# An empirical study of the multi-scale spatiotemporal Twitter user mobility patterns

**Abstract**

Understanding human mobility patterns is of great importance for urban planning, traffic management, and even marketing campaign. However, the capability of capturing detailed human movements with fine-grained spatial and temporal granularity is still limited. In this study, we extracted high-resolution mobility data from a collection of over 1.3 billion geo-located tweets. Regarding the concerns of infringement on individual privacy, such as the mobile phone call records with restricted access, the dataset is collected from publicly accessible Twitter data streams. In this paper, we performed an empirical study of the multi-scale spatiotemporal Twitter user mobility patterns in the contiguous United States during the year 2014.

Our approach included a scalable visual-analytics framework to deliver efficiency and scalability in filtering large volume of geo-located tweets, modeling and generating space-time user trajectories, and summarizing multi-scale spatiotemporal user mobility patterns.

We performed a set of statistical analysis to understand Twitter user mobility patterns across multi-level spatial scales and temporal ranges. In particular, Twitter user mobility patterns measured by the displacements and radius of gyrations of individuals revealed multi-scale or multi-modal Twitter user mobility patterns. By further studying such mobility patterns in different temporal ranges, we identified both consistency and seasonal fluctuations regarding the distance decay effects in the corresponding mobility patterns.

*Key words:* Geo-located tweets, mobility patterns, multi-scale spatiotemporal analysis, distance decay

## 1 Introduction

Understanding human mobility patterns is of great importance for a broad range of applications from urban planning [1], traffic management [2], and even the spatial spread of epidemic diseases [3]. Earlier research efforts relied

on low-resolution mobility data to understand human mobility patterns, such as using census records to study human migration patterns [4], or delivering questionnaires and asking volunteers to report the track of bank notes to infer human travel patterns [5]. However, such mobility data do not provide detailed human movements with fine-grained spatial and temporal granularity, which are usually aggregated and therefore are limited to capture mobility patterns of individuals [6,7]. In addition to the mobility data collected by GPS trackers [1,8] and mobile phone call records [6,9,10], emerging as a new mobility data source, today's pervasive Location Based Social Media (LBSM) platforms (e.g., Twitter and Foursquare) offer continuous spatial Big Data streams with massive amount of detailed and frequently updated user digital footprints in the form of real-world user trails and footprints [11]. A significant advantage of utilizing LBSM data streams as proxies for studying human mobility patterns is the large spatial coverage. For instance, researchers have used geo-located tweets for studying global mobility patterns [12], which is otherwise impossible for other mobility datasets (e.g., GPS traces and mobile phone call records). Regarding the concerns of infringement on individual privacy, such as the mobile phone call records with restricted access [13,14,7], the publicly available LBSM data streams offer unique opportunities for conducting reproducible and comparative scientific findings across different geographical regions.

Many recent studies have adopted LBSM data streams to study human mobility patterns. For example, they modeled and extracted trajectories of individuals and performed statistical analysis focusing on the distance decay effects in the collective user movements [6], which were used to reveal different travel modes [7], travel demands [15,16], and the impact of social connections [17]. These studies have provided strong supports for using LBSM data as proxies for studying mobility patterns of individuals and valuable insights into human mobility dynamics. However, the variations of movements in different spatial scales and temporal ranges are neglected in these studies, where the measurements of distances are either fixed in a certain time range or to a specific geographic region. For instances, the examinations on whether there are temporal (e.g., monthly or seasonal) changes within the movements or how the observed mobility patterns vary across different spatial scales (e.g., intra- or inter city or national levels) are lacking. Such insights are critical to advance our understandings of the collective mobility patterns for various applications, such as examining the mobility patterns across different cities [18], the spread patterns of disease [19,20] and touristic activities [12]. On the other hand, while the high-resolution spatiotemporal records from LBSM present unique research opportunities in this direction, the inherited large data volume poses significant data-intensive challenges for developing multi-scale spatiotemporal analysis approaches to dealing with the complexities in filtering movements of individuals, modeling and aggregating user trajectories at multiple spatial and temporal scales [21].

2

In this paper, we have employed a visual-analytics approach to exploring the Twitter user mobility patterns across multi-level spatial scales and temporal ranges in the continuous United States (i.e., excluding Alaska and Hawaii) during the year 2014. The mobility data is extracted from over 1.3 billion geo-located Twitter messages (i.e., tweets) from January 1 to December 31, 2014 over North America with over 6 million Twitter users and over 1 TB in file size. To address the data-intensive challenge embedded in this dataset, we have developed a scalable visual-analytics framework tailored to accommodate large volume of geo-located tweets. This framework is implemented based on high-performance distributed computing environment using Apache Hadoop (http://hadoop.apache.org/), which is an open source software framework to enable distributed processing of large datasets across computing clusters. Enabled by this framework, we have performed a set of statistical analysis to understand multi-scale spatiotemporal Twitter user mobility patterns. We have modeled the frequency of Twitter users visiting different locations to study the collective user visiting behaviors, where we have identified temporal similarities in the distributions. In particular, Twitter user mobility patterns measured by user displacements and radius of gyrations of individuals [6] have revealed different groups of Twitter users with multi-scale or multi-modal mobility patterns and multiple travel modes [7]. By further studying such mobility patterns in different temporal ranges, we have identified both consistency and seasonal fluctuations regarding the distance decay effects in the corresponding mobility patterns. In addition, our approach provides an interactive 3D virtual globe web mapping interface to enable exploratory geo-visual analytics for understanding the detailed Twitter user movement flows within a given spatial scale and time window.

The remainder of this paper is organized as follows. Section 2 describes the related work in the context of studying mobility patterns using LBSM data, in particular, the geo-located Twitter data. We focus on research challenges in using visual-analytics methods to enable multi-scale spatiotemporal analysis with massive movement datasets, including data management, multi-level spatiotemporal user trajectory modeling and visualization. Section 3 details the processes for extracting, aggregating and summarizing multi-level spatiotemporal Twitter user mobility patterns. Section 4 presents the case study of performing visual-analytics for seeking multi-scale spatiotemporal Twitter mobility patterns in the continuous United States of the year 2014. Section 5 concludes the paper.

## 2 Mobility Patterns in Location Based Social Media data

### 2.1 *Geo-located Twitter Data for Studying Large-scale User Movements*

To understand detailed mobility patterns of individuals, the capability to capture human movements with fine-grained spatial and temporal granularity is critical. In this connection, using GPS trackers tends to produce, to date, the most accurate records of individuals' movements regarding the accuracy of recorded user locations and update frequency [1]. However, such data are often limited in spatial scale (e.g., within a specific city or region) from a small group of people, for example, 226 and 182 volunteers participated in collecting such mobility data in [8] and [22] respectively. Other than tracking people directly, the vehicle-based GPS traces are often tied to specific vehicles (e.g. taxi), which are only accessible for a certain group of people [10].

Another approach from the literatures for studying human mobility is using mobile phone call data, such as Call Detail Records (CDR), where the locations of mobile users are estimated by cell tower triangulation with an accuracy in the order of kilometers [6,9,10]. Such a dataset can cover relatively large spatial scale [23,24] (e.g., national level) and a large portion of the population in the study region [10]. However, due to the concerns of infringement on individual privacy, mobile phone call data are not publicly accessible at all. Even such data were obtained in the mentioned studies, they came from various service providers covering different groups of users. These issues limit the capability for conducting reproducible scientific findings for mobility research, such as validating or extending the existing discoveries.

In this connection, it becomes increasingly popular for researchers to exploit the publicly accessible mobility data captured from today's pervasive Location Based Social Media (LBSM) platforms (e.g., Foursquare and Twitter). LBSM enables users to attach their current location as a geo-tag to the message they post, which is derived from either the GPS or Wi-Fi positioning with a high position resolution down to 10 meters [7]. A Big Data scenario emerges when millions social media users constantly post messages. In this study, geo-located Twitter data are chosen as a source for studying detailed mobility patterns. Compared to other LBSM platforms, Twitter is one of the most popular platforms and is been actively used in many countries. It provides a publicly accessible streaming API for easy data access [25]. Indeed, many other LBSM data can be collected from the data streams, such as Foursquare check-in data [26,16].

However, it is worth noting that there are some limitations and complexities in directly using LBSM data for studying human mobility patterns. For

example, comparing to GPS traces, the update frequency of an individual's location varies depending on when a user is posting a new geo-located message or check-in at a new place. Although geo-located tweets tend to provide geo-locations with high position resolution as aforementioned [7], the information regarding the quality of the geo-locations is absent in each tweet. This will contribute to the uncertainties in calculating the distance of Twitter user movements, especially in densely built environments. There is also a potential mismatch regarding the representativeness of the overall population since not all people use social media or send geo-located messages [10], the demographic information of the Twitter users cannot be easily identified. The derived mobility patterns may lead to an over or under-representation of the real-world human mobility patterns. Many studies started to look into the demographic information of LBSM data, in particular Twitter data [27,28]. Although the used methods are still arguable, these issues certainly require us to pose stricter criteria in understanding human mobility patterns using geo-located Twitter data. On the other hand, geo-located Twitter dataset presents some unique advantages that make it a valuable proxy for studying human mobility patterns. For example, the high-resolution location information enables to identify multiple travel modes in user mobility patterns [7]; the large spatial coverage enables to study global mobility patterns [12], which is almost impossible for other mobility datasets. More importantly, by continuously monitoring the geo-located Twitter data streams with large volume of detailed and frequently updated spatiotemporal records of Twitter users, it offers a great deal of potential for studying mobility patterns of large groups of individuals at different spatial scales (e.g., movements across cities, states or even countries) and temporal gratuity (e.g., weekly, monthly, and seasonal movements), which is one of the motivations for this study.

# References

[1] Zheng, Y., Li, Q., Chen, Y., Xie, X., Ma, W.Y., 2008. Understanding mobility based on GPS data. Proceedings of the 10th international conference on Ubiquitous computing. ACM, pp. 312–321.

[2] Jiang, B., Yin, J., Zhao, S., 2009. Characterizing the human mobility pattern in a large street network. *Physical Review E, 80*, 021136.

[3] Belik, V., Geisel, T., Brockmann, D., 2011. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X, 1*, 011001.

[4] Greenwood, M.J., 1985. Human migration: Theory, models, and empirical studies. *Journal of regional Science, 25*, pp. 521–544.

[5] Brockmann, D., Hufnagel, L., Geisel, T., 2006. The scaling laws of human travel. *Nature, 439*, pp. 462–465.

[6] Gonzalez, M.C., Hidalgo, C.A., Barabasi, A.L., 2008. Understanding individual human mobility patterns. *Nature*, *453*, pp. 779–782.

[7] Jurdak, R., Zhao, K., Liu, J., AbouJaoude, M., Cameron, M., Newth, D., 2015. Understanding Human Mobility from Twitter. *PLoS ONE*, *10*, e0131469.

[8] Rhee, I., Shin, M., Hong, S., Lee, K., Kim, S.J., Chong, S., 2011. On the levy-walk nature of human mobility. *IEEE/ACM transactions on networking (TON)*, *19*, pp. 630–643.

[9] Sevtsuk, A., Ratti, C., 2010. Does urban mobility have a daily routine? Learning from the aggregate data of mobile networks. *Journal of Urban Technology*, *17*, pp. 41–60.

[10] Kung, K.S., Greco, K., Sobolevsky, S., Ratti, C. 2014. Exploring universal patterns in human home-work commuting from mobile phone data. *PLoS ONE*, *9*, e96180.

[11] Thatcher, J., 2014. Living on fumes: Digital footprints, data fumes, and the limitations of spatial big data. *International Journal of Communication*, *8*, pp. 1765–1783.

[12] Hawelka, B., Sitko, I., Beinat, E., Sobolevsky, S., Kazakopoulos, P., Ratti, C., 2014. Geo-located Twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, *41*, pp. 260–271.

[13] Giannotti, F., Pedreschi, D., 2008. *Mobility, data mining and privacy: Geographic knowledge discovery*, Springer Science & Business Media.

[14] Crampton, J.W. 2014. Collect it all: national security, Big Data and governance. *GeoJournal*, pp. 1–13.

[15] Wu, L., Zhi, Y., Sui, Z., Liu, Y., 2014. Intra-urban human mobility and activity transition: Evidence from social media check-in data. *PLoS ONE*, *9*, e97010.

[16] Hasan, S., Zhan, X., Ukkusuri, S.V., 2013. Understanding urban human activity and mobility patterns using large-scale location-based data from online social media. Proceedings of the 2nd ACM SIGKDD international workshop on urban computing. ACM, 2013, p. 6.

[17] Cho, E., Myers, S.A., Leskovec, J. 2011, Friendship and mobility: user movement in location-based social networks. Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011, pp. 1082–1090.

[18] Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C., 2012. A tale of many cities: universal patterns in human urban mobility. *PLoS ONE*, *7*, e37027.

[19] Balcan, D., Colizza, V., Gonçalves, B., Hu, H., Ramasco, J.J., Vespignani, A., 2009. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the National Academy of Sciences*, *106*, pp. 21484–21489.

[20] Tamerius, J., Nelson, M.I., Zhou, S.Z., Viboud, C., Miller, M.A., Alonso, W.J., 2011. Global influenza seasonality: reconciling patterns across temperate and tropical regions.*Environmental health perspectives*, *119*, 439.

[21] Tsou, M.H., 2015. Research challenges and opportunities in mapping social media and Big Data.*Cartography and Geographic Information Science* **2015**, *42*, pp. 70–74.

[22] Zheng, Y., Xie, X., Ma, W.Y., 2010. GeoLife: A Collaborative Social Networking Service among User, Location and Trajectory.

[23] Becker, R., Cáceres, R., Hanson, K., Isaacman, S., Loh, J.M., Martonosi, M., Rowland, J., Urbanek, S., Varshavsky, A., Volinsky, C., 2013. Human mobility characterization from cellular network data.*Communications of the ACM*, *56*, pp. 74–82.

[24] Sobolevsky, S., Szell, M., Campari, R., Couronné, T., Smoreda, Z., Ratti, C., 2013. Delineating geographical regions with networks of human interactions in an extensive set of countries.*PLoS ONE*, *8*, e81707.

[25] Twitter, 2016. Twitter streaming API.*Available from: https://dev.twitter.com/streaming/overview.*

[26] Cranshaw, J., Schwartz, R., Hong, J.I., Sadeh, N.M., 2012. The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. ICWSM.

[27] Mitchell, L., Frank, M.R., Harris, K.D., Dodds, P.S., Danforth, C.M., 2013. The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place.

[28] Longley, P.A., Adnan, M., Lansley, G., others., 2015. The geotemporal demographics of Twitter usage. *Environment and Planning A*, *47*, pp. 465–484.

[29] Hägerstrand, T., others., 1985. Time-geography: focus on the corporeality of man, society, and environment. *The science and praxis of complexity*, pp. 193–216.

[30] Kwan, M.P., Lee, J., 2004. Geovisualization of human activity patterns using 3D GIS: a time-geographic approach. *Spatially integrated social science*, *27*.

[31] Andrienko, N., Andrienko, G., 2007. Designing visual analytics methods for massive collections of movement data. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *42*, pp. 117–138.

[32] MacEachren, A.M., Kraak, M.J., 2001. Research challenges in geovisualization. *Cartography and Geographic Information Science*, *28*, pp. 3–12.

[33] MacEachren, A.M., 2004. *How maps work: representation, visualization, and design*, Guilford Press.

[34] Andrienko, G., Andrienko, N., Wrobel, S., 2007. Visual analytics tools for analysis of movement data. *ACM SIGKDD Explorations Newsletter*, *9*, pp. 38–46.

[35] Black, A., Mascaro, C., Gallagher, M., Goggins, S.P., 2012. Twitter zombie: Architecture for capturing, socially transforming and analyzing the Twittersphere. Proceedings of the 17th ACM international conference on Supporting group work. ACM, pp. 229–238.

[36] Shvachko, K., Kuang, H., Radia, S., Chansler, R., 2010. The hadoop distributed file system. Mass Storage Systems and Technologies (MSST), 2010 IEEE 26th Symposium on. IEEE, 2010, pp. 1–10.

[37] Dean, J., Ghemawat, S., 2008. MapReduce: simplified data processing on large clusters. *Communications of the ACM*, *51*, pp. 107–113.

[38] Gao, H., Tang, J., Liu, H., 2012. Exploring Social-Historical Ties on Location-Based Social Networks. ICWSM.

[39] Buttenfield, B.P., McMaster, R.B., 1991. *Map Generalization: Making rules for knowledge representation*, Longman Scientific & Technical New York.

[40] Samet, H., 1984. The quadtree and related hierarchical data structures. *ACM Computing Surveys (CSUR)*, *16*, pp. 187–260.

[41] Clauset, A., Shalizi, C.R., Newman, M.E., 2009. Power-law distributions in empirical data. *SIAM review*, *51*, pp. 661–703.

[42] Reynolds, A., 2012. Truncated Lévy walks are expected beyond the scale of data collection when correlated random walks embody observed movement patterns. *Journal of The Royal Society Interface*, *9*, pp. 528–534.

[43] Zhao, K., Musolesi, M., Hui, P., Rao, W., Tarkoma, S., 2015. Explaining the power-law distribution of human mobility through transportation modality decomposition. *Scientific reports*, *5*.