Vol 17 No 8, August 2008
1674-1056/2008/17(08)/2874-07

Chinese Physics B

© 2008 Chin. Phys. Soc.
and IOP Publishing Ltd

An evolutionary model of urban bus transport network based on B-space*

Zhu Zhen-Tao(朱振涛)[a)b)†], Zhou Jing(周 晶)[a)],
Li Ping(李 平)[c)], and Chen Xing-Guang(陈星光)[a)]

[a)] *School of Management and Engineering, Nanjing University, Nanjing* 210093, *China*
[b)] *School of Economics and Management, Nanjing Institute of Technology, Nanjing* 211167, *China*
[c)] *Department of Basic Sciences, Nanjing Institute of Technology, Nanjing* 211167, *China*

(Received 11 November 2007; revised manuscript received 12 March 2008)

In this paper, an evolutionary model of bus transport network in B-space is developed. It includes the effect of the overlapping ratio of new route on network performance and overcomes the disadvantage, i.e. lack of economic consideration, in the evolutionary bus transport network model in P-space proposed by Chen *et al* (2007). The degree distribution functions are derived by using the mean-field method and the master equation method, separately. The relationship between the new stop ratio of a route, λ, and the error in exponential of degree distribution function from the mean-field method is developed as $\Delta\text{Slope}= \lambda/(1-\lambda)+\ln(1-\lambda)$. Finally, the bus transport networks of Hangzhou and Nanjing are simulated by using this model, and the results show that some characteristic index values of the simulated networks are closer to the empirical data than those from Chen's model.

Keywords: complex networks, Urban bus transport network, B-space, master equation method
PACC: 0590, 0175, 0250

1. Introduction

Ever since the last decade there has been a new movement of interest in the study of complex systems and complex networks with main focus on the analysis shifting from small networks to systems with thousands or millions of nodes, and with a renewed attention to the effect of dynamical structures of the systems on their functions.[1−30] The topological structure of urban public transport system, a complex system, is irregular in space and dynamically evolving with time. The analysis of topological properties of urban bus transport networks (BTN) is of significance in the research of urban transport systems. The pioneering papers[4,5] in 1998 and 1999 indicate the emergence of complex network theory, and attract many investigators from different fields to study the complex networks existing in their own fields. Since then, one has comprehensively investigated transport networks such as airport network, railway network and power grid, city road network and urban public transport network.[6−22] An evolutionary network[17] based on the evolution of the traffic flow has been proposed, discussing the cascaded failures in congested complex networks[18] and the topological characteristics of city public traffic networks[19−22] in some middle and large cities in China. Besides, several triangular patterns have been found to be important to BTN efficiency.[21]

Chen *et al*[16] have found that both the degree distributions and the distributions of 'the number of bus routes a stop joins' of BTNs for 4 Chinese cities in P-space are in the exponential function forms. On the other hand, the distributions of 'the numbers of stops in a bus route' follow unimodal functions, and have no effect on the above two types of distributions. Then they highlight that random attachment is a major growth mechanism of BTNs and introduce a random evolution model based on P-space, where a new bus route is built once a new site is added into the network. However, the constructors and administrators of BTNs often argue that routes do not follow the shortest path between their terminal stations, but rather meander through neighbourhoods and between sites of interest to efficiently cover the surface of the BTN network, namely, less bus routes meet more travel demands, in consideration of the constraints of

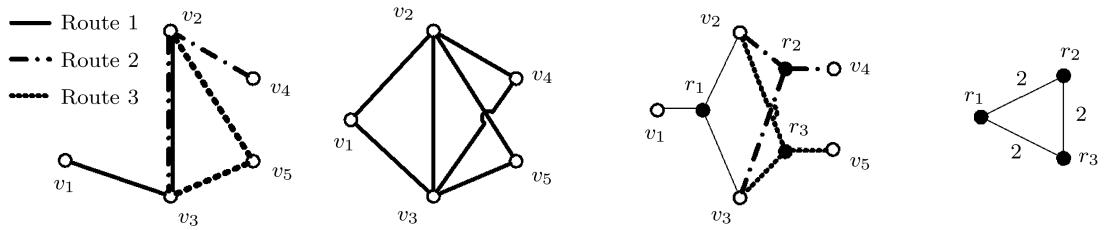
*Project supported by the National Natural Science Foundation of China (Grant No 70571033) and the State Key Development Program for Basic Research of China (Grant No 2006CB705500).
†E-mail: zztnit@yahoo.com.cn

land use, operation cost, and other factors. Hence, the assumption of Chen's model is obviously not consistent with these practical requirements, and may bring about considerable errors of the model in describing the number of routes and the average degree of nodes of actual BTNs in P-space (see n_{R} and $\langle k \rangle$ in Table 1), and therefore it should be modified.

2. Three types of mapping spaces in the analysis of BTN and their features

According to different meanings of nodes in BTNs, presentations of BTNs can be classified into three types: the stop network whose nodes represent bus stops, the bus route network whose nodes represent bus routes, and the mixing network whose nodes represent stops or bus routes. A BTN of three routes (r_1, r_2, and r_3) and its topological representations in P-, B-, and C-space are illustrated in Fig.1.

The P-space (Fig.1(b)) is also called 'space of transfer', in which all stops used by a route are fully interconnected and form a clique.[14] In the P-space, the node degree implies the number of stops that one can go to directly without any change, while the distance between two nodes equals the minimum number one has to change plus one. The P-space can better describe the accessibility and convenience of BTNs. Furthermore, it has a strong generalization due to ignoring the sequential order of the stops in a route.



(a) Three bus routes (b) P-space (c) B-space (d) C-space

Fig.1. Topologic illustration of A BTN of three routes in P-, B-, and C-space.

The bus route network, also called 'C-space' (Fig.1(d)), is a weighted network. In this space, two route nodes are connected when they share at least a common stop and the number of stops they share is defined as the weight of the link between the two routes. The node degree is defined as the number of routes that share at least one stop with this route, namely the number of routes on which the passengers can transfer. The node weight, defined as the sum of the weights of the links attached to the node, reflects the frequency that other routes appear during the journey of this route or the accumulative number of stops that this route shares with other routes, namely the number of changes.

It is remarkable that the mixing network, also called 'B-space' (Fig.1(c)), is much less studied than the above spaces. However, this space has its unique advantages that have been neglected so far. In the B-space, links exist only between a route node and a stop node. The degree of a route node implies 'the number of stops in a bus route', while the degree of a stop node, i.e. 'the number of bus routes that the stop joins', reflects the convenience extent of the stop when one travels to other bus stops. In fact, if one separates out these two types of nodes in the B-space and compute the degree distribution in each solo-type node group, one can directly obtain the distribution of 'the number of bus routes that a stop joins' and the distribution of 'the number of stops in a route'.[16] As the P-space and C-space can be obtained by projecting the B-space into the stop node set and route node set, respectively, the B-space has advantages and all information that the P-space and the C-space have. In this paper, we will derive the analytic distribution of 'the numbers of bus routes that a stop joins', $p(r)$, in the B-space. Here, r is the number of bus routes a stop serves while $p(r)$ is the probability that a randomly selected stop serves r bus routes.

3. Assumption and formulation of the model

As each new bus stop in the evolution of BTN is chosen for its own special reasons and the reasons are somewhat uncorrelated with each other, the random

attachment becomes a major tendency in a general sense. Therefore, the model developed in this paper is also a randomly evolving model, whose assumptions and relevant terminologies are as follows:

(1) The evolutionary BTN model is based on the B-space, and the asymmetric distribution of travel demands in the geographical space is omitted.

(2) Since simulation results[16] show that the distribution of 'the number of stops in a route', $p(\theta)$, has no influence on the distribution of 'the number of bus routes that a stop joins', $p(r)$, hence suppose that θ is a constant integer, which means that $p(\theta)$ takes a delta-function form. Therefore, the degree distribution to be discussed below refers to the distribution of stop node degree.

(3) A new bus route consists of b new stops and $\theta - b$ previously existing stops in the BTN. $R_i(t)$ is the number of routes that stop i joins at time t. The 'overlapping ratio of a route', κ, is defined as the ratio of the number of previously existing stops over the total stop number of the new route when the route is just built. The 'new stops ratio of a route', λ, is the ratio of new stop number to the total stop number of the route. So, $\kappa = 1 - \lambda$.

How to improve the economic efficiency of a BTN, with a high level of transfer convenience and the reliability of travel maintained, is a problem that deserves further investigating when one constructs new routes and new stops. Inspired by Ref.[16], we build a BTN evolutionary model taking into account the overlapping ratio of a new route to analyse its effects on the node degree distribution and network features of the randomly evolving BTN model.

Initially, there are m_0 stops in the model at $t = 0$, which are connected and form several cliques representing a number of routes. In each time step, a new route node and b stop nodes ($b = \lambda\theta$) are added first. Then $\theta - b$ old stop nodes are selected at random from the network. Subsequently these selected stop nodes and b new stop nodes are connected with the newly added route node respectively, forming a new route in B-space. The evolution of a BTN is illustrated in Fig.3. For example, there were two routes (γ_1,γ_2) in the original system. Each of them has 4 stops. Now, we add a new route node (γ_3) and two new stop nodes (n_7, n_8) to the system, randomly select 2 old stops in the system, and then connect these old and new stops with route node γ_3, respectively, to form a new route with $\lambda = 0.5$.

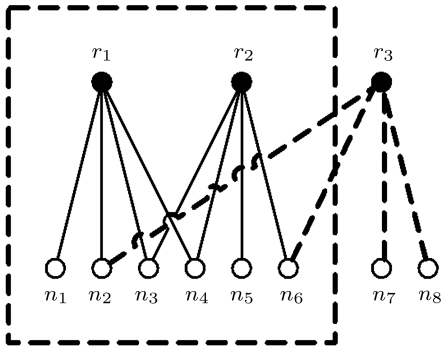


Fig.2. A schematic illustration for the evolution of a BTN.

According to the mean-field approach, we can write for stop node i

$$\frac{\partial R_i}{\partial t} = A\frac{1}{N(t)},$$

where, $A = \theta - b, N(t) = m_0 + bt$, $N(t)$ is the number of stop nodes in the system at time t, while R_i the degree of node i, i.e. the number of routes that stop i serves. Given the initial condition that stop node i is added to the system at time t_i with connectivity $R_i(t_i) = 1$, the solution of this equation is

$$R_i(t) = \frac{A}{b}\ln\left(\frac{m_0 + bt}{m_0 + bt_i}\right) + 1.$$

The probability that an arbitrary stop node has a connectivity $R_i(t)$ smaller than r, can be written as

$$\begin{aligned} &P(R_i(t) < r) \\ &= p\left(t_i > \frac{\mathrm{e}^{-\frac{b}{A}(r-1)}(m_0 + bt) - m_0}{b}\right). \end{aligned}$$

For the same number of nodes to be added to the system at equal time intervals in the mean-field approach, the probability density of t_i is

$$p_i(t_i) \approx \frac{1}{t}.$$

Then we obtain the degree distribution of stop node as follows:

$$\begin{aligned} p_A(r) &= \frac{\partial P(R_i(t) < r)}{\partial r} \\ &= \frac{b}{A}\mathrm{e}^{-(r-1)\frac{b}{A}} \\ &= \frac{\lambda}{1-\lambda}\mathrm{e}^{-\frac{\lambda}{1-\lambda}(r-1)}, \end{aligned}$$

where,

$$\lambda = b/\theta. \tag{1}$$

Since the mean-field method converts a discrete system into a continuous system and simplifies the probability density of t_i into a uniform distribution,

when multi-nodes are added to the evolutionary system at each time step, the entering time of the multi-nodes overlaps, which may cause considerable error in the degree distribution of stop node (Fig.3). For this reason, we will analyse this discrete system in the following by using the master equation approach. [29,30]

Assume that $p(r,s,t)$ represents the probability that the degree of stop node s is r at time t. The degree distribution of a stop node obeys the following master equation:

$$p(r,s,t+1) = \frac{A}{N(t)}p(r-1,s,t) + \left[1-\frac{A}{N(t)}\right]p(r,s,t). \quad (2)$$

The node degree distribution of the entire network is

$$p(r,t) = \frac{1}{N(t)}\sum_{s=1}^{N(t)} p(r,s,t).$$

Summing up Eq.(2) over s from 1 to $N(t)$, one obtains

$$N(t+1)\,p(r,t+1) - \sum_{s=N(t)+1}^{N(t+1)} p(r,s,t+1) = Ap(r-1,t) + [N(t)-A]p(r,t). \quad (3)$$

With the initial condition that stop nodes are born with one link to form a new bus route, $p(r,N(t-1)+i,t) = \delta_{r,1}$, $i = 1,\ldots,b, t > 0$, where $\delta_{r,1} = 1$ if $r = 1$, and $\delta_{r,1} = 0$ if $r \neq 1$, we obtain the following equation from Eq.(3):

$$N(t)[p(r,t+1) - p(r,t)] + bp(r,t+1) + Ap(r,t) - Ap(r-1,t) = b\delta_{r,1}.$$

If the steady state can be reached,[29] one obtains $(A+b)\,p(r) - Ap(r-1) = b\delta_{r,1}$. Then we obtain the stationary node degree distribution of the network,

$$p_B(r) = \frac{b}{A+b}\left(\frac{A}{A+b}\right)^{r-1} = \frac{b}{\theta}\left(\frac{\theta-b}{\theta}\right)^{r-1} = \lambda\left(\frac{1}{1-\lambda}\right)^{-(r-1)}. \quad (4)$$

We simulate several BTNs of 10000 nodes, and the obtained results are plotted in Fig.3. In the line-log scale, it is striking that fitting curves of simulated data by using $f(x) = ae^{-bx}$ overlap completely with $p_B(r)$ derived in the master equation approach, which fully demonstrates the effectiveness of master equation approach.

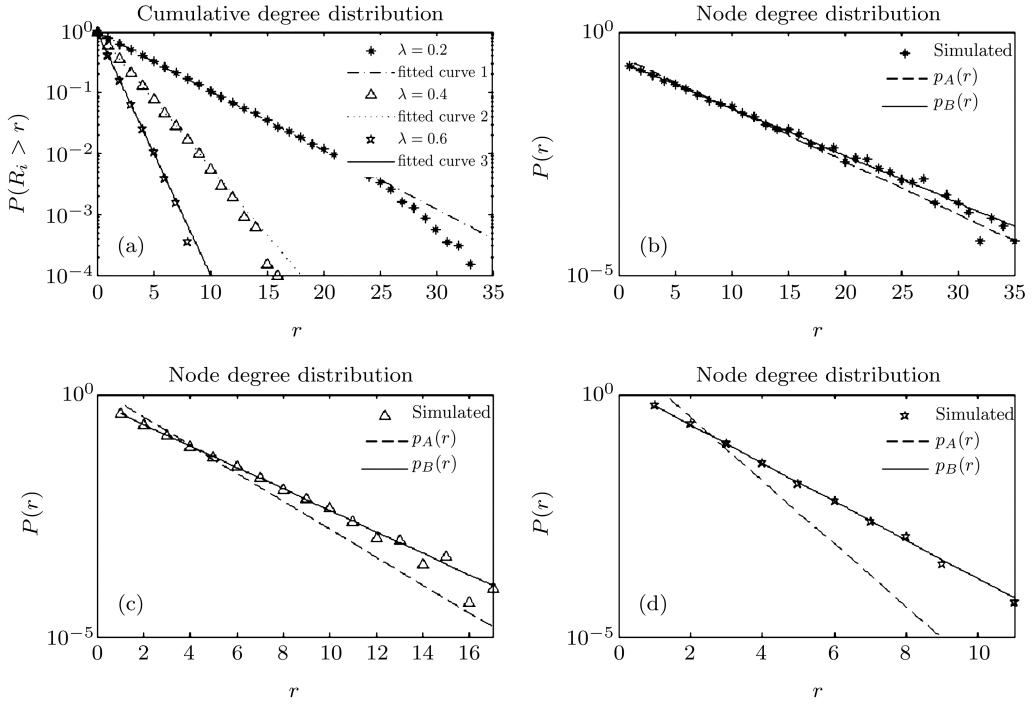


Fig.3. (a) The cumulative degree distribution and (b, c and d) degree distributions ($\lambda = 0.2, 0.4$ and 0.6) of simulated BTNs ($n = 10000$) calculated by using the mean field method ($p_A(r)$) and the master equation approach ($p_B(r)$), respectively.

To investigate the relationship between new stop ratio of a route, λ, and the error in degree distribution function caused by using the mean-field method, we compare the exponentials of $p_A(r)$ and $p_B(r)$, that is, the slopes of these two functions in the semilog $(\ln p(r) - r)$ coordinate system as follows:

$$\frac{\text{Slope}(\ln p_A(r))}{\text{Slope}(\ln p_B(r))} = \frac{b/A}{\ln(\theta/A)} = \frac{b/A}{\ln(1+b/A)},$$

Only when $b/A \to 0$, namely $b \ll \theta$, for instance, $b = 1 \ll \theta$, will these two slopes be nearly identical.

The difference between these two slopes is

$$\begin{aligned} \Delta\text{Slope} &= \text{Slope}(\ln p_B(r)) - \text{Slope}(\ln p_A(r)) \\ &= \frac{\lambda}{1-\lambda} + \ln(1-\lambda), \qquad (5) \\ \frac{\partial(\Delta\text{Slope})}{\partial\lambda} &= \frac{\lambda}{(1-\lambda)^2} > 0, \end{aligned}$$

where, $\lambda \in (0,1)$.

The difference between these two slopes becomes larger as λ increases. It can be learnt from Fig.4 that the error will dramatically increase when $\lambda > 0.5$. This means that the ratio of stops for small degrees in the BTN is overestimated while the ratio of stops for large degrees is underestimated by using the mean-field method.

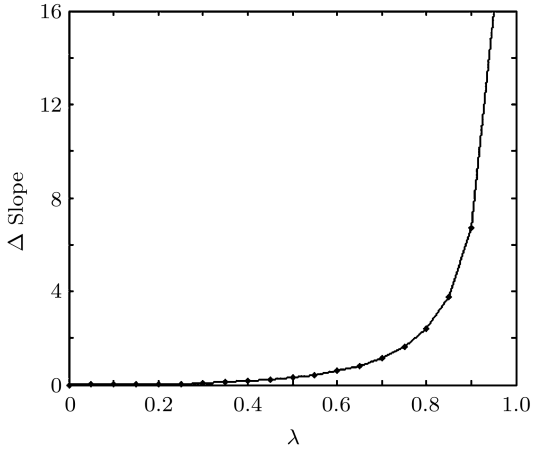


Fig.4. Difference in exponential between distribution functions as a function of λ.

4. Simulation analysis

We simulate the BTNs in Hangzhou and Nanjing by using the Chen model (denoted as Simul) and our model (denoted as Simu2) and compare some major characteristic index values of simulated BTNs by using the empirical data, in which Data refers to the empirical data,[16] with the results listed in Table 1. We define the 'average service efficiency of the network' as $\eta = \dfrac{n_v}{n_R \cdot \langle\theta\rangle}$, where n_v, n_R, and $\langle\theta\rangle$ are the total number of stops in the network, the number of routes and the mean number of stops per route respectively. $\eta \in (0,1)$. Given a certain number (n_v) of bus stops for a network, the larger the value of η, the less the bus routes that the bus company should maintain to satisfy the travel demands (i.e. cutting down the operation cost, thus improving the economic efficiency of the bus network). We also choose global efficiency E, clustering coefficient C, average path length APL and diameter D as major network indexes. These indexes are computed in the P-space since these indexes lack clear meaning in the B-space. Among these indices, the global efficiency E is defined as $E = \dfrac{1}{n_v(n_v-1)} \sum\limits_{\substack{i,j\in G \\ i\neq j}} \dfrac{1}{d_{ij}}$, which is a measure of how well the nodes communicate over the network,[7] where, d_{ij} is the length of the shortest path between nodes i and j, i.e. the minimum number of edges covered in order to go from node i to node j. Clustering coefficient of node i, C_i, is defined as a probability that two randomly chosen neighbours of node i possess a common link. If node i has k_i neighbours, there will be at most $k_i(k_i-1)/2$ edges between all pairs of these neighbours. C_i is proportional to the fraction of these edges that really exist, and it measures the local group cohesiveness of node i. C is the average of C_i over all nodes: $C = \dfrac{1}{n_v} \sum\limits_{i,j\in G} C_i$, and used to describe the probability that two randomly chosen neighbours of a stop can reach each other without any change. APL, the average shortest path length, is defined as $\text{APL} = \dfrac{1}{n_v(n_v-1)} \sum\limits_{\substack{i,j\in G \\ i\neq j}} d_{ij}$,[4] and it equals one plus the average number of times one has to change in order to go from one stop to another stop. Diameter D, the maximum of shortest path lengths between arbitrary pair of nodes of the network, equals one plus the maximal number of times one has to change to reach any other stop. The convenience of travel by bus in an urban city can be well described by APL and D.

We compare the simulation data from the Chen's model (Simul) and our model (Simu2) with the empirical data of BTNs in Hangzhou and Nanjing cited from Ref.[16]. Since the mean values of the number of bus stops per route, $\langle\theta\rangle$, in Hangzhou and Nanjing are 15.66 and 21.15, respectively, we let θ have fixed values of 16 and 21 in simulating BTNs for Hangzhou

and Nanjing respectively. We round off $\frac{n_v - \theta_0}{n_R - 1}$ to the nearest integer, b, the number of stop nodes added at each time step. Here, θ_0 is the number of the bus stops in the first bus route. To simplify the condition, we let $\theta_0 = \theta$. Table 1 lists the simulation results from the 50 runs of simulation computations and the relevant empirical data.

Table 1. Comparison of simulated and empirical data of BTNs for Hangzhou and Nanjing.

	Data†	Hangzhou		Data†	Nanjing	
		Simu1	Simu2		Simu1	Simu2
n_R	150	812	136	252	1744	250
n_v	827	827	826	1764	1764	1764
D	5	3	4	7	3	4
C	0.78	0.45	0.62	0.76	0.41	0.58
APL	2.6	1.83	2.33	2.74	1.85	2.34
$\langle k\rangle^*$	44.46	235.64	39.52	67.03	415.24	59.52
$\langle\theta\rangle$	15.66	16	16	21.15	21	21
λ	0.3477	0.0625	0.375	0.3283	0.0476	0.3333
η	0.3521	0.0637	0.3796	0.3310	0.0482	0.3360

*$\langle k\rangle$ is the average degree of nodes in the P-space.

†Data refers to the empirical data.

It is obvious that the simulation results of our model are closer to empirical data than those from the Chen's model for D, C or average change times between arbitrary pair of stops. For example, the relative errors of a major index, APL, of BTNs in Hangzhou and Nanjing reduce by 19.23% and 17.88%, respectively. In addition, we find that BTNs in these two cities have similar overlap ratios of a route ($\kappa = 1 - \lambda$; $\lambda = 0.3477$, and 0.3283, respectively).

To further investigate the relationship between the overlapping ratio of a route and BTN's features, we simulate 9 different BTNs, separately, with $\lambda = 0.1, \ldots, 0.9$. All of these networks each have a scale of 10000 stops and all the routes each have 20 stops. The averages of the BTN index values over the 50 runs of simulation computations are illustrated in Figs.5 and 6. Figure 5 shows that as the route overlapping ratio decreases (new stop ratio λ increases accordingly) the clustering coefficient and average service efficiency of the networks increase but the global efficiency declines. It is obvious that if a route contains more new stops, then the average service efficiency will be improved through cutting the operation cost. However, the meaning of the increase in clustering coefficient is complicated. In the P-space stops in a route each are connected with each other. This clique structure will have a considerable influence on the clustering coefficient when the route overlapping ratio decreases. This may affect the ability that clustering coefficient has to

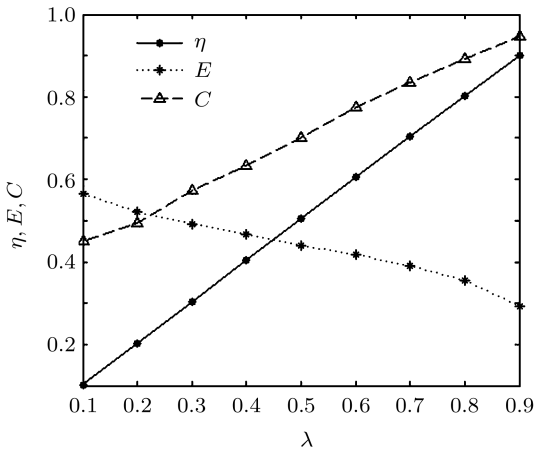


Fig.5. Network indexes (η, E, C) versus λ.

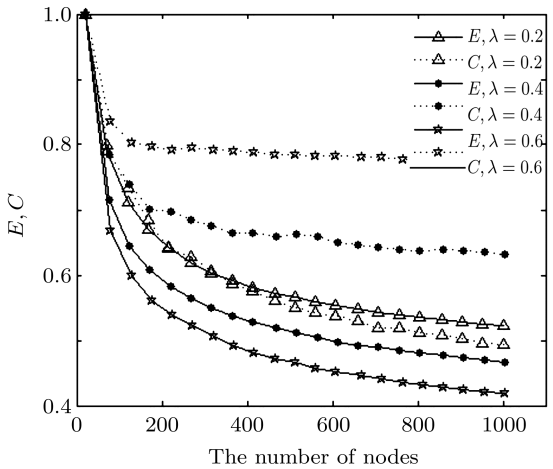


Fig.6. The evolution of network indexes (E, C) with node increasing.

describe the connectedness and changeability of the stops. Nevertheless, the decline of global efficiency and the increase of change times with the decline of the route overlapping ratio clearly imply that the accessibility, mobility, and convenience of the journey between arbitrary pair of stops will seriously deteriorate. For the purpose of optimizing these indexes, one should select $\lambda \in [0.3, 0.6]$. So, from this viewpoint, the structures of bus route systems in Hangzhou and Nanjing are both in an adequate condition in consideration of the current scales of the cities.

5. Conclusions

In this paper, we have proposed an evolutionary BTN model based on the B-space which is applicable to investigating the influence of the route overlapping ratio on the BTN performance, and we have accurately derived the degree distribution function in the B-space. We have found that the error in the exponent of degree distribution function, derived by using the mean-field method, will become considerably large with the overlapping ratio decreasing during the network evolution. This is due to the fact that the overlapping of entering times of multi-nodes deviates from the continuum assumption of the mean-field method. Finally, we have obtained the simulation results of the BTNs for Hangzhou and Nanjing, which are much close to the empirical data compared with those from the Chen Yong-Zhou's model. This indicates that our model can serve as a generic platform for the further studying of dynamic characteristics of evolutionary BTN such as dynamic robustness and spreading process in the B-space or in the P-space.

It should be pointed out that the B-space opens a novel understanding of public transport networks in the perspective of effective route construction. As one can straightly obtain two types of important distributions, i.e. the distributions of 'the number of bus routes that a stop joins' and 'the number of stops in a bus route', in the B-space, this unique feature of BTN makes it possible to discuss the relationship between the construction of routes under economic consideration and the performance of the network. The more complex construction of routes in the evolution of BTNs should be an interesting topic and worth investigating. In addition, the degree distribution function and other analyses obtained in this paper are extendable to similar networks such as collaboration networks, which can also be represented as bipartite graphs with two types of nodes: one type presents the actors, and the other denotes the acts of collaboration. For instance, in a large organization every project is accomplished with a fixed ratio between the numbers of freshmen and randomly selected veterans. The distribution of the numbers of projects that an actor has participated is just $p(r)$ that we have obtained.

References

[1] Arthur W B 1999 *Science* **284** 107
[2] Pearce N and Merletti F 2006 *Int. J. Epidemiol.* **35** 515
[3] Feng G L and He W P 2007 *Chin. Phys.* **16** 2825
[4] Watts D J and Strogatz S H 1998 *Nature* **393** 440
[5] Barabasi A L and Albert R 1999 *Science* **286** 509
[6] Albert R and Barabasi A L 2002 *Rev. Mod. Phys.* **74** 47
[7] Latora V and Marchiori M 2001 *Phys. Rev. lett.* **87** 198701
[8] Guimera R, Mossa S, Turtschi A and Amaral L A N 2005 *Proc. Natl. Acad. Sci.* U.S.A. **102** 7794
[9] Kurant M and Thiran P 2006 *Phys. Rev. Lett.* **96** 138701
[10] Crucitti P, Latora V and Marchiori M 2004 *Physica* A **338** 92
[11] Rosvall M, Trusina A, Minnhagen P and Sneppen K 2005 *Phys. Rev. Lett.* **94** 028701
[12] Kalapala V, Sanwalani V, Clauset A and Moore C 2006 *Phys. Rev.* E **73** 026130
[13] Sienkiewicz J and Holyst J A 2005 *Phys. Rev.* E **72** 046127
[14] Kurant M and Thiran P 2006 *Phys. Rev.* E **74** 036114
[15] Ferber C von, Holovatch T, Holovatch Yu and Palchykov V 2007 *Physica* A **380** 585
[16] Chen Y Z, Li N and He D R 2007 *Physica* A **376** 747
[17] Gao Z Y and Li K P 2005 *Chin. Phys. Lett.* **22** 2711
[18] Zheng J F, Gao Z Y and Zhao X M 2007 *Physica* A doi:101016/jphysa200707031
[19] Lu H P and Shi Y 2007 *Tsinghua Sci. & Technol.* **12** 204
[20] Zhao J S, Di Z R and Wang D H 2005 *Complex Syst. & Complexity* **2** 45 (in Chinese)
[21] Li P, Xiong X, Qiao Z L, Yuan G Q, Sun X and Wang B H 2006 *Chin. Phys. Lett.* **23** 3384
[22] Zhu C P, Xiong S J, Tian Y J, Li N and Jiang K S 2004 *Phys. Rev. Lett.* **92** 218702
[23] Wu Y, Xiao J H, Wu Z Y, Yang J Z and Ma B J 2007 *Acta Phys. Sin.* **56** 2037(in Chinese)
[24] Li K P and Gao Z Y 2007 *Chin. Phys.* **16** 2304
[25] Wu J J, Gao Z Y and Sun H J 2007 *Chin. Phys.* **16** 47
[26] Zhang P P, He Y, Zhou T, Su B B, Chang H, Zhou Y P, Wang B H and He D R 2006 *Acta Phys. Sin.* **55** 60 (in Chinese)
[27] Li J, Wang B H, Jiang P Q, Zhou T and Wang W X 2006 *Acta Phys. Sin.* **55** 4051 (in Chinese)
[28] Liu H K and Zhou T 2007 *Chin. Phys.* **56** 0106
[29] Dorogovtsev S N and Mendes J F F and Samukhin A N 2000 *Phys. Rev. Lett.* **85** 4633
[30] Dorogovtsev S N and Mendes J F F and Samukhin A N 2001 *Phys. Rev.* E **63** 062101