*AMSE JOURNALS-AMSE IIETA publication-2017-Series: Advances B; Vol. 60; N°1; pp 224-238*
*Submitted Jan. 2017; Revised March 15, 2017, Accepted April 15, 2017*
*https://doi.org/10.18280/ama_b.600114*

# Understanding Core Districts of City using Human Activity Data

*Duan Hu, **Jie Yang, ***Benxiong Huang

*The Key Laboratory of Fiber Optic Sensing Technology and Information Pro cessing, Ministry of Education, Wuhan University of Technology, China
(huduan@whut.edu.cn)
**The Key Laboratory of Fiber Optic Sensing Technology and Information Pro cessing, Ministry of Education, Wuhan University of Technology, China
(jieyang@ whut. edu. cn)
*** Communication software center, EIE dept, Huazhong University of Science and Technology, China

## Abstract

For "polycentric" urban systems, the role of city centers and their human flows's spatial influence on the surrounding area remain a challenging problem with many applications ranging from transportation planning to epidemiology. Firstly, we segmented urban area to different districts as the basic research unit and abstract latent activity contexts from large scale call logs in city using existing approaches. By proposing urban area as a inter-district network model of which edge is measured by temporal-spatial proximity among different districts stem from those latent activity contexts, we identified core districts of city by using a network-based method based on community detection and local centrality. We further analyzed the functional role of these core districts in city. It provides a feasible approach to uncover spatial structure of an urban system.

## Key words

City's core districts, community detection, human activity

## 1. Introduction

With the rapid development of urban-centered economy, today swelling cities is being filled with massive socioeconomic activities, urban area have gone through strong but heterogeneous sprawl [1]. In particular, the complicated transportation forms, various types of urban activities and the intrinsic complexity of human mobility patterns inevitably result in the maturity of emerging activity centers during the evolution of a city. In recent years, a set of studies that concerning functional urban space [2][3][4][5][6]which is using emerging urban movement data contribute to a good understanding of urban dynamics. As the spatial concentration of some part of urban functions, it is no doubt that urban functional regions which struggles to meet people's partial needs of massive socioeconomic activities has given rise to the process of process of polycentric urban transformation.

Moreover, the geographic, social constraints of human activities and mobility patterns [7][8][9] also yield in-depth insights to poly centers in urban area. People's daily life displays significant regularity referring to the inherent regularity in individual behaviors. Recent analyses of large-scale trajectories from mobile phone data [10] found that human experience a combination of periodic movement which is geographically limited and seemingly random jumps correlated with their social networks. Specially, this inherent character of human mobility [5] would enable urban socioeconomic activities show locality and hierarchy. For instance, for occasional impulse on splurge or group outing with friends, people would go to those recognized functional centers of metropolitan area. However, people prefer to go shopping or having fun near their house after work, even in weekends due to those city diseases such as population explosion and traffic congestion. It is true that for the countries like China and India, the phenomenon leads us to believe that, business centre or public transportation hubs are unconditional hot spots of global activity configuration in urban area, meanwhile some less attractive districts also would be at the core of local activity configuration. Those sub-centers can be considered as core districts of this "polycentric"urban structure.

In particular, city center that can be measured by the degree of spatial concentration of urban movement and human activity, certainly make significant insight into polycentric urban systems. A wide range of method [11][12][13]to identify various type of urban centers had been given in decades. For instance, functional centers can be identified by measuring the connectivity of individual centers to the whole urban system through human flows. Activity centers is strongly corrected with the peak of population density functions [13]. In [14], the author discuss "polycentric" urban spatial structure the identification of their poly-centers can only considered as geographical proximity of groups of station. The recent proliferation of massive ubiquitous sensors (such as GPS, mobile phone, smart card) plus the spectacular ability of researchers to identify city centers for urban studies and in-depth understood what role these centers play in the city.

In this paper, we attempt to uncover core districts of city using a network-based approach. Generally, above all else, the city has been segmented to a number of districts as basic research unit by a natural map segmentation method [14]. Massive individual latent activity geographical contexts [6][15] stemming from mobile call logs has been applied to measure temporal-spatial proximity among these districts. Furthermore, a district network model of which edge is measured by temporal-spatial proximity has been considered to understand core districts both using community detection technique [16] and local centrality indicator for complex network. Unlike the methodology research to discover different functional regions in a city that represent spatial concentration of certain urban function, our study focus on discovering the spatial concentration of all type of socioeconomic activities and in-depth study of some inherent nature of urban spatial structure. In addition, despite of the identification of core districts which is similar to those work refer to identify the city centers. We place more attention on locality and hierarchy of urban spatial structure.

In addition, as a case study, an anonymous mobile phone call data record dataset generated from Wuhan, an emerging metropolis in Central of China, has been used to analyze the role of core districts in the city. In this paper, we offer the following contributions:

1). In the first place we measured temporal-spatial proximity between any two districts using the quantity of their referred individual activity contexts. An inter-district network model with temporal-spatial proximity has been developed.

2). We developed an identification method for core districts of city based on community detection and local centrality indicator in network theory.

3). As a case study, we further analyzed the identified core districts in urban area of Wuhan.

Their functional role also has been illustrated using POI data.

## 2. Problem definition and methodology

In this section, we will discuss the characteristic of "polycentric" urban spatial structure more broadly in the perspective of the complex network. The new quantitative measurement named core district will be further defined.

### 2.1 Network modelling of urban system

A road network is usually comprised of some major roads like highways and ring roads, which naturally partition a city. We firstly apply a vector-based map segmentation method [14] to segment urban area of a city into region units in terms of major road network. Comparing with traditional image-processing method and grid-based methods, this method produces a more meaningful and easy-operated segmentation. Intrinsically, a segmented region should be an existing district of city, which is basically unit carrying human activities and also the origin and destination of a trip. We even believe that as a basic research unit, these districts are very well suited for studying the characteristic of urban spatial structure by modelling an inter-district network model.

Then, core district can be defined as district where nearby people are accumulated to perform certain activities which means that the core district should have most interactions with its neighborhood in inter-district network model.

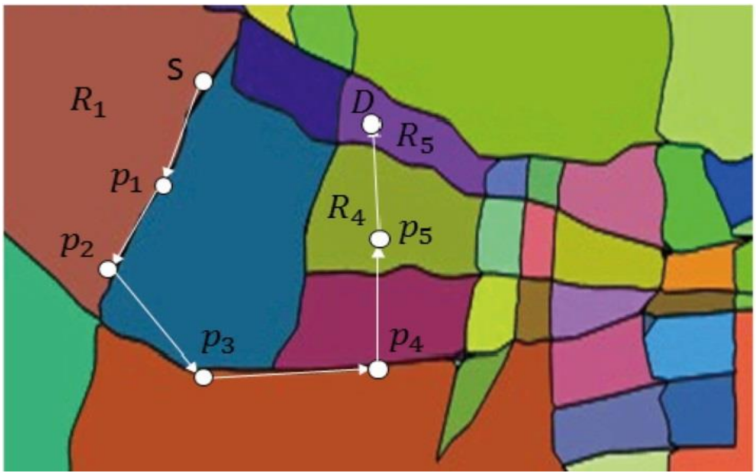


Fig.1. Trajectory mapping

In general, complex network can be used to model many types of interactions as the edges of network, including information or behavior transmission, attribute similarity, various types of contact. An intuitive interaction should be spatial proximity. Bordering districts are more closely related to one another. Either their functional configurations are so similar that they can attract the same people and meet their daily needs, or so complementary that there are existing fixed human flows between them. Note that spatial proximity highly relevant to a large number of shared-visitors and their migration patterns, it is reasonable to measure the relationship of two districts only using the number of trips between them. So we propose a novel type of interactions which is called temporal-spatial proximity using human activity data. Specifically, the number of trips between two districts will be used to describe this interaction.

Here we first need to count latent activity contexts for each individual. After the map segmentation of given urban area, the location history from activity log of individual $i$ can be mapped to a districts sequence by a simple rule in [15] as shown in fig.1. $T: S \to p_1 \to p_2 \cdots \to p_5 \to D$ can be mapped to the district sequence $D_1 \to 0 \to 0 \to 0 \to D_4 \to D_5$ as the latent activity contexts of individual $i$.

Table 1. The algorithm for inter-district network model

| Algorithm.1 |
|---|
| Input: Vector-based road network data $V$<br>individual activity contexts $S$ of all individuals $I$<br>Output: The weighted adjacency matrix $W$<br>district set $N$<br>1 segmented urban area to district set $N$ using the method in [14]<br>2 Return $N$<br>3 for each $i \in I$ , map check point sequence $S$ to the district sequence $D$ using the approach presented in [15].<br>4 calculating the weight $w_{xy}$ between any two districts $x$ and $y$<br>5 Return $W$ using equation(1) |

Given $T(x, y, i)$ represent the number of all trips from district $x$ to district $y$ in the latent activity context of $i$, the temporal-spatial proximity between district $x$ and $y$ can be evaluated by $w(x, y) = \sum_{i \in C(x,y)} T(x, y, i)$, $C(x, y)$ is the set of traveller that had at least one trip from district $x$ to district $y$ .We set the edge is undirected and the edge weight between district $x$ to district $y$ is identified to $w(x, y)$ due to the reason will be given in empirical analysis.

In short, we use an undirected graph with weight to model the temporal-spatial proximity among all districts in urban area. The algorithm has been presented in table.1.

The weighted adjacency matrix $W$ of this graph has been given as equation (1):

$$\begin{pmatrix} 1 & \cdots & w_{1n} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & 1 \end{pmatrix} \tag{1}$$

, and we consider $w_{xy} = 0$ if $x = y$.

## 2.2 The network-based method for identifying core districts

Core district can be conceptually regarded as center in local space of city. For instance, there are over one hundred and thousand IT workers in "Shang Di" which is a well-known innovation park in Beijing. Their migration pattern usually takes the subway to "Shang Di" from the nearby residential blocks such as "Hui Long Guan" or "Bei Yuan" in the morning of workday, having lunch or eating something in the surrounding stores, and getting home after 22'clock in the evening. In Sunday, most of them will meet their lives' needs in these residential blocks .In this case, "Shang Di" will be the core district of city in the first place, and have considerable potential to be a city center.

In the long time, many researchers used conventional measurements such as population density and activity diversity to identify city centers. Comparing with the above global measurements, the detection of core districts should consider some measurements with more locality. In this paper, with the concept of inter-district network model, the community detection technique and centrality indexes in complex network will be used for the identification of core districts. We will detect district groups in proposed network model using symmetric nonnegative matrix factorization firstly and identify the district at center of its group further.

**1). The detection of district group**

We choose symmetric nonnegative matrix factorization (SNMF) [15] as our tool to aggregate districts using adjacency matrix of inter-district network because of its powerful interpretability and close relationship between clustering methods. Assuming the given urban area has $k$ district groups. Here the goal is to factorize $X$ into the non-negative $n \times k$ matrix $U$ and the non-negative $k \times n$ matrix $V^T$ that minimize the following objective functions $J = \frac{1}{2}\left\| X - UV^T \right\|$, where $\|\bullet\|$ denotes the squared sum of all the elements in the matrix. Specially, due to our adjacory matrix is symmetric, $U = V$ , so, $J = \frac{1}{2}\left\| X - UU^T \right\|$.

Our destination is to solve the minimization problem: minimize $J$ with respect to $U$ under the constraints of $u_{ij} > 0$, where $0 \le i \le n, 0 \le j \le k$ . According to[17], $U$ can be solve by the following multiplicative update rule:

$$U_{ik} \leftarrow U_{ik}\left(\frac{1}{2} + \frac{(XU)_{ik}}{(2UU^TU)_{ik}}\right) \quad (2)$$

After convergence, the obtained $U^*$ is just the scale partition matrix of adjacency matrix $X$ of size $n \times k$, whose *i-th* row corresponding to the cluster membership of the *i-th* unit. We can further normalize $U^*$ to make $\sum U_{ij} = 1$, such that $U_{ik}$ corresponding to the posterior probability that the $i$-th unit belongs to the $k$-th group. In our experiment based on the adjacency matrix with Symmetric nonnegative matrix factorization method, we firstly assume the parameter $k = 20$ as a prior estimation on the maximum number of groups. The regions that have identical memberships from the 10 runs of this method have been colored as district groups since the SNMF result is initialization dependence.

**2). District centrality**

Based on the definition of core district, core district may play two different roles in their group. It may either be the district that have more interactions with other districts in same group, or keep a bridge role between other districts in same group and other groups former role used to be measured by local centrality [18]. In our study, degree centrality would be used to compute local centrality in this undirected graph with weight. For the latter, betweenness centrality [19] would describe the influence of every district in their group from a global perspective.

We detail these centrality measurements in proposed network model as follows:

Local weighted degree centrality (LWDC): in each district group of our undirected network with weight, LWDC of any district $i$ is the sum of the weight of its edges with other districts $j$ in same group. This measure has been formalized as follows:

$$c(i) = \sum_{j \in G, j \neq i}^{N_G - 1} w_{ij} \tag{3}$$

where $G$ is the group that district $i$ belong to, $w_{ij}$ is the weight value between district $i$ and $j$.

Weighted betweenness centrality (WBC): the sum of shortest weighted paths in whole inter-district network model.

$$g(i) = \sum_{s \neq i \neq d} \frac{\sigma_{sd}(i)}{\sigma_{sd}} \tag{4}$$

where $\sigma_{sd}$ is the sum of all shortest weighted path from start district $d$ . $\sigma_{sd}(i)$ is the sum of those paths that pass through $i$ ,the algorithm for identifying core district has been presented in table.2

Table 2. The algorithm for top-$k$ core district

| Algorithm.2 |
| --- |

Input: The weighted adjacency matrix $W$

the district set $N$ and the number of core district $k$

Output: core district set with WBC $i$ $S_w bc$

core district set with LWDC $S_l wdc$

1 calculating $g(i)$, $\sigma_{sd}$ and $\sigma_{sd}(i)$ for each district $i$ in $N$ using Dijkstra algorithm presented in [20] and equation (4)

2 sort all districts descending by $g(i)$ set top-$k$ districts as $S_w bc$

3 Return $S_w bc$

4 detecting all district groups $G_1$ to $G_k$ from $N$ using equation(2) ,calculating $c(i)$ for for each district $i$ in $N$ using equation (3)

5 sort the districts to each group $G_k$ descending by $c(i)$ set the first district in each group as the district in $S_l wdc$

6 Return $S_l wdc$

# 3. Case Study

## 3.1 Empirical Analysis on inter-district network model

**1). Data description:**

In this case study, as a well-known city in Central China, Wuhan will be used to understand the role of core districts of city. An anonymous mobile phone call data record (CDR) dataset will be used to abstract individual latent activity context. This dataset, which accounted for approximately 25% of the population and was collected by a Chinese telecom operator, are composed of massive call logs with temporal-spatial context in a time period (in our case study, the time period is 4 weeks). The map segmentation of Wuhan has been presented to use 2-level road network. As it shows in Fig.2, the urban area of Wuhan has been segmented to 321 valid districts. Noted that many regions are lake of river, avoiding the small connected areas induced by these unnecessary details such as bridges and lanes. Call logs located in each district have been aggregated as the latent activity contexts.

Fig.2 Map segmentation of Wuhan

**2).Inter-district network model:**

We develop the inter-district network model based on the map segmentation and latent activity contexts in Wuhan. In data pre-processing, the trips between any two districts stem from latent activity contexts always are directed. Subsequently, a trip pattern with bidirectional, approximate symmetrical has been observed in almost every origin-destination districts. It means that we can only considered single-direction trips and model these interactions among districts using an undirected graph with weight. The inter-district network model has been given in using algorithm.2 and has been shown in Fig.3. The statistics of proposed network model of Wuhan have been detailed in table.3.

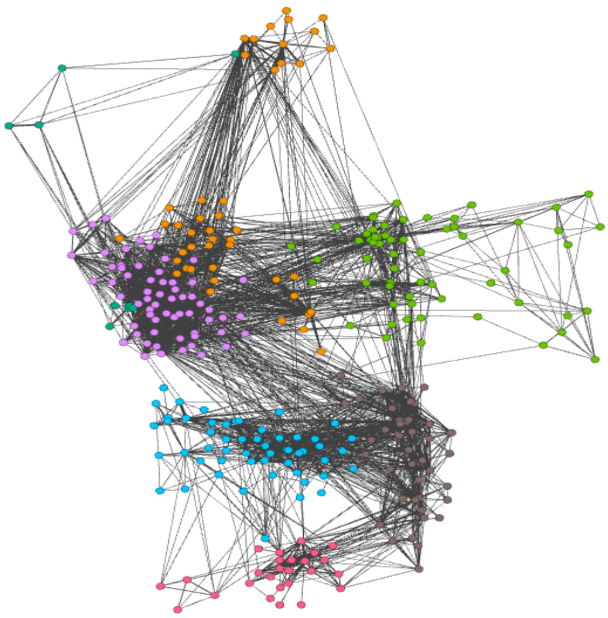

Fig.3 The inter-district network model of Wuhan

Table 3. The statistics of inter-district network

| Nodes | Edges | Average Degree | Avg.Weighted Degree | Groups |
|---|---|---|---|---|
| 321 | 3009 | 19.2 | 90267 | 6 |

## 3.2 Core districts of city

Then, we identify core districts from the network model in using algorithm.2. Firstly, the six district groups has been detected by SNMF method, these groups almost matched the administrative division of the city.as shown in Fig.4. Then, the core district of each group also is marked by red circle and its ranking in Global centricity which has been given in table.4. So, we can understand these core districts by their land mark in Baidu Map.

For instance, group A covers much of HanKou, which basically regard as the one of the important part of the city whose merging formed Wuhan. Its core district is JiangHan Road CBD which always is the most prosperous region of Wuhan in decades. Its ranking in WDC/WBC are 1/ 2.


Fig. 4 The configuration of core districts

Table 4. Top-$k$ core districts

| Core districts for each Group | | Ranking in Global Centricity | |
|---|---|---|---|
| Group | Core district | WDC | WBC |
| Group A | JiangHanLu CBD | 1 | 2 |
| Group B | NanHu Residential block | 2 | 4 |
| Group C | ZongGuan CBD | 4 | 10 |

| | | | |
|---|---|---|---|
| Group D | WangJiaWan CBD | 5 | 5 |
| Group E | SiMenKou CBD | 12 | 1 |
| Group F | TianHe Airport | 42 | 32 |

Wuchang is covered in group B and most of Wuhan Optical Valley, which both constitute traditional southern part of Wuhan. Its core district is NanHu residential block, which is the largest residential area in the south bank of the Yangtze River. And its ranking in WDC/WBC is 2 and 4.

In fact, we claim that, these core districts must be the center of their group and are quite important to meet the "polycentric" urban structure since their ranking in global centricity is high enough. Group F seems to be an exception only. However, its core district is Tianhe international airport, which is the busiest airport of central China as it is geographically located in the centre of China's airline route network.

Furthermore, we attempt to uncover the functional role of these core districts. We annotate six core district with some sematic terms, which can contribute the understanding of its real functions. Wuhan POI dataset from Baidu Map covers 187769 POIs from the year 2014, where each POI is associated with the information of this latitude, longitude and the category (see Table 5 for a complete list of categories), Fig.5 showed the POIs configuration of six core district. Restaurant, hotel and living service and public organizations accounts for a large percentage of the whole configuration.

Table 5. POI Category taxonomy

| Code POI category | |
|---|---|
| Administrative logos | Hospital |
| Road | Car service |
| Restaurant | Transportation facilities |
| Hotel | Banking and Insurance service |
| Shopping mall | Street furniture |
| Living service | Residence |
| Scenic spot | Corporate business |
| Leisure and entertainment | Public organizations |

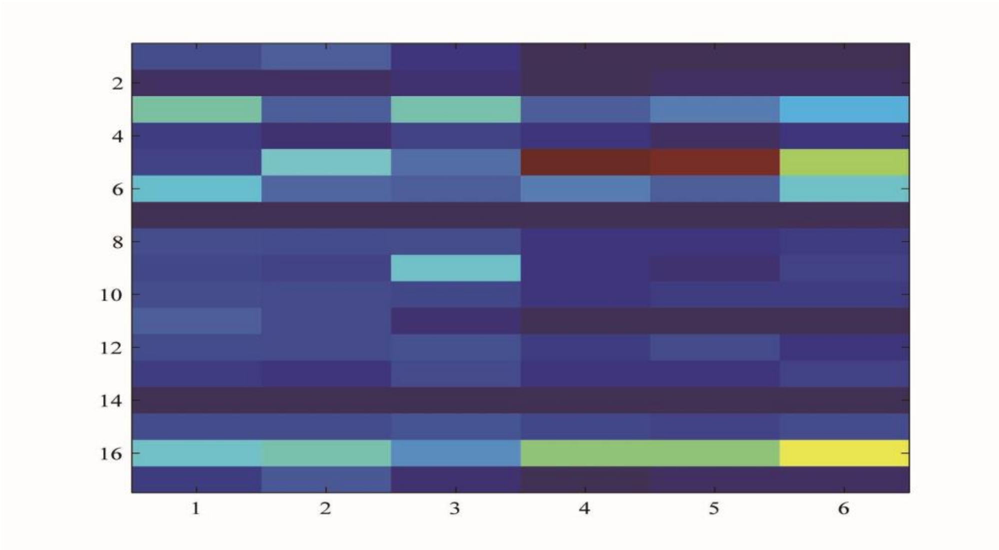


Fig.5 Functional role of core districts

## 4. Related Work

Our work focused on the spatial concentration of human activity patterns of city, the same theme named urban spatial structures have long been investigated in the past decades. Specially,in a case study of Singapore [11] explored the effectiveness of two traditional indices of city centers. [6] proposed a framework for discovering functional zones in a city using human trajectories, which implies socioeconomic activities performed by citizens at different times and in various places. Compared with above all mentioned studies, complex network's perspective is the topic we studied. We developed the inter-district network model using latent activity contexts stem from human activity data. In [21], the authors use similar methodology to apply telecommunication data for partitioning the Great Britain. However, their map segmentation is simple and the purpose is quite different from ours.

Furthermore, in [22], the authors use taxi trajectories extracted from GPS data and public transit data generated from smart card to propose a data-driven framework to discover functional zones in a city. A review on how to extract information from triangulated mobile phone signals for different goals in spatiotemporal analysis and urban modelling has been presented in [23]. In fact, despite of the difficulty to apply these latent activity trajectories still remain in analyzing in-depth urban dynamics, a wide range of method using individual geographical trajectories data to infer spatial structure of urban functions has been presented in recent years. Specially, cell phone plays a more important role in modern daily life and has many advantages over other sensor data especially in emerging cities. What is more, data volume of the call data record is large which give more opportunities for mining macro features in large scale areas, such as cities or states.

## 5. Conclusion

This paper explores core districts in "polycentric" urban system based on map segmentation and human activity data. Urban area has been segmented to different districts as the basic research unit. Massive latent activity contexts in city have been abstracted from large scale call detailed record data in city. By proposing urban area as an inter-district network model of which edge is measured by temporal-spatial proximity among different districts stemming from those latent activity contexts, core districts of city have been identified by using a network-based method based on community detection and local centrality. In a case study of Wuhan, the role of these core districts has been analyzed. We believe it provides a feasible approach to uncover spatial structure of an urban system.

In terms of future work, we intend to focus on two aspects to understand "polycentric" urban system based on human activity data from complex network's perspective. Firstly, structure hole spanners plays an important role in information diffusion network. We consider the related work would help uncover the role of core districts in urban space. Then, the application of core districts in offline marketing and propagation control strategy would be interesting.

## Acknowledgements

The authors are grateful for the support from the National Science Foundation of China (5147 9159).

## References

1. A. Anas, R. Arnott, K.A. Small, Urban spatial structure, 1998, Journal of economic literature, pp. 1426–1464.
2. R.A. Becker, R. Caceres, K. Hanson, J.M. Loh, S. Urbanek, A. Varshavsky, C. Volinsky, A tale of one city: Using cellular network data for urban planning, 2011, IEEE Pervasive Computing, vol. 10, no. 4, pp. 18–26.
3. S. Jiang, J. Ferreira, M.C. González, Clustering daily patterns of human activities in the city, 2012, Data Mining and Knowledge Discovery, vol. 25, no. 3, pp. 478–510.
4. C. Roth, S.M. Kang, M. Batty, M. Barthlemy, Structure of urban movements: polycentric activity and entangled hierarchical flows, 2011, PloS one, vol. 6, no. 1, pp. e15923.
5. T. Louail, M. Lenormand, M. Picornell, C.O. Garcia, R. Herranz, E. Frias-Martinez, J.J. Ramasco, M. Barthelemy, Uncovering the spatial structure of mobility networks, 2015, Nature Communications, vol. 6, pp. 6007.

6. X.X. Nicholas, J. Yuan, Y. Zheng, Discovering urban functional zones using latent activity trajectories, 2015, IEEE Transactions on Knowledge and Data Engineering, vol. 27, no. 3, pp. 712–725.
7. Y. Wang, N. J. Yuan, D. Lian, L. Xu, X. Xie, E. Chen, Y. Rui, Regularity and conformity: Location prediction using heterogeneous mobility data, in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2015, pp. 1275–1284.
8. X. Liu, Q. He, Y. Tian, W.-C. Lee, J. McPherson, and J. Han, Eventbased social networks: linking the online and offline social worlds, in Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2012, pp. 1032–1040.
9. K. Pelechrinis, P. Krishnamurthy, Location affiliation networks: Bonding social and spatial information, in Proceedings of the 2012 European conference on Machine Learning and Knowledge Discovery in Databases - Volume Part II, 2012, pp. 531–547.
10. B. Marta, C. Gonzalez, C. Hidalgo, Understanding individual human mobility patterns, 2008, Nature, vol. 453, no. 7196, pp. 779.
11. Z. Chen S.M. Arisona, X. Hunag, G. Schmitt, Identifying spatial structure of urban functional centers using travel survey data: A case study of singapore, in Proceedings of The First ACM SIGSPATIAL International Workshop on Computational Models of Place, ser. COMP '13. ACM, 2013, pp. 28:28–28:33.
12. J.F. Mcdonald, The identification of urban employment subcenters, 1973, Journal of Urban Economics, vol. 50, no. 3, pp. 187–202.
13. A. Vasanen, Functional polycentricity: Examining metropolitan spatial structure through the connectivity of urban sub-centres, 2012, Urban Studies, vol. 49, no. 16, pp. 3627–3644.
14. S. Zhao, H. Wu, L. Tu, B. Huang, Segmentation of urban areas using vector-based model, in 2014 IEEE 11th Intl Conf. on Ubiquitous Intelligence Computing, pp. 412–416,2014.
15. S. Chen, H. Wu, L. Tu, B. Huang, Identifying hot lines of urban spatial structure using cellphone call detail record data, in 2014 IEEE 11th Intl Conf. on Ubiquitous Intelligence Computing, 2014, pp. 299–304.
16. D. Hu, B. Huang, L. Tu, S. Chen, Understanding social characteristic from spatial proximity in mobile social network, 2015, International Journal of Computers Communications and Control, vol. 10, no. 4.

17. M.W. Berry, M. Brown, A.N. Langvill, V.P. Pauca, R.J. Plemmons, Algorithms and applications for approximate nonnegative matrix factorization, 2007, Computational Statistics and Data Analysis, vol. 52, no. 1, pp. 155–173.
18. S. Adal, X. Lu, M. Magdon-Ismail, Local, community and global centrality methods for analyzing networks, 2014, Social Network Analysis and Mining, vol. 4, no. 1, pp. 1–18.
19. L.C. Freeman, A set of measures of centrality based on betweenness, in Sociometry, 1977, pp. 35–41.
20. E.W. Dijkstra, A note on two problems in connexion with graphs, 1959, Numerische Mathematics, vol. 1, no. 1, pp. 269–271.
21. C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martino, R. Claxton, S.H. Strogatz, Redrawing the map of great Britain from a network of human interactions, 2010, Plos One, vol. 5, no. 12, p.e14248.
22. V.W. Zheng, Y. Zheng, X. Xie, Q. Yang, Towards mobile intelligence: Learning from gps history data for collaborative recommendation, 2012, Artificial Intelligence, vol. 184, pp. 17–37.
23. Y. Zheng, Trajectory data mining: An overview, 2015, Acm Transactions on Intelligent Systems and Technology, vol. 6, no. 3, pp. 1–41.