Monitoring the Spatial Spread of COVID-19 and Effectiveness of the Control Measures through Human Movement using Big Social Media Data: A Study Protocol

Zhenlong Li[1*], Xiaoming Li[2], Dwayne E. Porter[3], Jiajia Zhang[4], Yuqin Jiang[1], Bankole Olatosi[5], Sharon Weissman[6]

[1] Department of Geography, College of Arts and Sciences, University of South Carolina, Columbia, SC, USA

[2] Department of Health Promotion, Education, and Behavior, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

[3] Department of Environmental Health Sciences, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

[4] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

[5] Department of Health Services Policy and Management, Arnold School of Public Health, University of South Carolina, Columbia, SC, USA

[6] Department of Internal Medicine, School of Medicine, University of South Carolina, Columbia, SC, USA

[*]Corresponding Author

Zhenlong Li, Ph.D.
Associate Professor, Department of Geography
Director, Geoinformation and Big Data Research Lab
Big Data Health Science Center
Hazards and Vulnerability Research Institute
South Carolina SmartState Center for Healthcare Quality
University of South Carolina
709 Bull Street, Room 320, Columbia, SC 29208, USA
zhenlong@sc.edu
Phone: +1 803-777-4590

Keywords: big data; human movement; spatial computing; covid-19; artificial intelligence

ABSTRACT

Introduction: Human movement is among the essential forces that drive spatial spread of infectious diseases. To date, reducing and tracking human movement during the pandemic have proven effective in limiting the spread of COVID-19. Existing methods for monitoring and modeling the spatial spread of infectious diseases rely on various data sources as proxies of human movement, such as airline travel data, mobile phone data, and dollar bills tracking. However, intrinsic limitations of these data sources prevent us from systematic monitoring and analyses of human movement from different spatial scales (from local to global). Big social media data such as geotagged tweets have been widely used in human mobility studies, yet more research are needed to validate the capabilities and limitations of using such data for studying human movement at different geographic scales (e.g., from local to global) in the context of global infectious disease transmission.

Method and Analysis: This research will first develop a database with optimized spatiotemporal indexing to store and manage the multi-source datasets collected in this project. This database will be connected to our in-house Hadoop computing cluster for efficient big data computing and analytics. This research will then develop a novel data-driven approach, including innovative data models, predictive models, and computing algorithms, to effectively extract and analyze human movement patterns from big geotagged Twitter data for enhancing situational awareness and risk prediction in public health emergency response and disease surveillance systems. Research findings can help government officials, public health managers and emergency responders to answer critical questions during the pandemic regarding the current and future infectious risk of a state, county, or community and the effectiveness of the social/physical distancing practice in curtaining the spread of the virus.

Ethics and Dissemination: This research does not involve human subjects and received an exempt review from the Institutional Review Board (IRB). All data collected in this project are from public domains. Geotagged Twitter data are collected using the official Twitter Streaming Application Programming Interface (API). Twitter developer polices are strictly followed when collecting and dissimiating Twitter data. The raw individual geotagged tweets with exact latitude and longitude will not be published in any way including maps, technical report, or journal publications. All data collected in this project will be stored in an in-house Hadoop computing cluster hosted at the University of South Carolina with firewall protection and two-factor authentication & endpoint security. Results of this project will be disseminated as maps, summary graphics, news reports, and research articles.

Keywords: big data; human movement; spatial computing; covid-19; artificial intelligence

INTRODUCTION

The coronavirus disease 2019 (COVID-19), which is caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was originally detected in Wuhan, China, in December 2019. The World Health Organization (WHO) declared COVID-19 outbreak as a pandemic on March 11, 2020 [1] due to its rapid spread to several geographic regions. To contain the spread of COVID-19, unprecedented measures such as mass quarantines of cities such as Wuhan, and even lockdown of entire countries, such as Italy, have been taken. Due to the rapid human-to-human transmission of COVID-19, models or measurements that help understand potential infectious risk at different geographic levels can play an essential role for residents, medical workers, and governments. Such models can help local authorities and communities better allocate resources and efforts at a community-level. Meanwhile, it is equally important for policy makers and emergency responders to understand how people practice the social/physical distancing and how effective these control measures are at curbing the spatial propagation of virus.

Human movement is an important driver of the geographic spread of infectious diseases [2]. For example, studies on Severe Acute Respiratory Syndrome (SARS) [3], Middle East Respiratory Syndrome (MERS) [4], and influenza H1N1[5,6] all confirmed that airline travel was a major transmission mechanism at a large spatial scale. From a public health perspective, prediction and control of the spread of infectious disease benefits greatly from our growing capacity to quantify human movement [7]. COVID-19 has a high human-to-human transmission rate and can be transmitted during the pre-clinical incubation period. So far, limiting and tracking human movement during the outbreak has proven effective at reducing the spread of COVID-19 in different countries [8–10]. In this sense, monitoring and analyzing human movement patterns or population flows at different spatial scales (global, country, state, county, and community) is critical for us to gain better understanding on the current and future infectious risk at population level during the pandemic. Such situational awareness can help governments at all levels (local, state, federal, and international) proactively reallocate medical supplies and medical workforces to more vulnerable areas for better preparation and readiness in fighting the COVID-19 pandemic.

Existing studies have used various data sources to quantify human movement for infectious disease modeling. At large scale, airline data are important sources in understanding global transmission of infectious diseases. For example, global spread of SARS simulation models have been generated with airline data [11]. However, airline travel data have shown a limited usefulness in understanding transmission across short distances [12]. While airline data deepened understanding for transmission mechanism of infectious disease at large geographical scale, at small scale, the transmission mechanism can be different [13]. At local scale, mobile phone data were used as a measurement of human mobility and demonstrated its utility in understanding spatial transmission patterns of malaria [14], cholera [15], and influenza [16]. However, due to privacy issues, mobile phone data are limited in terms of accessibility and in situations where such data are available and sampled populations are often biased to users of sponsored carrier companies [17]. In addition, mobile phone data are often limited to a local region or at most a country and cannot provide a systematic global coverage. Besides mobile phone data, commuting patterns derived from census data also play important roles in understanding the spread patterns of virus at local scale [13,18].

With the increasing prevalence of location-enabled social media, geotagged Twitter data have been widely used in human mobility studies (e.g., [19–21]), yet limited research has been conducted to validate the potential and limitations of using these data for studying human movement at different geographic scales (e.g., from global to local) in the context of global infectious disease transmission. Meanwhile, the recent development of Artificial Intelligence (AI) has been proven useful in assisting diagnosis, drug analysis, data collection, and outbreak prediction [22]. Various types of neural network algorithms have demonstrated capacity in predicting HIV epidemic[23], influenza-like illness [24], and SARS [25]. However, the majority of these AI-based prediction algorithms have focused on mathematical models of trend development and outbreak identification, in which limited geospatial information (especially at different geographic scales) is considered. The recent COVID-19 pandemic provides us with a unique opportunity to explore innovative approaches to effectively use big Twitter data and AI-based algrithms, and examine their efficiency for enhancing situational awareness and risk prediction in public health emergency response and disease surveillance systems.

By leveraging the interdisciplinary team's collective expertise in spatio-temporal modeling, big data analytics, infectious disease, spatial epidemiology, and health promotion and behavior modification, we propose to develop a novel data-driven public health approach using big Twitter data and AI to monitor and analyze human movement at different spatial scales (from global to regional to local) for enhancing situational awareness and risk prediction in public health emergency response and disease surveillance systems. With the proposed approach, we aim to answer the following critical questions during the COVID-19 pandemic: (1) Where are people coming from and going to during the pandemic? We will answer this question by developing an Origin-Destination-Time data cube (ODT cube) to efficiently extract historical and near real-time population flows from worldwide geotagged tweets; (2) What is the current and future infectious risk of a country, state, or county? This will be estimated using a spatial-temporal fused neural network considering the historical human movement patterns and real-time population flows; (3) How well are people following the social/physical distancing orders? This question will be examined by performing spatial-temporal aggregation of the ODT cube at different spatial scales and temporal resolutions to quantify the human movement at different spatial scales; and (4) How effective are the social/physical distancing practice in curtailing the spread of the virus? We will answer this question by conducting spatio-temporal and geostatistical analysis (e.g., regression and correlation) for the aggregated population flows, the daily confirmed cases, and other factors such as face mask policies. The answers to these questions will be compiled as maps, diagrams, and news releases, technique reports, and peer-reviewed journal articles.

METHODS AND ANALYSIS

DATA COLLECTION AND DATABASE

This project will collect the following four types of data worldwide (where data are available): 1) geotagged Twitter data, 2) daily confirmed COVID-19 cases at the available highest spatial resolution for all countries, 3) most recent socioeconomic and demographic information (county level in the U.S. and similar level of administrative unit for other countries), and 4) data from other human mobility data sources such as mobile phone data-based human mobility, Google Mobility report, and Apple Mobility report. We have developed a computer program to stream geotagged tweets using Twitter's streaming application programming interface (API). Worldwide historical geotagged Twitter data collected by the team over the past five year will also be used to construct past population flows and identify spatio-temporal patterns of human movement. We will develop a database to store and manage the aforementioned multi-source datasets collected in this project. The database will be indexed with multi-level spatial scales (e.g., country, state, and county) and temporal resolutions (e.g., year, month, day) and will be connected to our in-house Hadoop computing cluster for efficient big data computing and analytics.

ANALYTIC APPROACH

Develop an Origin-Destination-Time data cube for efficient analysis of human movement from massive geotagged tweets with varying spatio-temporal scales

Data cube has been widely used to model high dimensional spatio-temporal data (e.g., [26,27]). We will develop an Origin-Destination-Time data cube (ODT cube) as a high-level conceptual model for quantifying the human movement across different places or locations over time (Figure 1) from billions of geotagged tweets. The ODT cube will serve as a foundation data model for efficiently conducting human movement analysis at different spatial and temporal scales. In the ODT cube, origin (O) and destination (D) is a set of places or locations (e.g., administrative boundaries such as county, state, and country, or latitude/longitude grids) that can be displayed with a map. Each cell in the data cube has a value that indicates the number of people moved from the origin location to the destination location during a specific time period (e.g., in an hour, a day, or a month). In other words, each cell value indicates the connection (measured by population movement) between two locations. With the cube, we can efficiently retrieve the number of people moved from O_i to D_j at time T_k.

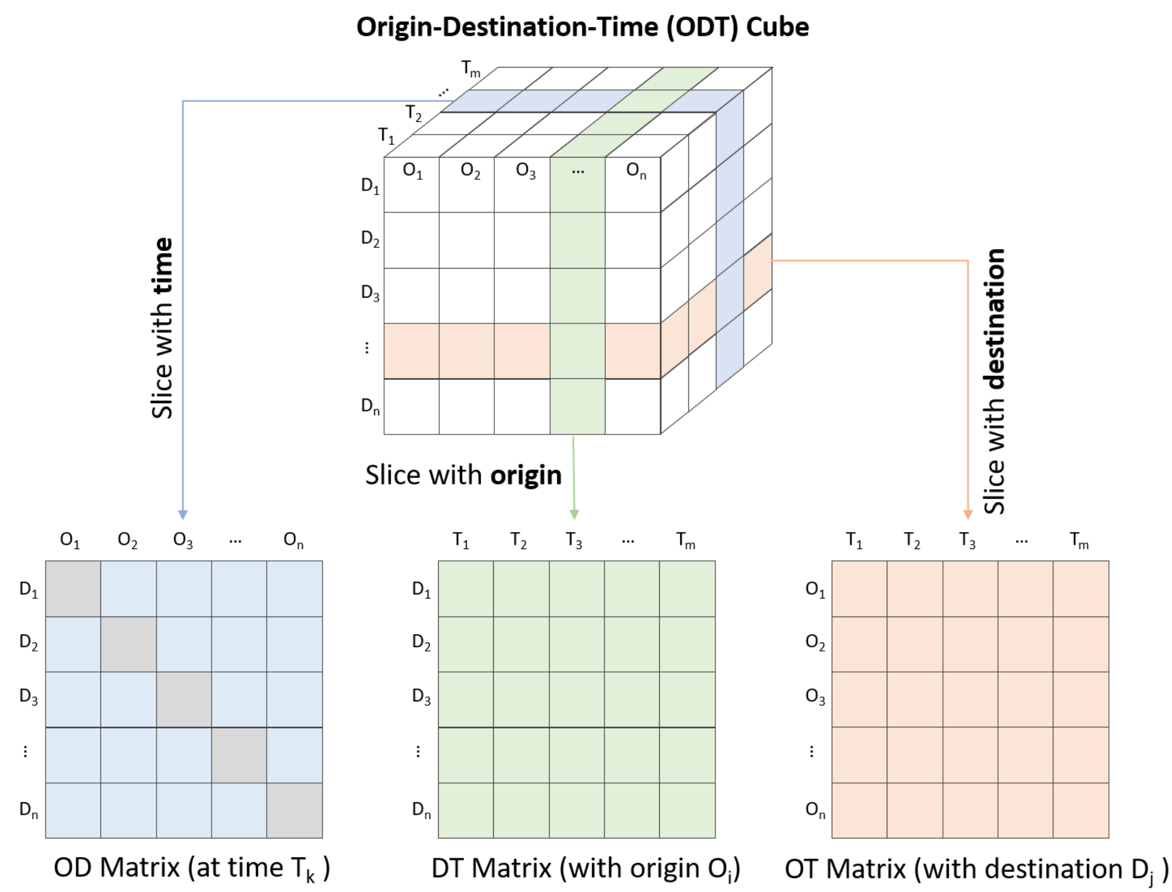

Figure 1. Illustration of Origin-Destination-Time Cube for modeling human movement

Three types of matrices will be derived from the data cube: origin-destination (OD) matrix quantifies the population flows between all the origin and destination locations during a time period. Destination-time (DT) matrix captures the number of incoming people to all destination locations from a specific origin location over a series of times. Similarly, an origin-time (OT) matrix captures the number of outgoing people from all origins to a specific destination over a series of times. In addition, the number of unique Twitter users can be calculated for a specific location over time. This enables us to efficiently conduct spatial-temporal aggregations of human movement at varying spatial and temporal resolutions.

The OD matrix is a $n \times n$ matrix, where n is the number of geographic entities included in the study. Column O_x and row D_x are the same location (x). An entry v_{ij} in this matrix represents number of people moving from origin i to destination j. It should be noted that human movements are directional. Therefore, v_{ij} and v_{ji} stand for two different spatiotemporal movements that they are likely to have different values. We define the values in the diagonal cells (grey cells in the OD matrix), v_{ii}, as the number of unique Twitter users in location i.

The process of constructing the ODT cube is extremely data- and computationally intensive because we need to perform massive number of point-in-polygon spatial operations, and the output will contain billions of connections. We will leverage our expertise in geospatial big data computing to perform the computation using an in-house Hadoop-based computing cluster. Based on the generated ODT cube, we will further derive a number of indices to quantify human mobility at varying spatio-temporal scales including county level daily number of Twitter visitors for a county, state level average travel distance, and place connectedness index between two counties.

Develop population-level infectious risk maps at different spatial scales based on population flows to enhance situational awareness

The ODT cube quantifies human movement among different places (e.g., US counties) during a given time period. Knowing such movement information is essential in assessing the infectious risk at the population level in a place. We propose to model the current infection risk of a given place (e.g., county) by integrating the following information: 1) population flows derived from the ODT cube during the recent time period among all places (e.g., past 14 days), 2) number of total infected cases for each place, and 3) socioeconomic and demographic variables that relate to the infection risk of this location (e.g., a county's population density, age, and race).

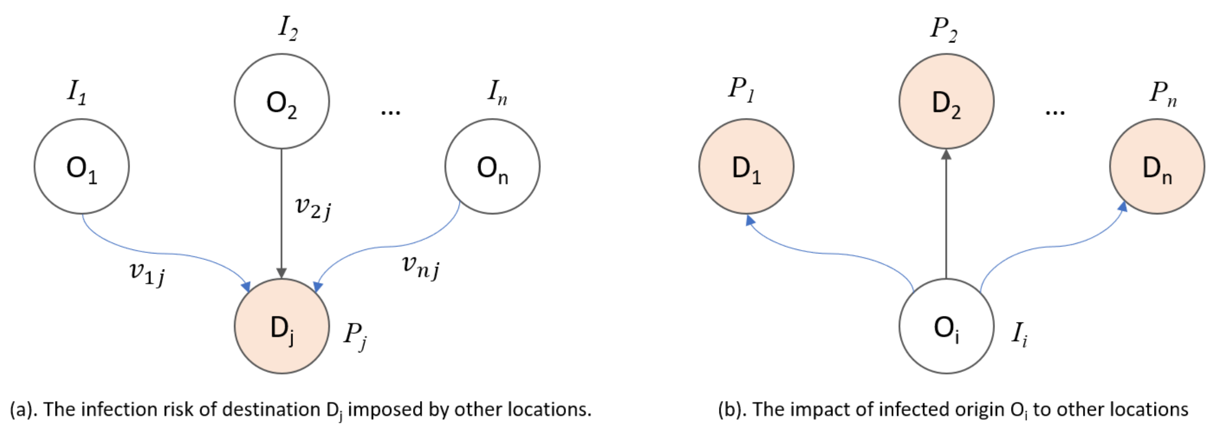

Figure 2. Illustration of (a) the infection risk modeling based on the incoming population to a location; and (b) the impact modeling of an infected location to other locations.

We will create an infection risk index for each place by combining the above mentioned factors. For example, suppose, based on the ODT cube, we observe a significant population flow from county *A* to county *B* during the past week and county *A* already has a number of infected cases, then the infectious risk for county *B* is high (people from highly infected area is likely to carry the virus). Note that the real scenario is more complex due to the fact that the risk of county B is also affected by other infected counties that have connections with county A and that population movement is not the only factor for the infectious risk. In other words, infection risk of destination D_j can be considered as a function of local factors (P_j), combined with population flow from each origin ($v_{1j}, v_{2j}, \ldots v_{nj}$) weighted by infected cases at each origin ($I_1, I_2, \ldots, I_n$) (Figure 2(a)). Risk index will be calculated for each location to produce an infectious risk map. Based on the ODT data cube, the risk map generation can be efficiently implemented using matrix computation. Such risk maps would be useful for targeting surveillance and outbreak control activities for a region.

Besides modeling the infection risk of a location using the incoming populations, we will also estimate the risk impact of an infected location on other locations. For example, since Italy was severely infected at the early stage of the pandemic, it would be helpful to understand where the outgoing population from Italy traveled to. As illustrated in Figure 2(b), we will build a model that combines the population movement information between the targeted location (O_i) to other locations ($D_1, D_2, \ldots, D_n$) as well as other factors associated with each location ($P_1, P_2, \ldots, P_n$). The output of the model would be a map showing the potential impact caused by the incoming populations from the targeted location (e.g., Italy).

Develop a predictive model to estimate future infectious risk using a fused neural network by considering both spatial patterns and temporal trends of the popoplation movement

In this research task, we aim to explore the feasibility and performance of a predictive model for future infectious disease potential at the US county level based on the following information: 1) historical population movement patterns among counties (based on historical tweets), 2) near real-time human movement information (from real-time twitter data streams), 3) the daily infection count of each county (will be collected/compiled each day), and 4) other socioeconomic/demographic factors. Specifically, the historical population movement patterns between any two counties at different time resolutions (daily, weekly, monthly) will be extracted from the ODT cube. The extracted movement patterns will be used as the proxy to infer future population movement among the counties.

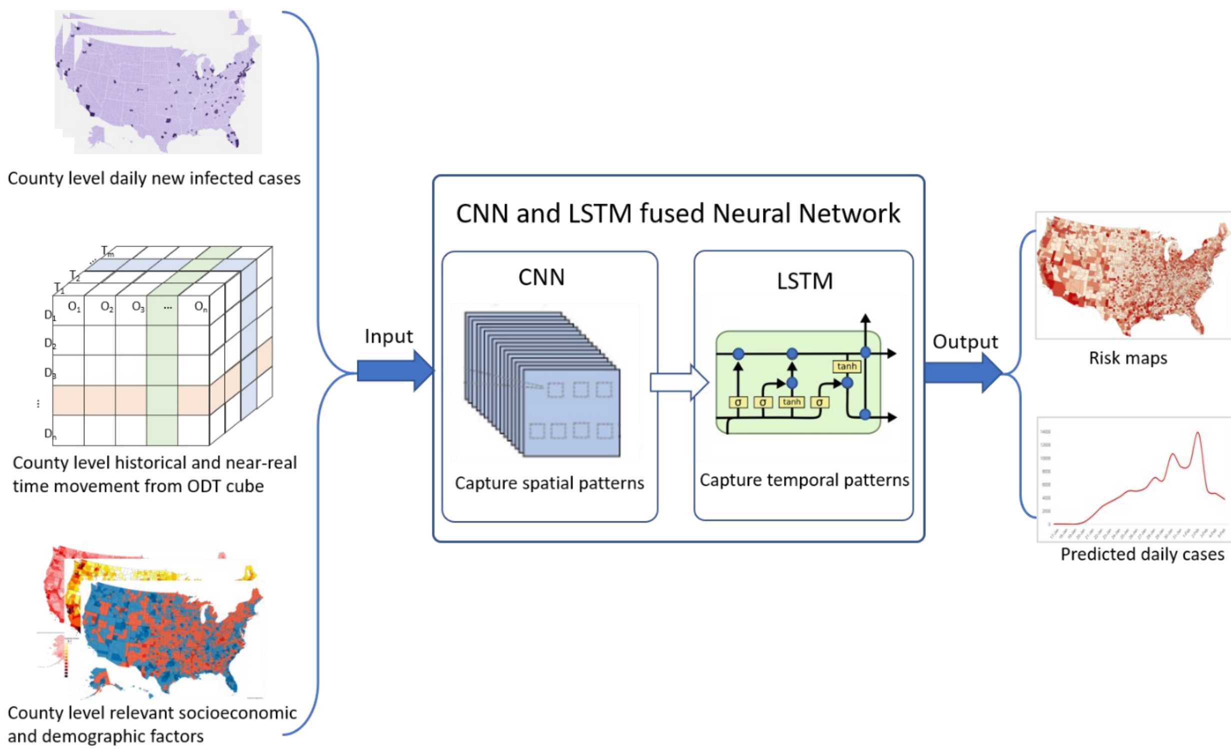

Figure 3. Conceptual architecture of the CNN-LSTM fused neural network for infectious risk prediction

Given the complex epidemiological and geographic processes among different infectious factors, we propose to use deep learning to learn the complex infectious processes from the large volumes and high dimensions of the input data. Deep learning is one type of machine learning in artificial intelligence. Unlike traditional machine learning, in which parameters of an algorithm (e.g., support vector machine) are configured by experts, deep learning determines these parameters by learning the patterns in a large amount of data based on artificial neural networks, which offers a promising solution in predicting the infectious risk. Specifically, we will develop a fused neural network that integrates two types of neural networks, CNN (convolutional neural network) and LSTM (long short-term memory recurrent neural network), to consider the spatial patterns and temporal trends simultaneously in the predictive model (Figure 3). The fused neural network will include a serious of CNN layers in the front end followed by LSTM layers with a Dense layer on the output. The locations in the ODT cube (e.g., counties) would be treated as pixels (neurons) in the CNN network to capture the spatial relationships and local patterns, and the temporal trend will be predicted with the LSTM network. Different combinations of socioeconomic/demographic factors will be tested during the model building, training, and validation process, and the combination yield highest accuracy will be used in the final model.

Patient and Public Involvement

Patients or the public WERE NOT involved in the design, or conduct, or reporting, or dissemination plans of our research.

DISCUSSIONS

In this paper, we report a research protocol that will utilize big social media data to derive information on human movement or population flows to monitor the spatial spread of COVID-19, quantify the effectiveness of the control measures, and predict the current and future infectious risk at various geospatial scales. We believe geotagged Twitter data are sufficient for studying the population flows in a large spatial scale with low or medium spatial resolutions, such as the movement between countries and between states in the U.S. For the county level, our previous studies indicate that these data perform well for examining human movement between different U.S. counties [28–30]. For finer resolutions than county, we have successfully conducted human mobility studies at the census tract level [21] and street/community level within a city [31]. However, we are aware that studies at a spatial resolution higher than city or county only work in highly populated areas since at this resolution we can only use the tweets with exact coordinates. Considering this issue, we will only perform the community level analysis for highly populated cities such as New York City. In addition, we will devise a method to extract finer location from the city level tweets by geocoding the location names in the tweet message.

Another limitation we would like to point out is that Twitter is not proportionally used by different population groups and thus shows demographic and socioeconomic biases as examined in a few studies [32–34]. However, this is less of a concern in this study since we will not examine how the population movements derived from Twitter data represent different population groups. Instead, we are more interested in how well the geotagged tweets sample can represent the overall population movement (e.g., the number of people travelled to a county during a time period). [19] confirms that geotagged tweets are exceptionally useful in quantifying country to country population movement. Our recent study suggests that the county level population movement derived from Twitter data can accurately reflect regular (such as holidays) and non-regular events such as the Hurricanes. [28]. In addition, we are comparing and integrating Twitter-derived human mobility with other human mobility data sources such as Apple mobility data, Google mobility data, and mobile phone data, to better understand human movement during the pandemic [35].

Lastly, geotagged Twitter data contains location information and may contain some personal information provided by the users directly. We are fully aware of the potential privacy concerns and will remove or mask the personal information when detected. The raw individual tweets with exact latitude and longitude will not be published in any way including maps, technical report, or journal publications. All data collected in this study will be stored in an in-house Hadoop computing cluster hosted in a secured server room at the University of South Carolina with firewall protection and two-factor authentication and endpoint security.

CONCLUSION

Human movement is among the essential forces that drive the spatial spread of COVID-19. During a global pandemic, monitoring and analyzing human movement patterns or population flows are critical for us to gain a better understanding into current and future infectious risk at the population level. This research aims to use big social media data (Twitter), artificial intelligence (AI), and spatio-temporal analysis to monitor and model the spatial spread of COVID-19 at different spatial scales (from local to regional to global) through the lens of human movement. Results of this study will not only provide enhanced situation awareness for the government at all levels, but also offer valuable contributions to building collective public awareness of the role people play in the evolution of the COVID-19 crisis.

The findings of the research may also have implications on policy domain by assisting the policy makers and general public to evalue the effectiveness of various control measures that aim to reduce the human movement during the pandemic. For example, the debate about the true effectiveness of social distancing as a public health tool in limiting COVID-19 transmission requires mobility research to generate evidence-based facts [36]. This is important especially in an era with mixed research findings about COVID-19 aerosolization [32,37,38] or the true effectiveness and costs of social distancing [39,40]. As universities and schools reopen, and traditional socialization activities like sporting/musical events resume, measuring and tracking the impact of human mobility takes on greater significance for populations.

We hope that the results can help government officials, public health managers and emergency responders to answer critical questions during the pandemic as elaborated above. Although this research is a response to the current COVID-19 pandemic, the proposed research will make significant contributions to the data sources, applications, models, and methodology in a variety of human mobility studies. This research is expected to have a broad impact on diverse fields that can benefit from a better understanding of human movement at varying spatial scales, such as infectious disease spread in public health, transportation, tourism, and economics.

ACKNOLEDGEMENTS: ZL, XL, and DP conceptualized and designed the study protocol. ZL, YJ, and JZ contributed to the methology for the study. ZL, XL, and BO are responsible for the study coordination. ZL, YJ, and JZ are responsible for data quality control, management, and analysis. SW, XL, and JY made substantial contributions to manuscript editing. All authors read and contributed to the writing of the study protocol and approved the final draft of the manuscript.

FUNDING: The study is supported by National Science Foundation (NSF) under grant 2028791 (PI: ZL), the National Institute of Allergy and Infectious Diseases (NIAID) of the National Institutes of Health (NIH) under grant R01AI127203-4S1 (PI: XL and BO) , and the University of South Carolina COVID-19 Internal Funding Initiative under grant 135400-20-54176 (PI: ZL). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of this article.

Conflict of Interest: All authors declare that they have no conflict of interest.

References

1. Archived: WHO Timeline - COVID-19 [Internet]. [cited 2020 Aug 28]. Available from: https://www.who.int/news-room/detail/27-04-2020-who-timeline---covid-19

2. Kraemer M, Golding N, Bisanzio D, Bhatt S, Pigott D, Ray S, Brady O, Brownstein J, Faria N, Cummings D. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. Sci Rep Nature Publishing Group; 2019;9(1):1–11.

3. Christian MD, Poutanen SM, Loutfy MR, Muller MP, Low DE. Severe acute respiratory syndrome. Clin Infect Dis The University of Chicago Press; 2004;38(10):1420–1427.

4. de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA. Commentary: Middle east respiratory syndrome coronavirus (mers-cov): announcement of the coronavirus study group. J Virol Am Soc Microbiol; 2013;87(14):7790–7792.

5. Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernández R, Lara-Puente JH, Castro-Peralta F, Cunha LF, Trovão NS, Lozano-Dubernard B. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife eLife Sciences Publications Limited; 2016;5:e16777.

6. Khan K, Arino J, Hu W, Raposo P, Sears J, Calderon F, Heidebrecht C, Macdonald M, Liauw J, Chan A. Spread of a novel influenza A (H1N1) virus via global airline transportation. N Engl J Med Mass Medical Soc; 2009;361(2):212–214.

7. Hancock PA, Rehman Y, Hall IM, Edeghere O, Danon L, House TA, Keeling MJ. Strategies for controlling non-transmissible infection outbreaks using a large human movement data set. PLOS Comput Biol Public Library of Science; 2014;10(9):e1003809.

8. Giordano G, Blanchini F, Bruno R, Colaneri P, Di Filippo A, Di Matteo A, Colaneri M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. Nat Med Nature Publishing Group; 2020;1–6.

9. Kraemer MU, Yang C-H, Gutierrez B, Wu C-H, Klein B, Pigott DM, Du Plessis L, Faria NR, Li R, Hanage WP. The effect of human mobility and control measures on the COVID-19 epidemic in China. Science American Association for the Advancement of Science; 2020;368(6490):493–497.

10. Thu TPB, Ngoc PNH, Hai NM. Effect of the social distancing measures on the spread of COVID-19 in 10 highly infected countries. Sci Total Environ Elsevier; 2020;742:140430.

11. Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world. Proc Natl Acad Sci National Acad Sciences; 2004;101(42):15124–15129.

12. Brockmann D, David V, Gallardo AM. Human mobility and spatial disease dynamics. Rev Nonlinear Dyn Complex Wiley Online Library; 2009;2:1–24.

13. Balcan D, Colizza V, Gonçalves B, Hu H, Ramasco JJ, Vespignani A. Multiscale mobility networks and the spatial spreading of infectious diseases. Proc Natl Acad Sci National Acad Sciences; 2009;106(51):21484–21489.

14. Wesolowski A, Eagle N, Tatem AJ, Smith DL, Noor AM, Snow RW, Buckee CO. Quantifying the impact of human mobility on malaria. Science American Association for the Advancement of Science; 2012;338(6104):267–270.

15. Finger F, Genolet T, Mari L, de Magny GC, Manga NM, Rinaldo A, Bertuzzo E. Mobile phone data highlights the role of mass gatherings in the spreading of cholera outbreaks. Proc Natl Acad Sci National Acad Sciences; 2016;113(23):6421–6426.

16. Tizzoni M, Bajardi P, Decuyper A, King GKK, Schneider CM, Blondel V, Smoreda Z, González MC, Colizza V. On the use of human mobility proxies for modeling epidemics. PLoS Comput Biol Public Library of Science; 2014;10(7):e1003716.

17. Wesolowski A, Buckee CO, Engø-Monsen K, Metcalf CJE. Connecting mobility to infectious diseases: the promise and limits of mobile phone data. J Infect Dis Oxford University Press; 2016;214(suppl_4):S414–S420.

18. Dalziel BD, Pourbohloul B, Ellner SP. Human mobility patterns predict divergent epidemic dynamics among cities. Proc R Soc B Biol Sci The Royal Society; 2013;280(1766):20130763.

19. Hawelka B, Sitko I, Beinat E, Sobolevsky S, Kazakopoulos P, Ratti C. Geo-located Twitter as proxy for global mobility patterns. Cartogr Geogr Inf Sci Taylor & Francis; 2014;41(3):260–271.

20. Jurdak R, Zhao K, Liu J, AbouJaoude M, Cameron M, Newth D. Understanding human mobility from Twitter. PloS One Public Library of Science San Francisco, CA USA; 2015;10(7):e0131469.

21. Hu L, Li Z, Ye X. Delineating and modeling activity space using geotagged social media data. Cartogr Geogr Inf Sci Taylor & Francis; 2020;47(3):277–288.

22. Sathya D, Sudha V, Jagadeesan D. Application of Machine Learning Techniques in Healthcare. Handb Res Appl Implement Mach Learn Tech IGI Global; 2020. p. 289–304.

23. Sibanda W, Pretorius P. Artificial neural networks-a review of applications of neural networks in the modeling of HIV epidemic. Int J Comput Appl 2012;44(16):1–9.

24. Hu H, Wang H, Wang F, Langley D, Avram A, Liu M. Prediction of influenza-like illness based on the improved artificial tree algorithm and artificial neural network. Sci Rep Nature Publishing Group; 2018;8(1):1–8.

25. Bai Y, Jin Z. Prediction of SARS epidemic by BP neural networks with online prediction strategy. Chaos Solitons Fractals Elsevier; 2005;26(2):559–569.

26. Guo D, Chen J, MacEachren AM, Liao K. A visualization system for space-time and multivariate patterns (vis-stamp). IEEE Trans Vis Comput Graph IEEE; 2006;12(6):1461–1474.

27. Yang L, Kwan M-P, Pan X, Wan B, Zhou S. Scalable space-time trajectory cube for path-finding: A study using big taxi trajectory data. Transp Res Part B Methodol Elsevier; 2017;101:1–27.

28. Martín Y, Cutter SL, Li Z, Emrich CT, Mitchell JT. Using geotagged tweets to track population movements to and from Puerto Rico after Hurricane Maria. Popul Environ Springer; 2020;1–24.

29. Martín Y, Li Z, Cutter SL. Leveraging Twitter to gauge evacuation compliance: Spatiotemporal analysis of Hurricane Matthew. PLoS One Public Library of Science San Francisco, CA USA; 2017;12(7):e0181701.

30. Jiang Y, Li Z, Cutter SL. Social network, activity space, sentiment, and evacuation: what can social media tell us? Ann Am Assoc Geogr Taylor & Francis; 2019;109(6):1795–1810.

31. Hu F, Li Z, Yang C, Jiang Y. A graph-based approach to detecting tourist movement patterns using social media data. Cartogr Geogr Inf Sci Taylor & Francis; 2019;46(4):368–382.

32. Li L, Goodchild MF, Xu B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. Cartogr Geogr Inf Sci Taylor & Francis; 2013;40(2):61–77.

33. Malik MM, Lamba H, Nakos C, Pfeffer J. Population bias in geotagged tweets. People 2015;1(3,759.710):3–759.

34. Jiang Y, Li Z, Ye X. Understanding demographic and socioeconomic biases of geotagged Twitter users at the county level. Cartogr Geogr Inf Sci Taylor & Francis; 2019;46(3):228–242.

35. Huang X, Li Z, Jiang Y, Ye X, Deng C, Zhang J, Li X. The characteristics of multi-source mobility datasets and how they reveal the luxury nature of social distancing in the US during the COVID-19 pandemic. medRxiv Cold Spring Harbor Laboratory Press; 2020;

36. Lewnard JA, Lo NC. Scientific and ethical basis for social-distancing interventions against COVID-19. Lancet Infect Dis Elsevier; 2020;20(6):631.

37. Evans M. Avoiding COVID-19: Aerosol Guidelines. ArXiv Prepr ArXiv200510988 2020;

38. World Health Organization. Modes of transmission of virus causing COVID-19: implications for IPC precaution recommendations: scientific brief, 27 March 2020. World Health Organization; 2020.

39. Thunström L, Newbold SC, Finnoff D, Ashworth M, Shogren JF. The benefits and costs of using social distancing to flatten the curve for COVID-19. J Benefit-Cost Anal Cambridge University Press; 2020;1–27.

40. Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures In The United States Reduced The COVID-19 Growth Rate: Study evaluates the impact of social distancing measures on the growth rate of confirmed COVID-19 cases across the United States. Health Aff (Millwood) 2020;10–1377.