Human Mobility Data in the COVID-19 Pandemic: Characteristics, Applications, and Challenges

Article *in* SSRN Electronic Journal · June 2021

DOI: 10.2139/ssrn.3851789

Human Mobility Data in the COVID-19 Pandemic: Characteristics, Applications, and Challenges

Tao Hu[1], Siqin Wang[2,*], Bing She[3], Mengxi Zhang[4], Xiao Huang[5,*], Yunhe Cui[6], Jacob Khuri[7], Yaxin Hu[8], Xiaokang Fu[9], Xiaoyue Wang[9], Peixiao Wang[9], Xinyan Zhu[9], Shuming Bao[10], Wendy Guan[1], Zhenlong Li[11]

1. Center for Geographic Analysis, Harvard University, Cambridge, MA 02138, USA. taohu@g.harvard.edu (T.H.);
2. School of Earth and Environmental Sciences, The University of Queensland, St Lucia, Australia, 4067. s.wang6@uq.edu.au (S.W.);
3. Institute for Social Research, University of Michigan, MI, USA, coolnanjizhou@gmail.com (B.S.);
4. School of Health, Ball State University, Muncie, Indiana 47304, USA; mzhang2@bsu.edu (M.Z);
5. Department of Geosciences, University of Arkansas, Fayetteville, AR, 72701, USA. xh010@uark.edu (X.H.);
6. University of Connecticut, Department of Geography, Storrs, CT 06269, USA; yunhe.cui@uconn.edu (Y.C.);
7. Independent Public Health Researcher, California, USA; jacobkhuri@outlook.com (J.K.);
8. School of Resource and Environmental Sciences, Wuhan University, Wuhan, Hubei 430079, China, yaya_cindy@whu.edu.cn (Y. H.);
9. State Key Laboratory of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, Hubei 430079, China, whistle_ant@whu.edu.cn (X.K.); 2017302590218@whu.edu.cn (X. W.); xinyanzhu@whu.edu.cn (X.Z.);
10. China Data Institute, Ann Arbor, MI, USA; sbao@umich.edu (S.B.)
11. Geoinformation and Big Data Research Laboratory, Department of Geography, University of South Carolina, SC, USA; zhenlong@sc.edu (Z.L.)

Abstract: The COVID-19 pandemic poses unprecedented challenges around the world. Many studies indicate that human mobility data provide significant support for public health actions during the pandemic. Researchers have applied mobility data to explore spatiotemporal trends over time, investigate associations with other variables, and predict or simulate the spread of COVID-19. Our objective was to provide a comprehensive overview of human mobility open data to guide researchers and policymakers in conducting data-driven evaluations and decision-making for the COVID-19 pandemic and other infectious disease outbreaks. We summarized the mobility data usage in COVID-19 studies by reviewing recent publications on COVID-19 and human mobility from a data-oriented perspective. We identified three major sources of mobility data: public transit systems, mobile operators, and mobile phone applications. Four approaches have been commonly used to estimate human mobility: public transit-based flow, social activity patterns, index-based mobility data, and social media-derived mobility data. We compared mobility datasets' characteristics by assessing data privacy, quality, space-time coverage, high-performance data storage and processing, and accessibility. We also present challenges and future directions of using mobility data. This review makes a pivotal contribution to understanding the use of and access to human mobility data in the COVID-19 pandemic and future disease outbreaks.

Keywords: COVID-19, public health, human mobility, open data, mobile phone, mobility index.

1. Introduction

Human mobility pertains to how people move across space and plays a crucial role in the spatiotemporal transmission dynamics of infectious diseases, including the coronavirus disease 2019 (COVID-19). The nature of person-to-person virus transmission of COVID-19 and the unprecedented global scale has led to the urgency of restricting human movement behaviors within and across national borders. Human mobility datasets characterize the pattern and trajectory of human activities, including, but not limited to, walking around local neighborhoods, driving to workplaces, or utilizing public transportation. Understanding human mobility with appropriate mobility data is crucial to urban planning (Zheng et al., 2020; Yuan et al., 2020; Qi et al., 2020), traffic forecasting (Jung et al., 2020; Goh et al., 2020; Pan et al., 2020), network applications (Zheng et al., 2020; Zhao et al., 2020; Rao et al., 2020), and epidemic control (Yang et al., 2020; Ni and Weng, 2009; Belik et al., 2011; Meloni et al., 2011; Changruenngam et al., 2020). As human mobility has a severe impact on epidemic spreading and the speed of disease spreading, measuring and examining human mobility have become increasingly important since the outbreak of COVID-19. Many COVID-19 studies have been published with common conclusions revealing that restricting human mobility (e.g., international travel bans, national border closures, lockdown orders, and limited gatherings) is the primary and effective strategy to reduce infections and curb the transmission of COVID-19 (Yang et al. 2020; Gatto et al. 2020; Kraemer et al. 2020; Liu et al. 2020a; Yabe et al. 2020). Now more than ever, the use of and access to human mobility data are imperative to controlling the spread of communicable and infectious disease outbreaks.

Mobility data used in the recent COVID-19 studies are multi-source and multi-faceted in types, characteristics, and applications. Apart from the traditional sources (e.g., survey and public transit system), the growth of telecommunication devices and mobile applications has significantly changed the way of data production and processing. Mobile phones utilize cell tower information and the Global Positioning System (GPS) for fine-grained location tracking. Billions of mobile users provide a large amount of spatiotemporal data that can be used to extract the trajectory of human movement activities. A growing volume of mobility data is collected and processed to generate new types of datasets via advanced algorithms. Data providers made such mobility data publicly available to facilitate COVID-19 research studies. There is a pressing need to review the human mobility data used in these studies to help researchers understand the spatiotemporal dynamics of the pandemic and to propose a future research agenda that will aid the prevention and control of disease outbreaks. To the best of our knowledge, there is no current review summarizing the human mobility data usage in recent COVID-19 studies.

To bridge this gap, we investigated and reviewed recent publications on COVID-19 and human mobility from a data-oriented perspective. The objectives of our review are threefold: 1) to classify and describe the mobility data used in COVID-19 studies in terms of data sources, measures, and characteristics; 2) to summarize how different types of mobility data have been used in COVID-19 studies; and 3) to highlight the challenges, recommendations, and future directions toward which we can orientate our collective efforts in utilizing mobility data in current COVID-19 and future infectious disease studies. To achieve the three objectives, we searched related literature by querying mobility dataset names in Google Scholar, including peer-reviewed journal articles, high-quality preprints, and working papers. Since the search terms may appear in either the title, abstract, or content, we conducted a manual selection of papers that applied human mobility data in their COVID-19 studies. We excluded papers that only described human mobility data without any application analysis. We summarized the mobility data from sources, categories, measures, and characteristics to estimations, applications, challenges, and future data usage of COVID-19 and other infectious disease modeling research. Our review makes a pivotal contribution to the current scholarship in public health and human mobility by presenting a primary summary of human mobility data highly pertinent to the current COVID-19 pandemic and future epidemic control.

2. Sources of Mobility Data

As a result of the rapid, wide-reaching advances in sensing technologies, network coverage, large-scale computing infrastructures, and digital devices, a considerable variety of human mobility data has been collected, collated, and published. The multi-source human mobility data contain rich, multi-faceted spatiotemporal information on human mobility patterns. We identified three primary sources of human mobility datasets used in the recent COVID-19 studies and summarized the publicly available datasets categorized within these three primary sources: 1) public transit systems; 2) mobile network operators; and 3) mobile phone applications.

2.1 Public Transit Systems

Mobility datasets retrieved from public transit systems, such as bus, train, ferry, metro, and air flight, represent the aggregated human mobility. There are two types of public transit data: the scheduled timetable and the actual travel records of passengers. The scheduled timetable data reflect the service capability and time, which can be potentially used to estimate the ridership and avenue if it is assumed to achieve the maximum service capability fully or proportionally (e.g., Watts et al., 2020; Zhuang et al., 2020). Most scheduled timetable data are publicly available on the website of transport authorities, governments, or private transport companies. In contrast, the actual travel records of passengers are usually retrieved from smart transit systems through the tapping of transit smart cards by passengers (Liu et al., 2019). As an example, the train dataset usually comprises the following information: trip/journey identification, smart card identification, timestamp, and alighting and boarding locations (longitude and latitude). Accordingly, an individual train passenger's mobility pattern can be reflected as the network distance between the boarding station being the origin and the alighting station being the destination with the recorded travel time (Carteni et al., 2020; Hu et al., 2020). Such smart-card data at the individual level can be provided by the transport and governmental authorities.

At the global level, air travel is the primary choice of public transit, and air flight data serve as the indicator of the mobility flow across countries and regions (e.g., Iacus et al., 2020; Haider et al., 2020). Air flight data include airport and airline data, online booking and trip information, and aircraft tracking data. Most of the data services provide either Application Programming Interface (API) or direct download links, and the historical travel data are usually accessible.

Train and bus data are usually available at the state or city level, whereas metro/subway data may be available at more locally precise scales (e.g., across and within a fare zone). However, the historical travel data for these public transit modes are typically not provided. Researchers often create crawlers to collect these data. At the city level, there are more diverse travel options by which people may choose to travel from one place to the next, including metro and bus. Most of the public transportation data from these transportation systems are provided by data service companies, but there is no standard data format. Metro and bus data are available at the station level (e.g., the total number of people tapping their travel smart cards at a specific station) rather than individual passenger level, mainly due to data privacy concerns. Aggregated datasets of train and bus travel information are available in some countries, such as the U.S. and China.

2.2 Mobile Network Operators

Mobile network operators can track people's locations in the Call Detailed Records (CDRs) that contain information about the time of the call and the cell tower to which mobile phones are connected when the call took place (Oliver et al., 2020). These human mobility data contain an abundant amount of information about mobile phone call locations; therefore, these data can help address many challenges, such as traffic congestion, public security, urban planning, and the control of pandemics (Zhao et al., 2016). However, due to data privacy issues, it is difficult to collect such data on a large scale, and these data are less likely to be publicly available. Thus, such data are usually aggregated to coarse geographic scales (e.g., county and state) for public release or further calculated as index-based data to indicate human mobility at the population level (Warren et al., 2020). In the meantime, many privacy-enhancing technologies have been proposed to ensure individual mobility data are anonymous and unidentifiable.

2.3 Global Positioning System (GPS)

The prevalence of telecommunication devices, particularly smartphones, enables the collection of individual mobility data through the GPS position of each mobile phone user. GPS-based telecommunication data can reveal an individual's daily mobility behavior, such as where and how long they remain in a particular location (Zhao et al., 2016). There are numerous installed mobile phone apps that record GPS locations once users enable the function of positioning. For example, navigation apps (e.g., Google Maps and Apple Maps) record real-time GPS locations when starting navigations. This information can help measure the access frequency to the points of interest (POI), such as residential areas, shopping centers, hospitals, and recreational and public facilities.

Another source to collect location data is the social media mobile apps (e.g., Facebook and Twitter). These data can only be collected if users enable the positioning function or if users input geographic information when they post content. Such social media data, usually termed as geotagged data, have become an emerging source of users' spatiotemporal information that can be further used to indicate users' mobility behavior. To minimize the spread of infectious disease as mobility restrictions are lifted, some contact-tracing applications have been developed and applied, such as C.A. Notify (https://canotify.ca.gov/) and Corona 100m (Dudden, 2020).

3. Mobility Estimations

Datasets collected from the three identified major sources have been processed and evaluated to indicate the pattern and magnitude of human mobility at the individual and aggregated levels. With a comprehensive consideration of the processing methods used to generate mobility data and the purpose of data usage, we classified the mobility measures utilized in the current COVID-19 studies into four different categories: 1) public transit data; 2) social activity data; 3) index-based mobility data; and 4) social media-derived mobility data. Table 1 lists and compares each group of mobility datasets in terms of provider, region, and scale, available time, origin-destination (OD) flow, availability, strengths, weaknesses, and related references. The OD flow indicates if the mobility value/index provides inter/intra-region movement information. The availability column shows how users can access each data set: private, public, or application submission. Some selected references which applied the mobility data sets for COVID-19 studies are listed as well. This section describes some typical mobility datasets within each category and the methods used to generate such mobility datasets.

Table 1. Summary of human mobility datasets in recent COVID-19 studies

Data Category		**Name and Provider**	**Region and scale**	**Available Time**	**OD Flow**	**Availability**	**First release after COVID-19**	**Strengths**	**Weaknesses**	**Selected References**
Public Transit System	Air flight	OpenSky-Network	Worldwide, mostly for Europe and North America	01/01/2019 ~ present	Yes	Public	Yes	Detailed tracking information	API limitations	Zhuang et al., 2020 Iacus et al., 2020
	Train	Transit system / dataset in different countries (e.g., China, U.S., Italy)	China, USA, Italy (by state, city, district)	Different for different regions	Yes	Public	No	Available in countries where there is a booking website	Real-time data but without history data o sometimes web crawling needed	Zhang et al.,2020; Carteni et al., 2020 Hu et al., 2020
	Metro	Transport authority (e.g., MATSim-NYC)	U.S.	Different for different regions	No	Public	No	Detailed ridership data, at the station level	Non-trackable; no route record	Zheng et al.,2020; Ahangari et al., 2020
Social Activity		Apple Mobility Trends Report	Worldwide/city, county, state	04/14/2020 ~ present	No	Public	Yes	global wide; one single file; data divided by country/region, sub-region, city	data source method (requests for direction in Apple Maps)	Huang et al., 2020b Kurita et al., 2021; Hadjidemetriou et al., 2020
		Google Mobility Reports	Worldwide/city, county, state	2/15/2020 ~ present	No	Public	Yes	global wide; one single file	not comparable among countries	Pepe et al.,2020; Delen et al., 2020; Rutz et al., 2020
		Foursquare Mobility Reports	U.S./county, state	02/19/202 ~present.	No	Submit Application	Yes	Available in 25 types of POI and by age group	Only available in U.S	Gao et al., 2020 Fathi-Kazerooni et al.,2020 Ding et al.,2020
		SafeGraph Mobility Reports	U.S./census tract, county, and state	01/01/2019 ~present	Yes	Submit Application	No	Varieties of data categories	Data are only availabl on Amazon S3	Li et al., 2020 Kang et al. 2020
Index-based Mobility Data		Cuebiq Mobility Index	U.S. at multiple geographic levels	01/01/2020	No	Submit Application	Yes	Available in DMA level; index allows counties to be compared to one another	Only available in U.S.	Fraiberger et al., 2020 Pepe et al., 2020
		Baidu Mobility Index	China/city and province	1/1/2020 ~ 5/7/2020 & 9/3/2020 ~ presen	Yes	Public	No	inter/intro-city mobility index	Not publicly accessible after May 7 2020, only available for Mainland China	Ze-Liang et al., 2020 Liu et al., 2020 Xu et al., 2020
		Descartes Lab Mobility Index	U.S./county and state	03/01/2020 – present	No	Submit Application	Yes	accurate positioning data (m50 score based on normalization methods)	Inter-city index not covered; only freely available in a certain period of time and scale	Warren et al., 2020 Gao et al., 2020; Chen et al., 2020

	Unacast Social Distancing Index	U.S.	02/24/2020 ~ present	No	Submit Application	Yes	Granular data, available down to specific data points; bias correction based on classifications of businesses	Since data is coming from third party sources, people may have to agree to consent on those sources	Brodeur et al., 2021
	University of Maryland Mobility Metrics and Social Distancing Index	U.S./county and state	01/01/2020 ~ present	No	Submit Application	Yes	Integrated and cleaned location data from multiple sources; be highly representative	Only available in the U.S.	Zhang et al., 2020 Lee et al., 2020 Ghader, et al., 2020
	Camber Systems Social Distancing Reporter	U.S./county	08/01/2020 ~ present	No	Submit Application	Yes	Integrating multiple data sources; less biased and more representative; easy to interpret	subject to calibration; only available in U.S. county level; no data before August 2020	Jeffrey et al., 2020
Social Media-Derived Mobility Data	Geotagged Tweets	Worldwide/any spatiotemporal scale	01/01/2018 ~ present	Yes	Public	No	Worldwide coverage, real-time, aggregation-flexible	Bias in population, low penetration	Huang et al., 2020a Li et al., 2021 Su et al., 2020
	Facebook Movement Range Maps	Worldwide	01/03/2020 ~ present	No	Submit Application	Yes	Machine-readable format that is global and free of charge	Only provided by mobile phone users who have enabled location history	Lau et al.,2020; Kuchler et al., 2020 Beria et al., 2021

3.1 Public Transit-based Mobility Data

Public transit data comprise metro, train, and air flight datasets. The air flight data are useful in estimating the population flow both between and within countries and regions. There are many ways to aggregate the data at multiple spatiotemporal scales based on the flights' origins and destinations. For example, OpenSky Network collected crowdsourced air traffic control data broadcast by airplanes and shared a dataset of global flight movements, which include aircraft identification number, model type, origin, destination, first-seen and last-seen timestamps, latitudes, and longitudes, etc. Researchers have aggregated the data to estimate the changes in air flight activities due to demand decline and travel restrictions (Iacus et al., 2020).

The data for other modes of public transit have a greater variation in terms of coverage and quality depending on the region. For the train data, the ctrip (https://www.ctrip.com/) and 12306 (www.12306.cn) are two major data sources (available only in China) that provide the scheduled timetable data. The passenger-level train data, however, are generally not available due to privacy and regulations, while aggregated datasets are available in some countries. As an example, the National Transit Database (https://www.transit.dot.gov/ntd) is a repository that stores financial and operating information of transit systems in the U.S. Some agencies, such as the Metropolitan Transportation Authority in the U.S., release daily turnstile data and aggregated ridership data (http://web.mta.info/developers/turnstile.html), both of which can be used to estimate travel volume.

3.2 Social Activity Data

Social activity data is another stream of mobility data that has been commonly utilized to reflect human social activity behavior with access to different places of interest, such as workplaces, residential areas, public transit, health care facilities, schools, shopping centers, and recreational and sports facilities. Most of such social activity data are GPS-derived metrics of foot traffic or POI access frequency sourced from large information technology (IT) companies with mapping services (e.g., Google and Apple). These data have been regularly updated and released to the public for the purpose of supporting public health authorities and the research community in response to COVID-19.

One of the mobility datasets most commonly used is Google Mobility Reports that provides the percentage change of place visit frequencies in six types of locations (workplaces, residential, parks, grocery and pharmacy, retail and recreation, and transit stations) compared to a pre-pandemic baseline value of mobility from January 3 to February 6, 2020 (Google LLC, 2020). Similarly, Foursquare Community Mobility Reports provides location data derived from U.S. foot traffic data, indicating the POI access frequency of people in different age groups through 32 categories of destinations (e.g., beach, bus station, hotel, etc.) (Foursquare Labs Inc., 2020). In contrast to Google and Foursquare data categorized by POI access, Apple Mobility Trends Report provides transport-based metrics that are generated by counting users' requests for directions in the Apple Maps app based on three types of transport modes (public transit, driving, and walking) compared to a pre-pandemic baseline value of mobility on January 13, 2020 (Apple Inc., 2020). Moreover, SafeGraph Foot Traffic data is an alternative mobility data with social distancing metrics determined by the night-time location of each mobile device. The POI access in SafeGraph data is categorized by industry and business (e.g., grocery stores, restaurants, bars, hotels); brand (e.g., Costco, Starbucks, Walmart); regions (e.g., metropolitan areas); and restaurant type (e.g., Chinese, Italian, Korean food) (SafeGraph, 2020).

3.3 Index-based Mobility Data

Unlike the social activity data, which is directly generated by POI access frequency or foot traffic, index-based mobility data are calculated as index metrics by data collected from the multiple sources described in Section 2. There are six types of index-based mobility data commonly used in the current literature, and these are Cuebiq Mobility Index; Baidu Mobility Index; Descartes Lab Mobility Index; University of Maryland Mobility Metrics and Social Distancing Index (UMMI); Camber Systems Social Distancing Reporter (CSSDR) (Camber,2020); and, Unacast Social Distancing Index. These are usually processed through different methods and aggregated at a particular spatial scale (e.g., state/province, county/city, and census tract) for public release.

Index-based mobility data usually employs advanced algorithms in the processing of data optimization to ensure data quality and reliability. Baidu Mobility Index (BMI) is the most widely used mobility data in COVID-19 related studies in China, and these data are provided by Baidu, which is a leading IT company in China. BMI is measured through gathering travel data based on more than 120 billion daily location requests from the Baidu Map app and other mobile phone apps that use Baidu's location services. BMI also provides a series of indices at different spatial scales, including within-city mobility intensity, population inflow and outflow ratio, intensity inter/intra-provinces, and top 100 cities (Baidu Inc., 2020). Cuebiq Mobility Index, which is also derived from mobile location data, quantifies how far mobile devices move each day by using a derivative factor indicating the distance between opposite corners of a box drawn around the observed locations (Cuebiq Inc., 2020). Descartes Lab Mobility Index (Descartes Lab, 2020) is calculated as the median of maximum distance mobility across all position reports within the threshold of 50 meters. It is optimized through a normalization method to improve the accuracy of positioning data (Warren, 2020). Similarly, Camber Systems Social Distancing Reporter (CSSDR) uses a locally estimated scatterplot smoothing regression to calculate two indices (e.g., the radius of gyration and median Shannon entropy index) that indicate how people are engaging in social distancing over time (Camber Systems, 2020). Unacast Social Distancing Index is calculated as the percent change of mobility in average distances traveled, the percent change in non-essential visitation, and the decrease in human encounters compared to the national baseline (Unacast, 2020). Lastly, the University of Maryland Mobility Metrics and Social Distancing Index (UMMI) contains six mobility metrics: the percentage of staying home; the percentage of reduction of all trips; work trips; non-work trips; travel distance; and out-of-county trips compared to the pre-pandemic benchmark through a multi-level weighting procedure (University of Maryland, 2020).

Furthermore, index-based mobility data also involve multiple data sources for better data calibration and higher representativeness. For instance, CSSDR data stored in Camber Systems' database merge various data sources to reduce errors and represent the larger population. Unacast Social Distancing Index is created and collated from third-party sources, such as Unacast's partners and the software development kit. UMMI, developed by the project team at the Maryland Transportation Institute at the University of Maryland, is based on integrated and cleaned mobile phone location data from multiple sources that represent person and vehicle movements (University of Maryland, 2020). In summary, compared to social activity data, index-based mobility data serve as an alternative to indicate human mobility as standardized and comparable indices available at multiple spatial scales and in different geographic contexts.

Although many human mobility indexes have been generated and applied in the U.S. studies, a few of these indexes estimate inter-regional population flow at different spatial scales. SafeGraph Foot Traffic data initially record POI access information in the census tract level as mentioned in section 3.2. This high-resolution human activity data provide the potential to derive a fine-scale inter-regions mobility index. Kang et al. (2020) estimated the daily and weekly dynamic OD population flows at three geographic scales: census tract, county, and state. The dataset is published in GitHub and is free to use. Such a high spatiotemporal resolution in a human mobility flow dataset may help monitor epidemic spreading dynamics, inform public health policy, and deepen our understanding of the change of human movement behavior within the current public health crisis. To promote the value of data, Li et al. (2020) developed a geospatial web portal, named Origin-Destination-Time (ODT) Flow Explorer, by extracting and aggregating the U.S. population flows from SafeGraph mobility data. The online tool helps extract and query, download mobility data, and visualize the population flows interactively.

3.4 Social Media-Derived Mobility Data

Commonly used social media platforms for gathering human mobility patterns include, but are not limited to: Twitter, Facebook, Instagram, WeChat (China), and Weibo (China). Compared with already summarized activity metrics at specific geographical scales (Section 3.2) and aggregated, constructed mobility index (Section 3.3), the raw social media records retrieved online are characterized by their aggregation-flexibility and scale-free nature. As a result, researchers are offered more freedom to construct desirable mobility indices at various spatiotemporal scales. This valuable (often extensive) user-generated information obtained from social media platforms and their derived mobility indices greatly facilitate (near) real-time human mobility monitoring during the COVID-19 pandemic in an active, less privacy-concerning manner (compared to the mobility records collected from phone records and smart cards).

Twitter has become the largest source of social media research data due to its policy of free access to about 1% of its total content (Martín et al., 2020). Many attempts have been made to utilize the derived geospatial contexts from social media posts, with the goal of assisting the monitoring of human mobility (Huang et al., 2020a; Li et al., 2020a; Huang et al., 2020b; Bisanzio et al., 2020; Peng et al., 2020; Zeng et al., 2021; Li et al., 2021). Huang et al. (2020a)conceptualized a mobility-based responsive index by integrating single-day distance and cross-day distance derived from massive Twitter posts worldwide, aiming to understand the dynamics of human mobility under implemented policies during the COVID-19 pandemic. To measure places connectivity, Li et al., (2020) proposed a global multi-scale place connectivity index (PCI) based on spatial interactions among places revealed by geotagged tweets. A PCI is defined as the normalized number of shared persons between two places during a specific time period. In response to the soaring needs of human mobility data, especially for the COVID-19 pandemic, Li's team developed origin-destination-time (ODT) flow platform which extracts, analyzes, and shares multi-scale human mobility index estimated by the PCI and population flows described above.

Facebook Movement Range Maps provide similar datasets with two different metrics, Change in Movement and Stay Put, both of which are generated from the mobility data of 27 million Facebook mobile app users with location history turned on. The Change in Movement metric looks at how many people are moving around and compares this number to a baseline period that predates most social distancing measures. The Stay Put metric looks at a fraction of the population that appears to stay within a small area surrounding their home for an entire day. Mobility records from social media only reflect the travel behaviors that users are willing to share and are dependent on the demographics of the local users in relation to the demographics of the local population. More data limitations regarding social media-derived mobility are discussed in Section 6.2.

4 Mobility Data Characteristics

We summarized the characteristics of the mobility datasets in terms of data privacy, quality, spatiotemporal coverage, storage, processing, availability, and accessibility. This summary of evidence aims to help researchers familiarize themselves with these mobility datasets and select the appropriate data for their studies.

4.1 Data Privacy

The privacy of human mobility measures varies across different types of datasets. Within public transit datasets, the scheduled timetable data often face fewer concerns about privacy issues. In contrast, the data with the actual trajectories of travelers pose greater risks to individual privacy. As a result, transport authorities and governments have put increasing efforts into privacy protection; the largely anonymous individual-level transit data is only issued to organizations or institutes for specific purposes (e.g., education or policy-making) through careful approval and confidential agreement. Social activity and index-based mobility data created by telecommunication signals, POI access, or mobile apps are subject to more data privacy challenges. They are more likely to be released as aggregated-level metrics based on a large volume of anonymous location data. Such aggregated data indicate the overall patterns or changes of human mobility in a particular spatial unit with less concern about identifiable user-based information. Governments, professional associations and organizations, data providers, and researchers have made joint efforts to improve the stringency and implementation of regulations in addition to ensuring the ethics clearance in the process of data sharing and manipulation. For example, data used for individuals' trajectories and identifications in COVID-19 related studies, such as contact-tracing data, are required to be re-coded. Additionally, a small volume of individual mobility data in less populated areas requires the addition of random spatiotemporal points to make such data unidentifiable (Zang and Bolot, 2011; De Montjoye et al., 2013; Xu et al. 2017).

4.2 Data Quality

The quality of mobility data depends on data types, and this makes it challenging to assess data quality without careful comparison studies. For example, the OpenSky-Network provides freely accessible airline flight data since 2019; however, it does not provide every global flight movement but only those ADS-B-equipped aircraft seen within the coverage (The OpenSky Network, 2020). Additionally, the monitored ADS-B-equipped aircraft data are mostly concentrated only in North America and Europe (Iacus et al., 2020). The smart-card-based public transit data usually contain a certain amount of missing data due to human errors made by smart card users (e.g., forgetting to tap in/out) or due to the malfunction of collection devices (e.g., tapping machines not working) (Liu et al., 2019). Furthermore, the producers of social activity data and index-based mobility data (e.g., telecommunication and IT companies, research institutes, and universities) consider that their data quality is reasonably controlled through data validation and calibration. However, most of these mobility data suffer from representativeness issues as the data are limited to only mobile phone users or app users whose locational function is turned on. For example, geotagged tweets only account for a small portion of the entire tweets (Jurdak et al., 2015; Martín et al., 2020). The accuracy of the geographic locations in geotagged tweets also varies as users may only geotag their posts at the city or state level rather than specific GPS coordinates. Therefore, more efforts are needed to refine, clean, anonymize, combine, and compare multiple mobility data types to ensure data quality and reliability.

4.3 Space-time Coverage

Human mobility data used in COVID-19 research vary greatly in spatial and temporal data coverage. For public transit data, air flight datasets are available at the global scale, while train data (e.g., in China, the U.S., and Italy) are available at the state, city, or district level. Metro and bus data are usually available at a finer scale (e.g., across fare zones or stations). For social activity data, Apple and Google mobility data are available globally at the state or city level in some countries. Apple data cover 63 countries with a timeline starting from January 13, 2020 (except for May 11-12, 2020), and Google data cover 131 countries with varying timelines based on the country. Foursquare Mobility Reports cover only the U.S. and are available at the county and state level with a timeline starting from January 1, 2020. Unlike the aforementioned datasets, Safegraph initially provides social activity data at the census tract level, and the data can be easily aggregated to the county and state level.

For index-based mobility data, Baidu Mobility Index only cover Mainland China from January 1, 2020 to May 7, 2020, and from September 22, 2020 to the present. Cuebiq Mobility Dashboard data only cover the U.S. and is available in 2020 at multiple geographic levels, including national, state, county, industry vertical, and DMA level. Descartes Lab Aggregated Mobility Index is available at the state level in the U.S. from February 16, 2020 to the present. Unacast Social Distancing Index is only available in the U.S. by state and county since February 24, 2020. The Mobility Metrics and Social Distancing Index and Camber Systems Social Distancing Reporter are both available in the U.S. at the county and state level covering the entire year of 2020.

In the COVID-19 pandemic, understanding human mobility spatial interaction patterns at different geographic scale has been critical for assessing the impact of non-pharmaceutical interventions. Public transit system provides an easy way to extract origins and destinations from timetables of airflight, train, metro, and bus. However, it only estimates OD flows of the transits, presenting a very small group of population. In contrast, Baidu Mobility Index offers inter-city and inter-provinces population flows index with comprehensive big data sources. However, it is only available in China. In the US, researchers estimated and shared the estimated daily dynamic OD population flows from SafeGraph at three geographic scales: census tract, county, and state (Kang, et al., 2020; Huang, et al., 2020a) and from Geotagged Tweets support the at county, state, and country level (Li, et al., 2020a).

4.4 High-Performance Data Storage and Processing

Human mobility data, generated passively or actively from the movement of billions of people around the world, are intrinsic Big Data that are characterized with the typical five challenging *V*s (volume, velocity, variety, veracity, and value) (Zikopoulos and Eaton 2011, Gudivada et al., 2015, Li, 2020). The value of human mobility data relies on the capability of extracting population movement patterns at various spatiotemporal scales and resolutions (local to global, real-time VS historical) from multi-source and multi-scale datasets in a space and time framework. The report-based mobility datasets from Google, Apple, Foursquare, and Facebook are highly processed with population movement information aggregated to a specific spatiotemporal scale (mostly showing the daily movement at the county or equivalent levels). While such datasets are often small in size and easy to handle, they offer limited flexibility for studies on different spatiotemporal scales and limited transparency in how these mobility data are generated. Other mobility data sources, such as geotagged social media data and cell phone data, may overcome such limitations by capturing individual human movements in much finer spatiotemporal resolutions. As these data are available in relatively raw formats with massive volumes, high-performance data storage and processing are needed to efficiently extract useful mobility information. For example, utilizing a Hadoop-based high-performance computing cluster coupled with a cube-based data storage model, Li et al. (2020a) extracted 591 million individual-level daily OD flows from 2.1 billion geotagged tweets worldwide and derived 9.7 billion daily OD flows at the U.S. block group level from SafeGraph data. With the increasing availability of and soaring demands for fine-scale human mobility data, more studies are needed to develop novel parallel and scalable computing algorithms and data models to manage, process, and analyze big mobility data in a high-performance computing environment. Furthermore, more studies are needed to develop interactive geospatial web portals (http://gis.cas.sc.edu/GeoAnalytics/pci.html) to allow researchers to query, extract, visualize, and share the derived human mobility data.

4.5 Data Accessibility

Conventional data sources, such as public transit system, have started to provide rich population flows data before the COVID-19 pandemic (see Table 1). To help researchers and governments worldwide with the response to COVID-19, technology companies and research institutions have made human mobility datasets publicly available after pandemic. There are several primary ways to publish or share these datasets: 1) online data dashboards via the official websites of data providers; 2) GitHub (https://github.com/); 3) Harvard Dataverse (https://dataverse.harvard.edu/); 4) user applications. Regarding online dashboards, technology companies, such as Google, Apple, and Facebook, build websites to present mobility data and provide direct data download links. Similarly, Cuebiq Mobility Index, Unacast Social Distancing, UMMI, and CSSDR can be freely accessed through browsing the dashboards in their official websites. GitHub is the largest and most advanced development platform that provides automatic data staging and publication regions. Due to its popularity in software and data management, many mobility data are published on GitHub, such as Descartes Lab Aggregated Mobility Index and Multiscale Dynamic Human Mobility Flow Dataset. The Harvard Dataverse is an online data repository in which research data is shared, preserved, retrieved, explored, and analyzed. The published data have Digital Object Identifier (DOI) and are open to all communities. The BMI data collected by the China Data Lab are shared on Harvard Dataverse (Hu, 2020a). The BMI is originally to explore inter-cities and inter-provinces population flows during the Spring Festival of China, so it has been published years ago before the COVID-19 pandemic. The fourth way of publishing and sharing mobility datasets is through user applications. Some data providers require users to submit an application indicating the purpose of data usage if users want to access the whole dataset. For example, SafeGraph provides free access to various COVID-19 relevant datasets after completing a specific form and signing a non-commercial agreement. In the same way, CSSDR can be accessed through a submitted application, but only users located within the U.S. can apply for data access.

5. Mobility Data Applications

In this section, we summarized the three major applications of human mobility data in COVID-19 studies: 1) revealing spatiotemporal trends and patterns; 2) examining the association of human mobility with COVID-19; and 3) integrating the simulation and predictions of COVID-19 transmission.

5.1 Spatiotemporal Trends

Exploring and examining the changing trends and patterns of human mobility in both spatial and temporal dimensions are the fundamental analyses in providing an overall picture of mobility changes across space and time. Mobility data are usually aggregated in a certain geographic unit and can be used to indicate the spatial differences of mobility, such as across states, counties, provinces, and cities. For that reason, a large amount of current COVID-19 studies used social activity data and index-based mobility data to conduct descriptive or exploratory analyses that reveal mobility patterns. BMI has been mainly used to reveal human mobility patterns in China during the period of intensive lockdown from the middle of January to April (Bao et al., 2020, Zhao et al., 2020, Liu et al., 2020a). Mobility pattern changes in the U.S. have been analyzed using different types of index-based mobility data, including Cuebiq Mobility Index (Fraiberger et al., 2020; Pepe et al., 2020, Bourassa et al., 2020; Aleta et al., 2020), UMMI (Zhang et al., 2020; Lee et al., 2020; Ghader et al., 2020; Xiong et al., 2020; Pan et al., 2020), Descartes Lab Mobility Index (Warren et al., 2020), and indices from SafeGraph (Huang et al., 2020c; Lamb et al., 2020; Levin et al., 2020). Social media-derived data have been used as well, such as geotagged tweets (Huang et al., 2020a; Li et al., 2020). Beyond the U.S., social activity data available at a global scale have been used to reveal mobility patterns in European countries (Vokó et al., 2020), Australia (Wang et al., 2020), Japan (Fraser et al., 2020), or in comparative studies across countries (Bryant and Elofsson, 2020).

The above mobility-based studies reveal some common findings contributing to the COVID-19 control and intervention as below. First, policy interventions including lockdown, travel restrictions, social distancing, and border control have effectively reduced the transmission of COVID-19 in different geographic contexts (e.g., China, US, and European countries by Djurović, 2020, Jiang and Luo, 2020, Kraemer et al. 2020, Dickson et al. 2020, Tobías, 2020). Second, the relationship between human mobility and the virus spread is temporal and spatial heterogeneity. Policy interventions, despite being globally effective in reducing both the spread of infection and its self-sustaining dynamics, have had heterogeneous impacts locally (O'Sullivan et al. 2020). Policy measures need to be adjusted across the different phases of the pandemic. The reduction of infection caused by mobility control is observed to be relatively weaker in places where the outbreak occurred later (Zhang et al. 2020). Third, mobility control is observed to have a time-lag effect on the virus transmission and such effect varies across the geographic contexts and the timeline of the pandemic. Studies across various countries reported that the efficacy of lockdown continues to hold over two weeks or even up to 20 days after a lockdown was implemented (Alfano and Ercolano 2020). The timing, effectiveness, and stringency of policy implementation are crucial for the success of COVID-19 control efforts in different countries. The early implementation of social and mobility restrictions is especially effective in lowering the peak value of new infections and reducing the infection scale (Kaur et al. 2020).

5.2 Association Analysis

In addition to exploring the spatial and temporal pattern of mobility changes, researchers have also enriched their analyses by examining the relationship between human mobility and other variables in specific dimensions. Some of these variables are the number of confirmed cases or deaths, reproductive rate, transmission rate, and doubling time within the pandemic context. Such a relationship has been mainly examined through correlation and regression analyses with the consideration of the different time-lag effects of human mobility on COVID-19 transmission (Wang et al., 2020). Mobility data used in such analyses are subject to different geographic contexts and spatial scales. For public transit data, air flight datasets have been utilized to estimate inter-country mobility, which has been one of the key drivers in disease transmission, especially at the early stages of the COVID-19 pandemic (Menkir et al., 2020; Bogoch et al., 2020). Metro data have been used to estimate mobility and examine the relationship between human mobility and ridership. In the case study of New York City (NYC), researchers used the transit network & schedule from General Transit Feed Specification (GFTS) data, NYC Bridge Traffic Volumes data, and MTA Subway Turnstile Data (Gao et al., 2020; Fathi-Kazerooni et al. 2020). Train data retrieved from China State Railway Group, National Transit Database, and Trenitalia have been used to investigate the incidence of transport accessibility during the spread of COVID-19 in China (Zheng et al. 2020; Hu et al. 2020; Zhang et al., 2020), U.S. (Ahangari et al., 2020; Anik Islam 2020), and Italy (Cartenì et al., 2021), respectively. Furthermore, all types of social activity data have been widely used to examine the relationship between human mobility and COVID-19 transmission, possibly due to their large spatial coverage and public availability at different spatial levels. For index-based mobility data, BMI has been extensively used in Chinese studies to examine how the virus spread was affected by the mobility from or to Wuhan and Hubei Province, where the first wave of COVID-19 appeared (Chen et al., 2020; Liu et al., 2020a; Xu et al., 2020; Shen 2020; Zhuang et al., 2020, Fang, et al., 2020). In the U.S. studies, Cuebiq Mobility Index and UMMI have been utilized to explore the relationship between mobility and virus spread; however, other types of index-based mobility data were relatively less used (Zhang et al., 2020; Lee et al., 2020; Ghader et al., 2020; Bourassa et al., 2020; Aleta et al., 2020). Finally, Twitter data have been more commonly used in emotional analyses based on the text and keywords from the content of tweets but relatively less used to examine the relationship between human mobility and COVID-19 transmission (Porcher et al., 2020; Xu et al., 2020; Li et al., 2020; Wang et al., 2020; Hu et al., 2019). Although, a few efforts have been made to link Twitter-derived human mobility with

COVID-19 cases. For instance, Zeng et al. (2021) examined the spatiotemporal relationship by associating Twitter-derived mobility and COVID-19 cases in the State of South Carolina. Li et al. (2021) explored the correlation between COVID-19 cases and the place connectivity index summarized from Twitter in Westchester County, New York, which was one of the early COVID-19 hotspots in the U.S.

5.3 Prediction and Simulation

Another dominant application of mobility data is modeling COVID-19 transmission by simulating and predicting the virus spread over a certain period of time. Mobility data have been integrated with epidemiological models or spatiotemporal models (e.g., a susceptible-infected-recovered (SIR) model or SIR-derived models), and these data have been controlled as parameters to simulate the COVID-19 cases based on different policy scenarios and to predict future trends of virus spread based on policy implications. For example, Zhou et al. (2020) built up a spatiotemporal epidemiological prediction model integrating the cellular automata model in the spatial dimension with the SIR model in the temporal dimension to inform county-level COVID-19 risk in the US. They utilized mobility measures and air distances between counties as the model parameters to indicate the spatial heterogeneity of COVID-19 transmission. Such an integration of accurate mobility measures with epidemiological modelling has been approved to improve and optimize model performance (see the additional citation added below). Public transit data (e.g., MTA ridership report; NYC subway turnstile) have been used to forecast the influence of COVID-19 on NYC's economy (Fathi-Kazerooni et al., 2020; Gharehgozli et al., 2020). Some studies utilized historical air flight passenger data and flight tracking data retrieved from VariFlight, SABRE, and OpenSky Network to predict COVID-19 cases and to estimate airplane passengers' number loss via different simulation scenarios (Zhuang et al., 2020; Iacus et al., 2020). Other studies utilized airplane passenger itinerary data from IATA and Cirium to simulate the potential risk and geographic range of COVID-19 spread within and beyond China and to predict the number of international COVID-19 cases arriving in China (Lai et al., 2020; Menkir et al., 2020). For index-based mobility data, BMI has been largely applied in disease modeling in Chinese studies (e.g., Zhao et al., 2020; Liu et al., 2020). For studies in the U.S., Cuebiq Mobility Index has been used to predict the COVID-19 trends in Boston (Aleta et al., 2020), and the UMMI has been applied to model the impact of stay-at-home orders on human mobility (Xiong et al., 2020). Other types of index-based data and social media-derived mobility data have been relatively less prevalent in modeling work, possibly due to the difficulties in data retrieval and processing (Li et al., 2020b).

6. Challenges and Future Directions

6.1 How to balance data privacy and data sharing?

Mobility data consists of location stamps about individuals. While it can help reveal the underlying patterns of human movement behaviors, it also poses a challenge to privacy protection as human movements are highly unique and predictable (Song et al., 2010; De Montjoye et al., 2013). The risk is even higher when different mobility datasets are merged, even with every dataset being anonymized (Kondor et al., 2018). Therefore, it is crucial to establish standards in the deposit, storage, processing, and distribution of mobility data. Researchers need to ensure that any identifiers from the datasets are removed before depositing the data. The storage of mobility data must be secure and must disallow any unauthenticated and unauthorized access. Data security in transit is also critical as this ensures that data is protected while being transferred in-between networks, such as during the upload, download, and data transmission steps of processing and backup. Due to the complexity of technologies involved that ensure data security and long-term preservation, it is often not practical for researchers to host the data on their own. After the data are processed, researchers may choose a trusted data repository in which their datasets could be deposited (Corrado, 2019).

To balance the needs of data sharing and privacy protection, researchers may process the data in a way that is suitable for multiple levels of access. Sensitive datasets should be held in a secure environment that allows other researchers to access the data through a remote desktop in a private network and with pre-installed analysis software. The output from the analysis needs to be vetted before releasing to researchers. After aggregation, the datasets may become less sensitive and more prone to distribution through the internet. Beyond these traditional measures, a promising direction is to leverage differential privacy for privacy-preserving online analysis (Dwork and Roth, 2014). Although it is not yet common, providing differential privacy as a service is expected from more vendors soon. This will enable researchers to collaborate more efficiently without sacrificing data privacy.

6.2 How multi-source mobility data are different in nature?

Human mobility is characterized by its multi-faceted nature, which has been evidenced by many studies (Gonzalez et al., 2008; Cui et al., 2018). For the COVID-19 pandemic, Huang et al. (2020b) cross-compared four mobility sources: Apple mobility trend reports, Google community mobility reports, mobility data from Descartes Lab, and Twitter-derived mobility data. Even though the similarity of reduced mobility trend has been observed after the declaration of COVID-19 as a pandemic, multi-source mobility datasets present unique and even contrasting characteristics. The findings of Huang's work coincide with the argument by many scholars that heterogeneous mobility data sources consist of different characteristics that reflect human spatial interactions and dynamics from different yet valuable perspectives (Huang et al., 2020b; Lau et al., 2019).

Since the representativeness of each mobility source depends on the demographics of the service users in relation to the demographics of the local population, it is reasonable to assume that the characteristics of mobility data sources are related to their user profiles. Thus, the similarity/disparity of multi-source mobility data can be explained by the spectrum of the population being captured. For example, mobility data retrieved from public transit systems are mode-specific and lack the holistic views of overall human spatial interactions. In addition, transit-based mobility records usually fail to assist human mobility monitoring when strong policy interventions are issued (e.g., the cancelation of public transit). Mobility data collected via a passive manner (mobile phones and wireless networks) tend to have high representativeness because of their high penetration ratios. Despite the broad spectrum of people that can be captured in these sources, the privacy and confidentiality concerns prevent the data from being released without a certain level of aggregation and anonymization (usually, proper authorization is also required). Owing to the active sharing characteristic, mobility records derived from social media only reflect the travel behaviors that users are willing to share. This can potentially add bias towards occasions that are regarded as "post-worthy" to users, such as visiting tourist attractions and attending memorable gatherings.

Despite the notable heterogeneity presented by the aforementioned sources, studies have shown that the fusion of multi-source mobility datasets can mitigate, to a large extent, the intrinsic bias within each mobility source and can provide a holistic view of mobility dynamics by capturing a broad spectrum of the population (Montero et al., 2019; Huang et al., 2018; Lau et al., 2019). A few data fusion efforts have been made as scholars have started to realize the fusion value of multi-source mobility datasets. For example, Zhang et al. (2014) integrated transit records and cellphone-derived positions to mitigate biased sampling via a systematic framework. Montero et al. (2019) proved that the integration of mobility sources leads to robust urban transportation models. Similarly, Huang et al. (2018) combined mobility records from urban transportation and mobile phone signals to achieve comprehensiveness via a high data penetration ratio. Despite these efforts, mobility data fusion is still rare in public health studies. The COVID-19 pandemic specifically demands rapid and comprehensive monitoring of multi-faceted mobility dynamics.

6.3 How to choose the appropriate mobility data in mobility applications?

Choosing appropriate mobility data is the fundamental and primary step in addressing human mobility-related questions during a pandemic such as COVID-19. The categorization of mobility data in our review can be used for human mobility studies at individual and aggregated levels as well as across various spatial and temporal scales. At the individual level, mobile phone positioning data or social media big data can be used to track individuals' trajectories, daily activities, exercise routines, and travel behaviors. Thus, this type of data is suitable for studies on contact tracing that aim to prevent the virus spread or to seek the origin and primary source of virus transmission in the context of COVID-19 and other disease outbreaks (Park et al. 2020; Salathé et al., 2020). Mobility studies at the aggregated level have more options in data sources, and an appropriate selection of aggregated data types is subject to the spatiotemporal scale. For example, global-scale or cross-country studies can utilize Facebook, geotagged tweets, Apple, and Google mobility data; however, these may not be suitable for studies in China where services provided by these companies are not available. While most other data sources cover various countries, the BMI only covers Mainland China. If studies aim to include the full timeline of the COVID-19 pandemic, Cuebiq Mobility Index, Descartes Lab Aggregated Mobility Index, and Unacast Social Distancing Index may be good options to include the entire year of 2020, whereas Google mobility data, Apple mobility data, and BMI would not be appropriate due to their limited coverage of time periods.

Public transit data are relatively useful when lockdown policies are not fully implemented. Although they only reflect the scheduled transit and not the actual usage of transit, they can be used as supplements to other types of mobility data. Such data can indicate connection and interaction across places under the restricted travel ban, and this information cannot be revealed by social activity mobility data. For example, the connection of air flights from China to other countries can serve as an essential indicator to show how the virus may have spread outside of China (Lau et al. 2020). Another example is using high-speed train data to simulate how the virus was transmitted from Wuhan to other regions in China during the early stages of the COVID-19 pandemic (Zhang et al., 2020).

Finally, social activity data are appropriate for studies on the different types of human mobility to various places and by various trip modes. Each of these datasets consists of different variables and indices to distinguish people's movement activities as a response to social restriction policies. For example, the Google mobility trends report provides data on mobility to parks, workplaces, and pharmacies, while the Apple mobility trends report reflects the mobility changes of walking, driving, biking, and public transit usage, all of which cannot be revealed by other types of mobility data.

6.4 How to integrate other data sources to enhance mobility applications?

With the rapid development of information techniques, a large amount of data can be created in an instant, and different data sources are widely applied from a variety of perspectives. How researchers can integrate different data sources to enhance human mobility applications in the COVID-19 pandemic and any future pandemics is a critical topic. Human mobility is strongly associated with regional socioeconomic indicators, such as income and poverty rate, and the relationship between mobility and socioeconomic status could vary among cities. Such association is influenced by the spatial arrangement of housing, employment opportunities, and human activities (Xu, 2018). Since the start of the COVID-19 pandemic, human movement behaviors have been completely disrupted, and these movement behaviors show very different patterns than ones before the pandemic. To keep essential services online, governments still require a considerable number of essential workers to physically travel to work, as opposed to most other types of employees who can work from home. Therefore, it is of vital importance to build a new model that could integrate human mobility and socioeconomic data across regions to re-evaluate the association between human movement behaviors and socioeconomic characteristics of the underlying population during the COVID-19 pandemic and any other disease outbreaks.

Apart from socioeconomic data, some big data sources have been widely applied in COVID-19 studies as well. For example, remote sensing data have helped detect air quality, traffic, and human activity changes with a large spatial coverage during different phases of the pandemic (Liu et al., 2020b; Fan et al., 2020; Chen et al., 2020). With remote sensing data, many classification systems are developed to detect different land use and land cover. Land use data shows how people use the landscape, and it can be grouped into five main classes: residential, agricultural, recreation, transportation, and commercial. The integration of high-resolution human mobility data derived from geotagged social media will help understand human movement patterns across different land use types. Compared with other mobility reports, this provides more types of places in more extensive spatial coverage within the U.S. as well as in some European and Asian countries. Such data integration may assist in the understanding of human activity trends within areas that lack comprehensive human movement tracking information.

Due to the varying levels of accuracy and different spatial and temporal resolution of mobility datasets and other spatial data, integrating multiple data sources is a complex issue (Torre-Bastida et al., 2018). One of the major challenges is to identify the items of incompliancy in a standard way. Researchers have proposed many methods and developed tools to integrate various spatial data. For example, Mohammadi et al. (2010) proposed a tool that facilitates data harmonization through the assessment of multi-source spatial datasets against many measures. Wiemann and Bernard (2014) applied Open Geospatial Consortium (OGC) and Semantic Web standards and used Linked data to build service-oriented spatial data fusion strategies. Multi-source data integration not only requires technical tools and considerations, such as the match of datasets geometrically, topologically, and with correspondence of attribute (Usery et al. 2005), but it also requires non-technical issues (e.g., institutional, policy, legal, and social mechanism) to facilitate the integration (Mohammadi et al., 2010; Torre-Bastida et al., 2018). Nevertheless, the effective integration of multi-source datasets has not been fully achieved. It is crucial to building a standardized framework to help researchers avoid the time-consuming and costly process of data integration and provide timely and high-quality data support for the COVID-19 studies.

7. Conclusion

The growing volume of human mobility data being collected and made available opens up new opportunities and challenges for analyzing, modeling, and predicting the spread of infectious diseases. By investigating recent COVID-19 studies from a data-oriented view, our review contributes a comprehensive summary of human mobility open data condensed into data sources, index measures, characteristics, applications, challenges, and future research directions. We propose four avenues for future studies along which the current discussions can be extended: data privacy and sharing, data similarity and replacement, the appropriation of data usage, and the integration of mobility data with other data sources to enhance their applications in the pandemic.

Despite that these open-sourced human mobility datasets (usually "big" in nature) have established a venue where rapid monitoring/supervision of human movement with high spatiotemporal scales can be achieved, essential contextual information on these monitored travels is often missing, as such information is unlikely to be released due to privacy concerns. Surveys are typical sources of human mobility data. Although they are not, in most cases, open-source data (thus, they are not in the scope of this study), they provide essential contextual information (e.g., purposes of travels, detailed location of destinations, demographic profiles, socioeconomic status, etc.) that contributes to a comprehensive story behind these movements. Numerous efforts have been made to take advantage of survey data to assist a better usage of mobility records collected from other means (Chen et al., 2021, Chen et al., 2016, Huff and McCallum). In the future, we will conduct a broader human mobility data review, combining both conventional small data and new big data. We encourage researchers and stakeholders to apply comprehensive human mobility data to further monitor, prevent, and control current and future public health emergencies.

Data Availability Statements

The public available mobility datasets, such as Google Community Mobility Report, Apple Mobility Data, Baidu Mobility Index, Foursquare Mobility Data, and Descartes Mobility Index, presented in the study are shared and updated by Spatial Data Lab at Harvard University on Harvard Dataverse (https://dataverse.harvard.edu/dataverse/data_ncov).

Acknowledgment

This study is supported by the NSF under Grant 1841403, 2027540, and 2028791.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

Aleta, A., Martin-Corral, D., y Piontti, A. P., Ajelli, M., Litvinova, M., Chinazzi, M., ... & Moreno, Y. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nature Human Behaviour, 4(9), 964-971.

Alfano, V. and S. Ercolano (2020). "The Efficacy of Lockdown Against COVID-19: A Cross-Country Panel Analysis." Applied Health Economics and Health Policy 18(4): 509-517.

Apple Inc. (2020, July 14). Mobility Trends Reports. https://covid19.apple.com/mobility

Baidu Inc. (2020, April 1). Baidu Migration Data. https://qianxi.baidu.com/

Bao, R., & Zhang, A. (2020). Does lockdown reduce air pollution? Evidence from 44 cities in northern China. Science of the Total Environment

Belik, V., Geisel, T., & Brockmann, D. (2011). Natural human mobility patterns and spatial spread of infectious diseases. Physical Review X, 1(1), 011001.

Beria, P., & Lunkar, V. (2021). Presence and mobility of the population during the first wave of Covid-19 outbreak and lockdown in Italy. Sustainable Cities and Society, 65, 102616.

Bisanzio, D., Kraemer, M. U., Bogoch, I. I., Brewer, T., Brownstein, J. S., & Reithinger, R. (2020). Use of Twitter social media activity as a proxy for human mobility to predict the spatiotemporal spread of COVID-19 at global scale. Geospatial health, 15(1).

Bogoch, I. I., Watts, A., Thomas-Bachli, A., Huber, C., Kraemer, M. U., & Khan, K. (2020). Potential for global spread of a novel coronavirus from China. Journal of travel medicine, 27(2), taaa011.

Brodeur, A., Cook, N., & Wright, T. (2021). On the effects of COVID-19 safer-at-home policies on social distancing, car crashes and pollution. Journal of environmental economics and management, 102427.

Bryant, P., & Elofsson, A. (2020). Estimating the impact of mobility patterns on COVID-19 infection rates in 11 European countries. PeerJ, 8, e9879.

Camber Systems, (2020). Social Distancing Reporter. https://covid19.cambersystems.com/

Cartenì, A., Di Francesco, L., & Martino, M. (2020). How mobility habits influenced the spread of the COVID-19 pandemic: Results from the Italian case study. *Science of the Total Environment*, *741*, 140489.

Cartenì, A., Di Francesco, L., & Martino, M. (2021). The role of transport accessibility within the spread of the Coronavirus pandemic in Italy. Safety science, 133, 104999.

Changruenngam, S., Bicout, D. J., & Modchang, C. (2020). How the individual human mobility spatio-temporally shapes the disease transmission dynamics. Scientific Reports, 10(1), 1-13.

Chen, C., Ma, J., Susilo, Y., Liu, Y., & Wang, M. (2016). The promises of big data and small data for travel behavior (aka human mobility) analysis. Transportation research part C: emerging technologies, 68, 285-299.

Chen, S., Li, Q., Gao, S., Kang, Y., & Shi, X. (2020). Mitigating COVID-19 outbreak via high testing capacity and strong transmission-intervention in the United States. MedRxiv.

Chen, X., Zhang, A., Wang, H., Gallaher, A., & Zhu, X. (2021). Compliance and containment in social distancing: mathematical modeling of COVID-19 across townships. International Journal of Geographical Information Science, 35(3), 446-465.

Chen, Z. L., Zhang, Q., Lu, Y., Guo, Z. M., Zhang, X., Zhang, W. J., ... & Lu, J. H. (2020). Distribution of the COVID-19 epidemic and correlation with population emigration from Wuhan, China. Chinese medical journal.

Corrado, E. M. (2019). Repositories, trust, and the coretrustseal. Technical Services Quarterly, 36(1), 61-72.

Cuebiq Inc. (2020, October 1) Cuebiq's COVID-19 Mobility Index (CMI). https://help.cuebiq.com/hc/en-us/articles/360041285051-Cuebiq-s-COVID-19-Mobility-Index-CMI-

Cui, Y., Xie, X., & Liu, Y. (2018). Social media and mobility landscape: Uncovering spatial patterns of urban human mobility with multi source data. Frontiers of Environmental Science & Engineering, 12(5), 1-14.

De Montjoye, Y. A., Hidalgo, C. A., Verleysen, M., & Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. Scientific reports, 3(1), 1-5.

Delen, D., Eryarsoy, E., & Davazdahemami, B. (2020). No place like home: cross-national data analysis of the efficacy of social distancing during the COVID-19 pandemic. JMIR public health and surveillance, 6(2), e19862.

Descartes Lab, 2020. Descartes Lab Mobility Index. https://www.descarteslabs.com/mobility/

Dickson, M. M., G. Espa, D. Giuliani, F. Santi and L. Savadori (2020). "Assessing the effect of containment measures on the spatio-temporal dynamic of COVID-19 in Italy." Nonlinear Dynamics 101(3): 1833-1846.

Djurović, I. (2020). "Epidemiological control measures and predicted number of infections for SARS-CoV-2 pandemic: case study Serbia march–april 2020." Heliyon 6(6): e04238.

Dudden, A., & Marks, A. (2020). South Korea took rapid, intrusive measures against Covid-19- and they worked. The Guardian, 20.

Dwork, C., & Roth, A. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, *9*(3-4), 211-407.

Fan, C., Li, Y., Guang, J., Li, Z., Elnashar, A., Allam, M., & de Leeuw, G. (2020). The impact of the control measures during the COVID-19 outbreak on air pollution in China. Remote Sensing, 12(10), 1613.

Fang, H., Wang, L., & Yang, Y. (2020). Human mobility restrictions and the spread of the novel coronavirus (2019-ncov) in china. *Journal of Public Economics*, *191*, 104272.

Fathi-Kazerooni, S., Rojas-Cessa, R., Dong, Z., & Umpaichitra, V. (2020). Time Series Analysis and Correlation of Subway Turnstile Usage and COVID-19 Prevalence in New York City. arXiv preprint arXiv:2008.08156.

Foursquare Labs, Inc. (2020, June 1). Location Visit Data. https://visitdata.org/index.html

Fraiberger, S. P., Astudillo, P., Candeago, L., Chunet, A., Jones, N. K., Khan, M. F., ... & Montfort, A. (2020). Uncovering socioeconomic gaps in mobility reduction during the COVID-19 pandemic using location data. arXiv preprint arXiv:2006.15195.

Fraser, T., & Aldrich, D. P. (2021). The dual effect of social ties on COVID-19 spread in Japan. Scientific reports, 11(1), 1-12.

Gao, J., Wang, J., Bian, Z., Bernardes, S. D., Chen, Y., Bhattacharyya, A., ... & Iyer, S. (2020). The Effects of the COVID-19 Pandemic on Transportation Systems in New York City and Seattle, USA. arXiv preprint arXiv:2010.01170.

Gao, S., Rao, J., Kang, Y., Liang, Y., & Kruse, J. (2020). Mapping county-level mobility pattern changes in the United States in response to COVID-19. SIGSpatial Special, 12(1), 16-26.

Gatto, M., Bertuzzo, E., Mari, L., Miccoli, S., Carraro, L., Casagrandi, R., & Rinaldo, A. (2020). Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proceedings of the National Academy of Sciences*, *117*(19), 10484-10491.

Ghader S, Zhao J, Lee M, Zhou W, Zhao G, Zhang L. Observed mobility behavior data reveal social distancing inertia. arXiv 2020; published online April 30, 2020. https://arxiv.org/abs/2004.14748 (preprint).

Gharehgozli, O., Nayebvali, P., Gharehgozli, A., & Zamanian, Z. (2020). Impact of COVID-19 on the Economic Output of the U.S. Outbreak's Epicenter. Economics of Disasters and Climate Change, 4(3), 561-573.

Goh, S., Lee, K., Park, J. S., & Choi, M. Y. (2012). Modification of the gravity model and application to the metropolitan Seoul subway system. Physical Review *E*, *86*(2), 026102.

Gonzalez, M. C., Hidalgo, C. A., & Barabasi, A. L. (2008). Understanding individual human mobility patterns. nature, 453(7196), 779-782.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. GeoJournal, 69(4), 211-221.

Google LLC. (2020, May 14). Google COVID-19 Community Mobility Reports. https://www.google.com/covid19/mobility/

Gudivada, V. N., Baeza-Yates, R., & Raghavan, V. V. (2015). Big data: Promises and problems. Computer, 3, 20–23.

Hadjidemetriou, G. M., Sasidharan, M., Kouyialis, G., & Parlikad, A. K. (2020). The impact of government measures and human mobility trend on COVID-19 related deaths in the U.K. Transportation research interdisciplinary perspectives, 6, 100167.

Haider, N., Yavlinsky, A., Simons, D., Osman, A. Y., Ntoumi, F., Zumla, A., & Kock, R. (2020). Passengers' destinations from China: low risk of novel coronavirus (2019-nCoV) transmission into Africa and South America. Epidemiology & Infection, 148.

Hu, M., Lin, H., Wang, J., Xu, C., Tatem, A. J., Meng, B., ... & Lai, S. (2020a). The risk of COVID-19 transmission in train passengers: an epidemiological and modelling study. Clinical Infectious Diseases.

Hu, T., Guan, W. W., Zhu, X., ..., & Bao, S. (2020b). Building an Open Resources Repository for COVID-19 Research, Data and Information Management, 4(3), 130-147. doi: https://doi.org/10.2478/dim-2020-0012

Hu, T., She, B., Duan, L., Yue, H., & Clunis, J. (2019). A systematic spatial and temporal sentiment analysis on Geo-Tweets. *IEEE Access*, *8*, 8658-8667.

Huang, X., Li, Z., Jiang, Y., Li, X., & Porter, D. (2020a). Twitter reveals human mobility dynamics during the COVID-19 pandemic. PloS one, 15(11), e0241957.

Huang, X., Li, Z., Jiang, Y., Ye, X., Deng, C., Zhang, J., & Li, X. (2020b). The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the U.S. during the COVID-19 pandemic. International Journal of Digital Earth. doi: 10.1080/17538947.2021.1886358.

Huang, X., Li, Z., Lu, J., Wang, S., Wei, H., & Chen, B. (2020c). Time-series clustering for home dwell time during COVID-19: what can we learn from it?. ISPRS International Journal of Geo-Information, 9(11), 675.

Huff, D., & McCALLUM, B. M. (2008). Calibrating the huff model using arcgis business analyst. ESRI White Paper, 1-33.

Iacus, S. M., Natale, F., Santamaria, C., Spyratos, S., & Vespe, M. (2020). Estimating and projecting air passenger traffic during the COVID-19 coronavirus outbreak and its socioeconomic impact. Safety Science, 104791.

International Air Transport Association. (2021), https://www.iata.org/en/about/

Jiang, J. and L. Luo (2020). "Influence of population mobility on the novel coronavirus disease (COVID-19) epidemic: based on panel data from Hubei, China." Global Health Research and Policy 5(1).

Jacsó, P. (2005). Google Scholar: the pros and the cons. Online information review.

Jeffrey, B., Walters, C. E., Ainslie, K. E., Eales, O., Ciavarella, C., Bhatia, S., ... & Riley, S. (2020). Anonymised and aggregated crowd level mobility data from mobile phones suggests that initial compliance with COVID-19 social distancing interventions was high and geographically consistent across the U.K. Wellcome Open Research, 5.

Jung, W. S., Wang, F., & Stanley, H. E. (2008). Gravity model in the Korean highway. *EPL (Europhysics Letters)*, *81*(4), 48005.

Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., & Newth, D. (2015). Understanding human mobility from Twitter. *PloS one*, *10*(7), e0131469.

Kang, Y., Gao, S., Liang, Y., Li, M., Rao, J., & Kruse, J. (2020). Multiscale dynamic human mobility flow dataset in the U.S. during the COVID-19 epidemic. Scientific data, 7(1), 1-13.

Kaur, S., H. Bherwani, S. Gulia, R. Vijay and R. Kumar (2020). "Understanding COVID-19 transmission, health impacts and mitigation: timely social distancing is the key." Environment, Development and Sustainability.

Kishore, N., Kiang, M. V., Engø-Monsen, K., Vembar, N., Balsari, S., & Buckee, C. (2020a). Mobile phone data analysis guidelines: applications to monitoring physical distancing and modeling COVID-19.

Kishore, N., Kiang, M. V., Engø-Monsen, K., Vembar, N., Schroeder, A., Balsari, S., & Buckee, C. O. (2020b). Measuring mobility to monitor travel and physical distancing interventions: a common framework for mobile phone data analysis. The Lancet Digital Health.

Kondor, D., Hashemian, B., de Montjoye, Y. A., & Ratti, C. (2018). Towards matching user mobility traces in large-scale datasets. IEEE Transactions on Big Data, 6(4), 714-726.

Kraemer, M. U., Yang, C. H., Gutierrez, B., Wu, C. H., Klein, B., Pigott, D. M., ... & Scarpino, S. V. (2020). The effect of human mobility and control measures on the COVID-19 epidemic in China. Science, 368(6490), 493-497.

Kuchler, T., Russel, D., & Stroebel, J. (2020). The geographic spread of COVID-19 correlates with the structure of social networks as measured by Facebook (No. w26990). National Bureau of Economic Research.

Kurita, J., Sugishita, Y., Sugawara, T., & Ohkusa, Y. (2021). Evaluating Apple Inc Mobility Trend Data Related to the COVID-19 Outbreak in Japan: Statistical Analysis. JMIR public health and surveillance, 7(2), e20335.

Lamb, M. R., Kandula, S., & Shaman, J. (2020). Differential COVID-19 case positivity in New York City neighborhoods: Socioeconomic factors and mobility. *Influenza and Other Respiratory Viruses*. doi:10.1111/irv.12816

Lau, B. P. L., Marakkalage, S. H., Zhou, Y., Hassan, N. U., Yuen, C., Zhang, M., & Tan, U. X. (2019). A survey of data fusion in smart city applications. Information Fusion, 52, 357-374.

Lau, H., Khosrawipour, V., Kocbach, P., Mikolajczyk, A., Schubert, J., Bania, J., & Khosrawipour, T. (2020). The positive impact of lockdown in Wuhan on containing the COVID-19 outbreak in China. Journal of travel medicine, 27(3), taaa037.

Lau, Max SY, et al. "Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA." Proceedings of the National Academy of Sciences 117.36 (2020): 22430-22435, https://doi.org/10.1073/pnas.2011802117

Lee M, Zhao J, Sun Q, Pan Y, Zhou W, Xiong C, Zhang L. Human mobility trends during the COVID-19 pandemic in the United States. arXiv 2020; published online May 4, 2020. http://arxiv.org/abs/2005.01215 (preprint).

Levin, R., Chao, D. L., Wenger, E. A., & Proctor, J. L. (2020). Cell phone mobility data reveals heterogeneity in stay-at-home behavior during the SARS-CoV-2 pandemic. *medRxiv*. doi: 10.1101/2020.10.31.20223776.

Li, Z. (2020). Geospatial big data handling with high performance computing: Current approaches and future directions. In High Performance Computing for Geospatial Applications (pp. 53-76). Springer, Cham.

Li, Z., Huang, X., Ye, X., & Li, X. (2020a). ODT flow explorer: Extract, query, and visualize human mobility. arXiv preprint arXiv:2011.12958.

Li, Z., Huang, X., Ye, X., Jiang, Y., Yago, M., Ning, H., ... & Li, X. (2021). Measuring Place Connectivity Using Big Social Media Data. arXiv preprint arXiv:2102.03991.

Li, Z., Li, X., Porter, D., Zhang, J., Jiang, Y., Olatosi, B., & Weissman, S. (2020b). Monitoring the Spatial Spread of COVID-19 and Effectiveness of Control Measures Through Human Movement Data: Proposal for a Predictive Model Using Big Data Analytics. JMIR Research Protocols, 9(12), e24432.

Liu, M., Ning, J., Du, Y., Cao, J., Zhang, D., Wang, J., & Chen, M. (2020a). Modelling the evolution trajectory of COVID-19 in Wuhan, China: experience and suggestions. Public health, 183, 76-80.

Liu, Q., Sha, D., Liu, W., Houser, P., Zhang, L., Hou, R., ... & Yang, C. (2020b). Spatiotemporal patterns of COVID-19 impact on human activities and environment in mainland China using night-time light and air quality data. Remote Sensing, 12(10), 1576.

Liu, Y., Wang, S., & Xie, B. (2019). Evaluating the effects of public transport fare policy change together with built and non-built environment features on ridership: The case in South East Queensland, Australia. Transport Policy, 76, 78-89.

Martín Y., Cutter S.L. Li Z, Emrich C., Mitchell, J.T. (2020) Using geotagged tweets to track population movements to and from Puerto Rico after Hurricane Maria. *Population and Environment* , https://doi.org/10.1007/s11111-020-00338-6

Meloni, S., Perra, N., Arenas, A., Gómez, S., Moreno, Y., & Vespignani, A. (2011). Modeling human mobility responses to the large-scale spreading of infectious diseases. Scientific reports, 1(1), 1-7.

Mohammadi, H., Rajabifard, A., & Williamson, I. P. (2010). Development of an interoperable tool to facilitate spatial data integration in the context of SDI. International Journal of Geographical Information Science, 24(4), 487-505.

Montero, L., Ros-Roca, X., Herranz, R., & Barceló, J. (2019). Fusing mobile phone data with other data sources to generate input OD matrices for transport models. Transportation research procedia, 37, 417-424.

Murugesan, S. (2007). Understanding Web 2.0. IT professional, 9(4), 34-41.

Nakada, L. Y. K., & Urban, R. C. (2020). COVID-19 pandemic: Impacts on the air quality during the partial lockdown in São Paulo state, Brazil. Science of the Total Environment, 730, 139087.

Ni, S., & Weng, W. (2009). Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks. Physical review E, 79(1), 016111.

Oliver, N., Lepri, B., Sterly, H., Lambiotte, R., Deletaille, S., De Nadai, M., ... & Vinck, P. (2020). Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle.

O'Sullivan, D., M. Gahegan, D. J. Exeter and B. Adams (2020). "Spatially explicit models for exploring COVID-19 lockdown strategies." Transactions in GIS 24(4): 967-1000.

Pan Y, Darzi A, Kabiri A, Zhao G, Luo W, Xiong C, Zhang L. Quantifying human mobility behavior changes in response to non-pharmaceutical interventions during the COVID-19 outbreak in the United States. arXiv 2020; published online May 4, 2020. http://arxiv.org/abs/2005.01224 (preprint).

Park, Y. J., Choe, Y. J., Park, O., Park, S. Y., Kim, Y. M., Kim, J., ... & Lee, J. (2020). Contact tracing during coronavirus disease outbreak, South Korea, 2020. Emerging infectious diseases, 26(10), 2465-2468.

Peng, Z., Wang, R., Liu, L., & Wu, H. (2020). Exploring urban spatial features of COVID-19 transmission in Wuhan based on social media data. ISPRS International Journal of Geo-Information, 9(6), 402.

Pepe, E., Bajardi, P., Gauvin, L., Privitera, F., Lake, B., Cattuto, C., & Tizzoni, M. (2020). COVID-19 outbreak response, a dataset to assess mobility changes in Italy following national lockdown. Scientific data, 7(1), 1-7.

Porcher, Simon, and Thomas Renault. "Social Distancing Beliefs and Human Mobility: Evidence from Twitter." arXiv preprint arXiv:2008.04826 (2020), https://arxiv.org/abs/2008.04826

Qi, G., Li, X., Li, S., Pan, G., Wang, Z., & Zhang, D. (2011, March). Measuring social functions of city regions from large-scale taxi behaviors. In *2011 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)* (pp. 384-388). IEEE.

Rutz, C., Loretto, M. C., Bates, A. E., Davidson, S. C., Duarte, C. M., Jetz, W., ... & Cagnacci, F. (2020). COVID-19 lockdown allows researchers to quantify the effects of human activity on wildlife. Nature Ecology & Evolution, 4(9), 1156-1159.

SafeGraph Inc. (2020, October 1). SafeGraph COVID-19 Data. https://www.safegraph.com/covid-19-data-consortium

Salathé, M., Althaus, C. L., Neher, R., Stringhini, S., Hodcroft, E., Fellay, J., ... & Eckerle, I. (2020). COVID-19 epidemic in Switzerland: on the importance of testing, contact tracing and isolation. Swiss medical weekly, 150(11-12), w20225.

Shen, J. (2020). Covid-19 and inter-provincial migration in China. *Eurasian Geography and Economics*, *61*(4-5), 620-626.

Song, C., Qu, Z., Blumm, N., & Barabási, A. L. (2010). Limits of predictability in human mobility. Science, 327(5968), 1018-1021.

Su, L., Hong, N., Zhou, X., He, J., Ma, Y., Jiang, H., ... & Long, Y. (2020). Evaluation of the secondary transmission pattern and epidemic prediction of COVID-19 in the four metropolitan areas of China. *Frontiers in medicine*, *7*.

Su, Yue, et al. "Examining the impact of COVID-19 lockdown in Wuhan and Lombardy: a psycholinguistic analysis on Weibo and Twitter." International Journal of Environmental Research and Public Health 17.12 (2020): 4552.https://doi.org/10.3390/ijerph17124552

The OpenSky Network.(2020). http://www.opensky-network.org

Tobías, A. (2020). "Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up." Science of The Total Environment 725: 138539.

Torre-Bastida, A. I., Del Ser, J., Laña, I., Ilardia, M., Bilbao, M. N., & Campos-Cordobés, S. (2018). Big Data for transportation and mobility: recent advances, trends and challenges. IET Intelligent Transport Systems, 12(8), 742-755.

Unacast. (2020). Social Distancing Scoreboard. COVID-19 Toolkit. https://www.unacast.com/covid19/social-distancing-scoreboard

University of Maryland (2020). University of Maryland COVID-19 Impact Analysis Platform,https://data.covid.umd.edu

Usery, E. L., Finn, M. P., Starbuck, M., & Center, M. C. M. (2005). Integrating data layers to support the national map of the United States. In *International Cartographic Conference. A Corua, Spain*.

Vokó, Z., & Pitter, J. G. (2020). The effect of social distance measures on COVID-19 epidemics in Europe: an interrupted time series analysis. GeroScience, 42(4), 1075-1082.

Wang, S., Liu, Y., & Hu, T. (2020). Examining the change of human mobility adherent to social restriction policies and its effect on COVID-19 cases in Australia. International journal of environmental research and public health, 17(21), 7930.

Warren, M. S., & Skillman, S. W. (2020). Mobility changes in response to COVID-19. arXiv preprint arXiv:2003.14228.

Watts, A., Au, N. H., Thomas-Bachli, A., Forsyth, J., Mayah, O., Popescu, S., & Bogoch, I. I. (2020). Potential for inter-state spread of Covid-19 from Arizona, USA: analysis of mobile device location and commercial flight data. Journal of travel medicine.

Xiong C, Hu S, Yang M, Younes HN, Luo W, Ghader S, Zhang L. Data-driven modeling reveals the impact of stay-at-home orders on human mobility during the COVID-19 pandemic in the U.S. arXiv 2020; published online May 2, 2020. http://arxiv.org/abs/2005.00667 (preprint).

Xu, F., Z. Tu, Y. Li, P. Zhang, X. Fu and D. Jin (2017). Trajectory recovery from ash: User privacy is not preserved in aggregated mobility data. Proceedings of the 26th international conference on world wide web.

Xu, P., Dredze, M., & Broniatowski, D. A. (2020). The Twitter Social Mobility Index: Measuring Social Distancing Practices With Geolocated Tweets. *Journal of medical Internet research*, *22*(12), e21499.

Xu, X., Wang, S., Dong, J., Shen, Z., & Xu, S. (2020). An analysis of the domestic resumption of social production and life under the COVID-19 epidemic. *PloS one*, *15*(7), e0236387.

Xu, Y., Belyi, A., Bojic, I., & Ratti, C. (2018). Human mobility and socioeconomic status: Analysis of Singapore and Boston. Computers, Environment and Urban Systems, 72, 51-67.

Yabe, T., Tsubouchi, K., Fujiwara, N., Wada, T., Sekimoto, Y., & Ukkusuri, S. V. (2020). Non-compulsory measures sufficiently reduced human mobility in Tokyo during the COVID-19 epidemic. Scientific reports, 10(1), 1-9.

Yang, C., Sha, D., Liu, Q., Li, Y., Lan, H., Guan, W. W., ... & Ding, A. (2020). Taking the pulse of COVID-19: A spatiotemporal perspective. International journal of digital earth, 13(10), 1186-1211.

Yuan, J., Zheng, Y., & Xie, X. (2012, August). Discovering regions of different functions in a city using human mobility and POIs. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 186-194).

Zang, H. and J. Bolot (2011). Anonymization of location data does not work: A large-scale measurement study. Proceedings of the 17th annual international conference on Mobile computing and networking.

Zeng, C., Zhang, J., Li, Z., Sun, X., Olatosi, B., Weissman, S., & Li, X. (2021). Spatial-temporal relationship between population mobility and COVID-19 outbreaks in South Carolina: A time series forecasting analysis. medRxiv. doi:10.1101/2021.01.02.21249119.

Zhang, C., C. Chen, W. Shen, F. Tang, H. Lei, Y. Xie, Z. Cao, K. Tang, J. Bai, L. Xiao, Y. Xu, Y. Song, J. Chen, Z. Guo, Y. Guo, X. Wang, M. Xu, H. Zou, Y. Shu and X. Du (2020). "Impact of population movement on the spread of 2019-nCoV in China." Emerging Microbes & Infections 9(1): 988-990.

Zhang L, Ghader S, Pack ML, Xiong C, Darzi A, Yang M, Sun Q, Kabiri A, Hu S. An interactive COVID-19 mobility impact and social distancing analysis platform. medRxiv 2020; published online May 5, 2020. DOI: https://doi.org/10.1101/2020.04.29.20085472 (preprint).

Zhang, D., Huang, J., Li, Y., Zhang, F., Xu, C., & He, T. (2014, September). Exploring human mobility with multi-source data at extremely large metropolitan scales. In Proceedings of the 20th annual international conference on Mobile computing and networking (pp. 201-212).

Zhang, Y., Zhang, A., & Wang, J. (2020). Exploring the roles of high-speed train, air and coach services in the spread of COVID-19 in China. Transport Policy.

Zhao, K., Tarkoma, S., Liu, S., & Vo, H. (2016, December). Urban human mobility data mining: An overview. In 2016 IEEE International Conference on Big Data (Big Data) (pp. 1911-1920). IEEE.

Zheng, Y., Liu, Y., Yuan, J., & Xie, X. (2011, September). Urban computing with taxicabs. In *Proceedings of the 13th international conference on Ubiquitous computing* (pp. 89-98).

Zhou, Y., Wang, L., Zhang, L., Shi, L., Yang, K., He, J., ... & Song, P. (2020). A spatiotemporal epidemiological prediction model to inform county-level COVID-19 risk in the United States. Special Issue 1-COVID-19: Unprecedented Challenges and Chances.

Zhuang, Z., Cao, P., Zhao, S., Lou, Y., Wang, W., Yang, S., ... & He, D. (2020). Estimation of local novel coronavirus (COVID-19) cases in Wuhan, China from off-site reported cases and population flow data from different sources. *medRxiv*.

Zhuang, Z., Zhao, S., Lin, Q., Cao, P., Lou, Y., Yang, L., & He, D. (2020). Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data. International Journal of Infectious Diseases, 94, 29-31.

Zikopoulos, P., & Eaton, C. (2011). Understanding big data: Analytics for enterprise class hadoop and streaming data. New York: McGraw-Hill.