Complex Network Analysis of Bus Network in Hanoi, Vietnam

Ngoc Dung BUI [a], Hieu Vu TRAN[b], Xuan Tich PHAM [c], Binh T. H. NGUYEN [d]

[a,b,c] *University of Transport and Communications, Hanoi, Vietnam*
[a] *E-mail: dnbui@utc.edu.vn*
[b] *E-mail:* ***hieutv@utc.edu.vn***
[c] *E-mail:* ***tichpx@utc.edu.vn***
[d] *University of Transport and Communication, Campus in Ho Chi Minh city;*
E-mail: binhnth@utc2.edu.vn

Abstract: In this paper, we introduce the complex network theory to analyze the public transportation system in Hanoi, Vietnam. The analysis focus on the statistical features of bus system in Hanoi based on the complex network properties such as degree distribution, closeness centrality, betweenness centrality and average path length. We also measure the robustness of the network under some attack strategies. The results show that the bus networks in Hanoi which operate efficiently with all stations with high degree are located in areas of high density population in which the number of transfers between any two stations is quite small. However, the robustness of the bus network is weak under the attack on the high degree stations, which may cause serious effect on the important stations.

Keywords: Bus Network, Public Transportation System, Network Analysis

1. INTRODUCTION

In the cities with large population, public bus is an important transportation method, which generally helps reduce traffic jam, air pollution and improve safety. Analyzing bus infrastructure and network has become more and more of an important work and has received much attention from research community. Commonly, bus system is represented as a network where each node is a station and edge is the connection between the two stations. To assess the operating as well as robustness of the bus network, including identifying the important node such as hubs or shortest paths, we need to apply complex network approach with bus network and then find the statistical properties of the network. Complex network has many advantages that can use to research public transport, complex network is suitable model to abstract public transport network. With many topological presentations, may kind of information can extract using complex network approach, these information are really important to understand the transport system and analyze the effective and robustness of the system (de Regt *et al.*, 2017). For instance with L space, properties of average shortest path, cluster, degree correlation… can be extracted meanwhile P space can be used to extract information of average transfer time (Zhang *et al.*, 2013) .The abstracted models of public transport network can also be used to stimulate attack and failure to network, this can predict damage’s effects, test the robustness of the transport system without causing any real damage (Zhang *et al.*, 2011).

Complex network have been a key tool to analyze the public transportation to assess the effective and find the optimal public transport network (Chatterjee *et al.*, 2014; Chen *et al.*, 2007; Derrible and Kennedy, 2010; Sarkar, 2013; Xu *et al.*, 2013; Zhang *et al.*, 2013). The topology and properties of complex network has been employed to compare bus systems between Brazil, Poland and China (De Bona *et al.*, 2016). The simulation and network statistical properties can be used to propose the solution to improve the network robustness and reduce the cost of time and transfer number (Zhang *et al.*, 2012). The global network characteristic and several centrality node measure can indicate the sensitive routes and stations (Haznagy *et al.*, 2015). However, complex network is dynamic and can result in difference in result due to the difference in size of cities and structure of transportation network. Therefore, there is a need to model the transportation network in specific city to look at the characteristic of its network.

Research of bus network is very important with all countries because these research can improve quality of most important transportation. There are already many research of bus system in many cities across Europa, USA, South America and also many developing countries such as China, India, however until now there is still no current research in Vietnam (de Regt *et al.,* 2017). The requirement of Hanoi bus network become more urgent because of the new expansion of Hanoi in 2008. After the expansion, population of the city double and land area is 3.6 times bigger, this huge increase of land area and population in a very short period of time can greatly effect Hanoi bus system. Hanoi bus system should be modified and expanded to form a broader and also more stability bus system that can meet the increase demand of passengers.

Hanoi city is one of biggest cities in Vietnam with more than 9 million in population but the current bus network only serves 23% of the city's transportation demand. There are 63 bus lines connect 13 districts but the movement demand is different between them. In this paper, we model the bus network in Hanoi as a complex network and analyze various statistical properties such as degree distribution, closeness centrality, betweenness centrality, clustering coefficient. We also analyze the robustness of the network by simulation of random failure and target attack the hubs. From the result, we can find out elements that affect the ability of bus system such as station position, path length as well as better understanding the topology properties of Hanoi bus.

This paper is organized as follow. Section 2 presents the method of network construction from the data. In Section 3, we analyze the properties of bus network in Hanoi. Finally, we make the conclusion in Section 4.

2. METHODOLOGY

Using complex network method, a public transport system network is abstracted as a graph $G = (V, E)$, where V is set of vertices or nodes and E is set of edges or connections. Commonly, the graph can be represented as adjacency matrix A_{ij}, where $A_{ij} = 1$ if there is an edge between node i and node j, otherwise, $A_{ij} = 0$. The complex network can be used to model the bus network where each bus stop is a node and the edge between two nodes means that there is a direct bus connection between them. The bus network is an undirected network if the bus lines are bidirectional. From the graph and complex network theory, we can measure various properties such as degree, degree distribution, closeness and betweenness centrality, average path length that describe the bus network. Degree, closeness and betweenness centrality is properties of each node and node with higher value of these properties can be consider playing a more important roles in a network. Degree distribution and average path length are property of whole network. Degree distribution can shows types of networks, random networks have Poissonian degree distribution, scale free networks have power law degree distibution (Cohen and Havlin, 2010). Average path length of the network can present efficiency of this network, normally network has small average path length even number of nodes and edges of this network are really huge (Albert *et al.*, 2000).

2.1 Degree Distribution

The degree of a node is the number of connections to this node. The degree of a node presents the importance of the node in the network. In transportation analysis, the degree of node, or bus station represents the number of the bus stations that can reach directly to this station. Mathematically, the degree k_i of the i^{th} bus station is proportional to the number of lines connected to this bus station.

$$k_i = \sum_{j}^{N} a_{ij} \tag{1}$$

where,

a_{ij} is an elementary of the adjacency matrix, if there is a direct edge connect node i and j then $a_{ij} = 1$, if there is no edge between node i and j then $a_{ij} = 0$,

N is number of nodes.

Degree distribution of a network is defined as the fraction of nodes that have degree k. Degree distribution describe the whole network nodes properties. With the degree, they can be used to find the hub in the bus network. In network theory, nodes with high degree are considered as a hub, that is, a main bus station in the city. The degree distribution $p(k)$ is the probability of a randomly selected node with k edges

2.2 Closeness and Betweenness Centrality

Closeness centrality of node i is calculated by the equation (2). The higher centrality, the more important of node. It means, the more central a station is the lower its total distance from all other stations.

$$C_c(i) = \frac{N-1}{\sum_{j \in V} d_{ij}} \tag{2}$$

where,

N is number of nodes,

d_{ij} is the shortest distance between node i and node j.

The betweenness of a node i is defined in the equation (3). Betweenness centrality measures the number of times a station acts as a bridge along the shortest path between two other stations. The higher betweenness centrality the more central of station.

$$C_b(i) = \sum_{j \in V} \sum_{k \in V, k > j} \frac{g_{jk}(i)}{g_{jk}} \tag{3}$$

where,

g_{jk} denotes the number of shortest paths between nodes j and k,

$g_{jk}(i)$ denotes the number of shortest paths between nodes j and k through node i.

2.3 Average path length

The path length or distance d_{ij} between two nodes (i, j) is the shortest path between them, measure in number of edges. The path length between two nodes is an essential in graph theory, that is, can be used in public transportation to find the shortest route from one station to another. The average path length of the graph measures the network information efficiency.

The network with low average path is likely to be travelled with short distance from any station to another.

$$l = \frac{2}{n \times (n+1)} \sum_{i>=j} d_{ij} \tag{4}$$

where,

n is number of nodes,

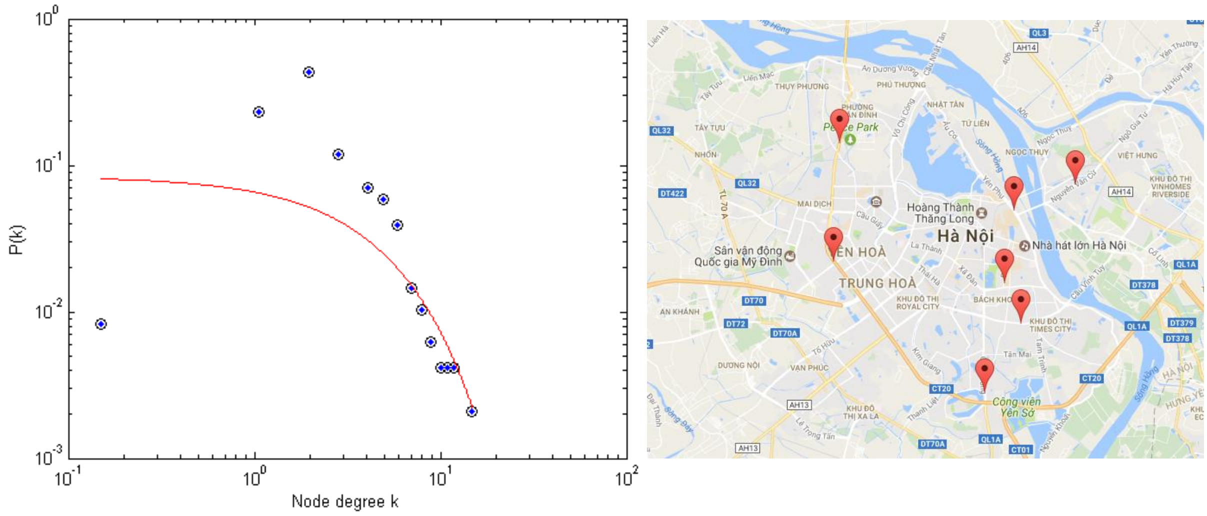
d_{ij} is the shortest distance between node i and node j.

2.4 Robustness of the network

Robustness of the network measures the ability of the network to continue to work well under the failure or targeted attack (Kooij and Ellens, 2013). In random failure, nodes are selected randomly and all the edges connected to these nodes are deleted. In targeted attack, nodes under some condition such as highest degree or closeness and betweennes are selected while all edges attached to them are deleted.

3. RESULTS AND DISCUSSION

In this paper, the data are collected from Hanoi bus website (**http://timbus.vn/fleets.aspx**) in 2017. The data is containt almost all lines and stations of Hanoi's bus network (exclude six lines to the suburban areas) which composed from 65 bus lines with 483 nodes and 643 edges. The average node degree is 2.66. The complex network model of Hanoi's bus network is an undirected network. Figure 1 shows the degree distribution of the Hanoi bus. The network show quite the same as partial power law degree distribution represented in the complementary cumulative distribution. The stations with a small number of connections with another station follow more closely with the red line, which has a scale-free. The stations following an exponential decay which are far from the red line will have more connection with another stations.



(a) (b)

Figure 1. Cumulative degree distribution of the bus stations (a) and the map of the highest degree stations (b)

The degree of a node measures the important of this node, the higher degree, and the more central of the node. The nodes with high degree are considered as a hub in the network, they are the main station in the city. Table 1 shows the top highest degree stations. Compare to the map in Figure 1b, they are the stations in the high density population.

Table 1. Top bus stations with highest degree

Node Degree	**Bus station name**
15	Giai Phong
12	Pham Van Dong
12	Minh Khai
11	Le Dai Hanh
11	Tran Nhat Duat
10	Nguyen Van Cu
10	Pham Hung

3.1 Closeness and Betweenness Centrality

Closeness centrality measure the closeness between one station and all other stations in the bus network. If the closeness centrality of the station is high, the station is important and it is in the central in the network because it is close to almost all other stations. Also, if the average closeness centrality of whole network is high, the bus network is operating efficiently.

Table 2 shows the bus station with highest closeness centrality, all are located on high density population. Also, from these stations, people can reach any stations with shortest distance.

Betweenness centrality measure the important of the station in the network. If the station is more important, it has a large number of lines that pass through this station. Table 3 shows the highest betweenness centrality of bus stations. Compared to the map, they are the main station in difference districts.

Table 2. Top bus stations with highest closeness centrality

Node Degree	**Bus station name**
0.0604	Le Duan
0.0597	Kham Thien
0.0595	Tran Hung Dao
0.0593	Tran Khanh Du
0.0592	Nguyen Van Cu
0.5090	Giai Phong

Table 3. Top bus stations with highest betweenness centrality

Node Degree	**Bus station name**
0.3032	Pham Van Dong
0.2843	Le Duan
0.2468	Giai Phong
0.2446	Nguyen Van Cu
0.2062	Cau Thang Long
0.2005	QL 3

3.2 Average path length

The average shortest path length is 9.2608 in L-space and 3.64 in P-space, respectively. In L-space representation, stations in each bus line are not directly connected. In P-space, every stations in each line are directly connected. Normally, L-space representation used to understand the relationship between the stations, and P-space representation used to measure the transfer between bus line in the network (Derrible and Kennedy, 2009). In Hanoi bus network, the average path length is smaller than 4, which means people need no more than 3 transfers when travel between two arbitrary stations.

3.3 Robustness of the network

In this study, we perform the robustness of bus network testing with random failure and targeted attack. The main distinction between random failures and targeted attack is the choice of edges to be deleted. Random failure simply selects a certain number of random edges to delete or destroy, on the other hand, targeted attacks selects edges of a certain property which in this case are edges connected to high degree nodes. The failure of complex systems of being resitant to damage is known as robusness. For the random failure, we perform edge deletion by randomly selecting 20% edges. The random failure does not affect the bus network much because important stations would not be broken. The node degree distribution is still follow the partial power law as in Figure 2a. The important stations are remained are: Tran Hung Dao, Pham Van Dong, and Minh Khai.

For the target attack, we perform Targeted attacks are implemented by arranging the nodes of the network according to degree. The edges of the highest degree nodes are then deleted until 20% of edges were deleted. All related edges are also removed. This attack strategy has seriously effect on the bus network by breaking the operation of the network. Most important stations fail and the node degree distribution does not follow the partial power law as in Figure 2b. The node degree of the station is decreased rapidly and the long line bus becomes high node degree and stations far from the central become more important such as Ngo Gia Kham, Duong Thanh, Phung Hung, Phan Dinh Phung, Cau Thanh Tri, Nguyen Chanh, Guot, Bo Ho, Tu Son, Cau Thang Long, QL32, Phung, Cau Phung, Hiep Thuan, Tam Hiep, Dong Ngac, QL32, Phu Thi, Y Lan, Binh Minh.

Random failure Targeted attack

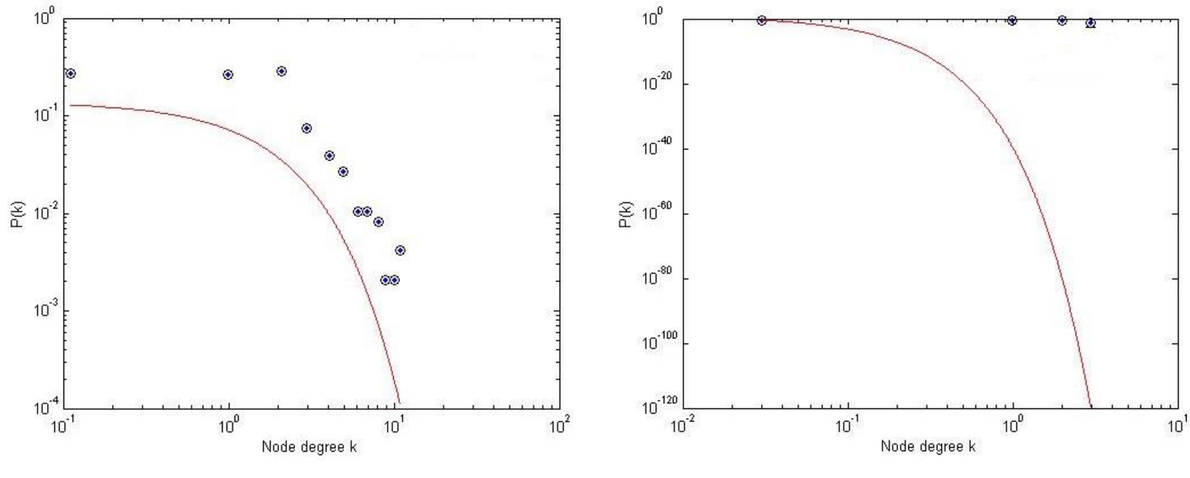


Figure 2. Cumulative degree distribution of the bus stations under random failure and targeted attack

4. CONCLUSION

In this paper, we analyzed the statistical properties of the bus network in Hanoi. Our analysis find out which Hanoi bus stations have high degree, closeness and betweenness centrality and can be considered as important bus stations. Also the relation of the bus lines is large so that people need less the number of transfer between bus line or transfer between two stations. We also stimulate targeted attack and random failure to bus network, the network is sensitive under attack. These are some suggestions to improve robustness of Hanoi bus network under failure and attack. The first suggestion is Hanoi bus network designers, engineers need to preserve some stations near high density population areas to reduce to effect of such damages. The second suggestion is important bus stations need careful plan in case any modifications or damages happen, these important stations are stations with high degree distributions, closeness and betweenness centrality and which were shown in tables 1, 2 and 3.

ACKNOWLEDGEMENTS

Research reported in this publication was supported by the ASEAN-European Academic University Network (ASEA-UNINET). We also would like to thank BSc. Hoang Thanh Tung, Cityme JSC, for bus network data collection.

REFERENCES

Albert, R., Jeong, H., Barabási, A. (2000), Error and attack tolerance of complex networks, *Nature*, vol. 406, no. 6794, pp. 378-82.

Chatterjee, A., Ramadurai, G. (2014) Scaling Laws in Chennai Bus Network. *4th International Conference on Complex Systems and Applications, France*, 137-141.

Chen, Y.-Z., Li, N., He, D.-R. (2007) A study on some urban bus transport networks. *Physica A: Statistical Mechanics and Its Applications*, 376(1-2), 747–754.

Cohen, R., Havlin, S. (2010), *Complex networks: structure, robustness and function,* Cambridge University Press, Cambridge, United Kingdom.

De Bona, A.A., Fonseca, K.V.O., Rosa, M.O., Luders, R., Delgado, R.B.S. (2016) Analysis of Public Bus Transportation of a Brazilian City Based on the Theory of Complex Networks Using the P-Space. *Mathematical Problems in Engineering*, 2016.

Derrible, S., Kennedy, C. (2009) Network analysis of world subway systems using updated graph theory. *Transportation Research Record: Journal of the Transportation Research Board*, 17-25.

Derrible, S., Kennedy, C. (2010) The complexity and robustness of metro networks. *Physica A: Statistical Mechanics and Its Applications*, 389(17), 3678–3691.

Haznagy, A., Fi, I., London, A., Nemeth, T. (2015) Complex network analysis of public transportation networks: a comprehensive study. *International Conference on Models and Technologies for Intelligent Transportation*, 371-378.

Kooij, R.E., Ellens, W. (2013) Graph measures and network robustness. *arXiv: 1311.5064*, 1-10.

Sarkar, S. (2013) Street network analysis for understanding typology in cities: Case study on Sydney CBD and suburbs. *State of Australian Cities Conference*, 1-9.

Xu, Q., Zu, Z.-H., Xu, Z.-J., Zhang, W.-D., Zheng, T. (2013) Space P based empirical research on public transport complex networks in 330 cities of China. *Journal of*

Transportation Systems Engineering and Information Technology, 13(1), 193–198.

Zhang, H., Zhao, P., Gao, J., Yao, X.-M. (2013) The analysis of the properties of bus network topology in Beijing basing on complex networks. *Mathematical Problems in Engineering,* 2013.

Zhang, L., Peng, L., Zhang, Y., Xue, S. (2012) Urban Bus Transport Network Optimization from Complex Network. *International Conference on Modelling, Identification and Control*, 1159-1164.

Zhang, J., Xu, X., Hong, L., Wang, S., Fei, Q. (2011), Networked analysis of the Shanghai subway network, in China, *Physica A: Statistical Mechanics and its Applications*, vol. 390, no. 23, pp. 4562-70.