Empirical investigation of topological and weighted properties of a bus transport network from China*

Shu-Min Feng(冯树民), Bao-Yu Hu(胡宝雨)†, Cen Nie(聂涔),
Xiang-Hao Shen(申翔浩), and Yu-Sheng Ci(慈玉生)

School of Transportations Science and Engineering, Harbin Institute of Technology, Harbin 150090, *China*

(Received 27 July 2015; revised manuscript received 20 November 2015; published online 18 January 2016)

Many bus transport networks (BTNs) have evolved into directed networks. A new representation model for BTNs is proposed, called directed-space P. The bus transport network of Harbin (BTN-H) is described as a directed and weighted complex network by the proposed representation model and by giving each node weights. The topological and weighted properties are revealed in detail. In-degree and out-degree distributions, in-weight and out-weight distributions are presented as an exponential law, respectively. There is a strong relation between in-weight and in-degree (also between out-weight and out-degree), which can be fitted by a power function. Degree–degree and weight–weight correlations are investigated to reveal that BTN-H has a disassortative behavior as the nodes have relatively high degree (or weight). The disparity distributions of out-degree and in-degree follow an approximate power-law. Besides, the node degree shows a near linear increase with the number of routes that connect to the corresponding station. These properties revealed in this paper can help public transport planners to analyze the status quo of the BTN in nature.

Keywords: bus transport network, directed and weighted network, empirical distributions, correlations

PACS: 05.90.+m, 01.75.+m **DOI:** 10.1088/1674-1056/25/3/030504

1. Introduction

There is a great interest in public transport networks as complex networks, involving the bus transport network and subway (or railway) network. Specifically, more and more public transport networks have been investigated in recent years: Sienkiewicz and Hołyst made a statistical analysis on public transport systems in 22 Polish cities.[1] Angeloudis and Fisk analyzed the world's twenty largest subway systems as complex networks.[2] von Ferber *et al.* presented a systematic survey of public transport networks of fourteen major cities in the world.[3,4] Xu *et al.* empirically studied complex features of public transport networks of 330 cities from China.[5] Statistical results showed that the public transport networks have some common features, such as exponential or power-law degree distribution, and typical small-world behaviors.

In previous studies, public transport networks were represented by three topological spaces:[1,6,7] space P, space L, and space R, which can be regarded as projections of a bipartite graph.[4,8,9] Some improved spaces were also proposed for actual public transport networks. Bi *et al.* proposed an improved space L (space S) to analyze urban rapid transit networks.[10] Yang *et al.* proposed an extended-space (ES) representation for bus transport networks based on space L (P), which comprised of ESL, ESP, and ESW networks.[11] To sum up, we noticed that most public transport networks were roughly described as undirected complex networks, in which the topological properties attracted more attention.

Recently, our investigation has discovered that the bus transport networks (BTNs) evolve into directed networks in some big cities (such as Harbin, Dalian from China), that is, there are lots of bus routes with different uplinks and downlinks. Zheng *et al.* built up a directed and weighted complex network model based on neighboring stops to analyze the topological properties of Beijing public transport network.[12] Therefore, it is necessary to establish a directed topological space for BTNs which can explore some new networked properties. Moreover, we find that nearly all the authors described BTNs as route networks or station networks and only focus on topological properties of the networks, and they commonly dismissed passenger flow on BTNs as another important composition of BTNs. We hold that some intrinsic properties of BTNs can be further revealed if passenger flow is considered in the networks.[13,14]

The remainder of this paper is organized as follows: Section 2 proposes a new representation model for BTNs, namely directed-space P. Section 3 details empirical distributions in directed and weighted BTN-H. Section 4 analyzes correlations and disparities in directed and weighted BTN-H. Section 5 gives a brief conclusion.

2. A new representation model for BTNs

In this paper, we have reported an empirical investigation of the BTN of Harbin City in 2013, which included 132 bus routes (excluding two loop routes), 993 stations and peak-hour

*Project supported by the National High Technology Research and Development Program of China (Grant No. 2014AA110304).
†Corresponding author. E-mail: hubaoyu87@163.com
© 2016 Chinese Physical Society and IOP Publishing Ltd

passenger production and attraction at each station. All the routes are divided into uplinks and downlinks. Specifically, we have counted the number of passengers who get on and get off the bus at each station in a morning peak hour by ride check, respectively, as shown in Fig. 1. Peak-hour passenger production at a station is the total number of passengers who get on the bus at the same station for each uplink and downlink. Peak-hour passenger attraction at a station is the total number of passengers who get off the bus at the same station for each uplink and downlink.

According to the investigation, it is such a common phenomenon that a bus route has different uplink and downlink in Harbin owing to single lanes and unidirectional passenger traffic, and its proportion is more than 36%. To describe the spatial characteristics of a directed BTN, we propose a new representation model for BTNs based on space P, called directed-space P.[11,12,15] The modeling can be divided into four steps as follows. A simple BTN is presented on the bus transport map, including two routes and nine stations (see Fig. 2(a)). The routes are divided into uplinks and downlinks. The bus transport map is abstracted into an extended bipartite graph, in which an uplink (or downlink) node and a station node are linked by an edge if the station is on the uplink (or downlink) (see Fig. 2(b)). A unipartite graph is obtained through down (up) projection from uplink nodes to station nodes (from downlink nodes to station nodes) (see Fig. 2(c)). The nodes are bus stations, and two nodes are linked by an edge if they are on the same uplink (or downlink) in Fig. 2(c). In the unipartite graph, the edges between two nodes are directed according to the direction of the corresponding uplink (or downlink) in Fig. 2(a), as shown in Fig. 2(d).

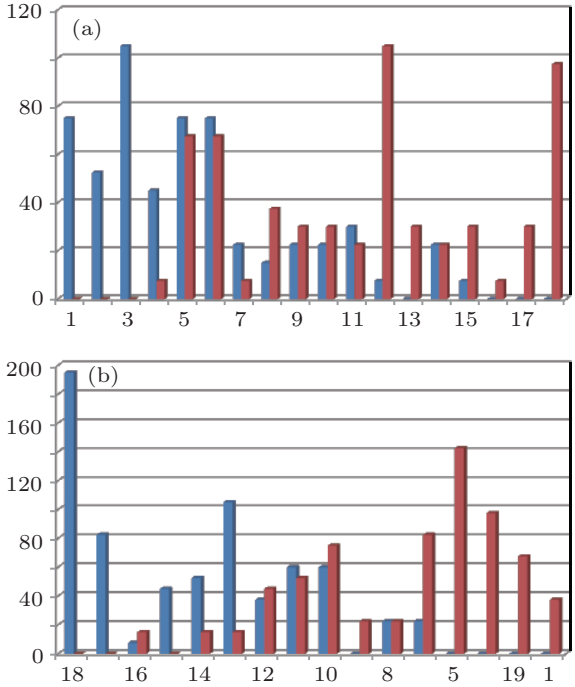


Fig. 1. (color online) The collected data of route 1 in investigation: (a) uplink; (b) downlink. From above: blue bar and red bar denote the number of passengers who get on and get off the bus. The station names are listed in Appendix A.

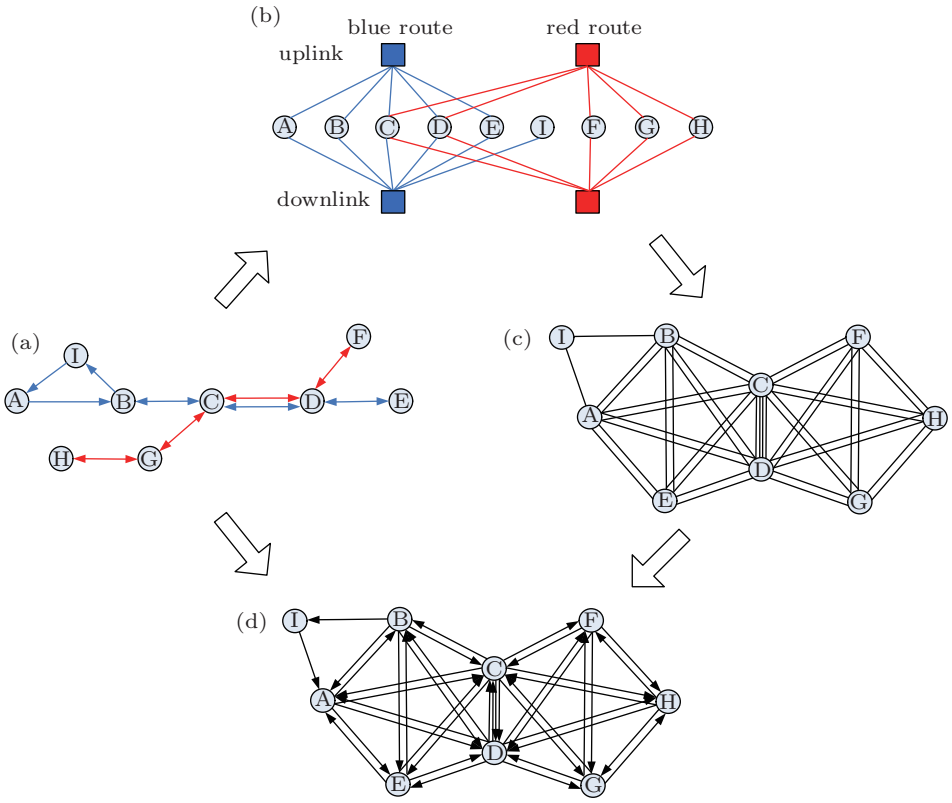


Fig. 2. (color online) (a) Bus transport map; (b) extended bipartite graph; (c) unipartite graph; (d) directed unipartite graph.

Figure 2(d) shows a directed and weighted network, which can be expressed by an asymmetrical adjacency matrix

$$\boldsymbol{A}^{\mathrm{w}} = \{a_{ij}^{\mathrm{w}}\}_{N\times N} = \begin{array}{c} \\ \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \\ \mathrm{E} \\ \mathrm{F} \\ \mathrm{G} \\ \mathrm{H} \\ \mathrm{I} \end{array} \begin{array}{c} \begin{array}{ccccccccc} \mathrm{A} & \mathrm{B} & \mathrm{C} & \mathrm{D} & \mathrm{E} & \mathrm{F} & \mathrm{G} & \mathrm{H} & \mathrm{I} \end{array} \\ \left(\begin{array}{ccccccccc} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 2 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 2 & 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \end{array}, \quad (1)$$

where a_{ij}^{w} is the number of adjacent edges (i.e., edge weight) from node i to node j, and N is the number of nodes in the graph.

3. Empirical distributions in directed and weighted BTN-H

BTN-H is described as a directed and weighted complex network by the proposed directed-space P and by giving each node weights. The network covers 993 nodes and 93682 directed edges. It means that each passenger has a travel route or routes on these directed edges. In particular, the travel route from node i to node j may be different from the travel route from node j to node i. Degree and node weight are important measures of centrality of a node in the network. Degree distribution, node-weight distribution, and relation between degree and weight are studied in detail to reveal topological and weighted properties of the network.

The node degree refers to in-degree and out-degree in a directed network. The in-degree of a node i is the number of edges with i as their terminal node (see Eq. (2)). The out-degree of a node i is the number of edges with i as their initial node (see Eq. (3)).

$$d_{\mathrm{in}}^{\mathrm{w}}(i) = \sum_j a_{ji}^{\mathrm{w}}, \quad (2)$$

$$d_{\mathrm{out}}^{\mathrm{w}}(i) = \sum_j a_{ij}^{\mathrm{w}}. \quad (3)$$

Figure 3 presents cumulative in-degree distribution and cumulative out-degree distribution. The two distributions follow an exponential law, with almost no difference. Exponential degree distributions are not surprising in most public transport networks.[5,6,8,16] A possible explanation behavior is revealed that new stations tend to be connected to the old ones at random during the expansion of BTN-H. From Fig. 3, we can see that BTN-H is an approximately symmetrical network, in other words, uplinks have almost the same stations with the corresponding downlinks in Harbin. Moreover, it can be inferred that those bus routes with different uplinks and downlinks have no impact on topological properties of BTN-H on the macro-scale.

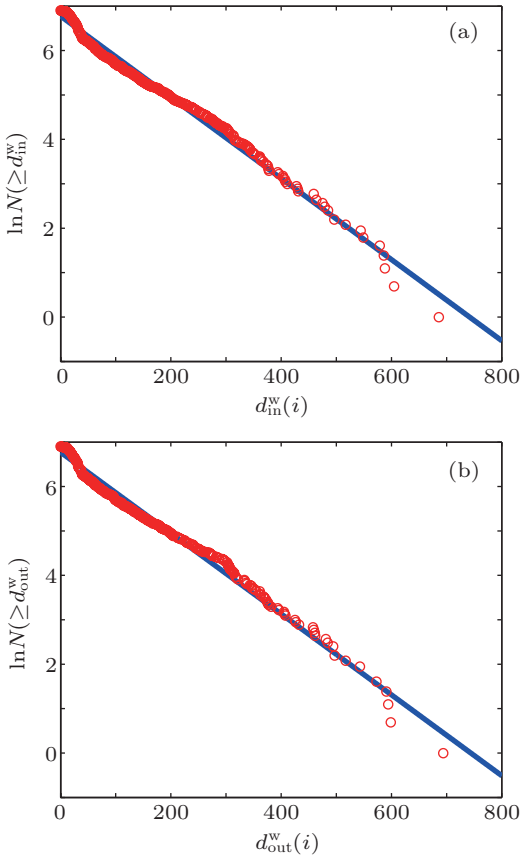


Fig. 3. (color online) (a) Cumulative in-degree distribution; (b) cumulative out-degree distribution.

Similarly, the node weight is divided into in-weight and out-weight. The in-weight $w_{\mathrm{in}}(i)$ of a node i is defined as peak-hour passenger attraction at the corresponding station. The out-weight $w_{\mathrm{out}}(i)$ of a node i is defined as peak-hour passenger production at the corresponding station. In terms of a node, in-weight and out-weight are usually unequal. Actually, the node weight is as important as its degree. If degree symbolizes the extrinsic importance of a node, then node weight symbolizes its intrinsic importance in the network. Put simply, the larger the node weight, the larger the corresponding degree. We can infer the BTN is designed reasonably according with passenger flow if degree distribution and node-weight distribution have the same law.

Figure 4 presents cumulative in-weight distribution and cumulative out-weight distribution, which have also an exponential decay. It indicates that passenger attraction (or production) is heterogeneously distributed at each station. From Figs. 3 and 4, we observe that both degree distribution and node-weight distribution follow an exponential law. This is not a coincidence because only the larger node degree can serve more passengers' travel. Therefore we have reasons to believe that BTN-H is designed reasonably. Additionally, we find that the out-weight distribution is more approximate to exponential distribution than the in-weight distribution. That may be because passengers' departure is given more consideration than passengers' arrival as BTN-H is designed.

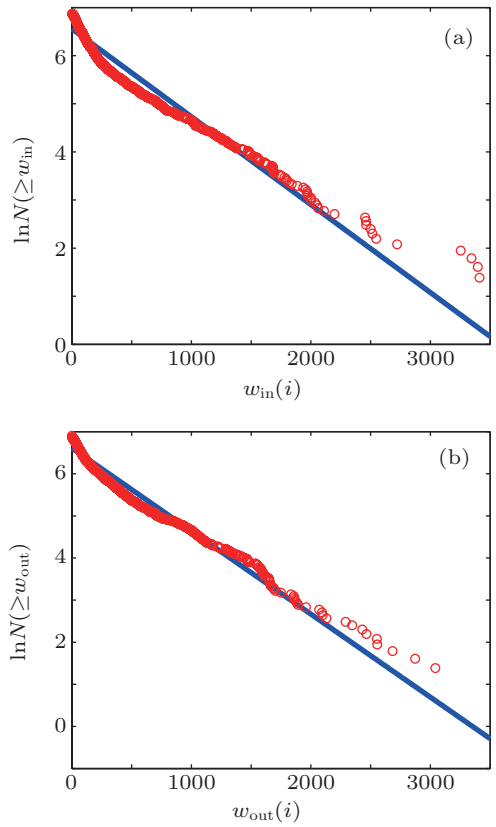


Fig. 4. (color online) (a) Cumulative in-weight distribution; (b) cumulative out-weight distribution.

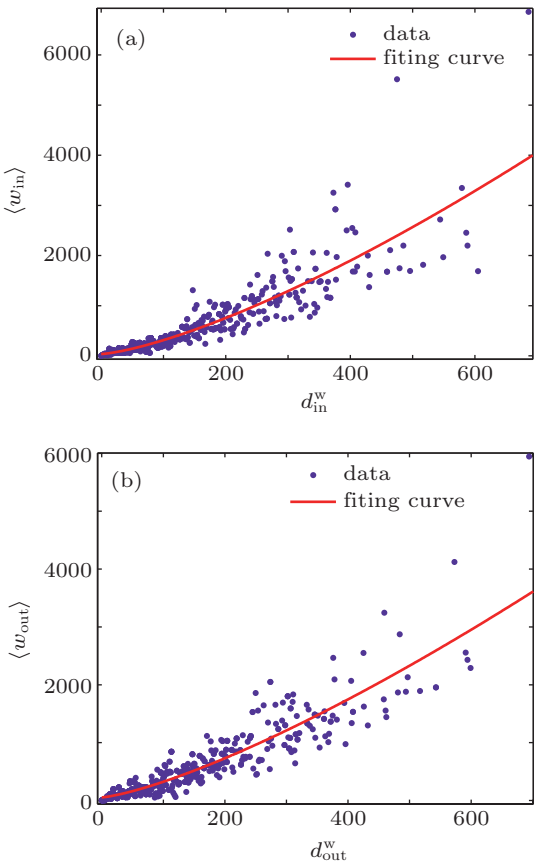


Fig. 5. (color online) (a) Relationship between in-degree and in-weight, $w\left(d^{\rm w}_{\rm in}\right) \propto \left(d^{\rm w}_{\rm in}\right)^{\beta_{\rm in}}$ with $\beta_{\rm in} = 1.374$; (b) relationship between out-degree and out-weight, $w\left(d^{\rm w}_{\rm out}\right) \propto \left(d^{\rm w}_{\rm out}\right)^{\beta_{\rm out}}$ with $\beta_{\rm out} = 1.325$.

To figure out the relationship between nodes' degree and weight, we investigate the dependence of $w_{\rm in}(i)$ on $d^{\rm w}_{\rm in}(i)$ and $w_{\rm out}(i)$ on $d^{\rm w}_{\rm out}(i)$, respectively. The dependence can be fitted by power functions

$$w\left(d^{\rm w}_{\rm in}\right) \propto \left(d^{\rm w}_{\rm in}\right)^{\beta_{\rm in}}, \tag{4}$$

$$w\left(d^{\rm w}_{\rm out}\right) \propto \left(d^{\rm w}_{\rm out}\right)^{\beta_{\rm out}}. \tag{5}$$

In Fig. 5, it is not unexpected that the node weight increases with its degree, $w_{\rm in}(i)$ versus $d^{\rm w}_{\rm in}(i)$ and $w_{\rm out}(i)$ versus $d^{\rm w}_{\rm out}(i)$. The power-law fits are given two parameters: $\beta_{\rm in} = 1.374$ and $\beta_{\rm out} = 1.325$. The values show that the node weight (i.e., passengers) has a faster growth than its degree.[17] Furthermore, there is a strong relationship between in-weight and in-degree (also between out-weight and out-degree), in spite of unavoidable fluctuations. This result is different from the tourism network in which only out-weight and out-degree have a strong relationship.[18] It can also be concluded that the topological structures of BTN-H are codetermined by passenger production and attraction at each station.

4. Correlations and disparities in directed and weighted BTN-H

We turn now to investigate degree–degree and weight–weight correlations between connected nodes in the network to illustrate whether the network has an assortative or disassortative behavior. An assortative network is one where high-degree (or high-weight) nodes have high-degree (or high-weight) nearest neighbors, while an disassortative network has high-degree (or high-weight) nodes that have low-degree (or low-weight) nearest neighbors.[18,19]

Degree–degree correlation can measured by the average nearest-neighbor degree. Considering that the network is directed, we define the average nearest neighbors' out-degree and in-degree for a given node as

$$d^{\rm w}_{\rm nout}(i) = \frac{1}{d^{\rm w}_{\rm in}(i)} \sum_{j\in v(i)} d^{\rm w}_{\rm out}(j), \tag{6}$$

$$d^{\rm w}_{\rm nin}(i) = \frac{1}{d^{\rm w}_{\rm out}(i)} \sum_{j\in v(i)} d^{\rm w}_{\rm in}(j), \tag{7}$$

where $v(i)$ is the set of nearest neighbors of node i. We calculate the average nearest neighbors' out-degree for the nodes with the same in-degree, and the average nearest neighbors' in-degree for the nodes with the same out-degree separately,

$$d^{\rm w}_{\rm nout}(k) = \frac{1}{N_{\rm in}(k)} \sum_{d^{\rm w}_{\rm in}(i)=k} d^{\rm w}_{\rm nout}(i), \tag{8}$$

$$d^{\rm w}_{\rm nin}(k) = \frac{1}{N_{\rm out}(k)} \sum_{d^{\rm w}_{\rm out}(i)=k} d^{\rm w}_{\rm nin}(i), \tag{9}$$

where $N_{\rm in}(k)$ is the number of nodes with in-degree k, $N_{\rm out}(k)$ is the number of nodes with out-degree k.

Analogously, weight–weight correlations can be measured by the average nearest-neighbor weight. We define the average nearest neighbors' out-weight and in-weight for a given node as

$$w_{\mathrm{nout}}(i) = \frac{1}{w_{\mathrm{in}}(i)} \sum_{j \in v(i)} w_{\mathrm{out}}(j), \tag{10}$$

$$w_{\mathrm{nin}}(i) = \frac{1}{w_{\mathrm{out}}(i)} \sum_{j \in v(i)} w_{\mathrm{in}}(j). \tag{11}$$

Actually, the value of the average nearest-neighbor degree (or weight) is more than 1. It also denotes the node's impact on its nearest neighbors, that is, the lower the neighbors' degree (or weight), the greater the node's impact.

Figures 6 and 7 show degree–degree correlations and weight–weight correlations in BTN-H. We find that the nearest neighbors of the nodes with low degree (or weight) have large fluctuations in the average degree (or weight). The fluctuations result in weak degree–degree (or weight–weight) correlations. In other words, BTN-H has no an assortative or disassortative behavior as the nodes have relatively low degree (or weight). But the neighbors' degree (or weight) has a negative trend, revealing that BTN-H has a disassortative behavior as the nodes have relatively high degree (or weight). Additionally, the negative trends demonstrate also that the nodes have a growing impact on their neighbors.

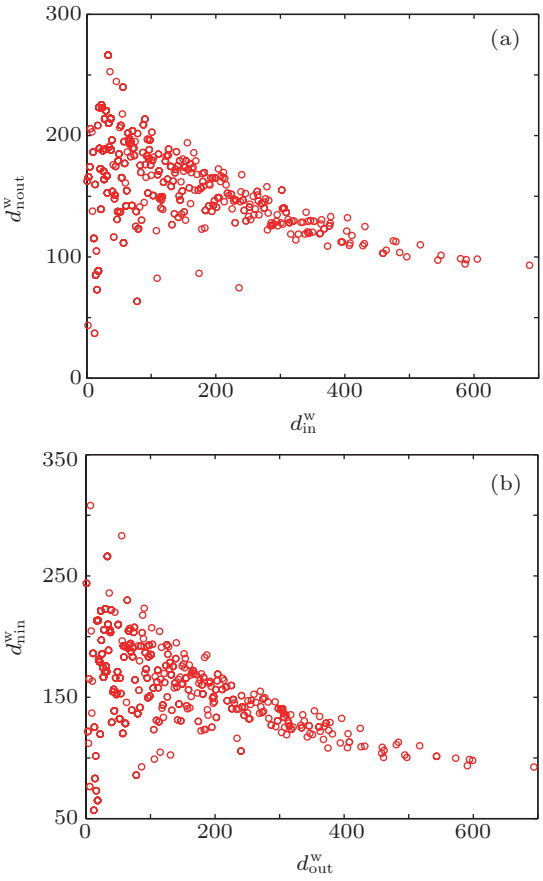


Fig. 6. (color online) Degree–degree correlations for (a) nodes' in-degree versus neighbors' out-degree and (b) nodes' out-degree versus neighbors' in-degree.

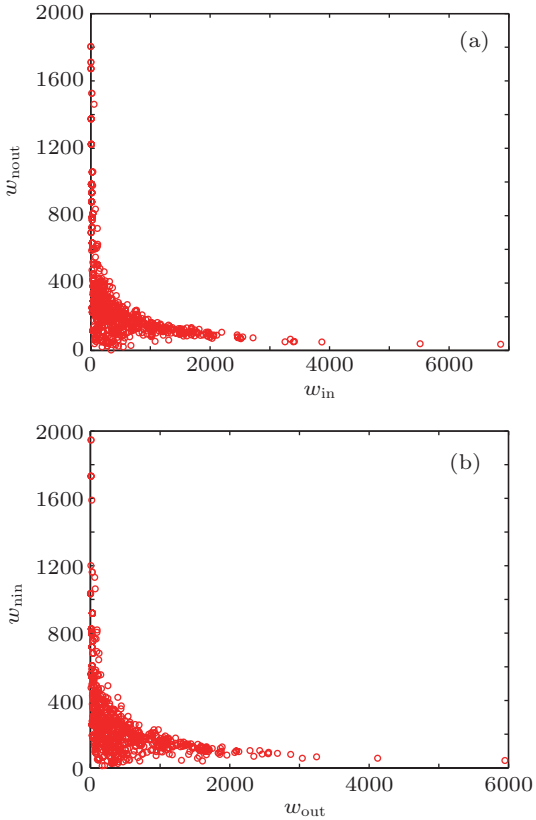


Fig. 7. (color online) Weight–weight correlations for (a) nodes' in-weight versus neighbors' out-weight and (b) nodes' out-weight versus neighbors' in-weight.

As another important topological parameter, we focus on the disparity of the node degree, which can measure the homogeneity of distribution of the node's adjacent edges. For the directed network, we define the disparities of out-degree and in-degree for a given node as

$$Y_{\mathrm{out}}(i) = \sum_{j \in v(i)} \left[\frac{a^{\mathrm{w}}_{ij}}{d^{\mathrm{w}}_{\mathrm{out}}(i)} \right]^2, \tag{12}$$

$$Y_{\mathrm{in}}(i) = \sum_{j \in v(i)} \left[\frac{a^{\mathrm{w}}_{ji}}{d^{\mathrm{w}}_{\mathrm{in}}(i)} \right]^2, \tag{13}$$

where a^{w}_{ij}, a^{w}_{ji} are defined in Eq. (1). The disparity is inversely proportional to the corresponding node degree if all the edge weights are equal. The disparity reaches the maximum 1 if a node's adjacent edges only connected with a neighbor.

We average the disparities of out-degree (or in-degree) for the nodes with the same out-degree (or in-degree)

$$Y_{\mathrm{out}}(k) = \frac{1}{|N_{\mathrm{out}}(k)|} \sum_{d^{\mathrm{w}}_{\mathrm{out}}(i)=k} Y_{\mathrm{out}}(i), \tag{14}$$

$$Y_{\mathrm{in}}(k) = \frac{1}{|N_{\mathrm{in}}(k)|} \sum_{d^{\mathrm{w}}_{\mathrm{in}}(i)=k} Y_{\mathrm{in}}(i). \tag{15}$$

Figure 8 shows the disparity distributions of out-degree and in-degree. We can see most nodes have very small disparities, illustrating that most stations have relatively homo-

geneous connections with other stations in BTN-H. Interestingly, each of the two distributions with a long tail follows an approximate power-law (see insets). We can infer that out-degree and in-degree have severely heterogeneous disparities for all the nodes, that is, there are significant differences between the tiny number of the nodes with maximum disparities and the great number of the nodes with minimum disparities.

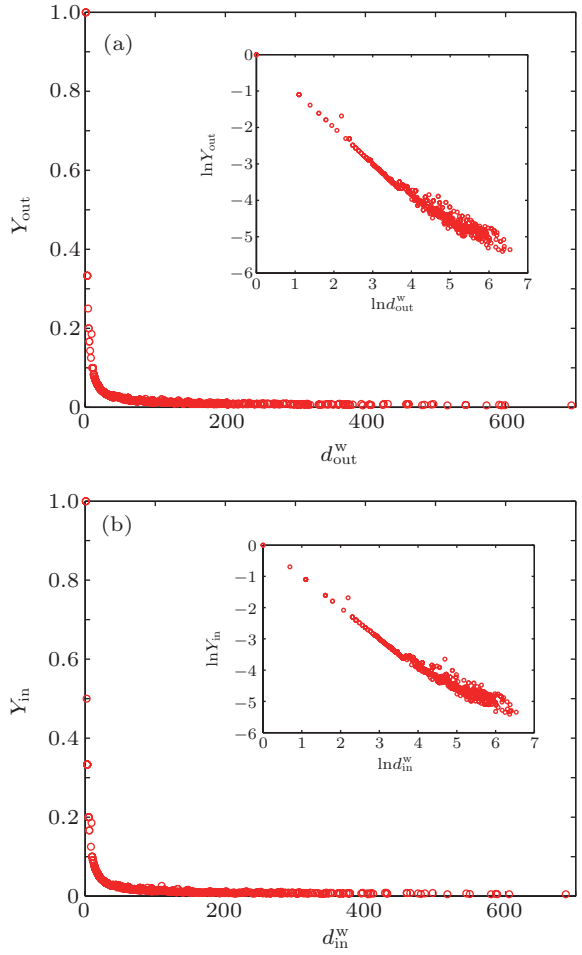


Fig. 8. (color online) (a) Disparity distribution of out-degree; (b) disparity distribution of in-degree.

Besides, the number of routes that connect to a station (NSS) is another important parameter measuring the centrality of the station (node). Here, we examine the degree-route correlation between the node degree and NSS. The node degree equals the sum of in-degree and out-degree as

$$d^{\rm w}(i) = d^{\rm w}_{\rm in}(i) + d^{\rm w}_{\rm out}(i), \tag{16}$$

and NSS is half of the number of uplinks and downlinks connected to the station. As seen in Fig. 9, we find that the node degree shows a near linear increase with the number of routes that connect to the corresponding station in BTN-H. It is reasonable to deduce that the two measures provide almost the same information for the node centrality.[17]

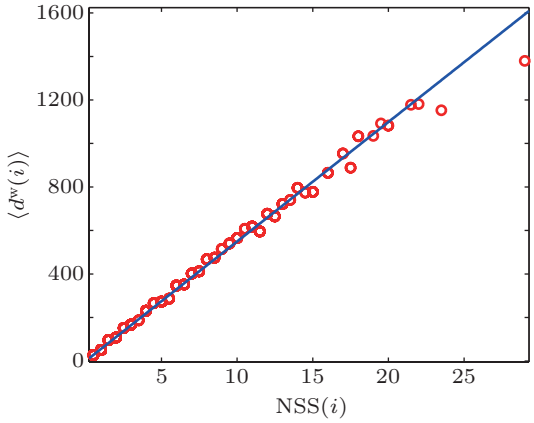


Fig. 9. (color online) Degree-route correlation in BTN-H.

5. Conclusions

In this paper, uplinks and downlinks are separately considered to investigate the bus transport network of Harbin (BTN-H). The directed-space P representation model for BTNs is proposed. BTN-H is described as a directed and weighted complex network by the new representation model and by giving each node weights. The topological and weighted properties are revealed as follows.

In-degree and out-degree distributions, and in-weight and out-weight distributions are presented as an exponential law, respectively. It is believed that BTN-H is designed reasonably. In-degree and out-degree distributions have almost no difference, illustrating that BTN-H is an approximately symmetrical network. A possible explanation behavior is also revealed that new stations tend to be connected to the old ones at random during the expansion of BTN-H. Out-weight distribution is more approximate to exponential distribution than in-weight distribution. That may be because passengers' departure is given more consideration than passengers' arrival as BTN-H is designed. There is a strong relation between in-weight and in-degree (also between out-weight and out-degree), in spite of unavoidable fluctuations. It can also be concluded that the topological structures of BTN-H are codetermined by passenger production and attraction at each station.

Degree–degree and weight–weight correlations are investigated to reveal that BTN-H does not have an assortative or disassortative behavior as the nodes have relatively low degree (or weight), while BTN-H has a disassortative behavior and the nodes have a growing impact on their neighbors as the nodes have relatively high degree (or weight). The disparity distributions of out-degree and in-degree follow an approximate power-law, inferring that most stations have relatively homogeneous connections with other stations in BTN-H. Besides, the node degree shows a near linear increase with the number of routes that connect to the corresponding station. It

can be deduced that the two measures provide almost the same information for the node centrality.

These properties revealed in this paper can help public transport planners to analyze the status quo of the BTN in nature. However, we have only collected peak-hour passenger production and attraction at each station. We will investigate the origin–destination (OD) passenger flow in future to explore more intrinsic properties of BTNs.

Appendix A

Table A1. The station names corresponding to the station numbers shown in Fig. 1.

Station number	Station name	Station number	Station name
1	CHELIANGWENHUAGONG	11	JINGYUSHIERDAOJIE
2	JINGWEISHIDAOJIE	12	SHISIYUAN
3	JINGWEIBADAOJIE	13	JINGYUERSHIDAOJIE
4	DAOLISHIERDAOJIE	14	SONGPUDAQIAO
5	ZHAOLINJIE	15	GENGXINJIE
6	MAIMAIJIE	16	SHIYOUCHANG
7	NANMALU	17	MUCAISHICHANG
8	CHENGDEGUANGCHANG	18	GANGWUJU
9	JINGYUTOUDAOJIE	19	GAOYIJIE
10	JINGYUQIDAOJIE		

References

[1] Sienkiewicz J and Hołyst J A 2005 *Phys. Rev. E* **72** 046127
[2] Angeloudis P and Fisk D 2006 *Physica A* **367** 553
[3] von Ferber C, Holovatch T, Holovatch Y and Palchykov V 2007 *Phys. A* **380** 585
[4] von Ferber C, Holovatch T, Holovatch Y and Palchykov V 2009 *Eur. Phys. J. B* **68** 261
[5] Xu Q, Zu Z H, Xu Z J, Zhang W D and Zheng T 2013 *J. Transp. Syst. Eng. Inf. Technol.* **13** 193 (in Chinese)
[6] Xu X P, Hu J H, Liu F and Liu L S 2007 *Physica A* **374** 441
[7] Chen Y Z, Fu C H, Chang H, Li N and He D R 2008 *Chin. Phys. B* **17** 3580
[8] Chen Y Z, Li N and He D R 2007 *Physica A* **376** 747
[9] Zhang Y, Jin X X, Zhang Y and Yao D Y 2007 *Syst. Eng.-Theory & Pract.* **49** (in Chinese)
[10] Bi X Y, Chen X W and Li Z M 2011 *J. Transp. Syst. Eng. Inf. Technol.* **11** 173 (in Chinese)
[11] Yang X H, Chen G, Chen S Y, Wang W L and Wang L 2014 *Transp. Res. Part A* **69** 1
[12] Zheng X, Chen J P, Shao J L and Bie L D 2012 *Acta Phys. Sin.* **61** 190510 (in Chinese)
[13] De Montis A, Barthelemy M, Chessa A and Vespignani A 2007 *Environ. Plann. B* **34** 905
[14] Ou Q, Jin Y D, Zhou T, Wang B H and Yin B Q 2007 *Phys. Rev. E* **75** 021102
[15] Feng S M, Hu B Y, Nie C and Shen X H 2016 *Physica A* **441** 85
[16] Zhang J H, Zhao M W, Liu H K and Xu X M 2013 *Physica A* **392** 1538
[17] Barrat A, Barthelemy M, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl. Acad. Sci. USA* **101** 3747
[18] Miguens J I L and Mendes J F F 2008 *Physica A* **387** 2963
[19] Soh H, Lim S, Zhang T Y, Fu X J, Lee G K K, Hung T G G, Di P, Prakasam S and Wong L 2010 *Physica A* **389** 5852