

## Prediction of Transportation Network Based on PageRank Algorithm

Tongbo ZHANG<sup>1</sup>, Guangli LI<sup>1</sup>, Yue XU<sup>1</sup>, Yang YANG<sup>1</sup>, Shuai LÜ<sup>1,a</sup>

<sup>1</sup> College of Computer Science and Technology, Jilin University, Changchun, 130012, China

<sup>a</sup>email: lus@jlu.edu.cn, corresponding author

**Keywords:** Data Mining; Complex Network Analysis; PageRank Algorithm; Congestion Prediction

**Abstract.** As a new research field, network science is gaining more and more attention with the networking of human society. Urban traffic is a kind of typical complex network, the analysis of its property is becoming a hot topic. The research of the urban traffic network congestion is presented in this paper. Through in-depth analysis of the urban traffic network features and a variety of circumstances which lead to traffic congestion, we build a complex network model which the nodes represent specific roads. In this model, we get the PageRank algorithm improved which in the field of search engines, and we apply it to a transportation network. Then we predict the situations of urban traffic network congestion with this model. Finally, comprehensive tests are conducted with the data from the monitoring of real roads. The experimental results show that the model proposed in this paper can simulate the change of traffic flow, which has a good forecast effect on the urban traffic congestion and also has a referential value to the planning of urban traffic.

### Introduction

With the development of human society, various kinds of infrastructures have been evolving into more and more complicated networking systems. As a typical complex network, the analysis of network properties on urban transportation network has become a research hotspot[1-2]. Traffic collision is one of the most ordinary problems in transportation network. There are large-scale collision phenomena appeared occasionally with the increasing amount of vehicles in city. So it is a livelihood issue to solve the congestion problem in urban traffic. According to the classical Braess paradox, adding new roads in city merely cannot certainly alleviate the traffic pressure. Instead, it may compound the congestion[3]. To make a better urban traffic planning, it is necessary to do elaborate analysis on some of the relations and the properties in urban traffic network.

In recent years, many cities have adopted multiple approaches to cope with the transportation collision. In order to analyze this phenomenon, various models are proposed. For some typical applications in urban traffic, Gao Z. Y. et al. gave a general analysis on the research meaning of urban transportation complexity by combining with characteristics such as travelers game and network structure etc.[4]. Lu S. F. et al. find out the order parameters of urban transportation complexity system and propose a macroscopic method which scientifically researches the traffic management decision by complex network[5]. Gu Q. et al. do the research on urban public transportation network by complex network, and compare the public traffic network systems among three big cities, i.e. Beijing, Shanghai and Hangzhou[6]. Zhang X. et al. analyze the topological property of the transportation network in Beijing through complex network theory and find out two different kinds of key nodes and the features of the whole network in urban public network, which provides new reference proposals for the optimization of city traffic network[7]. Li S. B. et al. work out the influence that the network topology structure made on traffic collision by the improved mesoscopic traffic flow, and they discuss the traffic transmission dynamics[8].

This paper builds a directed complex network model through the analysis on the monitoring data of all the intersections in urban transportation, and proposes a prediction model of urban traffic network collision. The first part illustrates the process of establishing complex network aiming at traffic monitoring data. The second part portrays how to predict the congestion in traffic network by the improved PageRank algorithm. The third part presents the test results and the analysis of the model. And the last part concludes the whole work of this paper and puts forward the future

research orientation.

## Complex Network Analysis of Urban Traffic

### Urban Traffic Network Model

A typical city transportation map mainly contains information of the intersections and the roads. The traditional one often turns intersections into nodes and turns roads into edges. The method like this can present the message in the map well but not precisely and detailed enough when it comes to the flow changes of urban transportation network. For congestion problem in urban transportation network, we build a network model which takes roads as nodes and takes connected relation among roads as edges.

**Nodes:** There are always two-way roads in urban traffic. In order to analyze the change of traffic volumes from different orientations, two nodes are used to represent two directions of a road.

**Edges:** In city traffic, the directions of vehicles heading at intersection primarily include going straight, turning left, turning right and turning around. We could establish the connected relations among roads.

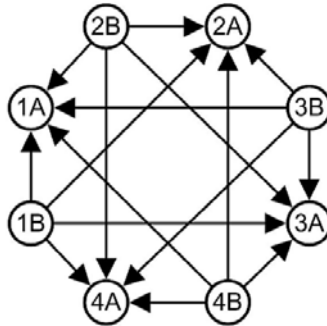


Fig.1. Example of Complex Network with Road-Node

Figure 1 is an example of the network model, where each road is represented by two nodes (such as 1A and 1B). Considering we standing at the node 1B, the edge <1B, 3A> represents the situation of going straight, the edge <1B, 2A> represents turning left, the edge <1B, 4A> represents turning right, the edge <1B, 1A> represents turning around.

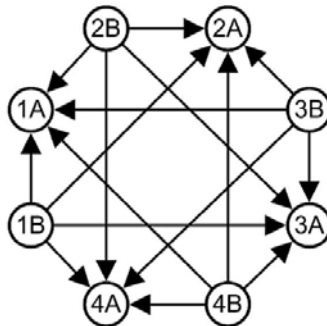


Fig.2. Example of Complex Network with Intersection-Node

### Complex Network Analysis

Complex network analysis has been widely applied in areas such as Internet, Mscbsc, communication network and electric network[9-11]. By further analyzing the complex network, we find out the inherent rules in diverse complex systems, which has great significance in understanding the features of complex systems. There are many important indexes in a complex system that could reflect the overall condition of a complex network. As for the complex network analysis methods applied by urban traffic, several crucial indexes as follows can be used:

**Degree:** The amount of edges that link directly to nodes in network is represented as degree. In actual traffic network, the more is the degree of a node, the more are roads linked to this node, which means the node is more vital to the traffic network.

**Weighted average degree:** In complex network, the weighted average degree refers to the weighted average of degree of all the nodes. We map this index to urban traffic complex network.

And the bigger weighted average indicates the more multi-way plane intersections in the city, which increases the indegree and outdegree of the node.

**Average Path Length:** In complex network, the distance between two nodes is represented as sum of the weights that on the shortest path between the two nodes, i.e. the geodesic distance or leap distance between the two nodes. The average path length of network is defined as the average of the distance between arbitrary two nodes. Map the index to traffic network, which represents the separation degree between nodes in network, i.e. the scale of traffic network. In general, the index represents the amount of roads between arbitrary two nodes, which to some extent reflects the shortest path and the least time needed by once resident trip.

**Network Diameter:** In network, the maximum of the shortest path length between arbitrary two nodes is known as network diameter. A bigger network diameter represents a longer distance that one resident trip needs, which also means a longer time consumed.

**Graph Density:** As for the directed complex network, the density of a network including several nodes is defined as the ratio between edges actually existing in network and mostly possible edges. A bigger graph density represents more path choices for resident trip, and it's much easier for nodes to reach mutually, which simplifies the distribution of urban traffic network planning.

**Clustering Coefficient:** Assume node  $i$  in network has  $k_i$  edges linking it with other nodes, then we name the  $k_i$  nodes as neighbors of node  $i$ . It's conspicuous that among the  $k_i$  nodes it is up to  $k_i(k_i-1)/2$  edges possibly. And among the  $k_i$  nodes, the ratio between actual existing edges  $E_i$  and the total number of edges  $k_i(k_i-1)/2$  is defined as a clustering coefficient of node  $i$ . The clustering coefficient indicates a compactness of the urban traffic network. For a node in traffic network, a low value indicates a sparse relation with its neighbor nodes. So the flux of this node tends to overflow when the traffic volume of neighbor nodes all go through it, which leads to situation like congestion.

After building a proper urban traffic network model, we can figure out the indexes above, which help us with the scientific analysis and discuss on the overall condition, thereby settling down the future planning of urban transportation.

## Congestion Prediction Based on PageRank

### PageRank in Transportation Network

Traffic volume is similar to network link, which both reflect the direction and influence from one node to another. PageRank is a classical algorithm to rank websites. We can get the final rankings through iterative computation for PageRank value of each node. The main thought can be represented by the formula below:

$$PR(i) = \frac{1-a}{n} + a \sum_{j \in I(i)} \frac{PR(j)}{|O(j)|} \quad (1)$$

Therein  $I(i)$  is a set of nodes directing to node  $i$ ,  $O(j)$  is a set of nodes directed by node  $j$ , and  $\alpha$  is damping factor. PageRank refers to a "random surfer model", which simulates a process that visitors randomly clicking on links will arrive at any particular page in the Internet.  $1-\alpha$  can be considered as the probability that visitor transfers from the current page to any other random page. In the urban transportation network, that cars at the intersection access from one road to another is an analogical process to the website redirecting case mentioned above, wherein the damping factor can be considered to be the situation when vehicles enter a parking lot or community after driving into one of the roads, explaining the inconstant traffic volume consequently.

The classical PageRank formula shows that the PageRank value of a node is uniformly distributed to the nodes it directing. In transportation network, however, the probabilities of vehicles driving to diverse directions are apparently different. To describe the change of traffic volume more precisely, usually the weight is introduced[12], accordingly the traditional PageRank formula is improved as below:

$$UTPR(i) = \frac{1-a}{n} + a \cdot \sum_{j \in I(i)} \frac{W_{ij} * UTPR(j)}{\sum_{k \in o(j)} W_{kj}} \quad (2)$$

Weights of edges in complex network are introduced in the improved formula, which caused that changes of traffic volume having among nodes get different weights, consequently revealing the flow changes in urban transportation network better. According to the various situations in traffic network, weights can be jointly determined by multiple factors such as driving direction, path length and intersection traffic time and so on. If the weights of different directions when vehicles redirecting at every intersection are same as  $W_{ij}$ , then the formula will degenerate to the conventional PageRank.

### Traffic Network Congestion Prediction

Aiming at the congestion problem in urban traffic network, a prediction algorithm based on the improved PageRank is designed in this paper. The algorithm simulates the process of flow changing in network traffic and finally works out the ranking list of probabilities of traffic congestions on each road. Hence, the congestion in city can be predicted by the ranking. The related pseudo-code is following as algorithm 1:

**Algorithm 1.** Urban traffic network congestion prediction algorithm

**Input:** Current urban traffic network data  $G = \langle E, V \rangle$

**Output:** Prediction ranking of congestion in urban traffic network RankList

1. Randomly initialize UTPR value as LastPR ;
2. **while** LastPR - CurPR > 0.001 **do**
3.     LastPR = CurPR;
4.     CurPR = {0};
5.     **for**  $i \leftarrow 1$  to  $n$  **do**
6.         **if** the out-degree of node  $i$  is 0
7.             **then** CurPR = CurPR + LastPR( $i$ ) /  $n$ ;
8.             **else foreach**  $\langle i, j \rangle \in E$  **do**
9.                 CurPR[ $j$ ] =  $W * \text{CurPR}[j] + \text{LastPR}(i) / n$ ;
10.     CurPR =  $\alpha * \text{CurPR} + (1 - \alpha) / n$ ;
11. RankList = Sort(LastPR, V);
12. **return** RankList;

The process of urban traffic network congestion prediction algorithm is similar to the traditional PageRank process. The algorithm first initializes the UTPR value and records it in LastPR. A comparison between the current UTPR value, CurPR, and the last UTPR value, LastPR, will be made during each iterative computation to determine whether the irritation terminates or not. The UTPR value in irritation updates according to formula 2. After the end of the irritation, rank the nodes by the ultimate UTPR value and then get the prediction ranking RankList. The algorithm is improved on basis of conventional PageRank and has an analogical computing procedure with it. When there is a large amount of data, it can be easily expanded to cloud computing cluster and to be processed in parallel[13].

### Experimental Results and Analysis

In this paper, we use the urban traffic network data in Tianchi Data Laboratory of Alibaba Group as test case. The data is the traffic volume's monitoring data at each intersection during a certain time period. Testing environment: Windows10 operating system, i7-3770 CPU, and the program is written in C++.

First of all, establish a complex network using roads as nodes by intersection monitoring data, as is shown in Figure 3.

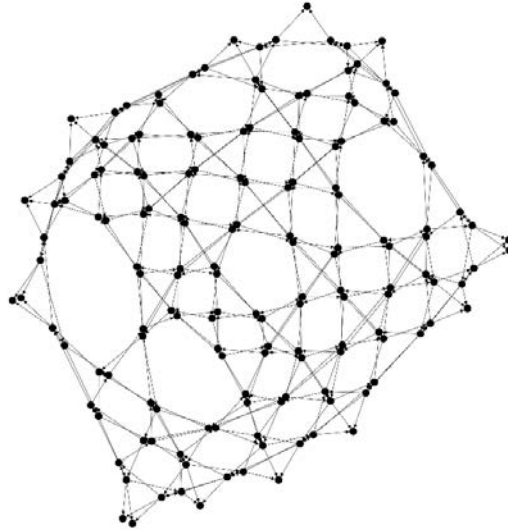


Fig.3. Complex Network of Urban Traffic

Next figure out the attributes in this complex network. Table 1 is the results.

Table 1 Attributes of Transportation Network

Complex Network Attribute	Attribute Value
Weighted Average Degree	0.903
Average Path Length	4.973
Network Diameter	10
Graph Density	0.021

It can be seen from Table 1 that the weighted average degree in city traffic network is 0.903, which indicates that the amount of multi-way plane intersections in complex network is relatively less. There are more crossroads and T-junctions in the network. The average path length is 4.973 or so, indicating less intersections that the vehicles need to pass from one road to another. The network diameter is 10, representing a small scale of the traffic network. The higher graph density represents more choices for resident trip. It becomes easier to travel among sites and it also facilitates residents to plan trip route more properly according to real-time traffic.

Rank the nodes in traffic network by prediction algorithm and then forecast the traffic congestion in city. The results are shown in figure 4.

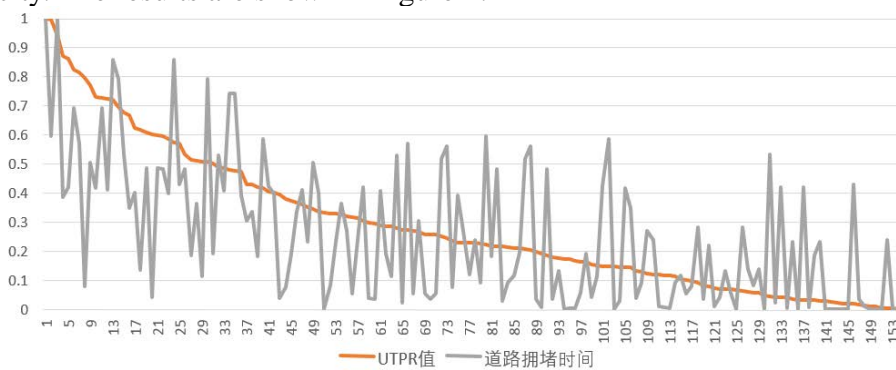


Fig.4. Prediction of Traffic Congestion

Figure 4 shows the results ranked by UTPR values in descending order after normalizing the congestion time and UTPR value of each node. From figure 3 we can see that as the UTPR values decline, the congestion time has a downward trend as well. The fluctuation is mainly derived from the solution to congestion that originally exists such as police assistance, traffic lights restriction etc.

## Conclusions

In this paper, we built a directed complex network model using roads as nodes through the further analysis on urban network features and different situations of traffic congestion. This model applied the improved PageRank to traffic network to realize the prediction of city traffic congestion. The model presented in this paper simulated the change of traffic flow in city well, which provided a reference value for urban traffic planning. We mainly made use of the probability of vehicles redirecting at intersection as the edge weight in network, and further added path length and surrounding etc. in traffic network, which made the urban traffic network model more precise.

## Acknowledgement

In this paper, the research was sponsored by the National Nature Science Foundation of China (No. 61300049, No. 61502197, No. 61503044), the Specialized Research Fund for the Doctoral Program of Higher Education of China (No. 20120061120059), the China Postdoctoral Science Foundation (No. 2011M500612), the Key Program for Science and Technology Development of Jilin Province of China (No. 20130206052GX), and the Natural Science Research Foundation of Jilin Province of China (No. 20140520069JH, No. 20150520058JH, No. 20150101054JC).

## References

- [1] Jianjun Wu, Ziyou Gao, Huijun Sun, Hui Zhao. Urban transportation system complexity[M]. Science Press, 2010.
- [2] Jianjun Wu, Shubin Li. Urban transportation system complexity based on complex network[J]. Shandong Science, 2009, 22(4): 68-73.
- [3] Xiaofan Wang, Xiang Li, Guanrong Chen. Network science: an introduction[M]. Higher Education Press, 2012.
- [4] Ziyou Gao, Jianjun Wu. Travelers game, network structure and urban traffic system complexity[J]. Complex Systems and Complexity Science, 2010, 7(4): 55-64.
- [5] Shoufeng Lu, Zhaosheng Yang, Ximin Liu. Research on urban traffic system based on complexity theory[J]. Journal of Jilin University (Engineering and Technology Edition), 2006, 36(3): 153-156.
- [6] Qian Gu, Xuhua Yang, Wanliang Wang, Bo Wang. Research on urban public transport networks based on complex networks[J]. Computer Engineering, 2008, 34(20): 266-268.
- [7] Xiao Zheng, Jianping Chen, Jiali Shao, Lidong Bie. Analysis on topological properties of Beijing urban public transit based on complex network theory[J]. Acta Physica Sinica, 2012, 61(19): 1-11.
- [8] Shubin Li, Jianjun Wu, Ziyou Gao, Yong Lin, Baibai Fu. Traffic congestion and propagation dynamics analysis based on complex network[J]. Journal of Physics, 2011, 60(5): 050701, 1-9.
- [9] Bogdan Danila, Yong Yu, John A. Marsh, Kevin E. Bassler. Transport optimization on complex networks[J]. Chaos: An Interdisciplinary Journal of Nonlinear Science, 2007, 17(2): 026102.
- [10] Xiaofen Jiang. The application of complex network model in mobile communication[J]. Electronic Test 2011 (4): 34-37.
- [11] Jin Shi, Guangyu Tu, Yi Luo. Power system complex network analysis and model improvement[J]. Proceedings of the CSEE, 2008, 28(25): 93-98.

[12] Wenpu Xing, Ghorbani Ali. Weighted pagerank algorithm[C]. Proceedings of Communication Networks and Services Research, Fredericton, NB, Canada, May 19-21, 2004: 305-314.

[13] Cheng Chen, Mengwei Zhan, Ying Li. Parallel pagerank algorithm based on the classification of web links[J]. Computer Application, 2015, 35(1): 48-52.