



## OPEN ACCESS

## EDITED BY

Huan Liu,  
Arizona State University, United States

## REVIEWED BY

Mohsen Bahrami,  
Massachusetts Institute of Technology,  
United States

Maria Jofre,  
Catholic University of the Sacred Heart,  
Milan, Italy

Federico Battiston,  
Central European University, Hungary

Adriana Manna,  
Central European University Wien, Austria, in  
collaboration with reviewer FB

## \*CORRESPONDENCE

Bruno Lepri  
✉ lepri@fbk.eu

RECEIVED 15 December 2022

ACCEPTED 10 May 2023

PUBLISHED 25 May 2023

## CITATION

Luca M, Campedelli GM, Centellegher S,  
Tizzoni M and Lepri B (2023) Crime, inequality  
and public health: a survey of emerging trends  
in urban data science.

*Front. Big Data* 6:1124526.

doi: 10.3389/fdata.2023.1124526

## COPYRIGHT

© 2023 Luca, Campedelli, Centellegher,  
Tizzoni and Lepri. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](#). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that  
the original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Crime, inequality and public health: a survey of emerging trends in urban data science

Massimiliano Luca<sup>1,2</sup>, Gian Maria Campedelli<sup>3</sup>,  
Simone Centellegher<sup>1</sup>, Michele Tizzoni<sup>3</sup> and Bruno Lepri<sup>1\*</sup>

<sup>1</sup>Mobile and Social Computing Lab, Bruno Kessler Foundation, Trento, Italy, <sup>2</sup>Faculty of Computer Science, Free University of Bolzano, Bolzano, Italy, <sup>3</sup>Department of Sociology and Social Research, University of Trento, Trento, Italy

Urban agglomerations are constantly and rapidly evolving ecosystems, with globalization and increasing urbanization posing new challenges in sustainable urban development well summarized in the United Nations' Sustainable Development Goals (SDGs). The advent of the digital age generated by modern alternative data sources provides new tools to tackle these challenges with spatio-temporal scales that were previously unavailable with census statistics. In this review, we present how new digital data sources are employed to provide data-driven insights to study and track (i) urban crime and public safety; (ii) socioeconomic inequalities and segregation; and (iii) public health, with a particular focus on the city scale.

## KEYWORDS

cities, crime, segregation and inequalities, public health, digital data

## 1. Introduction

Cities occupy only 3% of the global surface but are inhabited by more than 50% of the world's population.<sup>1</sup> Timely and accurate data are thus becoming fundamental for policymakers and municipalities to control cities' dynamics and respond to multiple societal challenges. In 2015, the United Nations set out 17 Sustainable Development Goals (SDGs)<sup>2</sup> that summarize the new challenges we have to face to guarantee everyone a better and more sustainable future. Examples of such goals are about guaranteeing good quality of (accessible) health and wellbeing, reduction of inequalities, and design of sustainable and safe cities and communities.

It is clear that big urban agglomerations have a pivotal role in the accomplishment of such goals as many of them are fundamentally related to human movements, displacement, and interactions (Glaeser, 2012; Sassen, 2019). More in general, it is known that human dynamics are related to the diffusion of viral diseases (Eubank et al., 2004; Colizza et al., 2007; Perkins et al., 2014), to the behavioral responses in case of natural disasters (Bohorquez et al., 2009; Bagrow et al., 2011), to the optimization of traffic volumes (Batty, 2013; Mazzoli et al., 2019), to the economic growth, innovation and social integration (Bettencourt et al., 2007; Pan et al., 2013; Schläpfer et al., 2014), and to the severity of air pollution and the consumption of energy, water and other resources (Bettencourt et al., 2007; Bettencourt and West, 2010).

1 <https://www.un.org/sustainabledevelopment/cities/>

2 <https://sdgs.un.org/goals>

To monitor the progress toward the aforementioned societal challenges, it is fundamental to have an always up-to-date picture of cities. In the past, institutions had to rely almost exclusively on census data and official statistics. However, both these data sources have some intrinsic limitations including (i) the time gap between the data collection and the actual availability of the data, and (ii) the frequencies and costs of the data collection campaigns (Lazer et al., 2009). Luckily, we are in the middle of a digital sensing revolution with billions of data that are generated every second and that can be employed to have an almost real-time picture of cities' dynamics at low costs. Examples of such data include tracks from GPS devices embedded in smartphones, vehicles, or boats, records produced by the communication between phones and the cellular network, and geotagged posts from social media platforms (Gonzalez et al., 2008; Zheng et al., 2010; Moreira-Matias et al., 2013; Spinsanti et al., 2013; Blondel et al., 2015; Cui et al., 2018). Finally, other data sources like satellite images, street networks and points of interest can provide precious information to integrate with human dynamics data in order to capture socio-economic aspects (Deville et al., 2014; Jean et al., 2016; Tatem, 2017; Weber et al., 2018; Yeh et al., 2020; Lepri et al., 2022).

In this review paper, we showcase and discuss how alternative data sources have been employed by researchers to study the relationship between human dynamics and three SDGs: (i) crime diffusion and public safety, (ii) socio-economic inequalities and segregation, and (iii) public health and disease diffusion. Also, we have decided to focus on the studies that investigate such dynamics in urban agglomerations. Thus, excluding studies that, for example, investigate the relationship between international mobility and the global diffusion of diseases, or map the socio-economic inequalities across countries.

Figure 1 provides a visual representation of the main topics covered in this review, highlighting how these topics are directly related to a set of SDGs. The main objective of this review paper is to examine how researchers make use of these novel digital data sources to develop new computational models and to derive insights that address important issues related to Urban Crime, Inequality and Public Health in urban environments.

The paper is structured as follows. We start in Section 2 by first describing the new sources of data used in these lines of research. Then, in Section 3, we discuss the implication of using computational methods for studying urban crime and public safety. After showing how researchers have used data sources like mobile phone data, social media, and others for a variety of aspects related to crime and security, we conclude the Section with a critical reflection. In Section 4 and Section 5, we follow a similar structure covering works related to socio-economic inequalities, segregation, and public health. Finally, in Section 6 we conclude the paper with a brief discussion.

## 2. Data sources

The digital age has revolutionized the world we live in, and simple actions such as clicking on a website, sending an email, paying with a credit card or making a phone call generate so-called digital traces. Digital traces track information about our daily behaviors and in the last decades this rich and vast

amount of information created new research opportunities to better understand and study human behavior (Lazer et al., 2009, 2020). In this section we describe the most common type of data used in this line of research.

### 2.1. Call detail records

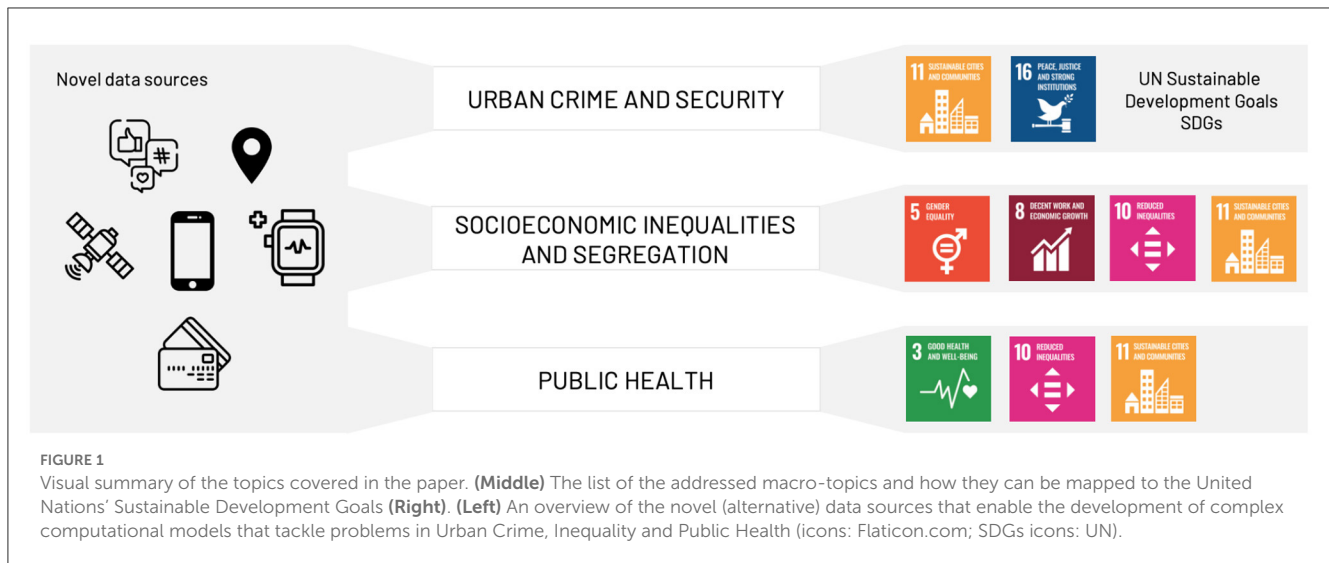
Telecommunication companies collect information regarding people's exchanges by means of Call Detail Records (CDRs), which contain real-world observations on how, when and with whom a person communicates (Blondel et al., 2015; Luca et al., 2021). A Call Detail Record is a tuple  $(u_o, u_i, t, d, A_o, A_i)$  which contains privacy-enhanced metadata about the caller  $u_o$ , the callee  $u_i$ , the timestamp  $t$  of when the call took place and its duration  $d$ .  $A_o$  and  $A_i$  represent respectively the outgoing and incoming Radio Base Stations (RBSs), namely the antennas that delivered the communication through the network. Mobile phone data may cover a large sample size on a national scale and aggregated mobility flows have been inferred by counting the number of users that move between RBSs or administrative units such as neighborhoods and municipalities (Calabrese et al., 2011). Since the position and the coverage area of each RBS are known, a user's telecommunication event represents a proxy of the user's geographic location. The precision of this location can vary from 100 m in urban areas to kilometers in rural areas. This approximation implies that the user's spatial and temporal resolution is given by where and when a user makes a call or sends an SMS leading to sparse and incomplete mobility trajectories. Nevertheless, CDRs have been proven valuable in studying and understanding human mobility (Gonzalez et al., 2008; Simini et al., 2012; Csáji et al., 2013; Blondel et al., 2015; Pappalardo et al., 2015).

A less common type of data used in the literature are the eXtended Detail Records (XDRs), which are generated by telecommunication companies when a user uploads or downloads data from the Internet using their phone's connection. A single event is a privacy-enhanced record  $(u, t, A, k)$  where  $t$  is the timestamp of the event,  $A$  is the RBS that managed the connection, and  $k$  the amount of uploaded/downloaded information. Given the higher frequency of mobile internet connections, XDRs reduce the problem of sparsity characterizing CDRs (Chen et al., 2019; Luca et al., 2022).

### 2.2. GPS location data

Location intelligence companies collect GPS location data of opt-in individuals from third-party mobile apps through a Software Development Kit (SDK) that captures user locations through GPS signals in Android and iOS devices. In general, a data point contains privacy-enhanced information like the *user identifier*, the *timestamp*, and geographic information such as *longitude* and *latitude*. In the last years, to help a prompt response to the COVID-19 pandemic, location intelligence companies such as Cuebiq,<sup>3</sup>

<sup>3</sup> <https://www.cuebiq.com/>



Unacast,<sup>4</sup> and Safegraph<sup>5</sup> made available several datasets for research purposes (Chang et al., 2021; Hunter et al., 2021; Lucchini et al., 2021; Aleta et al., 2022). The collected GPS location data generally provide more precise location information than CDRs. Unfortunately, location intelligence companies do not share details either on how the data is collected or from which mobile apps, potentially compromising their population representativeness.

On the same line, Big Tech companies such as Facebook (Data4Good<sup>6</sup>) and Google (Community Mobility Reports<sup>7</sup>) have provided GPS location data collected directly from their platforms. However, they have shared these data in an aggregated fashion both in time and space. Thus, the shared data have the advantage of covering all the countries the Big Tech companies operate in but are less precise than the GPS location data provided by location intelligence companies.

## 2.3. Social media data

Social media platforms such as Facebook, Instagram and Twitter facilitate the creation and sharing of content with posts that can contain text, videos, photos, etc. together with a timestamp and an (optional) geographic location. For example on Twitter, people can share geo-located tweets with either their precise geographical position (latitude and longitude) or the location suggested by the platform (e.g., restaurant, landmarks) where the user was located when the tweet was published. Platforms like Foursquare are location-based social networking websites where users share their locations by checking in at points of interest (POIs), such as restaurants, pubs, shops, museums. The users' location is thus available given that such venues have a geographical location (latitude and longitude). For most social media platforms, geotagged posts are downloadable through their

4 <https://www.unacast.com/>

5 <https://www.safegraph.com/>

6 <https://dataforgood.facebook.com/dfg/covid-19>

7 <https://www.google.com/covid19/mobility/>

Application Programming Interfaces (APIs). From spatial and temporal information present in social media posts is thus possible to infer users' mobility trajectories from their posts' history. Such data suffer from data sparsity problems with respect to datasets collected through mobile phones. Nevertheless, social media data have been proven valuable in modeling and dealing with different societal challenges such as urban crime, public health, unemployment (Wang et al., 2012; Broniatowski et al., 2013; Llorente et al., 2015).

## 2.4. Other data sources

### 2.4.1. Credit card transactions

Credit cards are universal across the world, but they have received relatively little attention to date. Since people's spending has become increasingly digitized, it is possible to capture consumer behavior at an unprecedented scale. Each credit card transaction generally consists of privacy-enhanced information such as a *user identifier*, the *timestamp* of the transaction, and the *transaction type* represented by the Merchant Category Code (MCC). Recent research has begun to use transaction records to provide insights on financial wellbeing (Singh et al., 2015), individual traits (Gladstone et al., 2019; Tovanich et al., 2021), purchase behavior in urban populations (Dong et al., 2017; Di Clemente et al., 2018) and segregation (Dong et al., 2020).

### 2.4.2. Satellite imagery

There exist several types of satellite imagery collected by governments and private companies and they can be mainly divided by their spatial, spectral, temporal, radiometric, and geometric resolutions (Campbell and Wynne, 2011). As an example, the Landsat Program represents the longest-running project for the acquisition of satellite imagery of Earth: they provide freely downloadable repeated (average return period of 16 days) imagery with a geometric resolution of 30 m for the entire planet. Satellite data has been proven useful for different tasks such

as tracking urbanization (Tatem, 2017; Strano et al., 2021) and forecasting diseases (Dister et al., 1997; Rogers et al., 2002; Ford et al., 2009).

### 2.4.3. Wearables

In the last few years, wearable sensors such as smartwatches have steadily grown in availability. These devices collect physiological and activity data such as heart rate, sleep, step count and calories burnt. This information can be exploited to track in almost real-time a person's health. As an example, recently smartwatches were used to investigate changes in physiological parameters in response to a COVID-19 infection or COVID-19 vaccines (Guan et al., 2022; Wiedermann et al., 2022).

### 2.4.4. Census

Census data is collected by governments to monitor and gather information about the population of a country. The data is then used to have a reliable picture of the current population, including important information such as demographics and socio-economic conditions. Despite the vast amount of data produced in the digital age, census data remains widely used since it can be jointly used (at an aggregate level) with newer sources of data such as CDRs and provide valuable insights on the general population.

## 3. Urban crime and security

This section delves into the benefits of using alternative data sources to address urban crime and security challenges. The section first provides a brief overview of how the availability of novel data and computational models have altered the landscape of crime research (Section 3.1). Then, Sections 3.2 and 3.2.2 detail various computational methods researchers have employed. In Sections 3.2.3 and 3.2.4, we explore how mobile phone data and social media data have been employed in this line of research. Finally, we conclude the section with critical reflections and potential future directions.

### 3.1. The computational contamination in research on crime

Over the last two decades, scholars from various fields and disciplines have focused on crime and public safety by leveraging the potential of computational methods and novel big data sources. This wave of methodological innovation transcended the traditional boundaries of criminology as a discipline, fostering the interest of social scientists as well as computer scientists, statisticians, applied mathematicians and physicists. As a result, the study of crime has been invested by contamination of approaches, techniques, and viewpoints (Brantingham and Brantingham, 2004; Groff and Mazerolle, 2008; Bogomolov et al., 2015; D'Orsogna and Perc, 2015; Bouchard and Malm, 2016; Faust and Tita, 2019; De Nadai et al., 2020; Hayward and Maas, 2021; Campedelli, 2022b).

Interestingly, the link between computational methods and the study of crime is not as recent as many scholars portray. For instance, Campedelli (2022b) noted how, despite attempts to rebrand such a relationship in terms of novelty, the dialogue between Artificial Intelligence (AI) and research on crime has roots that date back to the 1980s. The relationship in fact emerged decades ago as the result of two processes: the use of AI-based approaches for predictive purposes (Icove, 1986) and the exploration of AI as a tool for aiding sociological theorizing (Brent, 1988; Anderson, 1989; Woolgar, 1989).

Hence, while it is limiting to describe the link between computational methods and the study of crime only by focusing on the recent past, it is nevertheless true that recent years have led to an acceleration in this dialogue, at least in terms of scientific productivity. The reasons behind this fact are four-fold. First, administrative data in digital format have become more and more ubiquitous and easy to access. Second, the democratization of programming languages made it easier for criminologists and crime researchers without a computer science background to explore the potential of algorithmic methods. Third, the availability of other digital sources such as social media data, GPS data, and mobile phone data enriched the information horizon available to study crime. Fourth, following a trend that was already in place, governments and institutions in many Western countries pushed for data-driven solutions to reduce crime, thus increasing funding opportunities in academia as well as business opportunities in the digital and technological sectors. All these factors together made it easier for scholars to gather, process, and analyze data related to crime and security issues, substantially increasing the number of publications and projects over the years (Campedelli, 2020).

Methodology-wise, crime has been investigated through a plethora of different techniques and frameworks. Besides traditional statistical approaches that target either correlational or causal outcomes, geospatial modeling, network science, agent-based modeling, and machine learning have been the four main areas on which scholars have focused their attention. Virtually every area of criminology and crime research has been—to some extent—explored by computational approaches: from white collar crime (Ribeiro et al., 2018; Luna-Pla and Nicolás-Carlock, 2020; Kertész and Wachs, 2021) to terrorism (Moon and Carley, 2007; Chuang et al., 2019; Campedelli et al., 2021), from illicit drugs (Mackey et al., 2018; Magliocca et al., 2019; Sarker et al., 2019) to organized crime (Nardin et al., 2016; Troitzsch, 2017; Calderoni et al., 2021), from gun violence (Mohler, 2014; Green et al., 2017; Loeffler and Flaxman, 2018) to cyber-crime (Shalaginov et al., 2017; Duxbury and Haynie, 2018, 2020), from recidivism (Tollenaar and van der Heijden, 2013; Duwe and Kim, 2017; Berk and Elzarka, 2020) to predictive policing (Caplan et al., 2011; Mohler et al., 2011; Perry, 2013). Particularly, the dialogue between computational methods and the study of recidivism and predictive policing not only focused on technical innovations to optimize forecasting and predictive models, but also provoked vivid debates regarding critical issues of algorithmic accountability, fairness, and transparency (Lum and Isaac, 2016; Dressel and Farid, 2018; Richardson et al., 2019; Akpınar et al., 2021; Purves, 2022). In fact, although the computational analysis of crime has remained chiefly confined to the academic sphere, in some cases—such as predictive policing



and criminal justice risk assessment tools—algorithmic solutions have been deployed by courts and law enforcement agencies. In the US, where this transition from academia to the public and private sectors has been faster, data-driven tools to aid police agencies and courts have a long history (Berk, 2019). Yet, the rapid diffusion of novel tools, coupled with their secrecy, pushed scholars, activists and journalists to scrutinize the effects that these software have on high-stake settings, showing that these instruments often lead to disparate and unfair treatment against minorities, reinforcing discrimination and over-policing in policing and criminal justice. Two sides hence emerged: one populated by those defending the benefits and potential of computational approaches for predicting crime and recidivism (among other things), and those calling for either the elimination of such tools or their heavy regulation.

Within the kaleidoscope of areas in which the computational wave has spread, the study of urban crime has certainly fostered significant scholarly interest. Urban crime trivially embraces all those deviant and criminal behaviors occurring in urban settings and, therefore, can be seen as a higher-level category containing some of those previously mentioned, as the study of illicit drugs (when distributed or consumed in urban settings), the study of violent crime (when perpetrated in urban settings) or predictive policing itself, which by definition targets a certain urban area.

## 3.2. Computational methods, big data, and urban crime

### 3.2.1. The advantages in studying urban crime today

There are some specific reasons behind the fact that urban crime has attracted so much scholarly attention. First and foremost, one of the most popular regularities in the empirical study of crime is the so-called “law of crime concentration” (Weisburd, 2015). Inspired by the theoretical tradition on crime and place (Shaw and McKay, 1942; Cohen and Felson, 1979; Eck and Weisburd, 1995; Brantingham and Brantingham, 2004), the “law of crime concentration” states that most crimes in a city are concentrated in specific small areas, such as blocks, streets or neighborhoods. In other words, crime clusters spatially (Johnson, 2010). The dawn of this empirical finding dates back to the early seminal cartographic works of Quetelet (1831) in the Nineteenth century. Over the decades, scores of studies emerged in the context of routine activity (Cohen and Felson, 1979) and crime pattern theories (Brantingham and Brantingham, 1984) have verified these findings not only in the US but in many other countries all around the world (De Melo et al., 2015; Ye et al., 2015; Mazeika and Kumar, 2017; Breetzke, 2018; Favarin, 2018; Umar et al., 2020). Second, not only does crime cluster spatially, it also clusters temporally. It is in fact well known that the probability that a crime occurs is not homogeneous across time windows (Aaltonen et al., 2018; Holbrook et al., 2021; Piatkowska and Lantz, 2021). In general, crime has its own higher-level seasonalities and these temporal dynamics also vary across crime types (Yan, 2004; Linning et al., 2017; Aaltonen et al., 2018). In many cases, these two layers intersect—especially in the case of urban crime—creating spatio-temporal regularities that allow for deeper analytical scrutiny. In

general, spatial and temporal patterns create the conditions for deploying statistical and mathematical models taking advantage of the non-random data of criminal phenomena for forecasting and predictive purposes. A third reason behind the strong relationship between computational methods and urban crime is that both traditional and more novel data sources for studying crime both make spatial and temporal information available. In the US and other Western countries, for instance, data on reported crimes or calls for service are digitally recorded and easily accessible at the city level, combining information on the type of crime with information on the time and location of the offense. At the same time, data providers and tech companies sell or offer data on social media activity, mobile usage, public transportation, point of interest attendance, and GPS tracking. Overall, the availability of digital data beyond traditional administrative records has allowed scholars to expand the typical analytical frame in which crime—patterned but highly dynamic—is studied considering only fixed or static factors, variables, and conditions, such as the built environment or socio-economic characteristics. In light of this, the study of urban crime is aided by an arsenal of information that is rich—often way richer than the one available to the study of other crime contexts—and can be connected to other human phenomena that are known to be patterned, such as mobility flows. Fourth, the computational analysis of urban crime has straightforward practical consequences that transcend the pure research dimension. While the gap between empirical evidence and policy solutions may be wide for other areas of inquiry, the translation of empirical evidence to crime reduction strategies has always been much faster in the study of urban crime.

Scholars have taken advantage of these conditions and amassed a relevant number of studies with mainly two goals: disentangling crime correlates and forecasting or predicting crime trends and locations. The two goals are interrelated, as optimal forecasting can be achieved only through the selection of relevant correlates and, in turn, the study of correlates cannot be deemed independent from the need to optimize predictive performance.

### 3.2.2. Mobility, urban crime, and ecological networks

Taxi flows and mobility patterns, as proxied for instance by the analysis of activity at POI locations, have been critical components of recent studies targeting urban crime. Some of the works emerging in this area have been framed using agent-based models (ABMs). ABM refers to generative models that allow research to simulate social and criminal phenomena by incorporating empirical or artificial data to investigate research questions that are impractical to be investigated in the real world (e.g., for ethical or monetary reasons). Although ABMs pose several major limitations to the reliability of findings when simulations are not appropriately designed and cannot be validated (Groff et al., 2019; Campedelli, 2022a), when models are carried out properly they offer a compelling set of benefits for criminologists and crime researchers, including theory testing, scenario exploration, and long-term forecasting.

Within this line of research, Ross et al. (2020) propose a simulation model for offender mobility in New York City (NYC) using open data to simulate urban structure, location-based information to proxy human activity and taxi flow data

to proxy human mobility between different areas of the city. By comparing 35 different mobility patterns, the authors highlight the benefits of integrating taxi flow data with previous crime data and popular activity nodes to simulate offenders' mobility meaningfully. In another example, [Ross et al. \(2021\)](#) designed a model aimed at identifying drivers of relevant crime patterns through openly available static and dynamic geographical and temporal features, and proposed a data-driven decision-making process based on machine learning to allow artificial agents to decide whether to engage in crime based on their perception of the surrounding environment. Focusing again on NYC and targeting crime counts at the street level, the authors indicate the stability of high crime areas, in line with the criminological literature, and highlight the importance of the spatial environment in predicting crime hotspots. Agent-based modeling, however, is far from being the only methodological framework utilized in the study of urban crime through mobility data.

[Wang et al. \(2016\)](#), for instance, focused on Chicago to infer crime rates at the neighborhood level using POI and taxi flow data through more traditional statistical approaches like linear and negative binomial regression, indicating that including these information sources reduces prediction error by 17.6 percentage points. In a subsequent extension of the work, [Wang et al. \(2019\)](#) propose a graphically weighted regression approach for crime rate inference that aims at capturing the non-stationarity nature of crime across neighborhoods in the same urban context, i.e., Chicago, from 2010 to 2014. The assumption behind the analysis is that the same features may have different relationships to crime across different spatial contexts, thus involving a further layer of complexity in the dynamic nature of criminal phenomena. Chicago has been historically one of the US cities that attracted the highest scholarly attention in the study of deviance and crime ([Sampson, 2012](#)). In recent years, Chicago has also been the focus of several papers investigating crime from a computational perspective. Besides the articles mentioned above, others have explored the promises of sophisticated statistical techniques to shed light on the city's crime dynamics. [Papachristos and Bastomski \(2018\)](#), for instance, studied how criminal co-offending (measured via co-arrests data) generates pathways between neighborhoods in Chicago, creating a spatial network that facilitates the diffusion of crime in time and space. Their statistical analyses demonstrate that these "neighborhood networks" are stable over time, generated by various processes, including structural characteristics and social dynamics. Their work fits into a growing body of literature that assess the interdependency of neighborhoods within an urban context ([Peterson and Krivo, 2009](#); [Tita and Greenbaum, 2009](#); [Graif et al., 2014](#)), unfolding the connectivity of communities within cities, despite the belief that, given urban segregation and crime clustering, co-offending patterns should also be clustered. Relying on the interdependence of communities via spatial network representations, [Graif et al. \(2021\)](#) study the relationship between crime and commuting patterns across Chicago communities by concentrating on mobility patterns between job and home locations of the city residents. They mainly investigate whether exposure to workplaces characterized by higher disadvantage leads to an increase in local crime, suggesting that this relationship exists.

In other words, [Graif et al. \(2021\)](#) shows that disadvantage in the extra-local network of communities where citizens work is associated with higher crime levels in the communities where the same citizens live. [Sampson and Levy \(2022\)](#) depart from a similar theoretical perspective to show that a neighborhood's wellbeing is statistically dependent upon the wellbeing of the communities that their residents visit or the communities from which visitors come. Remarkably, their analysis outlines that mobility-based socioeconomic disadvantage explains rates of violence and homicide in Chicago neighborhoods. The combination of low scores of residential socioeconomic conditions of residents, visited communities, and visitors constitutes what the authors label "triple disadvantage," elaborating on how this concept is theoretically and technically valuable for explaining crime dynamics in Chicago ([Levy et al., 2020](#)).

### 3.2.3. Urban crime and mobile phone data

Mobility and people dynamics within urban contexts have also been investigated by means of mobile phone data. One of the first notable examples is the work by [Bogomolov et al. \(2014\)](#) in which mobile network infrastructure data on London, UK, is combined with traditional demographic information and geo-localized open data to show that human behavioral data significantly improve the prediction of crime hotspots. London has been the focus of another early work by [Traunmueller et al. \(2014\)](#) in which anonymized mobile telecommunication data are used to investigate urban crime theories. From such telecommunication data, authors extract quantitative proxies for mapping the presence of people in a given area and find that the age diversity and the ratio of visitors in a given area are negatively related to crime, in line with theoretical concepts proposed by Jane Jacobs ([Jacobs, 1961](#)) such as the one of "natural surveillance" and by Felson and Clarke ([Felson and Clarke, 1998](#)), i.e., ratio of young people. [Song et al. \(2019\)](#) utilized geocoded tracks of mobile phones to analyze if the intensity of population mobility among pairs of communities in a large Chinese city can help shed light on offenders' decision-making processes. The study explicitly considers thefts, and its outcomes suggest that such a measure of mobility leads to a higher predictive performance of theft locations compared to the traditional analysis of crime generators. By leveraging mobile phone data and fine-grained spatio-temporal data on violent crime in Manchester, UK, [Haleem et al. \(2021\)](#) proposed the use of the "exposed population-at-risk" concept to shed light on public crime hotspots on Saturday nights. [De Nadai et al. \(2020\)](#) sought instead to examine the link of socioeconomic conditions, built environment and mobility patterns with violent and property crime across multiple cities. The authors identify the focus on single urban contexts as one of the main shortcomings of the existing literature. They hence focus on four contexts with very different social, cultural, and urban characteristics—i.e., Bogotá, Boston, Chicago, and Los Angeles—to provide higher external generalizability of their findings. Mobility flows are proxied through the use of mobile phone data in the form of CDRs. The work shows that combining information on people, crime, places, and human

mobility produces better-performing models in terms of descriptive and predictive accuracy.

### 3.2.4. The role of social media

In the thriving literature focusing on mobility and crime, recent studies have also sought to unfold the potential of social media to capture the dynamic dimension of human behavior in urban contexts, in line with the hypothesis emerging from other studies in the same area of research, namely that resident population does not explain the complexity of the ecology of crime. Among social media platforms, Twitter has certainly received greater attention from scholars interested in urban crime. Wang and Gerber (2015) proposed the use of Twitter data for solving the problem of “next-place prediction,” thus seeking to estimate people’s individual trajectories. According to the authors, Twitter posts provide rich contextual information in the form of text that can be used to construct such individual trajectories even in cases when no direct reference to geospatial information is available. They hence present two models designed to extract geographic information from general texts allowing them to predict the type of venue a user will visit and the distance between the user and a given type of venue in the future. By leveraging this computational approach, they apply their methodology to test the correlation between next-place prediction and crime. Yang et al. (2018) included data from Twitter (and particularly data on the sentiment and topic of tweets) in their CrimeTelescope platform, a software intended to provide optimized crime hotspots prediction in New York. Besides Twitter data, CrimeTelescope also included information on urban infrastructure via POIs from Foursquare and historical crime data. The statistical outcomes of the study suggested that this multi-modal combination of data leads to better predictive performance (up to 5.2%) compared to traditional approaches only using data on past crimes. Malleson and Andresen (2015) highlighted that there is a relationship between the density of tweets in a given area and shifts in crime concentration. Similarly, Hipp et al. (2019) integrated geocoded Twitter data into models to capture the temporal ambient population in Southern California, arguing that social media data can be promising to test routine activity and crime pattern theories. Wo et al. (2022) examines the potential of four Twitter-derived measures to predict crime counts across more than 2,300 block groups in the city of Los Angeles. The aims of the study are specifically two. First, the authors seek to represent local human activity distinguishing between insider and outsider occupants of a neighborhood. Second, they analyze whether statistical relationships exist between property and violent crime and Twitter-derived measures of the ambient population in Los Angeles. Wo and co-authors conclude that Twitter is powerful in aiding research on ambient population and crime distributions at the spatial level. However, not all studies using Twitter data reached the same positive and promising conclusions. Tucker et al. (2021), in fact, critically tested whether geotagged Twitter data correlate with events of public violence and private conflict during weekday days, weekday nights, weekend days, and weekend nights in the city of Boston. The authors indicate that Twitter works as a proxy of human dynamics only for particular types of locations and activities, thus recommending caution in the use of tweets as

comprehensive sources for mapping the ambient population within a city.

### 3.2.5. Critical reflections

The study of crime has been impacted by the vast array of computational methodologies that have spread across social sciences in the last two decades, and urban crime in particular has benefited from this methodological contamination and the increasing availability of digital data sources coming from mobile phones, social media, transportation information, and other geolocalized trace data. This availability has opened many new possibilities to test criminological theories and improve predictive accuracy in terms of crime hotspots and crime patterns in space and time. Nonetheless, it should be noted that important caveats should be considered in critically evaluating the potential and relevance of novel data sources for the study of urban crime. Particularly, as noted by Browning et al. (2021), representativeness and generalizability of both mobility at the place- and person-level is a problem. Browning et al., for instance, argue that representativeness can be an issue when considering data that are collected based on voluntary choices of users and that this representativeness trivially poses a risk to the generalizability of findings across urban contexts (or even across areas within the same urban context). Hence, scholars must recognize this limitation and adopt strategies to mitigate it. Strategies may include methodological innovations in terms of weighting and result validation. Furthermore, the reliance on novel digital data may lead to an increase in the scholarly unbalance toward Western urban contexts. In fact, while accessibility to digital communication technologies is very high in the Western world, the scenario is very different for countries in other regions of the planet, reinforcing the abovementioned issue of representativity and generalizability when deploying these data sources outside the Western context. Finally, it is worth considering the societal implications of governmental decisions to incorporate mobile phones, GPS, POIs, and social media data into software designed for crime prediction, especially in non-democratic countries. In political contexts in which civil, political, and human rights are not sufficiently protected and guaranteed, the exploitation of multi-modal data sources may significantly increase the state of surveillance over citizens, causing detrimental effects on their liberties and wellbeing. Scholars should thus engage more in the ethical consequences of information systems engineered to collect as much data as possible to protect public safety and crime control allegedly.

## 4. Socioeconomic inequalities and segregation

This section explores how novel computational models can help in addressing socioeconomic inequalities and segregation-related issues. We organize the discussion as follows: Section 4.1 briefly summarizes how the increased availability of data and use of computational models benefit these research areas. Next, in Section 4.2, we describe novel computational methods developed to address socioeconomic inequalities and segregation with a focus on the role of GPS, mobile phones, and other data sources (Sections 4.2.1 and

4.2.2). Finally, we conclude the analysis of this line of research with critical reflections and future directions in Section 4.2.3.

## 4.1. The computational contamination in research on inequalities and segregation

Reducing inequalities is of crucial importance to guarantee a more sustainable and just future for our cities and societies. Indeed, socioeconomic inequalities and income segregation threaten access to health and negatively impact health population levels (Wilkinson and Pickett, 2006; Pickett and Wilkinson, 2015), prevent equal access to educational opportunities (Quillian, 2014; Logan and Burdick-Will, 2016), and hinder social and economic development (Neves et al., 2016). Moreover, they are intimately related to opportunities offered by neighborhoods (Sampson, 2012, 2019; Chetty et al., 2016; Manduca and Sampson, 2019; Hedefalk and Dribe, 2020) and human movements and interactions (Eagle et al., 2010; Chetty et al., 2016, 2022a,b; Wang et al., 2018; Dong et al., 2020). Before the advent of the digital era, socioeconomic inequalities were studied through surveys and census data. However, while census data provide a large-scale representativeness of the population, it lacks the ability to capture and provide an up-to-date picture of cities' dynamics and citizens' behaviors, routines, and habits (Lazer et al., 2009). As previously discussed, census data are collected every few years and are made publicly available several months after they were collected (Lazer et al., 2009). Instead, digital data provide alternative data sources that allow capturing different facets of human behavior: human interactions, human movements, and human encounters are just a few examples of human behavior which may play an essential role in investigating inequalities and segregation and that nowadays can be studied by means of mobile phone data (i.e., CDRs, GPS traces, etc.) and other digital traces (e.g., credit card transactions) (Lazer et al., 2009, 2020). In what follows, we discuss how several researchers, from economics and computational social science, have started using alternative data sources to study the daily behaviors and routines associated with socioeconomic inequalities and segregation in cities.

## 4.2. Computational methods, big data, and inequalities

As more people are moving to cities, governments have to deal with novel challenges like gentrification, unaffordability, segregation, and inequality (Glaeser et al., 2001; Florida, 2017). The place where a person lives can have substantial impacts on health (Wilkinson and Pickett, 2006), economic opportunities (Chetty et al., 2014, 2016), infrastructure and services accessibility (Glaeser et al., 2001; Reid et al., 2016; Florida, 2017), and many other aspects, both at a city and national scale (Chetty et al., 2014; Shelton et al., 2015). Thus, measuring inequalities and segregation with timely and accurate data is of paramount importance, and alternative data sources and ubiquitous technologies are starting to play a central role in deeply analyzing factors and behaviors associated to inequalities such as environmental inequalities (Brazil, 2022; Dass et al., 2022), social mixing and income segregation (Shelton et al.,

2015; Moro et al., 2021; Fan et al., 2023), and community resilience (Hong et al., 2021).

### 4.2.1. GPS and mobile phone data

Athey et al. (2020) developed a measure of experienced isolation (by race) to capture individuals' exposure to other (diverse) individuals using GPS data in the US. They found that the isolation individuals experience in their daily life is lower than the one measured by standard residential isolation metrics, especially in cities with higher levels of public transportation, density, education and income. Järv et al. (2015) moved beyond residential segregation to explore individuals' activity spaces, namely the locations visited by an individual because of their regular activities and routines. They exploited CDRs in the city of Tallin in Estonia to measure ethnic segregation (Estonian vs. Russian). They found that, for example, activity locations of Russian speakers tend to be more concentrated in regions with a prevalence of Russian-speaking communities. Xu et al. (2019) leveraged multiple urban datasets (e.g., CDRs records, housing prices and income data) to study citizens' segregation by their socio-economic status and its evolution in both physical and social (communication) spaces in Singapore. They found relatively higher levels of segregation in wealthier classes for both social and physical space. Hong et al. (2021) leveraged mobile phone data to measure the inequalities in community resilience to the Harvey hurricane in Texas. By measuring the mobility behavior of the individuals, the authors highlighted socio-economical and racial disparities in resilience capacity and evacuation patterns, suggesting the adoption of novel data-driven policies to prioritize equal allocations of resources to vulnerable neighborhoods. Another study (Dass et al., 2022) used mobile phone data, socio-demographic data, and infection rates information to measure accessibility to green spaces in Boston, at the beginning of the COVID-19 pandemic. The authors discovered inequalities, where communities with higher infections and higher prevalence of black residents experienced greater infection exposure per visit. Fan et al. (2023) employed mobile phone data of half a million people located in three different metropolitan areas in the US to study how people experienced social mixing in urban streets. The authors found that the density of people's street visits only explains the 26% of street-level diversity (e.g., social mixing), while the adjacent amenities, residential diversity, and income level explain the 44% of the designed diversity score. Also, Fan et al. (2023) shows that while streets densely visited tend to have more crime, diverse streets have fewer crimes. Moro et al. (2021) leveraged high-resolution mobility data of more than 4.5 million users in eleven big US cities to study income segregation. Previously, income segregation was studied using static residential patterns with high spatial resolutions. Thanks to the fine-grained mobility data, the authors found that the income segregation associated with places and individuals may significantly vary even for places that are close to each other. The authors proposed a model and showed that the experienced income segregation of individuals is associated with the exploration of new locations and places visited by visitors from different income groups. In general, Moro et al. (2021) highlights the importance of considering mobility patterns when we aim at measuring income



segregation. Yabe et al. (2023) investigated how social interactions (e.g., encounters) changed during the COVID-19 pandemic with respect to income diversity. The authors relied on a dataset of millions of mobile phone users in multiple US cities for a period of three years before and during the pandemic. Overall, Yabe et al. (2023) found that the diversity of individual-level urban encounters decreased significantly despite the fact that in 2021, indices related to aggregated mobility recovered to pre-pandemic levels. The authors argued that the pandemic could have long-lasting implications on urban income diversity. Brazil (2022) utilized mobile phone data to study the mobility behavior of individuals to uncover environmental inequalities. The author found that people from minority groups and poorer neighborhoods tend to travel to areas with greater levels of air pollution with respect to white and richer neighborhoods.

#### 4.2.2. The role of other data sources

Using hundreds of millions of geotagged tweets, Wang et al. (2018) have measured neighborhood isolation for 50 American cities finding that residents of black and Hispanic neighborhoods, despite their socio-economic status, are less exposed to either non-poor or white neighborhoods than residents of white neighborhoods. Using the same dataset, Phillips et al. (2021) have computed the social integration of 50 American cities with indices that measure the extent to which residents in each neighborhood travel to other neighborhoods in a city. They have shown that cities with greater population densities and less racial segregation have higher levels of structural connectedness. Using Twitter data and geo-located credit card transactions Morales et al. (2019) investigated segregation in the physical and online space together with their relationship. They show that physical and online interactions in urban areas are segregated by income and that information does not flow homogeneously across social classes in either the physical or the virtual space. In a follow-up study, Dong et al. (2020) found that segregation in urban and online interactions seems stronger than the residential ones. Indeed, while residential neighborhoods sometimes might consist of a mix of different socioeconomic groups, purchase activities and online interactions seem to take place more often between neighborhoods whose economic conditions are similar. Additionally, Dong et al. (2020) have shown that segregation increases with differences in socioeconomic status but this effect is asymmetric for shopping behaviors. In fact, the number of movements from poorer to wealthier neighborhoods is larger than vice versa. Hilman et al. (2022) instead leveraged check-in datasets collected from location-based social media and census information to study segregation levels in 20 cities in the US. The authors found an upwards bias for which people of a certain socioeconomic class mostly visited places in the same class with rare visits to locations from higher classes. Furthermore, this bias increases with socioeconomic status and correlates with metrics for estimating racial residential segregation. In recent work, Shelton et al. (2015) showed that geotagged Twitter data can be used to capture the socio-spatial relations of territories dynamically. In particular, the authors analyzed the data for the city of Louisville in Kentucky, US, to show that analyzing

the segregation of the city using static data is not sufficient to understand its dynamics.

Nicoletti et al. (2022) leveraged a combination of census data and geographical information from OpenStreetMap (OSM) to measure the accessibility to different POIs. The authors defined a metric of accessibility and found that, in more than 50 cities, their metric suggests that inequalities appear proportional to growth processes. Also, for 10 of the cities, low accessibility scores were associated with communities with a larger share of minorities and with lower income levels.

Inequalities can also be related to the possibilities of accessing transportation. For example, by analyzing Boston's BlueBikers program mobility data, Fraser et al. (2022) showed that some neighborhoods use bike-sharing programs more than others. However, by considering the underlying socio-demographic characteristics, it emerged that there are significantly different adoption rates with respect to race and income level. Fraser et al. (2022) also pointed out how, by analyzing the mobility network over time (e.g., 2011–2021), Boston's program is gradually reaching a broader range of neighborhoods.

In a couple of recent works, Chetty et al. (2022a,b) have leveraged data from Facebook on 21 billion friendships. In particular, in a first study Chetty et al. (2022a) have measured three types of social capital by postal code in the US: (i) connectedness, namely friendship between people with different characteristics (i.e., high income vs low income), (ii) social cohesion, which is the extent to which networks of friends are clustered in cliques, and (iii) civic engagement, which measures as rates of volunteering or participation in civic organizations. Interestingly, the share of high-income friends among people with low income (called economic connectedness by the authors) is one of the more powerful predictors of upward economic (e.g., income) mobility. Moreover, Chetty et al. (2022a) have also found that differences in economic connectedness can explain upward income mobility, even when controlling for other strong neighborhood-level predictors such as poverty rates, racial segregation, and inequality. In a companion study and again using Facebook data, Chetty et al. (2022b) have shown that about half of the social disconnection across different socioeconomic groups is explained by differences in exposure to people with high socioeconomic status in places such as schools. Instead, the other half is explained by a friendship bias, namely a lower tendency of low-income people to establish friendships with high-income individuals. This ability to disentangle differences in exposures and friendship bias is of paramount relevance for building effective interventions and strategies to increase economic connectedness and thus decrease income segregation and inequalities.

#### 4.2.3. Critical reflections

New methodologies and novel insights on segregation and inequalities in urban environments have been developed thanks to novel data sources that enabled the study of different facets of human behavior and interaction at spatio-temporal scales that were previously unavailable. While these new approaches have brought substantial advantages for a timely study of human dynamics, it is still difficult to use these new methodologies and data sources to

effectively measure the real impact of policies and interventions for reducing inequalities. In particular, this second aspect will require the development of a simulation framework able to generate different scenarios enabled by specific interventions. Moreover, given the proprietary nature of these novel data sources, it is often difficult to have access to longitudinal data that span multiple years and therefore it is problematic to track whether an intervention had an effective impact over time. An example of the potential of a study based on a longitudinal dataset is represented by Yabe et al. (2023), which using three years of data found that the COVID-19 pandemic still has long-lasting implications on urban income diversity in US cities. Furthermore, several studies are based on aggregated data and not on individual data for privacy reasons. This means that pieces of information such as poverty and deprivation are not directly available at the individual level and scholars have to develop proxies of such measures at an aggregated level (i.e., neighborhood) which may hinder the understanding of actual inequalities. As an example, Gündoğdu et al. (2019) had to move the two definitions of bridging and bonding social capital from an individual level to an aggregated level due to the unavailability of individual-level data.

## 5. Public health

In this section, we will examine how computational models can be useful in addressing public health-related issues. We will organize our discussion into several sections. First, we will briefly summarize how the availability of data and the use of computational models have impacted public health studies in Section 5.1. Then, we will analyze the benefits of using big data and new computational tools in Section 5.2.1. Next, we will explore how mobile phone data (Section 5.2.2), social media data (Section 5.2.3), and other data sources (Section 5.2.4) can be used to tackle these issues. Finally, we will wrap up this part of the research with a discussion of future directions and critical reflections in Section 5.2.5.

### 5.1. The computational contamination in research on public health

Public health is an inherently interdisciplinary field, whose key focus, preventing disease and promoting health, largely benefits from the contribution of different scholarly expertise, ranging from medicine to the social sciences, psychology and economics (Gavens et al., 2018). The past decades have seen an ever-increasing adoption of computational and digital methods in the research on public health, and many have been advocating for increasingly closer collaboration between computer scientists and public health scholars (Epstein, 2013).

In particular, major advances in the use of computational methods for public health have been witnessed in the area of infectious disease epidemiology, specifically to model and understand the patterns of disease dynamics and related causes. The use of mathematical models to study the spread of infectious diseases dates back to the seminal works of Ross (1916) and Kermack and McKendrick (1927) at the beginning of the 20th century, who first introduced the law of mass

action in epidemiology. Over the years, epidemic models became increasingly complex, pursuing a higher level of realism by introducing additional interacting components, such as spatially defined structures (Rvachev and Longini, 1985; Sattenspiel and Dietz, 1995), age-stratified contact patterns (Fumanelli et al., 2012), human movements on long and short scales (Colizza et al., 2007; Balcan et al., 2009) and, more in general, human behavior (Funk et al., 2010; Perra et al., 2011). At the same time, the development of such models has been possible thanks to the increasing availability of computing power, thus allowing the *in-silico* recreation of populations with unprecedented levels of detail. If, in the origins of mathematical epidemiology, models were based on the assumption of a single, closed and well-mixed population (Grassly and Fraser, 2008), modern epidemic models are usually structured as ABMs, simulating the daily routines of up to hundreds of millions of individuals and their close contacts, in households, schools, and workplaces, requiring large-scale computational infrastructures (Ferguson et al., 2006; Ajelli et al., 2010; Merler and Ajelli, 2010).

The growth of computing power has been matched by even faster growth in data availability. With the diffusion of ubiquitous technologies and the rise of the Internet era, the field of epidemiology has been rapidly contaminated by digital approaches leading to a newly defined “digital epidemiology” (Salathe et al., 2012). Digital epidemiology, in the definition given by Salathé (2018), refers to epidemiology that uses data that was generated outside the public health system, data that were not collected with a specific public health-related purpose. The first study that brought to worldwide attention the potential use of a novel digital data source in epidemiology described Google Flu Trends, a system to monitor flu activity in more than 25 countries based on search query data (Ginsberg et al., 2009). The service was shut down in 2015, but historical data are still available. Also, the same study sparked a significant controversy on the accuracy of such emerging models and their potential biases (Lazer et al., 2014a,b). Soon thereafter, studies on digital epidemiology started growing exponentially, using a variety of digital sources to track disease prevalence and design public health interventions (Althouse et al., 2015; Bansal et al., 2016). Many studies followed the seminal example of Google Flu Trends by integrating different web data sources to forecast flu activity (Polgreen et al., 2008; Shaman and Karspeck, 2012; Shaman et al., 2013; Yuan et al., 2013; Lampos et al., 2015). As other data sources became rapidly available, their use has been explored in a wide range of epidemiological applications. Mobile phone data have been used to measure human movements and inform both spatially structured epidemic models (Tatem et al., 2009; Wesolowski et al., 2012, 2016), and surveillance systems (Barlacchi et al., 2017). Other studies have further advanced epidemic modeling and forecasting by combining additional data streams such as social media data (Lampos et al., 2010; Zhang et al., 2015, 2017), internet media reports (Freifeld et al., 2008), wearable sensors (Isella et al., 2011; Viboud and Santillana, 2020), and satellite imagery (Bharti et al., 2016; Castro et al., 2021). Finally, the opportunity provided by the Web to directly engage users in scientific research has opened the path to participatory surveillance systems, moving beyond the initial paradigm of passively collected data sources (Paolotti et al., 2014; Smolinski et al., 2015; Brownstein et al., 2017).

In this context, the COVID-19 pandemic has marked a turning point for digital epidemiology. While before 2020, digital epidemiology has been mainly studied as a proof-of-concept with a few real-time applications, since the early days of the COVID-19 outbreak in China, digital approaches have played a crucial role across the whole pandemic life cycle (Oliver et al., 2020), ranging from predictive modeling (Poletto et al., 2020) to the population-scale deployment of digital contact tracing apps (Colizza et al., 2021).

In the next sections, we discuss in more detail the role of the different alternative data sources and modeling techniques in computational epidemiology with a specific focus on applications in the urban context. First, we briefly highlight the general advantages of studying urban public health using digital approaches and big data. After that, we highlight the roles of mobile phone data, social media data, and other novel data streams.

## 5.2. Computational methods, big data, and urban public health

### 5.2.1. The advantages of studying urban public health today

Modern epidemiology traces its roots to the foundational work of John Snow, who first identified the Broad Street pump as the source of the 1854 London cholera outbreak by mapping disease prevalence in Soho, and showing how cases occurred around this pump (Shiode et al., 2015). Since then, cartography and mapping spatial disease patterns have represented a fundamental tool for epidemiologists (Koch, 2011). In particular, the analysis of disease patterns in cities has attracted significant attention from scholars, as large metropolitan areas represent the main hubs of disease emergence and spreading (Ali and Keil, 2011; Connolly et al., 2021), a fact that has been well exemplified by the most recent global pandemics. Due to the high density of people and close proximity of living and working spaces in cities, infectious diseases can easily spread from person to person. Additionally, the crowded and often unsanitary living conditions in many cities can provide a conducive environment for the spread of infectious diseases. For example, inadequate access to clean water and sanitation facilities can lead to the proliferation of waterborne diseases such as cholera. Furthermore, the rapid pace of urbanization and population growth in many cities can put a strain on existing healthcare systems, making it more difficult to effectively detect and respond to outbreaks of infectious diseases (Neiderud, 2015).

In recent years scholars have investigated the role of urban features in the spread of infectious diseases, through a number of computational methods. On the one hand, computational epidemic models have been developed to capture the complexities of human behavior in the urban environment, from fine-scale human movements (Perkins et al., 2014) to contact networks (Eubank et al., 2004). On the other hand, many studies have explored the effect of city characteristics, such as urban population scaling laws (Bettencourt et al., 2007) on health outcomes (Rocha et al., 2015; Bilal et al., 2021) through a mix of theoretical and computational approaches (Schläpfer et al., 2014; Tizzoni et al., 2015). As digital trace data have become pervasive, providing researchers with a

tool to investigate human behavior at a high spatial resolution, several studies have advanced our understanding of the role of urban structures in the spread of epidemics. By combining novel data sources with spatially resolved records of disease incidence, researchers have shown that variations in mobility patterns and the associated spatio-temporal fluctuations in population size can predict variations in the dynamics of seasonal flu epidemics (Dalziel et al., 2013, 2018; Zachreson et al., 2018). Similarly, the hierarchical structure of cities, and their different organization as single-center or multi-center systems, has been shown to predict inter-city variations in the spreading dynamics of respiratory infections (Rader et al., 2020; Brizuela et al., 2021; Aguilar et al., 2022).

Furthermore, the availability of high-resolution digital sources has also enabled the study of determinants of non-communicable diseases and chronic health conditions in urban areas. In particular, the analysis of digital trace data has allowed the characterization of neighborhoods based on novel behavioral indicators, thus providing new metrics to explain the observed residents' health outcomes (Sadilek and Kautz, 2013). Mobile phone data, social media data, and remote sensing have been used to model the pulse of urban life at a scale and granularity that would be hard to achieve with traditional methods. Overall, research in this area has demonstrated that novel digital sources represent an invaluable tool to monitor the health conditions of cities, understand their dynamics and inform public health policies. In the following sections, we provide an overview of some relevant contributions, based on different data sources, to address public health issues in large metropolitan areas.

### 5.2.2. The role of mobile phone data

Location data generated by mobile phones have played a pivotal role in the modeling of human mobility and population settlements at international (Kraemer et al., 2020), national (Deville et al., 2014), and smaller spatial scales (Alessandretti, 2022). Since the early days of digital epidemiology, mobile phone data have represented an invaluable data source to connect empirical human mobility patterns and the spatial spread of infectious diseases. They have been used to calibrate epidemic models, understand disease spreading patterns, and evaluate intervention strategies against them (Wesolowski et al., 2016). Initial efforts to incorporate mobile phone derived mobility metrics into epidemic models have been mostly focused on large spatial scales, such as country-wide movements. This is the case, for instance, of seminal work by Tatem (2009), who leveraged mobile phone data collected in Zanzibar to estimate the relation between human mobility flows and parasite carrier movements and rates of malaria importation. The authors found that most of the people in Zanzibar traveled low-risk short distances but risk groups visiting higher-risk regions for extended periods could be identified. Similarly, Wesolowski et al. (2012) used mobile phone data and malaria prevalence information to estimate how people's movements were related to parasite importation between different regions. With their study, the authors were able to identify sources and sinks of imported infections and also to identify critical travel routes. Applications to city-scale epidemic scenarios have been generally more scarce, however, until the COVID-19 pandemic.

The COVID-19 pandemic has represented a defining moment for the use of mobile phone-derived data in epidemic modeling in cities. For the first time, high-resolution temporally resolved positioning data, collected from millions of users, became available to researchers. Such an unprecedented amount of information has fostered the development of a new generation of ABMs, with the ability to recreate synthetic populations of large urban areas with extraordinary realism, which was deemed impossible until a few years ago. While in 2011 [Cooley et al. \(2011\)](#) built their ABM of 7 million individuals living in New York City only based on the most recent census surveys, 10 years later, [Aleta et al. \(2020\)](#) could explicitly model the time-varying interactions of 100,000 people in the Boston Metropolitan Area, accounting for more than 5 million interactions in schools, workplaces, and households, derived from empirical co-location events. Their study showed that a response system based on enhanced testing and contact tracing could be an important tool to mitigate the spread of COVID-19, once social distancing measures were relaxed. In the following paper, [Aleta et al. \(2022\)](#) developed a similar model, integrating individual-level mobility data with socio-demographic information, to generate synthetic populations in New York City and Seattle, and simulate transmission events in more than 400,000 locations within the two cities. Such a detailed model allowed them to characterize the risk of COVID-19 transmission in different venues, identifying the most likely locations of super spreading events. In a similar effort, [Chang et al. \(2021\)](#) developed an epidemic ABM describing the mobility networks of ten of the largest US metropolitan areas. These networks mapped the hourly movements of 98 million people from census block groups, resulting in 5 billion dynamic edges. Simulations of COVID-19 spread on these large spatially-resolved networks demonstrated how higher infection rates among disadvantaged racial and socioeconomic groups were solely a result of differences in mobility in response to non-pharmaceutical interventions (NPIs). Studies of intra-city mobility during the pandemic were not limited to the US or Europe. As an example, [Gozzi et al. \(2021\)](#) investigated the dynamics of COVID-19 in the metro area of Santiago de Chile, using anonymized mobile phone data from 1.4 million users. By combining mobility traces and a compartmental epidemic model, they found that mobility responses to the lockdown were highly unequal in the city, with most deprived areas experiencing higher levels of mobility and, as a consequence, higher infection rates.

Other studies have used mobile phone data to investigate the socio-economic effects of NPIs in cities. [Bonaccorsi et al. \(2020\)](#) investigated the impact of the first lockdown measures in Italy, using a large-scale mobility dataset provided by Meta. They found that the impact on mobility was stronger in municipalities with higher fiscal capacity, while, at the same time, mobility reductions were larger in municipalities with higher income inequalities. Their results prompted fiscal interventions targeting the unequal effects of COVID-19 mitigation measures. On a similar note, [Gauvin et al. \(2021\)](#) used anonymized individual location data to study the mobility responses to COVID-19 in the neighborhoods of 3 major Italian cities. Their analysis uncovered the desertification of historic city centers, which persisted after the end of the first lockdown. Such a center-periphery gradient was mainly associated with differences in educational attainment. Similar results were

found by [Glodeanu et al. \(2021\)](#) who evaluated socio-economic disparities of mobility responses in the neighborhoods of Madrid.

As mobility restrictions were removed and social life returned to normal, other studies focused on persistent changes in human behavior, that followed the pandemic response. For instance, [Lucchini et al. \(2021\)](#) used location data to analyze how people changed their mobility patterns and person-to-person contacts in response to NPIs in the US. Interestingly, they found a persistent reduction in close contacts and in the number of venues visited, even after the lifting of COVID-19 mandates. Using crowdsourced mobility data from 45 million devices, [Li et al. \(2022\)](#) found evidence of aggravated social segregation in the 12 largest US metropolitan areas, as a consequence of the COVID-19 mobility restrictions. Other studies, instead, have investigated the effects of the pandemic on lifestyles and individual habits. A notable example is a work by [Hunter et al. \(2021\)](#), who investigated the effects of the COVID-19 pandemic on walking habits in 10 major metropolitan areas of the US. The authors used individual-level mobility data to identify changes in the walking behavior of more than 1.6 million anonymized mobile phone users. Their findings highlighted a dramatic decline in walking habits during the first wave of the pandemic. Moreover, they found that once restrictions were lifted, walking levels recovered to pre-COVID-19 measures in high-income areas, whereas low-income areas were still well below pre-COVID-19 levels.

Finally, recent modeling advances have further developed the field of large-scale ABMs by combining high-resolution mobility data with detailed information on the economic role of individuals, as workers and consumers. Recent work by [Pangallo et al. \(2022\)](#) developed an ABM of the New York-Newark-Jersey City Metro Area that is representative of the real population across multiple socio-economic characteristics, including their employment status, the industry they work in, and their ability to work from home. Parameterizing the model with privacy-enhanced location data, the authors could explore the complex tradeoff between health and economy with an unprecedented level of realism.

### 5.2.3. The role of social media

In digital epidemiology, social media data have always represented an important source of information to infer disease prevalence from health-related behaviors or symptoms reported by users ([Brownstein et al., 2009](#); [Aiello et al., 2020](#)). Among social media, Twitter is the one that has attracted the most attention from scholars, thanks to the public availability, and machine readability, of basically all its content ([Mejova et al., 2015b](#)). The most typical use of Twitter data involves the automatic identification of relevant tweets, through either keyword search or natural-language processing, to identify posts whose content is related to some health condition ([Paul and Dredze, 2011](#)). For instance, tweets posted by users who report Influenza-Like Illness (ILI) symptoms. Collected tweets are then used as input to predictive models that aim at reproducing some known baseline, such as the ILI trends reported by official public health surveillance.

While initial efforts in this direction were mostly focused on measuring aggregated statistics of disease prevalence at the



national level (Broniatowski et al., 2013; Gesualdo et al., 2013), the availability of geo-tagged social media data with GPS accuracy, in particular from Twitter, has allowed mapping users' health conditions at a very high spatial resolution, reaching the scale of a city (Sadilek et al., 2012). In a seminal paper, Nagar et al. (2014) used geo-referenced city-level Twitter data as a means of forecasting real-time ILI emergency department visits in New York City. They demonstrated that, at that spatial resolution, Twitter data could effectively capture the dynamics of flu in the boroughs of NYC. They also found that a model using the number of infection-related tweets outperformed one based on the number of web searches in predicting the number of ILI-related visits to emergency departments. Similarly, Lu et al. (2018) used Twitter data to model seasonal flu epidemics in the Boston Metropolitan Area.

Social media also represents a valuable data source to monitor non-communicable diseases and health habits in cities. A comprehensive study by Nguyen et al. (2016) developed a publicly available neighborhood-level dataset with indicators related to health behaviors and wellbeing in the USA. Interestingly, the authors found that greater happiness, and positivity toward physical activity and toward healthy foods, assessed via tweets, were associated with lower all-cause mortality and prevalence of chronic conditions such as obesity and diabetes and lower physical inactivity, and smoking. Similarly, Twitter data has been proven useful to map dietary habits in US cities, down to the level of census tract. A study by Gore et al. (2015) investigated how the obesity rate of an urban geographic area correlates with the contents of geo-tagged tweets in that area. In recent work by Sigalo et al. (2022), the authors analyzed about sixty-thousand geolocated food-related tweets collected across 25 cities, in the USA. They found associations between a census tract being classified as a food desert and an increase in the number of tweets in a census tract that mentioned unhealthy foods. Instagram, a video and photo sharing social network with more than 1 billion users worldwide, represents another relevant data source to investigate health habits, and in particular dietary choices, at scale. As an example, De Choudhury et al. (2016) showed that the textual content of Instagram posts predicts with high accuracy food deserts in the metropolitan areas of the US, while Mejova et al. (2015a) combined data from Instagram, Twitter, and Foursquare to correlate dietary choices and the prevalence of obesity across the USA. Another public health-relevant use of social media, which has been extensively explored by health departments in the US, is monitoring reports of foodborne illnesses. Two notable studies investigated the potential use of Twitter posts (Harris et al., 2014) and Yelp reviews (Harrison et al., 2014) to track food poisoning outbreaks in Chicago and New York City, respectively. Both studies demonstrated the high impact of using social media data to improve surveillance in collaboration with city public health authorities. Finally, Aiello et al. (2016) used a random sample of 17 million geo-referenced Flickr photos taken within Greater London between 2010 and 2015 to create a high-resolution map of the sound landscape of the city. They further leveraged such dataset to quantify the effects of noise on population health, correlating noise exposure levels with hypertension rates, at a very high spatial granularity (Gasco et al., 2020).

#### 5.2.4. The role of other data sources

Beyond mobile phone location data and social media, several studies have demonstrated the potential use of other digital sources for public health research. As already mentioned, multiple studies have leveraged search query data of services like Google Search (Ginsberg et al., 2009; Lampos et al., 2015), Baidu (Yuan et al., 2013), or Bing (Lampos et al., 2021), to monitor epidemics at scale. Internet search queries have not only been used to track the spread of infectious diseases but also to monitor other health conditions, for instance, mental health. As an example, Adler et al. (2019) combined official demographic statistics with data generated from Bing queries to gain insight into suicide rates per state in India as reported by the official census.

Another useful, but mostly untapped, data source to monitor disease incidence is Wikipedia pageview data. A landmark paper by McIver and Brownstein (2014) showed that the number of Wikipedia article views of specific health-related pages was a good predictor of ILI activity in the US. However, even though Wikipedia pageview data are geolocated, their availability with geo-encoded information is limited due to privacy reasons. Thus, city-scale studies are scarce. A notable example is a work by Tizzoni et al. (2020), who measured changes in awareness in the US during the 2016 Zika epidemic through geo-localized Wikipedia pageview data, at the level of US city. They examined the attention to Zika in 788 cities in the US with a population larger than 40,000 and found clear and distinct patterns of attention, varying with the exposure to the virus and the volume of media coverage.

Electronic records of retail market purchases represent a novel and interesting data stream, whose potential has been recently explored. Miliou et al. (2021) proposed to use retail market data to improve the forecasting of seasonal flu. In particular, the authors showed that by identifying some specific co-purchases of products, by specific customers, it is possible to model seasonal flu incidence in Italy, 4 weeks in advance, with improved accuracy with respect to an autoregressive baseline. Aiello et al. (2019) collected and analyzed a similar dataset, reaching an unprecedented level of spatial granularity. By analyzing 1.6 billion food item purchases and 1.1 billion medical prescriptions for the entire city of London over the course of one year, they showed that nutrient diversity and amount of calories are the two strongest predictors of the prevalence of three chronic health conditions: hypertension, high cholesterol, and diabetes.

#### 5.2.5. Critical reflections

The future of urban public health is undoubtedly going to be more and more digital. It is clear, however, that several challenges lie ahead, as has been evidenced by the adoption of computational and digital technologies during the COVID-19 pandemic. Thanks to increasing computing power and the availability of high-resolution behavioral data, computational models of epidemics are able to capture key determinants of transmission with impressive detail. However, they often lack a structural integration with socioeconomic dimensions that are known to affect epidemic outcomes. Recent studies have pointed out the need for equitable

approaches in digital epidemiology, to address socioeconomic gaps in disease surveillance and modeling (Buckee et al., 2021; Tizzoni et al., 2022). Future work should aim at reducing health disparities during health emergencies through closer collaboration between epidemiologists and social scientists, psychologists, and economists.

Of course, the use of passively collected digital traces in public health comes with significant privacy and ethical concerns. In an urban context, measuring behaviors at a high spatial granularity represents a key advantage with respect to traditional data sources. However, reaching a high granularity may imply a higher risk of data re-identification, especially with small sample sizes, thus putting individual privacy at risk. In the future, it will be important to understand what privacy-preserving mechanisms can be most effective in minimizing such risks while preserving the potential of data analysis, even at a high spatial resolution.

Finally, the use of novel digital sources requires a careful understanding of their limitations and their scope. For instance, mobile phone-derived mobility metrics have been proven useful to understand the dynamics of COVID-19 in the early phase of the outbreak, however, the relationship between mobility indicators and epidemic outcomes is not straightforward (Kishore, 2021). While mobile phone data are clearly useful to measure changes in human behavior and link them with epidemic dynamics, such link often varies over time, and understanding this varying relationship poses significant challenges to scholars and policymakers who may want to use mobile phone data to evaluate the effectiveness of interventions or forecast future epidemic trajectories. Further work is needed to define methods that can systematically assess the quality of mobile phone-derived mobility metrics and make them comparable across different settings and data providers.

## 6. Conclusions

Cities are the beating heart of our modern societies. With more than half of the world's population living there, multiple emerging societal challenges require modern solutions. In particular, measuring the efficiency of deployed policies and progress toward specific SDGs, it is fundamental to have an always up-to-date picture of human dynamics in cities (e.g., how people move, how they interact with each other). In this context, it is clear that a pivotal role is played by the data collected from alternatives (ubiquitous) data sources like mobile phones, social media, GPS traces, satellite images, wearable devices and many others. In our

review paper, we showcase how such alternative data are employed to monitor signs of progress toward some specific United Nations' Sustainable Development Goals. In particular, after a discussion about the different alternative data sources, we review how such information has been used to monitor urban crime and public safety. After that, we highlight the role of such data in reducing socioeconomic inequalities and segregation. Finally, we showed how they impacted research about public health. In all the sections, we start with a brief discussion about the advantages of using big data with respect to other techniques. Afterwards, we describe how different studies use such information. Finally, we conclude every section with some critical reflections about limitations and potential future directions.

## Author contributions

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

## Funding

BL and SC are supported by the PNRR ICSC National Research Centre for High Performance Computing, Big Data and Quantum Computing (CN00000013), under the NRRP MUR program funded by the NextGenerationEU.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## References

- Aaltonen, M., Kivivuori, J., and Kuitunen, L. (2018). Short-term temporal clustering of police-reported violent offending and victimization: examining timing and the role of revenge. *Crim. Justice Rev.* 43, 309–324. doi: 10.1177/0734016818761100
- Adler, N., Cattuto, C., Kalimeri, K., Paolotti, D., Tizzoni, M., Verhulst, S., et al. (2019). How search engine data enhance the understanding of determinants of suicide in India and inform prevention: observational study. *J. Med. Internet Res.* 21, e10179. doi: 10.2196/10179
- Aguilar, J., Bassolas, A., Ghoshal, G., Hazarie, S., Kirkley, A., Mazzoli, M., et al. (2022). Impact of urban structure on infectious disease spreading. *Sci. Rep.* 12, 1–13. doi: 10.1038/s41598-022-06720-8
- Aiello, A. E., Renson, A., and Zivich, P. (2020). Social media- and internet-based disease surveillance for public health. *Annu. Rev. Public Health* 41, 101. doi: 10.1146/annurev-publhealth-040119-094402
- Aiello, L. M., Schifanella, R., Quercia, D., and Aletta, F. (2016). Chatty maps: constructing sound maps of urban areas from social media data. *R. Soc. Open Sci.* 3, 150690. doi: 10.1098/rsos.150690
- Aiello, L. M., Schifanella, R., Quercia, D., and Del Prete, L. (2019). Large-scale and high-resolution analysis of food purchases and health outcomes. *EPJ Data Sci.* 8, 1–22. doi: 10.1140/epjds/s13688-019-0191-y

- Ajelli, M., Gonçalves, B., Balcan, D., Colizza, V., Hu, H., Ramasco, J. J., et al. (2010). Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC Infect. Dis.* 10, 1–13. doi: 10.1186/1471-2334-10-190
- Akpinar, N.-J., De-Arteaga, M., and Chouldechova, A. (2021). “The effect of differential victim crime reporting on predictive policing systems,” in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, FAccT '21* (New York, NY: Association for Computing Machinery), 838–849. doi: 10.1145/3442188.3445877
- Alessandretti, L. (2022). What human mobility data tell us about covid-19 spread. *Nat. Rev. Phys.* 4, 12–13. doi: 10.1038/s42254-021-00407-1
- Aleta, A., Martín-Corral, D., Bakker, M. A., Pastore y Piontti, A., Ajelli, M., Litvinova, M., et al. (2022). Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas. *Proc. Nat. Acad. Sci.* 119, e2112182119. doi: 10.1073/pnas.2112182119
- Aleta, A., Martín-Corral, D., Pastore y Piontti, A., Ajelli, M., Litvinova, M., Chinazzi, M., et al. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nat. Hum. Behav.* 4, 964–971. doi: 10.1038/s41562-020-0931-9
- Ali, S. H., and Keil, R. (2011). *Networked Disease: Emerging Infections in the Global City*. Hoboken, NJ: John Wiley & Sons.
- Althouse, B. M., Scarpino, S. V., Meyers, L. A., Ayers, J. W., Bargsten, M., Baumbach, J., et al. (2015). Enhancing disease surveillance with novel data streams: challenges and opportunities. *EPJ Data Sci.* 4, 1–8. doi: 10.1140/epjds/s13688-015-0054-0
- Anderson, B. (1989). On artificial intelligence and theory construction in sociology. *J. Math. Sociol.* 14, 209–216. doi: 10.1080/0022250X.1989.9990050
- Athey, S., Ferguson, B. A., Gentzkow, M., and Schmidt, T. (2020). *Experienced Segregation*. Technical report. Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w27572
- Bagrow, J. P., Wang, D., and Barabasi, A.-L. (2011). Collective response of human populations to large-scale emergencies. *PLoS ONE* 6, e17680. doi: 10.1371/journal.pone.0017680
- Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J. J., and Vespignani, A. (2009). Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Nat. Acad. Sci.* 106, 21484–21489. doi: 10.1073/pnas.0906910106
- Bansal, S., Chowell, G., Simonsen, L., Vespignani, A., and Viboud, C. (2016). Big data for infectious disease surveillance and modeling. *J. Infect. Dis.* 214(suppl\_4), S375–S379. doi: 10.1093/infdis/jiw400
- Barlacchi, G., Perentis, C., Mehrotra, A., Musolesi, M., and Lepri, B. (2017). Are you getting sick? Predicting influenza-like symptoms using human mobility behaviors. *EPJ Data Sci.* 6, 1–15. doi: 10.1140/epjds/s13688-017-0124-6
- Batty, M. (2013). *The New Science of Cities*. Cambridge, MA: MIT Press. doi: 10.7551/mitpress/9399.001.0001
- Berk, R. (2019). *Machine Learning Risk Assessments in Criminal Justice Settings*. Cham: Springer International Publishing. doi: 10.1007/978-3-030-02272-3
- Berk, R., and Elzarka, A. A. (2020). Almost politically acceptable criminal justice risk assessment. *Criminol. Public Policy* 19, 1231–1257. doi: 10.1111/1745-9133.12500
- Bettencourt, L., and West, G. (2010). A unified theory of urban living. *Nature* 467, 912–913. doi: 10.1038/467912a
- Bettencourt, L. M., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. (2007). Growth, innovation, scaling, and the pace of life in cities. *Proc. Nat. Acad. Sci.* 104, 7301–7306. doi: 10.1073/pnas.0610172104
- Bharti, N., Djibo, A., Tatem, A. J., Grenfell, B. T., and Ferrari, M. J. (2016). Measuring populations to improve vaccination coverage. *Sci. Rep.* 6, 1–10. doi: 10.1038/srep34541
- Bilal, U., de Castro, C. P., Alfaro, T., Barrientos-Gutierrez, T., Barreto, M. L., Leveau, P., et al. (2021). Scaling of mortality in 742 metropolitan areas of the Americas. *Sci. Adv.* 7, eabl6325. doi: 10.1126/sciadv.abl6325
- Blondel, V. D., Decuyper, A., and Krings, G. (2015). A survey of results on mobile phone datasets analysis. *EPJ Data Sci.* 4, 10. doi: 10.1140/epjds/s13688-015-0046-0
- Bogomolov, A., Lepri, B., Staiano, J., Letouze, E., Oliver, N., Pianesi, F., et al. (2015). Moves on the street: classifying crime hotspots using aggregated anonymized data on phone dynamics. *Big Data* 3, 148–158. doi: 10.1089/big.2014.0054
- Bogomolov, A., Lepri, B., Staiano, J., Oliver, N., Pianesi, F., Pentland, A., et al. (2014). “Once upon a crime: towards crime prediction from demographics and mobile data,” in *Proceedings of the 16th International Conference on Multimodal Interaction* (Istanbul: ACM), 427–434. doi: 10.1145/2663204.2663254
- Bohorquez, J., Gourley, S., Dixon, A., Spagat, M., and Johnson, N. (2009). Common ecology quantifies human insurgency. *Nature* 462, 911–914. doi: 10.1038/nature08631
- Bonaccorsi, 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. *Proc. Nat. Acad. Sci.* 117, 15530–15535. doi: 10.1073/pnas.2007658117
- Bouchard, M., and Malm, A. (2016). *Social Network Analysis and Its Contribution to Research on Crime and Criminal Justice, Volume 1*. Oxford: Oxford University Press. doi: 10.1093/oxfordhb/9780199935383.013.21
- Brantingham, P. J., and Brantingham, P. L. (1984). *Patterns in Crime*. New York, NY: Macmillan.
- Brantingham, P. L., and Brantingham, P. J. (2004). Computer simulation as a tool for environmental criminologists. *Secur. J.* 17, 21–30. doi: 10.1057/palgrave.sj.8340159
- Brazil, N. (2022). Environmental inequality in the neighborhood networks of urban mobility in us cities. *Proc. Nat. Acad. Sci.* 119, e2117776119. doi: 10.1073/pnas.2117776119
- Breetzke, G. D. (2018). The concentration of urban crime in space by race: evidence from South Africa. *Urban Geogr.* 39, 1195–1220. doi: 10.1080/02723638.2018.1440127
- Brent, E. (1988). Is there a role for artificial intelligence in sociological theorizing? *Am. Sociol.* 19, 158–166. doi: 10.1007/BF02691809
- Brizuela, N. G., García-Chan, N., Gutierrez Pulido, H., Chowell, G. (2021). Understanding the role of urban design in disease spreading. *Proc. R. Soc. A* 477, 20200524. doi: 10.1098/rspa.2020.0524
- Broniatowski, D. A., Paul, M. J., and Dredze, M. (2013). National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. *PLoS ONE* 8, e83672. doi: 10.1371/journal.pone.0083672
- Browning, C. R., Pinchak, N. P., and Calder, C. A. (2021). Human mobility and crime: theoretical approaches and novel data collection strategies. *Ann. Rev. Criminol.* 4, 99–123. doi: 10.1146/annurev-criminol-061020-021551
- Brownstein, J. S., Chu, S., Marathe, A., Marathe, M. V., Nguyen, A. T., Paolotti, D., et al. (2017). Combining participatory influenza surveillance with modeling and forecasting: three alternative approaches. *JMIR Public Health Surveill.* 3, e7344. doi: 10.2196/publichealth.7344
- Brownstein, J. S., Freifeld, C. C., and Madoff, L. C. (2009). Digital disease detection? harnessing the web for public health surveillance. *N. Engl. J. Med.* 360, 2153. doi: 10.1056/NEJMp0900702
- Buckee, C., Noor, A., and Sattenspiel, L. (2021). Thinking clearly about social aspects of infectious disease transmission. *Nature* 595, 205–213. doi: 10.1038/s41586-021-03694-x
- Calabrese, F., di Lorenzo, G., Liu, L., and Ratti, C. (2011). Estimating origin-destination flows using mobile phone location data. *IEEE Pervasive Comput.* 10, 36–44. doi: 10.1109/MPRV.2011.41
- Calderoni, F., Campedelli, G. M., Szekely, A., Paolucci, M., and Andrighetto, G. (2021). Recruitment into organized crime: an agent-based approach testing the impact of different policies. *J. Quant. Criminol.* 38, 197–237. doi: 10.21428/cb6ab371.d3cb86db
- Campbell, J. B., and Wynne, R. H. (2011). *Introduction to Remote Sensing*. New York, NY: Guilford Press.
- Campedelli, G. M. (2020). Where are we? Using Scopus to map the literature at the intersection between artificial intelligence and research on crime. *J. Comput. Soc. Sci.* 4, 503–530. doi: 10.31235/osf.io/853fx
- Campedelli, G. M. (2022b). *Machine Learning for Criminology and Crime Research: at the Crossroads*. Routledge Advances in Criminology, 1st ed. New York, NY: Routledge. doi: 10.4324/9781003217732
- Campedelli, G. M. (ed) (2022a). “Criminology at the crossroads? Computational perspectives,” in *Machine Learning for Criminology and Crime Research: At the Crossroads* (New York, NY: Routledge).
- Campedelli, G. M., Bartulovic, M., and Carley, K. M. (2021). Learning future terrorist targets through temporal meta-graphs. *Sci. Rep.* 11, 8533. doi: 10.1038/s41598-021-87709-7
- Caplan, J. M., Kennedy, L. W., and Miller, J. (2011). Risk terrain modeling: brokering criminological theory and GIS methods for crime forecasting. *Justice Q.* 28, 360–381. doi: 10.1080/07418825.2010.486037
- Castro, L. A., Generous, N., Luo, W., Pastore y Piontti, A., Martinez, K., Gomes, M. F., et al. (2021). Using heterogeneous data to identify signatures of dengue outbreaks at fine spatio-temporal scales across brazil. *PLoS Negl. Trop. Dis.* 15, e0009392. doi: 10.1371/journal.pntd.0009392
- Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., et al. (2021). Mobility network models of covid-19 explain inequities and inform reopening. *Nature* 589, 82–87. doi: 10.1038/s41586-020-2923-3
- Chen, G., Viana, A. C., Fiore, M., and Sarraute, C. (2019). Complete trajectory reconstruction from sparse mobile phone data. *EPJ Data Sci.* 8, 30. doi: 10.1007/978-981-15-0118-0
- Chetty, R., Hendren, N., and Katz, L. F. (2016). The effects of exposure to better neighborhoods on children: new evidence from the moving to opportunity experiment. *Am. Econ. Rev.* 106, 855–902. doi: 10.1257/aer.20150572
- Chetty, R., Hendren, N., Kline, P., and Saez, E. (2014). Where is the land of opportunity? the geography of intergenerational mobility in the united states. *Q. J. Econ.* 129, 1553–1623. doi: 10.1093/qje/qju022



- Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., et al. (2022a). Social capital: measurement and associations with economic mobility. *Nature* 608, 108–121. doi: 10.1038/s41586-022-04996-4
- Chetty, R., Jackson, M. O., Kuchler, T., Stroebel, J., Hendren, N., Fluegge, R. B., et al. (2022b). Social capital: measurement and associations with economic mobility. *Nature* 608, 122–134. doi: 10.1038/s41586-022-04997-3
- Chuang, Y.-L., Ben-Asher, N., and D'Orsogna, M. R. (2019). Local alliances and rivalries shape near-repeat terror activity of al-Qaeda, ISIS, and insurgents. *Proc. Nat. Acad. Sci.* 116, 20898–20903. doi: 10.1073/pnas.1904418116
- Cohen, L. E., and Felson, M. (1979). Social change and crime rate trends: a routine activity approach. *Am. Sociol. Rev.* 44, 588–608. doi: 10.2307/2094589
- Colizza, V., Barrat, A., Barthélemy, M., Valleron, A.-J., and Vespignani, A. (2007). Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS Med.* 4, e13. doi: 10.1371/journal.pmed.0040013
- Colizza, V., Grill, E., Mikolajczyk, R., Cattuto, C., Kucharski, A., Riley, S., et al. (2021). Time to evaluate covid-19 contact-tracing apps. *Nat. Med.* 27, 361–362. doi: 10.1038/s41591-021-01236-6
- Connolly, C., Keil, R., and Ali, S. H. (2021). Extended urbanisation and the spatialities of infectious disease: demographic change, infrastructure and governance. *Urban Stud.* 58, 245–263. doi: 10.1177/0042098020910873
- 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. *J. Urban Health* 88, 982–995. doi: 10.1007/s11524-011-9603-4
- Csáji, B. C., Browet, A., Traag, V. A., Delvenne, J.-C., Huens, E., Van Dooren, P., et al. (2013). Exploring the mobility of mobile phone users. *Phys. A Stat. Mech. Appl.* 392, 1459–1473. doi: 10.1016/j.physa.2012.11.040
- Cui, Y., Xie, X., and Liu, Y. (2018). Social media and mobility landscape: uncovering spatial patterns of urban human mobility with multi source data. *Front. Environ. Sci. Eng.* 12, 7. doi: 10.1007/s11783-018-1068-1
- Dalziel, B. D., Kissler, S., Gog, J. R., Viboud, C., Bjørnstad, O. N., Metcalf, C. J. E., et al. (2018). Urbanization and humidity shape the intensity of influenza epidemics in us cities. *Science* 362, 75–79. doi: 10.1126/science.aat6030
- Dalziel, B. D., Pourbouloul, B., and Ellner, S. P. (2013). Human mobility patterns predict divergent epidemic dynamics among cities. *Proc. R. Soc. B Biol. Sci.* 280, 20130763. doi: 10.1098/rspb.2013.0763
- Dass, S., O'Brien, D. T., and Ristea, A. (2022). Strategies and inequities in balancing recreation and covid exposure when visiting green spaces. *Environ. Plan. B Urban Anal. City Sci.* 23998083221114645. doi: 10.1177/23998083221114645
- De Choudhury, M., Sharma, S., and Kiciman, E. (2016). "Characterizing dietary choices, nutrition, and language in food deserts via social media," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work and Social Computing* (San Francisco, CA: Association for Computing Machinery), 1157–1170. doi: 10.1145/2818048.2819956
- De Melo, S. N., Matias, L. F., and Andresen, M. A. (2015). Crime concentrations and similarities in spatial crime patterns in a Brazilian context. *Appl. Geograp.* 62, 314–324. doi: 10.1016/j.apgeog.2015.05.012
- De Nadai, M., Xu, Y., Letouze, E., González, M. C., and Lepri, B. (2020). Socio-economic, built environment, and mobility conditions associated with crime: a study of multiple cities. *Sci. Rep.* 10, 13871. doi: 10.1038/s41598-020-70808-2
- Deville, P., Linard, C., Martin, S., Gilbert, M., Stevens, F. R., Gaughan, A. E., et al. (2014). Dynamic population mapping using mobile phone data. *Proc. Nat. Acad. Sci.* 111, 15888–15893. doi: 10.1073/pnas.1408439111
- Di Clemente, R., Luengo-Oroz, M., Travizano, M., Xu, S., Vaitla, B., and González, M. C. (2018). Sequences of purchases in credit card data reveal lifestyles in urban populations. *Nat. Commun.* 9, 1–8. doi: 10.1038/s41467-018-05690-8
- Dister, S. W., Fish, D., Bros, S. M., Frank, D. H., and Wood, B. L. (1997). Landscape characterization of peridomestic risk for lyme disease using satellite imagery. *Am. J. Trop. Med. Hyg.* 57, 687–692. doi: 10.4269/ajtmh.1997.57.687
- Dong, X., Morales, A. J., Jahani, E., Moro, E., Lepri, B., Bozkaya, B., et al. (2020). Segregated interactions in urban and online space. *EPJ Data Sci.* 9, 20. doi: 10.1140/epjds/s13688-020-00238-7
- Dong, X., Suhara, Y., Bozkaya, B., Singh, V. K., Lepri, B., Pentland, A. S., et al. (2017). Social bridges in urban purchase behavior. *ACM Trans. Intell. Syst. Technol.* 9, 1–29. doi: 10.1145/3149409
- D'Orsogna, M. R., and Perc, M. (2015). Statistical physics of crime: a review. *Phys. Life Rev.* 12, 1–21. doi: 10.1016/j.plev.2014.11.001
- Dressel, J., and Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Sci. Adv.* 4, eaao5580. doi: 10.1126/sciadv.aao5580
- Duwe, G., and Kim, K. (2017). Out with the old and in with the new? An empirical comparison of supervised learning algorithms to predict recidivism. *Crim. Justice Policy Rev.* 28, 570–600. doi: 10.1177/0887403415604899
- Duxbury, S., and Haynie, D. L. (2020). The responsiveness of criminal networks to intentional attacks: disrupting darknet drug trade. *PLoS ONE* 15, e0238019. doi: 10.1371/journal.pone.0238019
- Duxbury, S. W., and Haynie, D. L. (2018). The network structure of opioid distribution on a darknet cryptomarket. *J. Quant. Criminol.* 34, 921–941. doi: 10.1007/s10940-017-9359-4
- Eagle, N., Macy, M., and Claxton, R. (2010). Network diversity and economic development. *Science* 328, 1029–1031. doi: 10.1126/science.1186605
- Eck, J. E., and Weisburd, D. (eds) (1995). *Crime and Place: Crime Prevention Studies*. Monsey, NY: Willow Tree Pr.
- Epstein, J. A. (2013). Collaborations between public health and computer science: a path worth pursuing. *Am. J. Public Health Res.* 1, 166–170. doi: 10.12691/ajphr-1-7-4
- Eubank, S., Guclu, H., Anil Kumar, V., Marathe, M. V., Srinivasan, A., Toroczkai, Z., et al. (2004). Modelling disease outbreaks in realistic urban social networks. *Nature* 429, 180–184. doi: 10.1038/nature02541
- Fan, Z., Su, T., Sun, M., Noyman, A., Zhang, F., Pentland, A. S., et al. (2023). Diversity beyond density: Experienced social mixing of urban streets. *PNAS Nexus* 2, pgad077. doi: 10.1093/pnasnexus/pgad077
- Faust, K., and Tita, G. E. (2019). Social networks and crime: pitfalls and promises for advancing the field. *Ann. Rev. Criminol.* 2, 99–122. doi: 10.1146/annurev-criminol-011518-024701
- Favarin, S. (2018). This must be the place (to commit a crime). Testing the law of crime concentration in Milan, Italy. *Eur. J. Criminol.* 15, 702–729. doi: 10.1177/1477370818757700
- Felson, M., and Clarke, R. (1998). Opportunity makes the thief: practical theory for crime prevention. *Police Research Series, Paper 98*.
- Ferguson, N. M., Cummings, D. A., Fraser, C., Cajka, J. C., Cooley, P. C., Burke, D. S., et al. (2006). Strategies for mitigating an influenza pandemic. *Nature* 442, 448–452. doi: 10.1038/nature04795
- Florida, R. (2017). *The New Urban Crisis: How our Cities are Increasing Inequality, Deepening Segregation, and Failing the Middle Class-and What we can do about it*. London: Hachette UK.
- Ford, T. E., Colwell, R. R., Rose, J. B., Morse, S. S., Rogers, D. J., Yates, T. L., et al. (2009). Using satellite images of environmental changes to predict infectious disease outbreaks. *Emerg. Infect. Dis.* 15, 1341. doi: 10.3201/eid1509.081334
- Fraser, T., Van Woert, K., Olivieri, S., Baron, J., Buckley, K., Lalli, P., et al. (2022). *Cycling Cities: Measuring Transportation Equity in Bikeshare Networks*. Available at SSRN 4076776. doi: 10.2139/ssrn.4076776
- Freifeld, C. C., Mandl, K. D., Reis, B. Y., and Brownstein, J. S. (2008). Healthmap: global infectious disease monitoring through automated classification and visualization of internet media reports. *J. Am. Med. Inform. Assoc.* 15, 150–157. doi: 10.1197/jamia.M2544
- Fumanelli, L., Ajelli, M., Manfredi, P., Vespignani, A., and Merler, S. (2012). Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput. Biol.* 8, e1002673. doi: 10.1371/journal.pcbi.1002673
- Funk, S., Salathé, M., and Jansen, V. A. (2010). Modelling the influence of human behaviour on the spread of infectious diseases: a review. *J. R. Soc. Interface* 7, 1247–1256. doi: 10.1098/rsif.2010.0142
- Gasco, L., Schifanella, R., Aiello, L. M., Quercia, D., Asensio, C., de Arcas, G., et al. (2020). Social media and open data to quantify the effects of noise on health. *Front. Sustain. Cities* 2, 41. doi: 10.3389/frsc.2020.00041
- Gauvin, L., Bajardi, P., Pepe, E., Lake, B., Privitera, F., Tizzoni, M., et al. (2021). Socio-economic determinants of mobility responses during the first wave of covid-19 in Italy: from provinces to neighbourhoods. *J. R. Soc. Interface* 18, 20210092. doi: 10.1098/rsif.2021.0092
- Gavens, L., Holmes, J., Bühringer, G., McLeod, J., Neumann, M., Lingford-Hughes, A., et al. (2018). Interdisciplinary working in public health research: a proposed good practice checklist. *J. Public Health* 40, 175–182. doi: 10.1093/pubmed/idx027
- Gesualdo, F., Stilo, G., Agricola, E., Gonfiantini, M. V., Pandolfi, E., Velardi, P., et al. (2013). Influenza-like illness surveillance on twitter through automated learning of naïve language. *PLoS ONE* 8, e82489. doi: 10.1371/journal.pone.0082489
- Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., Brilliant, L., et al. (2009). Detecting influenza epidemics using search engine query data. *Nature* 457, 1012–1014. doi: 10.1038/nature07634
- Gladstone, J. J., Matz, S. C., and Lemaire, A. (2019). Can psychological traits be inferred from spending? evidence from transaction data. *Psychol. Sci.* 30, 1087–1096. doi: 10.1177/0956797619849435
- Glaeser, E. (2012). *Triumph of the City: How our Greatest Invention makes us Richer, Smarter, Greener, Healthier, and Happier*. New York, NY: Penguin Books. doi: 10.17323/1726-3247-2013-4-75-94
- Glaeser, E., Kolko, J., and Saiz, A. (2001). Consumer city. *J. Econ. Geogr.* 1, 27–50. doi: 10.1093/jeg/1.1.27
- Glodeanu, A., Gullón, P., and Bilal, U. (2021). Social inequalities in mobility during and following the covid-19 associated lockdown of the Madrid metropolitan area in Spain. *Health Place* 70, 102580. doi: 10.1016/j.healthplace.2021.102580



- Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A.-L. (2008). Understanding individual human mobility patterns. *Nature* 453, 779–782. doi: 10.1038/nature06958
- Gore, R. J., Diallo, S., and Padilla, J. (2015). You are what you tweet: connecting the geographic variation in America's obesity rate to twitter content. *PLoS ONE* 10, e0133505. doi: 10.1371/journal.pone.0133505
- Gozzi, N., Tizzoni, M., Chinazzi, M., Ferres, L., Vespignani, A., Perra, N., et al. (2021). Estimating the effect of social inequalities on the mitigation of covid-19 across communities in Santiago de Chile. *Nat. Commun.* 12, 1–9. doi: 10.1038/s41467-021-22601-6
- Graif, C., Freelin, B. N., Kuo, Y.-H., Wang, H., Li, Z., Kifer, D., et al. (2021). Network spillovers and neighborhood crime: a computational statistics analysis of employment-based networks of neighborhoods. *Justice Q.* 38, 344–374. doi: 10.1080/07418825.2019.1602160
- Graif, C., Gladfelter, A. S., and Matthews, S. A. (2014). Urban poverty and neighborhood effects on crime: incorporating spatial and network perspectives: neighborhood poverty, crime, and exposure networks. *Sociol. Compass* 8, 1140–1155. doi: 10.1111/soc4.12199
- Grassly, N. C., and Fraser, C. (2008). Mathematical models of infectious disease transmission. *Nat. Rev. Microbiol.* 6, 477–487. doi: 10.1038/nrmicro1845
- Green, B., Horel, T., and Papachristos, A. V. (2017). Modeling contagion through social networks to explain and predict gunshot violence in Chicago, 2006 to 2014. *JAMA Intern. Med.* 177, 326–333. doi: 10.1001/jamainternmed.2016.8245
- Groff, E., Johnson, S. D., and Thornton, A. (2019). State of the art in agent-based modeling of urban crime: an overview. *J. Quant. Criminol.* 35, 155–193. doi: 10.1007/s10940-018-9376-y
- Groff, E., and Mazerolle, L. (2008). Simulated experiments and their potential role in criminology and criminal justice. *J. Exp. Criminol.* 4, 187. doi: 10.1007/s11292-008-9058-0
- Guan, G., Mofaz, M., Qian, G., Patalon, T., Shmueli, E., Yamin, D., et al. (2022). Higher sensitivity monitoring of reactions to covid-19 vaccination using smartwatches. *NPJ Digit. Med.* 5, 1–9. doi: 10.1038/s41746-022-00683-w
- Gündoğdu, D., Panzarasa, P., Oliver, N., Lepri, B. (2019). The bridging and bonding structures of place-centric networks: evidence from a developing country. *PLoS ONE* 14, e0221148. doi: 10.1371/journal.pone.0221148
- Haleem, M. S., Do Lee, W., Ellison, M., and Bannister, J. (2021). The Sexposed population, violent crime in public space and the night-time economy in Manchester, UK. *Eur. J. Crim. Policy Res.* 27, 335–352. doi: 10.1007/s10610-020-09452-5
- Harris, J. K., Mansour, R., Choucair, B., Olson, J., Nissen, C., Bhatt, J., et al. (2014). Health department use of social media to identify foodborne illness-Chicago, Illinois, 2013-2014. *Morbidity Mortal. Wkly. Rep.* 63, 681.
- Harrison, C., Jorder, M., Stern, H., Stavinsky, F., Reddy, V., Hanson, H., et al. (2014). Using online reviews by restaurant patrons to identify unreported cases of foodborne illness-new york city, 2012-2013. *Morbidity Mortal. Wkly. Rep.* 63, 441.
- Hayward, K. J., and Maas, M. M. (2021). Artificial intelligence and crime: a primer for criminologists. *Crime Media Cult.* 17, 209–233. doi: 10.1177/1741659020917434
- Hedefalk, F., and Dribe, M. (2020). The social context of nearest neighbors shapes educational attainment regardless of class origin. *Proc. Natl Acad. Sci.* 117, 14918–14925. doi: 10.1073/pnas.1922532117
- Hilman, R. M., Iñiguez, G., and Karsai, M. (2022). Socioeconomic biases in urban mixing patterns of us metropolitan areas. *EPJ Data Sci.* 11, 32. doi: 10.1140/epjds/s13688-022-00341-x
- Hipp, J. R., Bates, C., Lichman, M., and Smyth, P. (2019). Using social media to measure temporal ambient population: does it help explain local crime rates? *Justice Q.* 36, 718–748. doi: 10.1080/07418825.2018.1445276
- Holbrook, A. J., Loeffler, C. E., Flaxman, S. R., and Suchard, M. A. (2021). Scalable Bayesian inference for self-excitatory stochastic processes applied to big American gunfire data. *Stat. Comput.* 31, 4. doi: 10.1007/s1222-020-09980-4
- Hong, B., Bonczak, B. J., Gupta, A., and Kontokosta, C. E. (2021). Measuring inequality in community resilience to natural disasters using large-scale mobility data. *Nat. Commun.* 12, 1–9. doi: 10.1038/s41467-021-22160-w
- Hunter, R. F., Garcia, L., de Sa, T. H., Zapata-Diomed, B., Millett, C., Woodcock, E., et al. (2021). Effect of covid-19 response policies on walking behavior in us cities. *Nat. Commun.* 12, 1–9. doi: 10.1038/s41467-021-23937-9
- Icove, D. J. (1986). Automated crime profiling. *FBI Law Enforcement Bulletin*, 55.
- Isella, L., Romano, M., Barrat, A., Cattuto, C., Lizzola, V., Van den Broeck, W., et al. (2011). Close encounters in a pediatric ward: measuring face-to-face proximity and mixing patterns with wearable sensors. *PLoS ONE* 6, e17144. doi: 10.1371/journal.pone.0017144
- Jacobs, J. (1961). *The Death and Life of Great American Cities*. New York, NY: Random House.
- Järv, O., Müürisep, K., Ahas, R., Derudder, B., and Witlox, F. (2015). Ethnic differences in activity spaces as a characteristic of segregation: a study based on mobile phone usage in Tallinn, Estonia. *Urban Stud.* 52, 2680–2698. doi: 10.1177/0042098014550459
- Jean, N., Burke, M., Xie, M., Davis, W. M., Lobell, D. B., Ermon, S., et al. (2016). Combining satellite imagery and machine learning to predict poverty. *Science* 353, 790–794. doi: 10.1126/science.aaf7894
- Johnson, S. D. (2010). A brief history of the analysis of crime concentration. *Eur. J. Appl. Math.* 21, 349–370. doi: 10.1017/S0956792510000082
- Kermack, W. O., and McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond.* 115, 700–721. doi: 10.1098/rspa.1927.0118
- Kertész, J., and Wachs, J. (2021). Complexity science approach to economic crime. *Nat. Rev. Phys.* 3, 70–71. doi: 10.1038/s42254-020-0238-9
- Kishore, N. (2021). Mobility data as a proxy for epidemic measures. *Nat. Comput. Sci.* 1, 567–568. doi: 10.1038/s43588-021-00127-7
- Koch, T. (2011). *Disease Maps: Epidemics on the Ground*. Chicago, IL: University of Chicago Press. doi: 10.7208/chicago/9780226449401.001.0001
- Kraemer, M. U., Sadilek, A., Zhang, Q., Marchal, N. A., Tuli, G., Cohn, E. L., et al. (2020). Mapping global variation in human mobility. *Nat. Hum. Behav.* 4, 800–810. doi: 10.1038/s41562-020-0875-0
- Lamos, V., Bie, T. D., and Cristianini, N. (2010). “Flu detector-tracking epidemics on twitter,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Berlin: Springer), 599–602. doi: 10.1007/978-3-642-15939-8\_42
- Lamos, V., Majumder, M. S., Yom-Tov, E., Edelstein, M., Moura, S., Hamada, Y., et al. (2021). Tracking covid-19 using online search. *NPJ Digit. Med.* 4, 1–11. doi: 10.1038/s41746-021-00384-w
- Lamos, V., Miller, A. C., Crossan, S., and Stefansen, C. (2015). Advances in nowcasting influenza-like illness rates using search query logs. *Sci. Rep.* 5, 1–10. doi: 10.1038/srep12760
- Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014a). *Google Flu Trends Still Appears Sick: An Evaluation of the 2013-2014 Flu Season*. Available at SSRN 2408560. doi: 10.2139/ssrn.2408560
- Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014b). The parable of google flu: traps in big data analysis. *Science* 343, 1203–1205. doi: 10.1126/science.1248506
- Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A. L., Brewer, D., et al. (2009). Computational social science. *Science* 323, 721. doi: 10.1126/science.1167742
- Lazer, D., Pentland, A., Watts, D., Aral, S., Athey, S., Contractor, N., et al. (2020). Computational social science: obstacles and opportunities. *Science* 369, 1060–1062. doi: 10.1126/science.aaz8170
- Lepri, B., Centellegher, S., and Nadai, M. D. (2022). “Understanding and rewiring cities,” in *European Conference on Advances in Databases and Information Systems* (Cham: Springer), 3–10. doi: 10.1007/978-3-031-15740-0\_1
- Levy, B. L., Phillips, N. E., and Sampson, R. J. (2020). Triple disadvantage: neighborhood networks of everyday urban mobility and violence in U.S. cities. *Am. Sociol. Rev.* 85, 925–956. doi: 10.1177/0003122420972323
- Li, X., Huang, X., Li, D., and Xu, Y. (2022). Aggravated social segregation during the covid-19 pandemic: evidence from crowdsourced mobility data in twelve most populated us metropolitan areas. *Sustain. Cities Soc.* 81, 103869. doi: 10.1016/j.scs.2022.103869
- Linning, S. J., Andresen, M. A., and Brantingham, P. J. (2017). Crime seasonality: examining the temporal fluctuations of property crime in cities with varying climates. *Int. J. Offender Ther. Comp. Criminol.* 61, 1866–1891. doi: 10.1177/0306624X16632259
- Llorente, A., Garcia-Herranz, M., Cebrian, M., and Moro, E. (2015). Social media fingerprints of unemployment. *PLoS ONE* 10, e0128692. doi: 10.1371/journal.pone.0128692
- Loeffler, C., and Flaxman, S. (2018). Is gun violence contagious? A spatiotemporal test. *J. Quant. Criminol.* 34, 999–1017. doi: 10.1007/s10940-017-9363-8
- Logan, J. R., and Burdick-Will, J. (2016). School segregation, charter schools, and access to quality education. *J. Urban Aff.* 38, 323–343. doi: 10.1111/juaf.12246
- Lu, F. S., Hou, S., Baltrusaitis, K., Shah, M., Leskovec, J., Hawkins, J., et al. (2018). Accurate influenza monitoring and forecasting using novel internet data streams: a case study in the boston metropolis. *JMIR Public Health Surveill.* 4, e8950. doi: 10.2196/publichealth.8950
- Luca, M., Barlacchi, G., Oliver, N., and Lepri, B. (2021). Leveraging mobile phone data for migration flows. *arXiv*. [preprint]. doi: 10.48550/arXiv.2105.14956
- Luca, M., Lepri, B., Frias-Martinez, E., and Lutu, A. (2022). Modeling international mobility using roaming cell phone traces during covid-19 pandemic. *EPJ Data Sci.* 11, 22. doi: 10.1140/epjds/s13688-022-00335-9
- Lucchini, L., Centellegher, S., Pappalardo, L., Gallotti, R., Privitera, F., Lepri, B., et al. (2021). Living in a pandemic: changes in mobility routines, social activity and adherence to covid-19 protective measures. *Sci. Rep.* 11, 1–12. doi: 10.1038/s41598-021-04139-1
- Lum, K., and Isaac, W. (2016). To predict and serve? *Significance* 13, 14–19. doi: 10.1111/j.1740-9713.2016.00960.x

- Luna-Pla, I., and Nicolás-Carlock, J. R. (2020). Corruption and complexity: a scientific framework for the analysis of corruption networks. *Appl. Netw. Sci.* 5, 13. doi: 10.1007/s41109-020-00258-2
- Mackey, T., Kalyanam, J., Klugman, J., Kuzmenko, E., and Gupta, R. (2018). Solution to detect, classify, and report illicit online marketing and sales of controlled substances via twitter: using machine learning and web forensics to combat digital opioid access. *J. Med. Internet Res.* 20, e10029. doi: 10.2196/10029
- Magliocca, N. R., McSweeney, K., Sesnie, S. E., Tellman, E., Devine, J. A., Nielsen, E. A., et al. (2019). Modeling cocaine traffickers and counterdrug interdiction forces as a complex adaptive system. *Proc. Nat. Acad. Sci.* 116, 7784–7792. doi: 10.1073/pnas.1812459116
- Malleson, N., and Andresen, M. A. (2015). The impact of using social media data in crime rate calculations: shifting hot spots and changing spatial patterns. *Cartogr. Geogr. Inf. Sci.* 42. doi: 10.1080/15230406.2014.905756
- Manduca, R., and Sampson, R. J. (2019). Punishing and toxic neighborhood environments independently predict the intergenerational social mobility of black and white children. *Proc. Natl Acad. Sci.* 116, 7772–7777. doi: 10.1073/pnas.1820464116
- Mazeika, D. M., and Kumar, S. (2017). Do crime hot spots exist in developing countries? Evidence from india. *J. Quant. Criminol.* 33, 45–61. doi: 10.1007/s10940-016-9280-2
- Mazzoli, M., Molas, A., Bassolas, A., Lenormand, M., Colet, P., Ramasco, J. J., et al. (2019). Field theory for recurrent mobility. *Nat. Commun.* 10, 3895. doi: 10.1038/s41467-019-11841-2
- McIver, D. J., and Brownstein, J. S. (2014). Wikipedia usage estimates prevalence of influenza-like illness in the united states in near real-time. *PLoS Comput. Biol.* 10, e1003581. doi: 10.1371/journal.pcbi.1003581
- Mejova, Y., Haddadi, H., Noulas, A., and Weber, I. (2015a). “# foodporn: obesity patterns in culinary interactions,” in *Proceedings of the 5th International Conference on Digital Health* (Florence: Association for Computing Machinery) 2015, 51–58. doi: 10.1145/2750511.2750524
- Mejova, Y., Weber, I., and Macy, M. W. (2015b). *Twitter: A Digital Socioscope*. Cambridge, MA: Cambridge University Press. doi: 10.1017/CBO9781316182635
- Merler, S., and Ajelli, M. (2010). The role of population heterogeneity and human mobility in the spread of pandemic influenza. *Proc. R. Soc. B Biol. Sci.* 277, 557–565. doi: 10.1098/rspb.2009.1605
- Miliou, I., Xiong, X., Rinzivillo, S., Zhang, Q., Rossetti, G., Giannotti, F., et al. (2021). Predicting seasonal influenza using supermarket retail records. *PLoS Comput. Biol.* 17, e1009087. doi: 10.1371/journal.pcbi.1009087
- Mohler, G. (2014). Marked point process hotspot maps for homicide and gun crime prediction in Chicago. *Int. J. Forecast.* 30, 491–497. doi: 10.1016/j.ijforecast.2014.01.004
- Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg, F. P., and Tita, G. E. (2011). Self-exciting point process modeling of crime. *J. Am. Stat. Assoc.* 106, 100–108. doi: 10.1198/jasa.2011.ap09546
- Moon, I.-C., and Carley, K. M. (2007). Modeling and simulating terrorist networks in social and geospatial dimensions. *IEEE Intell. Syst.* 22, 40–49. doi: 10.1109/MIS.2007.4338493
- Morales, A. J., Dong, X., and Bar-Yam, Y. SandyPentland, A. (2019). Segregation and polarization in urban areas. *R. Soc. Open Sci.* 6, 190573. doi: 10.1098/rsos.190573
- Moreira-Matias, L., Gama, J., Ferreira, M., Mendes-Moreira, J., and Damas, L. (2013). Predicting taxi-passenger demand using streaming data. *IEEE Trans. Intell. Transp. Syst.* 14, 1393–1402. doi: 10.1109/TITS.2013.2262376
- Moro, E., Calacci, D., Dong, X., and Pentland, A. (2021). Mobility patterns are associated with experienced income segregation in large us cities. *Nat. Commun.* 12, 1–10. doi: 10.1038/s41467-021-24899-8
- Nagar, R., Yuan, Q., Freifeld, C. C., Santillana, M., Nojima, A., Chunara, R., et al. (2014). A case study of the new york city 2012–2013 influenza season with daily geocoded twitter data from temporal and spatiotemporal perspectives. *J. Med. Internet Res.* 16, e3416. doi: 10.2196/jmir.3416
- Nardin, L. G., Andrighetto, G., Conte, R., Székely, A., Anzola, D., Elsenbroich, C., et al. (2016). Simulating protection rackets: a case study of the Sicilian Mafia. *Auton. Agent Multi. Agent Syst.* 30, 1117–1147. doi: 10.1007/s10458-016-9330-z
- Neiderud, C.-J. (2015). How urbanization affects the epidemiology of emerging infectious diseases. *Infect. Ecol. Epidemiol.* 5, 27060. doi: 10.3402/iee.v5.27060
- Neves, P. C., Afonso, O., and Silva, S. T. (2016). A meta-analytic reassessment of the effects of inequality on growth. *World Dev.* 78, 386–400. doi: 10.1016/j.worlddev.2015.10.038
- Nguyen, Q. C., Li, D., Meng, H.-W., Kath, S., Nsoesie, E., Li, F., et al. (2016). Building a national neighborhood dataset from geotagged twitter data for indicators of happiness, diet, and physical activity. *JMIR Public Health Surveill.* 2, e5869. doi: 10.2196/publichealth.5869
- Nicoletti, L., Sirenko, M., and Verma, T. (2022). Disadvantaged communities have lower access to urban infrastructure. *arXiv.* 50, 831–849. doi: 10.48550/arXiv.2203.13784
- Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., et al. (2020). Mobile phone data for informing public health actions across the covid-19 pandemic life cycle. *Sci. Adv.* 6, eabc0764. doi: 10.1126/sciadv.abc0764
- Pan, W., Ghoshal, G., Krumme, C., Cebrian, M., and Pentland, A. (2013). Urban characteristics attributable to density-driven tie formation. *Nat. Commun.* 4, 1–7. doi: 10.1038/ncomms2961
- Pangallo, M., Aleta, A., Chanona, R., Pichler, A., Martín-Corral, D., Chinazzi, M., et al. (2022). The unequal effects of the health-economy tradeoff during the covid-19 pandemic. *arXiv.* [preprint]. doi: 10.48550/arXiv.2212.03567
- Paolotti, D., Carnahan, A., Colizza, V., Eames, K., Edmunds, J., Gomes, G., et al. (2014). Web-based participatory surveillance of infectious diseases: the influenzanet participatory surveillance experience. *Clin. Microbiol. Infect.* 20, 17–21. doi: 10.1111/1469-0691.12477
- Papachristos, A. V., and Bastomski, S. (2018). Connected in crime: the enduring effect of neighborhood networks on the spatial patterning of violence. *Am. J. Sociol.* 124, 517–568. doi: 10.1086/699217
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nat. Commun.* 6, 1–8. doi: 10.1038/ncomms9166
- Paul, M., and Dredze, M. (2011). “You are what you tweet: analyzing twitter for public health,” in *Proceedings of the International AAAI Conference on Web and Social Media, Volume* (Barcelona: AAAI Press), 5, 265–272. doi: 10.1609/icwsm.v5i1.14137
- Perkins, T. A., Garcia, A. J., Paz-Soldán, V. A., Stoddard, S. T., Reiner Jr, R. C., Vazquez-Prokopec, G., et al. (2014). Theory and data for simulating fine-scale human movement in an urban environment. *J. R. Soc. Interface* 11, 20140642. doi: 10.1098/rsif.2014.0642
- Perra, N., Balcan, D., Gonçalves, B., and Vespignani, A. (2011). Towards a characterization of behavior-disease models. *PLoS ONE* 6, e23084. doi: 10.1371/journal.pone.0023084
- Perry, W. L. (2013). *Predictive Policing: The Role of Crime Forecasting in Law Enforcement Operations*. Arlington, VA: Rand Corporation. doi: 10.7249/RR233
- Peterson, R. D., and Krivo, L. J. (2009). Segregated spatial locations, race-ethnic composition, and neighborhood violent crime. *Ann. Am. Acad. Pol. Soc. Sci.* 623, 93–107. doi: 10.1177/0002716208330490
- Phillips, N. E., Levy, B. L., Sampson, R. J., Small, M. L., and Wang, R. Q. (2021). The social integration of american cities: network measures of connectedness based on everyday mobility across neighborhoods. *Soc. Methods Res.* 50, 1110–1149. doi: 10.1177/0049124119852386
- Piatkowska, S. J., and Lantz, B. (2021). Temporal clustering of hate crimes in the aftermath of the brexit vote and terrorist attacks: a comparison of Scotland and England and Wales. *Br. J. Criminol.* 61, 648–669. doi: 10.1093/bjc/zaaa090
- Pickett, K. E., and Wilkinson, R. G. (2015). Income inequality and health: a causal review. *Soc. Sci. Med.* 128, 316–326. doi: 10.1016/j.socscimed.2014.12.031
- Poletto, C., Scarpino, S. V., and Volz, E. M. (2020). Applications of predictive modelling early in the covid-19 epidemic. *Lancet Digit. Health* 2, e498–e499. doi: 10.1016/S2589-7500(20)30196-5
- Polgreen, P. M., Chen, Y., Pennock, D. M., Nelson, F. D., and Weinstein, R. A. (2008). Using internet searches for influenza surveillance. *Clin. Infect. Dis.* 47, 1443–1448. doi: 10.1086/593098
- Purves, D. (2022). Fairness in algorithmic policing. *J. Am. Philos. Assoc.* 8, 1–21. doi: 10.1017/apa.2021.39
- Quetelet, A. (1831). *Research on the Propensity for Crime at Different Ages*. Cincinnati, OH: Anderson.
- Quillian, L. (2014). Does segregation create winners and losers? Residential segregation and inequality in educational attainment. *Soc. Probl.* 61, 402–426. doi: 10.1525/sp.2014.12193
- Rader, B., Scarpino, S. V., Nande, A., Hill, A. L., Adlam, B., Reiner, R. C., et al. (2020). Crowding and the shape of covid-19 epidemics. *Nat. Med.* 26, 1829–1834. doi: 10.1038/s41591-020-1104-0
- Reid, E., Hamidi, S., Grace, J., and Wei, Y. (2016). Does urban sprawl hold down upward mobility. *Landsc. Urban Plan.* 148, 80–88. doi: 10.1016/j.landurbplan.2015.11.012
- Ribeiro, H. V., Alves, L. G. A., Martins, A. F., Lenzi, E. K., and Perc, M. (2018). The dynamical structure of political corruption networks. *J. Complex Netw.* 6, 989–1003. doi: 10.1093/comnet/cny002
- Richardson, R., Schultz, J., and Crawford, K. (2019). *Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice*. SSRN Scholarly Paper ID 3333423. Rochester, NY: Social Science Research Network.
- Rocha, L. E., Thorson, A. E., and Lambiotte, R. (2015). The non-linear health consequences of living in larger cities. *J. Urban Health* 92, 785–799. doi: 10.1007/s11524-015-9976-x
- Rogers, D. J., Randolph, S. E., Snow, R. W., and Hay, S. I. (2002). Satellite imagery in the study and forecast of malaria. *Nature* 415, 710–715. doi: 10.1038/415710a

- Ross, R. (1916). An application of the theory of probabilities to the study of a priori pathometry. part I. *Proc. R. Soc. London A* 92, 204–230. doi: 10.1098/rspa.1916.0007
- Ross, R., Kadar, C., Gerritsen, C., and Rouly, C. (2020). Simulating offender mobility: modeling activity nodes from large-scale human activity data. *J. Artif. Intell. Res.* 8, 541–570. doi: 10.1613/jair.1.11831
- Ross, R., Kadar, C., and Malleon, N. (2021). A data-driven agent-based simulation to predict crime patterns in an urban environment. *Comput. Environ. Urban Syst.* 89, 101660. doi: 10.1016/j.compenvurbysys.2021.101660
- Rvachev, L. A., and Longini Jr, I. M. (1985). A mathematical model for the global spread of influenza. *Math. Biosci.* 75, 3–22. doi: 10.1016/0025-5564(85)90064-1
- Sadilek, A., and Kautz, H. (2013). “Modeling the impact of lifestyle on health at scale,” in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining* (Rome: Association for Computing Machinery), 637–646. doi: 10.1145/2433396.2433476
- Sadilek, A., Kautz, H., and Silenzio, V. (2012). “Predicting disease transmission from geo-tagged micro-blog data,” in *Twenty-Sixth AAAI Conference on Artificial Intelligence* (Toronto: AAAI Press).
- Salathé, M. (2018). Digital epidemiology: what is it, and where is it going? *Life Sci. Soc. Policy* 14, 1–5. doi: 10.1186/s40504-017-0065-7
- Salathe, M., Bengtsson, L., Bodnar, T. J., Brewer, D. D., Brownstein, J. S., Buckee, C., et al. (2012). Digital epidemiology. *PLoS Comput. Biol.* 8, e1002616. doi: 10.1371/journal.pcbi.1002616
- Sampson, R., and Levy, B. (2022). The enduring neighborhood effect, everyday urban mobility, and violence in Chicago. *Univ. Chicago Law Rev.* 89, 2.
- Sampson, R. J. (2012). *Great American City: Chicago and the Enduring Neighborhood Effect*. Chicago, IL: Univ. of Chicago Press. doi: 10.7208/chicago/9780226733883.001.0001
- Sampson, R. J. (2019). Neighbourhood effects and beyond: explaining the paradoxes of inequality in the changing american metropolis. *Urban Stud.* 56, 3–32. doi: 10.1177/0042098018795363
- Sarker, A., Gonzalez-Hernandez, G., Ruan, Y., and Perrone, J. (2019). Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter. *JAMA Netw. Open* 2, e1914672. doi: 10.1001/jamanetworkopen.2019.14672
- Sassen, S. (2019). *Cities in a World Economy*. Van Nuys, CA: AGE Publications Inc. doi: 10.4135/9781071872710
- Sattenspiel, L., and Dietz, K. (1995). A structured epidemic model incorporating geographic mobility among regions. *Math. Biosci.* 128, 71–91. doi: 10.1016/0025-5564(94)00068-B
- Schläpfer, M., Bettencourt, L. M., Grauwin, S., Raschke, M., Claxton, R., Smoreda, Z., et al. (2014). The scaling of human interactions with city size. *J. R. Soc. Interface* 11, 20130789. doi: 10.1098/rsif.2013.0789
- Shalaginov, A., Johnsen, J. W., and Franke, K. (2017). “Cyber crime investigations in the era of big data,” in *2017 IEEE International Conference on Big Data (Big Data)* (Boston, MA), 3672–3676. doi: 10.1109/BigData.2017.8258362
- Shaman, J., and Karspeck, A. (2012). Forecasting seasonal outbreaks of influenza. *Proc. Nat. Acad. Sci.* 109, 20425–20430. doi: 10.1073/pnas.1208772109
- Shaman, J., Karspeck, A., Yang, W., Tamerius, J., and Lipsitch, M. (2013). Real-time influenza forecasts during the 2012–2013 season. *Nat. Commun.* 4, 1–10. doi: 10.1038/ncomms3837
- Shaw, C. R., and McKay, H. D. (1942). *Juvenile Delinquency and Urban Areas*. *Juvenile Delinquency and Urban Areas*. Chicago, IL: University of Chicago Press, xxxii, 451. doi: 10.2307/1334446
- Shelton, T., Poorthuis, A., and Zook, M. (2015). Social media and the city: rethinking urban socio-spatial inequality using user-generated geographic information. *Landsc. Urban Plan.* 142, 198–211. doi: 10.1016/j.landurbplan.2015.02.020
- Shiode, N., Shiode, S., Rod-Thatcher, E., Rana, S., and Vinten-Johansen, P. (2015). The mortality rates and the space-time patterns of john snow’s cholera epidemic map. *Int. J. Health Geogr.* 14, 1–15. doi: 10.1186/s12942-015-0011-y
- Sigalo, N., St Jean, B., Frias-Martinez, V., et al. (2022). Using social media to predict food deserts in the united states: infodemiology study of tweets. *JMIR Public Health Surveill.* 8, e34285. doi: 10.2196/34285
- Simini, F., González, M. C., Maritan, A., and Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature* 484, 96–100. doi: 10.1038/nature10856
- Singh, V. K., Bozkaya, B., and Pentland, A. (2015). Money walks: implicit mobility behavior and financial well-being. *PLoS ONE* 10, e0136628. doi: 10.1371/journal.pone.0136628
- Smolinski, M. S., Crawley, A. W., Baltrusaitis, K., Chunara, R., Olsen, J. M., Wójcik, O., et al. (2015). Flu near you: crowdsourced symptom reporting spanning 2 influenza seasons. *Am. J. Public Health* 105, 2124–2130. doi: 10.2105/AJPH.2015.302696
- Song, G., Bernasco, W., Liu, L., Xiao, L., Zhou, S., Liao, W., et al. (2019). Crime feeds on legal activities: daily mobility flows help to explain thieves’ target location choices. *J. Quant. Criminol.* 35, 831–854. doi: 10.1007/s10940-019-09406-z
- Spinsanti, L., Berlingiero, M., and Pappalardo, L. (2013). *Mobility and Geo-Social Networks*. Cambridge, MA: Cambridge University Press, 315–333. doi: 10.1017/CBO9781139128926.017
- Strano, E., Simini, F., De Nadai, M., Esch, T., and Marconcini, T. M. (2021). The agglomeration and dispersion dichotomy of human settlements on earth. *Sci. Rep.* 11, 1–10. doi: 10.1038/s41598-021-02743-9
- Tatem, A. J. (2009). The worldwide airline network and the dispersal of exotic species: 2007–2010. *Ecography* 32, 94–102. doi: 10.1111/j.1600-0587.2008.05588.x
- Tatem, A. J. (2017). Worldpop, open data for spatial demography. *Sci. Data* 4, 1–4. doi: 10.1038/sdata.2017.4
- Tatem, A. J., Qiu, Y., Smith, D. L., Sabot, O., Ali, A. S., Moonen, B., et al. (2009). The use of mobile phone data for the estimation of the travel patterns and imported plasmodium falciparum rates among zanzibar residents. *Malar. J.* 8, 1–12. doi: 10.1186/1475-2875-8-287
- Tita, G. E., and Greenbaum, R. T. (2009). “Crime, neighborhoods, and units of analysis: putting space in its place,” in *Putting Crime in its Place*, eds D. Weisburd, W. Bernasco, G. J. Bruinsma (New York, NY: Springer New York), 145–170. doi: 10.1007/978-0-387-09688-9\_7
- Tizzoni, M., Nsoesie, E. O., Gauvin, L., Karsai, M., Perra, N., Bansal, S., et al. (2022). Addressing the socioeconomic divide in computational modeling for infectious diseases. *Nat. Commun.* 13, 1–7. doi: 10.1038/s41467-022-30688-8
- Tizzoni, M., Panisson, A., Paolotti, D., and Cattuto, C. (2020). The impact of news exposure on collective attention in the united states during the 2016 zika epidemic. *PLoS Comput. Biol.* 16, e1007633. doi: 10.1371/journal.pcbi.1007633
- Tizzoni, M., Sun, K., Benusiglio, D., Karsai, M., and Perra, N. (2015). The scaling of human contacts and epidemic processes in metapopulation networks. *Sci. Rep.* 5, 1–11. doi: 10.1038/srep15111
- Tollenaar, N., and van der Heijden, P. G. M. (2013). Which method predicts recidivism best?: a comparison of statistical, machine learning and data mining predictive models: which Method Predicts Recidivism Best? *J. R. Stat. Soc. A* 176, 565–584. doi: 10.1111/j.1467-985X.2012.01056.x
- Tovanich, N., Centellegher, S., Seghouani, N. B., Gladstone, J., Matz, S., Lepri, B., et al. (2021). Inferring psychological traits from spending categories and dynamic consumption patterns. *EPJ Data Sci.* 10, 24. doi: 10.1140/epjds/s13688-021-00281-y
- Traunmueller, M., Quattrone, G., and Capra, L. (2014). “Mining mobile phone data to investigate urban crime theories at scale,” in *Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11–13, 2014. Proceedings*, eds L. M. Aiello, D. McFarland (Cham: Springer International Publishing) Lecture Notes in Computer Science, 396–411. doi: 10.1007/978-3-319-13734-6\_29
- Troitzsch, K. G. (2017). Can agent-based simulation models replicate organised crime? *Trends Organ. Crime* 20, 100–119. doi: 10.1007/s12117-016-9298-8
- Tucker, R., O’Brien, D. T., Ciomek, A., Castro, E., Wang, Q., Phillips, N. E., et al. (2021). Who tweets where and when, and how does it help understand crime rates at places? Measuring the presence of tourists and commuters in ambient populations. *J. Quant. Criminol.* 37, 333–359. doi: 10.1007/s10940-020-09487-1
- Umar, F., Johnson, S. D., and Cheshire, J. A. (2020). Assessing the spatial concentration of urban crime: an insight from Nigeria. *J. Quant. Criminol.* 37, 605–624. doi: 10.1007/s10940-019-09448-3
- Viboud, C., and Santillana, M. (2020). Fitbit-informed influenza forecasts. *Lancet Digit. Health* 2, e54–e55. doi: 10.1016/S2589-7500(19)30241-9
- Wang, H., Kifer, D., Graif, C., and Li, Z. (2016). “Crime rate inference with big data,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD ’16* (New York, NY: Association for Computing Machinery), 635–644. doi: 10.1145/2939672.2939736
- Wang, H., Yao, H., Kifer, D., Graif, C., and Li, Z. (2019). Non-stationary model for crime rate inference using modern urban data. *IEEE Trans. Big Data* 5, 180–194. doi: 10.1109/TBDDATA.2017.2786405
- Wang, M., and Gerber, M. S. (2015). “Using twitter for next-place prediction, with an application to crime prediction,” in *2015 IEEE Symposium Series on Computational Intelligence* (Cape Town, SA: IEEE), 941–948. doi: 10.1109/SSCI.2015.138
- Wang, Q., Phillips, N. E., Small, M. L., and Sampson, R. J. (2018). Urban mobility and neighborhood isolation in america’s 50 largest cities. *Proc. Natl Acad. Sci. USA* 115, 7735–7740. doi: 10.1073/pnas.1802537115
- Wang, X., Gerber, M. S., and Brown, D. E. (2012). “Automatic crime prediction using events extracted from twitter posts,” in *Social Computing, Behavioral - Cultural Modeling and Prediction, Lecture Notes in Computer Science*, eds S. J. Yang, A. M. Greenberg, M. Endsley (Berlin, Heidelberg: Springer), 231–238. doi: 10.1007/978-3-642-29047-3\_28
- Weber, E. M., Seaman, V. Y., Stewart, R. N., Bird, T. J., Tatem, A. J., McKee, J. J., et al. (2018). Census-independent population mapping in Northern Nigeria. *Remote Sens. Environ.* 204, 786–798. doi: 10.1016/j.rse.2017.09.024

- Weisburd, D. (2015). The law of crime concentration and the criminology of place. *Criminology* 53, 133–157. doi: 10.1111/1745-9125.12070
- Wesolowski, A., Buckee, C. O., Engø-Monsen, K., and Metcalf, C. J. E. (2016). Connecting mobility to infectious diseases: the promise and limits of mobile phone data. *J. Infect. Dis.* 214(suppl\_4), S414–S420. doi: 10.1093/infdis/jiw273
- Wesolowski, A., Eagle, N., Tatem, A. J., Smith, D. L., Noor, A. M., Snow, R. W., et al. (2012). Quantifying the impact of human mobility on malaria. *Science* 338, 267–270. doi: 10.1126/science.1223467
- Wiedermann, M., Rose, A. H., Maier, B. F., Kolb, J. J., Hinrichs, D., Brockmann, D., et al. (2022). Evidence for positive long- and short-term effects of vaccinations against covid-19 in wearable sensor metrics—insights from the German corona data donation project. *arXiv [preprint]*. doi: 10.48550/arXiv.2204.02846
- Wilkinson, R. G., and Pickett, K. E. (2006). Income inequality and population health: a review and explanation of the evidence. *Soc. Sci. Med.* 62, 1768–1784. doi: 10.1016/j.socscimed.2005.08.036
- Wo, J. C., Rogers, E. M., Berg, M. T., and Koyle, C. (2022). Recreating human mobility patterns through the lens of social media: using twitter to model the social ecology of crime. *Crime Delinq.* 00111287221106946. doi: 10.1177/00111287221106946
- Woolgar, S. (1989). “Why not a sociology of machines? An evaluation of prospects for an association between sociology and artificial intelligence,” in *Intelligent Systems in a Human Context: Development, Implications, and Applications* (Oxford: Oxford University Press, Inc.), 53–70.
- Xu, Y., Belyi, A., Santi, P., and Ratti, C. (2019). Quantifying segregation in an integrated urban physical-social space. *J. R. Soc. Interface* 16, 20190536. doi: 10.1098/rsif.2019.0536
- Yabe, T., Bueno, B. G. B., Dong, X., Pentland, A., and Moro, E. (2023). Behavioral changes during the COVID-19 pandemic decreased income diversity of urban encounters. *Nat. Commun.* 14, 2310. doi: 10.1038/s41467-023-37913-y
- Yan, Y. Y. (2004). Seasonality of property crime in Hong Kong. *Br. J. Criminol.* 44, 276–283. doi: 10.1093/bjc/44.2.276
- Yang, D., Heaney, T., Tonon, A., Wang, L., and Cudr-Mauroux, P. (2018). CrimeTelescope: crime hotspot prediction based on urban and social media data fusion. *World Wide Web* 21, 1323–1347. doi: 10.1007/s11280-017-0515-4
- Ye, X., Xu, X., Lee, J., Zhu, X., and Wu, L. (2015). Space time interaction of residential burglaries in Wuhan, China. *Appl. Geograp.* 60, 210–216. doi: 10.1016/j.apgeog.2014.11.022
- Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., et al. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in africa. *Nat. Commun.* 11, 1–11. doi: 10.1038/s41467-020-16185-w
- Yuan, Q., Nsoesie, E. O., Lv, B., Peng, G., Chunara, R., Brownstein, J. S., et al. (2013). Monitoring influenza epidemics in china with search query from baidu. *PLoS ONE* 8, e64323. doi: 10.1371/journal.pone.0064323
- Zachreson, C., Fair, K. M., Cliff, O. M., Harding, N., Piraveenan, M., Prokopenko, M., et al. (2018). Urbanization affects peak timing, prevalence, and bimodality of influenza pandemics in Australia: results of a census-calibrated model. *Sci. Adv.* 4, eaau5294. doi: 10.1126/sciadv.aau5294
- Zhang, Q., Gioannini, C., Paolotti, D., Perra, N., Perrotta, D., Quaggiotto, M., et al. (2015). “Social data mining and seasonal influenza forecasts: the fluoutlook platform,” in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases* (Cham: Springer), 237–240. doi: 10.1007/978-3-319-23461-8\_21
- Zhang, Q., Perra, N., Perrotta, D., Tizzoni, M., Paolotti, D., Vespignani, A., et al. (2017). “Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model,” in *Proceedings of the 26th international Conference on world wide web* (Perth: Association for Computing Machinery), 311–319. doi: 10.1145/3038912.3052678
- Zheng, Y., Xie, X., and Ma, W.-Y. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.* 33, 32–39.