

Digital epidemiology

Lesson 1

Michele Tizzoni

Dipartimento di Sociologia e Ricerca Sociale
Via Verdi 26, Trento
Ufficio 6, 3 piano



UNIVERSITÀ
DI TRENTO



Who I am

- Michele Tizzoni
- Email: michele.tizzoni@unitn.it (drop me an email to ask for a meeting)
- Ufficio: n°6, piano 3, Sociology Building
- Research interests: computational social sciences, computational epidemiology, complex networks
- Teaching AA 23/24
 - **Digital epidemiology (DISI)**
 - Introduction to big data and computational methods (DSRS)

Schedule

- Lectures will take place:
 - Every Tuesday, 11.30 - 13.30
 - Every Wednesday, 11.30 - 13.30
 - Rooms will vary, please check the timetable

Topics

- Introduction to basic concepts of epidemiology and public health
- Mathematical modeling of infectious diseases
- Network theory and epidemic modeling on networks
- Human mobility, space and metapopulation models
- Proximity networks from wearable sensors
- Digital public health surveillance, tools and study designs
- Social contagion
- Modeling behavioral changes in response to epidemics

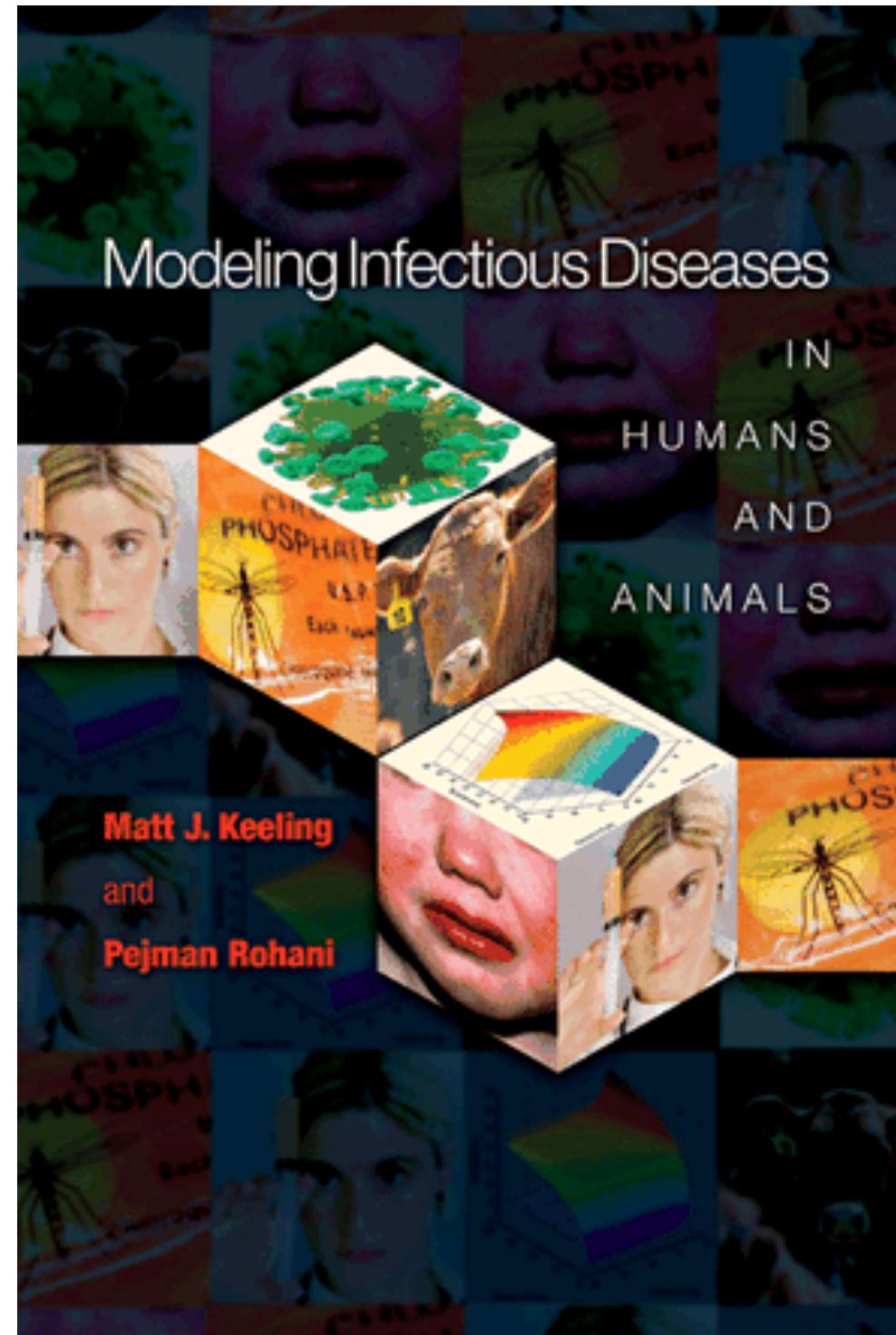
Coding

- The course will be structured as a mix of theory and coding laboratory
- I will present some coding examples that describe the concepts explained in the theory part
- Language used: Python
- Platform: Jupyter notebooks
- Python coding will not be required to pass the exam but concepts explained in the coding lab will be included as a subject of the exam

Materials

- I will upload slides and coding examples in my GitHub repository
 - [https://github.com/micheletizzoni/Digital epidemiology](https://github.com/micheletizzoni/Digital_epidemiology)
- The main reference for the whole course is the book by Marcel Salathé
 - Digital epidemiology: <https://www.digitalepibook.com/>
 - The book is free and publicly available

More references



Modeling Infectious Diseases in Humans and Animals

M.J. Keeling and P. Rohani

Princeton University Press

- An Introduction to Infectious Disease Modeling. Vynnycky and White. Oxford University Press.
- The rules of contagion. A. Kucharski
- Additional scientific papers will be mentioned during the lectures

Exam

- Oral exam on the topics taught during the course
- All students are asked to bring a short presentation (10-12 slides) to describe an in-depth study of their interest
- In-depth study can be: a research paper related to the course topics, the analysis of a dataset, a simulation exercise, etc.
- Examples of possible in-depth studies will be provided during the course

Introduction

The big picture

- ◆ The digital image of the world is tracking the world more and more closely.
- ◆ This allows us to use computation to extract patterns and establish causal inferences using tools from data mining, machine learning, statistics
- ◆ Mathematical modeling and forecast now happen on a data-rich landscape (e.g., mobility data, OSN data) and are fed by data streams from multiple sources
- ◆ We can assess our models against reality at unprecedented speed and scale, and feed back to models

New York City

1.1 billion taxi calls
6 years
350 Gb





facebook

December 2010
P. Butler

digital traces of human behavior

[Slide by C. Cattuto]

Italy:

97% of population owns a smartphone

84% has access to the internet

6+ hours a day online

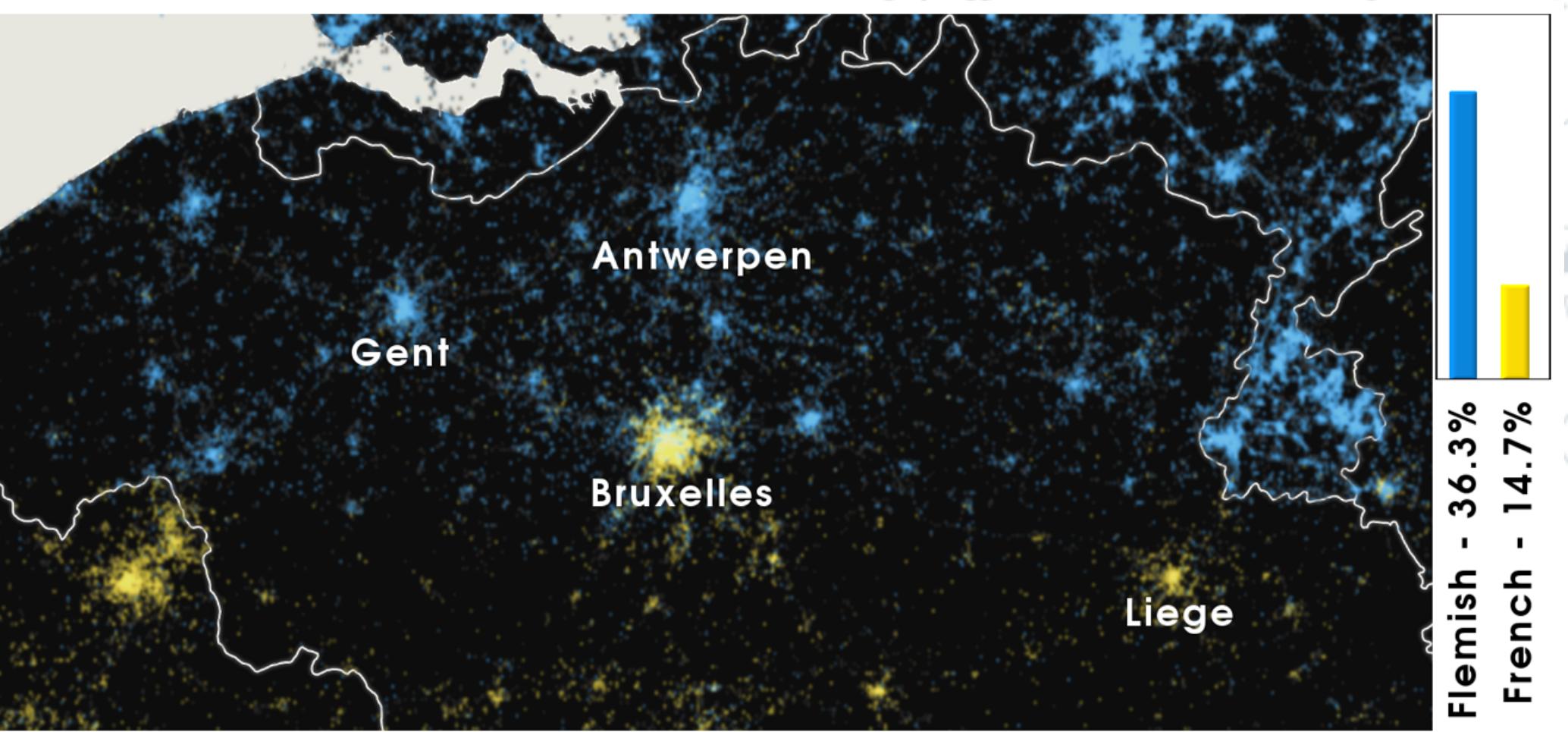
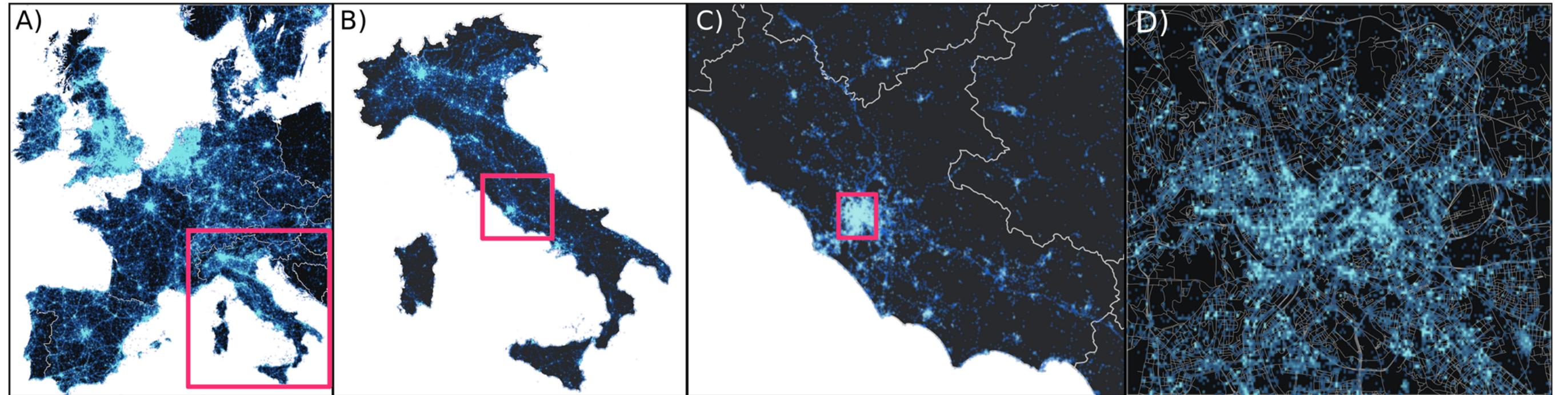
2 hours a day using social media

Digital traces

- historical view
- temporal horizon
- limited reproducibility
- limited context
- data protection challenges

- available as a side effect of many activities
- machine processable, pattern discovery
- high coverage, can work at scale

Digital demography



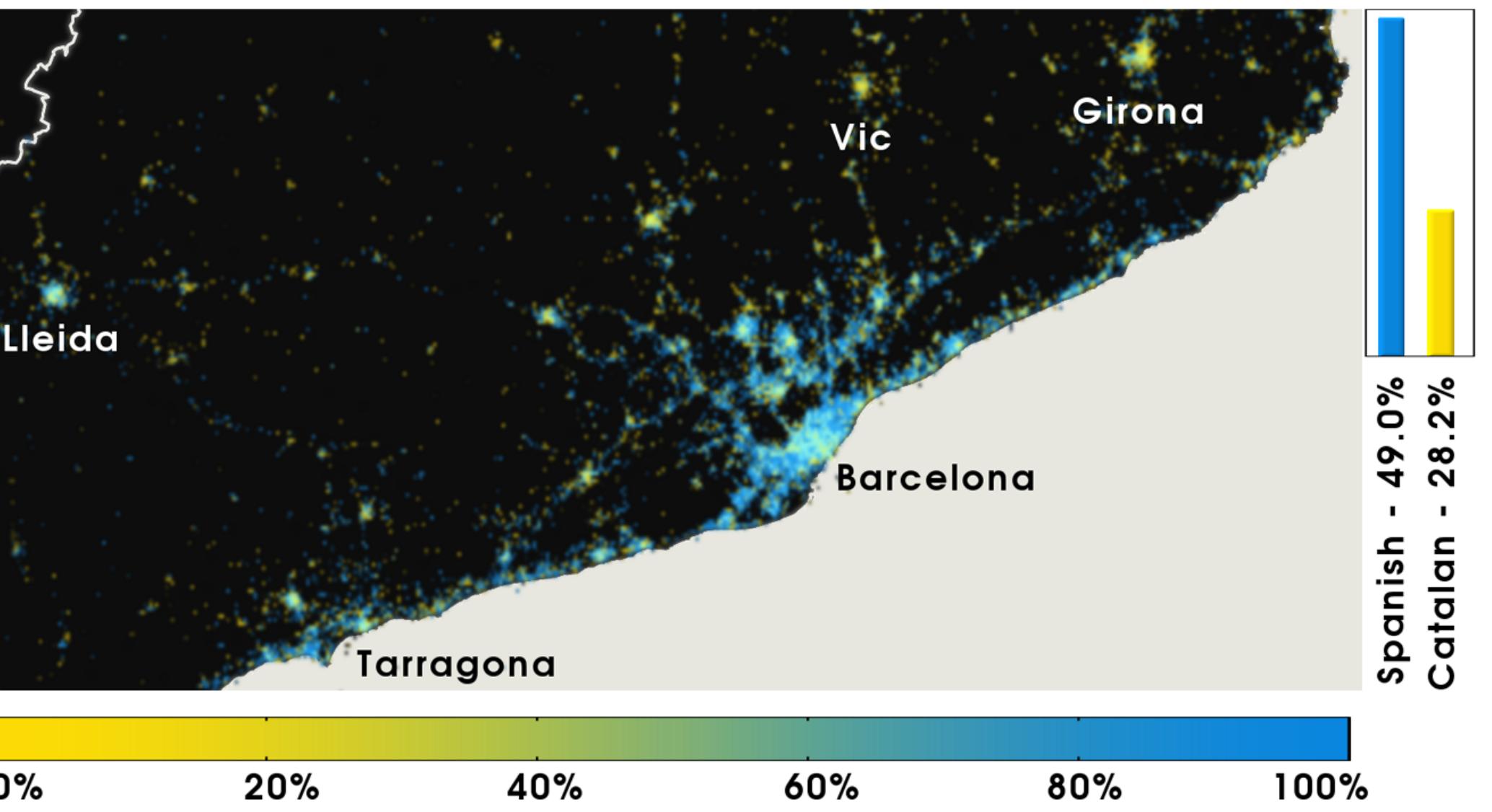
PLOS ONE

OPEN ACCESS PEER-REVIEWED
RESEARCH ARTICLE

The Twitter of Babel: Mapping World Languages through Microblogging Platforms

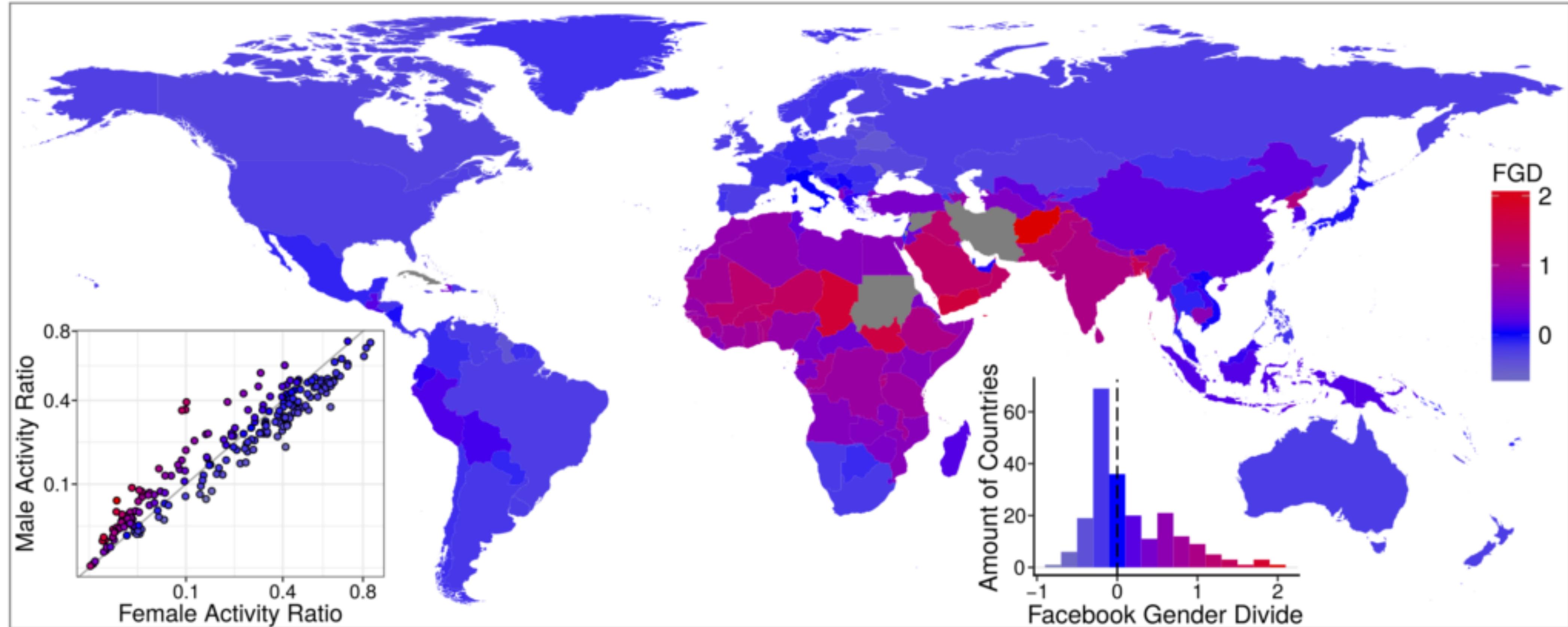
Delia Mocanu, Andrea Baronchelli, Nicola Perra, Bruno Gonçalves, Qian Zhang, Alessandro Vespignani

Published: April 18, 2013 • <https://doi.org/10.1371/journal.pone.0061981>



[Slide by C. Cattuto]

Gender inequalities



Garcia et al. PNAS 2018

[Slide by C. Cattuto]

Social capital



search: view: County ZIP Code High School College focus: try... West California Texas New York

SELECT A COUNTY TO SEE DATA

Economic Connectedness

Based on Current Income

40.3% of friends are high-income

Median: 50th

Low : 17.8%

High : 64.6%

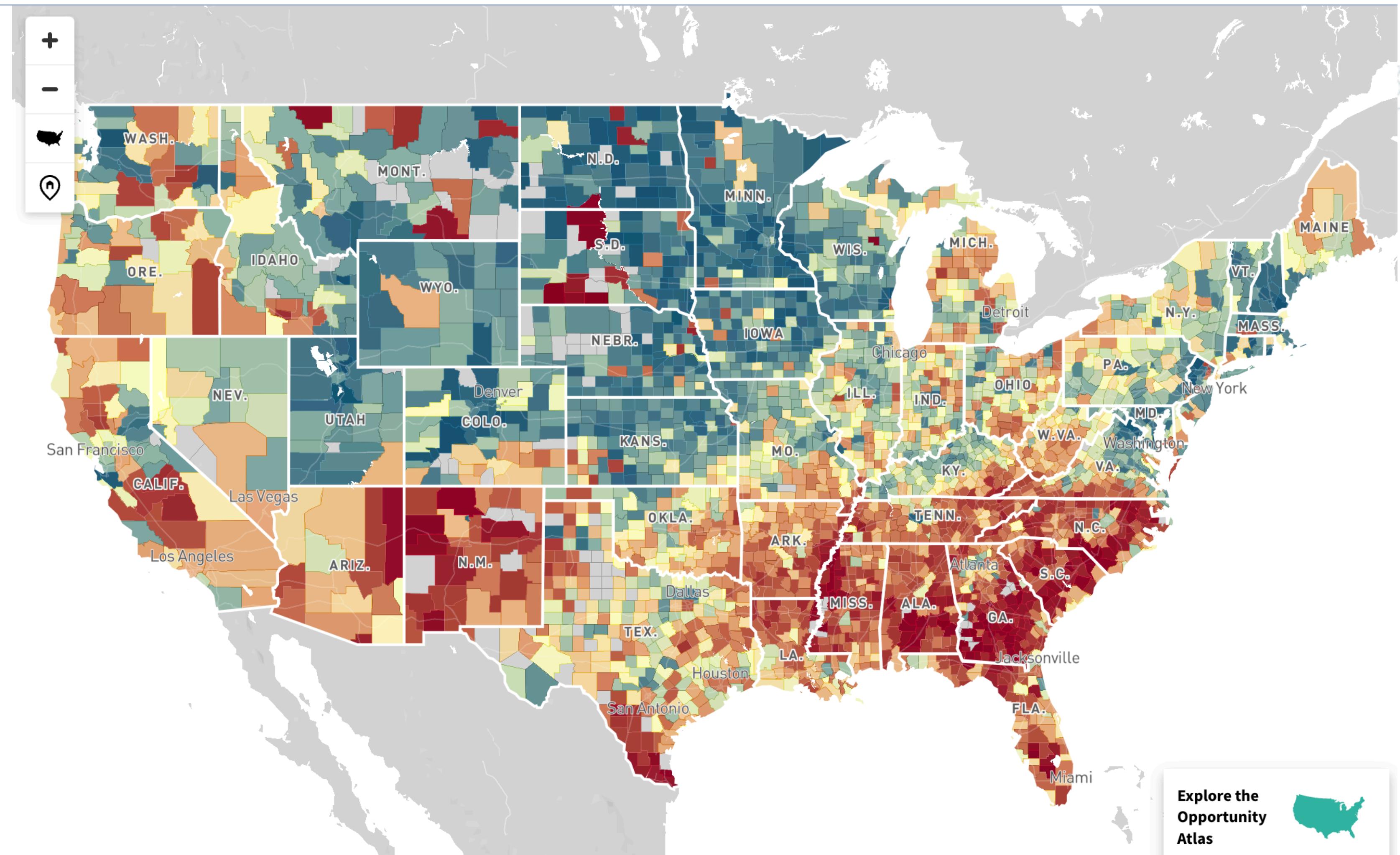
How does economic connectedness relate to economic mobility? >

What determines the level of economic connectedness? >

Cohesiveness

Civic Engagement

socialcapital.org



download data

download image

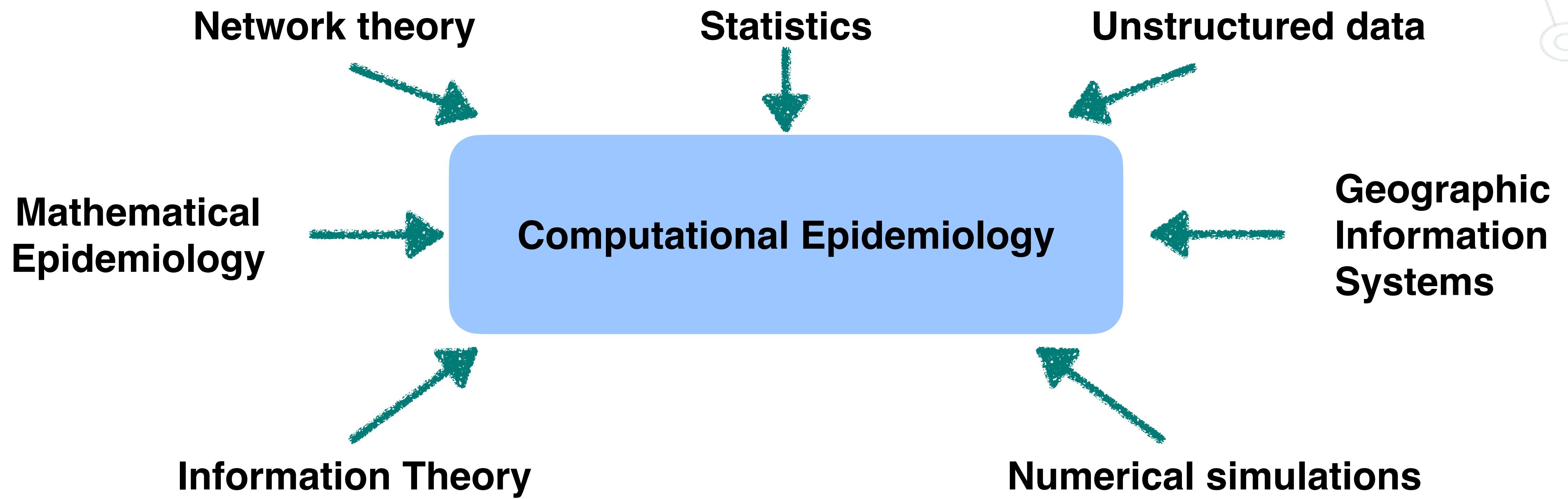
share

Explore the
Opportunity
Atlas

Digital Health

- ◆ User-generated content provides information about health habits, lifestyles, and attitudes towards health (nutrition, exercise, reproductive rights, drugs consumption, protective or risky behaviors)
- ◆ **Active** vs **passive** data collection
 - ◆ Passive: social media data, mobile phone traces
 - ◆ Active: participatory surveillance, crowdsourced experiments
- ◆ Data from wearables and sensors track individual health conditions at high resolution
- ◆ Statistics/machine learning models can be trained on such data-rich landscape to explain/predict the observed patterns of disease

Computational Epidemiology



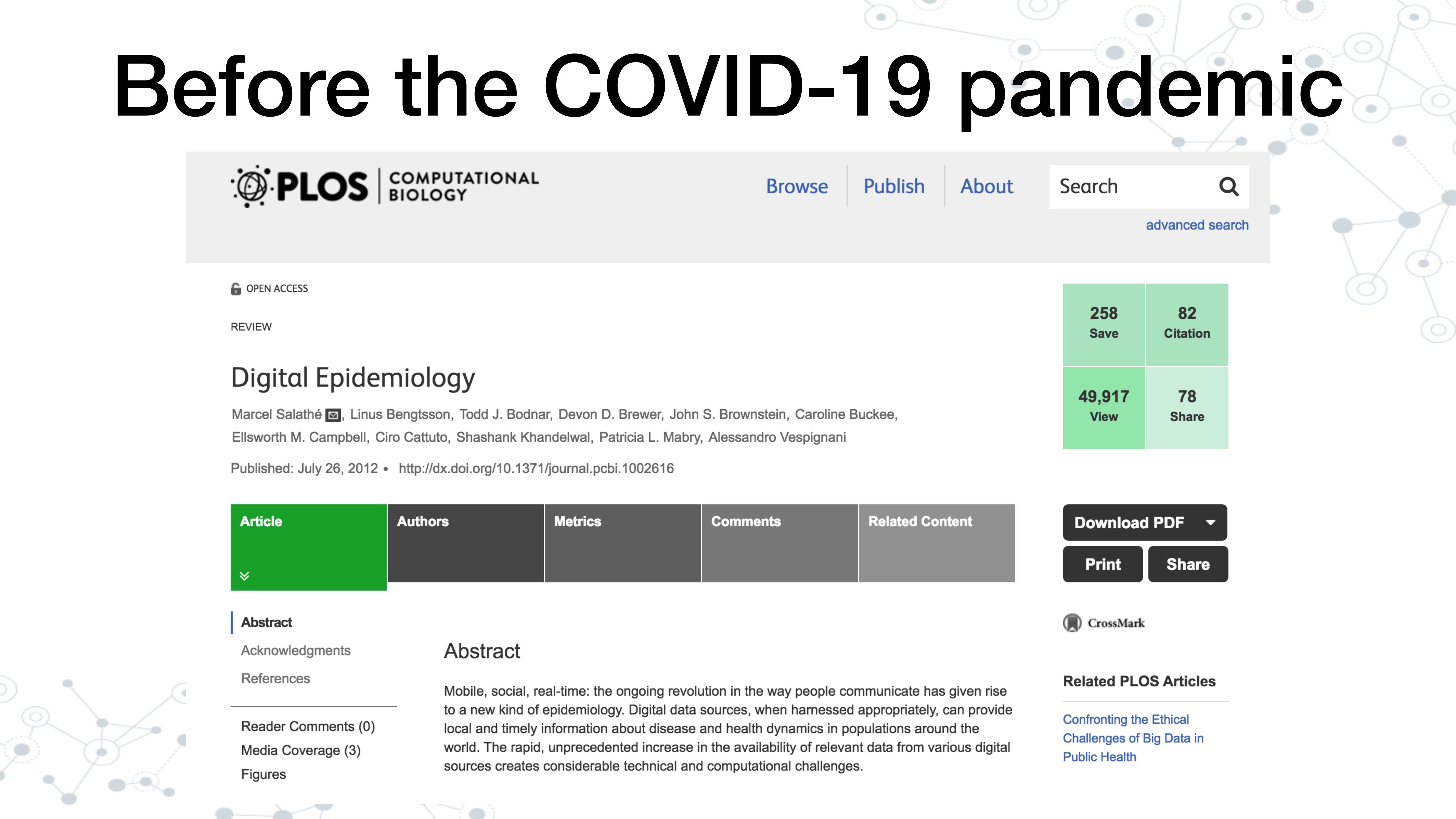
Research milestones

Before the COVID-19 pandemic

After the COVID-19 pandemic



Before the COVID-19 pandemic

A faint, light-gray network graph with numerous small, semi-transparent nodes connected by thin lines, forming a complex web-like structure.

PLOS COMPUTATIONAL BIOLOGY

Browse | Publish | About | Search | advanced search

OPEN ACCESS

REVIEW

Digital Epidemiology

Marcel Salathé , Linus Bengtsson, Todd J. Bodnar, Devon D. Brewer, John S. Brownstein, Caroline Buckee, Ellsworth M. Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L. Mabry, Alessandro Vespignani

Published: July 26, 2012 • <http://dx.doi.org/10.1371/journal.pcbi.1002616>

258 Save	82 Citation
49,917 View	78 Share

Article	Authors	Metrics	Comments	Related Content	Download PDF	Print	Share
---------	---------	---------	----------	-----------------	--------------	-------	-------

Abstract

Acknowledgments

References

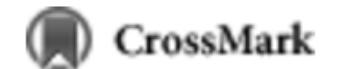
Reader Comments (0)

Media Coverage (3)

Figures

Abstract

Mobile, social, real-time: the ongoing revolution in the way people communicate has given rise to a new kind of epidemiology. Digital data sources, when harnessed appropriately, can provide local and timely information about disease and health dynamics in populations around the world. The rapid, unprecedented increase in the availability of relevant data from various digital sources creates considerable technical and computational challenges.



Related PLOS Articles

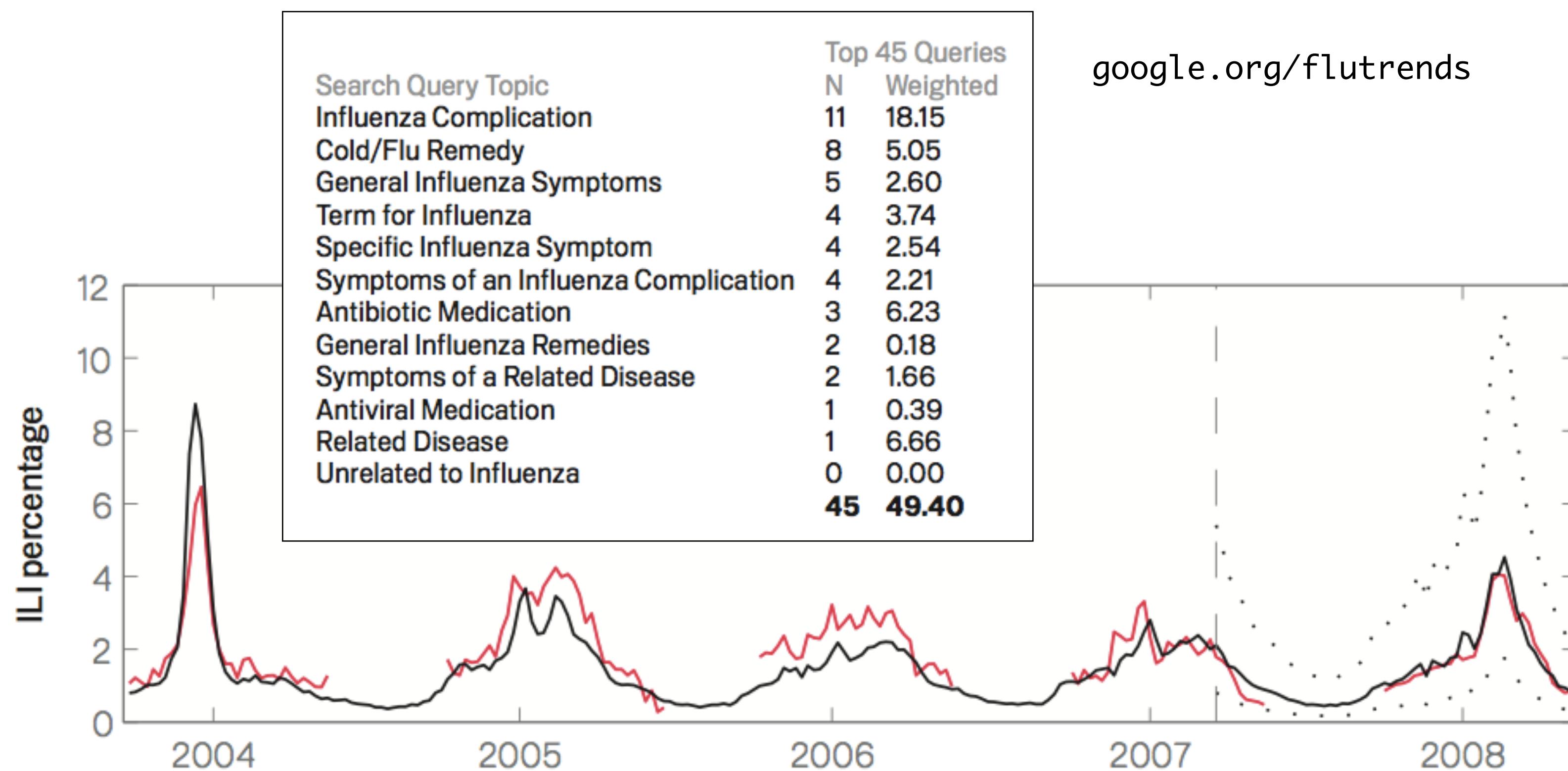
[Confronting the Ethical Challenges of Big Data in Public Health](#)



Detecting influenza epidemics using search engine query data

Jeremy Ginsberg¹, Matthew H. Mohebbi¹, Rajan S. Patel¹, Lynnette Brammer²,
Mark S. Smolinski¹ & Larry Brilliant¹

¹Google Inc. ²Centers for Disease Control and Prevention



google.org/flutrends

[Google.org home](#)

[Dengue Trends](#)

Flu Trends

Home

Select country/region ▾

[How does this work?](#)

[FAQ](#)

Flu activity

Intense

High

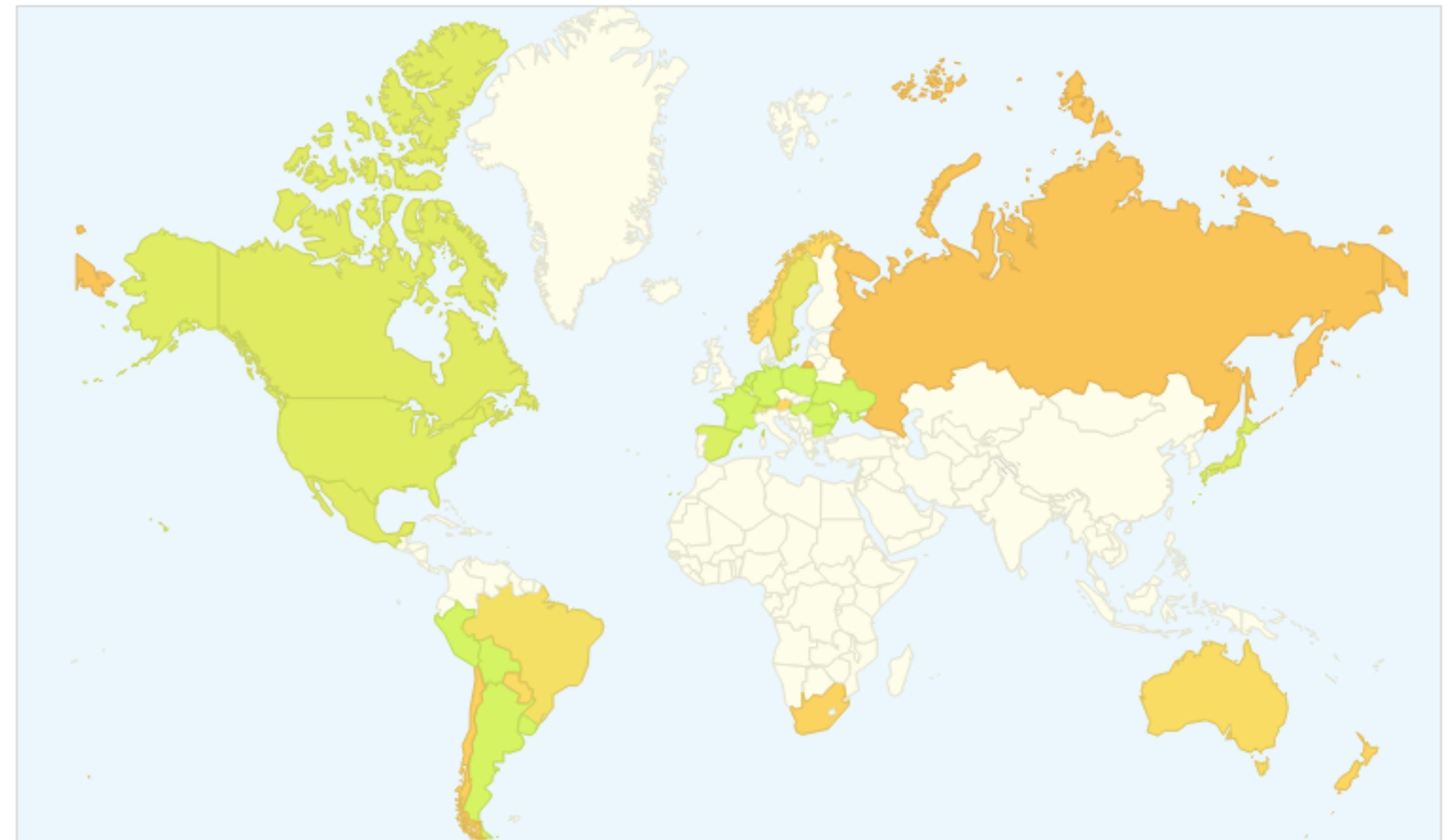
Moderate

Low

Minimal

Explore flu trends around the world

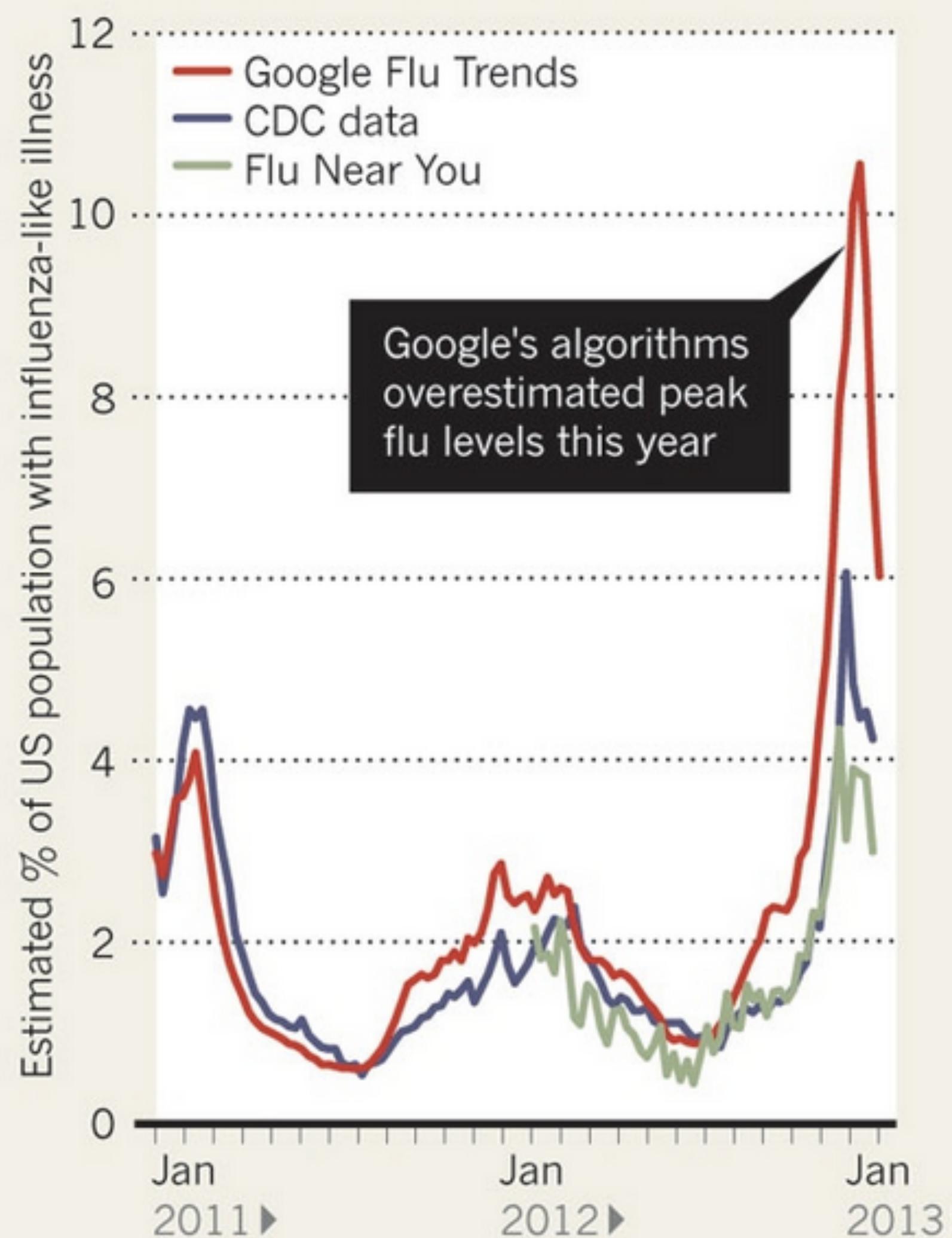
We've found that certain search terms are good indicators of flu activity. Google Flu Trends uses aggregated Google search data to estimate flu activity. [Learn more »](#)



[Download world flu activity data](#) - [Animated flu trends for Google Earth](#) - [Compare flu trends across regions in Public Data Explorer](#)

FEVER PEAKS

A comparison of three different methods of measuring the proportion of the US population with an influenza-like illness.



nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive

Archive > Volume 494 > Issue 7436 > News > Article

NATURE | NEWS

عربي

When Google got flu wrong

US outbreak foxes a leading web-based method for tracking seasonal flu.

Declan Butler

Science 14 March 2014:
Vol. 343 no. 6176 pp. 1203-1205
DOI: 10.1126/science.1248506

FT Magazine

Home

World ▾

Companies ▾

Markets ▾

Global Economy ▾

Lex ▾

Arts ▾

Magazine

Food & Drink ▾

House & Home ▾

Lunch with the FT

Style

Books ▾

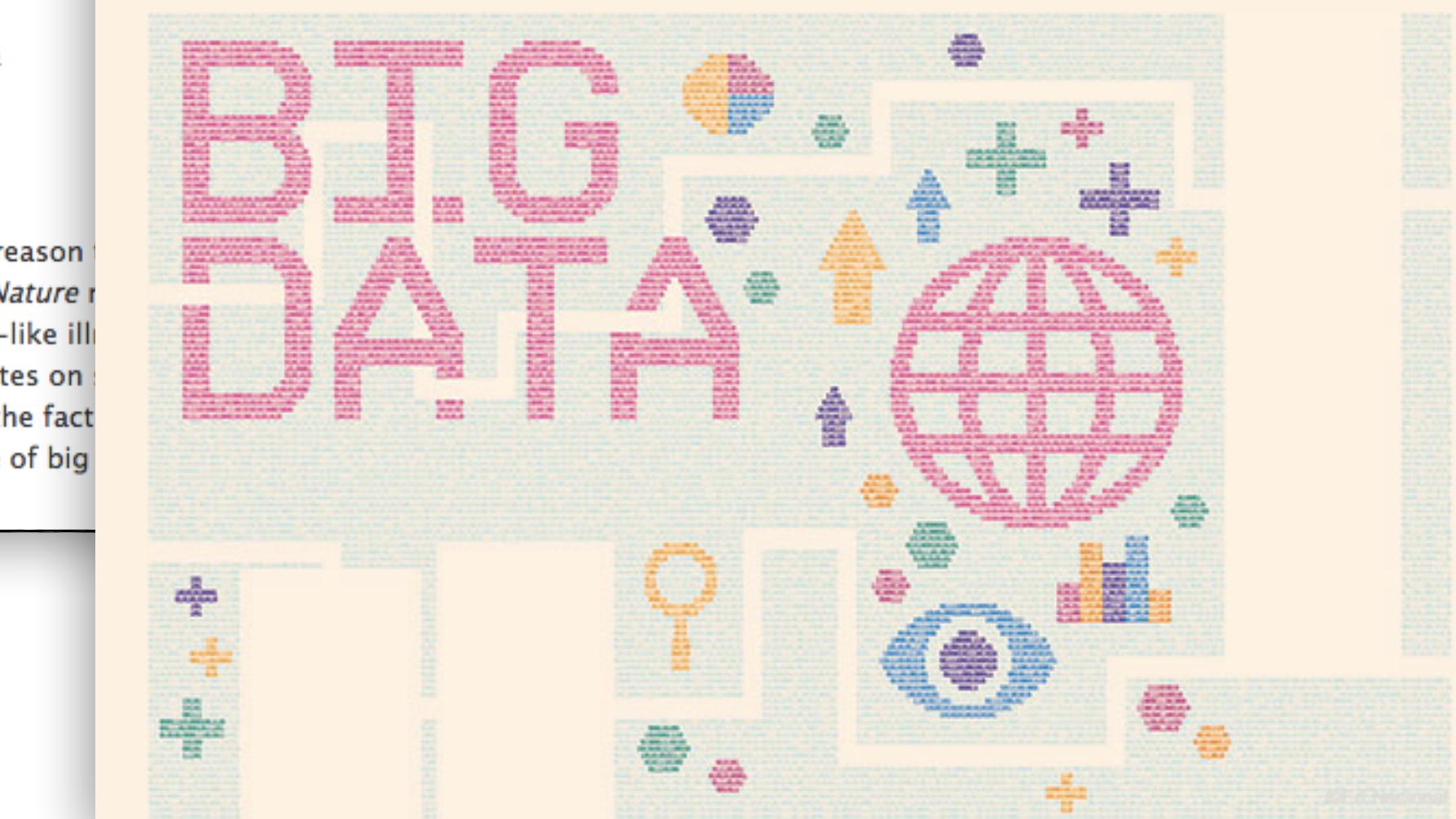
Pursuits

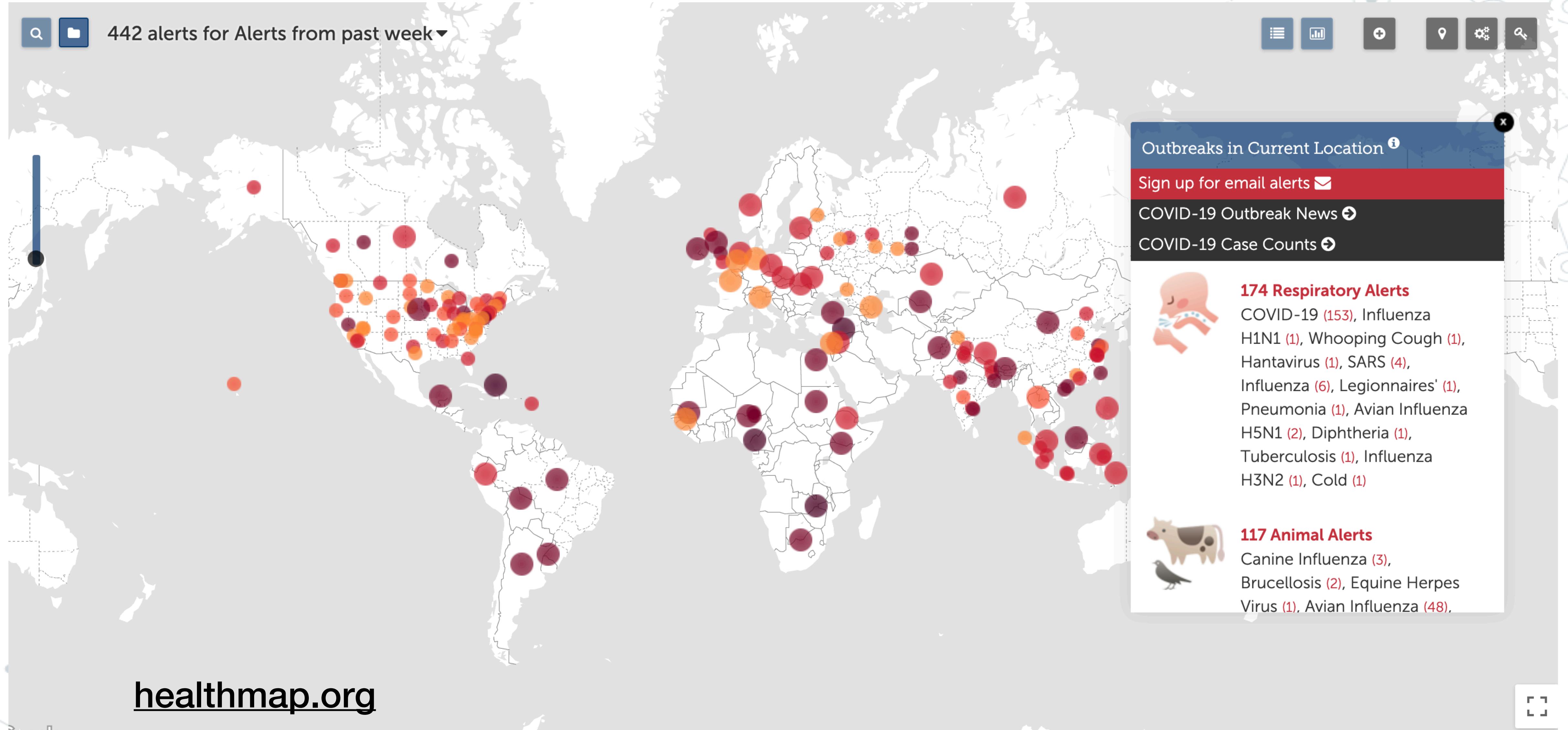
March 28, 2014 11:38 am

Big data: are we making a big mistake?

By Tim Harford

Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media





Google



Before the COVID-19 pandemic

Main areas of research:

- ◆ Forecasting disease incidence and prevalence using non-traditional data sources (web search queries, social media messages, Wikipedia page views)
- ◆ Detecting health relevant attitudes and behaviors from social media and Web sources
- ◆ Participatory surveillance: users' active reporting of health conditions
- ◆ Using mobile phone data to measure human movements relevant to disease transmission
- ◆ (And many others...)

Disease monitoring/forecasting

OPEN  ACCESS Freely available online



Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time

David J. McIver*, John S. Brownstein

Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

Abstract

Circulating levels of both seasonal and pandemic influenza require constant surveillance to ensure the health and safety of the population. While up-to-date information is critical, traditional surveillance systems can have data availability lags of up to two weeks. We introduce a novel method of estimating, in near-real time, the level of influenza-like illness (ILI) in the United States (US) by monitoring the rate of particular Wikipedia article views on a daily basis. We calculated the number of times certain influenza- or health-related Wikipedia articles were accessed each day between December 2007 and August 2013 and compared these data to official ILI activity levels provided by the Centers for Disease Control and Prevention (CDC). We developed a Poisson model that accurately estimates the level of ILI activity in the American population, up to two weeks ahead of the CDC, with an absolute average difference between the two estimates of just 0.27% over 294 weeks of data. Wikipedia-derived ILI models performed well through both abnormally high media coverage events (such as during the 2009 H1N1 pandemic) as well as unusually severe influenza seasons (such as the 2012–2013 influenza season). Wikipedia usage accurately estimated the week of peak ILI activity 17% more often than Google Flu Trends data and was often more accurate in its measure of ILI intensity. With further study, this method could potentially be implemented for continuous monitoring of ILI activity in the US and to provide support for traditional influenza surveillance tools.

Citation: McIver DJ, Brownstein JS (2014) Wikipedia Usage Estimates Prevalence of Influenza-Like Illness in the United States in Near Real-Time. PLoS Comput Biol 10(4): e1003581. doi:10.1371/journal.pcbi.1003581

Editor: Marcel Salathé, Pennsylvania State University, United States of America

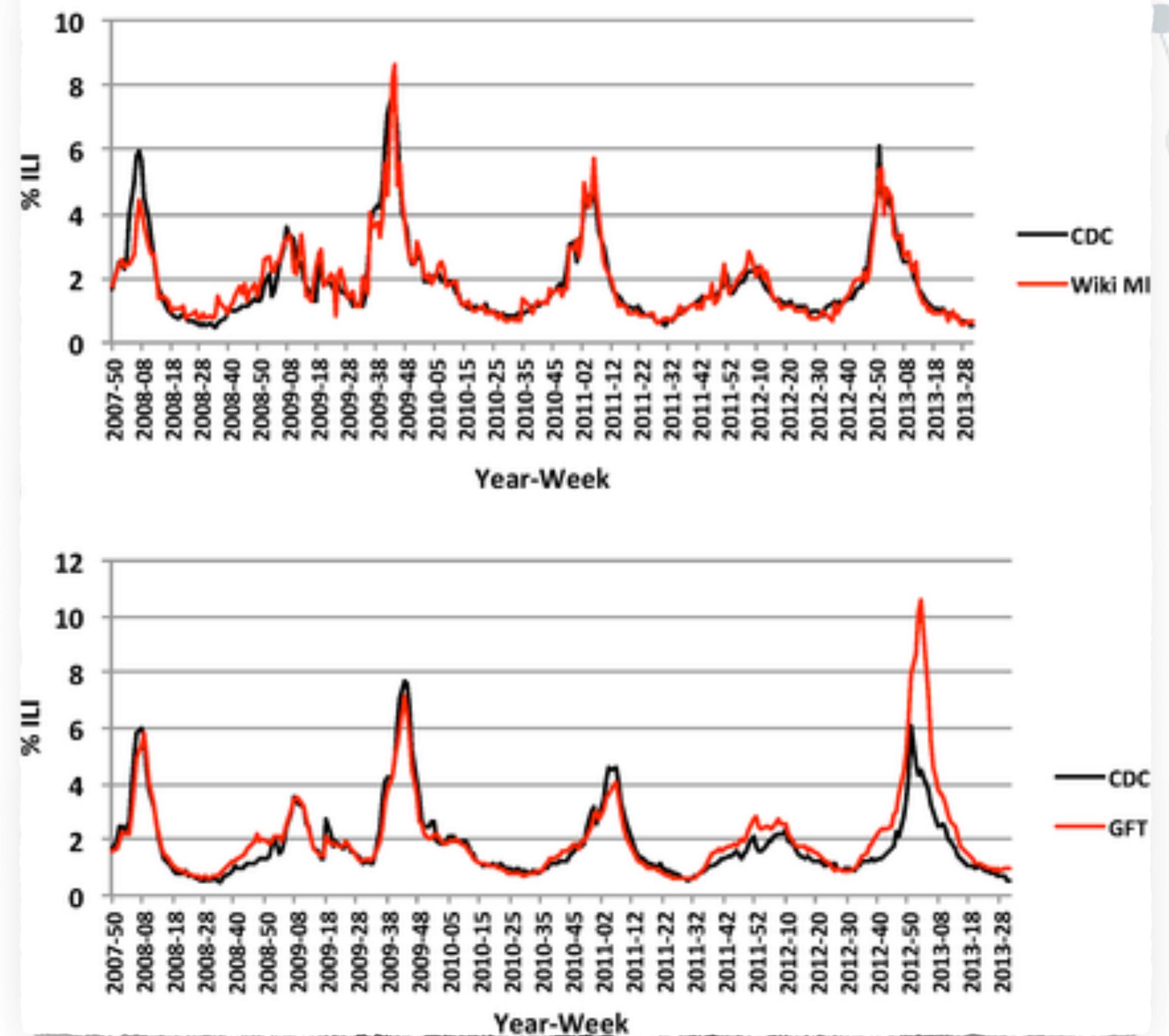
Received December 20, 2013; **Accepted** March 11, 2014; **Published** April 17, 2014

Copyright: © 2014 McIver, Brownstein. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was funded by the National Institutes of Health and National Library of Medicine 1R01LM010812-03. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: david.mciver@childrens.harvard.edu



Attitudes and behaviors

OPEN  ACCESS Freely available online

PLOS COMPUTATIONAL BIOLOGY

Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control

Marcel Salathé*, Shashank Khandelwal

Center for Infectious Disease Dynamics, Department of Biology, Penn State University, University Park, Pennsylvania, United States of America

Abstract

There is great interest in the dynamics of health behaviors in social networks and how they affect collective public health outcomes, but measuring population health behaviors over time and space requires substantial resources. Here, we use publicly available data from 101,853 users of online social media collected over a time period of almost six months to measure the spatio-temporal sentiment towards a new vaccine. We validated our approach by identifying a strong correlation between sentiments expressed online and CDC-estimated vaccination rates by region. Analysis of the network of opinionated users showed that information flows more often between users who share the same sentiments - and less often between users who do not share the same sentiments - than expected by chance alone. We also found that most communities are dominated by either positive or negative sentiments towards the novel vaccine. Simulations of infectious disease transmission show that if clusters of negative vaccine sentiments lead to clusters of unprotected individuals, the likelihood of disease outbreaks is greatly increased. Online social media provide unprecedented access to data allowing for inexpensive and efficient tools to identify target areas for intervention efforts and to evaluate their effectiveness.

Citation: Salathé M, Khandelwal S (2011) Assessing Vaccination Sentiments with Online Social Media: Implications for Infectious Disease Dynamics and Control. PLoS Comput Biol 7(10): e1002199. doi:10.1371/journal.pcbi.1002199

Editor: Lauren Ancel Meyers, University of Texas at Austin, United States of America

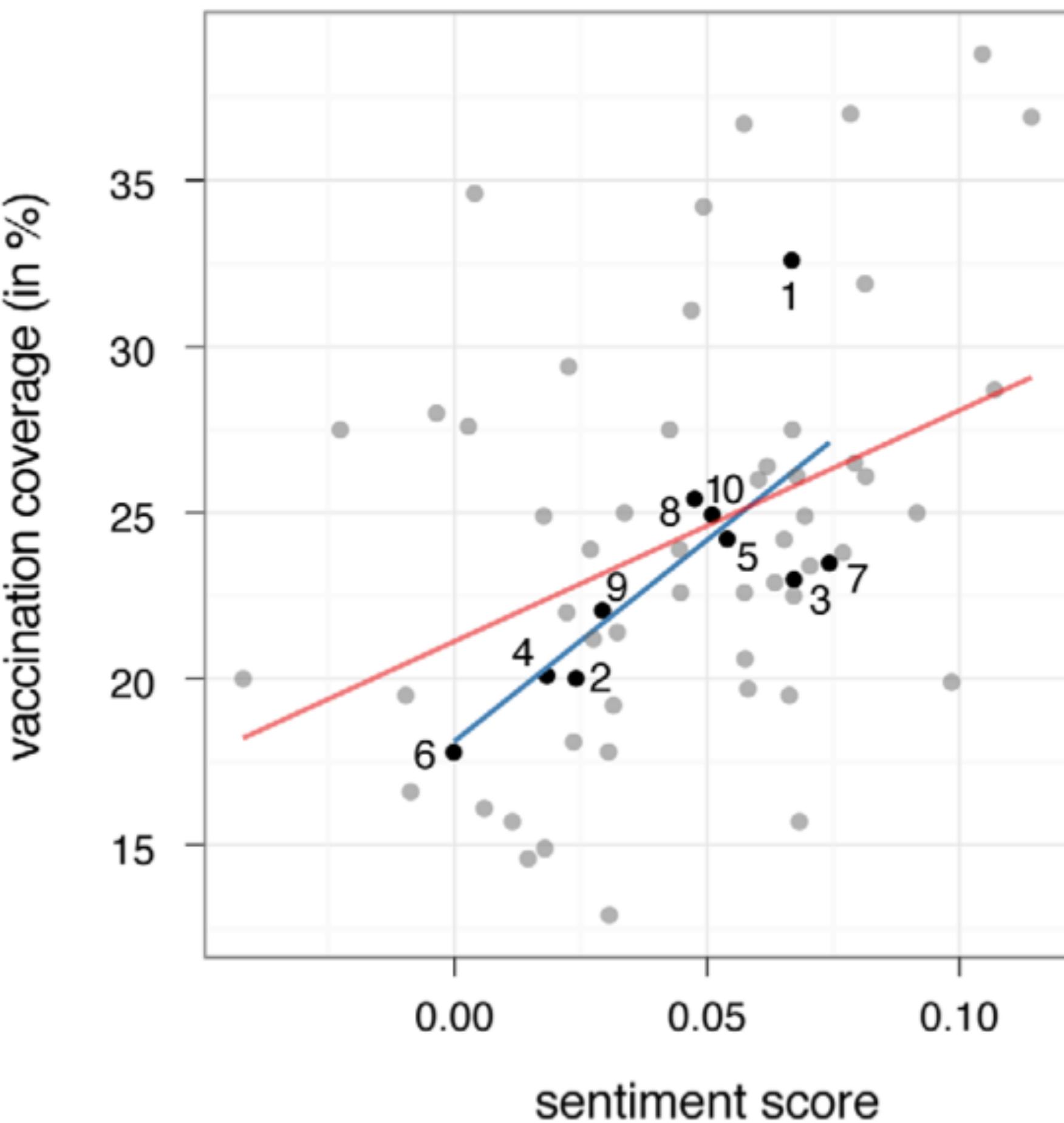
Received May 10, 2011; Accepted July 30, 2011; Published October 13, 2011

Copyright: © 2011 Salathé, Khandelwal. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: MS acknowledges funding from Society in Science: the Branco Weiss fellowship. <http://www.society-in-science.org/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: salathe@psu.edu



Participatory surveillance

The image shows the Influweb homepage. At the top, there's a navigation bar with links for Home, Il progetto Influweb, FAQ, Risultati, Entrà, and Registrati. A language selector "IT ▾" is also present. The main content area features a large image of a woman and a child. On the left, a white box contains the text "Benvenuto in Influweb" and "Aiutaci a monitorare il COVID-19 e l'influenza in Italia iscrivendoti al nostro studio". In the center, there's a "Outbreaks Near Me" section with a map of the United States and a count of "A community of **7,166,099** people tracking local COVID-19 and flu outbreaks". Below this is a large question "How are you feeling?". At the bottom, there are two buttons: a teal one for "Healthy, thanks!" and a red one for "Not feeling well". The background of the page features a light gray network graph. In the bottom right corner, there are logos for Boston Children's Hospital and Harvard Medical School.

Influweb

Home Il progetto Influweb FAQ Risultati Entrà Registrati IT ▾

Benvenuto in Influweb
Aiutaci a monitorare il COVID-19 e l'influenza in Italia iscrivendoti al nostro studio

Outbreaks Near Me United States (English) ▾

A community of **7,166,099** people tracking local COVID-19 and flu outbreaks.

How are you feeling?

Healthy, thanks! Not feeling well

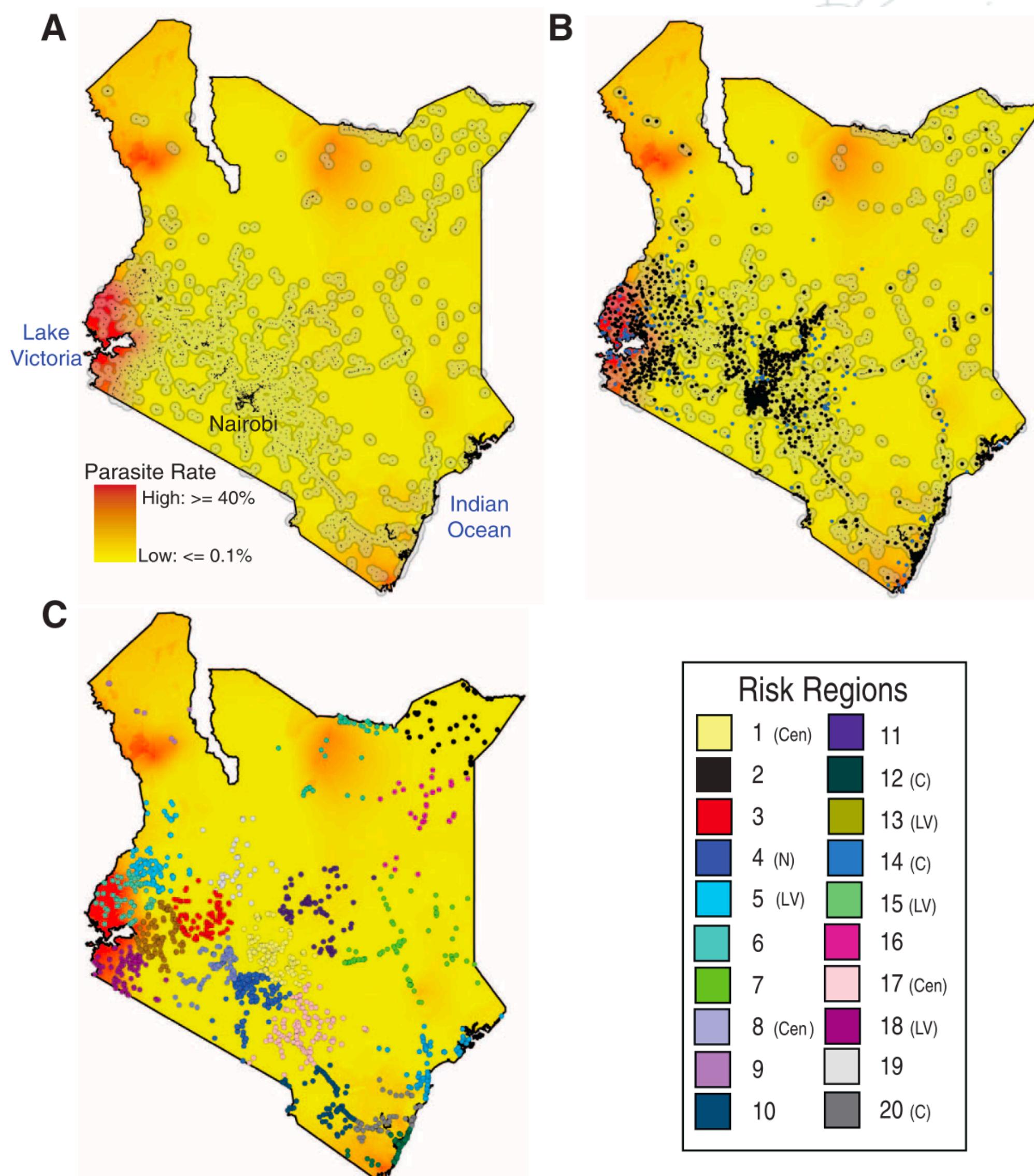
Boston Children's Hospital Where the world comes for answers HARVARD MEDICAL SCHOOL

Mobile phone data

Quantifying the Impact of Human Mobility on Malaria

Amy Wesolowski,^{1,2} Nathan Eagle,^{3,4} Andrew J. Tatem,^{5,6,7} David L. Smith,^{6,8} Abdisalan M. Noor,^{9,10} Robert W. Snow,^{9,10} Caroline O. Buckee^{4,11*}

Human movements contribute to the transmission of malaria on spatial scales that exceed the limits of mosquito dispersal. Identifying the sources and sinks of imported infections due to human travel and locating high-risk sites of parasite importation could greatly improve malaria control programs. Here, we use spatially explicit mobile phone data and malaria prevalence information from Kenya to identify the dynamics of human carriers that drive parasite importation between regions. Our analysis identifies importation routes that contribute to malaria epidemiology on regional spatial scales.



After the COVID-19 pandemic

The COVID-19 pandemic

- For the first time in history we could monitor in almost real-time behavioral changes during an outbreak
- Real-time analysis of changes in movements across regions and within a region
- Quantitative measure of social-distancing and physical proximity
- Changes in mobility patterns depending on the location, the time of the day, the day of the week, and socio-demographic factors



Human mobility



See how your community is moving around differently due to COVID-19

As global communities respond to COVID-19, we've heard from public health officials that the same type of aggregated, anonymized insights we use in products such as Google Maps could be helpful as they make critical decisions to combat COVID-19.

These Community Mobility Reports aim to provide insights into what has changed in response to policies aimed at combating COVID-19. The reports chart movement trends over time by geography, across different categories of places such as retail and recreation, groceries and pharmacies, parks, transit stations, workplaces, and residential.



Mobility Trends Reports

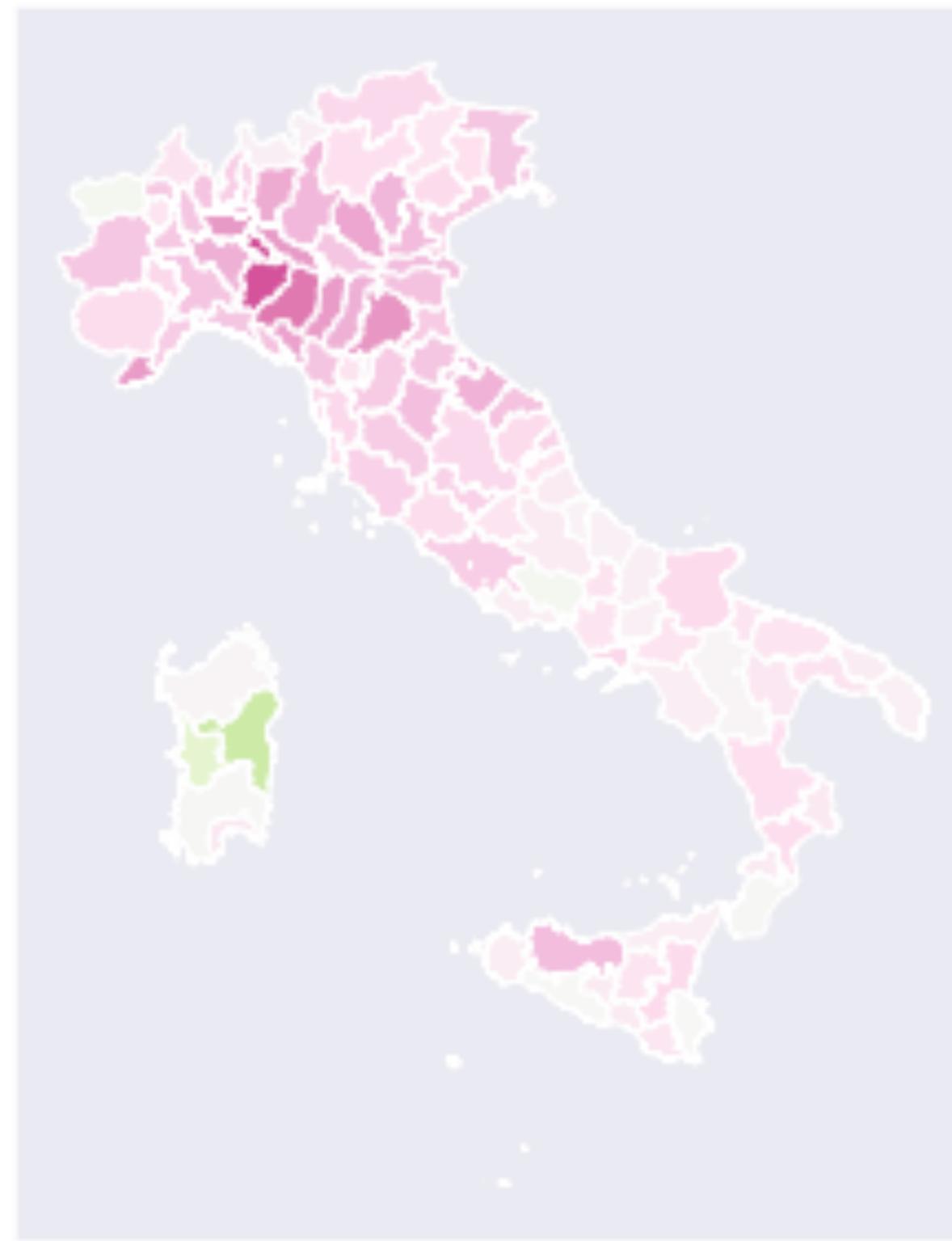
Learn about COVID-19 mobility trends. Reports are published daily and reflect requests for directions in Apple Maps. Privacy is one of our core values, so Maps doesn't associate your data with your Apple ID, and Apple doesn't keep a history of where you've been.



Human mobility



February 21-28

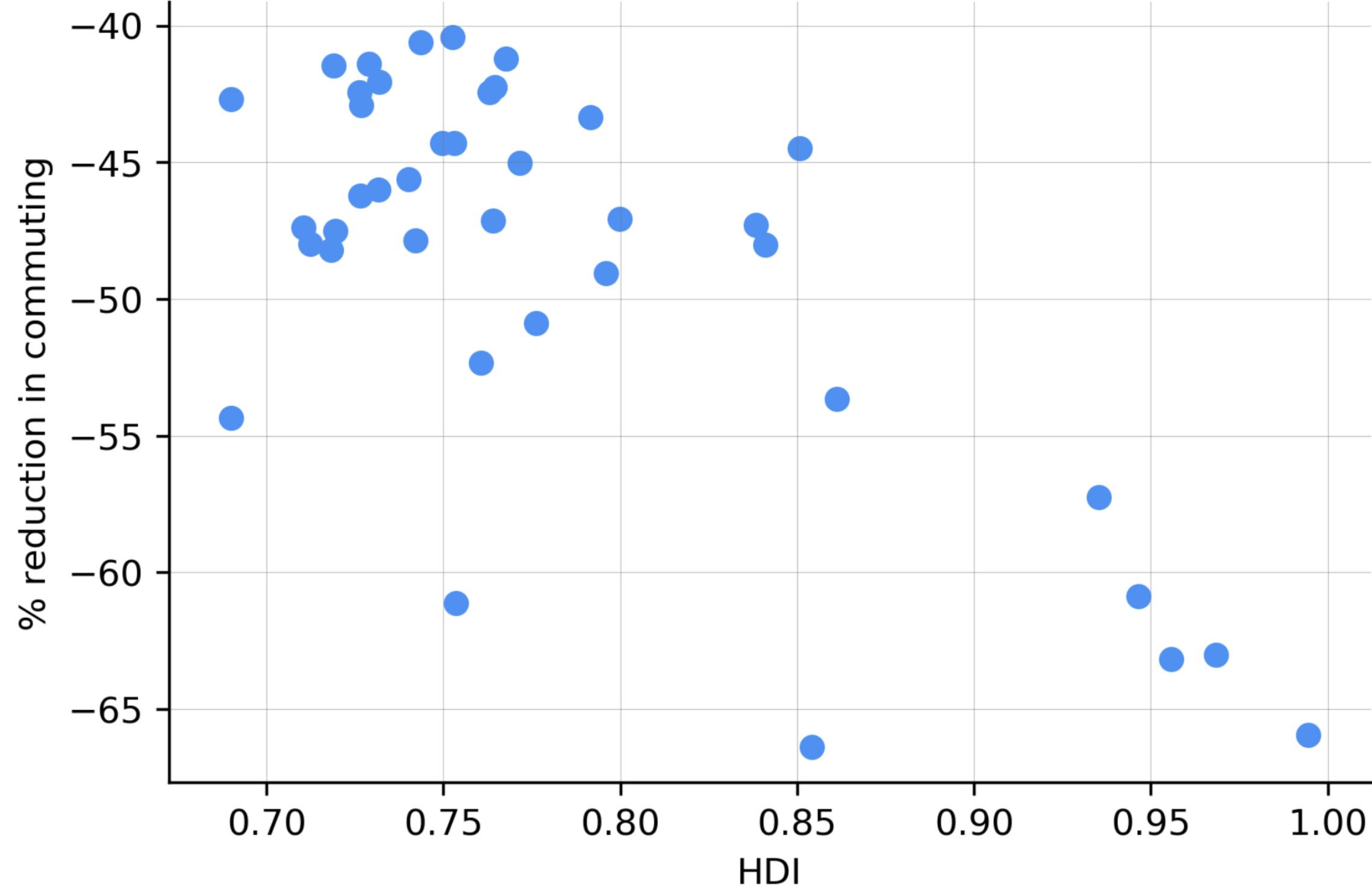
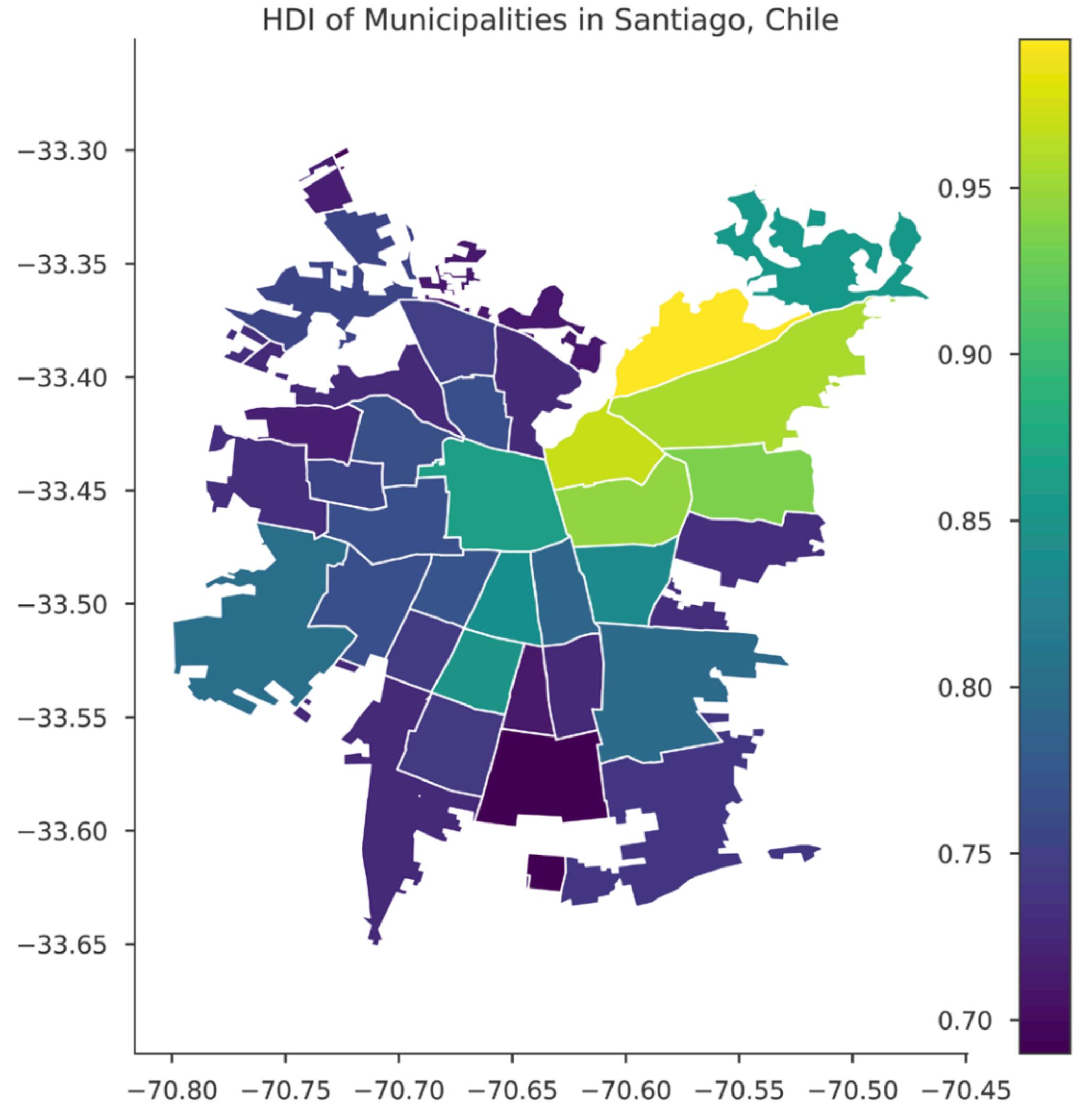


February 29 - March 6



March 7 - 13

Socio-economic effects



Digital contact tracing

RESEARCH

RESEARCH ARTICLE SUMMARY

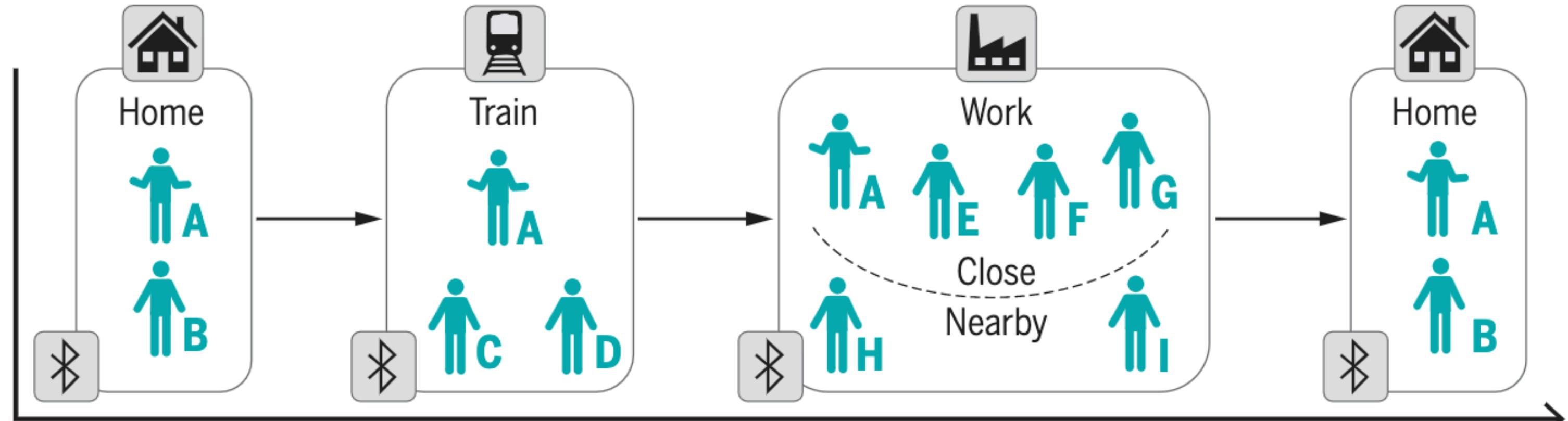
CORONAVIRUS

Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing

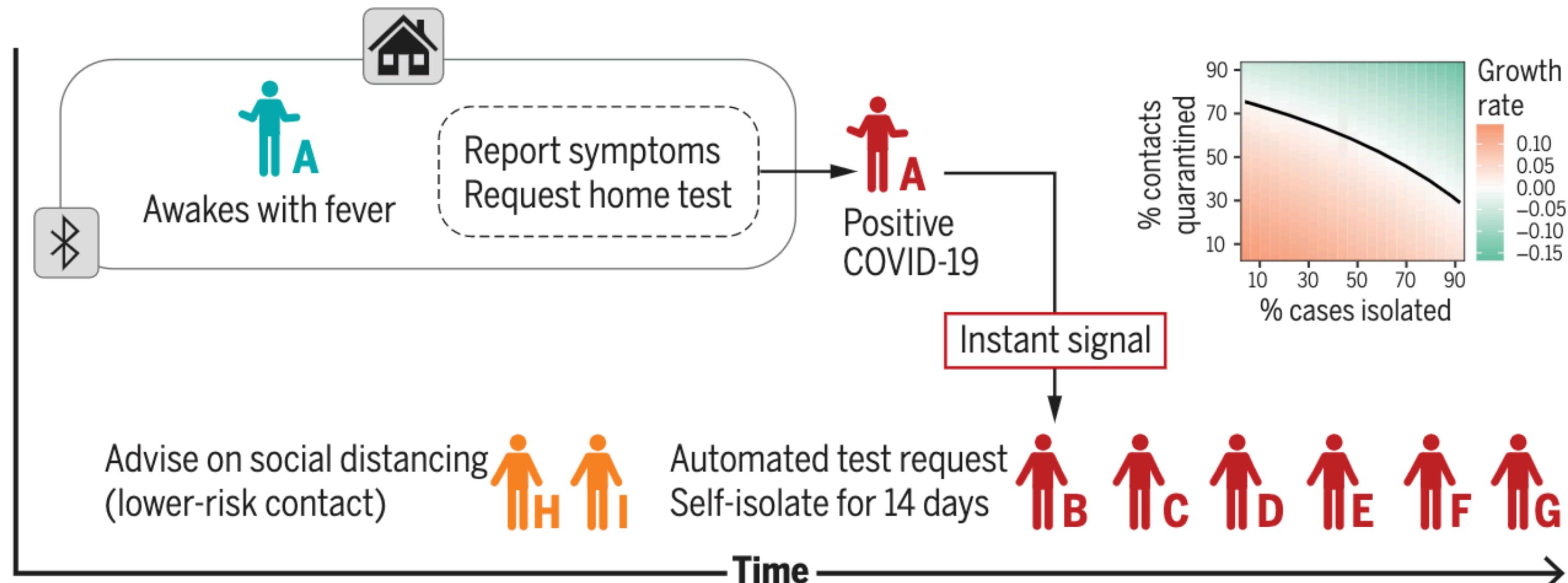
Luca Ferretti*, Chris Wymant*, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall†, Christophe Fraser†‡

Subject has COVID-19 infection. No symptoms

Day 1



Day 2



Computational modeling

nature

Explore content ▾ About the journal ▾ Publish with us ▾

nature > articles > article

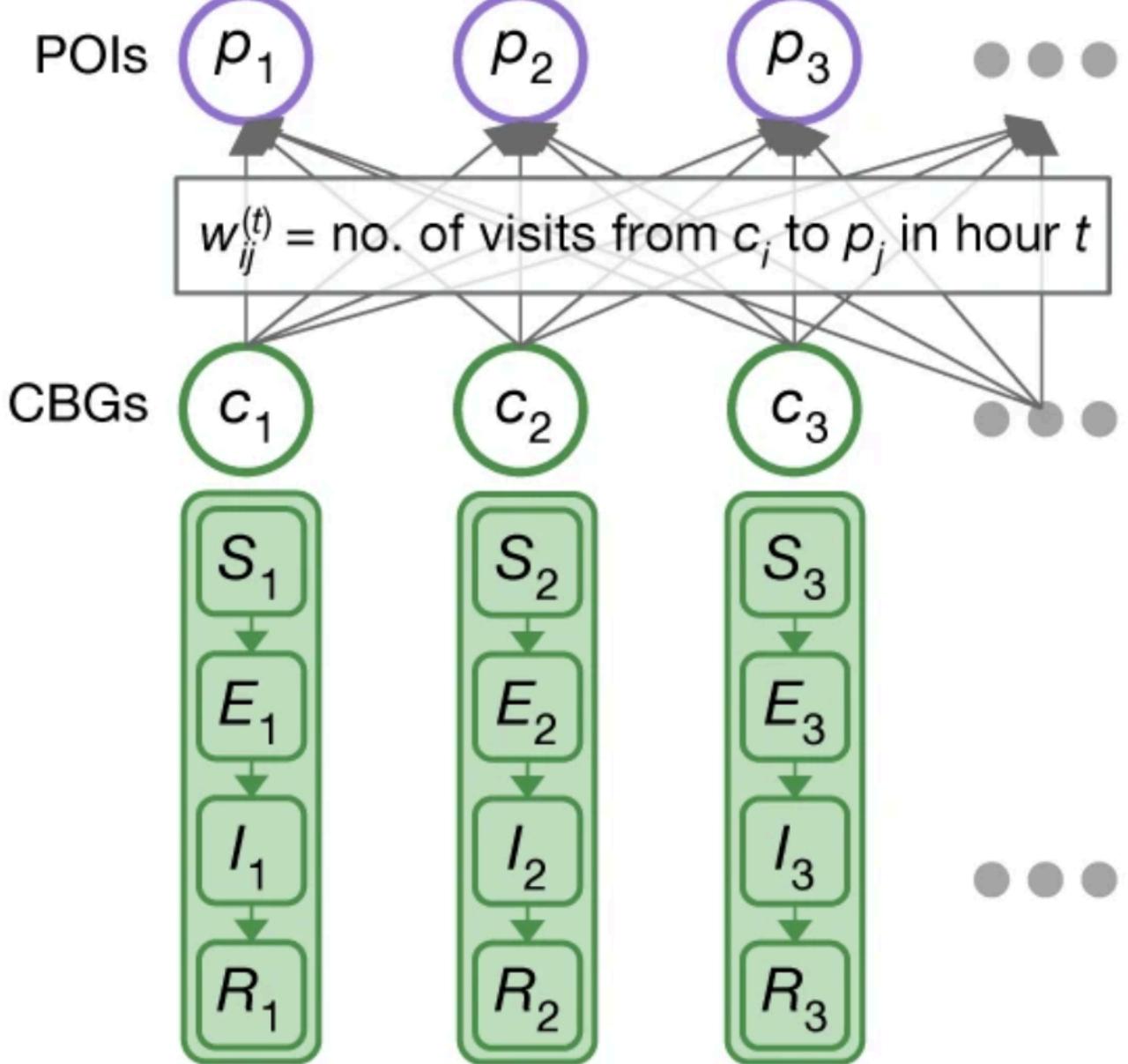
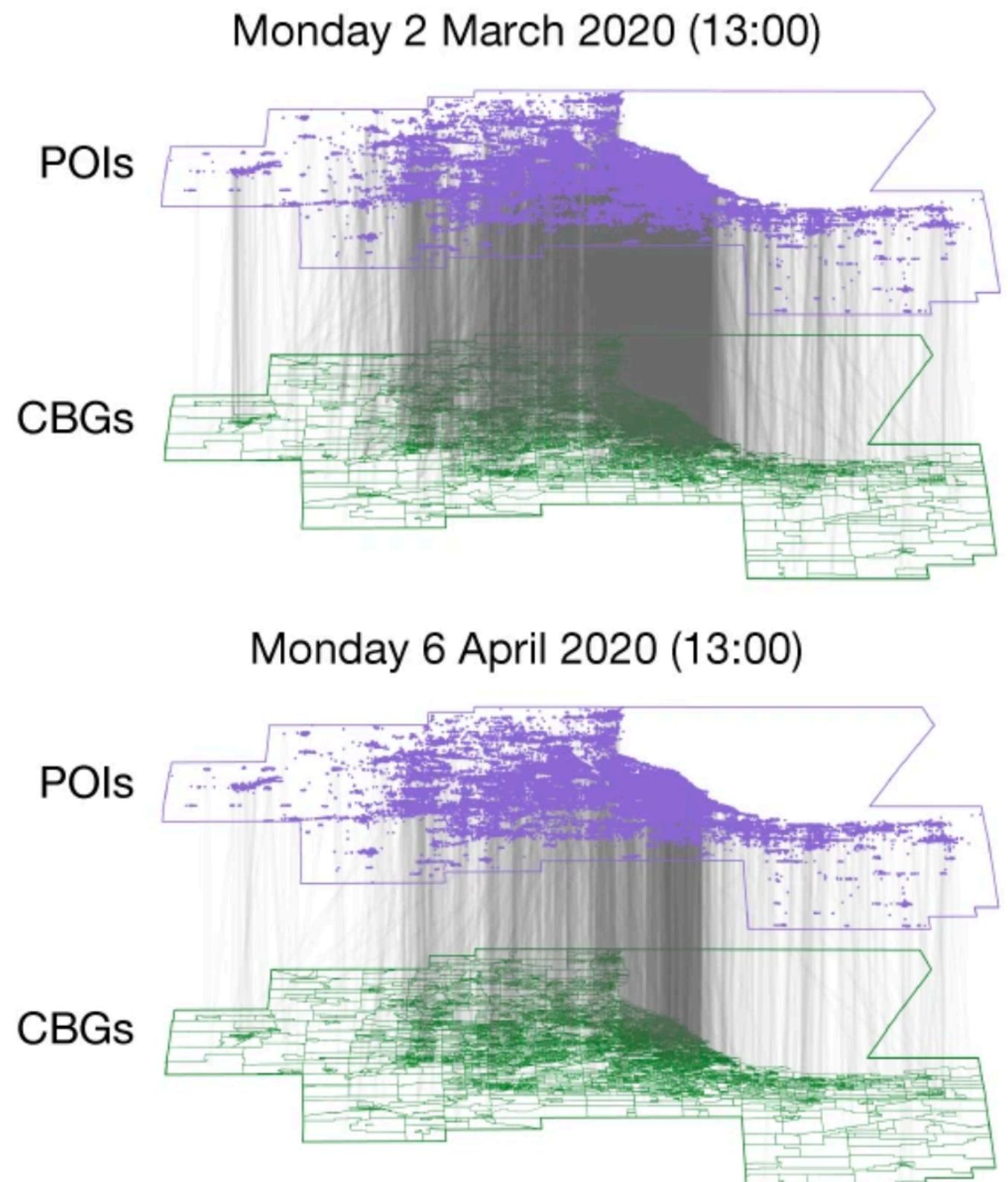
Article | Published: 10 November 2020

Mobility network models of COVID-19 explain inequities and inform reopening

Serina Chang, Emma Pierson, Pang Wei Koh, Jaline Gerardin, Beth Redbird, David Grusky & Jure Leskovec 

Nature 589, 82–87 (2021) | Cite this article

609k Accesses | 561 Citations | 11199 Altmetric | Metrics



Computational modeling

nature human behaviour

Explore content ▾ About the journal ▾ Publish with us ▾

[nature](#) > [nature human behaviour](#) > [articles](#) > [article](#)

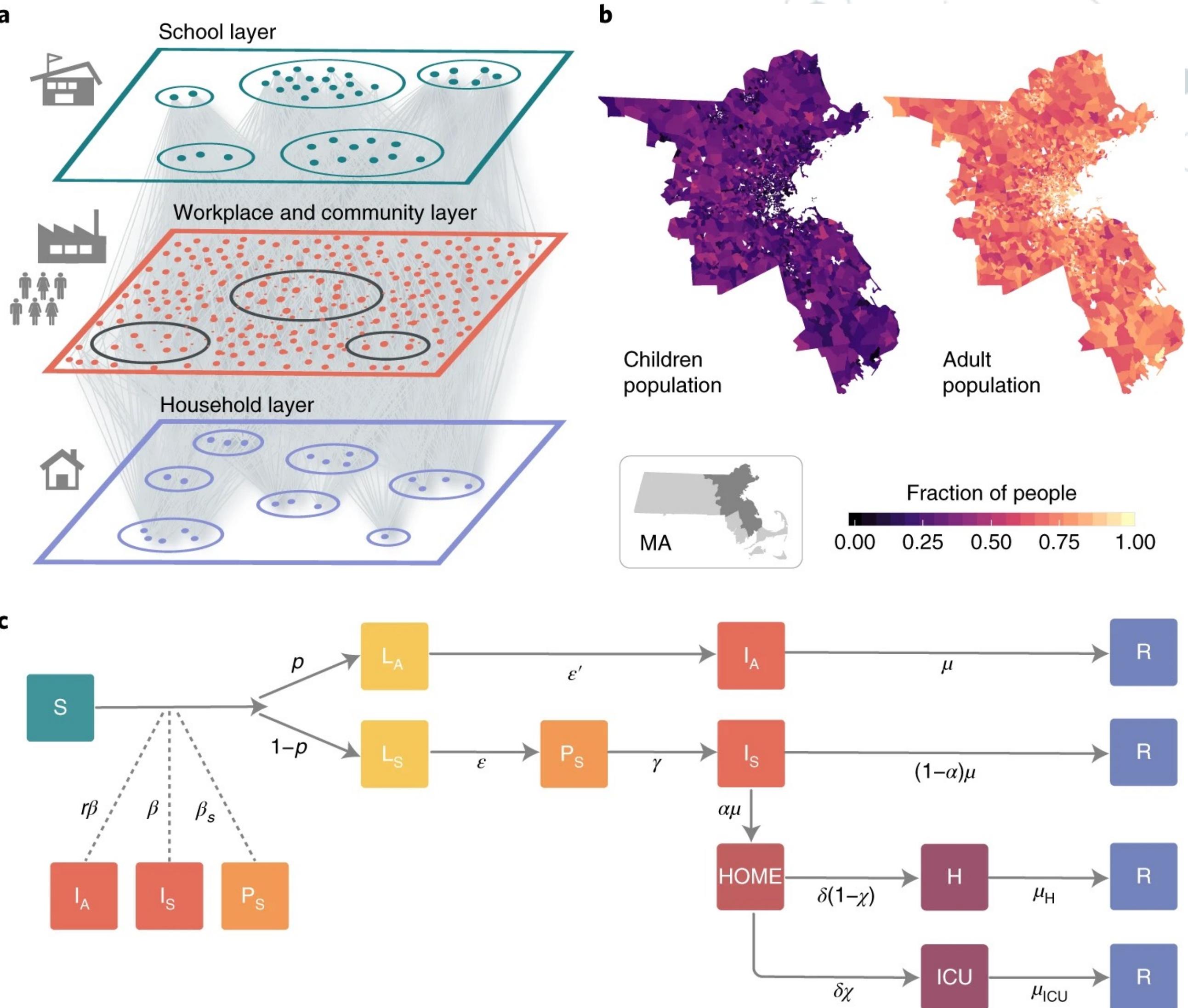
Article | Published: 05 August 2020

Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19

Alberto Aleta, David Martín-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E. Dean, M. Elizabeth Halloran, Ira M. Longini Jr, Stefano Merler, Alex Pentland, Alessandro Vespiagnani  Esteban Moro  & Yamir Moreno 

[Nature Human Behaviour](#) 4, 964–971 (2020) | [Cite this article](#)

41k Accesses | 358 Citations | 1103 Altmetric | [Metrics](#)



Setting up the Python environment

Requirements

- ◆ Python 3.x
- ◆ Jupyter notebooks
- ◆ Standard Python library, NumPy, SciPy
- ◆ NetworkX
- ◆ Geopandas