SUMMARY Understanding the structure and evolution of spatial-temporal networks is crucial for different fields ranging from urbanism to epidemiology. As location based technologies are pervasively used in our daily life, large amount of sensing data has brought the opportunities to study human activities and city dynamics. Ubiquitous cell phones can be such a sensor to analyze the social connection and boundaries of geographical regions. In this paper, we exploit user mobility based on large-scale mobile phone records to study urban areas. We collect the call data records from 1 million anonymous subscribers of 8 weeks and study the user mobility flux between different regions. First we construct the urban areas as a spatial network and use modularity detection algorithm to study the intrinsic connection between map areas. Second, another generative model which is widely used in linguistic context is adopted to explore the functions of regions. Based on mobile call records we are able to derive the partitions which match boundaries of the administrative districts. Our results can also catch the dynamics of urban area as the basis for city planning and policy making.

keywords: mobile call record, urban dynamics, topic model, location based network

1. Introduction

Human mobility has attracted substantial attentions of researchers for a long time. When people live and work in urban area, they leave behind trails which can be used to analyze the people’s living habits, social relationship and urban dynamics. While the traditional IoT (Internet of Things) technology such as wearable sensors, PAN (Personal area network), can be used for sensing human mobility, they are usually limited to small scale applications. However, recent advances in smartphone and big data are widely used for social sensing and analyzing human behavior, which extend the concept the IoT. Therefore in this paper, we exam a case of such new IoT applications using mobile phone call data.

More particularly, our work is aimed at analyzing the connections of urban area based on the flow of mobile phone traffic. We would like to examine whether telephone habits divide the urban area into districts or functional regions.

The ubiquitous sensors placed all around can reflect people’s mobility and the evolution of dynamic city, such as Internet usage, bank note transactions [1], GPS equipped vehicles [2] and city smart cards [3]. They store our timestamped coordinates when we use them. Consequently, location history data can be collected and analyzed to provide insights on social motivation. A few of different types of data has been used in human mobility studies, which lead us to develop modern applications and services in different ways from traditional approaches. For example, by studying the traffic flows of group from one area to another we can detect the traffic blocks or anomalies. Also the temporal-spatial structure can help to discover the activity patterns of students and workers based on travel surveys [4].

Recently, location acquisition technologies that are able to record the location history of individuals have become prevalent. With the rapid development of wireless communication technologies, mobile phone devices are widely used in our daily life while they funnel in an explosive amount of information among a large scale population. However, methods like GPS require end-user consent for the acquisition of monitoring and they are restricted in the size and deployment due to the cost. As a result of that researchers are more prone to study the GSM data. The mobile equipment can avoid both privacy risk of fine grained locations and battery limitations of continuous GPS samplings. The study of mobile phone records has brought forth many practical applications such as advertising, urban planning and traffic route recommendation.

The large amount of mobile call data can be a proxy to reveal social interactions and outline of urban area. Researchers study the mobility from two perspectives: individual and group. The former is inclined to uncover the human mobility patterns for personal recommendation [5], [6] while the latter may sense the surroundings from complex group behaviour. According to recent studies [7], [8], the mobile communication flows can be conceived as a large spatial network. It is shown that human mobility is closely coupled to the geography where people live in. In this paper, we study the mobility of groups of people to analyze the spatial-temporal correlations of different regions in urban area.

In this paper, we analyze mobile phone records collected from a local telecommunication operator which covers approximately 30 percent of the population in urban area in central China. The dataset provides virtual subscriber ID, time, duration, as well as identifiers of serving base station within a period of 8 weeks. We construct the whole urban city as a large spatial network and study the transitions of people in the following two approaches.

First, we adopt the modularity which is widely used in complex networks to partition the whole urban area. The proposed method aggregates transitions between differ-
ent regions and regards the counts of the transitions as the weights of links in the complex network graph. The result indicates that the geographical partitions overlap the administrative districts to a large extent.

Temporal information is another main aspect deserving research in human mobility study. Recently spatial-temporal feature has becoming an important issue to investigate the dynamic evolution of human mobility [9], [10]. So next, we use a topic based algorithm inspired by the similarity with linguistic context by considering the direction of transitions and temporal dimension features to explore the functional partitions of urban area.

The rest of this paper is organized as follows: In Sect. 2 we summarize the related works on urban study. In Sect. 3 we formulate the problem and give the framework of our system. In Sect. 4, we introduce the modularity algorithm and topic model for urban study. Results are explicitly analyzed in Sect. 5 followed by discussion and future work in final Sect. 6.

2. Related Work

A good understanding of human mobility patterns can yield insight into the dynamics of urban city, thereby enabling various applications for better living environment. With the evolution of urbanism and socialization, urban dynamics are becoming more complex due to the flexibility in human’s mobility patterns and social connections. As a consequence, it needs new methods and models to obtain a better understanding of urban area.

Due to the constraint of data acquisition and computer capability, traditional ways of human mobility studies are lack of efficiency and in small scale in population and region. Recently, modern technology has brought the opportunity to the understanding of urban area in terms of breadth and depth. With the widespread of location acquisition technology, human mobility has attracted attentions from many fields of studies such as social sensing, traffic engineering, and urban planning. For example, Pan et al. [11] detected the anomaly events based on trajectories of vehicles and social media. Traffic congestions sources can be inferred by studying the mobile phone records [12]. Base on their findings efficient strategies can be designed to alleviate traffic congestion to a great extent.

As we know, urban planning plays an important role in our daily life. Geographical partition of spatial area into subdivisions is essential for the administrative management, resource allocation, commercial recommendation and so on. Researchers have used various data sources to study the geographical partition problem. For example, in early time [13] used the landline telephone records to reflect the outlines of the country’s boundaries. In the meanwhile, Thiemann et al. [14] used records of bank notes circulation as a proxy of human movements to find the partial overlap between detected subdivisions and territorial state borders. Yin et al. [15] discovered similar geographical areas by studying the GPS-associated texts in social media.

However the above researches are in global or country scale. As majorities are spending most of their time in urban area, it is more realistic to explore the dynamics of urban area. For example, Yuan et al. [16] discovers functional regions in urban city combing GPS trajectories of taxicabs with points of interests. Their work may bring commercial applications such as geo-marketing. Due to the flexible human mobility in urban area, it is a difficult task of geographical partitions in such limited scale.

As we know, mobile wireless communication is the most widespread information technology these years. Such mobile phone records which catch the mobilities of large-scale population over several years, is in favor with many researchers as the proxy for human activity analysis [17] and urban study [18]. Results exploited from the mobile records can facilitate determining where to deploy infrastructure and how to reduce traffic burden. We refer the reader to [19] for details of advantages over other datasources.

In this paper, we aim to study the geographical partitions in urban area by using the mobile phone records. Our study construct a spatial based on human trajectories and explore the spatial connection of different regions. Furthermore we study the urban area using generative model as an ancillary to unveil the regions with similar functions. Compared with past researches, our contributions are as follows:

- We study the mobile transitions graph to analyze the geographical partitions and test it with administrative boundaries.
- We extend former method which solve the partition problem with community as weighted network and use the topic model as a good solution of learning urban area.
- Our results can catch the dynamics of urban city and help make proposals for urban planning in the future.

In the next section, we will formulate our problem and give explicit procedures of our work according to the framework presented in Fig. 1.

3. Framework

In this section, we develop a framework for region partitions by exploiting the user mobility (see Fig. 1). First we briefly introduce the features of our data set with proper data processing, and then segment the urban area into grid regions. Next we give a scenario in which trajectories are extracted from mobile records of individuals. Lastly we explain our network model as well as some definitions.

3.1 Data Set Description

The analysis described in this paper relies on logs collected from a local mobile operator. The dataset contains Call Detail Records (CDRs) of two months in recent years. Subscribers’ identifications are anonymized to subject to the strict data privacy agreements. Though the operator only has a market share of 30% in local mobile communication.
Assuming that the population is distributed uniformly in geography of urban city and service provided by particular operator is homogeneously penetrated with other operators, we make a reasonable deduction that our dataset can reflect the dynamics of all users’ mobility.

Whenever a subscriber initiates a voice call or sends a text message, a CDR is recorded in the repository by mobile operators for billing purpose. There is rich information in CDRs among which we only focus on the fields listed in Table 1.

Due to the existence of couriers, salesmen and telemarketers, the results may largely affected by their activities. So we need to ignore these users as outliers prior to the data analysis. Through statistical analysis we find such users either have a large range of movement with uniform distribution or have small ratio of incoming-calls and total calls [20]. So we successfully remove such subscribers considering the characteristics of their spatial distribution and calling patterns. We also limit our study in urban area covering the size of 17 by 27 square kilometers. After data filtering, there are approximately one million users and 800 million transactions in our database. Such a large scale source of data brings us unprecedented opportunity to study the city network compared with most traditional surveys in terms of population and time duration. After preprocessing we summarize the dataset with statistics in Table 2 as below:

### 3.2 Urban Area Segmentation

According to different resolutions, urban area segmentations are ranging from cell tower coverage, zip codes or other square-like spatial units. To reflect the dynamics of different spatial regions as well as reduce localization estimation noises due to the mobile traffic load balance. We adopt the method which divides the whole region into raster size [21] as a compromise. To get a finer granularity, urban area is split into similar uniform units $V$ of 0.01 degree of latitude and longitude each covers the area of $1.11 \times 0.94$ km$^2$. We term such a segmented area as a Grid Regions (GR) in the following context.

### 3.3 Scenario

When we extract the CDRs of a particular user and order them by time, the discrete samples can be concatenated as the representation of people’s life traces. A trajectory $T_r$ is a spatial-temporal sequence of way points $(l, t)$ ordered by time, where $l$ is the geographical coordinate and $t$ is the time stamp.

Here we give an example of how to form people’s trajectories in Fig. 2. When a mobile call event happened we draw a red dot which contains userid, cellid, timestamp and so on according to the serving cell tower nearby. We draw the users’ trajectories depicted as green line by concatenating the dots in time order. The whole urban area are segmented into raster regions according to the method introduced in Sect. 3.2. Then we calculate the number of distinct users who travel across different regions in unit time interval.
3.4 Spatial Network Model

People’s movements across grid regions can be a proxy to reflect the relationship between spatial areas. Generally we think that two areas are in close contract if there is large flow of population in between. Here we build a model based on human mobility and give relevant definitions as below.

In urban studies, various networks such as road network, contact network and transportation network can be defined to form the graphs which represent the connections between human or map regions. Characterizing these networks’ topological patterns can yield insight into the structure and function of urban city and society. In this paper, we will use the CDRs to reflect the human mobility and infer different geographic partitions to facilitate our understanding of urban dynamics. We explain a few concepts as follows:

**Definition 1 (Mobility Transition):** A mobility transition represents the displacement of people. It is extracted from a trajectory and contains \( l_i, t_i \) and \( l_j, t_j \) which represents the location and timestamp of start and end separately.

Here we construct a weighted graph \( G(V, E) \) in which nodes \( V \) indicate the geographic regions and links \( E \) the connections between pairs of nodes. To analyze the relationship between different geographic regions, we measure the weight of a link \( (u, v) \) by the aggregates of mobility transitions between region \( u \) and \( v \) in unit time. A link can be either undirected or directed. Generally two regions are more spatially coupled with the increasing of transitions in between. So we may cluster the highly connected component regions into groups or communities.

In the above, we construct a spatial network from the perspective of spatial coherence. But there is also potential to uncover the area with similar functions which maybe the cause for driving people to behavior differently. In our daily life, where people arrive or leave for largely depends on the time and function of regions. For example, people may go to restaurants after work when at supper time or go to office places from residence in the morning. Furthermore, two regions may have similar functions if crowds go to the two regions from similar functional regions or leave for similar ones. So the departure, arrival locations and time of mobility transitions correlate the people’s intentions of spatial activities. Here we give the definition of mobility items from transitions regarding the spatial and temporal features.

**Definition 2 (Mobility Item):** Mobility items are extracted from mobility transitions regarding the direction of transitions. For specific location \( l \), its departure mobility item is \( (l, l_i, t_i) \) and arrival mobility item is \( (l, l_j, t_j) \).

Our issue is how can we find the relation or structure of region groups something like communities or functional regions? To solve the problem, we will illustrate two methods in the next section.

4. Algorithms

We use two methods to study the urban region partition from different perspectives as mentioned in the framework introduction. We first discuss using modularity based community detection method to cluster Grid Regions for region partition. Then we explore the method of using topic model to discover a region’s function.

4.1 Modularity Based Community Detection

Mobility studies have recently become a hotspot in location based social network. It can help analyze the evolution of the network dynamics, the diffusion of epidemics in crowd and behaviors of anomalous events. In Sect. 3, a spatial network is constructed by using human transitions which capture interactions between spatial regions. It is well known that some regions may be more highly connected to each other than to the others. We refer to such set of nodes as community. Graph theory has been popularly used in detecting the communities of spatial network. Thanks to the researches in complex network, a few approaches such as KL, k-clique and modularity can be used to discover rich community structure.

Modularity based methods are widely used in complex network such as community detection in social network [22] and virus infection in genetic network. We look on the spatial area as a weighted network and use the modularity based algorithm to study urban area. Here the urban area are divided into a few communities without overlap. It is between −1 and 1 that the higher value of the modularity corresponds to good divisions of the network into subsets. The presence of communities which reflect the inequality in terms of ability to convey mobility flux can help the government design strategies for traffic and resource allocation.

Precise solution of the maximizing modularity is known to be a NP-hard problem which is computationally intractable [23]. According to the recent studies in social network community detection, a few algorithms have been proposed to obtain good performance in the optimization. Here we employ the Louvain algorithm MOA (Modularity Optimization Algorithm) introduced in [24] due to its fast speed and competitive accuracy in comparison with other community detection algorithms such as greedy optimization proposed in [25].

The modularity \( Q \) is defined as:

\[
Q = \frac{1}{2m} \sum_{i,j} \left[ A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j)
\]  

(1)

where \( A_{ij} \) represents the number of individuals travelling between two locations \( i \) and \( j \), \( k_i = \sum_j A_{ij} \) is the sum of the mobility flux attached to location \( i \), \( 2m = \sum_{ij} A_{ij} \), \( c_i \) is the community to which location \( i \) is assigned. \( \delta \) function \( \delta(i, j) \) is 1 if \( i = j \) and 0 otherwise.

Here we only consider the symmetric network in which
link weight is defined as aggregated flow of movements, so \( A_{ij} = A_{ji} \) when grid \( i \) and \( j \) are different. By definition we also don’t consider the loop edges from grid regions to themselves, so the diagonal items \( w_{ii} \) are always 0.

In our experiment, to simply our computation we add a threshold \( \text{freq}_\text{th} \), so the \( w_{ij} = 0 \) if link between node \( u \) and \( v \) is less than \( \text{freq}_\text{th} \), while \( w_{uv} = \text{links}/10 \).

In the first phase of MOA [24], the ordering of the nodes has influence on the modularity that is attained. So we can only get local maxima of the modularity each time we run the experiment. To get a better result, we choose the division with the maximum of the modularity.

The process of the MOA algorithm can be represented as a dendrogram which is a hierarchy graph like tree. Cutting through the dendrogram at different levels can get divisions into larger or smaller numbers of subsets. In our experiments, we get three levels with the final layer containing exactly 8 nodes. Due to lack of space, we just show the output of the last level.

We use the network analysis tool [26] which integrates MOA to produce the result. We run the experiment several times and obtain the maximum modularity values in the range \( 0.610 \pm 0.010 \). Among the experiments we choose highest modularity 0.620 as our result. We apply the MOA algorithm in the spatial network and the connectivity structure of urban area is visualized in Fig. 3. In the figure, each color corresponds to a community and radius of circle represents the activities level around the grid region. It is shown that the zones with bigger circles correspond to commercial zones which have larger in and out flows. We can also find that most active areas are not always in the geographical center of the communities. The link width measures the closeness of connection. A few links have a large span of long distance. We can clearly see that the whole network is divided diagonally and forms two main parts. However it gives us an intuitive understanding of urban partition, we will analyze with geographical environment in more detail in the next section.

4.2 Topic Model

Recently the dynamic mobility has become an important part of researches. One of the main concerns is that the changes are based on the time-varying mobility patterns. While human flow between different areas is changing along the time, the modularity approaches above may not reflect the temporal dynamic changes of the city. So we add temporal dimension to further study the correlation of urban regions.

The development of urbanization lead to different spatial regions to satisfy people’s needs in urban lives. Unlike aforementioned clustering method which can represent the range of the citizens’ daily activities, we also consider these regions have different functions, for example campus, entertainments, residential areas or commercial districts. People may go to different types of regions according to their lifestyles. However the function of a region may vary in different time periods of a day, so we consider there is a probability distribution of functions in each region. Considering the similarity, we draw an analogy between geographic regions and documents. As the document has multiple topics, the geographical region has multiple functions. The topic expressed by words is just like the region functionality being represented by mobility items. So we use topic model to study the spatial-temporal aggregate mobility flows.

First we give a brief introduction of the topic model with some formula definitions for clarity and second we map our model into topic discovering problem in semantic context.

Topic model is widely used in document context assuming a generative probabilistic layer named topic between the documents and words [27]. That is to say, the documents are composed of random mixtures over latent topics and topic \( Z \) is defined to be a distribution over a fixed vocabulary of terms \( w \). Assume there are \( K \) topics, \( \theta \) is the size of the vocabulary, \( \alpha \) is a positive \( K \)-vector, \( \beta \) is a \( K \times V \) matrix and \( \phi = \beta_k \) is a distribution over the vocabulary. Let \( \theta_d \) be the topic mixture for \( d \)th document where \( \theta_{dk} \) is the topic proportion for topic \( k \) in the \( d \)th document. For each document in the corpus, generative process can be described as follows:

1. Randomly pick a distribution over topics: draw \( \theta_d \sim \text{Dir}(\vec{\alpha}) \), which determines how topics are mixed together in forming each document. Here, \( \text{Dir}(\cdot) \) is the Dirichlet distribution.
2. For each word in the document, a single topic \( z_n \) is chosen according to the selected distribution \( \theta_n \).
3. Choose a word \( w_n \) from \( \text{Dir}(\cdot) \), which is a conditional probability of a specific word \( w_n \) conditioned on the unobserved topic \( z_n \), \( \beta \) is the topic distribution over words.

The latent topic model can be shown with a graph model in Fig. 4.
Next we define our topic model in analogy with topic discovering in text documents. We regard the grid regions as spatial documents and the two-tuple mobility term \((r; t)\) as the basic word. Here, the whole day is divided into 12 splits each with equal time interval of 2 hours. Note that the time intervals are not necessarily equal. Because it may weaken the features in low activities. For example, during the time range of 0 to 6 a.m., people’s activities densities are weak if dividing the time into 3 splits.

Apparently, user’s daily mobility has statistical regularity which is in a cycle of 7 days, so we plot the activities map at weekly scale in Fig. 5. From the figure, we find that activities from Monday to Friday are similar in shape and amplitude. But activities in weekdays are obviously different from those in weekends. Because in our daily life, most people work on weekdays while on weekends, people may rest or engage in entertainment. In addition the locations when performing the activities in weekends are also different from those in weekdays. So we divide a week into weekdays and weekends [28] to reflect the differences of mobility patterns.

For specific grid region \(r\), there are two different mobility items: origins and destinations. Origins are the mobility items starting from other regions and destinations are the mobility items to other regions. In our model, both two types of mobility items contribute to revealing the relation of regions. By distinguishing such two types of mobility, the spatial network is considered as a bidirectional network. Such consideration which was neglected in the above MOA is important to analyze the population flow between different regions.

So, there are two types of mobility items here: Arriving and Leaving. Intuitively in weekdays people usually go for work in the morning and return home in the evening. Furthermore, two places may have similar function if they are the origins or destinations of most people from similar regions at particular time interval. For each grid region, we aggregate all the corresponding arriving and leaving transitions and count the number of similar transitions as frequency. So each region consists of a series of mobility items with frequency. Then the model can be estimated using Bayes inference [29].

Finally we obtain a topic distribution of each cell region \(\theta_{grid} = (\theta_1, \theta_2, \cdots, \theta_k)\) where subscript grid denotes the grid label. To know the relationship between different grid regions, we perform the Partitioning Around Medoids (PAM) algorithm [30] on the k-dimensional vector \(\theta_{grid}\). To get satisfied classification result, we use silhouette index [31] to judge the cluster number. It maximizes inter-cluster distances while minimizing the intra-cluster distances. Here we perform experiments with different cluster number \(k\) multiple times and choose the results with maximum average silhouette value.

5. Result and Analysis

The urban city we study is in central of china. A long river (Yangtze River) and one of its branches traverse the whole city and split it into three main parts. The city has many lakes and the road network is not as the regular like boxy area. With the development of the city, it is now comprised by 7 administrative districts according to city planning bureau\(^1\). To better analyzing the geographical partitions, we obtain the administrative map boundaries from Baidu Map API and redraw it on top of the Google satellite map (see Fig. 6). Administrative districts are plotted with different colors.

To give further exploration, we also integrate the MOA in our own system to visualize the map cohesive regions. As is shown in Table 3, we obtain four layers: layer 0 is the initial ground where each grid region constitutes a cluster, in layer 1, 33 clusters are detected, and in layer 2, 9 clusters. Layer 3 is with 8 clusters. Because modularity can be no further improved, layer 3 with 8 clusters is the final result. We also get a mobility score with 0.62035 which verify the result in Fig. 3.

By comparing the Fig. 7 (a) and Fig. 6, we find that MOA result resembles administrative districts in the left part. They both contains 4 partitions. When considering the right part, we see MOA splits the Hongshan districts into 3 partitions while the district in Fig. 6 is a very large region.

| Level | Clusters | Score |
|-------|----------|-------|
| 0     | 354      | 0.518 |
| 1     | 33       | 0.619 |
| 2     | 9        | 0.620 |
| 3     | 8        | 0.620* |

*Actually the modularity of level 3 is a little larger than level 2, the scores look the same as we round to the nearest thousandth.

\(^{1}\)The boundaries of districts are from website: http://en.wikipedia.org/wiki/Wuhan#Administrative_divisions
Then we draw the partition results using the LDA methods. By comparison between the boundaries of LDA partitions and administrative districts, a high degree of similarity can also be found. Such as partitions 1, 2 and 3 in LDA are detected.

5.1 Evaluation Metric

To evaluate how well our result match the political districts, a classical measurement index $B$ [32] is used to quantify the overlap of partitions. When the two clusters are completely unrelated the index is strictly larger than zero and the perfect match between two partitions will have $B = 1$. We calculate the overlap ratio of our results with respect to the administrative districts using the $B$ index. We get $B = 0.477$ for the MOA and $B = 0.419$ for the LDA. To give a baseline, we randomly shuffle the regions given the administrative partitions and get $B = 0.185$ which is much less than our results. It is shown that our results can partly reflect the administrative planning.

5.2 Analysis

Through observations, we discover that the geographical partitions are consistent with the city planning in large part of the urban area. There still exists some deviations between our results and administrative districts. In reality, the Hongshan District is too big and may not function well to keep up with the development of urban city. Both of partitions $\Delta$, $\Phi$ in Fig. 7 (a) and partition $\Psi$, $\Theta$ in Fig. 7 (b) geographically coincide with new emerging partitions named Nanhu and Guanggu separated. Now the two partitions are not outlined as exclusive districts but they are actually qualified to be competent divisions compared with other administrative sub-divisions. From the satellite map, we can clearly see that they are in the area of the largest Hongshan district. In Nanhu area, there is a large area of water in the middle of the partition. In recent years, it indeed absorbs a large amount of population and plays an important role in the evolution of urban city. The other new emerging area named Guangu (Optics Valley of China) is particular partition zone where a lot of campuses and hi-tech companies are located. It is coincided with the blueprint of Wuhan East Lake High-Tech Zone. As MOA and LDA both find such hidden territorial regions, we suggest that new districts to be planned with supporting policy.

There is higher degree of similarity in the comparison of MOA and LDA with administrative districts. Intuitively, we think it is because they both consider the mobility flux between different regions in spatial dimension. But there are also some distinctions between two methods. It seems that

---

*Optics Valley of China Planning: http://en.wehdz.gov.cn/planning/1722.jhtml*
MOA does better in matching the boundaries of administrative districts. But we can also obtain additional results by using the LDA.

As the generative model can reflect the topics of a document, the LDA method here can also reflect the functional regions. For example, in partition ⑤ there are many schools which are indicated by the school icons and our results successfully reveal the education area. Compared with MOA, partition ② is a new emerging partition which largely match Qingshan District in Fig. 6. It represents the industry area because there exist a large steel mil with many factories scattered in this area. Workers’ residence is far away from factories and they leave home for work in the morning and back home in the evening. LDA discovers the partition by considering the spatial-temporal features of workers’ commuting. We also find some regions with rare people in dark yellow along the left edge. These areas may be parks or under development. For example, region ④ and ⑥ are parks while ② is under construction when the dataset was collected. Because a large proportion of mobile transitions are in short distance, the functional partitions are coupled with spatial proximity. Our result based on topic model can only find a fraction of regions with similar functions. To improve the performance, we are going to deduce the OD (Origin-Destination) matrix to study the land usage of human mobility activities in the near future.

6. Conclusion

Administrative planning may affect people’s daily life, but as the city evolves, people’s life pattern may change. From the beginning, city’s planning may affect the activities of people living in, but as time goes on, administrative boundaries may vary according to the changes of users’ lifestyles. Prevalent mobile technology offers us an opportunity to study the human mobility in a cost-efficient way based on large scale of population.

In this paper, we analyzed the mobility of mobile phone subscribers based on their spatial and temporal features. We have used MOA and LDA to detect the spatial connection of urban city regions and evaluated their performances with the real administrative districts. It is shown that these methods can be adopted a useful tool to tackle issues such as detecting optimal homogeneous territorial units. Our results can be of guidance for urban planners, analysts in understanding homogeneous sub-divisional systems.

We find that both MOA and LDA methods are capable of reflecting the administrative districts at a certain degree. The two methods show similar output of partitions but with obvious differences. It seems that the result by MOA matches well with administrative districts than LDA. By the way, both algorithms happen to uncover another new district which is not shown on the administrative planning map. Furthermore, the LDA can find somewhat functions of regions which is coherent with reality. In this study we have shown that the aggregates of mobile users can be proxy to differentiate the functions of regions. Because the origin-destination can be hard to extracted, the regions’ function features are weak than other OD-based study. Our research result can be improved by using such formatted data or prior acknowledge.

In the present work, there is slight deviations with the reality, indicating that there is room for improvement. Such bias might be due to limited information and overlook the motivation of users and coarse resolution of localization. In addition, the precision of B has been limited by the lack of O-D (Origin-Destination) data. For future research, email records will be added as a complement to the study of different types of activities. Our data has higher degrees of uncertainty inferring trip purposes. We will investigate this in more detail. We hope that our results can be of an alternative for studying human mobility patterns.

Acknowledgments

This work is supported by the Young Scientist Fund of National Natural Science Foundation of China (Grant No.61202303) and the National Science and Technology Major Projects of China (Grant No.2012ZX03001035-003).

References

[1] D. Brockmann, L. Hufnagel, and T. Geisel, “The scaling laws of human travel,” Nature, vol.439, no.7075, pp.462–465, Jan. 2006.
[2] Y. Zheng, Y. Liu, J. Yuan, and X. Xie, “Urban Computing with Taxi-cabs,” Proceedings of the 13th International Conference on Ubiquitous Computing, UbiComp ’11, pp.89–98, New York, NY, USA, 2011.
[3] S. Hasan, C. Schneider, S. Ukkusuri, and M. González, “Spatiotemporal patterns of urban human mobility;” J. Statistical Physics, vol.151, no.1-2, pp.304–318, 2013.
[4] S. Jiang, J. Ferreira, and M.C. Gonzalez, “Discovering urban spatial-temporal structure from human activity patterns;” Proc. ACM SIGKDD International Workshop on Urban Computing - UrbComp ’12, p.95, New York, New York, USA, 2012.
[5] M.C. González, C.a. Hidalgo, and A.L. Barabási, “Understanding individual human mobility patterns,” Nature, vol.453, no.7196, pp.779–782, June 2008.
[6] C. Song, Z. Qu, N. Blumm, and A.L. Barabási, “Limits of predictability in human mobility,” Science, vol.327, no.5968, pp.1018–1021, Feb. 2010.
[7] A. Montis, S. Caschili, and A. Chessa, “Commuter networks and community detection: A method for planning sub regional areas,” The European Physical Journal Special Topics, vol.215, no.1, pp.75–91, 2013.
[8] M. Barthélémy and M. Barthelemy, “Spatial networks,” Physics Reports, vol.499, no.1-3, pp.1–101, Feb. 2011.
[9] P. Grindrod, M.C. Parsons, D.J. Higham, and E. Estrada, “Communityability across evolving networks,” Phys. Rev. E, vol.83, no.4, 46120, April 2011.
[10] C.a. Hidalgo and C. Rodríguez-Sickert, “The dynamics of a mobile phone network,” Physica A: Statistical Mechanics and its Applications, vol.387, no.12, pp.3017–3024, May 2008.
[11] B. Pan, Y. Zheng, D. Wilkie, and C. Shahabi, “Crowd sensing of traffic anomalies based on human mobility and social media,” Proc. 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, SIGSPATIAL’13, pp.344–353, New York, NY, USA, 2013.
[12] P. Wang, T. Hunter, A.M. Bayen, K. Schechtner, and M.C. González,
“Understanding road usage patterns in urban areas,” Scientific reports, vol.2, pp.1–5, Jan. 2012.

[13] C. Ratti, S. Sobolevsky, F. Calabrese, C. Andris, J. Reades, M. Martinu, R. Claxton, and S.H. Strogatz, “Redrawing the map of great Britain from a network of human interactions,” PLoS ONE, vol.5, no.12, e14248, 2010.

[14] C. Thiemann, F. Theis, D. Grady, R. Brune, and D. Brockmann, “The structure of borders in a small world,” PLoS ONE, vol.5, no.11, e15422, 2010.

[15] Z. Yin, L. Cao, J. Han, C. Zhai, and T. Huang, “Geographical topic discovery and comparison,” Proc. 20th International Conference on World Wide Web, WWW ’11, pp.247–256, 2011.

[16] J. Yuan, Y. Zheng, and X. Xie, “Discovering regions of different functions in a city using human mobility and POIs categories and subject descriptors,” Proc. 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD ’12, pp.186–194, 2012.

[17] G. Di Lorenzo and F. Calabrese, “Identifying human spatio-temporal activity patterns from mobile-phone traces,” 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), pp.1069–1074, 2011.

[18] F. Calabrese, M. Colonna, P. Lovisolo, D. Parata, and C. Ratti, “Real-time urban monitoring using cell phones: A case study in Rome,” Trans. Intell. Transport. Sys., vol.12, no.1, pp.141–151, 2011.

[19] C. Ratti, R.M. Pulselli, S. Williams, and D. Frenchman, “Mobile landscapes: Using location data from cell phones for urban analysis,” Environment and Planning B: Planning and Design, vol.33, no.5, pp.727–748, 2006.

[20] Z.Q. Jiang, W.J. Xie, M.X. Li, B. Podobnik, W.X. Zhou, and H.E. Stanley, “Calling patterns in human communication dynamics,” Proc. National Academy of Sciences, vol.110, no.5, pp.1600–1605, 2013.

[21] J. Candia, M.C. González, P. Wang, T. Schoenharl, G. Madey, and A.L. Barabási, “Uncovering individual and collective human dynamics from mobile phone records,” J. Physics A: Mathematical and Theoretical, vol.41, no.22, 224015, June 2008.

[22] M.E.J. Newman and M. Girvan, “Finding and evaluating community structure in networks,” Phys. Rev. E, vol.69, no.2, 26113, 2004.

[23] U. Brandes, D. Delling, M. Gaertler, R. Görke, M. Hoefer, Z. Nikoloski, and D. Wagner, “On finding graph clusterings with maximum modularity,” vol.4769, pp.121–132, 2007.

[24] V.D. Blondel, J.L. Guillaume, R. Lambiotte, and E. Lefebvre, “Fast unfolding of communities in large networks,” J. Statistical Mechanics: Theory and Experiment, vol.2008, no.10, P10008, 2008.

[25] M.E.J. Newman, “Fast algorithm for detecting community structure in networks,” Phys. Rev. E, vol.69, no.6, 66133, 2004.

[26] M. Bastian, S. Heymann, and M. Jacomy, “Gephi: An open source software for exploring and manipulating networks,” 2009.

[27] D.M. Blei, “Probabilistic topic models,” Commun. ACM, vol.55, no.4, pp.77–84, April 2012.

[28] E.B. Fowlkes and C.L. Mallows, “A method for comparing two hierarchical clusterings,” J. American Statistical Association, vol.78, no.383, pp.553–569, Sept. 1983.

Feng Xiang received B.S degree in Electronics and Information Engineering from Huazhong University of Science and Technology in 2008. Now he is a Ph.D student in Huazhong University of Science and Technology. His research interests include social network, urban computing and machine learning.

Benxiong Huang is a professor in Department of Electronic and Information Engineering in Huazhong University of Science of Technology. He is the laureate of Fund for Distinguished Young Scholar of Hubei province, senior member of China computer association, Commission of Young Computer Scientists & Engineers Forum (YOCSEF). His research interests cover communication system, next generation mobile internet, signal processing and social computing.

Lai Tu received the B.S in Communication Engineering and Ph.D. degree in Information and Communication Engineering from Huazhong University of Science and Technology, Wuhan, China, in 2002 and 2007 respectively. From 2007 to 2008/12, he worked as a postdoc fellow in the Dept. of EIE. in Huazhong University of Science and Technology. He is now working in the Dept. EI. in Huazhong University of Science and Technology. His research areas include network communication, mobile computing and networking.

Duan Hu received B.S degree in Electronic and Information Engineering from Huazhong University of Science and Technology in 2007. She studied as a visiting student in Computer Engineering of Wayne State University, Detroit, Michigan from 2010–2011. Her research interests include cloud computing, data mining, social network.