has comprehensively reviewed its derivative versions. In previous studies, the spatial distribution of demand is usually assumed as static and unchanged over time. However, epidemic-induced demand for medical resources, including the exposed and infected population, significantly varies over space and time (Weissman et al., 2020). Matching the availability of medical resources should consider the demand and supply at different epidemic stages simultaneously. Recent studies considered time-varying dynamic demand resulted from daily routines and human mobility for measuring accessibility (Ma, Luo, Wan, Hu, & Peng, 2018; Xia et al., 2019). Although these studies incorporated daily spatial behaviors of individuals reflecting the different spatial distributions of demand at distinct periods, these considerations cannot comprehensively capture the dynamic nature of epidemic-induced demand in time and space. In other words, the dynamics of exposed and infected populations reflect the time-varying medical demand, and spatial clusters of an epidemic may

F.-Y. Kuo · T.-H. Wen (✉)

Department of Geography, National Taiwan University, Taipei, Taiwan

e-mail: wenthung@ntu.edu.tw

also shift over time. Therefore, to avoid inappropriate resource allocation, it is necessary to incorporate the spatial-temporal dynamics of epidemic-induced demand into assessing medical resource availability.

The spatial mechanic dynamics of an epidemic is usually simulated by the family of susceptible-exposed-infectious-recovered (SEIR) meta-population models (Wang & Wu, 2018). An SEIR model separates the population into different stages of infection, and each individual will shift from one stage to the next due to epidemic propagation (Tsai, Huang, Wen, Sun, & Yen, 2011). The meta-population, the extended concept, highlights that the propagation speed and transmission risk of an epidemic would be different at distinct places due to local environmental heterogeneity and human movements among locations. These local risk factors may trigger an outbreak to diffuse across the surface (Lima, De Domenico, Pejovic, & Musolesi, 2015). Therefore, the SEIR meta-population model framework can be effectively used for representing the spatial-temporal dynamics of an epidemic. Nonetheless, the capacity and availability of medical resources for the exposed and infected population may influence consecutive epidemic propagation and resource accessibilities. High accessibility of medical resources in some areas may protect local at-risk populations and reduce disease transmission. Then, these resources in epidemic-controlled areas may provide neighboring regions to mitigate the prolongation of an epidemic spreading. On the other hand, low accessibility may aggravate an epidemic and enlarge the need and consumption of medical resources, which lowers resource availability.

As the dynamics of resource accessibility and the epidemic dynamic process interact, they cannot be measured independently. Therefore, we proposed an integrated mathematical model, named Epi-RA, to combine two components, including epidemic dynamic process and resource accessibility, to measure the medical resource availability of the COVID-19 epidemic. The epidemic dynamic process is developed based on the SEPIA model proposed by Gatto et al. (2020). The model considers the isolation measure, and the pre-symptomatic stage of infection, the essential characteristics of the COVID-19 epidemic. Moreover, resource accessibility is implemented from the two-step floating catchment area (2SFCA) model, which is widely used to measure spatial accessibility (Guagliardo, 2004; Wang, 2012). The community screening/testing capacity of COVID-19 is regarded as the medical resource availability in this study. The Taipei metropolitan area, one of the major cities in East Asia, is used as a study region to demonstrate the proposed model's feasibility and effectiveness.

15.2 Methods and Data

15.2.1 Integrated Model Framework

We proposed an integrated model, named Epi-RA, which combines epidemic dynamic process (Epi) and resource accessibility (RA), to evaluate the changing

ration levels of medical resources during different epidemic periods. Figure 15.1 illustrates the framework of our integrated model. It demonstrates how epidemic dynamic processes and resource accessibility interacted with each other. Our model divides the epidemic dynamic process into eight disease progression stages, including susceptible (S), exposed (E), pre-symptomatic (P), severe symptomatic (I), mild/no symptomatic (A), isolated (Q), recovered (R), and dead (D). As individuals at the infectious stages of P, I, and A can transmit the pathogen to others, those could reflect the epidemic-induced demand which needs to acquire the COVID-19 rapid tests. The spatial distribution of time-varying demand can be determined from the epidemic dynamic modeling process in time and space. Subsequently, resource accessibility modeling is used to estimate the resource availability of the epidemic-induced demand at a specific period. The matching results of resource availability influence the succeeding transmission propagation, which can be revealed from the following epidemic dynamic process.

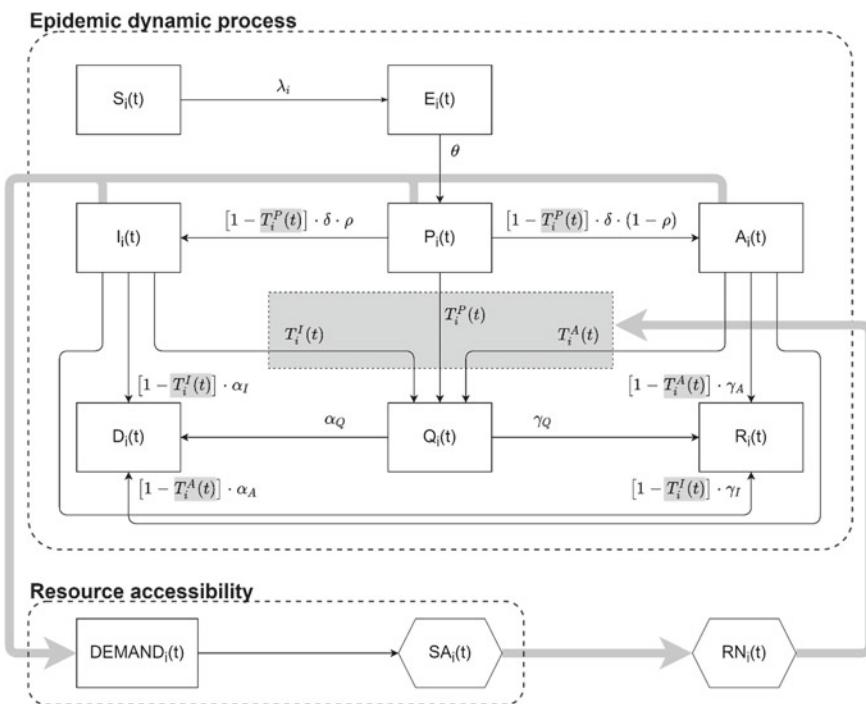


Fig. 15.1 The framework of the integrated model. The subscript i represents a specific location, and (t) represents the situation at time step t . The descriptions of each parameter are shown in Table 15.1

15.2.1.1 Resource Accessibility Sub-Model

The resource accessibility sub-model is implemented to evaluate the COVID-19 testing resource availability for epidemic-induced demand, including the individuals in pre-symptomatic, severe symptomatic, and mild/no symptomatic stages. The demand varies over space and time through epidemic dynamic process modeling (as shown in Eq. 15.1).

$$DEMAND_i(t) = v_P P_i(t) + v_I I_i(t) + v_A A_i(t), \quad (15.1)$$

where $P_i(t)$, $I_i(t)$, and $A_i(t)$ are the number of people in pre-symptomatic, severe symptomatic, and mild/no symptomatic stages at location i at time t ; v_P , v_I , and v_A denote the proportion of people who are required to be tested from these three groups, respectively.

Based on these dynamic demands, we used the generalized 2SFCA (Wang, 2012) to measure the time-varying spatial accessibility (SA) for every location:

$$SA_i(t) = \sum_{j=1}^n \frac{S_j f(d_{ij})}{\sum_{k=1}^m DEMAND_k(t) f(d_{kj})} \quad (15.2)$$

where S_j denotes the amount of resource that facility j can provide; d_{ij} denotes the geographical distance between location i and facility j ; $f(d_{ij})$ is a function transforming d_{ij} into a weight to represent the interaction strength between these two places. We adopted the negative-power transformation ($f(d_{ij}) = d_{ij}^{-2}$), which is one of the commonly used transformations in the literature (Kwan, 1998). According to Eq. 15.2, the accessibility varies across both the study region and the epidemic period.

15.2.1.2 Epidemic Dynamic Process Sub-Model

The epidemic dynamic process sub-model is to simulate the spatial-temporal dynamics of disease transmission for estimating epidemic-induced demand. Equation 15.3 shows details of our epidemic process component, which is developed based on the SPEIA model proposed by Gatto et al. (2020):

$$\frac{dS_i}{dt} = -\lambda_i(t) \cdot S_i \quad (15.3)$$

$$\frac{dE_i}{dt} = \lambda_i(t) \cdot S_i - \theta E_i$$

$$\frac{dP_i}{dt} = \theta \cdot E_i - \{T_i^P(t) + [1 - T_i^P(t)] \cdot \delta\} P_i$$

$$\begin{aligned}\frac{dI_i}{dt} &= [1 - T_i^P(t)] \cdot \rho \cdot \delta \cdot P_i - \{T_i^I(t) + [1 - T_i^I(t)](\alpha_I + \gamma_I)\} I_i \\ \frac{dA_i}{dt} &= [1 - T_i^P(t)] \cdot (1 - \rho) \cdot \delta \cdot P_i - \{T_i^A(t) + [1 - T_i^A(t)](\alpha_A + \gamma_A)\} A_i \\ \frac{dQ_i}{dt} &= T_i^P(t) \cdot P_i + T_i^I(t) \cdot I_i + T_i^A(t) \cdot A_i - (\alpha_Q + \gamma_Q) \cdot Q_i \\ \frac{dR_i}{dt} &= [1 - T_i^I(t)] \cdot \gamma_I I_i + [1 - T_i^A(t)] \cdot \gamma_A A_i + \gamma_Q Q_i \\ \frac{dD_i}{dt} &= [1 - T_i^I(t)] \cdot \alpha_I I_i + [1 - T_i^A(t)] \cdot \alpha_A A_i + \alpha_Q Q_i\end{aligned}$$

In the sub-model, susceptible people (S) become exposed to infection upon contact with infectious people (in pre-symptomatic (P), severe symptomatic (I), or mild/no symptomatic (A) stages). The force of infection at location i at time t is denoted as Eq. 15.4:

$$\lambda_i(t) = \sum_j \frac{\beta_P \cdot P_j(t) + \beta_I \cdot I_j(t) + \beta_A \cdot A_j(t)}{S_j(t) + E_j(t) + P_j(t) + I_j(t) + A_j(t) + R_j(t)} \cdot \frac{w(j, i)}{N_j} \quad (15.4)$$

where β_P , β_I , and β_A are the transmission rates corresponding to three different infectious compartments, respectively; $w(j, i)$ denotes the daily average number of people traveling from location j to location i ; N_j denotes the population number in location j . The fraction, $w(j, i)/N_j$, reflects how the epidemic in location j affects the epidemic in location i through residents' movements. In Eq. 15.4, j is allowed to equal to i , which reflects local transmissions occurring in location i .

In our study, exposed people (E) are assumed to be latently infected. In other words, they are not infectious until they become (at rate θ) pre-symptomatic (P) after the incubation period. For those non-isolated pre-symptomatic people, some of them then progress (at rate δ) to develop severe symptoms (I) with probability ρ , and some of them may show mild or still no apparent symptoms (A) with probability $1 - \rho$. The people in these two stages, despite they are isolated, will finally recover (at the rate γ_I or γ_A) or die (at the rate α_I or α_A). Similarly, isolated people (Q) also end the process of infection with recovery or death at the rate γ_Q or α_Q , respectively.

Subject to the accessibility of testing reagents, some infected people who need isolation could become non-isolated infectious people and influence the following epidemic propagation. The geographical accessibility sub-model dynamically decides these three floating proportions, including individuals in P, I, and A stages. The estimated spatial accessibility (Eq. 15.2) represents the average amount of resource that one individual in demand at location i at time t can acquire. This accessibility score may be more than one if a location is full of resource supply or no large demand. Thus, we transform this accessibility score into the ration level (RN) through Eq. 15.5:

$$RN_i(t) = \min(1, SA_i(t)) \quad (15.5)$$

This equation ensures that $RN_i(t)$ ranges from 0 to 1, with one means all demand for reagent resources can be satisfied. Then, the proportions that infected people shift from three infectious compartments to the isolation compartment are formulated as Eq. 15.6:

$$T_i^P(t) = v_P \cdot RN_i(t) \cdot u \quad (15.6)$$

$$T_i^I(t) = v_I \cdot RN_i(t) \cdot u$$

$$T_i^A(t) = v_A \cdot RN_i(t) \cdot u$$

where v_P , v_I , and v_A , as stated above, represent the proportion of people who are required to be tested from P, I, and A groups; u is the accuracy of the COVID-19 reagent. The values of $T_i^P(t)$, $T_i^I(t)$, and $T_i^A(t)$ not only differ among distinct places but also dynamically change over time. Thus, Eq. 15.6 integrates spatiotemporal dynamics of accessibility into the epidemic dynamic process. Table 15.1

Table 15.1 Descriptions and values of model parameters

Parameter	Description	Value	Reference
β_P	Transmission rate from P to S	3.7859 day ⁻¹	*
β_I	Transmission rate from I to S	0.1287 day ⁻¹	*
β_A	Transmission rate from A to S	0.1249 day ⁻¹	*
$1/\theta$	Incubation period	3.32 days	*
$1/\delta$	Latent period	0.75 days	*
ρ	Probability to be severe symptomatic (I)	0.25	*
γ_I	Recovery rate of I	0.0698 day ⁻¹	*
γ_A	Recovery rate of A	0.1396 day ⁻¹	*
γ_Q	Recovery rate of Q	0.1047 day ⁻¹	*
α_I	Death rate of I	0.0413 day ⁻¹	*
α_A	Death rate of A	0.0207 day ⁻¹	-
α_Q	Death rate of Q	0.0207 day ⁻¹	-
v_P	Demand proportion of P	0.1	-
v_I	Demand proportion of I	0.5	-
v_A	Demand proportion of A	0.3	-
u	Accuracy of reagent	0.9	-

Note parameter values with * mark refer to Gatto et al. (2020). The other parameter values are assumed by the authors

presents detailed descriptions of parameter settings, and the settings of the COVID-19 epidemic process refer to the study that proposed SEPIA model (Gatto et al. 2020).

15.2.2 Materials

15.2.2.1 Study Area

To demonstrate our Epi-GA model's feasibility, we used the Taipei metropolitan area, one of the major cities in East Asia, as the case study. Our study area covers approximately 1362.57 square kilometers of land and has a population of approximately 6.34 million, 27.5% of Taiwan's total population. In 2015, the Department of Rapid Transit Systems (DORTS) of the Taipei City Government divided this area into 543 traffic analysis zones (TAZs). The average size of a TAZ is about 2.52 square kilometers, and the average population is 12,500. These 543 TAZs are used as spatial units of demand locations in this study. Figure 15.2a displays spatial distributions of population densities. The downtown area with high population density locates at the center of our study area, and most peripheral areas have low population density.

15.2.2.2 Data

The supply facilities are represented by 22 community COVID-19 testing stations established by Taiwan CDC (Fig. 15.2a), and these locations are district hospitals or regional hospitals. In Taiwan, people who once contacted COVID-19 confirmed cases or show significant symptoms, such as fever or dyspnea, can only be tested in these official testing stations. Before the testing process is finished, people have to be temporarily quarantined. If the testing result is negative, the person can leave the station; otherwise, s/he has to be isolated immediately until recovering. The testing process usually takes a few hours, which is shorter than the time unit in our model (day); thus, we ignored this waiting time. We assumed that each testing station provides 500 reagents for COVID-19 testing each day.

The human movement data is from an official survey conducted by DORTS in 2015, the fourth version of the “Taipei Rapid Transit Systems Demand Model” (TRTS-IV). Based on the 543 TAZs, the daily average number of residents moving from any one TAZ to another were surveyed (Fig. 15.2b). According to the official report (Chou, Ko, Lee, Chen, & Lee, 2018), the error of movement amount on most flow links are smaller than 10%, and only a few links have larger errors yet remaining within 20%.

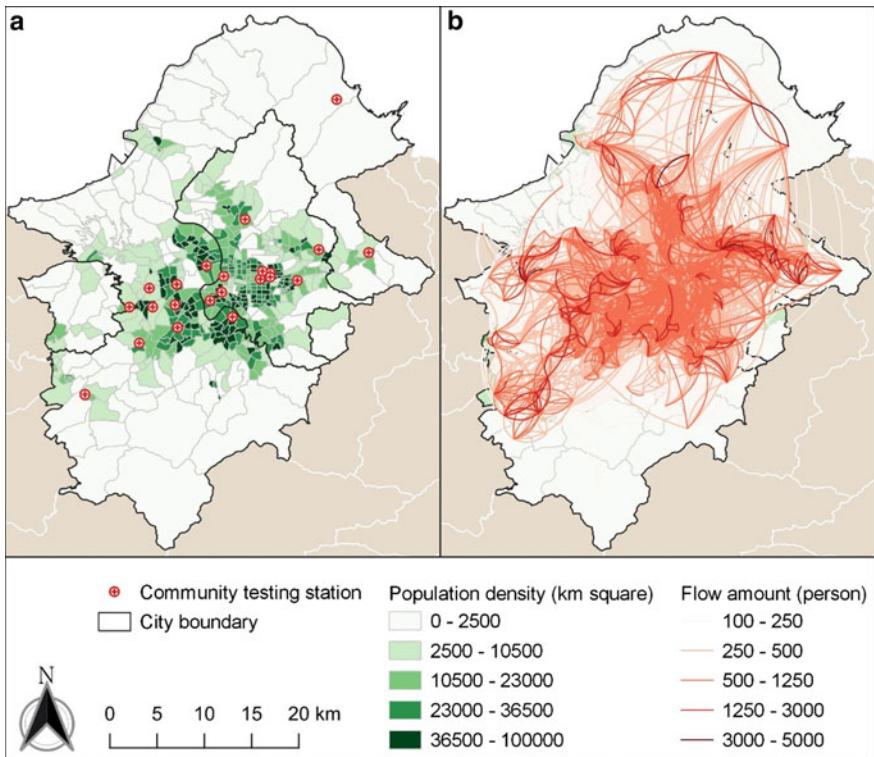


Fig. 15.2 Spatial distributions of **a** population density and locations of community COVID-19 testing stations, and **b** Population daily movements

15.3 Results

We presented the temporal progression of epidemic-induced demand for COVID-19 testing reagents, including individuals in the pre-symptomatic (P), severe symptomatic (I), or mild/no symptomatic (A) stages, estimated from the Epi-RA model (Fig. 15.3a). The temporal pattern shows the exponential growth of the demand is initiated after the 20th day of the first infected person and a significant decrease in the demand after the 26th day. Figure 15.3b, c, and d illustrates the spatial distributions of the demand on the 20th (exponential growth), 26th (epidemic peak), and 32th (significant decrease) day. These figures illustrate how the demand dynamically changes over space. In the initial stage of exponential growth, the demand concentrated on Taipei downtown areas (Fig. 15.3b). As the epidemic peak is reached on the 26th day, not only in downtown areas but also in peripheral urban areas have a high demand for testing reagents (Fig. 15.3c). In the last stage of the epidemic, the demand in downtown areas declines, but the demand in peripheral urban areas remains high.

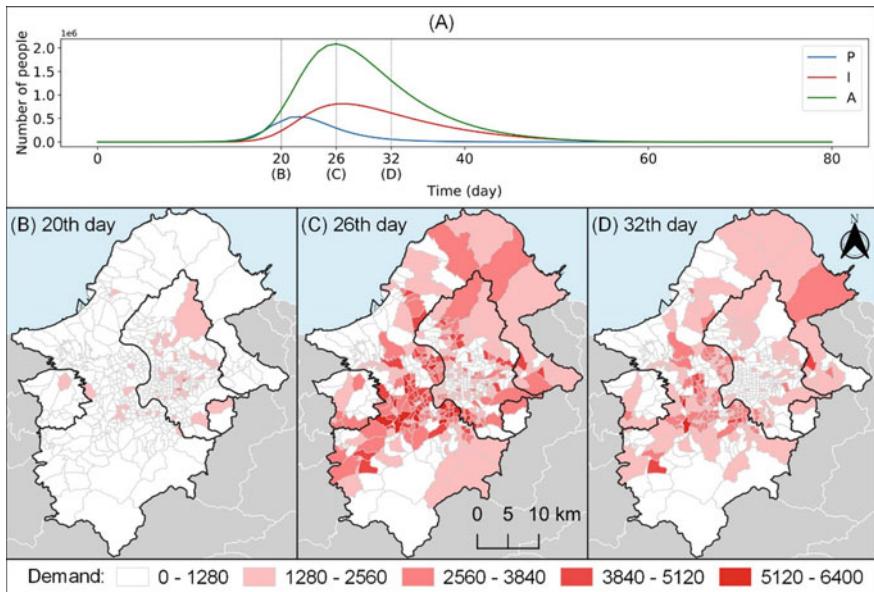


Fig. 15.3 Temporal progression and spatial distributions of epidemic-induced demand for COVID-19 testing reagents, including individuals in the pre-symptomatic (P), severe symptomatic (I), or mild/no symptomatic (A) stages

Figure 15.4 displays spatial distributions of the ration level of testing reagents in three periods of an epidemic. As the epidemic exponentially grows in the initial stage, the ration level in peripheral urban areas drops immediately, but testing reagents in the city centers is sufficient (Fig. 15.4a). As the epidemic reaches the peak, the demand

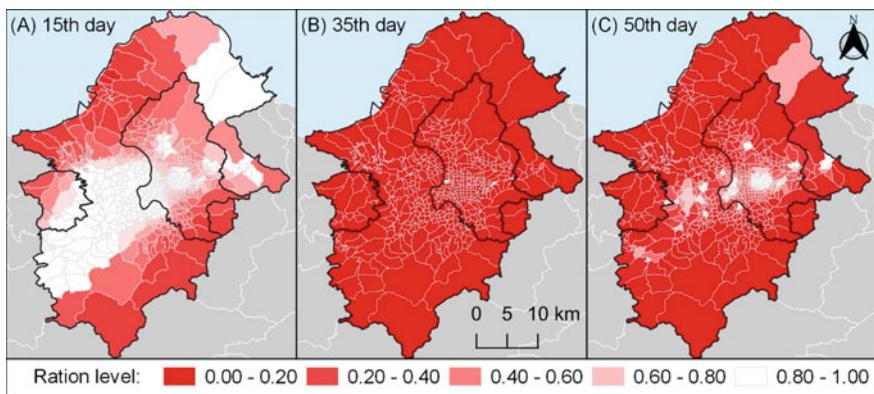


Fig. 15.4 Spatial distributions of the ration level of testing reagents in three periods of an epidemic: **a** the 15th day, **b** 35th day, and **c** 50th day

is more significant than the scheduled volume of testing reagents in all the areas. The volume of reagents is challenging to be provided to all the infectious people at the epidemic stage. The deficit in testing reagents could underestimate cumulative cases and intensify the severity of the following epidemic. In the last stage of the epidemic, the downtown areas start the recovery and return to a high ration level. Most urban peripheral areas are the latest ones recovering from resource deficiency (Fig. 15.4c). Our results capture each TAZ's ration level over time, and the changing rate is also different geographically.

We used two indicators to measure the temporal variations of ration levels of testing reagents in each TAZ (Fig. 15.4): insufficient onset time and the deficit duration. The insufficient onset time is defined as when the TAZ's ration level begins to be lower than 100%. The deficit duration represents the duration of the ration level in a TAZ to recover to 100%. These two indicators are highly negatively correlated (Pearson's correlation coefficient $r = -0.82$), namely that a TAZ with a longer duration of resource deficit usually encounters insufficient resources earlier. Figure 15.5 depicts the spatial distributions of insufficient onset time and the deficit duration across the study region. Our results indicate that regions with the earlier insufficient onset and longer duration of the deficit are mostly located in the outskirt of the

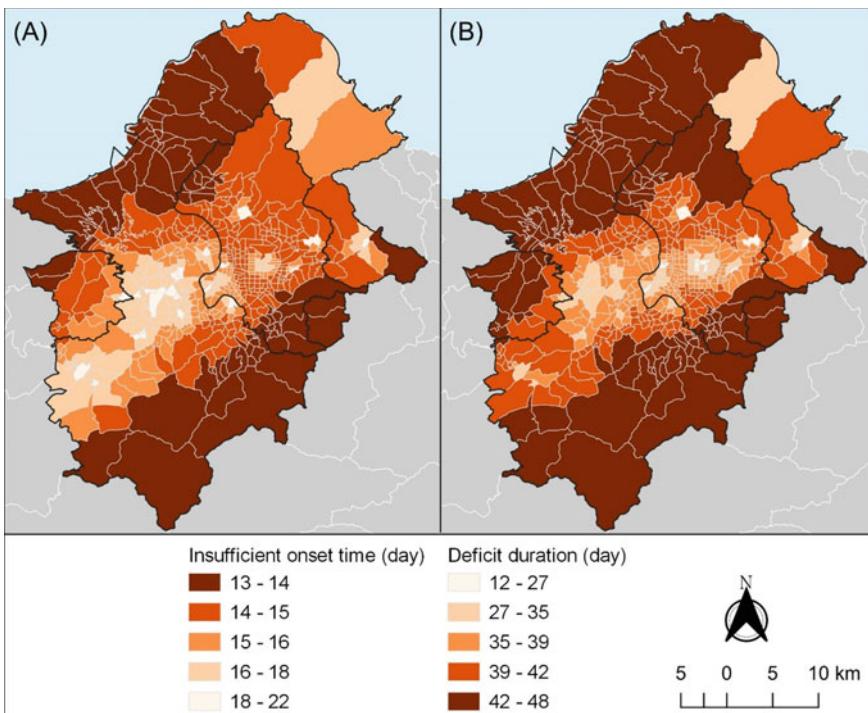


Fig. 15.5 The spatial distribution of (A) insufficient time and (B) insufficient duration across the study region

metropolitan. However, some downtown areas (north-west corner in Fig. 15.5) could also have the potential of resource deficits (earlier onset and longer duration), which could be more vulnerable than peripheral urban areas.

15.4 Discussions and Conclusion

We developed an integrated mathematical model, named Epi-RA, which combines epidemic dynamic process and resource accessibility to assess the availability of COVID-19 testing reagents in space and time. Our model framework captures the dynamic nature of an epidemic. It demonstrates that the resource demand for testing reagents and each TAZ's ration level varies over space and time. The specific regions with the earlier insufficient onset and longer duration of the deficit are identified. These areas are located in the outskirt of the metropolitan and some of the downtown areas. The results could support the health authority to implement the resource allocation schemes for confronting possible pandemic.

Methodologically, our Epi-RA model captured the spatial-temporal dynamics of epidemic-induced demand and proposed the integrated accessibility modeling to deal with time-varying demand. Previous studies on measuring geographic accessibility often used population census as the potential demand for resources, which is regarded as static spatial distribution and assumed unchanged over time (Yun et al., 2020). Population census represents population distribution in the nighttime situation and cannot reflect appropriately spatial behaviors of people seeking resources in the daytime (Fransen, Neutens, De Maeyer, & Deruyter, 2015). Recent studies highlighted the importance of considering people's routine mobility to capture more realistic resource-seeking behaviors in measuring geographical accessibility (Ma et al., 2018; Xia et al., 2019). However, their time-varying demand did not take the frequency of people's contacts into account to not capture the dynamics of disease transmission. Our results from the Epi-RA model represent more reasonable spatial distributions of the time-varying ration level of COVID-19 testing reagents. It could provide critical information for the health authority to decide how effectively allocating community testing reagents in time and space.

Our results identified some specific locations are more vulnerable to infections due to the deficit of community testing reagents (Fig. 15.5). The low density of testing facilities and small community testing reagents in metropolitan areas' outskirts makes it reasonable that exposed individuals in these areas are likely to encounter the shortage of testing reagents. On the other hand, some downtown areas with a high density of testing facilities also can encounter testing resource shortages. Our results from the epidemic dynamic process reveal the areas where people more easily get infected could result from many contacts happening in public places (Leung, Jit, Lau, & Wu, 2017). In other words, the exponential growth of an epidemic causes a surge of infected populations and the need for pathogen testing in a short time. Therefore, some downtown areas may require more testing reagents than we ever

expected. The finding is also consistent with the study on the accessibility of Seoul city's emergency rooms in South Korea (Yun et al., 2020).

The deficit in medical resources could increase the severity of disease transmission. Matching the availability of medical resources should consider the demand and supply over space and time simultaneously. During the COVID-19 epidemic, the health authority must allocate medical resources appropriately. Our integrated modeling framework differentiates well-supplied and under-resourced areas. Geographically, infectious disease is quickly diffused to locations which well connected through spatial interactions. Implementing appropriate resource allocation from well-supplied to under-resourced areas could prevent vulnerable areas from high-damage epidemic centers. Similar suggestions to confront COVID-19 are also discussed in recent studies (Ranney, Griffeth, & Jha, 2020). The under-resourced areas identified in this study could be enhanced by more capacity of medical resources against a large-scale epidemic in advance. For example, locating the appropriate locations for building field hospitals (or makeshift hospitals) are widely used in many countries to house more infected persons (Arango, 2020; Chen et al., 2020; des Déserts, Mathais, Luft, Escarment, & Pasquier, 2020; Wallis, Gust, Porter, Gilchrist, & Amaral, 2020). In summary, these findings from the Epi-RA model provide more comprehensive insights into mobilizing urgent medical resources against large-scale epidemics in time and space.

There are some limitations in this study. First, the age-structured population is not considered in our model framework. The elderly people who are vulnerable groups possess lower accessibility than young people because of low motilities (El Bcheraoui et al., 2018; Horner, Duncan, Wood, Valdez-Torres, & Stansbury, 2015). Thus, distinguishing the demand for medical resources among different age groups could improve the measurement of spatiotemporal accessibility. Previous studies have incorporated the age-structure into either the FCA model (Hashtarkhani et al., 2020) or SEIR meta-population model (Tsai et al., 2011). It could be helpful to extend our model framework. Second, socioeconomic status and transportation mode may influence the individual mobility behaviors to access the medical resource (Kim, Kim, Paul, & Lee, 2020) and the risk of exposure to infection (Cooley et al., 2011; Scargrough, Holt, Hill, & Kafle, 2019). Incorporating high-resolution geo-demographics and contact structures into our integrated model warrants further investigation. Last but not least, the amount of supply is assumed to be fixed in this study. The health authority may dynamically raise the amount according to the situation of epidemic progression (Tanne et al., 2020), thus our results may underestimate the availability of medical resources.

Acknowledgements This research was supported by grants from the Ministry of Science and Technology in Taiwan (MOST 108-2638-H-002-002-MY2) and the National Health Research Institutes (MR-109-CO-14). The authors also acknowledge the financial support provided by the Infectious Diseases Research and Education Center, the Ministry of Health and Welfare (MOHW), and National Taiwan University (NTU). The funders had no role in the study design, data collection and analysis, or manuscript preparation.

References

- Arango, C. (2020). Lessons learned from the coronavirus health crisis in Madrid, Spain: How COVID-19 has changed our lives in the last 2 weeks. *Biological Psychiatry*, 88(7), e33–e34.
- Chen, S., Zhang, Z., Yang, J., Wang, J., Zhai, X., Bärnighausen, T., & Wang, C. (2020). Fangcang shelter hospitals: A novel concept for responding to public health emergencies. *The Lancet*, 395(10232), 1305–1314.
- Chou, W.-R., Ko, C.-H., Lee, S.-W., Chen, Y.-Y., & Lee, T.-Y. (2018). A study of Taipei rapid transit system model version 4S (TRTS-4S). *Journal of Rapid Transit Systems and Technology*, 53, 1–45.
- Cooley, P., Brown, S., Cajka, J., Chasteen, B., Ganapathi, L., Grefenstette, J., et al. (2011). The role of subway travel in an influenza epidemic: A New York City simulation. *Journal of Urban Health*, 88(5), 982–995.
- El Bcheraoui, C., Mokdad, A. H., Dwyer-Lindgren, L., Bertozzi-Villa, A., Stubbs, R. W., Morozoff, C., et al. (2018). Trends and patterns of differences in infectious disease mortality among US counties, 1980–2014. *JAMA*, 319(12), 1248–1260.
- Fransen, K., Neutens, T., De Maeyer, P., & Deruyter, G. (2015). A commuter-based two-step floating catchment area method for measuring spatial accessibility of daycare centers. *Health and Place*, 32, 65–73.
- Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., & Rinaldo, A. (2020). Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, 117(19), 10484–10491.
- Guagliardo, M. F. (2004). Spatial accessibility of primary care: Concepts, methods and challenges. *International Journal of Health Geographics*, 3, 3.
- Hashtarkhani, S., Kiani, B., Bergquist, R., Bagheri, N., VafaeiNejad, R., & Tara, M. (2020). An age-integrated approach to improve measurement of potential spatial accessibility to emergency medical services for urban areas. *The International Journal of Health Planning and Management*, 35(3), 788–798.
- Horner, M. W., Duncan, M. D., Wood, B. S., Valdez-Torres, Y., & Stansbury, C. (2015). Do aging populations have differential accessibility to activities? Analyzing the spatial structure of social, professional, and business opportunities. *Travel Behaviour and Society*, 2(3), 182–191.
- Ji, Y., Ma, Z., Peppelenbosch, M. P., & Pan, Q. (2020). Potential association between COVID-19 mortality and health-care resource availability. *The Lancet Global Health*, 8(4), e480.
- Kim, H., Kim, D., Paul, C., & Lee, C. K. (2020). The spatial allocation of hospitals with negative pressure isolation rooms in Korea: Are we prepared for new outbreaks? *International Journal of Health Policy and Management*, 9(11), 475–483.
- Kwan, M. P. (1998). Space-time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis*, 30(3), 191–216.
- Leung, K., Jit, M., Lau, E. H., & Wu, J. T. (2017). Social contact patterns relevant to the spread of respiratory infectious diseases in Hong Kong. *Scientific Reports*, 7, 7974.
- Lima, A., De Domenico, M., Pejovic, V., & Musolesi, M. (2015). Disease containment strategies based on mobility and information dissemination. *Scientific Reports*, 5, 10650.
- Ma, L., Luo, N., Wan, T., Hu, C., & Peng, M. (2018). An improved healthcare accessibility measure considering the temporal dimension and population demand of different ages. *International Journal of Environmental Research and Public Health*, 15(11), 2421.
- McGrail, M. R. (2012). Spatial accessibility of primary health care utilising the two step floating catchment area method: An assessment of recent improvements. *International Journal of Health Geographics*, 11(1), 50.
- McGrail, M. R., & Humphreys, J. S. (2014). Measuring spatial accessibility to primary health care services: Utilising dynamic catchment sizes. *Applied Geography*, 54, 182–188.
- Ranney, M. L., Griffith, V., & Jha, A. K. (2020). Critical supply shortages—The need for ventilators and personal protective equipment during the Covid-19 pandemic. *New England Journal of Medicine*, 382, e41.

- Remuzzi, A., & Remuzzi, G. (2020). COVID-19 and Italy: What next? *The Lancet*, 395(10231), 11–17.
- Scargrough, A. W., Holt, M. M., Hill, J., & Kafle, R. C. (2019). Is there a relationship between income and infectious disease: Evidence from Cameron County. *International Journal of Community Well-Being*, 2, 3–13.
- Tanne, J. H., Hayasaki, E., Zastrow, M., Pulla, P., Smith, P., & Rada, A. G. (2020). COVID-19: How doctors and healthcare systems are tackling coronavirus worldwide. *BMJ*, 368, m1090.
- Tsai, Y.-S., Huang, C.-Y., Wen, T.-H., Sun, C.-T., & Yen, M.-Y. (2011). Integrating epidemic dynamics with daily commuting networks: Building a multilayer framework to assess influenza A (H1N1) intervention policies. *SIMULATION*, 87(5), 385–405.
- Wallis, N., Gust, C., Porter, E., Gilchrist, N., & Amaral, A. (2020). Implementation of field hospital pharmacy services during the COVID-19 pandemic. *American Journal of Health-System Pharmacy*, 77(19), 1547–1551.
- Wang, F. (2012). Measurement, optimization, and impact of health care accessibility: A methodological review. *Annals of the Association of American Geographers*, 102(5), 1104–1112.
- Wang, L., & Wu, J. T. (2018). Characterizing the dynamics underlying global spread of epidemics. *Nature Communications*, 9, 218.
- Weissman, G. E., Crane-Droesch, A., Chivers, C., Luong, T., Hanish, A., Levy, M. Z., et al. (2020). Locally informed simulation to predict hospital capacity needs during the COVID-19 pandemic. *Annals of Internal Medicine*, 173(1), 21–28.
- Xia, T., Song, X., Zhang, H., Song, X., Kanasegi, H., & Shibasaki, R. (2019). Measuring spatio-temporal accessibility to emergency medical services through big GPS data. *Health and Place*, 56, 53–62.
- Yun, S. B., Kim, S., Ju, S., Noh, J., Kim, C., Wong, M. S., & Heo, J. (2020). Analysis of accessibility to emergency rooms by dynamic population from mobile phone data: Geography of social inequity in South Korea. *PLoS ONE*, 15(4), e0231079.

Chapter 16

Health Resilience Among European Countries in the Face of Pandemic: Reflections on European Countries' Preparedness for COVID-19



Yijing Li

16.1 Introduction

The global coronavirus pandemic spread throughout 2020, with countries in Europe experiencing tremendous socioeconomic impacts, including a cliff-like drop in citizens' mobility, overcrowding of hospitals in metropolitan cities, an economic recession, massive furloughs and an increase in unemployment, as well as a significant transition to more remote working. The European Commission called for collective efforts among EU partner countries to tackle the coronavirus pandemic with a robust and targeted EU response plan (European Commission 2020a), intended to not only address the immediate health crisis and the resulting humanitarian needs, but a long-term goal to strengthen the partner countries' resilience against such health crises. To do so would require a comprehensive picture of the status quo of national health resilience (Regional Office for Europe WHO 2017) capability.

The goal of this chapter is to capture the nature of the crisis resulting from the current coronavirus pandemic among selected European countries and further evaluate their individual resilience capability against such a crisis using an adapted place-based model. This model takes advantage of cutting-edge techniques in machine learning and spatial data analysis to fulfil the following objectives: (1) organize evidence of the coronavirus pandemic's spread and detrimental impacts among 35 selected European countries, (2) evaluate their resilience capabilities against the pandemic by collocating indicators defined in the model, (3) estimate the optimal geographical unit among case studies at which to enhance the resilience capability against health crisis and finally, (4) provide information not only on how to build-up domestic health resilience capacity but also to integrate the international community to jointly fight against health crises.

Y. Li (✉)

CUSP London, King's College London, London, UK

e-mail: yijing.li@kcl.ac.uk

This analytical work tried to consider resilience capability from a regional¹ perspective among nations, encompassing indicators for coronavirus health risk, national preparedness and domestic vulnerability to further cluster European countries into groups ranked by resilience. This ranking was performed with the assumption that countries in the same cluster may deploy similar strategies and policies on resilience improvement against health crisis. However, given the place-based feature of the clustering model, it is appropriate to consider geographical scale effects when evaluating the efficiency of countries' strategies. Countries in the same prominent cluster group will be selected for a case study, with the aim of identifying the optimal geographical scale (either within-country regional,² provincial or city) at which health resilience capacity plays the role well. Finally, succinct suggestions for policymaking are made based on the results.

16.2 Theoretical Background

The term *resilience* was first proposed by Holling (1973) as a “measure of the persistence of systems and their ability to absorb change and disturbance and still maintain the same relationships between populations or state variables”. There is no universal definition of resilience, in this study, resilience is defined as a target region's (country, province or city) capacity to absorb disturbance from the pandemic and return to its normal state before the disturbance (Klein et al. 2003). Theoretically, levels of resilience against a hazard vary with the level of preparedness and vulnerability in various societies, reflecting each society's capability to draw upon internal resources and competencies (Paton and Johnston 2001) to manage the challenges encountered. In other words, the conceptual model for national resilience against the COVID-19 pandemic, considered as a health hazard, should be underpinned by a set of indicators for hazard risks, and the inherent vulnerability and preparedness of the object studied.

A hazard event's risks are normally measured in terms of exposure to the event. For the COVID-19 pandemic, risk was measured by epidemiological indicators—specifically, confirmed cases and the death rate, according to European Centre for Disease Prevention and Control (<https://www.ecdc.europa.eu/en/covid-19-pandemic>).

The studied object's inherent vulnerability to hazard can be measured by demographic indicators for vulnerable groups (Lindell and Whitney 2000), which serve as variables for environmental limitations, marginalized political status, and limited social network access (Omer and Alon 1994; Schwarzer 1994). During the COVID-19 pandemic, the demographics most affected may be deprived groups, measured by the percentage of domestic population living in poverty, inequality in terms of the human development index and number of international migrants or refugees

¹“Regional” here refers to multiple-country regions.

²“Within-country regional” here refers to regions within a specific country, which is bigger than provincial and smaller than national.

(measured by remittance inflows). In addition, the hard-hit tourism industry can be considered to have suffered a significant global impact from COVID-19; hence, inbound tourism expenditure serves as an ideal indicator for national economic vulnerability (Davradakis et al. 2020) to COVID-19.

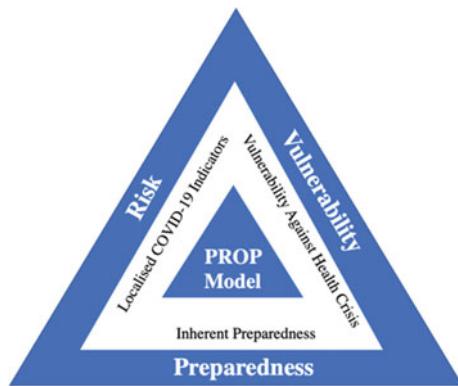
Preparedness has been referred to as self-efficacy and competence in adjusting to hazards, and prepared action plans reflect invested perseverance and efforts (Paton and Johnston 2001). Social connectivity and health environment variables were selected in this study as indicators for preparedness at the national level: the human development index (HDI)³ to measure the life health status, number of hospital beds available, current health expenditure in terms of GDP and the number of mobile phone and fixed broadband subscriptions.

To better compare and interpret the similarities and disparities of European countries' resilience to COVID-19, this study has considered Cutter's (1996) hazards-of-place model of vulnerability for its embrace of using geographic elements to evaluate regional risk vulnerability. Cutter used GIS to apply the model to some case studies in 2000 and further proposed Social Vulnerability Index (SOVI) in 2003, which emphasizes the importance of geographical units for analyses of regional vulnerability, as well as their essential role for emergency relief, preparedness, and prevention strategies. Numerous vulnerability assessment models and techniques—both theoretical and practical—have been developed (Adger 2006; Fussel 2007; McLaughlin and Dietz 2008; Polsky et al. 2007). All emphasize the importance of place-based studies, especially to understand how cross-scalar dynamics influence the vulnerability of a place, and to support the building local resilience capacity (Cutter et al. 2008).

However, from the hazards research perspective, pandemic processes and impacts can be more localized, rendering consideration of the spatial unit of analysis whilst aggregating local information into broader within-country regional comparison essential. For example, there is consensus about some characteristics with potential impacts on pandemic resilience, like the spatial distribution of population density and healthcare services/facilities. Building upon and adapting the Disaster Resilience of Place (DROP) model (Cutter et al. 2008), a Pandemic Resilience of Place (PROP) Model had been proposed to recognize the limitations of applying universal resilience models over varied geographical scales, and further to suggest the optimal spatial/administrative scale for executing pandemic-related policies. The assumption for the PROP model is that policies intended to address the rapidly evolving COVID-19 pandemic should be flexible enough to consider the influence from varied spatial units when evaluating the social healthcare services resilience of places. Hence, the PROP index includes not only inherent conditions but exogenous factors and geographic scales to assess a policy's efficiency.

³Human Development Index (HDI) measuring average achievement in basic dimensions of human development including long and healthy life (technical notes, http://hdr.undp.org/sites/default/files/hdr2020_technical_notes.pdf).

Fig. 16.1 Pandemic resilience of place (PROP) model



16.3 Data and Methodology

Adapted from the DROP model (Cutter et al. 2008), locational resilience capacity against natural hazards could be interpreted as a compounded “triangular” effect from exposure to hazards (normally measured in terms of an event’s risk), the system’s inherent vulnerability and the corresponding preparedness. For the situation of the current COVID-19 pandemic, the localized impact could also be represented by a Pandemic Resilience of Place (PROP) model (Fig. 16.1), with the aim to compare relative resilience against a pandemic event like COVID-19.

Indicators were derived from various sources (Table 16.1) for the 35 target countries. The method applied to evaluate their comparative resilience capacity is the most frequently used clustering technique in machine learning, naïve K-means clustering (Pelleg and Moore 1999), to categorize the target countries into seven groups using the indicators listed in Table 16.1, upon run Python scripts using K-means algorithm of *scikit-learn* package.

Table 16.1 Indicators for PROP model at the national level

Dimension	Indicator	Data sources
Risk	Average number of daily confirmed COVID-19 cases	The Center for Systems Science and Engineering (CSSE) at JHU https://github.com/CSSEGISandData/COVID-19 (24th January 2020–28th June 2020)
	Average daily death rate due to COVID-19	The Center for Systems Science and Engineering (CSSE) at JHU https://github.com/CSSEGISandData/COVID-19 (24th January 2020–28th June 2020)

(continued)

Table 16.1 (continued)

Dimension	Indicator	Data sources
Vulnerability	Proportion of people in poverty in 2018 ^a	UNEP human development Reports (2019) on global preparedness and vulnerability dashboards at https://hdr.undp.org/en/content/global-preparedness-and-vulnerability-dashboards
	Immediate economic inflows remittance in 2018 ^b	
	Inbound tourism expenditure in 2018 ^c	
	Percentage of inequality in HDI in 2018 ^d	
Preparedness	Human development index in 2018	
	Number of physicians in 2018 ^e	
	Number of nurses and midwives in 2018 ^f	
	Number of hospital beds in 2018 ^g	
	Healthcare expenditure per GDP (%) in 2018 ^h	
	Number of fixed broadband subscriptions in 2018 ⁱ	
	Number of mobile phone subscriptions in 2018 ^j	

^aPercentage of the population living below the international poverty line of \$1.90 (in purchasing power parity terms [PPP]) a day (World Bank, 2020; World Development Indicators database. Washington, DC. <https://data.worldbank.org>. Accessed 02 April 2020)

^bRemittances inflows for earnings and material resources transferred by international migrants or refugees to recipients in their country of origin or countries in which the migrant formerly resided (World Development Indicators database. Washington, DC. <https://data.worldbank.org>. Accessed 15 July 2019)

^cInbound tourism expenditure by international visitors, including payments to national carriers for international transport (World Tourism Organization, 2020; Compendium of Tourism Statistics dataset, Madrid, https://statistics.unwto.org/method_notes_tourism_stat_database_2019ed. Accessed 02 April 2020)

^dPercentage difference between the inequality human development index value and the human development index value

^eAccessibility to physicians measured by the number of medical doctors (physicians) per 10,000 population, both generalists and specialists

^fAccessibility to nurses and midwives, measured by the number of professional nurses, professional midwives, auxiliary nurses, auxiliary midwives, enrolled nurses, enrolled midwives and other associated personnel, such as dental nurses and primary care nurses per 10,000 people (World Health Organization, 2019; Global Health Observatory. www.who.int/gho/. Accessed 15 July 2019)

^gAvailability of number of hospital beds per 10,000 people

^hExpenditure on healthcare goods and services per GDP (World Bank, 2019a; World Development Indicators database. Washington, DC. <https://data.worldbank.org>. Accessed 15 July 2019)

ⁱFixed broadband subscriptions per 100 people to high-speed access to the public Internet, at downstream speeds equal to, or greater than, 256 kbit/s (International Telecommunication Union, 2019)

^jMobile phone subscriptions per 100 people

Countries in the cluster identified as “prominent” were selected for further investigation at a finer geographical scale (the region, province, and city levels), with the aim to select an optimum geographical scale for future improvements to resilience capacity. The datasets used for these case studies are location data on health sites scraped from the Global Health Sites Mapping project by OpenStreetMap at <https://healthsites.io/>; the proxy for human density and activities from cell towers locational data at <https://www.opencellid.org>; and the geographical boundary layers at within-country regional, provincial and city scales from <https://gadm.org/>.

16.4 Analysis and Results

16.4.1 Clustering of Resilience to COVID-19 Pandemic

16.4.1.1 COVID-19 Risk in 35 European Countries

The time-series (24 January–28 June 2020) trend for accumulated confirmed COVID-19 cases and deaths, in each of the targeted 35 European countries is depicted in Fig. 16.2. It could be observed from Fig. 16.2 that, the confirmed COVID-19 cases and death cases among 35 countries vary by volume significantly. For example, the six countries in the first row had fewer than 4,000 confirmed cases and fewer than 300 deaths, whilst the seven countries in the last row had up to 300,000 confirmed cases with up to 45,000 deaths.

These 35 countries were divided into three categories by their time-series trend patterns:

- (1) Rates of confirmed COVID-19 cases and deaths were similar. Countries falling into this category may exhibit relatively good but already tight healthcare services; it may also raise the attention to policymakers for simultaneous actions, in reducing the spread rate of coronavirus pandemic to avoid any sooner spikes of confirmed cases. The listed countries may have Croatia, Estonia, Hungary, Austria, Moldova, Norway, the Czech, Poland, Belgium, the Netherlands, Italy, Spain, France and the United Kingdom.
- (2) Rates of death increased more quickly than confirmed cases. Countries in this category (Bosnia and Herzegovina, Bulgaria, Greece, Finland, Denmark, Switzerland, Ireland and Romania) may have been experiencing a lack of sufficient healthcare resilience to meet the emerging requirements from an increased number of confirmed COVID-19 cases; and required further policies to improve hospital facilities if fatality rates were to be reduced.
- (3) Rate of confirmed cases increased faster than deaths, which could be due to the lower level of confirmed cases in certain countries (e.g. Albania, Montenegro, Slovakia, Slovenia, Iceland and Latvia) but also may indicate a healthcare system that performs better in treating confirmed COVID-19 cases, enabling a

(The blue curve against the left y axis represents confirmed cases; the red curve against the right y axis recorded deaths)

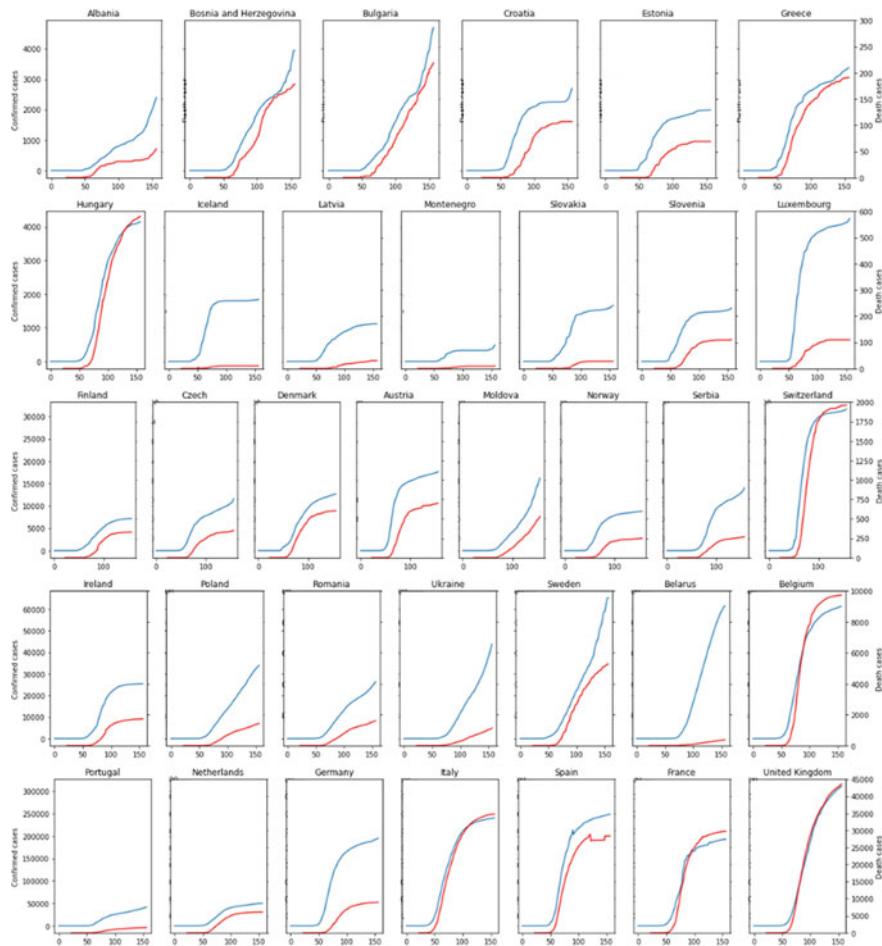


Fig. 16.2 Confirmed cases and deaths from COVID-19 in 35 target countries between 24 January and 28 June 2020 (The blue curve against the left y-axis represents confirmed cases; the red curve against the right y-axis recorded deaths)

higher recovery rate for those admitted to hospitals (e.g. Luxembourg, Serbia, Bahrain, Ukraine, Sweden, Belarus, Portugal and Germany).

The ratio between confirmed COVID-19 cases and deaths can be considered an indicator of the risk the crisis imposed on target countries. Thus, based on the PROP model for resilience evaluation, the next step is to evaluate the other two dimensions on target countries' inherent preparedness and vulnerability to the pandemic.

16.4.1.2 Evaluation of Preparedness

As listed in Table 16.1, the seven indicators utilized to measure national preparedness for pandemics are the human development index (HDI), healthcare infrastructure (i.e. number of physicians, nurses, and hospital beds and expenditure on health in terms of GDP) and internet connection infrastructure (i.e. mobile phone and fixed broadband subscriptions). Four sample indicators have been presented as choropleth maps in Fig. 16.3; a blue–yellow–red colormap is used to signify deciles, with bluer countries better prepared and redder countries more unprepared.

Figure 16.3 suggested that preparedness weakens from North to South, and from West to East, with the south-eastern region being recognized as the especially unprepared areas. Countries with higher HDI values are clustered in the north-west. These countries had relatively consistent spatial distributions of Internet coverage, which was expected to facilitate citizens' adaption to the pandemic; however, their preparedness in healthcare infrastructures varied significantly regarding the spatial distribution patterns. For example, Germany and the Netherlands were well prepared for all indicators, whilst Norway and Sweden were less prepared to provide enough hospital beds for the COVID-19 pandemic. The United Kingdom was the least prepared of this group in terms of access to physicians, enough hospital beds and rate of mobile phone subscriptions.

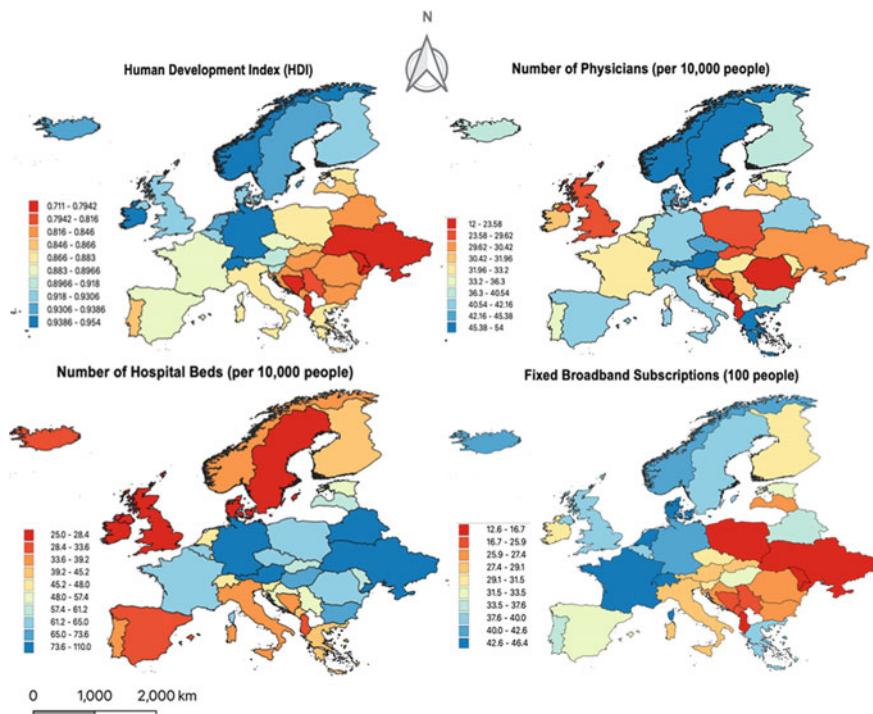


Fig. 16.3 Comparison of preparedness among 35 European countries

16.4.1.3 Vulnerability

Those four indicators chosen to evaluate the national vulnerability against health crisis are the proportion of people in poverty, the immediate economic inflows remittance, the inbound tourism expenditure and the percentage of inequality in HDI. The choropleth maps in Fig. 16.4 also use a blue–yellow–red colormap to signify deciles for vulnerability, with bluer countries less vulnerable, and hence with higher pandemic resilience, whilst redder countries are more vulnerable. The rationale for the selection of indicators is that more economically deprived individuals are assumed to be lack of sufficient access to necessities during lockdown and thus have higher vulnerability due to difficulty in maintaining social distancing due to their work, deprived housing conditions and the attendant negative mental health impacts (DeLuca et al. 2020).

In general, the northern parts of the region (including Iceland, Norway, Sweden, Finland and Denmark) were less vulnerable than those in the southern portion, which negatively affect the resilience capability. Countries in the Western portion of the region, including the United Kingdom and France, were comparatively resilient in terms of their economy and tourism industry, but vulnerable in terms of domestic

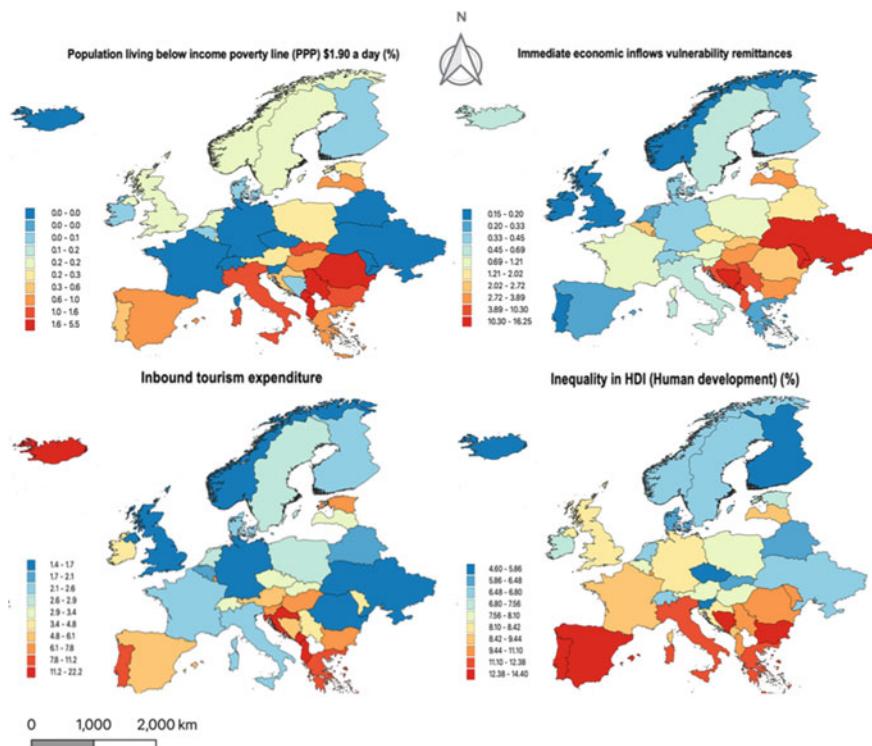


Fig. 16.4 Comparison of vulnerability among 35 European countries

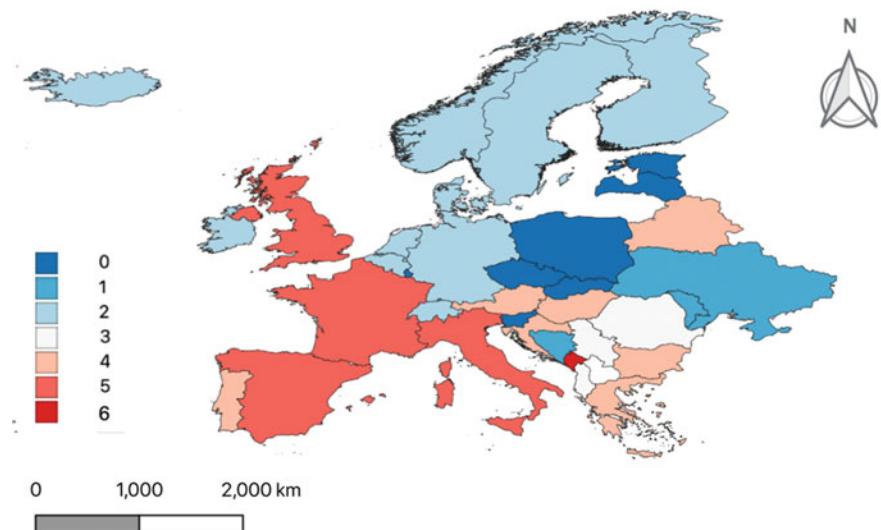


Fig. 16.5 Clustering of resilience among 35 European countries

inequality in the HDI; Spain and Italy had higher vulnerability than other western countries.

16.4.1.4 Clusters of Resilience Among 35 European Countries

The final resilience clustering results from applying the K-means algorithm of machine learning package *scikit-learn* to the standardized indicators for the three dimensions in the PROP model are shown in Fig. 16.5.

Four countries in cluster #5—the United Kingdom, France, Italy and Spain—are rated as “outstanding” in terms of resilience. Consideration of the results reflected in each PROP dimension for COVID-19 pandemic risk, the preparedness indicators and the vulnerability indicators, led these four countries to be selected for case studies to identify the optimum geographical scale for executing strategies and policies to improve resilience capacity.

16.4.2 Effect of Geographic Scale

To better prepare for future responses to health crisis like COVID-19 pandemic, it is of paramount importance to improve local resilience capacity, which can be supported by evaluation of the PROP model’s dimensions: risk exposure, preparedness and vulnerability. This suggested the need to determine the optimum geographical scale for the improvement of resilience capabilities.

Table 16.2 Statistics on data for case study countries

Country	Cell towers	Healthcare sites (excluding dentists)
United Kingdom	1,585,236	17,458
Italy	1,238,529	17,323
France	1,828,870	25,388
Spain	948,782	22,298

The four “prominent” countries from the previous discussion—the United Kingdom, France, Spain and Italy—were selected as case studies to facilitate the evaluation of scale effect. The local resilience index against health crises was reflected by per-capita access to healthcare services. However, the population density data may vary between geographical units in each case study and also generate massive work for data collection. To maintain consistency between data sources for population density, it was proxied by the number of cell towers, with that value considered to “provide quality assurance (QA) of Local Authority District (LAD)-level residential population counts” (Office for National Statistics 2017).

Upon calculating the healthcare services per capita by dividing the number of health sites by the number of cell towers (as listed in Table 16.2) in each target area, the local resilience spatial dependence could be further visualized by LISA map (Anselin 1995), which is widely utilized in measuring spatial autocorrelation at varied geographical scales (within-country regional, provincial and city; Fig. 16.6). In this analysis, adjacent neighbours were defined with those at first order by Queen’s contiguity (Anselin 1995) to create the spatial weight matrix.

The varied definition of geographically administrative levels among case study countries (i.e. provinces are called departments in France, and counties in the United Kingdom) was adjusted based on the LISA maps in Fig. 16.6, following a consistent scale-down definition of region, province and city. The coloured legends indicate the varied types of spatial dependence; besides the grey areas without any significant spatial dependence, high-valued areas surrounded by high-valued neighbours is denoted as HH clusters (in red) and low-valued areas surrounded by low-valued neighbours as LL clusters (dark blue). In contrast, high-valued area surrounded by low-valued neighbours are denoted as HL outliers (orange) and low-valued areas surrounded by high-valued neighbours as LH outliers (light blue). The expected result is a LISA map with multiple types of spatial dependence relations (i.e. a map with at least HH, LL and insignificant areas, rather than one with a universal spatial distribution pattern, which might be taken as too idealized to be realistic).

As shown in Fig. 16.6, spatial dependence patterns were more observable at the provincial level for all four countries, and significant spatial patterns were evident at the finer city level only in France and the United Kingdom. At the provincial level:

- (1) In the United Kingdom, LL clusters are evident around the greater London area and in counties in the northwest (e.g. Birmingham and Manchester) and northern Yorkshire. The spatial pattern indicated their lower resilience against health crises not only for themselves but for neighbouring areas, which had

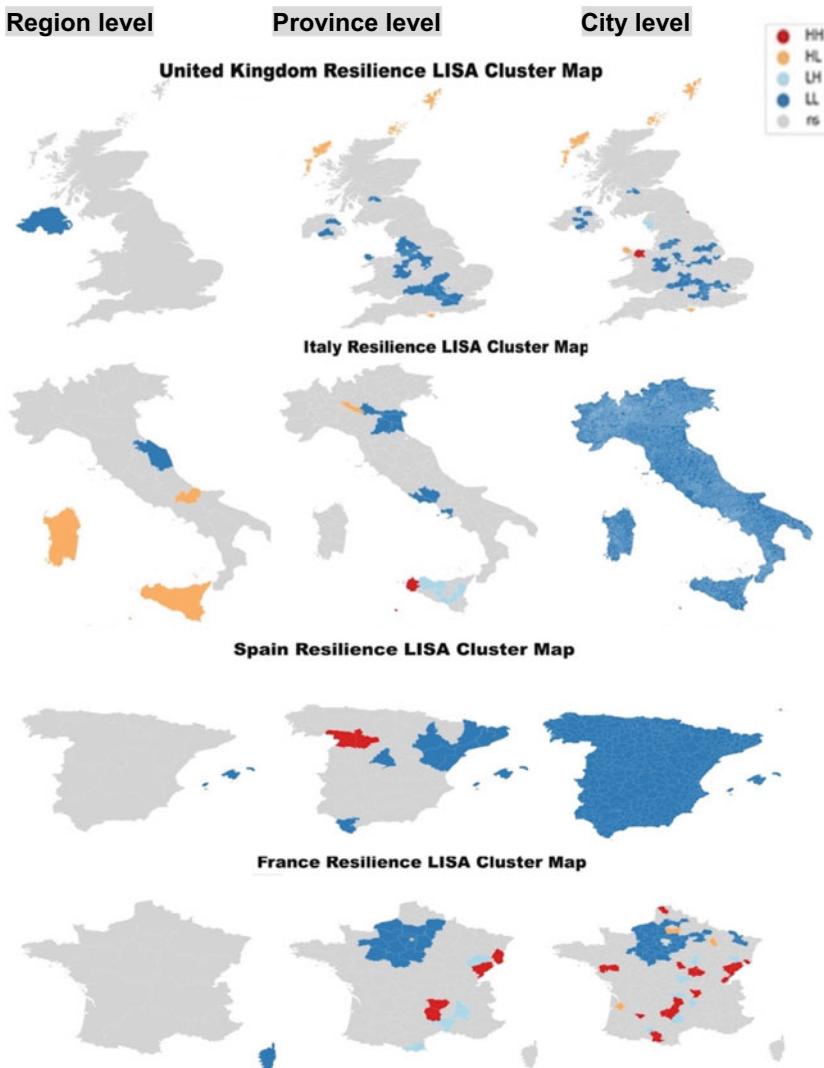


Fig. 16.6 Case studies: comparison of health resilience by scale

a compounding effect on lowering resilience, leading to an urgent need for local-level policies and measures.

- (2) In Italy, the neighbouring provinces around Bologna and Rome were identified as cold spot clusters (LL), with low health resilience and thus presumably higher vulnerability to health crises. In contrast, Trapani province in the far south turned out to be a hot spot cluster (HH), with high resilience to health crises.

- (3) In Spain, areas around Barcelona in the north, Madrid in the central region and Cadia in the south exhibited LL cluster features, and hence require localized measures to build up health resilience capability; Zamora and Valladolid in the northwest of Spain presented better health resilience performance.
- (4) In France, the clusters were significant in northern France, with LL spatial patterns of health resilience capability, requiring urgent improvements, whilst some provinces in the east, like Haut Rhin and Dovbs, and Lozere in the south, have better (HH) resilience against the crisis, which may suggest further promotion of their localized practises and policies.

16.5 Discussion and Conclusions

A recent report from the European Commission (2020b, p. 8; 2020c) called for joint strategic measures to contain the health crisis, with an increase of the capacity and the resilience of healthcare system one major objective. The results of this analysis are consistent with the resilience building policies in that report, in emphasizing the decisive roles of sufficient hospital capacity, well-trained health care staff (doctors, nurses, etc.), as well as well-protected financing capacity for the healthcare system. In addition, this research further steers building of resilience towards the place-based resilience evaluation (PROP) model, by suggesting that resilience capability in Europe reflects three majors: exposure to external crisis risk; inherent systematic short-term response capacity, or preparedness; and inherent long-term bounce-back capacity, or vulnerability. It was also suggested that all 35 selected European countries develop prompt, responsive policies and reactions to such pandemic based on their respective preparedness. The interrelationships among the three dimensions in the PROP model were reflected succinctly in the result that the varying levels of risks imposed by the COVID-19 pandemic may be influenced by countries' inherent demographic features, such as population density and age pyramid; indicators measuring preparedness (e.g. access to physicians and number of hospital beds) had strong influences on local fatality rates; and vulnerability indicators selected at the national level tend to measure more on the economic recovery capability and policies in the long term. Thus, in general, policies and measures to strengthen national health resilience should embrace both long- and short-term capability building to save lives both during the COVID-19 pandemic and in the future.

The study also could be utilized as the evidence for needed actions to be taken in the immediate future, both at the national and provincial scales to increase the resilience in that:

- (1) Regional collaboration against the coronavirus pandemic is imperative. Resilience clustering of 35 European countries on basis of PROP model revealed dispersed distributions of countries falling into the same cluster, which testified to the interdependency among partner countries in the region. This highlights their shared responsibility to take action jointly and urgently—for

- instance, to set up a regional network and platform for rapid epidemiological surveillance (European Commission 2020b).
- (2) Strengthening a national health care system is demanding. Countries, especially those with healthcare systems with higher vulnerability and risk levels, should take immediate actions to strengthen their preparedness and response capacities. These may include setting up collaborations with other highly resilient neighbouring countries to, for example, supply critical medical equipment, facilities and supplies, train public health workers, and provide enough protective personal equipment for those workers.
 - (3) Provincial-scale resilience capability building will offer optimum efficiency domestically. Evaluation of the spatial scale effect among the four case-study countries (the United Kingdom, Spain, Italy, and France) suggested it is best to execute both pandemic-mitigation measures and future resilience improvement policies at the provincial scale for optimum efficiency. The results also emphasized the importance of locally adapted resilience capability building against health crises, on basis of national policies and measures.
 - (4) It is of paramount importance to maintain communication and awareness efforts regarding protective measures and hygiene, regardless of geographical scale. Such efforts will facilitate a speedy and thorough response in future crises, and can help overcome the misunderstandings induced by fragmented information, which will otherwise continue to endanger lives and make the region collectively vulnerable.

There are some limitations of this analysis. For example, the limited indicators selected for PROP model, the relatively short research period (focusing on the first half of 2020), and limited number of case study countries due to data accessibility may result in biased results and conclusions. It is also acknowledged that the mismatch of data collection periods (i.e. indicators for preparedness and vulnerability are all from 2018, whilst the coronavirus indicators used data collected in the first half of 2020) may have caused distortion in the results. However, considering the comparative nature of this study, and especially its use of a clustering algorithm and spatial dependence-based method, its results are appropriate. Despite this study's limitations, its results shed light on the topic and open the potential for future replication or explorative work along the same lines.

References

- Adger, W. N. (2006). Vulnerability. *Global Environmental Change*, 16, 268–281.
- Anselin, L. (1995). Local indicators of spatial association-LISA. *Geographical Analysis*, 27(2), 93–115.
- Cutter, S. L., Barnes, L., Berry, M., Burton, C., Evans, E., Tate, E., & Webb, J. (2008). A place-based model for understanding community resilience to natural disasters. *Global Environmental Change*, 18, 598–606.
- Cutter, S. L. (1996). Societal vulnerability to environmental Haz-ards. *Progress in Human Geography* 20(4), 529–539. <https://doi.org/10.1177/030913259602000407>.

- Davradakis, E., Santos, R., Zwart, S., & Marchitto, B. (2020). The EIB COVID-19 economic vulnerability index—An analysis of countries outside the European Union. *European Investment Bank*. Retrieved November 8, 2020, from https://www.eib.org/attachments/thematic/the_eib_covid-19_economic_vulnerability_index_en.pdf.
- DeLuca, S., Kalish, M., & Papageorge, N. W. (2020). *The unequal cost of social distancing*. Retrieved November 25, 2020, from <https://coronavirus.jhu.edu/from-our-experts/the-unequal-cost-of-social-distancing>.
- European Commission—Press Release. (2020a). *Coronavirus: EU global response to fight the pandemic*. Retrieved August 6, 2020, from https://ec.europa.eu/commission/presscorner/detail/en/ip_20_604.
- European Commissions. (2020b). *EU global response to COVID-19*. Retrieved November 10, 2020, from https://ec.europa.eu/international-partnerships/topics/eu-global-response-covid-19_en.
- European Commission. (2020c). Joint European Roadmap towards lifting COVID-19 containment measures. https://ec.europa.eu/info/sites/info/files/communication_a_european_roadmap_to_lifting_coronavirusContainment_measures_0.pdf. Accessed 10 Oct 2020.
- Fussel, H. M. (2007). Vulnerability: A generally applicable conceptual framework for climate change research. *Global Environmental Change*, 17(2), 155–167.
- Holling, C. S. (1973). Resilience and stability of ecological systems. *Annual Review of Ecology and Systematics*, 4, 1–23.
- Klein, R. J. T., Nicholls, R. J., & Thomalla, F. (2003). Resilience to natural hazards: How useful is this concept? *Environmental Hazards*, 5(1–2), 35–45.
- Lindell, M. K., Whitney D. J. (2000). Correlates of household seismic hazard adjustment adoption. *Risk Analysis* 20(1), 13–25. <https://doi.org/10.1111/0272-4332.00002>
- McLaughlin, P., & Dietz, T. A. (2008). Structure, agency and environment: Toward an integrated perspective on vulnerability. *Global Environmental Change*, 18, 99–111.
- Office of National Statistics. (2017). *ONS methodology working paper series number 13—Comparing the density of mobile phone cell towers with population estimates*. Retrieved October 15, 2020, from <https://www.ons.gov.uk/methodology/methodologicalpublications/generalmethodology/onsworkingpaperseries/onsmethodologyworkingpaperseriesnumber13comparingthedenistyofmobilephonencelltowerswithpopulationestimates#executive-summary>.
- Omer, H., & Alon, N. (1994). The continuity principle: A unified approach to disaster and trauma. *American Journal of Community Psychology*, 22, 273–287.
- Paton, D., & Johnston, D. (2001). Disasters and communities: Vulnerability, resilience and preparedness. *Disaster Prevention and Management*, 10(4), 270–277.
- Pelleg, D., & Moore, A. (1999). Accelerating exact k-means algorithms with geometric reasoning. In U. Fayyad, S. Chaudhuri, & D. Madigan (Eds.), *KDD'99: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Diego, California, United States, August 1999 (pp. 277–281). New York: ACM Press.
- Polksy, C., Neff, R., & Yarnal, B. (2007). Building comparable global change vulnerability assessments: The vulnerability scoping diagram. *Global Environmental Change*, 17, 472–485.
- Regional Office for Europe, World Health Organisation. (2017). *Strengthening resilience: A priority shared by Health 2020 and the sustainable development goals*. Retrieved July 15, 2020, from <https://www.euro.who.int/en/countries/monaco/publications/strengthening-resilience-a-priority-shared-by-health-2020-and-the-sustainable-development-goals-2017>.
- Schwarzer, R. (1994). Optimism, Vulnerability, and self-beliefs as health-related cognitions: A systematic overview. *Psychology & Health*, 9(3), 161–180.
- UNDP. (2019). *Human development report 2019. Beyond income, beyond averages, beyond today: Inequalities in human development in the 21st century*. New York. Retrieved July 7, 2020, from <https://hdr.undp.org/en/content/human-development-report-2019>.

Chapter 17

Improving Public Transportation Safety in COVID-19 Era Through Crowdsourcing Technique



Qisheng Pan, Zhonghua Jin, and Tao Tao

17.1 Background

Public transportation systems in major cities around the world have been plagued since the novel coronavirus disease 2019 (COVID-19) first broke out in Wuhan, the capital city of Hubei Province, China in the end of 2019 (Huang et al. 2020). When the COVID-19 pandemic quickly spread out to surrounding cities, many municipal governments imposed lockdown policies and human mobility restrictions. As the center of the pandemic, Wuhan ordered a complete lockdown on January 23rd, 2020. Within hours, orders were issued to lockdown cities and towns near Wuhan. A total of 57 million people in Hubei Province were required to stay at home with travel restrictions. In this extreme case, public transportation stopped operation in Wuhan and its surrounding cities completely for several months.

Different from China, many western countries hesitated to order total lockdown or travel restrictions. For example, Sweden had no lockdown policy at all and allowed people voluntarily adjusted their travel behavior without any government interventions (Born et al. 2020; Oum and Wang 2020). The U.S. government did not impose lockdown or domestic travel restrictions at the initiation and early spread of the COVID-19. When the White House issued new guidelines on March 16th, 2020 to slow down the rapid spread of coronavirus, the federal government allowed each state to decide when to start and when to end the lockdown period independently. Urban public transit either stopped services all together or reduced services significantly in all American metropolitan areas since then.

Q. Pan (✉) · T. Tao

Department of Public Affairs and Planning, College of Architecture, Planning, and Public Affairs

(CAPPA), University of Texas, Arlington, USA

e-mail: qisheng.pan@uta.edu

Z. Jin

Department of Urban Planning and Environmental Policy, Texas Southern University, Houston,
USA

After the U.S. began to re-open in the early of May 2020, the restrictions on businesses and public facilities were gradually lifted across the country. However, the ridership of transit service has remained much lower than the normal levels because many people have chosen either to work from home or drive alone to avoid taking public transit for fear of getting infected by the COVID-19 virus. Many U.S. Metro agencies predict that it may take years before their transit ridership returns to the pre-COVID-19 levels. Since passenger fare accounts for over one-third of total public transit revenue according to American Public Transportation Association (APTA), large drop of ridership has resulted in significant reduction of transit services, much less revenue generated by transit fares, and sizable budget cuts.

Currently, both transit service providers and transit riders worry about the risks for transmitting coronavirus in transit service, especially before the vaccine and the cure for COVID-19 are ready. It is critical to adopt effective mechanisms to address these concerns and protect transit riders as well as transit drivers in services during the pandemic. Public transit planners and decision makers are facing big challenges, such as collecting epidemiological information timely, analyzing the data quickly, and turning the research findings to actions and policies, so as to predict, prevent, monitor and respond to the COVID-19 properly. Traditional data collection approaches like onboard survey do not meet these urgent needs.

In a recent empirical study, Argente et al. (2020) estimated that the change in commuting patterns reduces the number of COVID-19 cases by 200,000 and the number of deaths by 7,700 in Seoul, South Korea after the public disclosure of spatial information collected from the individuals who were tested positive for COVID-19. But this study did not measure the costs of losing privacy from the disclosure of COVID-19 cases.

Crowdsourcing works as a model to collect information from individual users and share the information with the others through internet on a voluntary basis. It has been proposed to collect and share information for the ongoing pandemic (Desai et al. 2020; Chodera et al. 2020; Freeman et al. 2020; Leung and Leung 2020). It offers an opportunity to obtain symptom and other related information right from travelers and allows the information to flow between parties smoothly (Chunara et al. 2012, 2013). This novel technique enables the collection and dissemination of epidemic data to transit planners and decision makers directly and improves spatial and temporal resolution for epidemic monitoring in transit system. New Jersey Transit has tried to adopt crowdsourcing techniques in its mobile app to tell riders about the crowd level of incoming bus or subway train, which may help them to avoid close contact and reduce the spread of the COVID-19 virus.¹

This research intends to review current public transit status in the U.S. and the existing studies on the connections between public transit and COVID-19 dissemination. Based on the newly available data and the advance in technology, it proposes a conceptual framework of crowdsourcing technique to collect and share the information of COVID-19 between transit operators and riders who voluntarily participate

¹<https://statescoop.com/nj-transit-tests-crowdsourcing-feature-to-reduce-rail-and-bus-congestion/>.

in supporting public transit services, validate the crowdsourcing data through fusion and analysis of multiple data sources, and quickly find and notify the close contacts. The information of the infected riders and those in danger (i.e. confirmed patients, suspected patients, fever patients, close contacts) is comprehensively examined to understand the sources and the trajectories of the infection, estimate the risk of infection, and conduct scenario simulations, etc., so as to support the prediction, monitoring, early warning and response of COVID-19 in public transit system. The policy implications are also discussed at the end of the study.

The rest of the chapter is organized as the following sections. Section 17.2 reviews the public transit status in the U.S. and the challenges transit is facing in the COVID-19 era. Section 17.4 describes the crowdsourcing techniques and proposes a framework for a crowdsourcing-based public transportation information system (CB-PTIS) to prevent COVID-19 transmission in public transit system. Section 17.5 discusses policy implications and draws conclusions.

17.2 Public Transit in the US

17.2.1 *The Important Role of Public Transit in Urban Transportation*

In many U.S. metropolitan areas, public transit plays a critical role in economy development and well-being of society, especially to the transit dependent population. Many urban residents in central city take public transit routinely for daily activities while many suburban commuters ride buses or trains to workplace in urban core through transit centers or park-and-ride facilities. Access to public transit services provides affordable mobility across all social classes, especially to the residents in large metropolitan areas (Griffin and Sener 2016). There are over 30 million daily trips made by public transit in the US, and the trips within the nation's top one hundred largest metropolitan areas account for about 95% of all transit passenger miles traveled (Tomer et al. 2011).

The American Public Transportation Association (APTA) reported that the public transportation industry in the U.S. is worth over \$74 billion dollars. There are over 6,800 public transit providers providing more than 400,000 jobs.² Though the mode share of public transit is relatively low, the total number of riders is fairly large. In 2019, there were almost 10 billion trips made by public transit in the U.S. Public transit development also generates a variety of economic returns, including the increase of property values, the promotion of business sales, and the creation of job opportunities, etc. APTA estimated that every one-dollar investment in public transit generates about five time more economic returns.

²<https://www.apta.com/news-publications/public-transportation-facts/>.

17.2.2 The Challenges of Public Transit in the US

Though public transit is a critical component of urban transportation system, it has a much lower share of trips comparing to private vehicles in the developed countries. In the U.S. large metropolitan areas, there are only 7% or about 6.5 million of workers taking transit in their daily commute (Tomer et al. 2011).

Public transit network is distributed unevenly in different parts of the U.S. The large metropolitan areas in the West and the Northeast have higher density and larger coverage of public transit than those in the South. About 45% U.S. population have no access to public transportation. It's been recognized that the public transit ridership in major U.S. metro areas has become flat or declining over the past years (Graehler et al. 2019). There are a variety of reasons behind it, such as suburbanization, road network development, increase of auto ownership, quality of services, and people's desire for high quality of life, etc. In recent years, innovative travel means such as Uber or Lyft has emerged as new alternatives to travel and complemented the fixed-route transit network (Hall et al. 2018).

The public transit ridership, especially the bus ridership, had been dropping gradually even before this pandemic (Fig. 17.1). Many cities had plans to change transit services in the areas with poor ridership. One argument is that public transit rarely

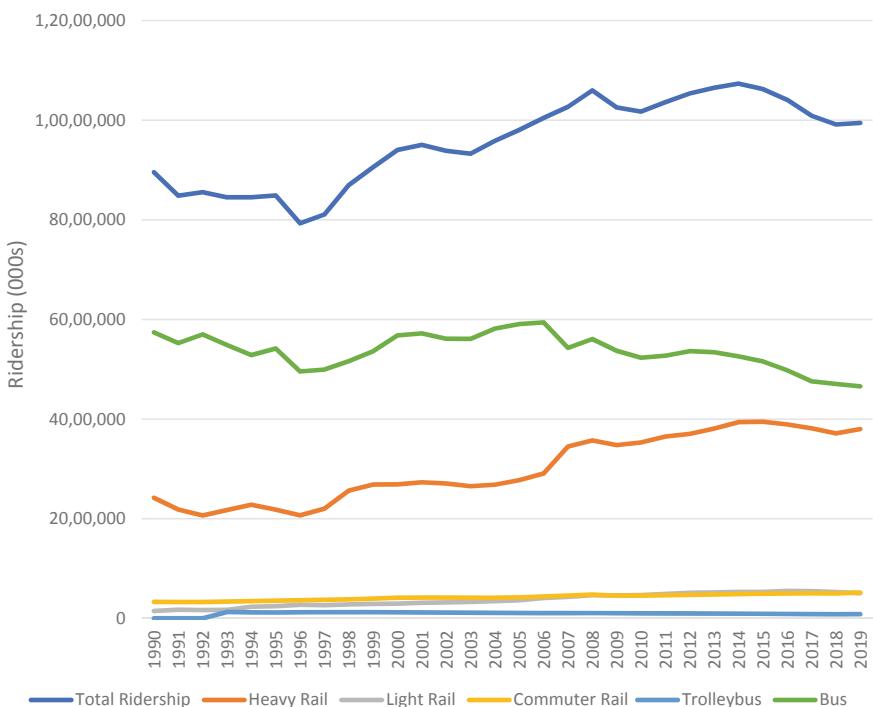


Fig. 17.1 Change of transit ridership, 1990–2019

makes profit and most of transit services are subsidized by local tax money. Harford (2006) conducted benefit-cost analyses on the public transit systems in 81 urbanized areas. He concluded that for a conservative estimate, less than one-third of the areas had a benefit-cost ratio greater than one, even in those urban areas having the largest number of population and high density of transit network.

17.2.3 The Impacts of COVID-19 on Public Transit Ridership and Revenue

The unexpected COVID-19 pandemic makes the situation of public transit operations much worse than before. Because many people worry that public transit is overcrowded and prone to create an environment for virus spreading, they avoid taking public transit for commuting and other activities. Since March 2020, many companies and public entities have started to promote telecommuting and work from home (WFH).

According to the APTA data, the total number of riders decreased from 2.5 billion to 0.6 billion and total transit passenger trips dived by 76.01% while the heavy rail ridership dropped by 87.39%, light rail ridership decreased by 74.44%, commuter rail ridership fell by 89.76%, trolley bus ridership lowered by 86.04%, and bus transit ridership reduced by 65.04% in the second quarter of 2020 comparing to the same period of 2019 (APTA 2020).

Though many local transit services around the U.S. gradually recovered, transit ridership is still much lower than that before the pandemic. For example, Houston METRO reported that the ridership of its local bus, light rail, park and ride decreased by 46%, 54%, and 87%, respectively, in September 2020 comparing to the same month in 2019. Based on a day-to-day ridership comparison, New York MTA had its subway ridership reduced at a range of 64.8–72.6% and bus ridership decreased about 46–55% between Oct. 27, 2020 and Nov. 5, 2020, comparing to the equivalent days in 2019.

The revenue of public transit in the U.S. comes from four major sources, i.e. transit agencies generated funds, local government funding, state government funding, and federal governmental funding. According to APTA'S Public Transportation Investment Background Data in 2013, the largest part of directly generated revenue comes from passenger fares, which accounted for 42.7% of all operating revenue and 23.7 percentage of all capital revenue. Local government funding from state sale tax or income tax accounted for 22.2% of operating revenue and 18.4% of capital revenue. State revenues or state financial assistance collected tax or fees comprised about 26% of operating revenue and 16% of capital revenue. The federal assistance from the federal government accounted for almost 9% of operating revenue and 41% of capital revenue in 2015 (APTA 2015). The spread of COVID-19 has not only made the transit ridership drop, but also caused dramatical cuts of transit revenue.

In September 2020, the APTA requested the U.S. Congress to provide at least \$32 billion federal assistance to support transit industry. However, it's uncertain whether the additional federal fund is available or when it will be in place at current circumstance of high unemployment rates and significant economic drop.

To avoid overcrowding and ensure transit safety, many transit systems have reduced their capacity and started to provide regular sanitizing and cleaning, which yields much less daily revenue and generates higher expenses. Even though that transit systems are subsidized by local tax revenue, the poor economy has made it harder for transit to maintain its pace. Any impact of funding cut on the transit operation would primarily affect the transit dependent groups or disadvantaged population groups, especially the minority in low-income neighborhoods. Without necessary public services, many urban residents who depend on the public transit system may lose their access to job opportunities and other socio-economic resources.

17.3 Studies on the Connection Between Public Transit and COVID-19 Dissemination

To impose lockdown or resume economy in the COVID-19 era, it is important to examine the balance between public safety and economic development. In the cases without mandatory lockdown and travel restrictions, people are unlikely to stop making person-to-person contacts voluntarily due to job responsibilities, family needs, and many other important reasons. They are weighting the benefits and costs of personal contacts through different modes of transportation (Fenichel et al. 2011). It is feasible for employees in some professions to work from home, such as those in information technology or financial services. However, many others have to earn a living by taking the risk of exposure to COVID-19, especially those low income and minority people. Many of them belong to transit dependent group.

There are a growing number of studies on public transit safety in the COVID-19 pandemic era. Because of the crowded environment on bus or subway train, public transit has been considered an incubator for coronavirus transmission (Qian et al. 2020). Some regarded public transit as a major disseminator of coronavirus infection (Harris 2020) while the others reported no direct correlation between the use of urban public transit and the transmission of COVID-19 (Schwartz 2020).

Desmet and Wacziarg (2020) examined the determinants of heterogeneity of COVID-19 cases and deaths across the U.S. between March 15, 2020 and June 29, 2020. They adopted the data of COVID-19 cases and deaths at county level reported by the New York Times and the public transportation variables obtained from the 2018 American Community Survey 5-Year Estimates. They reported that the counties with a higher proportion of individuals going to work by public transit have significantly higher severity of COVID-19.

Harris (2020) blamed the subway system for the dissemination of COVID-19 infection in the early stage of the epidemic in New York City through simple correlations between the number of newly reported COVID-19 cases and subway turnstile entries between March 1, 2020 and April 3, 2020. As the author admitted, however, scientific reviewers may argue that cause-and-effect of coronavirus transmission in subway system is difficult to prove. Furthermore, both the commuters in the subway system and the MTA workers lacked necessary protections during the initial takeoff of the massive epidemic in March 2020. The situation was improved after April 2020.

Knittel and Ozaltun (2020) adopted both multiple linear regression and negative binomial mixed models to correlate U.S. county-level COVID-19 death rates with several sets of key variables, including the modes of commuting between April 4, 2020 and May 27, 2020. They found that all the modes except for biking are associated with higher death rates than telecommuting while public transit is the mode correlating with highest death rates. But the authors also emphasized that their study only identifies correlation rather than causal relationship between the variables. They inferred that some of the higher death rates correlated to public transit use may come from public transit itself while some come from the day-to-day interactions of the transit riders at their workplace. But they did not explore further the contributions of public transit and interactions at the workplace respectively. The authors also highlighted a greater need to sterilize and impose socially distance in public transit systems and at workplace.

Glaeser et al. (2020) examined the effects of mobility on COVID-19 spread. They obtained COVID-19 weekly cases by zip code in New York City (NYC) and cross-sectional data for Boston, Philadelphia, Atlanta, and Chicago. The COVID-19 cases were related to the mobility estimated from subway turnstile data for NYC and SafeGraph cellular phone data for all these cities. Their study found that every 10% drop in mobility leads to a 20% reduction of COVID-19 cases per capita. The regression using the NYC Metropolitan Transit Authority (MTA)'s turnstile data reported that a 10% fall in public transit trips is associated with 4% fewer COVID-19 cases per capita. However, the authors emphasized that their results do not provide a clear estimate of the impact of public transportation use on the spread of COVID-19 because their instruments were not specified for public transit.

In contrast to the analyses above at aggregated population and transportation levels, there are a few studies about the risk of COVID-19 transmission between individual travelers in certain mass transportation modes. Hu et al. (2020) examined the transmission rates of COVID-19 on high-speed trains in China. They reported that passengers' risk of catching virus is highly dependent on how close they sit to an infected person and how long they travel together. Passengers sitting near an infected person have much higher probability to catch the virus than those sitting farther away. The average infection rates also increase by 0.15% by each additional hour traveling with a COVID-19 patient. The authors suggested that passengers on train should sit at least two seats apart from each other and have travel time less than three hours to keep a low infection rate. Hu et al. (2020) also admitted that their study may overestimate the risk of COVID-19 transmission on a high-speed train because

they cannot obtain the data about social relationships of the passengers or the use of protective equipment.

There are also many studies reporting that public transit is not the major source for the COVID-19 transmission. Almagro and Orane-Hutchinson (2020) obtained the incidence rates of COVID-19 and the number of tests performed in April 2020 from NYC's Department of Health (DOH). They also acquired demographic, occupation, and commuting related information at zip code level from the US Census American Community Survey. Their study found that the length of commute and the use of public transportation have no significant impact on the rates of COVID-19 after controlling occupations.

Public Health France traced 150 virus “clusters” detected between May 9 and June 3 in Paris, France and found that none of them was located in transportation system including public transit.³ Similar epidemiological studies in Vienna, Austria in July 2020 and Tokyo, Japan in May 2020 also reported that none of their virus clusters was traced back to their public transit systems. Another example came from Hong Kong, China, in which most people travel regularly by public transit. By July 2020, the total number of COVID-19 cases in Hong Kong was less than 2,000, which is so small in a population over 7.5 million. It was credited to travelers’ self-protective behaviors and the adoption of certain preventative mechanisms such as traveling alone and wearing masks.

After a comprehensively review of the studies on the COVID-19 transmission in public transit, Schwartz (2020) reported that COVID-19 case rates are mainly connected to community spread rather than public transit ridership rates in the empirical cases. He pointed out that the occupation of the commuters affects the probability of virus infection far more than their travel modes because the essential workers have higher case rates regardless of their commute mode. It concluded that there is no direct correlation between the use of public transit and coronavirus transmission.

In a summary, the relevant studies in the literature have not reached an agreement on the connection between public transit and COVID-19 dissemination. A strong relationship between these two variables has been found in some studies, but most of them emphasize that it is a correlation rather than causal relationship (Knittel and Ozaltun 2020) or the instrument is not designated for public transit (Glaeser et al. 2020). Many other studies such as Almagro and Orane-Hutchinson (2020) and Schwartz (2020) have not found a connection between public transit and COVID-19 dissemination. They claim that public transit is not the major contributor of COVID-19 transmission.

³<https://www.leparisien.fr/societe/coronavirus-pourquoi-aucun-cluster-n-a-ete-detecte-dans-les-transports-05-06-2020-8330415.php>.

17.4 Crowdsourcing for Improving Public Transportation Safety in the COVID-19 Era

17.4.1 *Introduction to Crowdsourcing and Its Applications*

As a term, crowdsourcing was originally used by Jeff Howe to describe a new business structure benefited from technology advances in an article of *Wired* in June 2006 (Brabham 2013). It refers to the process of obtaining and sharing knowledge, goods, or services in a large group of people through online requests in an efficient and low-cost way. It is also the practice of engaging the services of the crowd to enter or deliver information for a specific task or activity (Howe 2009). It has been considered as a problem-solving model to leverage collective intelligence for public good through online communication (Brabham 2010). The technology can be utilized through social media like Facebook or Twitter, smartphone apps, or dedicated platforms for real time monitoring, planning, and research purposes. The promotion of crowdsourcing on well-known networks allows manipulators to gain access to a large user base in a short period of time.

The concept of crowdsourcing has spread out from business to many other fields. In public transit studies, crowdsourcing technique has been initially adopted to solicit ideas and information in public participation programs. As an example, Next Stop Design launched in 2009 as an online competition project funded by Federal Transit Administration tested crowdsourcing for transit planning in a public participation context. It allowed participants to submit their bus stop shelter designs and vote for the best one of their peers (Brabham 2010; Brabham et al. 2010). In a relevant study, crowdsourcing was adopted in Cincinnati bike-share plan to learn the challenges and explore its implementation in non-motorized modes (Afzalan and Sanchez 2017).

Crowdsourcing technique was also utilized in disastrous outbreaks, such as the 2012 and 2013 Colorado wildfires for resource allocation and health status report, the cholera in Haiti's 2010 earthquake (Bengtsson et al. 2011), and the dengue in Thailand and Indonesia (Desai et al. 2020).

Many existing studies on the spatial and temporal dynamics of infectious diseases are based on the epidemic data collected from the public, which lags far behind the rapid development of the epidemic. Crowdsourcing technique has been adopted to expedite data collection and analysis process in biomedical informatics and health-information technology since 2010 (Desai et al. 2020). In crisis management for bioinformatics and ecology, data collected from individuals provide heterogeneous views and supplementary solutions to the existing systems in normal or interrupted situations (Lakhani et al. 2013; Anderson et al. 2002; Meymaris et al. 2008; Bengtsson et al. 2011). In infectious disease surveillance, crowdsourcing enables the assembly of the symptom and related information right at the point of care (Chunara et al. 2012).

Though crowdsourcing has been implemented in many fields including public transit and infectious disease surveillance. There are great challenges in predicting, monitoring and responding to the infectious diseases in public transit. An urgent

need is to address the issues in data acquisition, including timeliness, convenience, authenticity, accuracy and completeness, etc.

17.4.2 The Techniques Addressing the Security Concerns About Crowdsourcing

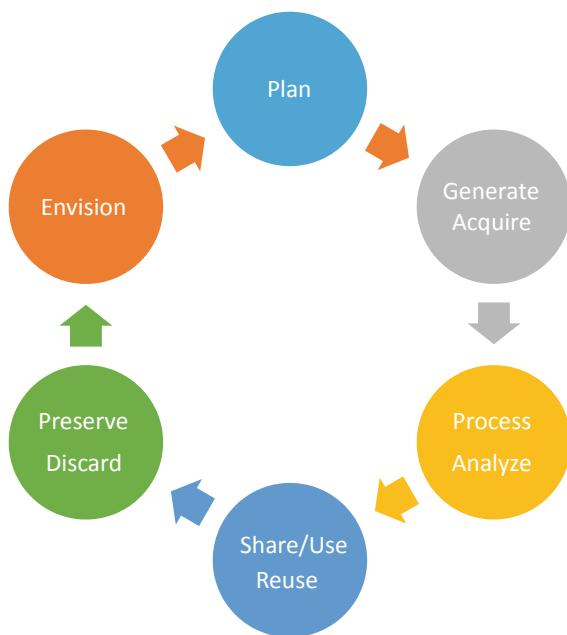
Crowdsourcing has the advantages to collect and assemble large amount of information from various sources. However, certain risks are associated with mismanaging such large-scale datasets. Information security and privacy protection are the most concerned aspects for research data collection as potential data breach can cause severe social economic damages to the stakeholders. In 2019, the National Institute of Standards and Technology (NIST) has published its Research Data Framework (RDaF) that is under construction by the Information International Associates (IIA). The preliminary RDaF was expected to be submitted in December 2020 to the Community of Interest, including the representatives from government, academia, private foundations, international organizations, non-government advocacy, and scholarly publishers.

This framework was originally inspired by the Framework for Improving CRITICAL Infrastructure Cybersecurity that was issued by the NIST in 2014. It is structured under the NIST Cybersecurity and Privacy Frameworks. It consists of three key components, i.e. the Framework Core, the Framework Profiles, and the Implementation Tiers. Currently, the core functions as the main body carry out data management, categorize data needs and activities, and also set data standards and guidelines. The Framework Profiles act as an outreach organism that facilitates the communication between different level of users and stakeholders. The RDaF is considered as a data management framework that maps the key elements of research data, including Who, What, Where, Why, and When, etc. It provides a dynamic guidance of research data management by structuring customizable strategies, following the multiple steps of plan, generate/acquire, process/analyze, share/use/reuse, preserve/discard, and envision (Fig. 17.2). The RDaF is also recognized as a consensus document that helps the users/stakeholders to understand the risk and cost-benefit of the research data and work as a tool of research data management (NIST 2020).

Though it has a framework ready for implementation, the RDaF is still under the review process of the Community of Interest. One of its key components, the Framework Implementation Tiers, is not yet included in the final report of the Preliminary RDaF, instead it is supposed to be published in the next version. It is claimed that the implementation tiers would help the organizations to assess and keep tracking their data management status and create roadmap for future activities.

In addition to socio-economic and security concerns, data accuracy and completeness are also very important in transit research and operations. Digital Science has addressed this issue since 2016 in their series of reports about sharing, using, and redistributing open data. Digital Science is an organization that provides smart tools

Fig. 17.2 RDaF description
(NIST 2020)



for research data solutions to researchers, institutions, publishers, funders, as well as industrial users. The organization has created a cloud-based repository named Figshare, where academics can publish their research openly, so as to encourage academia researches to build their work on each other's databases. Digital Science has also published a number of open data reports to discuss the issues and concerns of data accuracy and completeness in data sharing. In 2019, Digital Science published its fourth report, i.e. the State of Open Data 2019, that claims itself as the longest running longitudinal studies on open data.

According to the report, there are constant fears of data sharing and misuse in today's research environment, and the barrier to study data sharing and reuse is a matter of trust. By surveying data users, the 2019 report listed the top three data sharing concerns as the misuse of data, copyright and licensing, and proper credit or acknowledgement. The survey also revealed that over 2,000 respondents have concerns about misuse of their research data. It reported that 65% of respondents provided their data for sharing either privately or publicly in 2019, similar to the percentage of respondents in 2016, i.e. 67%, but lower than that in 2017 and 2018 (both are about 74%). Additionally, a majority of respondents, i.e. about 79%, were supporters for a national mandate to make primary research openly available (Digital Science 2019).

The solution for data sharing is to follow the concept of FAIR Data, i.e. Findable, Accessible, Interoperable, and Reusable. It claims that adopting such policy has a high cost, but the creation of such norm would bring extraordinary reward as open science is ultimately an investment, rather than a cost.

17.4.3 *Crowdsourcing Techniques Developed for Dealing with Infectious Diseases*

Traditionally, the official surveillance system for epidemic infectious diseases is based on the counts and reported cases of epidemic diseases from medical system and front-line medical professionals. However, this procedure has two major limitations. (1) Only moderate and severe cases are reported in the monitoring systems while those with mild and self-healing cases are not, which may result in the failure of monitoring and responding at the early stage, especially for the infectious diseases having a long incubation period, such as COVID-19; (2) Because the most important task of medical staff is to treat and help patients, it is difficult for them to track the cases before they are approached by researchers or statistician.

Geospatial data and tools are essential for emergency management (NRC 2007). The geographic information created by public participation, namely crowdsourcing (Howe 2006) or Voluntary Geographic Information (VGI) (Goodchild 2007), characterize as convenient access and dynamic updates, is an important supplement to traditional data collection. A framework based on crowdsourcing or VGI plays a unique role in predicting, monitoring, and responding to major infectious diseases.

Goodchild (2007) first proposed the concept of VGI in response to the phenomenon that individuals spontaneously use the Internet to create, collect, and disseminate geospatial information. At that time, the emergence of both Wikimapia and OpenStreetMap has given the public an ability to create their own geographic information. Goodchild and Glennon (2010) worried about the quality of VGI but they also emphasized the advantages of VGI, especially when responding to emergent situation like infectious diseases.

Sui et al. (2013) described public participation in geographic science, the applications of VGI in geographic knowledge production, and the relationship between VGI and traditional geographic information in data acquisition and dissemination. Boulos et al. (2011) discussed the tendency, standards and applications of crowdsourcing technique, citizen sensing, and sensor web in public and environmental health monitoring and crisis management.

In the context of information communication environment represented by mobile Internet and social networks, Gao et al. (2013) proposed a systematical framework and explored the applications of crowdsourcing technique in disaster emergency management. Becker and Bendett (2015) also investigated the implementation of crowdsourcing technique in handling disaster and emergency. Liu and Liu (2016) discussed the application of crowdsourcing from social public crisis management. However, none of these studies are directly related to the monitoring, prevention and response to major infectious diseases.

In a study of SARS-COV and seasonal influenza in 2002/2003, Boulos and Geraghty (2020) highlighted that geographic information systems (GIS) can be used to map the disease transmission on social media in real-time, to predict and visualize the risks using travel data, and to display the spatiotemporal information of the trajectory and the behavior of super spreaders, which are all essential mechanisms

to deal with epidemics in a timely and effectively manner. These applications of GIS visualization functions on mobile devices can be used to track and map the spread of COVID-19 in 2020.

In the field of epidemic monitoring, one successful application of crowdsourcing or VGI was the “Google Flu Trend” platform run by Google between 2009 and 2013. By assembling google search data with simulation models, the Google Flu Trends platform made predictions on influenza and dengue fever in 25 countries around the world with very good results, which highly match with the CDC data (Ginsberg et al. 2020; Cook et al. 2011). In the following years, more public health and disease detection platforms have emerged basing on crowdsourcing or VGI technique, such as GrippeNet.fr in France and Flu Near You in the U.S.

Recently, many online platforms were developed to track the global statistics of COVID-19 with real time information and network interactions. Based on ArcGIS Online, the framework developed by the Center for Systems Science and Engineering at Johns Hopkins University provided network interactive functions to track real-time spread of new coronary pneumonia in the world. People can query the confirmed cases of new coronary pneumonia in real time. The data comes from the World Health Organization (WHO), the United States Centers for Disease Control and Prevention (CDC), the Chinese Health Commission, the European Centers for Disease Control and Prevention, and the Chinese online medical resource Dingxiangyuan (DXY.cn) (Dong et al. 2020). However, this platform is not opened to the public or allows public participation in the contribution of spontaneous data. Neither crowdsourcing nor spatial analysis functions are provided on this platform. The platform for searching the close contact persons provides some query functions while personal information is entered through VGI with crowdsourcing technique, but the quality of crowdsourced data is not guaranteed.

As a forefront technique described by Goodchild and Glennon (2010), crowdsourcing has provided a way for data acquisition through mobile devices such as cellphones or tablet and the grassroot participation in relatively low cost. It can be used to improve efficiency of data acquisition through mobile devices. The issues of data authenticity and accuracy have been addressed with the fusion of multiple data sources.

17.4.4 Crowdsourcing Applications in Private and Public Transportation Services

Public transport provides many urban residents, especially transit dependent population, essential services to gain access to necessary resources, such as employment opportunities, social activities, medical care, and recreational opportunities (FHWA

2002). As the service providers, public transportation agencies are expected to evaluate and meet the demands of their passengers. Crowdsourcing allows public transportation agencies to achieve their goals with the participations of customers who tell public transport service providers their needs, requirements, and preferences, etc.

An important application of crowdsourcing in public transportation is to provide real time public transport information. Many public transportation agencies post real-time transport information with vehicle location through their websites and apps automatically. This information can be improved by crowdsourced data. However, its applications in public transportation are limited because it takes much complex procedures for development than standard travel planning applications. Some mobile applications have been developed by commercial or private vendors to carry out multiple functions, such as problem reports, multi-mode connections, tickets, and local information, etc. The applications available in market include Transit Apps, Moovit, Swiftly and Tiramisu, etc.

Transit App⁴ is designed for aggregating and mapping real-time public transit data on mobile devices. It operates in over 175 metropolitan areas around the world and aims to minimize the demands of private vehicles in cities (Peters 2019). Crowd-sourced data is adopted by Transit app to pinpoint the location of buses and trains. It also offers users the schedules of the available services of multiple transportation modes, including bus and rail.

As a real-time worldwide public transit app, Moovit⁵ collects both crowdsourced data and official data of public transit to assist riders on trip planning, route maps, and schedules, etc. It also provides data and APIs to transportation service companies and transit agencies. It intends to fill the gaps in the areas where official transit information is not available.

Swiftly⁶ is a transit mobile application that serves as a crowdsourcing platform to encourage riders to share information about crowding, delays, and disruptions, etc. It combines real-time transit operation information with crowdsourced data from riders together to provide accurate real-time data of passenger information and vehicle operations. It is also able to visualize historical transit data as charts or graphs to facilitate knowledge extraction and decision making.

Tiramisu⁷ is one of the first crowdsourcing-based public transport applications developed in 2011. It initiated the ideas of asking riders to provide information such as how full their bus is and whether there is any wheelchair space left. It also allows transit riders to share the trajectory of their trips to improve the estimates of arrival times, bus fullness, and rider experience (Yoo et al. 2010; Steinfeld et al. 2011).

In addition to private vendors, public transit service providers also implement crowdsourcing technique in their operations. As a commuter shuttle service company in New Zealand, Chariot plans new routes based on demographic information and

⁴<https://transitapp.com/>.

⁵<https://moovit.com/>.

⁶<https://www.goswift.ly/>.

⁷<http://www.tiramisutransit.com/>.

crowdsourced data (Cutler 2014; Suzdaltshev 2014). It obtains crowdsourced information from their riders to provide assistance on route development in two ways: First, they adopt algorithms to plan transit routes according to user demands; Second, they ask riders to vote on new routes. If a proposed route gets enough supporters within a month, Chariot will launch the new route in a week (Lazo 2017).

Harris et al. (2016) summarized the applications of crowdsourcing technique on transit services in four aspects: (1) Providing a channel to obtain real-time travel data from riders through a passenger information system while riders use their mobile devices to report and receive information about network conditions, such as the delays and the quality/capacity of transit services; (2) Delivering public transport service information in the areas where traditional sensor networks are not available, for example, in the rural areas; (3) Reporting service quality issues to bus service providers and other riders in public transportation network; (4) Creating an interactive community among public transportation riders and service providers.

17.4.5 Conceptual Framework and Strategies of a Crowdsourced Public Transportation Information System for Preventing the Transmission of COVID-19

Based on the existing studies and available techniques, this study proposes a conceptual framework and relevant strategies for a crowdsourcing-based public transportation information system (CB-PTIS). It aims to collect crowdsourced data together with accurate real-time information as input to both transit service providers and riders to ensure safe environment and quality services on public transportation during the COVID-19 era. The CB-PTIS has several key components that help to implement the principles and enforce the regulations to prevent the spread of the coronavirus disease, which is demonstrated in Fig. 17.3.

One of the key components is the crowd, who are transit riders, service providers, and any other participants or stakeholders in public transportation system. Through mobile devices or an appropriate interface, the crowdsourced data or the inputs from the crowd include the information about events, incidents, activities, and quality of services of public transportation system that may affect the transmissions of COVID-19 and the safety of the system.

Mobile devices, i.e. smartphones, are the necessary component for the communication between crowd and public transportation information system. They are employed as an interface for collection and the dissimilation of crowdsourced public transportation information. They are used to gather data from the crowd either passively or actively. The spatial location and the trajectory of the crowd are collected and transmitted through the built-in GPS passively and automatically while the crowd enter the information of their own or the others and the relevant information about public transportation system actively.

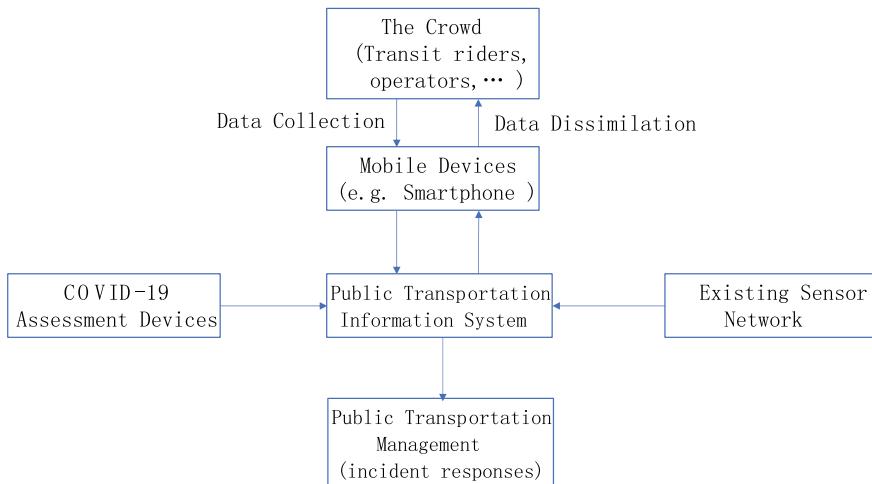


Fig. 17.3 Conceptual framework for a crowdsourcing-based public transportation information system (CB-PTIS) for preventing COVID-19 transmission

Another important component is the assessment devices with specific functions to block the transmission of COVID-19, such as non-contact temperature assessment, face mask/covering detection, and crowding measure, etc. Although a non-symptomatic person with COVID-19 may be infectious without a fever or any other symptoms, a higher-than-normal temperature is considered an important indicator to identify a person who may be infected by COVID-19. As a simple, cheap, and effective way to limit the spread of COVID-19, face mask or covering is recommended or required in many public indoor spaces, including public transportation facilities. Social distance is also one of the best ways to reduce the spread of the virus. The devices with non-contact temperature assessment, face mask/covering detection, and crowding measure devices can be installed at the entrance of subway and bus stations, or on railcars and buses. The real-time information collected by the devices is employed for enforcing federal, state, and local regulations for social distance, face masks and respirators during the COVID-19 public health emergency. Some real-time information like overcrowding status can also be obtained from the crowd through crowdsourcing technique. The disseminated information about public transportation system allows the passengers waiting at stations to make a reasonable decision about keeping or changing their trip plans after weighing the risks and benefits of the trip.

Public transportation information system (PTIS) plays a key role in the framework. It has the necessary functions for data collection and analysis. In addition to process and dissimilate the real-time information collected from the existing sensor network, it also incorporate crowdsourced data from the crowd inputs and the measures of COVID-19 assessment devices to block the possible transmission of COVID-19 and ensure the safety of public transportation network.

Based on data analysis results, the PTIS compiles user-friendly information that is disseminated to the subscribed crowd or allows the crowd to gain access through social media, text messages, public transportation websites, radio, television, and other media interfaces. The PTIS information includes the crowding level, cleanliness, frequency, arrival time, and quality of services about bus or train, etc. On the other side, it makes recommendations to policy makers or planners about public transportation management, especially the quick response to the COVID-19 incidences.

Though the crowdsourcing-based transit applications are more complex than standard transit planning applications, the technology preparation enables public transportation agencies to add COVID-19 assessment devices to their sensor network and deliver the relevant information to their riders or service providers to make decisions timely for COVID-19 prevention. It also allows to incorporate crowdsourced COVID-19 data and real-time information to the existing mobile apps, such as Transit Apps, Moovit, Swiftly and Tiramisu. The mechanisms developed in these public transportation applications, such as the functions for reporting quality of service, can be adopted by riders and service providers to actively present and share their feedbacks on preventing the transmission of COVID-19.

Data collection is a complex process that involves many responsible parties and stakeholders, such as data collectors, archivists, processors, managers, and end users, etc. It is critical to maintain data accuracy and security to minimize the potential risk of data misuse. One major concern is to protect ‘privacy during the entire process from data collection to data publication.

Many countries have imposed regulations on research data collection. For example, in the U.S., the Food and Drug Administration (FDA) requires research data involving human subjects comply with the regulations of Institutional Review Board (IRB) to protect the rights and welfare of humans. Researchers who collect and use human subjects are required to complete the formal IRB training before they can collect data and perform analysis. In the European Union, the regulatory body for personal information protection is the General Data Protection Regulations (GDPR), powered by the Regulation 2016/6793 of the European Parliament and the Council. However, different countries may have its different regulatory bodies and policies regarding to personal information. In China, personal information protection is a complex issue because private information in different sectors is monitored and supervised by different government regulatory bodies, including the Cybersecurity Administration of China (CAC), the Ministry of Public Security (MPS), the National Medical Products Administration (NMPA), and the Ministry of Science and Technology (MST), etc. Among them, some have very tight regulations while the others have much loosed policy on personal information collection and usage. In data collection process, this framework needs to consider the standards and policies of different countries and regions because data collected from one region may not comply with those collect from the other places. It is impossible for the US and Western countries to adopt many of the effective practices in China and East Asia for COVID-19 data collection and contact tracing.

Another important issue about COVID-19 data collection and processing is the rampant misinformation overwhelming the internet. The misinformation on COVID-19 has been used by hackers in cyberattacks. It has been challenging for the PTIS to validate the crowdsourced data and control data quality. The crowdsourcing data are usually validated in multiple steps through the comparison of multiple data sources. The crowd may check and approve the information submitted by the others. The data collected from registered or validated users is more reliable than those submitted by the guests or the users that have not been validated. There are also some local and global newspapers, medias, and platforms reporting on COVID-19 misinformation at a daily base or even at real-time, such as the Washington Post Factchecker,⁸ NewsGuard,⁹ and Snopes,¹⁰ etc. The PTIS will check these platforms periodically and compare the crowdsourced data against those from the trustworthy sources. The crowdsourced data will be adopted by the PTIS after they are validated in the CB-PTIS system.

17.5 Policy Implications and Conclusions

As Harris et al. (2016) pointed out, very little research has been carried out to learn customer preference and network efficiency of crowdsourcing applications. There are even fewer studies on the crowdsourcing applications on public transit in the COVID-19 era. This research intends to fill the gap by exploring the adoption of crowdsourcing technique to block the transmission of COVID-19 in public transit and improve public transportation safety.

Based on the reviews of existing studies, crowdsourcing is considered as a promising technique to meet the challenges that public transportation system is facing now on preventing the transmission of COVID-19 among transit riders. The security concerns about the crowdsourcing technique are addressed in the new research data framework, such as NIST's new RDaf and Digital Science's cloud-based repository Figshare. The adoption of crowdsourcing technique in epidemic disease monitoring and disaster management has demonstrated its capability in dealing with the public health emergency. The extensive applications of crowdsourcing technique also highlight its potential in stopping the dissemination of COVID-19 in public transportation services. The conceptual framework proposed in this study illustrates the relationship between the crowd, COVID-19 assessment devices, existing sensor network, and public transportation information system.

To implement the framework, some policy implications are formulated as follows: (1) Crowdsourcing is not limited as the inputs of the crowd. It also includes the engagement of a public transportation community that collaborates and contributes to block the transmission of the COVID-19; (2) Body temperature check, facial mask

⁸<https://www.washingtonpost.com/news/fact-checker/>.

⁹<https://www.newsguardtech.com/>.

¹⁰<https://www.snopes.com/>.

or covering, social distance, hand hygiene are the general practices for protecting riders and operators from the COVID-19 on public transit. A combination of assessment devices and crowdsourced information is necessary to ensure the safety by blocking the transmission of COVID-19 in public transit services; (3) Public transportation information system (PTIS) plays a key role in collecting and processing data. It is critical to incorporate the crowdsourced COVID-19 data and the real-time information of transit services to the PTIS in the battle of fighting against the transmission of COVID-19 in public transportation system.

This study also has some limitations. The principles and procedures of private data collection, storage, and usage are important topics in crowdsourcing applications. However, due to the scope of the work, this study has not addressed the relevant legislation information, copyright, and data ownership, etc.

Validation and quality assessment of crowdsourced information has also raised up a lot of attentions because of rampant misinformation about public transportation in the COVID-19 era. It is important for public transportation agencies to balance risk and reward of crowdsourced information through a trust model, which incorporates the credibility about the contributors of crowdsourced data, the assessment of spatial parameters, the indicators of temporal quality, semantic quality, and data integrity, etc. (Harris et al. 2016). We plan to explore the validation and assessment of crowdsourced data in our future studies.

References

- Afzalan, N., & Sanchez, T. W. (2017). Testing the use of crowdsourced information: Case study of bike-share infrastructure planning in Cincinnati, Ohio. *Urban Planning*, 2(3), 33–44.
- Almagro, M., & Orane-Hutchinson, A. (2020). *The determinants of the differential exposure to COVID-19 in New York City and their evolution over time*. New York University Working Paper.
- American Public Transportation Association (APTA). (2015). *Public transportation investment background data* (11th ed.). Retrieved from <https://www.apta.com/research-technical-resources/transit-statistics/funding-data/>.
- American Public Transportation Association (APTA). (2020). *Public transportation ridership report*. Retrieved from <https://www.apta.com/wp-content/uploads/2020-Q2-Ridership-APTA.pdf>.
- Anderson, D. P., Cobb, J., et al. (2002). An experiment in public-resource computing. *Communications of the ACM*, 45(11), 56–61.
- Argente, D. O., Hsieh, C.-T., & Lee, M. (2020). *The cost of privacy: Welfare effect of the disclosure of COVID-19 cases*. National Bureau of Economic Research Working Paper 27220.
- Becker, D., & Bendett, S. (2015). Crowdsourcing solutions for disaster response: Examples and lessons for The US Government. *Procedia Engineering*, 107, 27–33.
- Bengtsson, L., Lu, X., Thorson, A., Garfield, R., & von Schreeb, J. (2011). Improved response to disasters and outbreaks by tracking population movements with mobile phone network data: A post-earthquake geospatial study in Haiti. *PLoS Medicine*, 8(8), e1001083. <https://doi.org/10.1371/journal.pmed.1001083>.
- Born, B., Dietrich, A. M., & Müller, G. J. (2020). *Do lockdowns work? A counterfactual for Sweden*. CEPR Discussion Paper No. DP14744.
- Boulos, M. N. K., & Geraghty, E. M. (2020). Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and

- associated events around the world: How 21st century GIS technologies are supporting the global fight against outbreaks and epidemics. *International Journal of Health Geographics* 11;19(1), 8.
- Boulos, M. N. K., Resch, B., Crowley, D. N., Breslin, J. G., Sohn, G., Burtner, R., et al. (2011). Crowdsourcing, citizen sensing and sensor web technologies for public and environmental health surveillance and crisis management: Trends, OGC standards and application examples. *International Journal of Health Geographics*, 10, 67.
- Brabham, D. C. (2010). *Crowdsourcing as a model for problem solving: Leveraging the collective intelligence of online communities for public good*. Doctoral dissertation, the University of Utah.
- Brabham, D. C. (2013). *Crowdsourcing*. Cambridge: MIT Press.
- Brabham, D. C., Sanchez, T. W., & Bartholomew, K. (2010). Crowdsourcing public participation in transit planning: Preliminary results from the next stop design case. In *TRB 89th Annual Meeting Compendium*.
- Chodera, J., Lee, A. A., London, N., et al. (2020). Crowdsourcing drug discovery for pandemics. *Nature Chemistry*, 12, 581. <https://doi.org/10.1038/s41557-020-0496-2>.
- Chunara, R., Freifeld, C. C., & Brownstein, J. S. (2012). New technologies for reporting real-time emergent infections. *Parasitology*, 1(1), 1–9.
- Chunara, R., Smolinski, M. S., & Brownstein, J. S. (2013). Why we need crowdsourced data in infectious disease surveillance. *Current Infectious Disease Reports*, 15(4), 316–319. <https://doi.org/10.1007/s11908-013-0341-5>.
- Cook, S., Conrad, C., Fowlkes, A. L., & Mohebbi, M. H. (2011). Assessing Google flu trends performance in the United States during the 2009 influenza virus A (H1N1) pandemic. *PLoS ONE*, 6(8), e23610. <https://doi.org/10.1371/journal.pone.002361>.
- Cutler, K. M. (2014). *As a cohort of bus startups emerge, chariot looks to source new routes through crowdfunding*. TechCrunch.
- Desai, A., Warner, J., Kuderer, N., et al. (2020). Crowdsourcing a crisis response for COVID-19 in oncology. *Nature Cancer*, 1, 473–476. <https://doi.org/10.1038/s43018-020-0065-z>.
- Desmet, K., & Wacziarg, R. (2020). *Understanding spatial variation in COVID-19 across the United States*. National Bureau of Economic Research Working Paper 27329.
- Digital Science. (2019). *The State of Open Data 2019*, Digital Science. Retrieved from https://digitalscience.figshare.com/articles/The_State_of_Open_Data_Report_2019/9980783.
- Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infectious Diseases*. Published online Feb 19. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).
- Fenichel, E., Castillo-Chávez, C., Ceddia, G., Chowell, G., Parra, P., Hickling, G., et al. (2011). Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 6306–6311.
- FHWA. (2002). *2002 status of the nation's highways, bridges, and transit: Conditions & performance, Chap. 14, The importance of public transportation*. Report to Congress.
- Freeman, E. E., McMahon, D. E., Fitzgerald, M. E., Fox, L. P., Rosenbach, M., Takeshita, J., et al. (2020). The American Academy of Dermatology COVID-19 registry: Crowdsourcing dermatology in the age of COVID-19. *Journal of the American Academy of Dermatology*, 83(2), 509–510.
- Gao, G., Ma, L., Wang, J., et al. (2013). Crowdsourcing compatible disaster emergency management system. *Computer System Applications*, 22(11), 31–36.
- Ginsberg, J., Mohebbi, M., Patel, R., et al. (2020). Detecting influenza epidemics using search engine query data. *Nature*, 457, 1012–1014. <https://doi.org/10.1038/nature07634>.
- Glaeser, E. L., Gorback, C., & Redding, S. J. (2020). *How much does COVID-19 increase with mobility: Evidence from New York and four other U.S. cities*. NBER Working Paper 27519, Cambridge, MA.
- Goodchild, M. F. (2007). Citizens as sensors: The world of volunteered geography. *Geo Journal*, 69(4), 211–221.
- Goodchild, M. F., & Glennon, J. A. (2010). Crowdsourcing geographic information for disaster response: A research frontier. *International Journal of Digital Earth*, 3(3), 231–241.

- Graehler, M., Mucci, R. A., & Erhardt, G. D. (2019, January) Understanding the recent transit ridership decline in major US cities: Service cuts or emerging modes. In *Transportation Research Board 98th Annual Meeting, Washington, DC*.
- Griffin, G. P., & Sener, I. N. (2016). Public transit equity analysis at metropolitan and local scales: A focus on nine large cities in the US. *Journal of public transportation*, 19(4), 126–143. <https://doi.org/10.5038/2375-0901.19.4.8>.
- Hall, J. D., Palsson, C., & Price, J. (2018). Is Uber a substitute or complement for public transit? *Journal of Urban Economics*, 108, 36–50.
- Harford, J. D. (2006). Congestion, pollution, and benefit-to-cost ratios of US public transit systems. *Transportation Research Part D: Transport and Environment*, 11(1), 45–58.
- Harris, J. E. (2020). *The subways seeded the massive coronavirus epidemic in New York City*. National Bureau of Economic Research Working Paper 27021.
- Harris, D., Smith, D., O'neil, C., & Severinsen, J. (2016). *The role of real-time crowdsourced information and technology in supporting traveler information and network efficiency*. New Zealand Transport Agency research report 593.
- Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine*, 14(6), 1–4.
- Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business*. New York: McGraw-Hill.
- Hu, et al. (2020). The risk of COVID-19 transmission in train passengers: An epidemiological and modelling study. *Clinical Infectious Diseases*, 1057. <https://doi.org/10.1093/cid/ciaa1057>.
- Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., et al. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*, 395(10223), 497–506.
- Knittel, C. R., & Ozaltun, B. (2020). *What does and does not correlate with COVID-19 death rates*. NBER Working Paper No. 27391, Cambridge, MA.
- Lakhani, K. R., Boudreau, K. J., et al. (2013). Prize-based contests can provide solutions to computational biology problems. *Nature Biotechnology*, 31(2), 108–111.
- Lazo, L. (2017). *Start-ups find footing with crowdsourced bus service in cities with ailing transit*. The Washington Post.
- Leung, G. M., & Leung, K. (2020). Crowdsourcing data to mitigate epidemics. *The Lancet Digital Health*, 2(4), e156–e157.
- Liu, Y., & Liu, Y. (2016). Theoretical research on crowdsourcing crisis management-new directions for social public crisis management. *Exploration*, 2, 100–115.
- Meymaris, K., Henderson, S. et al. (2008). *Project BudBurst: Citizen science for all seasons*.
- National Research Council (NRC). (2007). *Successful response starts with a map: Improving geospatial support for disaster management*. Washington, DC: National Academies Press.
- NIST. (2020). *Research data framework (RDaF)*. National Institute of Standards and Technology. Retrieved from https://www.nist.gov/system/files/documents/2020/12/01/Preliminary%20RaF%20final_12-01-2020.pdf.
- Oum, T. H., & Wang, K. (2020). Socially optimal lockdown and travel restrictions for fighting communicable virus including COVID-19. *Transport Policy*, 96, 94–100.
- Peters, A. (2019). *This transit app now can figure out your route even if you have no service*. Fast Company.
- Qian, H., Miao, T., Liu, L., et al. (2020). *Indoor transmission of SARS-CoV-2*. Retrieved from <https://www.medrxiv.org/content/10.1101/2020.04.04.20053058v1.full.pdf>.
- Schwartz, S. (2020). *Public transit and COVID-19 pandemic: Global research and best practices*. Report prepared for the American Public Transportation Association.
- Steinfeld, A., Zimmerman, J., Tomasic, A., Yoo, D., & Aziz, R. D. (2011). Mobile transit rider information via universal design and crowdsourcing. In *Transportation Research Board 90th Annual Meeting*, Washington, DC, USA (p. 16).
- Sui, D., Elwood, S., & Goodchild, M. F. (2013). Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice. *The Canadian Geographer*, 59(1), e19–e20.
- Suzdaltsev, J. (2014). *Crowdsourced bus lines: A viable MUNI alternative?* The Bold Italic.

- Tomer, A., Kneebone, E., Puentes, R., & Berube, A. (2011). *Missed opportunity: Transit and jobs in metropolitan America*.
- Yoo, D., Zimmerman, J., Steinfeld, A., & Tomasic, A. (2010). *Understanding the space for co-design in riders' interactions with a transit service*. Presented at the Proceedings of the 28th International Conference on Human Factors in Computing Systems, Atlanta, Georgia, USA.

Chapter 18

Outlook and Next Steps: Understanding Human Dynamics in a Post-pandemic World—Beyond Mapping COVID-19 in Space and Time



Daniel Sui and Shih-Lung Shaw

Ever since the SARS-CoV-2 virus was first reported in late 2019, there has been an explosive growth in the use of mapping and geospatial analytical tools by researchers, journalists, and the public at large to communicate and analyze the COVID-19 global pandemic (Sui 2020). Chapters presented in this edited volume by a group of international interdisciplinary scholars report some major exciting advances in the breadth and depth of mapping the pandemic's impact in space and time. Like all other scientific and scholarly work, findings reported in this book are works in progress and may serve as stepping stones for others to continue the exploration in the future.

In this final chapter, we summarize the major findings and conclusions from the previous 17 chapters and identify some of the major paradoxes and complexities revealed from the growing literature on mapping COVID-19. We then outline the next steps and discuss major areas of future research before we wrap up this volume with our concluding thoughts.

18.1 Summary

Using a plethora of theoretical perspectives, data sources, and analytical/modeling methods, the previous 17 chapters in this volume present a snapshot of advances in interdisciplinary research towards a better understanding of the spatial and temporal dynamics of the COVID-19 global pandemic. At the time of the completion of the

D. Sui (✉)

Geography and Public & International Affairs, Burruss Hall, Suite 340, 800 Drillfield Drive
Virginia Tech, Blacksburg, VA 24061, USA

e-mail: dsui20@vt.edu

S.-L. Shaw

Department of Geography, University of Tennessee, Knoxville, TN 37996-0925, USA
e-mail: sshaw@utk.edu

manuscripts for this book (February 2021), the global trends in COVID-19 mortality and morbidity were still grim despite multiple countries working around the clock to vaccinate their citizens. Needless to say, we may well see some new trends and patterns by the time this book is in print because of the rapidly evolving situation. Nonetheless, we believe that the findings reported in this volume represent a significant contribution to the growing literature on mapping COVID-19 in space and time. Some of the themes and conclusions are worth highlighting, because they will help us better understand the new human dynamics in a post-pandemic world and can lay the foundation for future research on understanding global pandemics in the coming years.

- (1) The COVID-19 pandemic is a fatal threat to not only biological health, but also to mental health, emotional well-being, and economic stability. These effects were distributed unequally, with those at the lower social and economic ladder rungs suffering the most. Mapping the crisis-upon-crisis caused by COVID-19 requires a holistic framework that integrates the social, economic, political, and emotional aspects of the population concerned, ideally linked to an integrated, broadly defined framework on vulnerability and resiliency.
- (2) The spread of COVID-19 follows the power law (more popularly known as the 80–20 rule or Pareto principle), which predicts that—like population—the vast majority of cases are in a small fraction of communities. Because the spatial distribution of cases so closely resembles the underlying spatial distribution of populations, there may be questions about the effectiveness of lockdown and social distance measures. On the other hand, the observation that there are not more significant mini-outbreaks of COVID-19 (the ht-index of cases does not exceed that of population) implies that lockdown and social distance measures do indeed have some effect. Future work is needed to tease out more geographic as well as policy implications of the power/scaling law of COVID-19.
- (3) The so called 4T (testing, tracking, tracing, and treatment) models proved to be quite effective as reflected in the very low infection and death rates in East Asia despite extremely high population density. These measures are implemented less effectively (or not at all) in the Western world in general—and the U.S. in particular—due to privacy laws and other ethical concerns. This book proposes a useful framework to mitigate these privacy concerns by using ethical and privacy principles of geospatial data handling. Similar to the policy changes in the U.S. following the 9/11 terrorist attacks, this framework aims to strike a delicate balance between privacy protection and the effectiveness of contact tracing and tracking in real-time.
- (4) Methodologically, we are also pleased to see the new advances of spatial and temporal analysis at a much finer granularity using mobile phone data, shortest path at the street level, and space–time kernel density (STKDE) techniques using the CyberGIS framework. Advances in improved methods and analysis of human mobility patterns will certainly provide more credible information of the virus spread in real time. Further, a virtual crypto place can be established by deploying blockchain technology. Not merely serving as a virtual memorial

for one of the earlier whistleblowers for COVID-19 in China, this work has far broader conceptual and methodological implications than what was discussed in the chapter by Zhao and Huang. We hope this pioneering work will stimulate further discussions on using this decentralized approach for improving data quality, privacy protection, and resistance of surveillance by authorities.

- (5) Despite the repeated warnings by experts and luminaries, most countries are woefully inadequately prepared to deal with global pandemics like COVID-19. Multiple chapters also address a broad range of policy issues and public health crisis preparedness. Instead of pursuing relentless growth, more than ever we need to focus on resiliency and security. Rather than relying on a one-size-fits-all approach, local responses should be more elastic to reflect both the national policy and local circumstances, as demonstrated by the epidemic resource allocation (EPI-RA) framework in this book. Furthermore, citizen mobility by air and on the ground is closely linked to how fast the virus can spread, crowdsourcing approach (as demonstrated by <https://howwefeel.org>) can be used to collect data under NIST's released research data framework (RDaf) (<https://www.nist.gov/programs-projects/research-data-framework-rdaf>).

18.2 Paradoxes and Complexities of Human Dynamics During and After the Pandemic

Few people paid serious attention to the news in early 2020 when the Chinese government announced the lockdown of the city of Wuhan on January 23, 2020, where SARS-COVID-2 virus was first discovered in late December 2019. Even through February, when countries around the world, including South Korea and Italy, began implementing massive shutdowns, most countries considered the pandemic a far-off threat. However, by March 11, 2020 when the World Health Organization determined that COVID-19 was a global pandemic, it began to upend the world at the global level with such a magnitude and speed, many countries and their citizens were caught off-guard. The global economy suddenly came to a virtual halt as more and more people across the globe were ordered by their government to isolate or self-quarantine. Within weeks, people of different ages quickly became experts of Zoom or on-line meeting platforms. Students and teachers moved on-line for virtual classes. Patients saw their doctors and therapists using telemedicine. Companies changed policies that would normally have taken them years to revise. Cities turned avenues into pedestrian walkways and sidewalks into cafes. Governments opened their coffers in ways that were once unimaginable and could lead to much greater willingness to invest in the future.

Compared to past pandemics in history, human dynamics during the COVID-19 have exhibited both striking similarities and some fundamental differences (Poos 2020). Governments in many countries downplayed the risks in the beginning, with some politicians even vehemently denied the pandemic despite the overwhelming mortality and morbidity data. The public resented stay-home or shelter-in-place

orders, especially without a long-term plan for how to re-open. Some groups continued to challenge the effectiveness of wearing face coverings, social distancing measures, and washing hands and surfaces. Racism flourished, with minority groups singled out as scapegoats for the spread of the virus. All of these led to unnecessary loss of life.

However, we must be mindful that we are living in a different world today than 100 years ago or even the more distant past, especially considering the new technological environment we are living in right now. While decades of studies have revealed that all the major issues and challenges (e.g. climate change, pandemics, cyber security) we face today are global in nature, and yet our reactions and policy prescriptions tend to have a growing nationalist focus. This came into sharp relief in the global fight over the distribution of personal protective equipment (PPE) supplies in summer 2020 and growing vaccine nationalism in 2021 (Bieber 2020; Bollyky and Bown 2020). Barriers have also been erected to slow down global collaboration and data sharing in COVID-19 research, consistent with the global trend in the rise of borders and deglobalization during the past five years.

Under the backdrop of this broader paradoxical context, we are experiencing multiple other conflicting trends related to more specific aspects of this pandemic:

- (1) Technology has enabled the explosive spread of both truth and fiction. The pace of scientific and technological breakthroughs in understanding and combating this virus (such as the speedy sequencing of the SARS-COVID-2 virus, unprecedented speed of developing new vaccines) has accelerated. At the same time, we see the meteoric rise of conspiracy theories among various population groups globally. This explosion of lies has further contributed to the denigration and diminishing role of professionals and experts in addressing pressing crises (Nichols 2020), which is indeed very disturbing, disruptive, and harmful along multiple fronts. Concomitantly, citizens across the globe have also declining trust in government (Devine et al. 2020).
- (2) Technology has vastly improved capabilities to collect and process a swelling ocean of big data. Scientific and scholarly publications have exploded, and yet significant percent of scientific and scholarly publications can't be reproduced/replicated—in some fields at rates of almost 70% (Sui and Kedron 2020). This further undermines the respect for science, making truth indistinguishable from the tsunami of misinformation, alternative facts, and blatant lies.
- (3) Despite the proliferation of ways to digitally connect, people are more divided than ever. As more people are connected by the Internet and various platforms of social media, there is an unprecedented level of division and polarization with growing hostilities among different groups in society from local to global levels (Klein 2020). Similarly, economic disparities are larger than ever due to the twin crisis of the pandemic and the crippled economy. In the U.S., the wealth of the top 1% of Americans has grown to \$1.5 trillion within the one-year period since the pandemic began, equivalent of 15 times of the wealth of the bottom 50% of the population (Beer 2020). The preliminary mortality and morbidity data of COVID-19 also lay bare the astonishing disparity among different

racial/ethnic, socioeconomic, education, age, and gender groups. Minorities suffer disproportionately the mortality and morbidity rate. Some have suggested that the effectiveness of the 4 T model during the pandemic is evidence for the benefits of authoritarianism as governments suspend or reduce democratic freedoms and civil liberties.

We now should come to the full grip of the true meaning of McLuhan's (1962) vision of a "global village." On the one hand, we are now connected globally by media/communication technologies through collapsing space and time. On the other hand, we are simultaneously becoming more like village dwellers with growing tribalism in our thinking and behavior. While the interactions and connections at the global scale have enhanced significantly due to modern transportation and communication technologies, opinions and thoughts among different countries, regions, and locales are becoming more polarized. For example, we now are well connected via the Internet, which enables us to access all kinds of information, but we tend to read certain news and information more than before due to the specific links pushed to us based on our reading history analyzed by artificial intelligence using big data analytics. Our thinking therefore becomes more polarized even though the Internet offers a huge amount of diverse information—a phenomenon known by behavioral economists as confirmation bias (Thaler 2015). In other words, we may be closer to each other in relative and relational space, but we also become farther apart in mental space (Shaw and Sui 2019). The dynamics of COVID-19 pandemic therefore should not be understood and studied only from the traditional concept of physical space such as social distancing and human mobility. Instead, it is critical to take a holistic framework that integrates the social, economic, political, and mental aspects of the population concerned.

- (4) While division leaves society less equipped to face global challenges, the dual threats of cyber security and growing environmental vulnerability are reaching tipping points. The pandemic has suddenly merged the function of home, school, and workplace, which isn't just a nightmare for parents, but also a nightmare from a cyber security perspective since much of these activities use less secure home computers and connections, giving bad actors more vulnerabilities to exploit. We have observed a spike in phishing attacks, malspams, and ransomware attacks as attackers are using COVID-19 as baits to impersonate brands thereby misleading employees and customers. The most recent SolarWinds attack serves as another potent reminder that malicious actors can capitalize on this centralized/decentralized structure to exploit even some of the most secure government servers.

With the continued growth of the Internet of Things and the rapid global expansion of 5G network coverage, different kinds of sensors, computers, cameras, and other devices are all connected in cyber space. Yet globally, there is no broadly accepted industry standard for ensuring low-cost, commodity devices are secure and regularly updated to protect against hacking and attacks by state actors and lone wolves.

The vulnerabilities in cyber space are matched by vulnerabilities in physical environment in the broader context of global climate change. Starting with catastrophic forest fire in Australia, disastrous hurricanes in North America, and devastating droughts in Africa and South America, the year 2020 turned out to be one of the hottest years and also one of the most turbulent years in terms of climate change in human history. Indeed, both cyber security and environmental vulnerability have reached a tipping point, moving forward we must take a holistic view and address this dual challenge effectively and continue to improve the infrastructure to get ourselves prepared for the potential catastrophes in both the virtual and physical worlds, which are intricately linked with one other.

We are indeed living in a pivotal moment in human history. To address any of these paradoxes or challenges effectively, out of the box, united thinking is required. The scientific community can certainly play its role, and none of these issues can be addressed by the research community alone, but we must spare no efforts to get our voice heard.

18.3 Outlook and Next Steps: Beyond Mapping COVID-19 in Space and Time

We need robust and reliable science and technologies more than ever to address these paradoxical and conflicting challenges, but we must emphasize that although science and technologies are necessary, they are insufficient to address these challenges we face effectively if we do not rethink the human dynamics at the fundamental level and undergo fundamental social, economic, political, and cultural changes. To facilitate these fundamental changes, we need to push for a human-centered research and innovation agenda with the development goal for a healthy, prosperous, equitable, and sustainable society to better meet human needs. With this goal in mind, we outline the following four general areas for future studies of human dynamics in a post-pandemic world.

1. Promoting one-health towards whole health: Towards a new convergence

According to the One Health Initiative Task Force (OHITF) (https://www.avma.org/sites/default/files/resources/onehealth_final.pdf), one health is defined as the collaborative efforts of multiple disciplines working locally, nationally, and globally, to attain optimal health for people, animals and our environment. The ultimate goal of one health research is to improve human health. The emerging paradigm of the whole health initiative is to move beyond focusing on human biological health, which often leads to focus predominantly on diseases. The concept of whole health centers around a holistic integration of human biological, mental, and spiritual health, and always starts by posing the question on what matters the most for each individual first. Whole Health empowers and equips people to take charge of their physical, mental and spiritual well-being and live their lives to the fullest.

Although there are still many things we do not know about the SARS-CoV-2 and its various mutations (Christaki 2020), what is certain is that it cannot be addressed as a single threat to health. This pandemic redoubles what previous pandemics illustrated—human health is intimately connected to biological pathogens in the animal world. To improve human health, we must continue to learn how to live in harmony with other species and environment from local to global level. Furthermore, as several chapters in this book have reported, the pandemic's toll on people's mental health and emotional well-being is as severe, if not more so, as its impacts on people's biological health.

We strongly believe that the current global pandemic provides us a golden opportunity to promote one health towards whole health. The convergence approach (<https://www.convergencerevolution.net>) can be used as a catalyst to link scholars in multiple disciplines to conduct transdisciplinary research under the umbrella concept of one health and whole health that transcends the concept of absolute space and also considers the concepts of relative space, relational space, and mental space. As demonstrated in other groundbreaking convergence health research, there are great potentials for academic researchers to partner with healthcare industry and health policy makers to achieve stakeholder synergy—a defining characteristics in convergence research.

2. Harnessing technological advances for a more equitable and just future

The COVID-19 global pandemic occurred in the midst of the so-called 4th industrial revolution, led by advances in artificial intelligence, block chain, cloud and quantum computing, and big data analytics. Even before the pandemic, scholars have warned the new class bifurcation during this post-industrial era—the emerging creative class (who are leading the inventions and innovations for the digital economy) (Florida 2014) vs. the useless class (who are not only unemployed or underemployed but also essentially unemployable) (Harari 2020). Multiple government and industry-sponsored reports had already predicted the future of work will be disrupted heavily by advances in robotics, AI, and machine learning. More than half of the jobs in certain industries will be eliminated by 2030 (McGowan and Shipley 2020; Susskind 2020).

The COVID-19 pandemic has accelerated the adoption of cutting-edge technologies in terms how we work, how we educate our children, and how we interact with one other. How do we harness all the wonderful technological advances for a more equitable and just future for all? Especially in the context of the future of work, what does it mean to work in the age of pervasive robotics? How can we explore innovative ways to make people still relevant in the age of smart machines so that we can all live a dignified life with purpose and meaning—whatever these purposes and meaning may be defined according to individual circumstances. Instead of making humans serve the needs of machines, we need to be vigilant and always put human needs first, perhaps with Maslow's hierarchy of human needs as a blue print. Should everyone be given a universal basic income? How should we adjust the current education system and become more effective to promote life-long learning? Without addressing these

fundamental questions, we are likely to perpetuate the current trends of ever widening gap between the haves and have-nots, which are the root of growing polarization and increasing violence.

3. Starting a global data reckoning to combat misinformation

Shortly after the first decade of the twenty-first century, the private sector controls more data about citizens' activities and what is going on in the society than the government. In the U.S., we know now that tech giants like Google, Facebook, Amazon, Twitter, Apple, and Microsoft control more data than the U.S. Census Bureau or the Bureau of Labor Statistics. Likewise, Baidu, Tencent, and Alibaba (which are China's counterparts of Google, Facebook, and Amazon) control more data about Chinese citizens than most the state apparatus combined. Data have been called the new oil in the digital age. Because of its importance, there have been growing efforts by many countries in the world to promote data protectionism, which is, unfortunately, in many ways similar to the vaccine protectionism. Recent data protectionism has been moved to efforts to shut down platforms (e.g. China's efforts to block Google and Facebook in China; the U.S. efforts to shut down WeChat and Tik Tok in the U.S. etc.) entirely to prevent these tech giants to collect data in the first place. The danger of data protectionism, an integral part of global trade war, can have serious implications for digital global economy and the global research community.

A recent study by EuroAsia Group (<https://www.eurasia-group.net/issues/top-risks-2021>) concluded that the disturbing trend of data protectionism is so alarming that it has called for a global data reckoning by industry leaders as well as citizens alike. This global data reckoning couldn't come at a better time as we are trying to address so many global challenges. Without a steady flow of quality data at the global level, the basic function of global society—from global commerce and research collaboration to personal interactions and educational efforts—could be in jeopardy. Needless to say, progress on global challenges brought by the COVID-19 global pandemic could be delayed, and even derailed.

Moving forward, we need to do everything we possibly can to reverse the disturbing trend of data protectionism. More research should focus on developing meaningful national and international data framework to facilitate meaningful data flow. NIST's framework of research data (<https://www.nist.gov/programs-projects/research-data-framework-rdaf>) and the effort led by Harvard on university-industry data partnership (King and Persily 2019) are excellent starting points for our global data reckoning effort.

Equally important, the global data reckoning should also entail concerted efforts to combat misinformation. The spread of misinformation in recent years has gone viral, having become a parallel epidemic in cyberspace (aka infodemic) to the pandemic in the physical world. Governments and industries have made efforts to combat misinformation such as promoting fact checking websites, adding warning labels to misleading statements, and shutting down accounts of super spreaders of misinformation. It is an imperative for the research community to set an aggressive research

agenda to join force with industry partners and government agencies to mount an effective offense to reduce and eventually root out misinformation.

This misinformation spread is proliferated through visual media, which have been shown to be more convincing than words alone. Maps are no different, and can be used to spread lies (Monmonier 2018). While the broader use of web-based maps has become an effective way to educate the general public about different aspects of the spread of SARS-CoV-2, preliminary studies have already shown that various on-line mapping platforms and dashboards have been a source of misinformation during COVID-19 (Mooney and Juhasz 2020). Web maps have been widely misused for delivering public information on this fast moving, epidemiologically complex, and geographically unbounded process. For example, the inconsistent use of spatial aggregation across countries gives a false impression of the virus' spread (e.g., HealthMap: <https://www.healthmap.org/covid-19/>). We need more rigorous research to detect and discard geospatial misinformation. Just as importantly, we must educate the general public on how to interpret geospatial information and visualization accurately and intelligently.

4. Bolstering environmental security holistically: Towards an O2O perspective

To address the dual challenges of cyber security and a growing vulnerability of the physical environment, we need to take a holistic view on environmental security by linking the on-line with the off-line world (O2O). In the digital marketing literature, O2O has a rather narrow definition, describing how to design on-line systems to entice consumers within a digital environment so that they can make purchases of goods or services from physical businesses (off-line). As illustrated in a recent spatial framework (Shaw and Sui 2019), the on-line and off-line worlds are increasingly connected in multiple complex ways as more and more human activities and interactions are conducted on-line. Moving forward, an O2O perspective will help us bolster environmental (cyber as well as physical) security holistically.

The anthropogenic and highly decentralized nature of both cyber security and climate threats further illustrates the benefits of the O2O perspective to translate scholarship from one environment to the other. The polycentric governance structure developed by Ostrom (2012) for global environmental governance has profound implications for addressing the threats of cyber security. Preliminary work has shown the greater relevance of one to another (Cassotta et al. 2020). Geopolitically, recent studies have shown that efforts by both governments and industry to develop global rules for state behavior in cyberspace have proven futile and ineffective (EuroAsia Group 2021). Targeted economic sanctions and “naming and shaming” have also not deterred the bad behaviors by state actors in cyber space. The polycentric framework, developed for global environmental governance, shows promises for addressing cyber security issues where these other approaches have failed.

Another connection between cyber and climate threats is related to one silver lining of the COVID-19 pandemic: significant positive environmental impacts because of

the dramatic reduction in material and energy consumption triggered by the COVID-19 lockdown and enabled by a greater reliance on digital connection rather than physical interaction. These unintended positive environmental consequences of COVID-19 illustrated two points loud and clear, especially to all the climate change skeptics and deniers: (1). Humans can survive without the level of material and energy consumption during the pre-COVID years; (2). We are major contributors to the deterioration of the fragile global environment and we can improve the global environment if we have the political will to change our behavior in terms how we work and live. Of course, we need a more secure cyber on-line environment to facilitate the flow of quality information to ensure the full function of the digital economy and all the physical infrastructure that supports the economy and the off-line physical environment. Hopefully, we draw some essential elements from how we work during the pandemic to develop a blueprint for the best practices for how we work in the future.

18.4 Concluding Remarks: New Consilience and Solidarity in the Making

Sometimes, the greatest realist is an idealist (Zakaria 2020).

In a real sense all life is inter-related. All men are caught in an inescapable network of mutuality, tied in a single garment of destiny. Whatever affects one directly, affects all indirectly. I can never be what I ought to be until you are what you ought to be, and you can never be what you ought to be until I am what I ought to be... This is the inter-related structure of reality.

—Martin Luther King Jr., Letter from Birmingham Jail

Consistent with geography's first law, which emphasizes everything is related to everything else, it is indeed inspiring to reread Dr. Martin Luther King's letter from the Birmingham jail today when we are contemplating how we can work together to address all the global challenges we face along multiple fronts. Keenly recognizing the "inescapable network of mutuality" and "interrelatedness of all communities" should be the number one guiding principle when we are exploring our ways out of this dark moment in human history.

Moving forward, we not only need to assume the ontological position about the interconnectedness of nature and consilience of human knowledge (Wilson 1999) at the philosophical level but also pursue convergence research aggressively at the operational level with a transdisciplinary approach and stakeholder synergy. We should also aspire to set our goal for a healthy, efficient, equitable, and sustainable world from local to global level through human-centered innovation. Anything short of this high ideal will not serve us well in the long run, and will, in fact, create more problems to come back to haunt us down the road, as have demonstrated time and again throughout human history. To innovate (with a focus on meeting the hierarchy of human needs) means to be free, and to be free means we must make a concerted

effort to liberate ourselves from the prisons of technology and the outdated social, economic, political, and cultural system we are in.

Acknowledgements The authors would like to gratefully acknowledge (with implicating)—Dr. Hui Kong provided useful references for this chapter, and Dr. Laurel Miner read and commented an earlier draft of this chapter.

References

- Beer, T. (2020). Top 1% of U.S. households hold 15 times more wealth than bottom 50% combined. <https://www.forbes.com/sites/tommybeer/2020/10/08/top-1-of-us-households-hold-15-times-more-wealth-than-bottom-50-combined/?sh=20c429155179>.
- Bieber, F. (2020). Global nationalism in times of the COVID-19 pandemic. *Nationalities Papers*, 1–13. <https://doi.org/10.1017/nps.2020.35>
- Bollyky, T.J., & Bown, C. P. (2020). Vaccine nationalism will prolong the pandemic: A global problem calls for collective action. Retrieved January 3, 2021, from <https://www.foreignaffairs.com/articles/world/2020-12-29/vaccine-nationalism-will-prolong-pandemic>.
- Cassotta, S., Sidortsov, R., Pursiainen, C., Pettersson, M., & Goodsite, M. E. (2020). Climate change, environmental threats and cybersecurity in the European High North. Retrieved January 8, 2021, from <https://isdp.eu/content/uploads/2020/01/Climate-Change-Environmental-Threats-and-Cybersecurity-in-the-European-High-North.pdf>.
- Christakis, N. A. (2020). *Apollo's arrow: The profound and enduring impact of coronavirus on the way we live*. New York, NY.: Little, Brown Spark.
- Devine, D., Gaskell, J., & Jennings, W. (2020). Trust and the coronavirus pandemic: What are the consequences of and for trust? An early review of the literature. *Political Studies Review*. Retrieved January 24, 2021.
- EuraAsia Group. (2021). Top risks 2021. Retrieved January 17, 2021, from <https://www.eurasia-group.net/issues/top-risks-2021>.
- Florida, R. (2014). *The rise of the creative class*. New York, N.Y.: Basic books.
- Harari, Y. N. (2020). *21 lessons for the 21st century*. New York, NY: Random House.
- King, G., & Persily, N. (2019). A new model for industry-academic partnerships. *PS: Political Science and Politics*. Publisher's Version Copy at <https://j.mp/2q1IQpH>.
- Klein, E. (2020). *Why we're polarized*. New York, N.Y.: Simon & Schuster.
- McGowan, H. E., & Shipley, C. (2020). *The adaptation advantage: Let go, learn fast, and thrive in the future of work*. New York, NY.: Wiley.
- McLuhan, M. (1962). *The Gutenberg galaxy: The making of typographic man*. Toronto: University of Toronto Press.
- Monmonier, M. (2018). *How to lie with maps* (3rd ed.). Chicago: University of Chicago Press.
- Mooney, P., & Juhasz, L. (2020). Mapping COVID-19: How web-based maps contribute to the infodemic. *Dialogues in Human Geography*, 10(2), 265–270.
- Nichols, T. (2020). *The death of expertise: The campaign against established knowledge and why it matters*. New York, NY.: Oxford University Press.
- Ostrom, E. (2012). Polycentric systems: Multilevel governance involving a diversity of organizations. In *Global environmental commons: Analytical and political challenges in building governance mechanisms*.
- Poos, L. R. (2020). Lessons from past pandemics: Disinformation, scapegoating, and social distancing. <https://www.brookings.edu/blog/techtank/2020/03/16/lessons-from-past-pandemics-disinformation-scapegoating-and-social-distancing/>.

- Shaw, S. L., & Sui, D. Z. (2019). Understanding the new human dynamics in smart spaces and places: Towards a spatial framework. *Annals of AAG*, 110(2), 339–348. <https://www.tandfonline.com/>.
<https://doi.org/10.1080/24694452.2019.1631145>.
- Shaw, S.-L., & Sui, D. Z. (Eds.). (2018). *Human dynamics research in smart and connected communities*. Cham, Switzerland: Springer International Publishing AG.
- Sui, D. Z. (2020). Mapping pandemics: COVID-19 and beyond. Retrieved January 23, 2021, from <https://scholarworks.uark.edu/hnrclepc/3/>.
- Sui, D. Z., & Kedron, P. (2020). Reproducibility and replicability in the context of contested identities of geography. *Annals of the AAG*. <https://doi.org/10.1080/24694452.2020.1806024>.
- Susskind, D. (2020). *A world without work: Technology, automation, and how we should respond*. New York, NY: H. Holt & Co.
- Thaler, R. H. (2015). *Misbehaving: The making of behavioral economics*. New York: W. W. Norton & Company.
- Wilson, E. O. (1999). *Consilience: The unity of knowledge*. New York, N.Y.: Vintage Books.
- Zakaria, F. (2020). *Ten lessons for a post-pandemic world*. New York, NY: W.W. Norton & Co.