Using Mobile Phone Data to Explore Spatial-Temporal Evolution of Home-Based Daily Mobility Patterns in Shanghai

Zhicheng Liu
School of Information Science and Engineering
Southeast University
Nanjing, China
zhichengliu@seu.edu.cn

Jinbin Yu
School of Information Science and Engineering
Southeast University
Nanjing, China
jinbinyu@seu.edu.cn

Weiting Xiong
School of Architecture
Southeast University
Nanjing, China
xiongweiting90@126.com

Jian Lu
School of Information Science and Engineering
Southeast University
Nanjing, China
lujian1980@seu.edu.cn

*Junyan Yang**
School of Architecture
Southeast University
Nanjing, China
yjy-2@163.com

*Qiao Wang**
School of Information Science and Engineering
Southeast University
Nanjing, China
qiaowang@seu.edu.cn

* To whom correspondence should be addressed.

***Abstract*—This paper aims at investigating home-based daily mobility patterns in Shanghai. The dataset consists of Data over Signaling (DoS) from 107,100 anonymous mobile phone subscribers in Shanghai over 9 days in different seasons, which contains spatial-temporal information of subscribers. Daily mobility pattern is characterized as motif of the individual's daily trajectory in this paper. Homes of subscribers are recognized with a priori knowledge. Motifs are extracted from each individuals' daily trajectories. In this way, we have revealed the spatial-temporal evolution of home-based daily mobility patterns in Shanghai. We find that the spatial distribution of home-based daily mobility patterns inside the enclosed area of Middle Ring Road in weekdays diffuses to the enclosed area of Outer Ring Road in weekends. However, the spatial distribution of home-based daily mobility patterns in the areas outside Outer Ring Road seems to be invariant to weekdays and weekends and is more active than that inside the enclosed area of Outer Ring Road. These active areas include affordable housing communities, such as New Gucun Big Homeland and Xinkai Homeland. The phenomenon presented in this paper may correlate with socioeconomic factors in different regions of Shanghai and worth further investigation.**

Keywords—Human mobility; Mobile phone data; Motifs; Shanghai

This research is supported by National Natural Science Foundation of China under Grant No. 51578128.

I. Introduction

Human mobility is of great concerned by urban planners and transportation planners [1-5]. The reason is that understanding human mobility in cities can help to improve urban accessibility and quality of life [6]. In past, investigations on human mobility were mainly based on collecting survey data in small sample of citizens. This traditional approach is faced with the challenge of rapid urbanization progress [7]. The small sample size of survey data and expensive cost of data collection can no longer meet this trend [7]. However, the emergent field of big data has provided researchers with powerful tools and new approach to tackle this challenge [8].

Big data has enabled researchers to investigate human mobility in various aspects [9-11]. Reference [9] exploits 3-month GPS data from 12,000 taxicabs in Beijing and POI datasets to discover regions of different functions. Functions of each regions can be viewed as the topics of each documents in a generative statistical model, called latent Dirichlet allocation. Human mobility represented by Origin-to-Destination (OD) matrix extracted from GPS data of taxicabs is in analogy to words of documents. Thus, mobility patterns can reveal the functions of the regions. This method can be used to calibrate land use planning. Reference [10] demonstrates how daily location data recorded from mobile phones help to identify importation routes that contribute to malaria epidemiology in

Kenya. The daily locations of 14,816,521 Kenyan mobile phone subscribers between June 2008 and June 2009 were collected and merged with malaria prevalence map from 2009. Sources and sinks of imported infections were discovered by identifying the dynamics of human carriers. The result of this research could improve malaria control programs. Reference [11] explores the method to develop transient OD matrix by mobile phone Call Detail Records (CDR) data collected from 2.87 million subscribers of Dhaka, Bangladesh. The estimated transient OD matrix is validated by traffic counts from 13 key locations. Compared with traditional approaches, this approach benefits from big data of human mobility. It is more economic and is capable of providing transportation authority with transient view of traffic flows.

Our study aims at investigating home-based daily mobility patterns in Shanghai. We collected Data over Signaling (DoS) from 107,100 anonymous mobile phone subscribers in Shanghai over nine days. Individuals’ daily mobility patterns are characterized as daily mobility network (motif) [12] in our study. The daily motifs are extracted from daily trajectories of each individuals. Then, homes of each individuals are detected by our algorithm. We combine individuals’ mobility networks and their homes to represent the spatial-temporal evolution of daily mobility patterns.

II. DATA DESCRIPTION

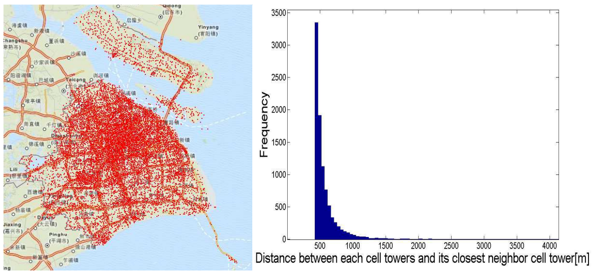
TABLE I. DATES OF THE SELECTED NINE DAYS

Season	Weekday(WD) or Weekend(WE)	Date
Autumn	WD	2012.9.19
	WE	2012.9.23
Winter	WE	2012.12.23
	WD	2012.1.17
Spring	WE	2013.4.7
	WD	2013.4.10
Summer	National Holiday	2013.5.1
	WE	2013.7.14
	WD	2013.7.18

Data over Signaling (DoS) is designed to control the telecommunication process between mobile phones and telecommunication infrastructures such as cell towers and mobile switching centers. The locations (i.e. cell tower) and timestamps of the mobile phone subscribers are recorded once a signaling event occurs. Thus, DoS can be used to study human mobility in that it records spatial-temporal information of the mobile phone subscriber. In this research, we obtain DoS from one of Shanghai local mobile phone service carriers. 107,100 anonymous mobile phone subscribers are sampled from the entire Shanghai local mobile phone subscribers and we record nine days (see table I) of signaling events (see table II) of these subscribers (the subscribers remain unchanged). Notice that a cleaning procedure has been conducted by our data provider before we reached the data. The procedure is as follows: (1) Each days is partitioned into 24 non-overlapping time slots and the length of each time slots is one hour. (2) A mobile phone subscriber’s location in each time slot is assigned as the most frequently visited cell tower during this time slot. If there is no DoS recorded in this time slot, the location is assigned as the last recorded location. Though this procedure may filter out some useful information such as the time duration a subscriber stayed in one tower, the recorded signaling events include periodic location update which means a subscriber’s location will be updated periodically. In this scenario, the updated period is approximately one hour. Therefore, the time resolution of each subscribers’ records is one hour. The spatial resolution of subscriber location is at cell tower level. Our data set records 9291 cell towers. Fig. 1 shows the cell towers distribution over Shanghai and the histogram of distance between each cell towers and its closest neighbor cell tower.

TABLE II. THE RECORDED SIGNALING EVENTS

Event	Explanation
Calling	The subscriber makes a phone call.
Called	The subscriber receives a phone call.
Sending text message	The subscriber sends a text message.
Receiving text message	The subscriber receives a text message.
IMSI attach	The subscriber turns on the phone.
IMSI detach	The subscriber turns off the phone.
Periodic location update	The mobile phone automatically and periodically connect to closest cell tower to update its location.
Normal location update	This event happens when the subscriber move across a Location Area (LA).
Data service	The event includes following data service: ACTIVE PDP CONTEXT、DEACTIVE PDP CONTEXT、Modify PDP Context、RADIO STATUS、RAU-NORMAL、RAU-PERIODIC、GPRS ATTACH、GPRS DETACH、TD ATTACH、TD DETACH、TAU-NORMAL、TAU-PERIODIC.
Switching within Base Station Controller (BSC)	BSC switches a logical channel if the connecting signal in current channel is weak.
Switching across BSC due to downlink signal quality and strength variation	This event may happen when the location of subscriber is on boundary of BSC.
Switching across BSC due to other factors	These factors include hardware failure, improper data configuration, etc.
Switching across Mobile Switching Center (MSC) due to downlink signal quality and strength variation	This event may happen when the location of subscriber is on boundary of MSC.
Switching across BSC due to other factors	The same as that of BSC.
Conversation ends	The subscriber rings off.
Cell reselection in idle state	The mobile phone reselect cell in idle state so as to balance the random access burden.
Others	Unrecognized events.

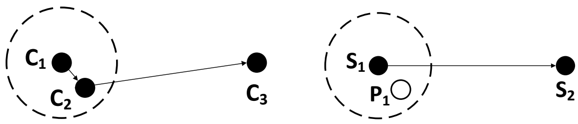


(a) (b)

Fig. 1. Recorded cell towers: (a) Cell tower distribution over Shanghai. (b) Histogram of distance between each cell towers and its closest neighbor cell tower.

III. METHEDOLOGY

A. Stay Point Detection

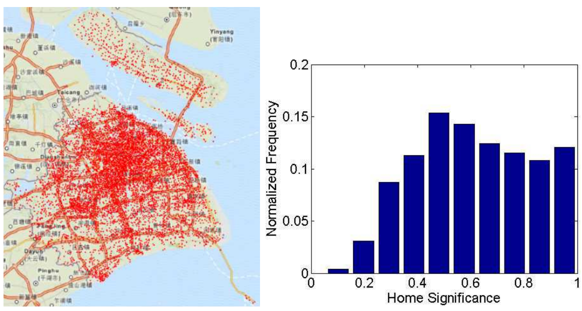


(a) (b)

Fig. 2. Mechanism of stay point detection: (a) Original trajectory. (b) Trajectory filtered by stay point detection.

Our goal is to understand daily human mobility patterns in Shanghai. Thus, our first step is to extract stay points from each individual's daily trajectory and eliminate the pass-by points. We exploit the method demonstrated in [13]. However, due to the coarse time resolution of our data, we only take the spatial condition to detect stay points. Here, the spatial neighboring threshold is 650 meter which implies that if two points from a trajectory lie within 650 meter, these two points will be accounted as one place which coordinated as the location of the first point of these two. This method will bring some bias towards extracting motif, but the result shown in the next section seems to be robust to this stay point detection method.

B. Home Detection



(a) (b)

Fig. 3. (a) Detected homes. (b) Histogram of home significance.

Home is usually the place where individual's daily trip starts and ends. Thus, it's important to detect subscribers' homes from their trajectories so as to understand mobility patterns in geographical semantics. In this scenario, we detect each individual's home by recognizing the most frequent cell tower at which the individual stays from 1 a.m. to 6 a.m. during the observed nine days. We regard this cell tower as the home cell tower and term this frequency of each individuals as home significance, i.e. the frequency of the most frequent cell tower that are visited by individual from 1 a.m. to 6 a.m. We define home significance in the sense that not every citizens maintain a regular daily life, e.g. taxi drivers and cleaners. It will be biased if we detect their homes by this method. The home significance is used to eliminate biased sample in latter section. Fig. 3 shows the detected home cell towers and the histogram of home significance of 107,100 subscribers.

C. Extracting Individual Motifs

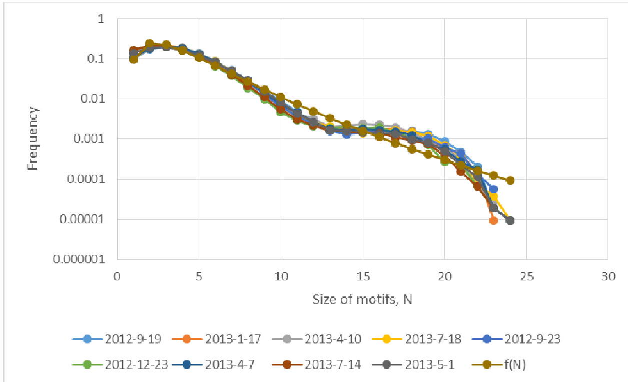


Fig. 4. Empirical distribution of the size of motif.

Network motif is the subnetwork that recurs within the network more often than expected at random [14]. Human mobility trajectory can be viewed as a network and its motif is an essential tool to reveal human mobility patterns [12]. In this research, we filter the individual daily trajectories to extract stay points and eliminate the pass-by points. These stay points of individual's daily trajectory compose a directed network (graph) and unravel the individual activities pattern. Since the selected nine days (see table 1) are discontinuous, we make an assumption that, for each individuals, the daily trajectory ends at the same cell tower where the trajectory starts. This means if the last stay point and the first stay point of a daily trajectory mismatch, we manually add the first stay point to the end of the trajectory. After this procedure, we recognize individual daily motif by the algorithm presented in [15].

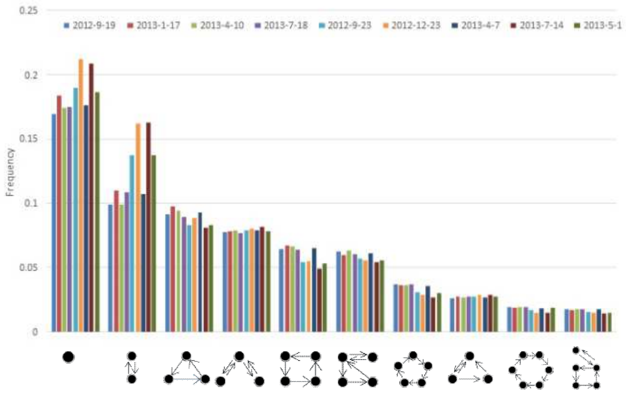


Fig. 5. Empirical distribution of top 10 frequent motifs.

The aggregate daily motif should be investigated. First, we explore the distribution of the size of motifs, i.e. the number of visiting places, N. The result is shown in Fig. 4. Notice that for N < 14, the number of daily visited locations can be approximated as log-normal distribution

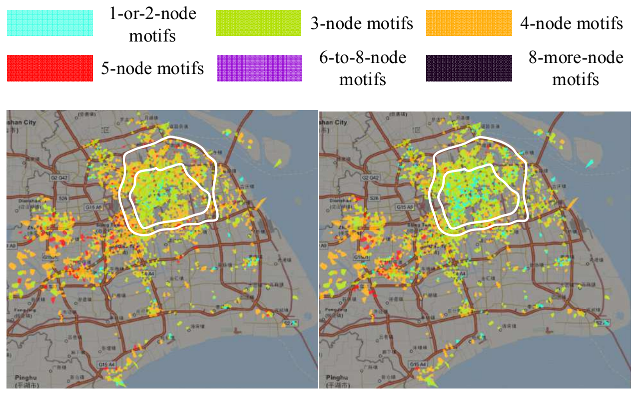
$$f(N)=\frac{1}{N\sigma(2\pi)^{1/2}}\cdot\exp(-\frac{\ln N-\mu}{2\sigma^2})$$

with μ= 1.18 and σ=0.59 . However, for N > 13, the empirical distribution distorted from being log-normal. Instead, it stops decreasing as N increases. This phenomenon will be discussed in discussion section. Second, we investigate the distribution of different motifs. In Fig. 5, we exhibit the top 10 frequently recurred motif. Approximately 70 per cent of the subscribers' daily mobility patterns can be explained by these ten motifs.

IV. RESULT

In this research, we correlate each individuals' homes and their motifs to analyze the spatial distribution of motifs in weekdays and weekends. It's important for urban planners to understand how citizens' neighborhood may have an effect on their daily travels [16]. The home detection algorithm that we utilize cannot convincingly detect homes of subscribers who do not lead a regular life. In this scenario, we define a priori regular life as taking sleep at home between 1 a.m. and 6 a.m. To filter out subscribers who lead irregular life, we select the subscribers whose home significances are greater than 0.3, see Fig. 3. We also eliminate the home cell tower of which population is less than 11.

Fig. 6 illustrates the spatial distribution of daily mobility motifs in Shanghai during weekdays and weekends. We aggregate motifs of the same size, i.e. the number of nodes, into a class. Due to the coarse time resolution of our data, twenty-four classes can be retrieved at most. For each home cell tower, we calculate the frequency distribution of each class of motifs of which individuals' homes lie in this cell tower. We assign the median of motif class distribution of this cell as the most significant mobility pattern of each cell tower. In this way, the mobility patterns of each cell towers are represented as the median of motif class distribution.



(a) (b)

Fig. 6. Spatial distribution of home based mobility motifs in Shanghai. (a) Mobility motifs in weekdays. (b) Mobility motifs in weekends. The outer white ring is the Outer Ring Road and the inner white ring indicates the Middle Ring Road. To improve color contrast enhancement, we deliberately set the color of blank Voronoi cell to light gray.

Fig. 6(a) shows the spatial distribution of the median of motif class distribution of each cell tower in weekdays (i.e. four days, see table I). To clearly demonstrate the spatial distribution, we use Voronoi diagram to approximately show the spatial coverage of each cell towers. The color of each Voronoi cells represents the class of the motifs. In order to uncover the geographic background, we also set transparency to Voronoi cell. In Fig. 6(a), we can clearly observe that inside the enclosed area of Middle Ring Road, specified by the inner white ring in Fig. 6(a), the mobility patterns are dominated by 1-node, 2-node and 3-node motif class. Notice that most commonly seen real world motifs in weekdays are 2-node and 3-node work related motif [15]. We can infer that during weekdays, most citizens living in enclosed area of Middle Ring Road have simple mobility patterns. This phenomenon may be caused by the intense circadian rhythm of citizens living inside the enclosed area of Middle Ring Road. However, subscribers who live outside Middle Ring Road exhibit more complex mobility patterns. Their mobility patterns are dominated by 4-node and 5-node motifs.

Due to the different mobility patterns of individual activities in weekdays and weekends, in Fig. 6(b), we show the spatial distribution of the median of motif class distribution in weekends (i.e. 4 days, see table I). We find that 1-node motif concentrated inside the enclosed area of Middle Ring Road, which indicates that residents living in the central area of Shanghai tend to have stay-at-home mobility patterns in weekends. Comparing Fig. 6(a) and 6(b), we discover that the pattern of spatial distribution of motifs diffuses from the inside area of Middle Ring Road to the inside area of Outer Ring Road. The dominant motifs are 2-node and 3-node motifs in the area between Middle Ring Road and Outer Ring Road during weekends. The spatial distribution of the median of

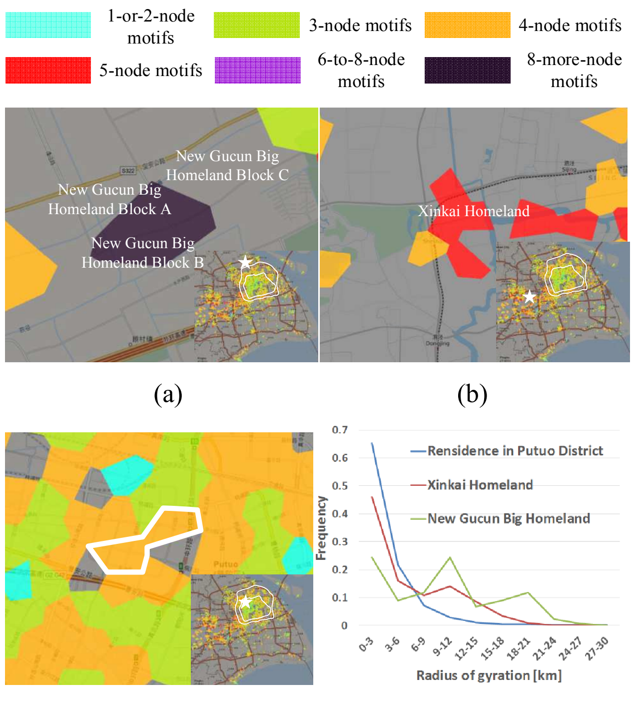


Fig. 7. Neighborhoods with high median value of motif class distribution. (a) New Gucun Big Homeland. (b) Xinkai Homeland. (c) Residence in Putuo district used as reference. (d) Distribution of radius of gyration of these 3 neighborhoods.

motif class distribution in holiday, which is not shown in Fig. 6, is almost the same as that in weekends.

What we find spatial distributions of motifs in weekdays and weekends have in common is that the regions outside the Outer Ring Road are always dominated by 4-node and 5-node motifs. This phenomenon may correlate with the socioeconomic factors in these areas. In Fig. 7, we zoom into the neighborhoods within which the median of motif class distribution is relatively high. These neighborhoods include affordable housing communities which are built up by government and rent to low-income families, such as New Gucun Big Homeland, Xinkai Homeland. From Fig. 7(a) and 7(b), New Gucun Big Homeland and Xinkai Homeland exhibit relatively high median value of motif class distribution, indicating that, residents living in these two areas tend to travel more actively. To show their active travel patterns, distributions of radius of gyration [17] are shown in Fig. 7(d). Radius of gyration is a metric to capture the characteristic travel distance of mobility patterns [17]. In order to demonstrate the singularity of the distribution of radius of gyration in New Gucun Homeland and Xinkai Homeland, residents of a home cell tower in Putuo District are introduced as a reference. The location of reference cell tower is shown in Fig. 7(c). We can observe that residents living in New Gucun Homeland and Xinkai Homeland have higher probability to take lager radius of gyration than those living in Putuo District. This phenomenon may correlate with regional socioeconomic factors. In Gucun town, over 70 per cent of residents are renters and relocated households from rural area [18]. However, the proportion of these two groups of people in other regions is less than 20 per cent [18]. We suggest that further relationship between regional socioeconomic factors and mobility patterns should be investigated provided that such data is accessible.

V. DISCUSSION

The data we utilize in this research is cleansed by our data provider. The time resolution of our data is 1 hour. Thus, some activities may not be recorded such as picking up children. Also, we cannot detect the activity of which time duration is less than 1 hour. This coarse time resolution leads to the inaccurate detection of stay points of individual trajectory. Some pass-by points will not be eliminated by our method, since we cannot detect time duration that individual stay nearby this point and its distance from other points of trajectory is greater than our threshold value. This defect will result in larger size of motif for some subscribers who traveled through a long distance and left a DoS record during her/his journey. Therefore, the distortion of the distribution of the size of motif can be explained (see Fig. 4). For $N > 13$, an anomalous fat tail occurs in the distribution. These subscribers have at least 13 non-adjacent locations recorded in a single day, which implies their active travel behaviors. For these active travelers, lots of pass-by points occur in their trajectories as shown in Fig. 8. Our stay point detection algorithm will fail to eliminate these pass-by points. Thus, the size of motifs extracted from trajectories tends to be greater than expected for $N > 13$ and smaller than expected for $N < 14$. However, this defect will not influence our result for the following reason. Even though up to 24 classes of motifs could be found, we did not distinguish the motif classes of which the number of nodes are greater than 8, see Fig. 6. This means subscribers who are potential to be active travelers are assigned to the same class provided their real motif size is larger than 8.

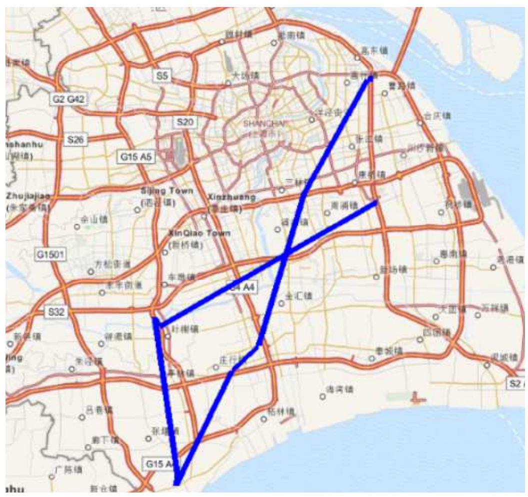


Fig. 8. Daily trajectory of an anonymous active traveler.

VI. CONCLUSION

Understanding human mobility patterns related to their homes is important for urban planner and policy maker to understand how the built environment may influence individuals' travels. In this research, we exploit DoS of 107,100 anonymous mobile phone subscribers over nine days to detect their home locations and to extract individuals' mobility patterns. By synthesizing the home location and individual's mobility pattern, we have revealed the spatial-temporal evolution of daily mobility patterns in Shanghai.

From macroscopic scale, we discover that the mobility patterns within the area enclosed by Middle Ring Road are dominated by 1-node, 2-node and 3-node motifs during weekdays and this spatial distribution patterns diffuse to the enclosed area of Outer Ring Road during weekends. However, the daily mobility patterns outside the Outer Ring Road are dominated by 4-node and 5-node motifs and seem to be invariant to weekdays and weekends.

From mesoscopic scale, regions which contain affordable housing communities exhibit active mobility patterns, such as New Gucun Big Homeland and Xinkai Homeland. We argue that this phenomenon may correlate with socioeconomic status of citizens living in different regions of Shanghai. This relationship needs further investigation.

In conclusion, we utilize state-of-the-art data mining techniques to provide urban planners with knowledge of spatial-temporal evolution of daily human mobility patterns in Shanghai.

REFERENCES

[1] M. Batty, *The New Science of Cities*. MIT Press, 2013.

[2] S. Jiang, J. Ferreira, and M.C. González, "Clustering daily patterns of human activities in the city," Data Min. Knowl. Discov., vol. 25, no. 3, pp. 478–510, Apr. 2012.

[3] Transportation Research Board, "Activity-based travel demand models: a primer," 2015.

[4] A.R. Pinjari and C.R. Bhat, "Activity-based travel demand analysis," A Handb. Transp. Econ., no. 1, pp. 1–36, 2011.

[5] J. Hao, M. Hatzopoulou, and E. Miller, "Integrating an activity-based travel demand model with dynamic traffic assignment and emission models," Transp. Res. Rec. J. Transp. Res. Board, vol. 2176, pp. 1–13, 2010.

[6] K. Lynch, The Image of the City. Cambridge, MA: The MIT Press, 1964.

[7] M. Batty, "Cities as complex systems: scaling, interactions, networks, dynamics and urban morphologies," in Encyclopedia of Complexity and Systems Science, vol. 1, R. Meyers, Ed. Berlin, DE: Springer, pp. 1041–1071, 2009.

[8] National Research Council. *Frontiers in Massive Data Analysis*. Washington, DC: The National Academies Press, 2013.

[9] J. Yuan, Y. Zheng and X. Xie. "Discovering regions of different functions in a city using human mobility and POIs". In Proc. of the SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '12), ACM Press, 186-194, 2012.

[10] A. Wesolowski et al. "Quantifying the impact of human mobility on malaria". Science 338(6104):267–270, 2012.

[11] M.S. Iqbal, C.F. Choudhury, P. Wang and M.C. González, "Development of origin-destination matrices using mobile phone call data," Transport. Res. C 40, 63–74, 2014..

[12] C.M. Schneider, V. Belik, T. Couronné, Z. Smoreda and M.C. González, "Unravelling daily human mobility motifs," J. R. Soc. Interface 10,20130246, 2013.

[13] Y. Zheng, "Trajectory data mining: an overview," ACM Trans. Intell. Syst. Technol., vol. 6, no. 3, pp. 29:1–29:41, May 2015.

[14] U. Alon, "Network motifs: theory and experimental approaches". Nat. Rev. Genet. 8, 450–461, 2007.

[15] S. Jiang, J. Ferreira, and M.C. González, "Activity-based human mobility patterns inferred from mobile phone data: A case study of Singapore," in Proc. UrbComp, Sydney, NSW, Australia, pp. 1–9, Aug. 10, 2015.

[16] R. Ewing and R. Cervero, "Travel and the built environment -- a meta-analysis," J. Am. Plan. Assoc., vol. 76, no. 3, pp. 265–294, 2010.

[17] M.C. González, C.A. Hidalgo and A.-L. Barabási, "Understanding individual human mobility patterns," Nature, volume 453, 2008.

[18] D. Gan, D. Wang and W. Zhu, "Research on the residents' commuting feature of large-scale residential district in suburban Shanghai:A case study of Gucun, Baoshan District," Geographical Research, 34(8): 1481-1491, 2015.