

Detection and Analysis of Urban Area Hotspots Based on Cell Phone Traffic

Xiaoqing ZUO

Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming, China
Email: zuoxq@163.com

Yongchuan ZHANG

Faculty of Land Resource Engineering, Kunming University of Science and Technology, Kunming, China
Email: zhangyongchuan1@hotmail.com

Abstract—This paper is to explore new ways to better understand urban system emphasizing on detection and analysis of urban area hotspots through cell phone traffic data. Firstly, according to the characteristics of GSM network, Voronoi cellular network is introduced to determine the service area of the base station. Then two visualization methods are discussed through analyzing the distribution of cell phone traffic data. Next, activity patterns of residents in urban areas are discussed in macro-perspective. Additionally, an algorithm to detect hotspots of urban areas is proposed, and through the usage of the algorithm, many hotspots are found, and then, some typical hotspots are analyzed. Meanwhile, an experiment is conducted based on real cell phone traffic data during the month on February 2011, covering the entire area of Kunming, China.

Key words —cell phone traffic, urban visualization, hotspot detection, urban dynamics, human activity

I. INTRODUCTION

The development of science and technology, especially in the area of wireless communication, has greatly changed the way people live, and also changed the structure of city [1]. So people have to change the view of city. As Batty [2] argued that it is possible to conceive of cities as being clusters of 'spatial events', which take place in time and space, where the event is characterized by its duration, intensity, volatility, and location. And interactions may be in time and space between events, leading to clusters and other aggregations. Therefore, understanding of the urban system cannot remain at the stage regarding the urban system as a collection of many buildings, but extends to the activity of urban residents. The traditional methods, for example, surveying, cannot meet this demand, especially when confronting the distribution of the residents' activities and urban dynamics [3, 4]. Due to the high penetration of cell phone ("The operation conditions of the communications industry in June 2011" [5], which is published by the

Industry Ministry of China, shows that by the end of May 2011, the cell phone subscriber had reached as much as 900 million in China), cellular network (generally GSM network) provides a new opportunity to monitor dynamics and detect hotspots in urban areas.

In recent years, some scholars have explored this field by taking cell phone as detection device for urban dynamic monitoring [6-12]. They can be divided into micro- and macro- aspects. At the micro level, the cell phone is used to track individual subscribers by the means of using mobile positioning technology to locate the cell phone for the purpose of extracting traffic information and the way people communicate with each other and so forth through the subscriber's cell phone records [7, 8]. However, there is a serious drawback of privacy violation in this approach leading it not easy to its popularity. At the macro level, by analyzing aggregate data which is collected at the base station level, the activity pattern and the interesting lifestyle of the overall resident in urban area are revealed. And a good example is that, the collaboration between SENSEable City Laboratory of MIT and Telecom Italy in 2006 created a new era of exploring urban dynamics [4, 9 and 10]. The aggregate data can be more readily available, because it will not result in violation of personal privacy. But most of these studies remain in the display of urban structures, and there is no in-depth study of the resident's activity pattern and hotspots in city which many organizations and individuals may be interested in.

This paper briefly first of all introduces the Cell-ID positioning method, then, Voronoi cellular network is introduced to roughly determine the base station's service area. And then, two visualization methods are discussed through analyzing the distribution of cell phone traffic data. Afterwards the activity patterns of residents in urban is discussed from the macro-perspective through the visual view. Additionally, an algorithm to detect the hotspot of urban is proposed, and by using the algorithm many hotspots are found and the typical are analyzed. Meanwhile, an experiment is conducted by examining real cell phone traffic data during the month on February 2011 covering the entire area of Kunming, China.

Manuscript received September 14, 2011; revised November 25, 2011; accepted, January 12, 2012.

Supported by China NSFC (NO.41061043).

Corresponding author: Xiaoqing ZUO, zuoxq@163.com

The remainder of this paper is organized as follows. In section 2, the background is introduced including a division method of the base station's service area and the cell phone traffic data. Section 3 proposes two visualization methods and a hotspot detection algorithm. Section 4 is the visualization and hotspot detection experiments. Section 6 presents the conclusion of this paper.

II. BACKGROUND

A. Determination of Base Station's Service Area

The base station mentioned in this paper refers to the Base Transceiver Station (BTS), which handles the radio-link between cell phone and the network. And the service coverage of a BTS is called a cell. In order to manage and maintain the base station, cell phone service operator keeps the information of each base station including Cell-ID, Site-ID, LAC, Longitude, Latitude and direction. Table 1 shows the information of a base station.

Through base station's positioning information, we can estimate the position of a cell phone, which is connected with a base station, within the service area of an antenna. Therefore, how to determine the service area of base station is a problem. Classically, hexagonal mesh is used

to represent the GSM network, but Baert and Sem [13] argued that Voronoi cellular network appears to be the more suitable model. After studying and considering the features of Voronoi cellular network and GSM network, Voronoi cellular network is advised to represent base station service area, and the area is also known as a cell.

Voronoi cellular network, also known as Thiessen polygons, is composed of a set of continuous polygons which is made up of perpendicular bisect of lines connected by a two adjacent points. If the Euclidean space is divided into a set of polygons by a set of distinguishable points with the principle of nearest, each point should relate to the nearest region. The strength of signals from the base station is changing all the time affected by many factors, so is the service area. For simplicity, we suppose that signals from base station are the same and the service area of a base station is changeless.

Thus, the distributional Voronoi cellular network of the base station can be created based on position of the base station. The Voronoi cellular network which is created to represent the communication cellular network is shown in Fig. 1.

TABLE I.
INFORMATION OF A BASE STATION

Field Name	Meaning	Example	Notation
Site-ID	Base station ID	4	A base station usually has three cells(sectors)
Cell-ID	Cell ID	44	Index number of Non-Omnidirectional antenna
LAC	Location area code	35015	Used to detect the location of mobile terminals
Longitude	Longitude	102.374593	
Latitude	Latitude	25.1533798	
Direction	Direction of the antenna	135	Direction of the Non-Omnidirectional antenna

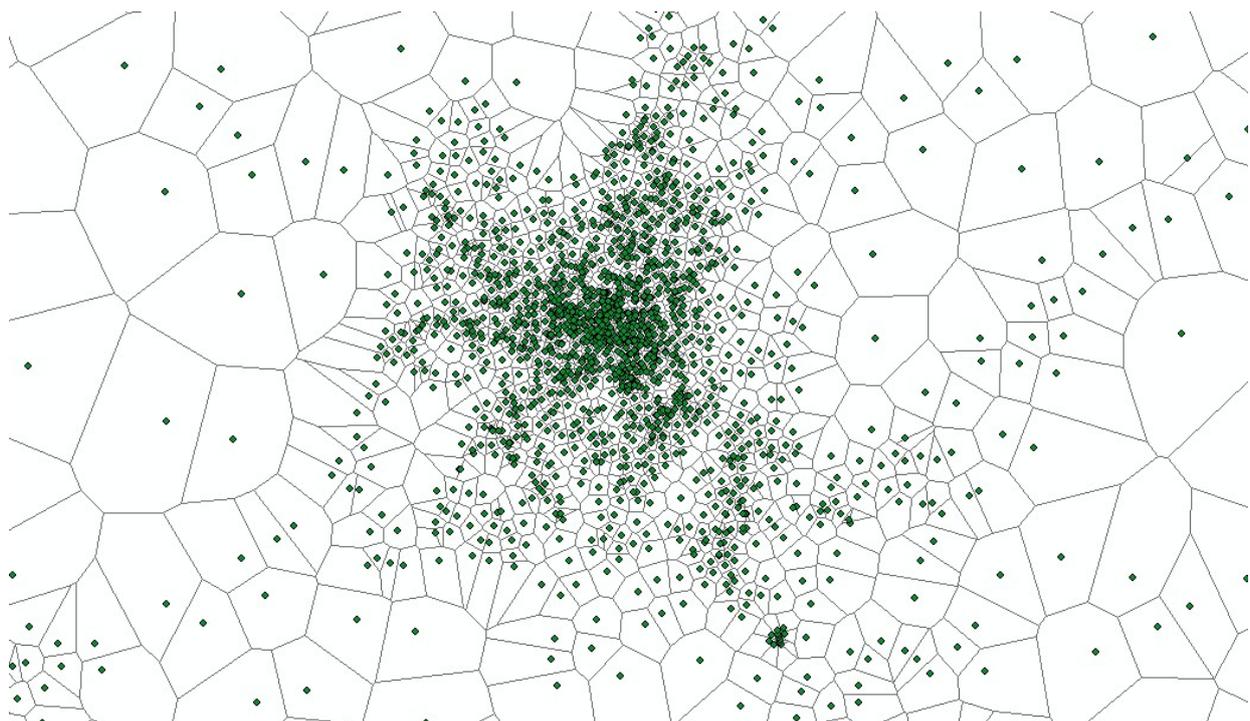


Figure 1. Voronoi cellular network of the base station

B. Cell Phone Traffic Data

Cell phone traffic refers to the product of a number of communications and their duration in a particular communication cell. It is also known as Erlang, the standard unit of measurement of cell phone traffic (1Erlang represents one cell phone one hour of phone use). Cell phone traffic data is usually recorded at the base station level. It is a macro indicator that reflects the frequency of the call activity in a cell. It is an aggregated and anonymous data that will lead to no privacy invasion. In other words, it is impossible to deduce individual information from it. Moreover, this kind of data can be collected with no modification to the GSM network and no impact on the routine operation of the network. From the definition, we know that cell phone traffic data are space-related, and their location can be determined according to the Cell-ID. So it can reflect the intensity of communication activities in different places. Taking advantage of this, the distribution of activities in the city over time can be understood, and people's activities pattern can be deduced.

III. VISUALIZATION METHODS & DETECTION ALGORITHM

A. Visualization Method of Cell phone Data

Data visualization is a way to analyze and display data aiming at helping people better understand and use data. Visualization of traffic data can help us understand and analyze the rule of the spatial and temporal distribution of communication activity which reflects the activity pattern of urban residents. A simple visualization method is to make cell phone traffic data be the elevation dimension of the Voronoi cell, which represents the base station's service area, according to the Cell-ID. This method is capable to show the spatial distribution of the communication activities, while the distributed space expressed is discontinuous and not conducive to present their dynamic change. Another visualization method is to translate cell phone traffic data into different colors according to their values. Firstly, divide the studied area into small squares (taking into account the amount of calculation and the smoothness of the interpolation surface, the study area is divided into 30m × 30m grids in our experiment below). Secondly, regard all cell phone traffic data as sampling point in the position of the base station. Thirdly, with the knowledge of geostatistics, take a particular interpolation method (In this paper, the classical Kriging interpolation method is adopted) to put corresponding cell phone traffic data into each square created at the first step. Then, translate the cell phone traffic data in each square into colors according to the value. Finally, overlap a GIS map to it. Now, a map which can show the spatial distribution information of communication activity at a time comes is created. Moreover, putting the maps in time order, the temporal and spatial structure and dynamics of the call activities in the city is noted. Based on it, particular patterns of human life in the city can be deduced and analyzed.

B. Hotspot Detection Algorithm

In this paper, hotspot refers to the hotspot of communication activity where its cell phone traffic data display at high level for a long time and form contrast with other areas. The purpose of hotspot detection is to find these hotspots out automatically by some methods. Considering the situation that the cell phone traffic data is low in most areas, the article learns from the k-th Nearest Neighbor (KNN) outlier detection algorithm in the field of data mining, and proposes a kind of detection algorithm to find out hotspots. The fundamental idea of the algorithm is based on the sparse and dense degree around a data object to identify hotspots. The notations involved in the algorithm are as follows:

- (1) C for the cell set, c_i shows the cell numbered i ; T for the time series, as the record cycle of cell phone traffic data is 1 hour, so we can get 24 observation points on one day, t_k means the first k hours on one day; DA For the date set, d_j says the date is j ; So that, $erlang\{c_i, d_j, t_k\}$ means the c_i cell's cell phone traffic data in t_k hours on d_j .
- (2) k is the interested neighbor number, n is the preset number of city hot spots, they are variables by artificially controlled.
- (3) D is a 24-dimensional data sets, it shows value of gathered cell phone traffic data in a cell in 24 hours on one day. p, q means the data object in D, which Cell-ID and the index of it identified any distinguishable cell.
- (4) $d_{x,y}$ is the distance between the data object, and the Euclidean distance can be applied to express it. Such as the distance between the data object p and q .

$$d_{p,q} = \sqrt{\sum_{k=0}^{23} (erlang\{c_p, d_j, t_k\} - erlang\{c_q, d_j, t_k\})^2} \quad (1)$$
- (5) $Dk(p)$ represents the k -th nearest neighboring distance of the data object p . First calculate the distance between p and others, then put these distances in ascending order, as a result, the k -th distance is $Dk(p)$. The basic process of this algorithm as shown in Fig. 2.

The n output objects mean n hotspots, the obtained hotspots can be linked through Cell-ID to Voronoi cells which are shown on the map and then highlight them in bright color.

IV. EXPERIMENT

The experiment sets Kunming, a central city in China, as its study area, collecting about 700 million pieces of cell phone traffic data from 2/1/2011 to 28 of this city as the experiment data.

A. Cell Phone Traffic Spatial-Temporal Distribution

We create the Voronoi cellular network of the base station based on the position of the base stations in Kunming.

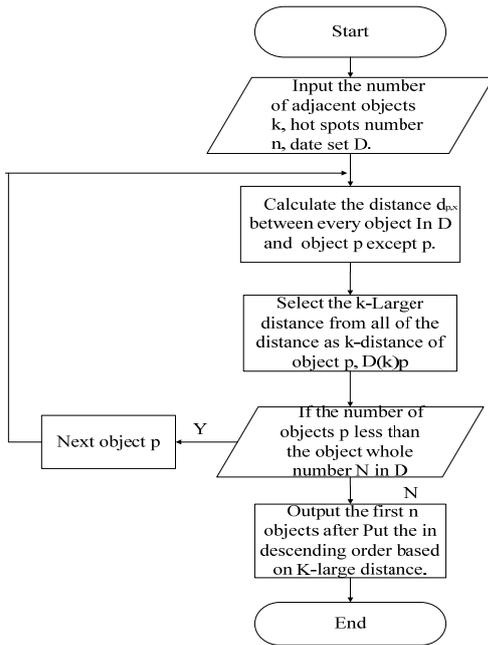


Figure 2. Process of hot detection algorithm

Then let cell phone traffic data be the elevation dimension of the Voronoi cell according to the Cell-ID. Fig. 3 briefly shows the distribution scenario of the communication activity at 10am in February 22, 2011.

Based on this scenario presented by Fig. 3, we can easily find the distribution of communication activity in the city. The overall trend of communication activity is decreasing from the center to the round of the city, while with fluctuation in city center. Based on this situation, we can speculate that most of people’s activities centralize in center of the city, while part of them distributed in different region due to different function area.

Fig. 3, which can display the call activity distribution in a specific time period, is a screenshot for urban structure at the perspective of cell phone traffic. In fact, the call activities in different regions change over time. Through the study of the cell phone traffic data that we collect for one month, we find the call activities in the city show regular pattern in time and space.

Fig. 4 displays variation of the cell phone traffic data in different areas over time within a week. This statistical

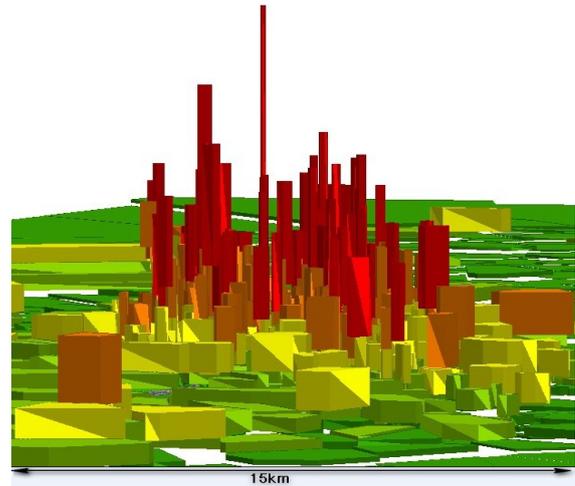


Figure 3. Distribution of call activity at 10am,22/2/2011

chart clearly shows the call activity is different in different areas at the same time, while the trend is consistent over time. Moreover, although the changes of different areas are not the same, it cycles by day. If we apply the gradual-changed color to sign out the amount of cell phone traffic and display the traffic by hours on map and order these maps in time series, the dynamic temporal and spatial regular pattern of communication activity in the city can be displayed directly.

Fig. 5 is an example of cell phone traffic variation on February 22, 2011, displaying the city’s spatial distribution and evolution of call activity. It shows that the cell phone traffic data spreads around the first and second cycle road areas of the city, and the north is larger than the south. At the beginning of a day, the most call activities are in the edge of the city, and then slowly gather to the center, and finally spread to the edge of the city in the night.

Combining to Fig. 4, we can know that the cycle of the call activity starts to rise from the morning and maintains at perch level during the daytime, then falls in the night. The cell phone traffic data is closely linked to people’s call activity, so, the changes of the cell phone traffic data reveal the residents’ activity distribution and evolving regulation.

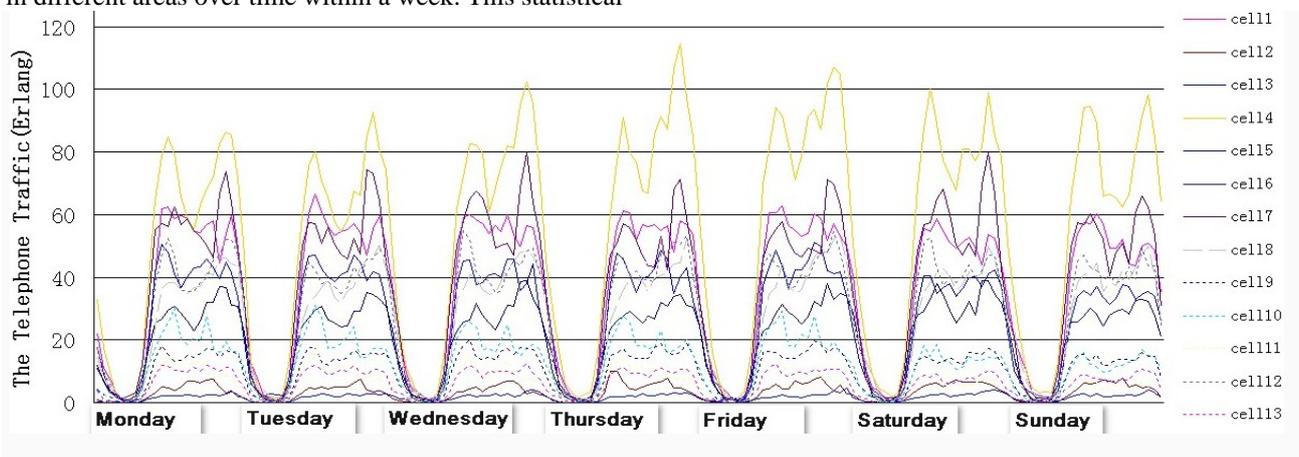


Figure 4. Sample of cell phone traffic data in different region in the third week in February, 2011

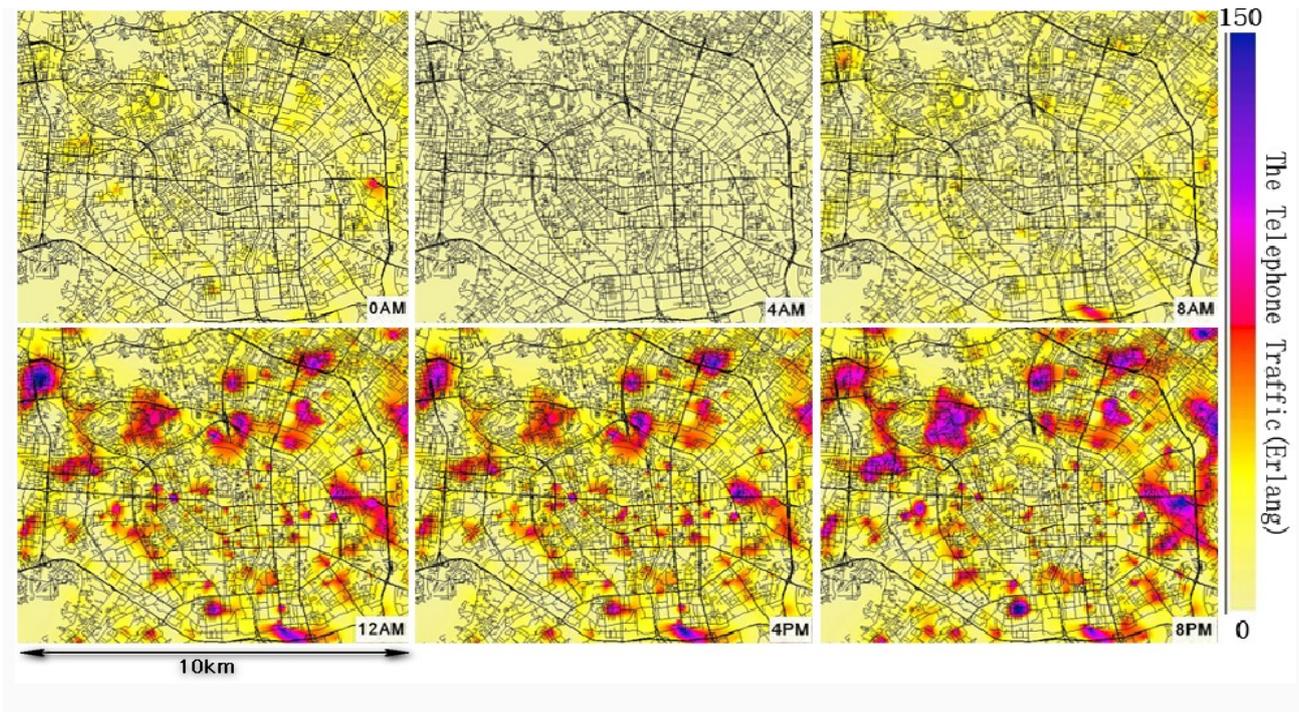


Figure 5. Variation of call activity in 22/2/2011

This rule may be described like this: most residents live on the edge of the city and they go to city center to work every day in working days and gather in the center, then come back home at night. However, this rule cannot be directly displayed by conventional method.

B. Hotspot Detection & Typical Hotspots Analysis

The cell phone traffic data of Kunming on 2/21/2011 is selected as study materials. Applying the detection algorithm proposed in this paper, hotspots are detected setting the number of hotspots $n = 100$, the number of neighbors $k = 10$. The result is shown in Fig. 6.

In Fig. 6, the detected result is placed on the top of a road map. From the Fig. we can note that hotspots in this city are mainly distributed between the first cycle road and the second cycle road, also, some hotspots located in center area of this city. The background knowledge of this city can contribute to find the distribution regulation of those hotspots. Based on the monthly hotspot exploration of traffic data, we found that the distribution of hotspots is relatively fixed in time and space. That is to say, the hotspot areas distribute in specified area in different time periods. What's more, the hotspots can be divided into different types according to their characters. Paragraphs below will select some typical hotspots as example to search the regular pattern of human activity in urban areas. To set a day as a cycle within one week, and combined with the statistical graph, the activity pattern of different hotspots is discussed. Fig. 7 to Fig. 14 are the curve graphs that display the cell phone traffic in different hotspots over time.

Fig. 7 demonstrates the situation of Residential areas. The selected hotspot is one of the bigger residential areas in Kunming named Chunhui area. The cell phone traffic curve's characteristics of this area are as follows: curve is

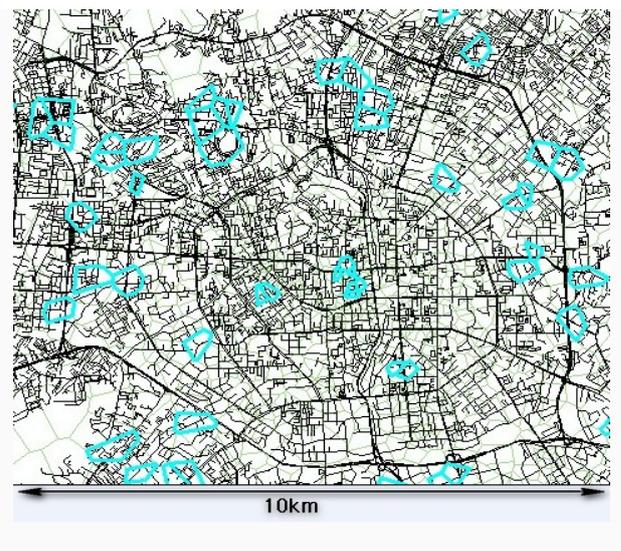


Figure 6. A screenshots of hotspots detection on 2/2/2011

much smooth and the shape is the letter "M" type. The cell phone traffic in period from 14:00 to 17:00 is the bottom of "M". The maximum of cell phone traffic is at 21:00, and after that, cell phone traffic drops sharply, still, cell phone traffic is very high about 2:00. These characteristics reveal people's living habits in residential area. Activities start about 7:00, and the lively times are about 11:00 in the morning and 21:00 in the evening. People in this area keep activities to late night while some others don't.

Fig. 8 exhibits the situation of railway stations. The railway station of Kunming, which is detected out, is the most important transportation junction of the city, and also one of the most typical hotspots of cell phone traffic. The cell phone traffic curve's characteristics of this area

are as follows: cell phone traffic begins to rise steadily about 4:00 from 0, and the curve is a litter sharper; the shape is also like the letter "M"; the maximum value is around 15:00; From the time 10:00 to 16:00, the cell phone traffic of this area changes significantly; And at 13:00 in the afternoon comes to the tiny point; After 21:00, cell phone traffic drops sharply, and about 23:00 tends to zero. People's activity in this area is not successive because of the train schedule. And we can see that lively activity time span is bigger than others while people start the journey very early in the morning and end it late in the evening.

Fig. 9 shows the situation of another important transportation junction of the city, the airport of Kunming. It is also a typical hotspot. Cell phone traffic in this area is highest in all selected typical hotspots, and its characteristics are as follows: the shape of the cell phone traffic curve is similar to the shape of "n"; cell phone traffic begins to sharply increase from 5:00; traffic from 7:00 to 8:00 is like a small step, and cell phone traffic is high from 12:00 to 19:00; it drops sharply after 19:00 and tends to zero after 1:00. People's activities in this area are also not successive, and no concavity in the afternoon indicates that activities in the airport are always busy.

Fig. 10 is the situation of a recreation area. The selected hotspot is the biggest recreation place in Kunming named Kun Du. There distributes all kinds of bars, parties, and it's the most typical hotspot. Curve characteristics of the change of cell phone traffic in this area is similar to the shape of letter "U"; cell phone traffic increases slowly from 5:00, and sharply increases from 19:00, and soon reaches the maximum value at 22:00; after that, cell phone traffic starts to drop and tends to zero until dawn. These characteristics indicate that this area is very popular during the night period. People usually go there at about 19:00, and leave at about 2:00. We can guess that there may come to many big shows during that time.

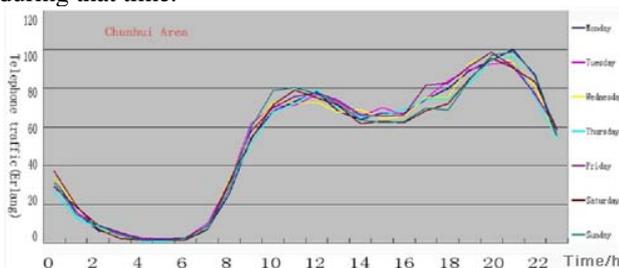


Figure 7. Graph of cell phone traffic of typical hotspots in a week over time—Residential Area

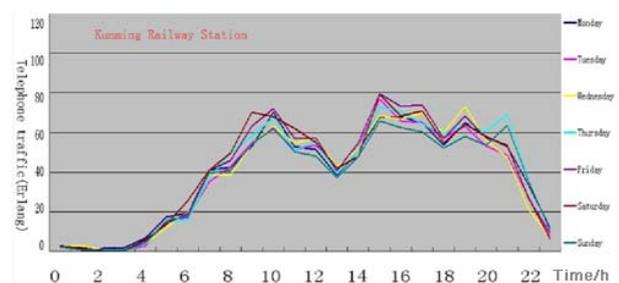


Figure 8. Graph of cell phone traffic of typical hotspots in a week over time—Railway Station

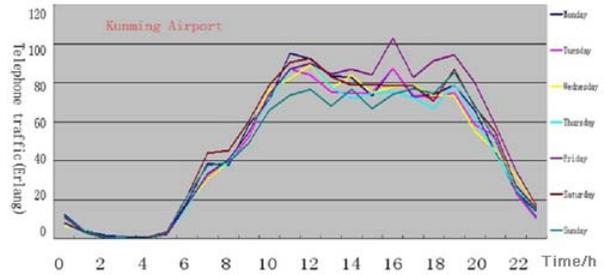


Figure 9. Graph of cell phone traffic of typical hotspots in a week over time—Airport

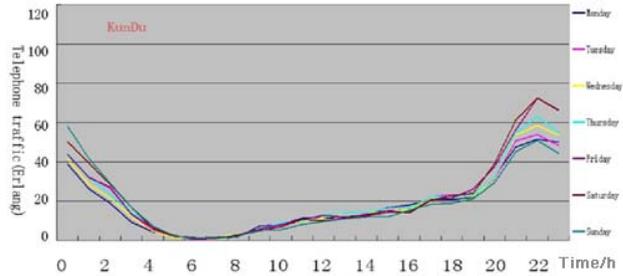


Figure 10. Graph of cell phone traffic of typical hotspots in a week over time—Recreation place

Fig. 11 displays the situation of cell phone traffic changing of office area. The most obvious characteristic of the changing curve of cell phone traffic in this area is that the cell phone traffic on weekdays and at weekend are different. Cell phone traffic in the weekend is lower than the weekdays. At lunch time it exists small cell phone traffic period, and after 18:00, cell phone traffic drops sharply and slowly drops after 19:00. This chart shows that most people do not go to work in this area at weekend.

Fig. 12 shows the situation of Business district. The selected business district is located in the city's center named NanPing Pedestrian Street. Cell phone traffic in this area begins to rise about 8:00, declines a little in the noon, and urged to the peak (about 16:00), and then drops sharply; after 18:00, it keeps higher value until 20:00, then falls sharply, and it goes zero about 23:00. The remarkable place of the shapes is at Saturday, in this day cell phone traffic does not drop at 10:00 to 20:00, and at about 20:00 goes the highest peak cell phone traffic. From above features, we see that Saturday is the most relaxed day of a week and people could go shopping till night.

Fig. 13 demonstrates the situation of a Park. Yuan Tong Shan Park, which is detected out, is a place where people rest and play. In the densely populated city, the park naturally becomes a hotspot. We select a park in the center of the town. The characteristics of the park are cell phone traffic rises slowly from 8:00, reaches peak about 11:00, but not appear obvious wave like other areas at noon, and then drops again from 18:00. And the overall cell phone traffic value in this area is relatively small. These features characterize people's activities in the park, and maybe this park is not popular among people.

Fig. 14 exhibits the situation of school. School is an area that population is relatively concentrated, so cell

phone traffic in the school area is high. The area around Yunnan Normal University is selected as a typical hotspot, cell phone traffic in this area begins to rise from 7:00 and about 12:00 it reaches a peak. It drops slightly in the evening and then rise again to remain a high value last 22:00, and drop to zero about 1:00. So, they reflect the living habit of students.

Overall, the curves of cell phone traffic in a particular area at different days are almost the same suggesting that it is stable for people to be active in a particular area. In addition, change of cell phone traffic over time is different because of the different places and this reflects people living model vary with locations.

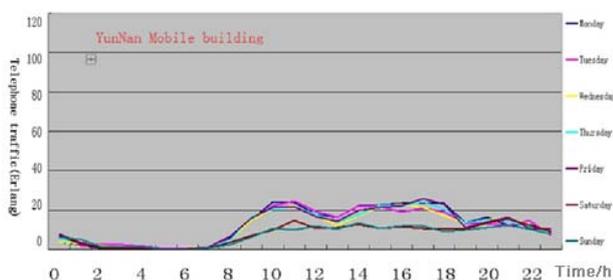


Figure 11. Graph of cell phone traffic of typical hotspots in a week over time—Office Area

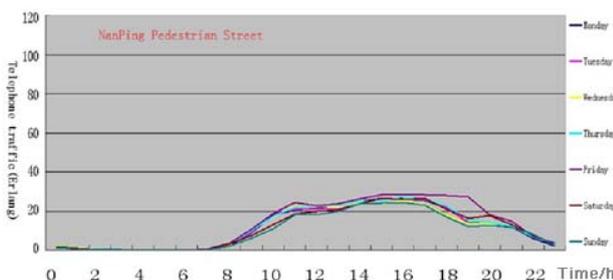


Figure 12. Graph of cell phone traffic of typical hotspots in a week over time—Business District

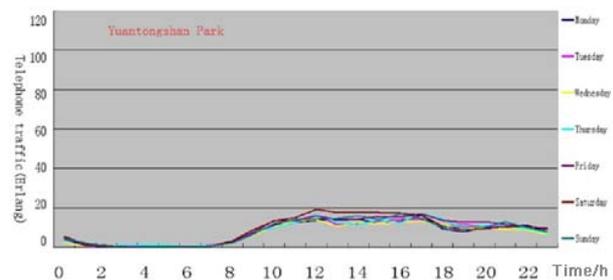


Figure 13. Graph of cell phone traffic of typical hotspots in a week over time—Park

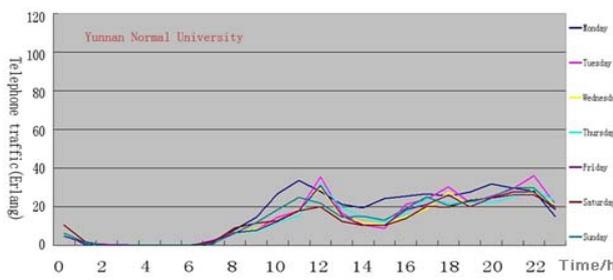


Figure 14. Graph of cell phone traffic of typical hotspots in a week over time—School

V. CONCLUSIONS

A new way to enhance the understanding of urban system is addressed by analyzing cell phone traffic data generated by the communication activity. Results showed that it is feasible to explore urban system through the methods presented above based on cell phone data. Cell phone traffic data provide us with a new perspective to view structure and dynamics of urban system. And through some methods, for example, the detection algorithm proposed in this paper, spatiotemporal hotspots can be found timely. By visualization of cell phone traffic data, the space morphology and dynamic spatiotemporal structure of Kunming are captured. By hotspots detection algorithm, some hotspots in Kunming which can indicate popular places or some important events are found, in which some organization or people may be interested in. And typical detected out hotspots discussed above can reveal the living pattern in these areas.

Owning to lack of social activity data of Kunming, we did not assess the efficiency of the methods proposed above. And the detection technique and analysis presented aforementioned only deals with cell phone activity, and the further relationship between cell phone activity and human living activity needs further study. Also, section 3.2, in the hotspot detection algorithm, the number of hotspots is an input parameter. It is possible to use another clustering algorithm that can automatically detect the number of clusters, but we think our method is more flexible.

In the future we would like to focus on real-time hotspot study. If we can find out new hotspots which related to special event in real time, it can play a significant role in public safety in case of emergencies.

ACKNOWLEDGMENT

This work was supported by the National Nature Science Foundation of China (NO.41061043).

REFERENCE

- [1] B. Luca and D. Martin, "Mobility Environments and Network Cities," *Journal of Urban Design*, Vol.8, No.1, pp.27-43, 2003.
- [2] Batty, "Thinking about cities as spatial events," *Environment and Planning B*, 29, pp.1-2, 2003.
- [3] J. Matthew and R. Carlo, "Urban Ritual in Rome: Characterizing the City with High-Resolution Cell Phone Data," <http://reality.media.mit.edu/publications.php>, 2007. [accessed 13 September, 2011]
- [4] R. Jonathan, C. Francesco, S. Andres, and R. Carlo, "Cellular Census: Explorations in Urban Data Collection," *MIT Senseable City Lab*, 2007
- [5] http://www.cnii.com.cn/yy/content/201106/22/content_887992.htm, [accessed 13 September, 2011]
- [6] R. Francisca, et al, "Real Time Rome," *IEEE transactions on intelligent transportation systems*, vol.12, NO.1, March 2011.
- [7] N. Caceres, J. P. Wideberg, F. G. Benitez, "Review of traffic data estimations extracted from cellular networks," *Intelligent Transportation Engineering IET*, 2008.

- [8] Y. Zheng, L. Z. Zhang, X. Xie, W. Y. Ma, "Mining Correlation between Locations Using Human Location History," *In Proceedings of GIS*, pp.472~475, 2009.
- [9] R. Carlo, M. P. Riccardo, W. Sarah, F. Dennis, "Mobile Landscapes: Using Location Data from Cell Phones for Urban Analysis," *Environment and Planning B: Planning and Design* 33(5), pp.727 – 748, 2007.
- [10] J. B. Sun, J. Yuan, Y. Wang, H. B. Si, X. M. Shan, "Exploring space–time structure of human mobility in urban space," *Physica A* 390 (2011), pp.929–942.
- [11] M. P. Riccardo, R. Carlo, T. Enzo, "City out of Chaos: Social Patterns and Organization in Urban Systems," *International Journal of Ecodynamics. Vol. 1, No. 2 (2006)*, pp. 125–134.
- [12] C. Francesco, R. Jon, R. Carlo, "Eigenplaces: analyzing cities using the space-time structure of the mobile phone network," *Environment and Planning B: Planning and Design* 2009, volume 36, pp. 824 - 836.
- [13] A. Baert, D. Sem, "Voronoi Mobile Cellular Networks: Topological Properties." In: *ISPDC 2004:3rd International Symposium on Parallel and Distributed Computing/HeteroPar 04: 3rd International Workshop on Algorithms, Models and Tools for Parallel Computing on Heterogeneous Networks*, proceeding, pp.29-35.



mining, 3DGIS and image processing.

Xiaoqing ZUO received the M.Sc. degree (2001) in GIS from Kunming University of Science and Technology and Ph.D. degree (2004) in GIS from Wuhan University. Now he is a professor in Faculty of Land Resource Engineering, Kunming University of Science and Technology. His research interests are in the areas of spatial data



Yongchuan ZHANG is pursuing a M.Sc. degree in Cartography and Geographic Information Engineering from Kunming University of Science and Technology. His main interest is in the use of cell phone data for road traffic estimation and urban dynamics analysis.