

The Consistency Analysis of Urban Land Function with Communication Networks and Human Behavior Patterns

Jing Li¹, Xuewei Li¹, Wenjun Wang¹, Shuyu Shi¹, Ruixin Song¹

¹School of Computer Science and Technology, Tianjin University, Tianjin 300350, China

Abstract. Effectively analyzing multi-attribute urban land function(ULF) makes government plan the use of urban lands more reasonable, such as the arrangement of the location of hospitals, markets, restaurants and so on. Previously, the general methods of analyzing function of urban lands are related to social investigation, however, there exists shortcoming of these kinds methods. For example, most of methods above can not implement quantitative calculation. In this paper, we intend to produce a new method to make consistency analysis of ULF with communication networks and human behavior patterns. Based upon the key indexes of communication networks and human behavior dynamics, we used the classifier of decision-making tree to achieve our goals. The results show that our method is feasible and excellent.

Keywords: communication networks, human behavior dynamics, urban land function.

1. Introduction

In recent years, there are more and more researches regarding to city planning and predicting ULF. ULF is functional attributes of city lands which are divided by urban road network, including education, residence, finance and so on. When it comes to making prediction of ULF, there are a great deal of significant approaches which are based on sociology [1]. Though these methods have made some contributions to the accuracy promotion of prediction, the sociological means are to some extent determined by personal or group's experience and it's hard to make sensible strategy with too many choices available.

Some researchers with geography and city planning background have studied cities from different scales and views based upon data sets of geography, social media and check-in, etc. Their works include macro researches such as city form, area association, and micro researches [2]. Quantitative analysis of related data sets will make it more effective to enhance the accuracy of some predictions. Nowadays, with the development of telecommunication industry, the coverage of mobile phone is almost the full-scale of society and the records of mobile phone users enable researchers to construct a directed communication network. And many characteristics of our lives, even "motifs" and regularities of human behavior, have been extracted from these social media data sets [3]. For example, Xu etc. studied the core-based dynamic community detection method based on mobile social networks [4]. Based on the data of smart-phone, Do, D. proposed contextual conditional models for human mobility prediction in real life [5]. Multi-threaded Community Detection in Social Networks by Riedy J etc [6], Bayesian Nonparametric Models [7] etc.

Conveniently, we study the communication networks based on the mobile phone data set and the indexes of communication networks in each land we study are mainly related to communities detected in this land. The results of community detection show individuals who connect each other closely tend to be in the same group. Also for the functions of urban land being closely related to the human communication links and behavior rules, we implemented the experiment. From the results of classification, we make consistency analysis of ULF with communication networks and human behavior patterns.

¹⁺ Corresponding author. Tel.: 15620562570 (Jing Li)
E-mail address: gxllj@tju.edu.cn

2. Method

2.1. The profile of data

The data we used are mobile phone data from 2014.12 to 2015.11 of a metropolis in China and lands data of the city.

For the mobile phone data, there are two main parts. The communication records are the logs of calls and messages and the mobile phone heartbeat data is that can keep track of individual positions. The count of former kind of data is about 1.1 billion per month and 1.3 billion per month for the latter kind of data.

The data of lands include 1678 lands with the attribute, for example, land 1 and its attribute Education Facility. There are 22 distinct attributes we used. The empirical study is based on the massive data above and the enormous quantity of data increase reliability of our research.

2.2. The processing of data

There are three steps of data processing. First, we defined the resident population. Through the periodic position statistic of each person by telecommunication base stations in the city at the frequency once an hour from 21:00pm to 7:00am the next day during a month, if the total records of one person by all the base stations were more than 77 times, putting it another way, the person has stayed in this city for almost 7 days in a month. As a result, he or she would be recorded as resident population and we chose the position with the most times recorded by all base stations as the person's address. According to mobile phone heartbeat data and lands data we found which land a ordinarily resident belonged to.

Then, we calculated the indexes of communication networks in order to reveal some key phenomena of communication, especially the communities in each land. By detecting the communities, we finally located each person into single community. Based on the results of community detection, we intersected the coverage of communities and lands, thus we were able to know the affiliation of calls and messages of users in each land. Then we calculated the key indexes of communities and communication networks for each land, messentropy, callmessnmi, calldegree, messdegree, callcomcnt, messcomcnt, callentropy. Callentropy means the complexity of calls communities in per land and there is similar meaning with messentropy. The higher value of entropy[8] in one land means that the higher complexity of communities in the land. We calculated entropy of each land to represent the complexity of communities in the land. Based on the entropy calculating, We calculated normalized mutual information[9] expressed by callmessnmi to measure the similarity of the results of community detection by the separate data of message and calls, the higher value of callmessnmi means the higher similarity of two results. Calldegree and messdegree are respectively represent the total number of calls links or message links of people in per land. Callcomcnt and messcomcnt are the amount of communities detected by calls data and message data in per land. The above indexes are necessary to evaluate the basic state of communication networks in lands.

At last, we calculated the indexes of human behavior dynamics both in individual level and land level. From the perspective of human behavior dynamics, we calculated the indexes and then through the mapping of people and lands, we studied the human behavior patterns in the level of land. For individual level, the eight indexes needed to be calculated. MessOutCnt represents the number of a person calls out. MessInCnt is the number of a person receiving calls. There are the similar meaning with CallOutCnt and CallInCnt. The above four indexes are about the basic statistics of individual communication. MessInInter stands for the sum of internals of any two adjacent times for a person's receiving messages. MessOutInter is the sum of internals of any two adjacent times for a person sending messages. The similar meaning with CallOutInter and CallInInter. MessInter represents the sum of intervals of any two messages for one person, no matter the person sends or receives messages. Its meaning is similar with CallInter. For the land level, the MessOutCnt, MessInCnt, CallOutCnt, CallInCnt in each land respectively represent the arithmetic average of corresponding indexes for all individuals in the certain land. When calculating MessInInter in the level of land, we only summed up the intervals of people who receives messages more than twice. It's also similar with the last five indexes in the land level, MessOutInter, MessInter, CallOutInter, CallInInter and CallInter. The six key indexes in one land all reflect the features of human communication in the certain land.

In total, we calculated 19 necessary indexes. Based on the results of calculating indexes and basic

attributes of lands, we made consistency analysis of the indexes with ULF by a classification.

3. Results

Based on the key indexes of communication networks and human behavior dynamics calculated, we will show the results in three parts.

3.1. The connection of indexes and urban lands

In this part, we will illustrate the connection of key indexes with lands. In the fig1(a), the probability distributions of callmessnmi in twelve months, which measure the similarity of calls communities and message communities detected in each land, show that the majority of results of community detection which are based on the data of calls and messages are similar in most lands, namely, the behavior of calls is corresponding to the behavior of messages in most communities of almost each land and this feature of corresponding has been stable for the year. The fig1 (b) explains that with the growth of index entropy in a certain range, the complexity of communities rises and the number of lands are being more. The complexity of communities in most lands are relatively high. But if index entropy is over a certain value, the number of lands will be fewer. Also, when the value of entropy is relatively low, the two distributions of entropy exist certain differences. It could be seen that people prefer the way of calls than messages.

Fig2 (a) shows the distributions of degree, for most lands, there are low communication connectivity or low population. If some lands are with similar population, the higher degree means the higher communication connectivity. If two persons connect each other by the way of calls or message, they will be regarded as connected. what's more, the two distributions are similar under a range of degree values, but when the values of degree grows to be a certain scope, the distributions exist obvious differences and the land numbers under both indexes toboggan. In the fig2(b), on the whole, there are relatively small number of calls and message communities in nearly each land and the total quantity of calls communities and message communities are similar.

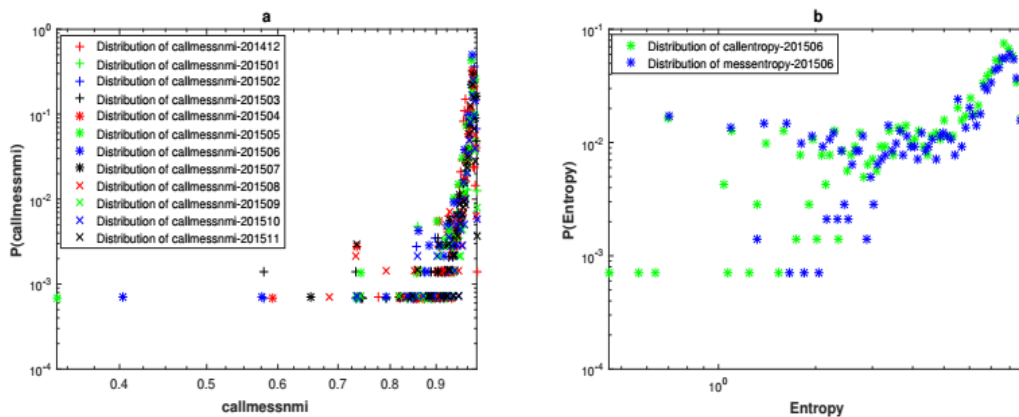


Fig. 1: Fig1 (a) is based upon mobile phone data in twelve months while fig1(b) uses the data in 2015.06.

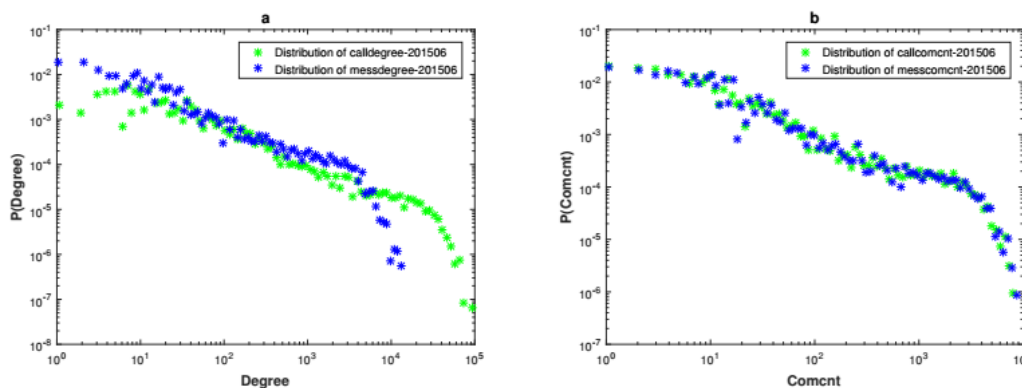


Fig. 2: Fig2 (a) and fig2(b) are both based on mobile phone data in 2015.06. Fig2(b) describes the two distributions of calls community counts(callcomcnt), the counts of community detected by calls and message community counts(messcomcnt).

In fig3 (a), for most lands, the intervals of communication are short and the status of long communication intervals is infrequent. Fig3(b) clearly reveals that people prefer to call than using messages and the number of out-going call and message sending is larger than phone-in and messages receiving. For all four indexes, there are four respective distributions with obvious differences, in other words, human behavior of four communication ways are unlike.

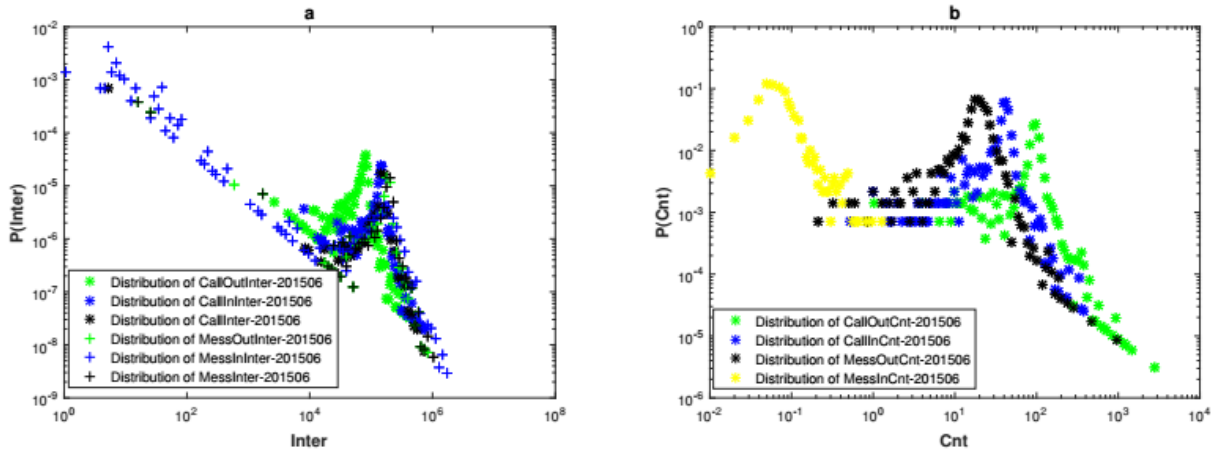


Fig. 3: The two sub-figures here use the data of mobile phone in 2015.06. In the fig3(a), abbreviations are as follows, interval(inter), calls out interval(CallOutInter), calls in interval(CallInInter), calls interval(CallInter), message in interval(MessInInter), message out interval(MessOutInter) and message interval(MessInter). Fig3(b) describes the distributions of communication counts of four communication ways in each land.

3.2. The urban land function prediction by the key indexes

In this part, the connection between indexes and lands will be explained by the way of classifying. We used the data during 2015.03 and 2015.07, the indexes of communication networks, human behavior patterns and lands to predict the UFL. The basic attributes of lands here used are population and area. Addition of the two attributes can improve the accuracy of prediction in a slight degree. The method of decision-making tree there used is the way of classification.

Table 1: The result of classification by all indexes

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.803	0.195	0.738	0.803	0.769	0.834	Residential Facility
0.671	0.12	0.665	0.671	0.668	0.814	Financial Facility
0.516	0.013	0.516	0.516	0.516	0.813	Research Facility
0.567	0.015	0.642	0.567	0.602	0.831	Public Facility
0.528	0.018	0.549	0.528	0.538	0.815	Primary Education Facility
0.453	0.003	0.618	0.453	0.523	0.796	Middle Education Facility
0.456	0.006	0.598	0.456	0.517	0.797	Medical Facility
0.738	0.004	0.702	0.738	0.72	0.892	Logistic Facility
0.64	0.012	0.706	0.64	0.671	0.86	Government Facility
0.63	0.004	0.763	0.63	0.69	0.849	Higher Education Facility
0.551	0.004	0.623	0.551	0.585	0.85	Office Facility
0.482	0.005	0.582	0.482	0.527	0.784	Entertainment Facility
0.667	0.006	0.697	0.667	0.681	0.886	Service Facility
0.444	0.002	0.645	0.444	0.526	0.797	Hotel Facility
0.613	0.001	0.655	0.613	0.633	0.853	Adult Education Facility
0.528	0.006	0.57	0.528	0.548	0.832	Traffic Facility
0.6	0.001	0.706	0.6	0.649	0.874	News Facility
0.381	0.003	0.5	0.381	0.432	0.778	Military Facility
0.383	0.009	0.519	0.383	0.441	0.76	Industry Facility
0.733	0.001	0.647	0.733	0.688	0.866	Storage Facility
0.676	0.001	0.735	0.676	0.704	0.851	Sports Facility
0.542	0.002	0.52	0.542	0.531	0.853	Library Facility

The accuracy of classification with all the indexes to predict the classes of UFL is 68.3792% and Kappa statistic[10] is 0.5747. The result shows the close connection between UFL and indexes, namely, the method

of prediction is effective and meaningful. The detail results show in table 1.

3.3. The urban land function prediction by the key indexes

It's clear that the key indexes reflect the main feature of communication networks and human behavior patterns. And through the results above, we can conclude the consistency of ULF with communication networks and human behavior patterns, in the other word, it is feasible to predict ULF by the key indexes of communication networks and human behavior dynamics. Also from the function of urban lands, we could infer the communication links and human behavior in a degree.

4. Conclusion

It's clear that ULF is closely related to communication networks and human behavior patterns, in other words, city planning should be on the basis of human communication behavior and other daily performance. It's possible that there are more suitable indexes not being used while the accuracy of prediction is satisfactory. Based on the study of key indexes, we can do more researches, for example, studying the variation trends of indexes in the lands to discover some meaningful phenomena, doing researches on the laws of human behavior dynamics in some specific functional lands.

5. Acknowledgment

This work was supported by the Major Project of National Social Science Fund(14ZDB153), the major research plan of the National Natural Science Foundation (91224009,51438009), the National Science and Technology Pillar Program (2013BAK02B06 and 2015BAL05B02), Tianjin Younger Natural Science Foundation (Grant no: 14JCQNJC00400).

6. References

- [1] Giles-Corti B, Donovan R J, "The relative influence of individual, social and physical environment determinants of physical activity," *Social science & medicine*, 2002, 54(12): 1793-1812.
- [2] Li M X, Xie W J, Jiang Z Q, "Communication cliques in mobile phone calling networks," *Journal of Statistical Mechanics: Theory and Experiment*, 2015, 2015(11): P11007.
- [3] Qin S M, Verkasalo H, Mohtaschemi M, "Patterns, entropy, and predictability of human mobility and life," *PloS one*, 2012, 7(12): e51353.
- [4] Xu H, Hu Y, Wang Z, "Core-based dynamic community detection in mobile social networks," *Entropy*, 2013, 15(12): 5419-5438.
- [5] Do, Trinh Minh Tri, and Daniel Gatica-Perez, "Contextual conditional models for smartphone-based human mobility prediction," *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 2012.
- [6] Riedy J, Bader D A, Meyerhenke H, "Scalable multi-threaded community detection in social networks," *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW)*, 2012 IEEE 26th International. IEEE, 2012: 1619-1628.
- [7] Chen Y, Wang X L, Xiang X, "Network structure exploration via Bayesian nonparametric models," *Journal of Statistical Mechanics: Theory and Experiment*, 2015, 2015(10): P10004.
- [8] Est évez, P. A., Tesmer, M., Perez, C. A., Zurada, J. M. Normalized mutual information feature selection. *IEEE Transactions on Neural Networks* 20, 189–201 (2009).
- [9] Thompson, J. R, "Estimating equations for kappa statistics," *Statistics in Medicine* 20, 2895–2s906 (2001).
- [10] Fawcett, T, "An introduction to roc analysis," *Pattern Recognition Letters* 27, 861–874 (2006).