Does Where You are Matter? A Visual Analytics System for COVID-19 Transmission Based on Social Hierarchical Perspective

Jianing Hao*
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
jhao768@connect.hkust-gz.edu.cn

Xibin Jiang*
The Hong Kong University of Science and Technology
Hong Kong, China
xjiangax@connect.ust.hk

Qing Shi
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
brantqshi@ust.hk

Wei Zeng
The Hong Kong University of Science and Technology (Guangzhou)
Guangzhou, China
weizeng@ust.hk

ABSTRACT

The COVID-19 pandemic has emerged as a global health crisis necessitating multidisciplinary efforts to address its profound social and economic repercussions. Big cities like Hong Kong face significant challenges in controlling the spread of pandemic for the higher COVID-19 infection and death rates. In densely populated urban areas, due to high heterogeneity in socioeconomic backgrounds, residents are more likely to reside in different social and physical environments, which may lead to a high discrepancy of individual's COVID-19 infection risk. Combining social hierarchical perspectives and a Multiple coordinated view (MCV) visualization system, this paper aims to provide a comprehensive investigation into the spatial and temporal dynamics of COVID-19 in Hong Kong during 2020, with a focus on understanding how the living environment shapes individuals' infection risk. Our study provides three analysis levels, including case, neighborhood and district, to facilitate the exploration of spatial and temporal patterns of virus transmission. Through interactive visualization analysis and exploration, we show that individuals from deprived neighborhoods exhibit a higher likelihood of contracting coronavirus and there exists segregation of virus transmission among different social classes. Moreover, through case studies, we analyze the travel records of 8000+ confirmed cases and find there exists spatial stratification of infectious risk. Leveraging our proposed interactive visualization system, policymakers and stakeholders can make more informed decisions to effectively manage and contain the spread of infectious pandemics like COVID-19.

*Both authors contributed equally to this research.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

CCS CONCEPTS

• **Human-centered computing → Visual analytics**.

KEYWORDS

neighborhood deprivation, visual analytics, COVID-19

ACM Reference Format:
Jianing Hao*, Xibin Jiang, Qing Shi, and Wei Zeng. 2018. Does Where You are Matter? A Visual Analytics System for COVID-19 Transmission Based on Social Hierarchical Perspective. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX).* ACM, New York, NY, USA, 9 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

The emergence of the Coronavirus disease 2019 (COVID-19) in December 2019 has had a profound global impact, with a staggering 765.9 million confirmed cases and 6.92 million deaths reported worldwide[1]. Beyond the devastating loss of life, such a pandemic has inflicted significant social and financial costs for most countries, such as education closure [21], increased poverty and inequality [42], as well as enormous public financial expenditure [1].

Some studies indicate that the living environment, intervened with one's socioeconomic status, can affect his or her infection risk from COVID-19 [12, 15]. Higher concentration of confirmed cases is found among groups who live in deprived areas with fewer distribution of epidemic prevention facilities [24]. Several studies show that substantially elevated COVID-19 incidence rates in the most disadvantaged urban areas [2], and individuals from disadvantaged neighborhoods are more vulnerable to infectious diseases compared to those from neighborhoods with more favorable socioeconomic conditions [12]. Factors contributing to this vulnerability include limited access to healthcare and hygienic facilities, insufficient vaccination services [7], overcrowded living conditions [22] and etc. Higher population density has also been associated with increased COVID-19 infection risks and prolonged epidemics [34]. According to previous research, the social and physical environment greatly influences an individual's risk of COVID-19 infection, it is necessary to consider the disparity of vulnerable condition for specific group when analyzing the trend of COVID-19 infection.

[1]https://covid19.who.int/

To help manage and analyze the vast amount of infectious disease cases, visual analytics tools are of great importance. An example of such a tool is the COVID-19 dashboards developed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [13]. It provides real-time information on the number of confirmed COVID-19 cases, deaths and recoveries across countries, enabling users to comprehend and analyze the situation effectively. In addition to the overall presentation of data, some studies have focused on decision-support visualizations [18, 41, 49], which analyze various epidemic transmission scenarios and decision-making processes to enhance mitigation strategies against the impact of COVID-19. However, most existing studies primarily concentrate on general trends or decisions regarding COVID-19, often based on comprehensive and long-term observations. This approach may overlook the real-time dynamics of the situation, including virus updates and variations in infection rates among different groups. Incorporating multidisciplinary perspectives into the spatial-temporal view of COVID-19 transmission and combining rich interactive technologies can effectively help to show the dynamic changes of the pandemic and to explore strategies or solutions from epidemiology and public policy perspectives.

In this work, we present a comprehensive investigation into the spatial and temporal dynamics of COVID-19 in Hong Kong during 2020, to find out how the social and physical environment in which individuals reside influences their COVID-19 infectious risk. By developing a Multiple coordinated view (MCV) [9] visualization system, we show that individuals living in deprived neighborhoods exhibit a higher likelihood of COVID-19 infection and there exist spatial stratification of infectious risk on the analysis of the travel records of 8000+ confirmed cases. Further, by closely examining the groups and locations visited by individual cases, we observe different spread patterns of the COVID-19 among different social classes. To gain a comprehensive understanding of the transmission dynamics and offer targeted opinions on epidemic prevention, we explore and analyze the data from three levels: case, neighborhood (District Council Constituency Area, DCCA in our study) and district. Our analysis mainly focuses on the initial one-year spread of COVID-19 in 2020 because it is a proper stage to observe the transmission patterns of the virus before a sound epidemic prevention system is built for most social groups.

The major contributions of this paper are summarized as follows:

- We provide a novel social hierarchical perspective into the COVID-19 pandemic by visually examining how individuals' infection risk is influenced by their residential location. This theoretical angle contributes to the understanding of similar epidemics in the future.
- We develop a Multiple Coordinated View visualization system, enabling *spatio-temporal* and *multi-level* (i.e., case, neighborhood, and district) analysis of the coronavirus spread patterns. This system focuses on the micro-mechanisms of viral transmission among cases and sub-networks within a single city, distinguishing it from other COVID-19 visualization systems.
- Based on the analysis conducted, suggestions are provided for policymakers and specialists involved in the governance and prevention of the COVID-19 pandemic. Specifically, attention is drawn to individuals living in deprived environments, and timely prophylactic measures are proposed, particularly for densely visited areas.

2 RELATED WORKS

2.1 Health Disparities

Numerous studies have highlighted the varying infection rates of the COVID-19 virus among groups with different socioeconomic background [3, 24]. Bambra et al. [6] proved that there are inequalities in COVID-19 morbidity and mortality rates. Several studies point out that an individual's socioeconomic status plays a significant role in determining their exposure to COVID-19, as it influences their strategies and attitudes towards COVID-19 [4, 39]. Other studies indicate that people who work in a low-paid job are more likely to undertake key tasks during the COVID-19 pandemic and thus face a higher risk of being exposed to the virus [6]. Other social and economic reasons can denote the difference of COVID-19 infection risk, such as whether one lives below the poverty line, lacks in health knowledge and has higher susceptibility to the pandemic [19, 29].

Eliminating health disparities remains a crucial imperative for both public health and society [5, 45]. Concerns of racial health inequity worsening during COVID-19 were raised [26, 43], and community-level investigations demonstrated that individuals from disadvantaged socioeconomic backgrounds may face an elevated risk of COVID-19 infection [10]. Research has found the association between COVID-19 infection rates in low socioeconomic status and income among racial and ethnic minority groups [23, 37]. Specifically, studies have found a positive correlation between COVID-19 risk and Area Deprivation index (ADI) [31]. This correlation can be attributed to barriers in accessing medical care and a lack of resources in rural hospitals and communities, which further contribute to health disparities [35].

2.2 COVID-19 Visualization

Various systems have been developed to analyze and visualize epidemics and pandemics [8]. Preim and Lawonn [33] conducted a detailed survey of visual analytics tools designed for managing infectious diseases, chronic diseases, and other healthcare issues. Among the commonly used tools for tracking the COVID-19 pandemic, epidemiological maps play a prominent role in collecting and visualizing data on confirmed cases, deaths, and recoveries across spatial and temporal scales. 'Dashboards' have emerged as a popular tool for this purpose. One notable example is the COVID-19 Dashboard developed by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) [13]. This interactive web-based world map shows the real-time information on the location and number of confirmed COVID-19 cases, deaths, and recoveries for all affected countries. Zeng et al. [46] developed a semi-automatic layout adaption method that supports the COVID-19 dashboard on mobile devices. Additionally, the World Health Organization (WHO) has created an intuitive public dashboard resource that globally maps COVID-19 cases and deaths while facilitating between geographic regions.

Beyond dashboards, some other tools have been developed to monitor COVID-19 epidemiology. For instance, Worldometers.info

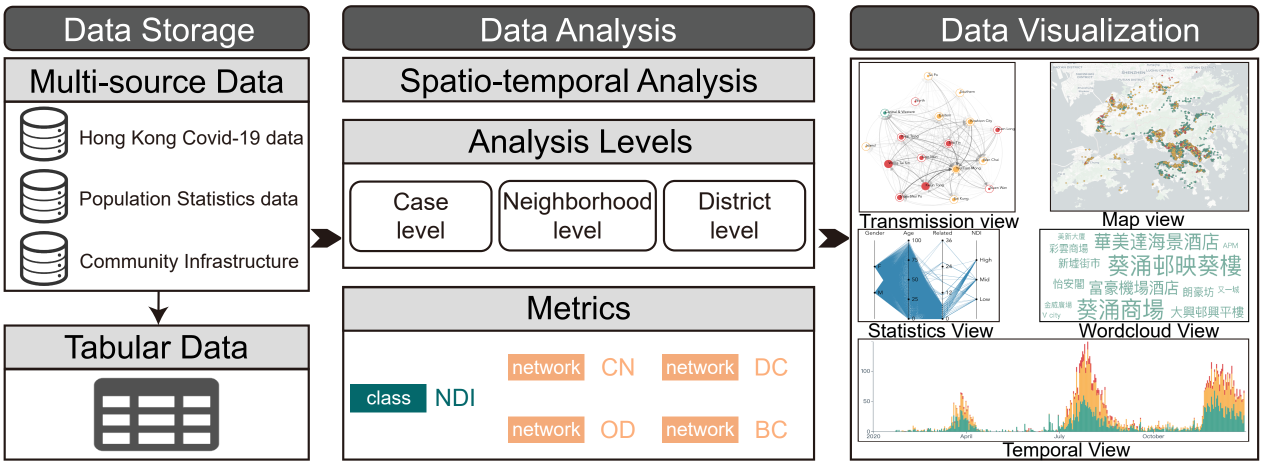


Figure 1: Overview of the system workflow which consists of three modules: Data Storage, Data Analysis, and Data Visualization.

is a popular real-time spreadsheet-like data-sharing website, which collects, aggregates and performs statistical analyses on official-sourced data. Modeling and visualizing the spread of COVID-19 under various conditions have also received attention to understand the spread patterns [32, 44]. Furthermore, decision-support visualizations [18, 41, 49] have been proposed to aid in the management and facilitate comparisons between different epidemic spread scenarios and decision measures, aiming to enhance decision-making processes in managing COVID-19 situation.

2.3 Network Visualization

Research on network layout has been dominated by force-directed strategies because they produce aesthetically pleasing node layouts and ensure reasonable link visibility [16]. Given the dense and massive nature of epidemic contact-tracing data, it is crucial to employ intuitive representations that allow users to dig out insights. Network visualization offers a compelling approach for representing contact-tracing data by incorporating topological information, node relations, and metrics that characterize individual nodes. Numerous studies have applied network visualization techniques to analyze various aspects of the COVID-19 pandemic.

One notable area of investigation involves the analysis of social networks to understand the spread of COVID-19 through human-to-human contact. For instance, Zhao et al. [48] examined the public's hotspots of concern regarding the COVID-19 epidemic on Sina Microblog, analyzing the trends of hot topics related to the pandemic. Such an approach enables governments and health departments to obtain timely access to public responses and implement appropriate measures for epidemic prevention and control.

Another significant focus has been on analyzing the COVID-19 virus itself, using network visualization to explore its genetic structure and evolution. Network analysis and visualization can complement traditional modeling techniques, providing enhanced estimations of COVID-19 pandemic risks and timely evidence to information preparedness plans [40]. For example, Milano et al. investigated the impact of the clinical evolution of COVID-19 pandemic and the spread of different variants between regions by studying the events in Italy [27, 28].

3 SYSTEM OVERVIEW

We follow a user-centered design process to develop and improve our visual analytics system iteratively. We work closely with two domain experts long engaged in research on social stratification and spatial analysis. Through exchanging ideas with experts, we collect several groups of data and design the system workflow (Figure 1).

The system consists of three modules: data storage module, data analysis module, and data visualization module. For our mission, we collect, store, and process multi-source data related to COVID-19 epidemic and neighborhood characteristics in Hong Kong. We provide 3 analysis levels in data analysis module: case level, neighborhood level, and district level. Besides, we choose Neighborhood Deprivation Index (NDI) as the key indication for social group classification and calculate the network quantitative metrics to help users analyze the network and explore insights (Sec. 4.2). In the data visualization module, we provide 6 coordinated views and rich interactions that allow a user to explore various and (maybe) new COVID-19 stories in an interactive fashion (Sec. 5).

4 DATA DESCRIPTION AND PROCESSING

4.1 Data Source and Description

Since the outbreak of COVID-19, Hong Kong has been the city with the highest COVID-19 death rate in the world, one of the cities most challenged by this unprecedented pandemic. In this work, we set Hong Kong as the typical city to research. The basic dataset used in our system is the **2020 Hong Kong COVID-19 historical data** published by the Centre for Health Protection (CHP) of the Department of Health (DH) and the Lands Department of Hong Kong. This dataset contains 25140 pieces of data, each of which has 24 attributes. The dataset includes basic information about every case, such as gender, age, and resident type. Moreover, the dataset includes temporal information and highlights every case's spatial information, detailing the exact building and track of the case.

Besides the Hong Kong COVID-19 historical data in 2020, we introduce the following subsidiary datasets in order to gain a deep insight into the society and conduct reliable social analysis:

- **2016 Population by-census Statistics data** is government public data released by the Hong Kong government in 2016. This census took District Council Constituency Areas (DCCA) as the smallest statistical unit. It provides information on the population in each neighborhood as well as the overall population distribution of Hong Kong.
- **2020 Hong Kong Geographic Community Infrastructure data** is obtained from the GeoCommunity Database of the Hong Kong government in 2020. The data reflects the improvement level of public facilities in various city regions. It further offers insights into the extent of community facility integration, population density, and economic development within each region. It is used as an evaluation indicator of the development level of a social area.

4.2 Metrics

In order to combine the multi-source data to support the identified analytical tasks, we select the following metrics related to social class and network, and conduct data processing.

4.2.1 Quantitative Metrics. Visual analysis aims to represent relationships between cases graphically. However, static infectious networks can still be hard to understand and might require processing (interactive filtering and browsing) and metrics that help identify and analyze patterns. To better filter and mine information from the network, we calculate three main metrics that provide quantitative insights regarding case node connections:

- **Number of contacts of a given case node (CN)** is measured by the number of contacts of a given node.
- **Out-degree (OD)** measures the number of outgoing edges for a given node, reflecting the number of infections of this given case.
- **Degree centrality (DC)** is measured by the number of direct contacts $X_{a,ga\neq g}$ of a given node divided by the total number of nodes n. It measures the number of links occurred upon a node, meaning that case nodes with a high degree centrality often serve as 'hubs' in the pandemic transmission network [47].

$$DC(a) = \frac{\sum_{a=1,\backslash a\neq g}^{n} X_{a,g}}{n-1} \quad (1)$$

- **Betweenness centrality (BC)** characterizes the relevance of case node for establishing short path with other nodes. The case nodes with a high betweenness centrality serve as 'bridges' that connect many different COVID-19 infected groups in the network.

$$BC(a) = \sum_{g,h\in V\backslash g,h\neq a} \frac{p_{g,h}(a)}{p_{g,h}} \quad (2)$$

  where $p_{g,h}$ is the number of paths that connect nodes g and h, while $p_{g,h}(a)$ is the number of paths tat connect nodes g and h through node a.

4.2.2 Neighborhood Deprivation Index (NDI) based Social Classification. To further analyze the socioeconomic gradient of the COVID-19 infectious risk, this research classifies all the confirmed cases into different social groups based on the deprivation index of the neighborhood where they live. The previous studies point out that people are segregated into various neighborhoods according to their ethnic background [25], housing affordability [36], occupation [11] and other socioeconomic-related issues. Therefore, the geographic units where one lives can approximately reflect his/her socioeconomic status. Further, the living environment also has impact on their risk of COVID-19 infection by provision of facilities and epidemic prevention within each neighborhood. Since we do not have individual socioeconomic attributes of these confirmed cases, we roughly evaluate them with the information of DCCA neighborhoods where they live.

Neighborhood Deprivation Index (NDI) is used to measure economic or social shortage in a given geographical area [38]. Initially designed for resource arrangement and urban planning, NDI has been applied in health inequalities assessment in the last decade [14, 17, 20]. Generally, material deprivation and social deprivation are two dimensions of NDI and different measurements are conducted in previous research based on their data accessibility [38]. In this study, we apply the Neighborhood Deprivation Index (NDI) in the measurement of both the neighborhood and individual socioeconomic status.

Previous research provides various versions of the NDI measurement [14, 20, 30]. We find that five core indications compose the multidimensional NDI in most previous research, including the income level, the labor force level, the education level, the housing ownership, and the living environment. Referring to these indicators, we use five DCCA attributes to measure NDI, including the percentage of households whose monthly median income is less than 10000, the percentage of those of economically active age who are unemployed, the percentage of people (aged 5+) who cannot write in English, the number of persons sharing one room and the percentage of households renting a house. The Cronbach's alpha of these five indicators is 0.9039. Principal Component Analysis (PCA) is applied to synthesize the NDI and the gained eigenvalue is 3.6 with a variance accounted cumulative percentage of 0.72. After normalizing the NDI, the DCCA neighborhoods are divided into low-deprivation neighborhoods, middle-deprivation neighborhoods and high-deprivation neighborhoods.

5 VISUALIZATION

To address the analytical tasks, we design an MCV (Multiple Coordinated View) system (as shown in Figure 2) that primarily incorporates five views (Control Panel, Temporal View, Transmission View, Map View, Wordcloud View and Statistics view), integrated with a set of interactions to facilitate system exploration.

5.1 Control Panel

In the control panel (Figure 2-A), we first display the three levels on NDI and the three levels of network analysis. We use color to encode three different social classes. Green indicates low neighborhood deprivation index, yellow indicates middle neighborhood deprivation index, and red indicates high neighborhood deprivation index. Users can select the analysis level they want to explore based on their analysis tasks. Both the transmission view and map view change with the selected analysis level. The network visualization can be filtered by four main metrics that provide quantitative insights regarding node connections (as mentioned in Sec. 4.2). The

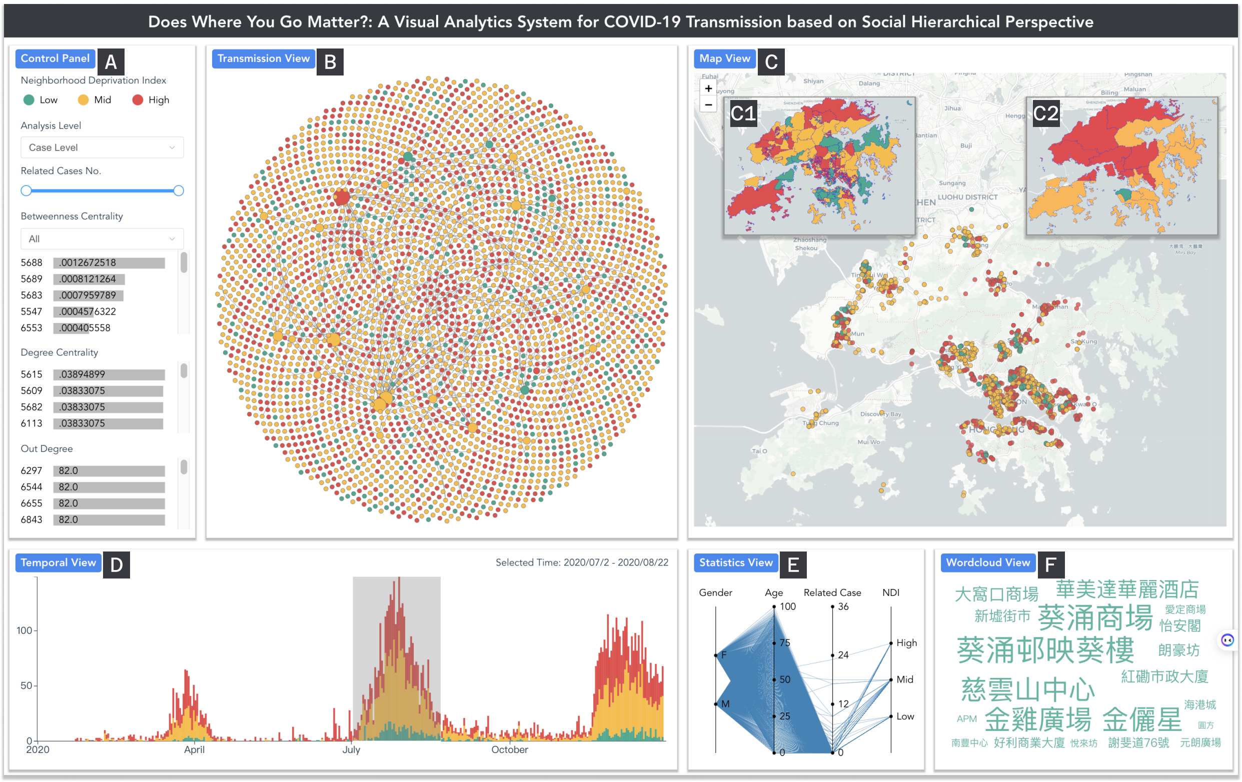


Figure 2: The interface of our proposed system which consists of 6 main views: (A) Control Panel allows user to filter cases according to metrics. (B) Transmission View supports three levels of network exploration. (C) Map View shows the spatial distribution of confirmed cases. (D) Temporal View displays the number of new confirmed cases per day. (E) Statistics View displays detailed information about cases, and (F) Wordcloud View presents the top 50 visited buildings.

summary information of these metrics is presented with bar charts at the bottom. These quantitative parameters enable comparative studies among cases and complement intuitive visual information provided by the graph representation. Users can click on the case they are interested in and get the connection relationship of that case in the transmission view.

5.2 Transmission View

In transmission view (Figure 2-B), we support exploration and analysis at three levels: case level, neighborhood level, and district level. We draw this network based on the Force Directed algorithm for visual clarity. In the case-level network, each node represents a confirmed case. The radius of the node indicates the number of cases related with the corresponding case (CN). The larger the node, the more cases are associated with the case. If one case is infected by another, there exists a directed edge between these two nodes. The color of a node codes the deprivation type of the neighborhood where the case lives. Green represents the case that lives in the low-deprivation neighborhoods, yellow represents the middle-deprivation neighborhoods, and red represents high-deprivation neighborhoods. Through analyzing the case-level network, users can explore infection paths of interest.

In the neighborhood- and district-levels networks, each node denotes one neighborhood/district. The number of cases occurred in this neighborhood/district is represented by the radius of the center circle in each node. The larger the center circle is, the more cases occur in this neighborhood/district. The color of a neighborhood node indicates its NDI, while the color of a district node is determined by the NDI of the most numerous neighborhoods in the district. Since people are more active in their living area, there may be many self-loops in neighborhood- and district-level networks. We mainly focus on the spread of COVID-19 among neighborhoods and districts, so we ignore the self-loops when drawing networks. If there is a movement path from one area to another, there exists a directed edge between the two corresponding nodes. The shade of the edge reflects the number of confirmed cases with that locus. The darker the color of the edge is, the more typical the infection trajectory is.

In this view, we design and realize the basic zoom in, zoom out, and scroll interaction. We implement two kinds of selection: hover over and click. When hovering over a node in the network, the network will be converted to the graph diagram consists the corresponding node and its related nodes. Thus, the user can quickly go through the nodes they are interested in by the hover-over interaction. Besides, users can also select a node by clicking it. On the basis of the hover-over operation, the click interaction will trigger the update of the map view and statistics view.

5.3 Map View

The map view (Figure 2-C) shows the geographical distribution of confirmed cases. In this view, each point represents a confirmed case. The color of the point represents the NDI of the case. Besides, map view could help users understand the relationships between different analysis levels. When the user selects the neighborhood level, the neighborhood borders, which is the boundaries of DCCA in our study, are drawn on the map and each area is colored according to their NDI respectively (Figure 2-C1). When the user selects the district level, the 18 districts are depictured on the map panel and the color of each district is also determined by the NDI of the majority of the neighborhoods within this district (Figure 2-C2) . For example, if one district has 10 neighborhoods in high deprivation, 10 in middle deprivation and 11 in low deprivation, we will color the district as red representing low deprivation. In the map view, zooming and dragging are realized to explore specific areas. In addition, users can hover over the neighborhood/district they are interested in to get the number of cases confirmed in this area.

5.4 Temporal View

In the temporal view(Figure 2-D), we design a stack histogram to display the fluctuation of the daily number of new cases in Hong Kong throughout the year and the change of the proportion of each deprivation neighborhood in this process. The X-axis represents the time, while Y-axis represents the number of new cases per day. The color of each bar encodes different deprivation neighborhoods, as defined in the control panel. Brushing is supported in the temporal view for users to filter cases within a specific time period. Specifically, users can drag and stretch the slider to get the visual content update in transmission view and map view at different periods.

5.5 Statistics View

The statistics view (Figure 2-E) provides details about the nodes in the transmission view. In this view, we display the basic information of the selected node using the parallel coordinate chart. The information about nodes is different at different analysis levels. For example, in the case-level network, if a user wants to explore the details of a case, he/she can click on this node to get its infectious sub-network in the transmission view, and get specific information about the case in the statistics view, including gender, age, number of related cases, and deprivation index (as Figure 2-E shows). Through the display of the highlighted path in the parallel coordinate chart, the user can know the specific information of the selected case and whether the case belongs to a specific group.

5.6 Wordcloud View

To analyze the locations of the case distributed and buildings visited by confirmed cases during the selected time period, we design the word cloud (Figure 2-F) for visited buildings. The more confirmed cases that have visited the building are, the bigger its name is in the word cloud. When a user clicks on a building's name in the word cloud, the building's location is displayed on the map view. Combined with map view and work cloud view, we can mine typical COVID-19 spatial spreading patterns.

6 CASE STUDY

6.1 Overall Spread Pattern of COVID-19 Cases

The COVID-19 observation period in this research spans from February to the end of 2020. By observing the temporal view (Figure 2-D), we identify three waves of virus transmission based on the stack graph's peaks. The first wave is from the mid-March to mid-April, followed by the second wave from early July to late September, and the third wave starts from late November. Compared with the first wave, the second and third waves show higher peak values and longer durations. More interesting insights can be found in the perspective of social hierarchy. As mentioned in the Control panel (Sec. 5.1), confirmed cases are assigned with three different colors. Across all three waves of virus transmission, a greater number of cases are observed from high-deprivation neighborhoods. This findings reflects the heightened susceptibility of socioeconomically disadvantaged groups to COVID-19, which aligns with previous research, Poor people in Hong Kong are more likely to live in densely populated dwellings, facilitating the easier spread.

Next, we screen the time horizons of the three waves in the temporal view and compare the different case networks (Figure 3 b1-b3). Firstly, we observe the transmission chain is much clear in the first wave, while the second and third waves consist of more single cases. It suggests that it is easier to trace the virus's source at the early stages of the pandemic, but becomes increasingly challenging as individuals become infected through accidental encounters. Secondly, during the first and second waves, nodes with a higher degree centrality mostly live in high-middle-deprivation neighborhoods, while larger green nodes emerge in the third wave. It tells that with the development of COVID-19, people from advantaged neighborhoods became easier to be the "hub" of the virus transmission, though groups in high-middle-deprivation neighborhood act as "hubs" consistently. Finally, there exists segregation of coronavirus transmission in the COVID-19 pandemic. Figure 4 (a) and Figure 4 (b) are two subnetworks showing the transmission patterns among related cases when node 6702 and 8546 are selected. These nodes either are related to other nodes in the same color with them, which reveals that virus are more likely to be transferred within the same social group living in the similar neighborhood.

Finally, by combining the case-level map view(Figure 3 c1-c3) and transmission view, we can understand the dynamic changes in the spatial distribution of confirmed cases in Hong Kong. When we display the first wave, we observe that cases during this wave are mainly distributed in Hong Kong Island and Kowloon Island, but fewer in New District. After brushing the second wave, more cases can be found in Kowloon Island and New District, which means the virus moves toward the northern area. However, if we select cases in the third wave, we can observe that they are distributed more evenly in the three districts.

6.2 District Level of COVID-19 Spreading

To get a close insight at the visiting patterns among different districts in Hong Kong, we can observe the district-level network, which provides us with the visiting situation between every two districts by confirmed cases. There are 18 nodes in the district-level network, respectively representing the 18 districts of Hong Kong (as Figure 5 shows). As can be seen from the color of directed edges

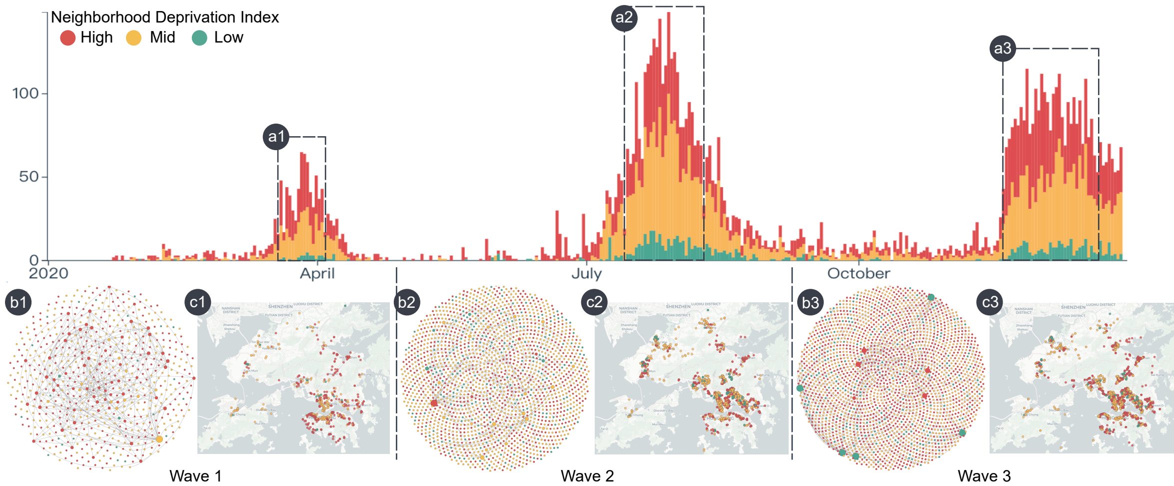


Figure 3: The spread pattern of the COVID-19 in Hong Kong in 2020. There are three main spreading waves (a1-a3). b1-b3 shows respectively the case-level transmission view corresponding to these three waves. c1-c3 shows the respectively the case-level map view corresponding to these three waves.

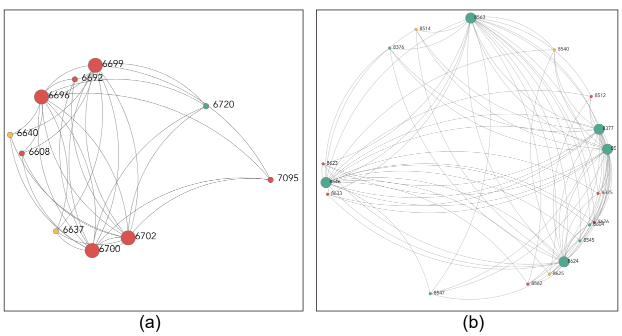


Figure 4: The two subnetworks are selected from the third wave. Subnetwork (a) is obtained by selecting the largest red node in the transmission view, while (b) by selecting the largest green node.

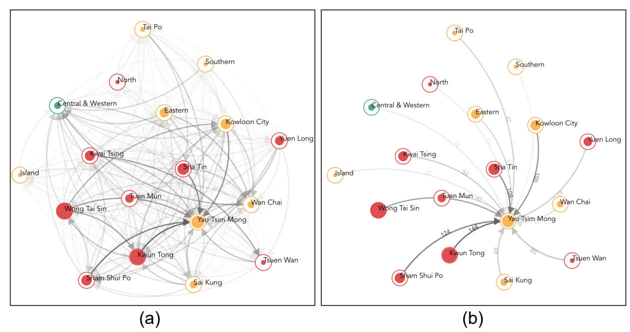


Figure 5: (a) The district-level network. Each node represents one district in Hong Kong. (b) The district-level sub-network of Yau Tsim Mong district.

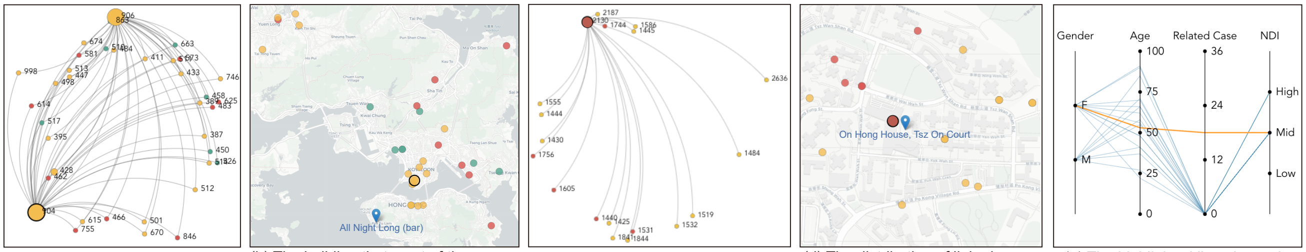
in the network, Yau Tsim Mong, Central and Western are mostly visited by cases from other districts. Generally, there are more visiting observations between districts close to each other in geography. For example, cases in Sham Shui Po and Kowloon City visit Yau Tsim Mong more because they are neighbors of each other. A similar pattern can be found between Tsuen Wan and Kwai Tsing, Kwun Tong and Wong Tai Sin.

We also investigate the locations of the confirmed cases distributed. Generally, population density is positively associated with infectious case numbers in a specific area. Figure 2-F shows us the top 50 buildings visited by confirmed cases, most of which are non-residential buildings. Golden Rooster Square is the site that developed the "Dancing Group" including more than 700 confirmed cases. The Tsz Wan Shan Shopping Center is a shopping center surrounded by public housing dwellings, which accommodate low-income people. While most of these buildings are for shopping and recreation, the remaining buildings are almost public housing resided by groups in high deprivation neighborhoods, such as Kam Wan House, Choi Fai Estate, Tsz Ching Estate in Ching Hong House and Kwai Shing Estate. Therefore, people who live in public housing are easier to be exposed to COVID-19 infectious risk.

6.3 The Sub-network Analysis

The COVID-19 virus spreads from person to person, which provides the basis for sub-network analysis in this case. Besides the general pattern of COVID-19 in Hong Kong, many infectious groups have been reported by the government since the outbreak of COVID-19, such as the "Hot-pot Group" or "Nursing Home Group[2]". We aim to study the structure and dynamics of groups to find out how the virus spreads between individuals, especially from the social hierarchical perspective. As we mentioned in Sec. 6.2, people who visit some recreation sites may be infected by other cases, as a result of which a group may include people who do not know each other. However, in many cases, the virus is more likely to be transferred between acquaintances and relatives. In the case-level network, we notice that the sizes of some nodes are bigger than other nodes because they are connected to more other cases. It can be observed in the control panel that the node with the maximum output degree

[2]https://topick.hket.com/article/2698722



(a) The case-level subnetwork of node 994 and its linked nodes.
(b) The building that most of these nodes have visited is All Night Long.
(c) The case-level subnetwork of node 2130 and its linked nodes.
(d) The distribution of linked cases and the most visited building.
(e) The highlighted line shows the information of node 2130.

Figure 6: The sub networks and spatial distributions of case 994 and case 2130 with their related cases.

(the most infectious case) is case 994. We then click this node, and the case-level network shows us other nodes linked to it (as Figure 6 (a) shows). In the corresponding map view, we notice the location of the clicked node and their related nodes are highlighted. We also find that most of these related cases have visited one site, which is the All Night Long (Figure 6 (b)), the most popular place for beers and dancing. Thus, we deduce that it is likely that these cases are linked together in some bars in All Night Long. After checking with the group information published by the Hong Kong government, we confirmed our hypothesis, and it is the "Bar and Band group" which appears in March 2020[3]. This process builds our confidence that our system is valid and reliable to be consistent with real situations.

We further explore another node with a high degree centrality. We select node 2130 because it not only has a high degree centrality, but also is in red, which means this case is in high deprivation status. The case-level network of node 2130 and its related nodes are shown in Figure 6 (c). Combined with the visiting sites of these cases, we find that all of these cases have one overlap, which is a nursing home(Figure 6 (d)). Refer to the statistic view information(Figure 6 (e)), We also notice that most of these cases are old people and many of them live in these sites. Again we check the reported date of these cases and confirmed that this group is the "Nursing Home group". Besides this finding, we also notice that all these older cases are from high-deprivation neighborhoods, and we do not find any green nodes in the network. It reflects that it is less likely for the deprived group to infect the group from wealthy neighborhoods. While it is more likely that the virus can be transmitted from the advantaged group to the disadvantaged group. For example, we also find that some Philippines maids are infected by their employers who visit some place of creation.

7 DISCUSSION

Export Feedback. Incorporating a social hierarchical perspective, this study offers multiple views of the COVID-19 pandemic in Hong Kong throughout 2020. Our findings validate previous research by confirming the existence of infection risk disparities among different socioeconomically deprived groups. We uncover the reasons why a group from high-deprived neighborhoods is more vulnerable to the coronavirus and find that the buildings mostly visited are located in high-deprived neighborhoods. Furthermore, there exists a segregation of virus transmission in the COVID-19 pandemic, with individuals classified into closed loops based on their deprivation levels of residence. Finally, vulnerable groups heavily rely on public transportation, thereby facilitating the virus's spread among groups from high-deprivation neighborhoods. Our study provides a comprehensive understanding of how the living space shapes their infection risk. To gather expert feedback, We presented our system to collaborating experts and engaged in discussions regarding our findings. The experts expressed enthusiasm for the exciting patterns and insights we uncovered, and they recommended the richness of interaction provided by our system.

Future Work. There are more issues residing on the COVID-19 pandemic besides what we find in this study. Due to the data limitations, we do not provide detailed information of the confirmed cases except for where one visit and several demographic indications. We are aware that it needs a more detailed investigation into the micro mechanism of how individuals intervened with other persons and groups during such a pandemic event. For example, some neighborhoods saw more confirmed cases even though they stand out in socioeconomic conditions. On the other hand, more explanation can be put into the transmission of the virus among neighborhoods and districts. Besides spatial proximity and transportation, other factors such as economic interaction, political ideology as well as urban morphology can also provide insightful explanations.

8 CONCLUSION

By using data from various resources, this study adopts a social hierarchical perspective to investigate the dynamics of COVID-19 in Hong Kong during 2020. We develop an interactive Multiple Coordinated View (MCV) visualization system to facilitate multi-level spatial-temporal analysis. The findings of this study are as follows: Firstly, groups from deprived neighborhood areas are more vulnerable to the COVID-19 virus and we can find more confirmed cases at each wave of pandemic. Secondly, we identify the socioeconomic gradient in COVID-19 infectious risk within Hong Kong. Some districts, especially those with more low-deprivation neighborhoods, such as the public housing setting, have a higher concentration of confirmed cases. In contrast, wealthier neighborhoods show fewer cases. Thirdly, we discover that geographical proximity and transportation play significant roles in facilitating the movement of confirmed cases among different districts. Finally, which is more surprising, segregation of virus transmission between different social hierarchies is found in the process of coronavirus transmission. Our findings provide valuable insights for the early prevention and control measures of infectious diseases and offer guidance for minimizing the disparity of infection risk between deprived and advantaged groups in densely populated cities like Hong Kong.

[3]http://paper.wenweipo.com/2020/04/04/HK2004040006.htm

REFERENCES

[1] David R Agrawal and Aline Bütikofer. 2022. Public finance in the era of the COVID-19 crisis. *International Tax and Public Finance* 29, 6 (2022), 1349–1372.

[2] Isabel Aguilar-Palacio, Lina Maldonado, Sara Malo, Raquel Sánchez-Recio, Iván Marcos-Campos, Rosa Magallón-Botaya, and Mª José Rabanaque. 2021. COVID-19 inequalities: Individual and area socioeconomic factors (Aragón, Spain). *International Journal of Environmental Research and Public Health* 18, 12 (2021), 6607.

[3] Faheem Ahmed, Na'eem Ahmed, Christopher Pissarides, and Joseph Stiglitz. 2020. Why inequality could spread COVID-19. *The Lancet Public Health* 5, 5 (2020), e240.

[4] Maria de Fátima Carvalho Alves, Maria da Luz Lima Mendonça, Janice de Jesus Xavier Soares, Silvânia Da Veiga Leal, Menilita Dos Santos, Júlio Monteiro Rodrigues, and Edna Duarte Lopes. 2021. Knowledge, attitudes and practices towards COVID-19: A cross-sectional study in the resident cape-verdean population. *Social sciences & humanities open* 4, 1 (2021), 100184.

[5] Ana I Balsa, Naomi Seiler, Thomas G McGuire, and M Gregg Bloche. 2003. Clinical uncertainty and healthcare disparities. *American Journal of Law & Medicine* 29, 2-3 (2003), 203–219.

[6] Clare Bambra, Ryan Riordan, John Ford, and Fiona Matthews. 2020. The COVID-19 pandemic and health inequalities. *J Epidemiol Community Health* 74, 11 (2020), 964–968.

[7] Gabriela Cuevas Barron, George Laryea-Adjei, Vaira Vike-Freiberga, Ibrahim Abubakar, Henia Dakkak, Delanjathan Devakumar, Anders Johnsson, Selma Karabey, Ronald Labonté, Helena Legido-Quigley, et al. 2022. Safeguarding people living in vulnerable conditions in the COVID-19 era through universal health coverage and social protection. *The Lancet Public Health* 7, 1 (2022), e86–e92.

[8] Lauren N Carroll, Alan P Au, Landon Todd Detwiler, Tsung-chieh Fu, Ian S Painter, and Neil F Abernethy. 2014. Visualization and analytics tools for infectious disease epidemiology: a systematic review. *Journal of biomedical informatics* 51 (2014), 287–298.

[9] Xi Chen, Wei Zeng, Yanna Lin, Hayder Mahdi Al-maneea, Jonathan Roberts, and Remco Chang. 2021. Composition and Configuration Patterns in Multiple-View Visualizations. *IEEE Transactions on Visualization and Computer Graphics* 27, 2 (2021), 1514–1524. https://doi.org/10.1109/TVCG.2020.3030338

[10] A Das, S Ghosh, K Das, T Basu, M Das, and I Dutta. 2020. Modeling the effect of area deprivation on COVID-19 incidences: a study of Chennai megacity, India. *Public Health* 185 (2020), 266–269.

[11] Niki Dickerson. 2008. Occupational and residential segregation: The confluence of two systems of inequality. *Labor Studies Journal* 33, 4 (2008), 393–411.

[12] Finn Diderichsen, Lia Giraldo da Silva Augusto, and Bernadete Perez. 2019. Understanding social inequalities in Zika infection and its consequences: a model of pathways and policy entry-points. *Global Public Health* 14, 5 (2019), 675–683.

[13] Ensheng Dong, Hongru Du, and Lauren Gardner. 2020. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases* 20, 5 (2020), 533–534.

[14] Rodrigo Javier Durán and Miguel Ángel Condorí. 2019. Deprivation index for small areas based on census data in Argentina. *Social Indicators Research* 141 (2019), 331–363.

[15] Paul Farmer. 1996. Social inequalities and emerging infectious diseases. *Emerging infectious diseases* 2, 4 (1996), 259.

[16] Thomas MJ Fruchterman and Edward M Reingold. 1991. Graph drawing by force-directed placement. *Software: Practice and experience* 21, 11 (1991), 1129–1164.

[17] Carlos Gradín, Olga Cantó, and Coral Del Río. 2017. Measuring employment deprivation in the EU using a household-level index. *Review of Economics of the Household* 15 (2017), 639–667.

[18] Lucien Hardy. 2020. Counting risk increments to make decisions during an epidemic. *arXiv preprint arXiv:2006.11244* (2020).

[19] Gunness Harlem. 2020. Descriptive analysis of social determinant factors in urban communities affected by COVID-19. *Journal of Public Health* 42, 3 (2020), 466–469.

[20] Sabrina Havard, Séverine Deguen, Julie Bodin, Karine Louis, Olivier Laurent, and Denis Bard. 2008. A small-area index of socioeconomic deprivation to capture health inequalities in France. *Social science & medicine* 67, 12 (2008), 2007–2016.

[21] Jacob Hoofman and Elizabeth Secord. 2021. The effect of COVID-19 on education. *Pediatric Clinics* 68, 5 (2021), 1071–1079.

[22] Kenji Karako, Peipei Song, Yu Chen, and Wei Tang. 2020. Analysis of COVID-19 infection spread in Japan based on stochastic transition model. *Bioscience trends* 14, 2 (2020), 134–138.

[23] Leo Lopez, Louis H Hart, and Mitchell H Katz. 2021. Racial and ethnic health disparities related to COVID-19. *Jama* 325, 8 (2021), 719–720.

[24] Marc Marí-Dell'Olmo, Mercè Gotsens, M Isabel Pasarín, Maica Rodríguez-Sanz, Lucía Artazcoz, Patricia Garcia de Olalla, Cristina Rius, and Carme Borrell. 2021. Socioeconomic inequalities in COVID-19 in a European urban area: two waves, two patterns. *International journal of environmental research and public health* 18, 3 (2021), 1256.

[25] Douglas S Massey and Nancy A Denton. 1988. The dimensions of residential segregation. *Social forces* 67, 2 (1988), 281–315.

[26] Sergio Meneses-Navarro, María Graciela Freyermuth-Enciso, Blanca Estela Pelcastre-Villafuerte, Roberto Campos-Navarro, David Mariano Meléndez-Navarro, and Liliana Gómez-Flores-Ramos. 2020. The challenges facing indigenous communities in Latin America as they confront the COVID-19 pandemic. *International Journal for Equity in Health* 19, 1 (2020), 1–3.

[27] Marianna Milano, Giuseppe Agapito, and Mario Cannataro. 2022. Application of CCTV Methodology to Analyze COVID-19 Evolution in Italy. *BioTech* 11, 3 (2022), 33.

[28] Marianna Milano, Chiara Zucco, and Mario Cannataro. 2021. COVID-19 Community Temporal Visualizer: A new methodology for the network-based analysis and visualization of COVID-19 data. *Network Modeling Analysis in Health Informatics and Bioinformatics* 10 (2021), 1–38.

[29] Luiza Nassif-Pires, Laura de Lima Xavier, Thomas Masterson, Michalis Nikiforos, Fernando Rios-Avila, et al. 2020. *Pandemic of inequality*. Technical Report. Levy Economics Institute.

[30] Michael Noble, Helen Barnes, Gemma Wright, and Benjamin Roberts. 2010. Small area indices of multiple deprivation in South Africa. *Social indicators research* 95 (2010), 281–297.

[31] Evrim Oral, Susanne Straif-Bourgeois, Ariane L Rung, Edward S Peters, et al. 2020. The effect of area deprivation on COVID-19 risk in Louisiana. *PLoS One* 15, 12 (2020), 1–12.

[32] Alexandru Popa. 2020. A decision support system for optimizing the cost of social distancing in order to stop the spread of COVID-19. *arXiv preprint arXiv:2004.02807* (2020).

[33] Bernhard Preim and Kai Lawonn. 2020. A survey of visual analytics for public health. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 543–580.

[34] Benjamin Rader, Samuel V Scarpino, Anjalika Nande, Alison L Hill, Ben Adlam, Robert C Reiner, David M Pigott, Bernardo Gutierrez, Alexander E Zarebski, Munik Shrestha, et al. 2020. Crowding and the shape of COVID-19 epidemics. *Nature medicine* 26, 12 (2020), 1829–1834.

[35] Maritza Vasquez Reyes. 2020. The disproportional impact of COVID-19 on African Americans. *Health and human rights* 22, 2 (2020), 299.

[36] Jonathan Rothwell and Douglas S Massey. 2009. The effect of density zoning on racial segregation in US urban areas. *Urban Affairs Review* 44, 6 (2009), 779–806.

[37] Laura J Samuel, Darrell J Gaskin, Antonio J Trujillo, Sarah L Szanton, Andrew Samuel, and Eric Slade. 2021. Race, ethnicity, poverty and the social determinants of the coronavirus divide: US county-level disparities and risk factors. *BMC Public Health* 21 (2021), 1–11.

[38] Carmen Sánchez-Cantalejo, Ricardo Ocana-Riola, and Alberto Fernández-Ajuria. 2008. Deprivation index for small areas in Spain. *Social Indicators Research* 89 (2008), 259–273.

[39] Muna Shifa, David Gordon, Murray Leibbrandt, and Mary Zhang. 2022. Socioeconomic-related inequalities in COVID-19 vulnerability in South Africa. *International Journal of Environmental Research and Public Health* 19, 17 (2022), 10480.

[40] Mike KP So, Agnes Tiwari, Amanda MY Chu, Jenny TY Tsang, and Jacky NL Chan. 2020. Visualizing COVID-19 pandemic risk through network connectedness. *International Journal of Infectious Diseases* 96 (2020), 558–561.

[41] Mathias Unberath, Kimia Ghobadi, Scott Levin, Jeremiah Hinson, and Gregory D Hager. 2020. Artificial Intelligence-Based Clinical Decision Support for COVID-19–Where Art Thou? *Advanced Intelligent Systems* 2, 9 (2020), 2000104.

[42] Navdar Doğuş Uzun, Mustafa Tekin, Emre Sertel, and Alpay Tuncar. 2020. Psychological and social effects of COVID-19 pandemic on obstetrics and gynecology employees. *Journal of Surgery and Medicine* 4, 5 (2020), 355–358.

[43] Zhicheng Wang and Kun Tang. 2020. Combating COVID-19: health equity matters. *Nature medicine* 26, 4 (2020), 458–458.

[44] Chuang Yang, Zhiwen Zhang, Zipei Fan, Renhe Jiang, Quanjun Chen, Xuan Song, and Ryosuke Shibasaki. 2022. Epimob: Interactive visual analytics of citywide human mobility restrictions for epidemic control. *IEEE Transactions on Visualization and Computer Graphics* (2022).

[45] Teo You Yenn. 2022. *This is what inequality looks like.* Ethos Books.

[46] W. Zeng, X. Chen, Yihan Hou, Lingdan Shao, Zhe Chu, and Remco Chang. 2023. Semi-Automatic Layout Adaptation for Responsive Multiple-View Visualization Design. *IEEE Transactions on Visualization and Computer Graphics* (2023).

[47] Junlong Zhang and Yu Luo. 2017. Degree centrality, betweenness centrality, and closeness centrality in social network. In *2017 2nd international conference on modelling, simulation and applied mathematics (MSAM2017)*. Atlantis press, 300–303.

[48] Yuxin Zhao, Sixiang Cheng, Xiaoyan Yu, and Huilan Xu. 2020. Chinese public's attention to the COVID-19 epidemic on social media: observational descriptive study. *Journal of medical Internet research* 22, 5 (2020), e18825.

[49] Anatoly Zhigljavsky, Ivan Fesenko, Henry Wynn, Roger Whitaker, Kobi Kremnizer, Jack Noonan, and Jonathan Gillard. 2020. A prototype for decision support tool to help decision-makers with the strategy of handling the COVID-19 UK epidemic. *medRxiv* (2020), 2020–04.