Study on Properties of Traffic Flow on Bus Transport Networks

XU-HUA YANG, JIU-QIANG ZHAO, GUANG CHEN, AND YOU-YU DONG
College of Computer Science and Technology
Zhejiang University of Technology
Hangzhou, 310023, China
P.R.China
xhyang@zjut.edu.cn

Abstract: - In this paper, we investigate the statistical properties of traffic flow on the urban bus transport network (BTN) in the city Hangzhou of China. To alleviate the congestion and simulate properties of traffic flow on BTN, we construct a model using Hangzhou BTN which considers the short walking distance between two nearby stations and the maximum transfer frequency. We select four parameters to evaluate the characteristics of traffic flow on the BTN. We find that the results obtained from the simulation model are similar with the theoretical model irrespective of congestion. If congestion is considered, the situation is different. Besides, if we increase the departure frequency of the buses with some large passenger flow lines, it will lead to the critical packets generating rate increasing on the whole network tremendously. It indicates that buses with larger passenger flow are more important in the BTN, so we can increase the departure frequencies to improve the network performance. Our results provide a method to minimize the congestion on BTN.

Key-Words: - BTN, traffic flow, passenger flow, relieving congestion

1 Introduction

During the last few years, studies of transportation networks have drawn lots of attention of network scholars[1]-[8]. These transportation networks are called public transportation networks, in which the nodes in the networks represent stations while the edges are the links between two stations which mean that we can reach from one station to another directly. These public transportation networks such as bus networks[9]-[12], railway networks[13],[14] and airport networks[15],[16] affect our daily life and nation’s economy in many respects[17], so they become the study focus of the scholars.

In this paper, we investigate the urban bus transportation network (BTN) in city Hangzhou of China. To our knowledge, the urban bus system is essential to maintain the normal operation of the city and takes an important role in our daily life. In the BTN, the nodes are bus stations and the edges are links connected by the bus routes between two stations. Through the research of the BTN, we can obtain many statistical properties of the bus networks. Although at present there have been a lot of previous works on the BTN, they mostly focus on the static characteristics of the networks, the dynamic characteristics are rarely discussed. The problem of traffic flow on BTN is one of the important aspects in the research as it directly reflects the performance of the bus networks. The traffic flow of networks has been widely researched by many scholars[18]-[21], however they often explore it from abstract perspective, or use the abstract model in their research, or the definition of traffic flow is too simple (for example they only consider node congestion or edge congestion). Although abstract method can help us find the common characteristics in different kinds of networks, it may also lead to the
results obtained from the model conflict with the actual network. In our paper, we construct a model of traffic flow under some assumptions which agree with the actual situation to analysis the properties of traffic flow of bus networks and then compare them with the empirical investigations obtained from bus systems in city Hangzhou. The conclusions should be applicable for many real systems and helpful for the corresponding research of traffic flow.

This paper is organized as follows. In the next section, the model of traffic flow is described in details. In section 3, the results of the simulation of theoretical model and computer simulation are provided and discussed. Section 4, presents our conclusions.

2 Model

In the problem of traffic flow of networks, the packet is a basic concept. Every packet is generated with information of its sources and destinations. We name the pair of the initial node and the terminal node to be the OD pair. Once a packet is generated, it will be added to the networks and enters the life cycle. Then it tries step by step from the initial node to reach the terminal node along one certain route. The life cycle is over after the packet reaches its destination, then the packet is removed from the network. During its life cycle, the move forward of the packet along the given path generates traffic flow. In fact, the research of traffic flow comes down to analysis of the properties of traffic flow on network components (nodes and edges) under different situations.

It is worth noting that the topology structure of networks, the distribution of OD pairs and the choice of the packet path all of them will affect the traffic flow on networks. In Fig. 1, it shows that a bilateral unweighted network exhibits different properties under the influence of the three factors mentioned above.

So the analysis of the traffic flow should focus on these three factors. In order to obtain results agree with the actual situations, we must make these factors approach to the actual situations. From the abstract view, BTN has many common properties with lots of actual networks, such as small world characteristic, exponent degree distribution or power degree distribution and so on. But BTN has its particular properties. Different from many networks, the basic components of BTN are routes. In the actual motion of BTN, the basic components can be represented by buses which are interval driving on every route with fixed traffic capacity. This feature indicates that we should apply different method in study of BTN compared with traditional model.

Fig.1 A bilateral unweighted network influenced of three factors. (a). the topology structure of the network. (b). under even distribution of OD pairs, the distribution of the traffic flow as all the packets move forward along the shortest path. (c). under given distribution of OD pairs (the packets between node 3 and node 4 is double than the packets between other pairs), the distribution of the traffic flow as all packets move forward along the shortest path. (d). under even distribution of OD pairs, the distribution of the traffic flow as all packets move forward along the given path (the shortest path which excludes the edge between node 2 and node 3).

For BTN, we can’t change its topology structure. For the distribution of the OD pairs of the network, we can obtain it through statistical method. So we pay our attention on the choice of the packet path. It is reasonable to use the shortest path along which packets move forward in the networks to approximately represent the actual path in traditional traffic flow models, such as in communication networks. But, for BTN, the packets represent passengers who travel through BTN. Passengers
choose path has nothing to do with the links in networks but based on the basic components of BTN, namely, the routes. This makes the actual path often different from the shortest path, as showed in Fig.2. So in the BTN model, we have to define a passenger path to satisfy actual situation.

![Fig.2](image)

Fig.2 A BTN network formed by two groups of four routes. Each group is connected by solid lines with different colors and contains two routes (the uplink and the downlink). Due to the stations in both routes of one group are the same, so this network is symmetric. The node 2 and node 8 is connected by purple dotted line which represents the distance between them is smaller than the walking distance threshold, so they can walk to each other. The distance of the each line represents the physical distance between two stations. According to the shortest path, node 1 should move to node 6 along the path 1-2-3-6, in fact passengers move along the path 1-2-3-4-5-6. Because the former case needs to take two buses, instead the latter case just needs one bus.

Traditionally, BTN can be mapped into two spaces: space P and space L [2], [9], [21]. In both spaces, one node represents one bus stations. In space P an edge connects a pair of stations when there is a bus traveling between them. If passengers have to exchange routes or buses, the pair of stations is connected by more edges. In space L, an edge between two stations exists if they are consecutive stations on the route. Space L reflects the topology of BTN while space P shows the characteristic of BTN transfer. Only if we find the shortest path between two nodes in space P, we can determine the least transfer times between the corresponding stations.

This method doesn't take into account the weight of network, we trivially find the least transfer times but the path length may be very long. In this paper, we propose weighted space P network (WSP) in which the weight of random edge is equal with the minimum value of the actual driving distance on the route between the corresponding stations. Similarly, we define weighted space L network (WSL). We can obtain the shortest transferred path based on the least transfer times through evaluating the shortest path of WSP.

But the shortest transferred path obtained from WSP may also have difference with the actual situations. Because in addition to the same station transfer, there is also transfer in different stations. In the latter case, passengers get off one bus at one station and walk to another station to take the next bus. Obviously, different station transfer considers more about spatially characteristic than the WSP network; it may help reduce the transfer diameter. Therefore in order to achieve the transfer in different stations, we propose another network WSW which contains the walking information between any two stations. In fact, passengers will never accept long distance walking, so we just store the pair of nodes between which the walking distance is below a threshold (in this paper, we set the threshold as 0.5km).

Combined WSW and WSP and based on the principle of priority on walking, we get the best transferred path between any two stations. Fig.3 shows the path tree generated from node 1 in Fig.2.

![Fig.3](image)

Fig.3 The path tree of node 1. The solid line represents the shortest path from one station to another by bus; the dotted line represents the eliminated path. The path of the solid line is not
marked completely, for example from node 1 to node 4; the path can be through the red line from node 1 via node 2 and node 3 to node 4.

In our method, we not only give the stations on the way from the initial station to the terminal station, but also recessively give the bus lines from one station to the next. For example, in Fig.3, there is blue line between node 1 and node 7, which indicates that line 2 is the most optimal path. Though we can get the most optimal path, but it is just a kind of expectation. In practice, when passengers travel from station $i$ to station $j$, they may not choose the shortest path, instead they choose the path which first reach node $i$ and can reach node $j$ directly at the same time the path length is similar to the shortest path length. Because of passengers don’t like to take a roundabout bus. So we set a threshold $T$, when a passenger at one station $i$ wants to reach another station $j$ by a bus, if the ratio of the path length between two stations with the length between two corresponding nodes in space $P$ is smaller than $T$, he chooses this bus. Here we define the set of routes between any two neighbor stations $j$ and $k$ which meets above requirement as $LS_{jk}$, we will use this set in the subsequently analysis.

The path defined above meets with the actual situation, so we choose this path as the passenger path and analysis the characteristic of traffic flow on BTN. First we define the distribution of OD pairs as matrix $POD$. $POD$ is a $n \times n$ matrix in which $n$ represents the number of nodes of the network, for BTN it means the number of stations. A component $POD_{ij}$ is the probability that randomly generate a packet whose initial node is $i$ and terminal node is $j$. The matrix $POD$ can be obtained by sampling survey of passengers at every station. Combined matrix $POD$ and passenger path, we can get the betweenness of every station and every link in the WSL of BTN. We define the betweenness of any node $i$ as $NB_i$ represents the number of packet paths via this node. In the WSL, the betweenness $EB_{ij}$ of any link which connects node $i$ and node $j$ represents the number of packet paths via this link. The equation (1) shows more details of the definition of $NB_i$ and $EB_{ij}$ in which $P_{mn}$ represents the path of OD pair $m,n$.

$$
NB_i = \sum_{i \in P_{mn}}^{i \neq m,i \neq n} POD_{mn}
$$

$$
EB_{ij} = \sum_{ij \in L_{mn}} POD_{mn}
$$

For BTN, $NB_i$ and $EB_{ij}$ reflect the situation of passenger flow on stations and routes (where is congested and where is free in the network), this information can help to optimize the BTN network but it can’t reflect the characteristics of traffic flow completely. That’s because the basic components of BTN are routes and buses on the routes, research without the basic components makes no sense. If we can get more information of passenger flow of every bus route (which routes are congested and which routes are free in the BTN), then we can comprehend the characteristics of traffic flow more. We define the betweenness $SB_{ijk}$ of any section $j,k$ on route $i$ as the number of packet paths via this section and shared by this section. Apparently, for any bus route $i$ which belongs to the $LS_{jk}$ set of one section $j,k$, the shorter its departure interval time is, the more likely to share packet path on this section. For one route, we pay our attention on the section with the largest passenger flow. We define its flow as $SBB_i$. $SB_{ijk}$ and $SBB_i$ can be expressed in equation (2) in which $L_i$ represents the departure
The betweenness $SB$ reflects the traffic flow on a certain section of one bus route. There may be many buses on one bus route, so we can obtain the characteristics of traffic flow of every bus on a certain section. For any bus belongs to the bus route $i$, we define $SBF_{ijk}$ as the average traffic flow on any section $j, k$ of this bus. Obviously, $SBF_{ijk}$ is proportional to $SB_{ijk}$ and also proportional to its departure interval time. Because if the departure interval time is less, there are more buses running at the same time, the passenger flow assigned to every bus is fewer. For one route $i$, we focus on the section which have the biggest passenger flow on every bus, we define it as $SBBF_i$. $SBF_{ijk}$ and $SBBF_i$ can be expressed in equation (3).

$$SBF_{ijk} = SB_{ijk} \cdot \frac{L_i^{-1}}{L_i} \sum_{l \in LS_k} POD_{mn} \cdot \frac{1}{\sum_{l \in LS_k} L_i}$$

$$SBBF_i = \max(SBF_{ijk})$$

Definitions of $NB, EB, SBB, SBBF$ in the above give us index to analysis the characteristics of traffic flow on BTN network.

### 3 Simulations

From the Internet we get the information of BTN in Chinese city Hangzhou which contains routes, stations and so on. We show the information in Table.1. According to these information, we can generate WSP, WSL and WSW networks and get transferred paths among stations. Based on the information, we respectively simulate the traffic flow of BTN in Hangzhou from theoretical model and computer simulations.

The strategy of generation of OD pairs plays a vital role in the simulations of traffic flow, so we have to set the distribution of the OD pairs of passengers. In the process of simulations, different generation strategy of OD pairs will produce different results. We use the distribution of static OD pairs to conduct our theoretical simulation. In our distribution, every time step we choose the starting point of a new passenger according to the priority of out-degree of stations. Namely, it is more likely to be chosen if the value of the out-degree is bigger. Whereas, the terminal point is chosen by the priority of the in-degree of stations, namely it is more likely to be chosen if the value of the in-degree is bigger. In this paper, we define the out-degree($od$) of any station $i$ as the number of stations which can be reached from station $i$ through matrix WSL or WSW. The in-degree($id$) is defined as the number of stations which reach the station $i$ through matrix WSL or WSW. Therefore the possibility is a station chosen to be the start point and the possibility is a station chosen to be the terminal point. They can be respectively expressed in equation (4).

$$P_{od} = \frac{od_i}{\sum_j od_j}$$

$$P_{id} = \frac{id_i}{\sum_j id_j}$$

<table>
<thead>
<tr>
<th>City</th>
<th>SN</th>
<th>RN</th>
<th>AS</th>
<th>AL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hangzhou</td>
<td>2750</td>
<td>509</td>
<td>20.34</td>
<td>4.19</td>
</tr>
</tbody>
</table>

Table 1 | The parameters of BTN in Hangzhou. SN is the total number of stations. RN is the total number of routes. AS is the average number of stations belong to one bus route. AL is the average number of bus routes distributed to one bus station.
This selection strategy is more realistic due to passengers will be more in places where buses are intensive. According to $P_o$ vector and $P_d$ vector composed by each station, we can get trip distribution matrix $POD$ as we show it in equation (5). For every element $POD_{ij}$ in the matrix $POD$, it can be expressed as $POD_{ij} = P_o \cdot P_d^T$.

$$POD = P_o \cdot P_d^T$$ (5)

We still need to do further treatment on the matrix $POD$. In addition to setting the diagonal elements as 0, we also need to set those elements which can be reached by walking from starting point to terminal point directly or passengers have to take more than three buses as 0. In the above two cases, passengers can’t reach the destination only by bus. The dealt matrix $POD$ is normalized to get distributing matrix of OD pairs which will be used in our simulations.

With the normalized $POD$ matrix and paths obtained by calculation, we can simulate the characteristics of traffic flow on BTN in Hangzhou as our theoretical model results. In order to simulate the actual situation better, we use our new strategy which shortens the bus departure time to half in the busiest 10% bus lines to simulate, namely we divide bus routes into two groups, group A and group B which contains the busiest 10% bus lines. We set the departure frequencies of group B twice than of group A. The ratio of the number of routes in group A and in group B is 9:1. We obtain the characteristics of BTN in Hangzhou with our theoretical model. The result is showed in Fig. 4.

Fig. 4| The characteristics of traffic flow in BTN with the theoretical model. (a). the cumulative probability distribution of $NB$, note that the parameter $NB$ is normalized, namely is the ratio of $NB$ with the maximum $NB$. (b). the cumulative probability distribution of $SBBF$, note that the parameter $SBBF$ is normalized, namely is the ratio of $SBBF$ with the maximum $SBB$. (c). the sorting distribution of $SBB$ and $SBBF$, which are shown in the blue and red line respectively. (d). the relation between $SBB$ and $SBBF$. Note that all parameters are normalized.
The subgraph (c) in Fig.4 indicates that the maximum of parameter $SBBF$ is 0.5, because of the departure frequencies of group A is twice than those of group B. The subgraph (d) in Fig.4 indicates our theoretical analysis which contains two straight lines. The lower one represents the route of group B while the higher one represents the route of group A. Furthermore, slope of the lower line is one half of the higher one, because the departure frequency of group B is twice than of group A. The result agrees with our former theoretical analysis.

Next we continue to do computer simulation of traffic flow on BTN. In the program, a bus is restricted by passengers and move along the given bus route. Therefore we need three characters, the bus, the passengers and the system. For the running bus, it has four actions, drive, reach, drop off passengers and pick up passengers. For the passengers, it has four actions, walk, reach, get on the bus and get off the bus. For the system, it also has four actions, generate new passengers, generate new buses, recycle reached passengers and recycle reached buses. Our model is worked by time steps, first we assume that different buses drive at the same speed and different passengers walk at the same speed and their speed don’t change over the time. Meanwhile, we ignore the time of buses reaching the station and driving out and the time of passengers getting on and getting off. In order to ensure these three characters running well, we have to rule that the driving distance of a bus in one time step is no longer than the minimum value of the links in WSP network. In every time step, the details are showed as follows:

(1) The system generates the same number of new passengers according to matrix $POD$ and sends new buses for the routes which have new demand based on the departure time interval.

(2) Exploit the driving distance in one time step to update all the driving buses. For the buses which have reached the stations, we record their reaching state, passengers getting off and the change of the state of the passengers. Moreover, for the buses which have reached the destination, we recycle them.

(3) Exploit the walking distance of passengers in one time step to update the state of walking passengers.

(4) According to the record, passengers of corresponding stations get on and change their state.

Based on the process above, the passengers who don't reach the destination have three states: walking, waiting or on one bus. These three states correspond to three sets: the set of walking passengers, the set of waiting passengers and the set of passengers on bus. Obviously, there is only one set of walking passengers, the set of waiting passengers while every driving bus has a set of passengers on bus. The changes of state of passengers mentioned above are all about extracting passengers from one set and putting them into another set (Besides the changes also contain the recovery of passengers who have reached their destination).

With the MATLAB software, we do the simulation on BTN. We propose a new parameter, the passenger capacity of bus which helps to better understand the congestion on BTN. In practice, congestion will happen on BTN due to the passenger capacity of one route is limited. When a bus stop at one station, if passengers who want to get on are more than the rest of passenger capacity after some passengers get off, some passengers can’t get on the bus. If the situation is serious, it causes congestion. To deal with this problem, we randomly choose the same number of passengers from the waiting set of the bus route to the rest of passenger capacity to get on the bus.

After introducing the biggest passenger capacity, the congestion of BTN appears. An important
parameter to evaluate the congestion is the critical packets generating rate $R_c$. It measures the network’s capacity. At $R_c$, the phase transition occurs when the network’s state changes from freedom to congestion. When the packet generate rate $R$ is less than $R_c$, the number of new generating packets are equal with new reaching packets that means the network is free. When the packet generate rate $R$ is larger than $R_c$, due to the limitation of the capacity of nodes or edges, the number of new generating packets is larger than new reaching packets. The number of packets in life cycle becomes larger and larger and finally causes the congestion. The traffic flow can be expressed as follow:

$$
\eta(\rho) = \lim_{\rho \to \infty} \frac{C\langle \Delta W \rangle}{\rho N \Delta t} \quad (6)
$$

Where $\Delta W = W(t+1) - W(t)$, $\langle \Delta W \rangle$ represents the average value of $\Delta W$. $W(t)$ represents the number of packets exist at time $t$. $N$ represents the total number of nodes in the network. Apparently, when $R$ is less than $R_c$, $\langle \Delta W \rangle = 0$ and $\eta = 0$, the network is in free. When $R$ is larger than $R_c$, $\eta > 0$, the network congest.

We choose the same POD matrix and path with the theory simulation and set the biggest passenger capacity of every bus to be the same. Then we simulate and obtain the critical packets generating rate of BTN in Hangzhou. Next we simulate two typical states of network, the free state and the congested state. The results are showed in Fig.5 and Fig.6. At the free state, we can find that the result of computer simulation is similar with the results in the theoretical model. But because of the limitation of the biggest passenger flow of any section on one route and the biggest passenger flow of every bus on the section, the result at congested state in the computer simulations are different from the theoretical model. As we all know, the capacity of a bus is limited, so as subgraph (c) in Fig.6 shows, $SBB$ and $SBBF$ are limited by the biggest passenger flow of a section and the capacity of a bus. But due to the departure frequencies of group B are twice than group A, the parameter $SBB$ is also doubling.
Fig.5] The characteristics of traffic flow in BTN with the simulation model at free state. (a). the cumulative probability distribution of $NB$, note that the parameter $NB$ is normalized, namely is the ratio of $NB$ with the maximum $NB$. (b). the cumulative probability distribution of $SBBF$, note that the parameter $SBBF$ is normalized, namely is the ratio of $SBBF$ with the maximum $SBB$. (c). the sorting distribution of SBB and SBBF. Parameter SBB and SBBF are also normalized, which are shown in the blue and red line respectively. (d). the relation between $SBB$ and $SBBF$. Note that all parameters are normalized.
The characteristics of traffic flow in BTN with the simulation model at congested state. (a). the cumulative probability distribution of $NB$, note that the parameter $NB$ is normalized, namely is the ratio of $NB$ with the maximum $NB$. (b). the cumulative probability distribution of $SBBF$, note that the parameter $SBBF$ is normalized, namely is the ratio of $SBBF$ with the $SBB$. (c). the sorting distribution of SBB and $SBBF$, note that which are shown in the blue and red line respectively. (d). the relation between $SBB$ and $SBBF$. Note that all parameters are normalized.

Next, we continue to compare the distributions of theoretical model and simulation model at free state and at congested state, the results are showed in Figure 7. From the subgraphs (a) and (b), we can find that the theoretical model and the simulation model have approximate distributions. From the subgraphs (c) and (d) in Figure 7, we can find that the distributions of the two parameters which value the characteristics of traffic flow of the networks tend to be different, which means the state of the networks have important influence on the traffic flow.

Finally, we calculate the critical packets generating rate using three different strategies and make three comparisons among three different strategies, which are shown in table 2.

<table>
<thead>
<tr>
<th>strategy</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>CG</td>
<td>22.93</td>
<td>23.39</td>
<td>46.23</td>
</tr>
</tbody>
</table>

In Table 2, CG is the critical packets generating rate. Strategy 1 indicates that the departure frequency of all lines is the same. Strategy 2 indicates that the departure frequencies of the 10% lines which are randomly selected are doubling. Strategy 3 indicates that the departure frequencies of the 10% lines which have the biggest passenger flow are doubling. In this paper, we use Strategy 3 as our strategy to simulate.
We can find that when the departure frequencies of the 10% lines which have the biggest passenger flow are doubling, the critical packets generating rate doubles correspondingly. It indicates that buses with large passenger flow are more important, we should shorten the bus departure time appropriately to improve the BTN properties.

Fig.7| The comparison of the distributions of parameters obtained from the theoretical model and the simulation model at free state and congested state. (a). the comparison of $NB$ at free state. (b). the comparison of $SBBF$ at free state. (c). the comparison of $NB$ at congested state. (d). the comparison of $SBBF$ at congested state. Note that all parameters are normalized.

4 Conclusions

In our work, we propose a model for traffic flow on BTN, and simulate the characteristics of the traffic flow from two aspects: the theoretical and computer simulation. The simulation results tally well with our theoretical analysis irrespective of congestion. If congestion is considered, the situation is different. Therefore, we should take congestion into account when studying the BTN. In the model we introduce some parameters to measure the properties of traffic flow on BTN, such as $NB, EB, SBB$ and $SBBF$. These parameters help us to optimize the BTN even if we only know the matrix $POD$. For example, based on $SBBF$, we reduce the departure frequency of buses on less important routes and increase the departure frequency of buses on important routes to reasonably assign resources and improve the ability to deal with congestion. We believe that such insights will be helpful exploring the ways to optimize problems of traffic flow.
5 Acknowledgements

The work is supported by the National Natural Science Foundation of China under Grant No. 61374152, the commonweal application technique research project of Zhejiang Province under Grant No. 2012C23126.

References:


[17] Neri F., Quantitative Estimation of Market Sentiment: a discussion of two alternatives,

