Title: Measuring Variability of Mobility Patterns from Multiday Smart-card Data

Author: Chen Zhong Ed Manley Stefan Muller Arisona Michael Batty Gerhard Schmitt

PII: S1877-7503(15)00059-9
DOI: http://dx.doi.org/doi:10.1016/j.jocs.2015.04.021
Reference: JOCS 364

To appear in:

Measuring Variability of Mobility Patterns from Multiday Smart-card Data

Chen Zhong[1, 2], Ed Manley[1], Stefan Muller Arisona[2, 3], Michael Batty[1], and Gerhard Schmitt[2]

[1]*Central for Advanced Spatial Analysis, University College London, U.K.*
[2]*Future Cities Laboratory, Architecture department, ETH Zurich, Switzerland.*
[3] *Institute of 4D Technologies, FHNW, Switzerland*
{c.zhong, ed.manley, m.batty}@ucl.ac.uk.com, stefan.arisona@fhnw.ch, schmitt@arch.ethz.ch

Abstract

The availability of large amounts of mobility data has stimulated the research in discovering patterns and understanding regularities. Comparatively, less attention has been paid to the study of variability, which, however, has been argued as equally important as regularities, since variability identifies diversity. In a transport network, variability exists from person to person, from place to place, and from day to day. In this paper, we present a set of measuring of variability at individual and aggregated levels using multi-day smart-card data. Statistical analysis, correlation matrix and network-based clustering methods are applied and potential use of measured results for urban applications are also discussed. We take Singapore as a case study and use one-week smart-card data for analysis. An interesting finding is that though the number of trips and mobility patterns varies from day to day, the overall spatial structure of urban movement always remains the same throughout a week. This finding showed that a systemic framework with well-organized analytical methods is indeed necessary for extracting variability that may change at different levels and consequently for uncovering different aspects of dynamics, namely transit, social and urban dynamics. We consider this paper as a tentative work towards such generic framework for measuring variability and it can be used as a reference for other research work in such a direction.

Keywords: Variability, smart-card data, spatial analysis, clustering, network

1 Introduction

There is a rapid rise of interests in data-related research in recent years, mostly because that never before has it been so easy to collect such cheap and large amounts of mobility data in fairly good spatiotemporal resolution. Research that depended on simulated data for discovery universal human mobility patterns (Szell et al., 2012), can now use real-world data as alternatives due to abundant large data sets generated by human agents. The human mobility data enables us to have a closer and more direct look into spatiotemporal patterns of human mobility (Agard et al., 2006; Liang et al., 2009; Zhong, Huang, et al., 2014), and regularities and/or scaling laws (Gonzalez et al., 2008; Schneider et al., 2013; Simini et al., 2012; Song et al., 2010).

Equally important as regularities is the variability, which however, has received much less attention comparatively. Jones and Clarke (1988) once discussed the significance of variability. In their paper, they measured the multi-day variability of travel behaviors and considered that the insights of the measuring could improve the modeling of travel behavior and the assessment of policy impacts. Similar work is conducted in (Ahas et al., 2010), in which, temporal variability of mobility are analyzed to better understand urban life. “The main distinction can be drawn between work days and weekend days. There are also more specific movement characteristics for each day, such as the impact of weekend activities on Friday and Monday; Saturday is influenced by the impacts of working days or by the shopping day mentality”. More related work looking into daily rhythms can be found in Axhausen’s work (Axhausen et al., 2002; Schönfelder & Axhausen, 2010). Early works on measuring variability were quite often limited by unsufficient samples of data, which is no longer the case nowadays.

This paper considers variability as an important factor that identifies diversity in reality. More specifically, regularity can be used to find clusters and variability tells what the clusters are by distinguished features. Moreover, an irregularity may help us to identify critical points or factors that trigger an event. In a transport network, variability exists from day to day, from person to person, and from place to place. The way to measure variability depends on the granularity of data, the richness of information and the specific questions we aim to answer. It is more reasonable to propose a generic framework, which is our final goal, rather than a limited set of fixed indices.

Based on such thinking, this paper is a tentative work that measures variability of urban mobility patterns across different levels, namely individual, aggregated and overall from smart-card data which is a newly available type of human mobility data. The level of details is the core part in building a generic framework. The specific contributions of this work are twofold, firstly, we give a summary of typical analytical methods that are suitable for measuring variability at different levels for certain urban applications. The analytical methods are classical ones which have long been proved useful and stable in the field of computer science but been applied in the field of urban computing only recently. Such summary could be references to other research work in this direction. Secondly, examples of measuring variability by the proposed methods using smart-card data in Singapore are presented. It is more than a demonstration of the methodology, because using insights about transit, social and urban dynamics of Singapore is common practice by the planning agencies nowadays. The rest of the paper is organized by the following sequences: in section 2, we describe the data sets and preliminary data processing; in section 3, we present results of analysis at three levels; section 4 concludes the paper.

2 Study Materials

Singapore is an island city-state in Southeast Asia with an area of 710.2 km2 as shown in Figure 1. The current population of Singapore including non-residents is approximately 5.4 million. The built public transportation system has greatly contributed to the urban development and helps shape the city. Nowadays, about half of the population in Singapore are using public transportation daily, generating more than 5 million travel records per day. At the time that we conducted this research, there are about 4770 bus stops and Mass Rapid Transit (MRT) train stations in use, as detected from our data sets, covering the whole land of Singapore geographically.

This paper uses one-week smart-card data collected in April 2014. The smart card data were collected by an automatic fare collection system. While the main purpose of such system is to collect fare, they also produce large quantities of records on daily travelling (Pelletier et al., 2009). The smart card data used in this paper record each ride as a pair of tap in and tap out events. Each record (a ride) contains attributes including trip id, passenger id, card types, boarding and alighting time, boarding and alighting location, distance, fare, and a transfer index associated with transferring to next rides.

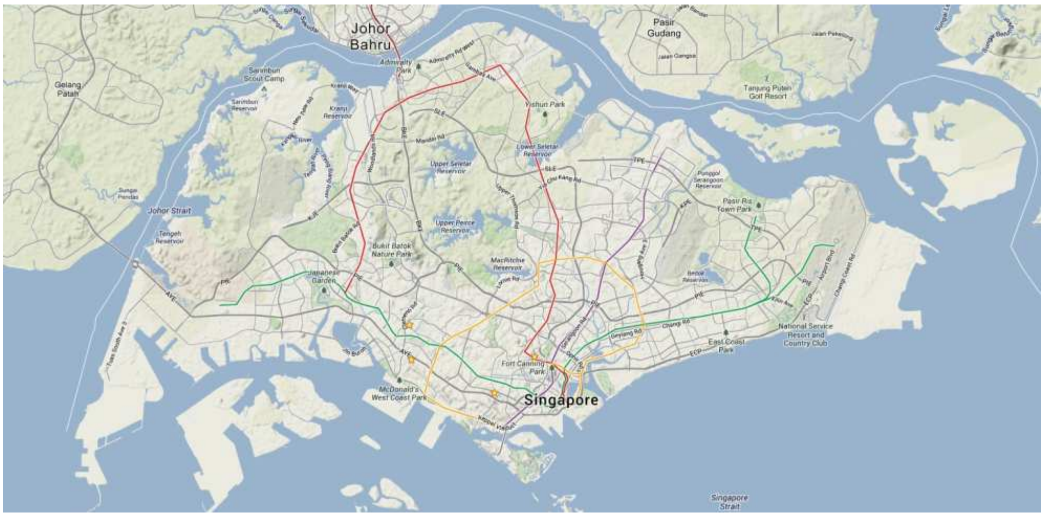


Figure 1. Case study area: Singapore. Image from google maps, 2014

Preliminary data processing is conducted to format the data into certain structures for later analysis. As indicated, the original records are rides, which are segments of a trip. In the first step, we construct trips from rides by sorting them in time sequences using attributes, including journey id, card id, alighting time, alighting location, boarding time, boarding location, and transfer index. In the second step, we derive an Origin-Destination (O-D) matrix by counting numbers of trips between locations. In the third step, a spatial network is constructed from the O-D matrix, taking locations as nodes, trips between locations as edges and number of trips as weight of edges. The total number of rides, trips, and edges in three data structures are shown in Table 1.

3

4

5 Measuring the Variability

In this section, we demonstrate different methods of measuring variability using processed smart-card data. The methods are organized by different levels of details, looking into individual behavior from trips, collective effects from data aggregated by spatial locations and overall structure of urban movement from a network system. We give only intuitive interpretations of the analyzed results, rather than rigorous scientific verifications, since the emphasis of this paper is the way of measuring and the potential use of measured results in urban applications.

5.1 Variability of Individual Mobility Patterns

Spatiotemporal analysis of mobility patterns is always the first step to go with transportation data. Through simple statistics, temporal mobility patterns can be easily detected. Figure 2 shows a profile of trip starting-time during one week with y-axis indicating the number of trips per hours and x-axis as a timeline (by hour). From this profile, peak hours can be identified and the difference between weekdays and weekend are clearly shown. The morning and evening peak disappears on weekends and an even more interesting observation is the small peak emerging on Friday lunchtime. Some local context can also be found in this simple profile. There is a small morning peak appeared on Saturday morning, mostly because many companies in Singapore have only half day-off on Saturdays.

Figure 3 is a correlation matrix of the temporal patterns of each day. X and Y-axis are index of different days in a week. We take the temporal profile of trip starting time in Figure 2 as attributes of a day. The color map clearly tells the similarity of daily mobility patterns. The red color means higher correlations, indicating that people have very similar travel behaviors on these two days. In reverse, blue color means dissimilarity. The correlation value changes gradually from left to right and from Friday to Sunday indicating that lifestyles are gradually changing from working mode to leisure mode. Such matrix could also be generated with a more refined temporal resolution, for instance using hours instead of days as a temporal unit to observe daily activities patterns.

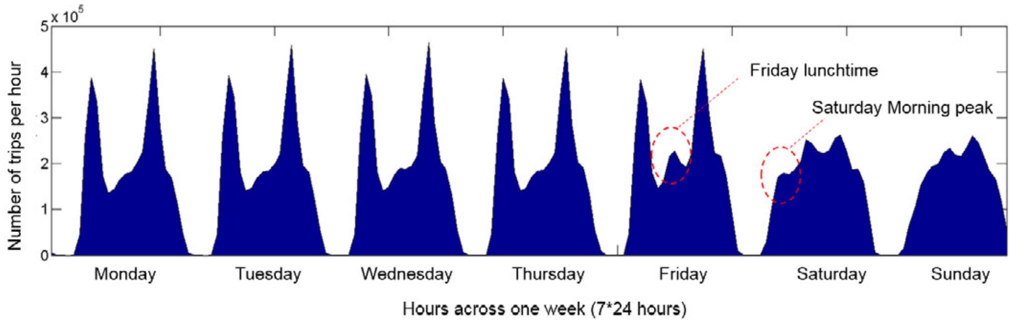


Figure 2. Number of trips during 7*24 hours.

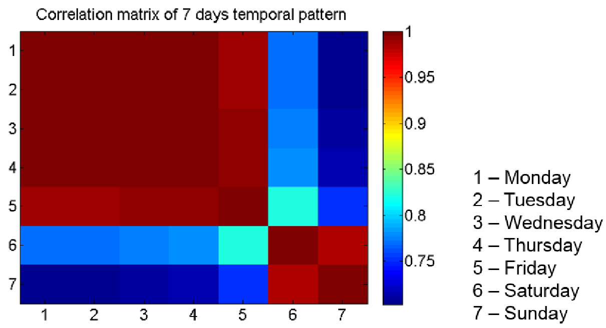


Figure 3. Correlation matrix of temporal patterns of days during a week.

5.2 Variability of Aggregated Mobility Patterns

Aggregating trips by locations of bus stops and train stations helps project numerical trip records into geographic space that variability from place to place can be identified. Figure 4 shows profiles of aggregated temporal mobility patterns at three randomly selected bus stops. All of them shows a unique pattern that differentiates them from each other. The most straightforward reason causing such variability could be mixed-land use in surrounding areas as indicated in related work (Hasan et al., 2013; Zhong et al., 2013). The stop that has only one morning peak is most likely to be surrounded by residential areas. Stops with multiple peaks during a day and similar patterns throughout an entire week are likely within mixed-use areas that provide diverse services to attract people coming at any time of a day. Stops with a significant peak occurred at around 6PM are most likely to be working places. A normal way of verification is to group places with similar patterns and verify our prediction of land uses to ground truth.

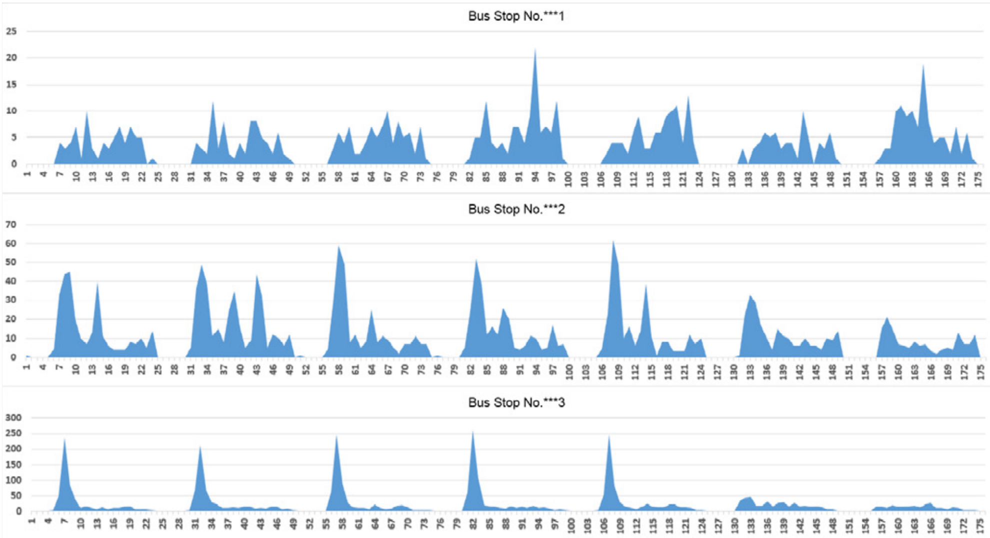


Figure 4. Profile of temporal patterns of trip starting time at three selected bus stops.（x-axis hours of one week, y-axis number of people in a hour）

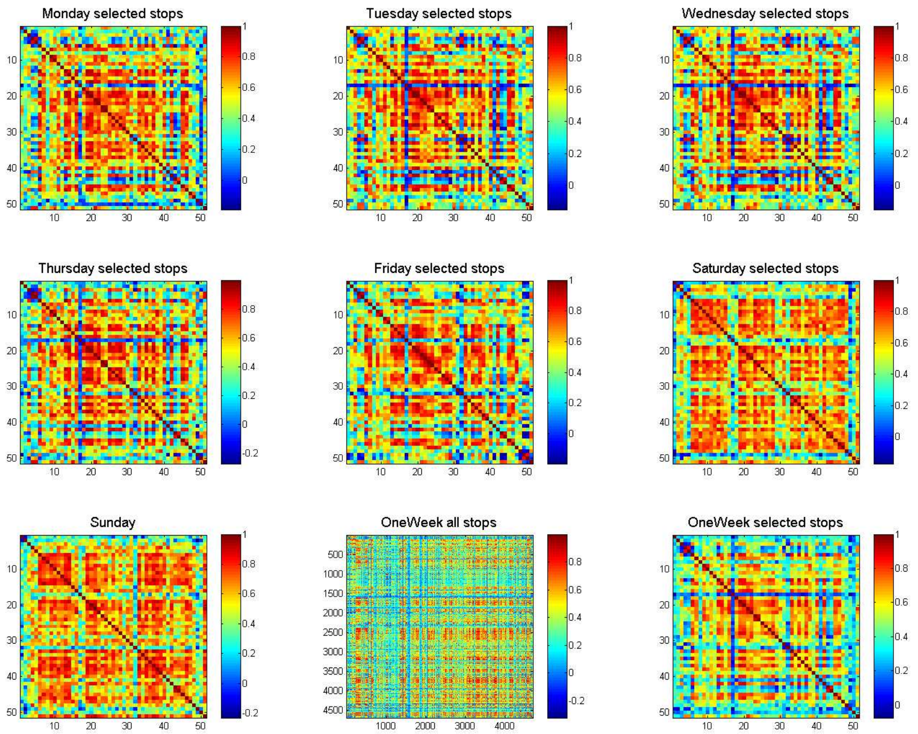


Figure 5. Correlation matrix of stops of their temporal pattern profile. x and y axis are index of stops.

Similar to Figure 3, Figure 5 is a set of correlation matrix of the aggregated temporal patterns of each stops or stations. X and Y-axis are ids of stops and stations in our case. We treat the profile of temporal patterns across one day or one week as attributes of a stop. The color map tells the value of correlations. The red color indicates higher correlation of two stops/stations, which might be grouped together. Based on our previous assumption about the impact of land uses on travel time, the stops, which have high correlations between each other, are the ones have similar land use types. With a one-day correlation matrix, we are able to find locations with similar land uses. The essential idea is quite similar to related papers using K-means or self-organizing map (SOM) to cluster locations (Soto & Frías-Martínez, 2011). Furthermore, with variations of correlation values during one-week, we can found the changes and therefore tell more information about different components of mixed land use, the dynamic urban functions, furthermore, to identify some interesting places. As you easily found from any plot in Figure 5, there are always clusters of continuous grids with very high correlation value. The stops that have very closed ids are also near to each other geographically due to the coding system. They tend to form a neighborhood with very similar land uses. They represent important locations and together show spatial structure of a city.

5.3 Variability of Spatial Networks: The City as a System

At this level of measuring, we look at the overall urban movement. We consider a city as a system and areas within a city are linked by urban flows. In the context of daily mobility, the urban flows are represented by trips between areas. A spatial network is then constructed from an O-D matrix of daily trips as we previously introduced in Section 2. Actually, the network analysis approach has been widely applied and proved useful with various types of flow data including mobile phone calls records, airlines and money flow (Barthélemy, 2011; Ratti et al., 2010; Thiemann et al., 2010; Zhong, Arisona, et al., 2014).

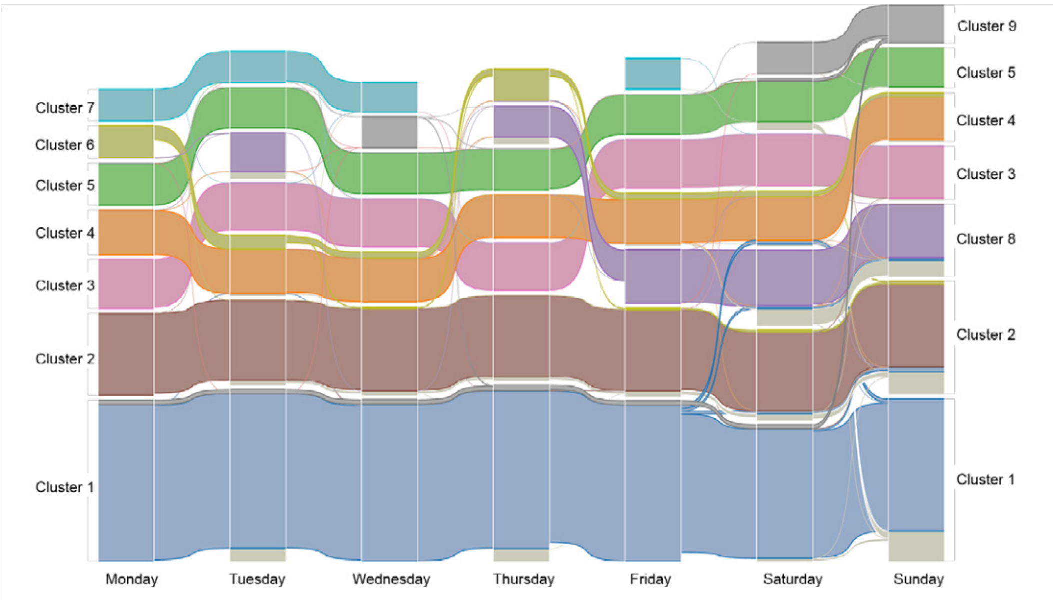
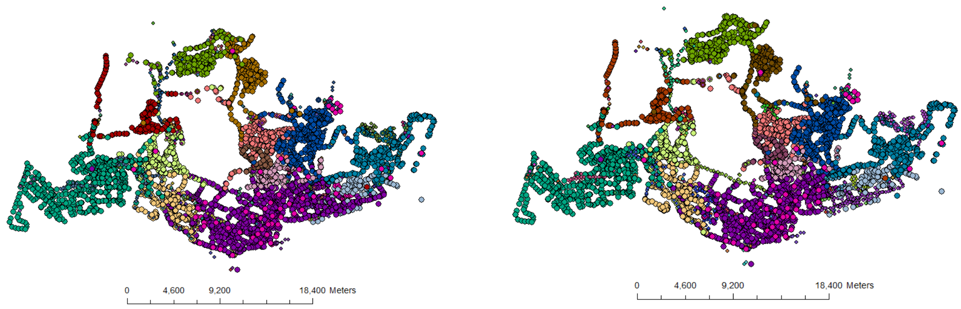


Figure 6. Results of Community detection and ranking of Page Rank values from Monday to Sunday. The exchanging flows indicates that areas function differently throughout a week.

In Figure 6, we show results of community detection and PageRank centrality as examples out of many possible centrality measuring. Community detection was often used in social network analysis to find closed linked social ties, here is used to group intensively interacted stops into a cluster. PageRank, which is originally used to measuring the importance of a web page has been redefined to measure the attractiveness of detected clusters in a city (Zhong, Arisona, et al., 2014). From bottom to up, the rectangles denote detected clusters composed by groups of stops or stations. From left to right are results of seven days. The exchange flows in between shows the changing grouping of clusters as well as changing value of their corresponding values of attractiveness. Local changes can be observed from the changing ranking of clusters and flows. Based on this observation, we can then identify certain special places, which changes their roles in terms of urban functions on different days of a week.

An interesting finding is that though exchanging flows exist locally, the global structure always remains the same, regardless of the slightly different number of trips or changing travel times and travel distance from weekdays to weekends. This observation is further presented by a geographic mapping of clusters shown in Figure 7. Different colors refer to different clusters composed by a group of stops and stations. The points (referring to stops or stations) which have the same color belong to the same cluster. Based on the theory of community detection in complex network analysis, points generate more inter-flows inside their own cluster than intra-flows between other clusters. Interpreted in the context of urban mobility, locations inside a cluster interact more between each other and together they form a neighborhood/local system. Based on this theory, we can partition the urban system into several local systems, a structure of city appear consequently.



Monday **Sunday**

Figure 7. Detected communities using smart-card data on Monday (left) and Sunday (right) shows almost the same spatial partitions of urban space.

6 Conclusion

Analysis of automatically generated data such as smart-card data of daily human mobility activities, gives us a good profile of city and citizen's life-style. Its potential is still under exploration. In this paper, we show that variability can be measured from day to day and location to location from such data set. Rather than proposing completely new methods of measuring, some of the analytical methods we were used are classical ones in data mining and have already been applied in various forms in previous related works. They have been proved useful in given contexts, and in this paper, we organized them by the logic of level of details. More specifically, variability has been measured as an indicator of diverse life styles at individual level, of different components of land uses at aggregated level, and of changing spatial structures of overall urban movement. Taking Singapore as a case study, we measured the (1) variability of temporal patterns by descriptive statistical analysis; (2) variability of correlations between stops by correlation matrix and clustering methods; and (3) variability of spatial network structure by complex network analysis, using multi-day smart-card data. We found that, in the case of Singapore, variability of mobility patterns exist at individual and spatial aggregated scale, however, the overall structure of urban movements remains almost the same. Whether or not it is a universal phenomenon should be further verified using data from other cities.

As indicated, this paper is considered as preliminary work providing directions towards establishing a generic framework with well-organized analytical methods and potential urban applications, because huge potentials of such rich urban data is still under exploration. For instance, using the same clustering method but aggregating data by different attribute could give insight to other aspects. With even finer granularity of mobility data, we can detect changes in even higher temporal resolutions and predict short-term changes. Building linkage among various urban data sets such as census, economy, and cohort enables us to find out causality of urban phenomenon. In the future work, our goal is to further extend the proposed generic framework based on the essential logic of level of details with more advanced analytical methods and to explore extensive potential urban applications. This work will certainly contribute to the theory of smart cities.

Acknowledgement

The authors would like to acknowledge the valuable comments from the anonymous reviewers. This research work was established at the Singapore-ETH Centre (SEC) and the Centre for Advanced Spatial Analysis, and is co-funded by the Singapore National Research Foundation (NRF), ETH Zurich and the European Research Council under 249393-ERC-2009-AdG. The authors would like to express their sincere gratitude to the Singapore Land Transport Authority and Transportation Module in Future Cities Laboratory for supporting this research and providing the required data.

References

Agard, B., Morency, C., & Trépanier, M. (2006). *Mining public transport user behaviour from smart card data.* Paper presented at the 12th IFAC Symposium on Information Control Problems in Manufacturing-INCOM.

Ahas, R., Aasa, A., Silm, S., & Tiru, M. (2010). Daily rhythms of suburban commuters' movements in the Tallinn metropolitan area: Case study with mobile positioning data. *Transportation Research Part C: Emerging Technologies, 18*(1), 45-54.

Axhausen, K. W., Zimmermann, A., Schönfelder, S., Rindsfüser, G., & Haupt, T. (2002). Observing the rhythms of daily life: A six-week travel diary. *Transportation, 29*(2), 95-124.

Barthélemy, M. (2011). Spatial networks. *Physics Reports, 499*(1), 1-101.

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

Hasan, S., Schneider, C., Ukkusuri, S., & González, M. (2013). Spatiotemporal Patterns of Urban Human Mobility. *Journal of Statistical Physics, 151*(1-2), 304-318. doi: 10.1007/s10955-012-0645-0

Jones, P., & Clarke, M. (1988). The significance and measurement of variability in travel behaviour. *Transportation, 15*(1-2), 65-87. doi: 10.1007/BF00167981

Liang, L., Anyang, H., Biderman, A., Ratti, C., & Jun, C. (2009, 4-7 Oct. 2009). *Understanding individual and collective mobility patterns from smart card records: a case study in Shenzhen.* Paper presented at the Intelligent Transportation Systems, 2009. ITSC '09. 12th International IEEE Conference on.

Pelletier, M. P., Trépanier, M., & Morency, C. (2009). *Smart card data in public transit planning: a review*: CIRRELT.

Ratti, C., Sobolevsky, S., Calabrese, F., Andris, C., Reades, J., Martino, M., . . . Strogatz, S. H. (2010). Redrawing the map of Great Britain from a network of human interactions. *PloS one, 5*(12), e14248.

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface, 10*(84), 20130246.

Schönfelder, S., & Axhausen, K. W. (2010). *Urban rhythms and travel behaviour: spatial and temporal phenomena of daily travel*: Ashgate Publishing, Ltd.

Simini, F., González, M. C., Maritan, A., & Barabási, A.-L. (2012). A universal model for mobility and migration patterns. *Nature, 484*(7392), 96-100. doi: 10.1038/nature10856

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics, 6*(10), 818-823. doi: 10.1038/nphys1760
Soto, V., & Frías-Martínez, E. (2011). *Automated land use identification using cell-phone records.* Paper presented at the Proceedings of the 3rd ACM international workshop on MobiArch.
Szell, M., Sinatra, R., Petri, G., Thurner, S., & Latora, V. (2012). Understanding mobility in a social petri dish. *Scientific Reports, 2*, 1-6. doi: 10.1038/srep00457
Thiemann, C., Theis, F., Grady, D., Brune, R., & Brockmann, D. (2010). The structure of borders in a small world. *PloS one, 5*(11), e15422.
Zhong, C., Arisona, S. M., Huang, X., Batty, M., & Schmitt, G. (2014). Detecting the dynamics of urban structure through spatial network analysis. *International Journal of Geographical Information Science, 28*(11), 2178-2199. doi: 10.1080/13658816.2014.914521
Zhong, C., Huang, X., Müller Arisona, S., Schmitt, G., & Batty, M. (2014). Inferring building functions from a probabilistic model using public transportation data. *Computers, Environment and Urban Systems, 48*, 124-137.
Zhong, C., Müller Arisona, S., Huang, X., & Schmitt, G. (2013). *Identifying spatial structure of urban functional centers using travel survey data: a case study of Singapore*. Paper presented at the Proceedings of The First ACM SIGSPATIAL International Workshop on Computational Models of Place, Orlando FL, USA.

Highlights

- We proposed a framework for measuring variability based on level of details.
- We presented analytical methods for measuring variability at different levels.
- A case study of Singapore was conducted using one-week smart-card data.
- Insights were made into the transit, social and urban dynamics in Singapore.

Table 1. Preliminary Smart-Card data processing.

Date	**Number of rides (original records)**	**Number of Trips (with invalid records removed)**	**Number of Edges (constructed spatial network)**
Saturday (06/04/2013)	5041982	3599742	252968
Sunday (07/04/2013)	4529983	3295580	233109
Monday (08/04/2013)	5781244	4026319	283055
Tuesday (09/04/2013)	5894184	4119277	284758
Wednesday (10/04/2013)	5983855	4174802	288356
Thursday (11/04/2013)	5923847	4143741	284523
Friday (12/04/2013)	6261704	4405694	290437