# Gravity Model for Transportation Network Based on Optimal Expected Traffic

Jiang-Hai Qian and Ding-Ding Han*

School of Information Science and Technology, East China Normal University, Shanghai 200241, China
ddhan@ee.ecnu.edu.cn

**Abstract.** We propose a spatial network model for transportation system based on the optimal expected traffic. The expected traffic represents the prediction of the flow created by two vertices and is calculated by the improved gravity equation $w_{ij} = K\frac{M_i^\alpha M_j^\alpha}{D_{ij}^\gamma}$. The model maximizes the total expected traffic of the network. By changing the two parameters $\alpha$ and $\gamma$ which controls the fitness and the geographical constraints, the model can vary its topology from the star-like network to the decentralized road-like network. The simulation for the Chinese city airline network reproduced many properties of the real network. In the end of this paper the relationship of the expected traffic and the real traffic is discussed.

**Keywords:** spatial network, expected traffic, gravity.

## 1 Introduction

Since the initial studies on the small-world phenomenon by Watts and Strogatz [1] and the scale-free property by Barabasi and Albert [2], lots of achievements on complex network have been gotten. And our research group have studied some works [3]. Most previous works focus on the topological properties of the network. However many networks are those embedded in the real space whose nodes occupy a precise position in Euclidean space and whose links are constrained by the geographic distance. The typical examples are the transportation systems ranging from river [4] to airport [5,6,7], street [8], railway and subway [9]. To model these spatial networks, geographical ingredient is demonstrated to play an important role on the network's topology. In the previous studies the large cost to establish long-distance link is considered to be the main reason that causes the nodes to connect to their geographical neighbors [10,11,12,13]. In addition to the spatial preference, the topology preferential attachment, namely nodes with larger degree have larger probability to be linked, is also important in the formation of the complex network [14,15,16]. Such mechanism can form hubs, the well-connected nodes, which usually reduce the diameter of the network.

* Corresponding author.

J. Zhou (Ed.): Complex 2009, Part I, LNICST 4, pp. 514–524, 2009.

Another interesting view of modeling the spatial network is to consider the intrinsic attributes of nodes [18,17]. The intrinsic attributes represent the fitness of nodes to win edges and are interpreted as, for example, capacity, social skills, activity levels, information contents and population of cities, etc. Based on this idea, Naoki Masuda and Hiroyoshi Miwa proposed a non-growing geographical threshold model which generalizes a variety of models such as the Boolean model and the gravity model [19]. R.Xulvi-Brunet and I.M.Sokolov interpret the nodes' intrinsic attributes as their different interaction range [20]. The nodes with large interaction range have larger probability to cover more nodes and gain more links. The novel idea combines the intrinsic attribute and the geographical influence naturally.

These works provide some guidelines in modeling spatial network. However they all concentrate on the cost of constructing networks while the effect of traffic on network's design is paid less attention. Whereas traffic may be an even more important factor because it represents the efficiency of the network. If the traffic between nodes can be predicted in some way, it is likely to construct the network efficiently. Inspired by this idea and its significance, we propose a simple spatial network model. The model is to maximize the whole expected traffic of the network, indicating the highest efficiency that the network may gain. The expected traffic is measured by the gravity equation. In the end of this paper, the relationship between the expected traffic and real traffic is discussed.

## 2 Expected Traffic and Gravity

Traffic in the real-world network is demonstrated to be strongly correlated to its topology. The empirical evidence coming from the studies on metabolic and airline network has shown that the traffic between nodes has the following form [21,22]:

$$w_{ij} \sim x_{ij}(k_i k_j)^{\theta} \ . \tag{1}$$

where $x_{ij}$ is a random number and $\theta$ is a positive exponent. $k_i, k_j$ are the degree of node $i$ and $j$ respectively and $w_{ij}$ is the weight or traffic between them. This result indicates the traffic can be measured after the topology has been known. However if the aim is to predict the traffic before the network is constructed, it seems useless.

Motivated by the studies on the intrinsic attributes of nodes in modeling complex networks [18,17], we consider the effect of node's fitness on the traffic. It is believed and demonstrated that nodes with better fitness usually gain more links. For example, in airline network, cities with large population are usually hubs. To satisfy this basic fact that node's degree usually has a positive correlation with its fitness, here the correlation of fitness and degree is simply assumed to be following the form:

$$k \sim M^{\beta} \ . \tag{2}$$

where $M$ and $k$ are respectively the fitness and the degree of node while $\beta$ is a positive exponent. According to Eq(1) and Eq(2) the relationship of traffic and the fitness of nodes is easily to be obtained:$w_{ij} \sim x_{ij}(M_i M_j)^\alpha$, where $M_i, M_j$ represent the fitness of node $i$ and $j$ and exponent $\alpha = \beta\theta$. On the other hand, for spatial network the traffic decreases with the geographical distance since long distance journey spends more time and cost. So the traffic is assumed to follow $w_{ij} \sim \frac{1}{D_{ij}^\gamma}$, where $D_{ij}$ is the geographical distance between node $i$ and $j$. In this paper, it is defined as the Euclidean distance. Thus the traffic between two nodes can be described by a gravity equation:

$$w_{ij} = K \frac{M_i^\alpha M_j^\alpha}{D_{ij}^\gamma} \ . \tag{3}$$

where $K$ is a constant coefficient, $\alpha, \gamma$ are two tunable parameters which determine the impact of fitness and geographical distance on the traffic.

Eq(3) has been confirmed by a very recent empirical study [23] and the gravity model is considered as a suitable form in describing interaction of particles in geographical space when the physical gravity or similar mass interaction is active [19]. The gravity model provides us a way to predict the traffic of network, because once the nodes are sited and the fitness is assigned, the traffic of any pair of nodes can be calculated even when no link exists. For this reason, we call $w_{ij}$ described by Eq(3) the expected traffic. However we emphasize that Eq(3) only describes the flow created by vertices i and j, that is, the traffic which origins from i(or j) and ends at the other. It does not include the traffic created by other pair of nodes but travels through link (i,j). Thus the expected traffic could be different from the real weight. The relationship of the expected traffic and the real traffic will be discussed in the section 5.

## 3 Gravity Model for Transportation Network

When people prepare to construct a network, what do they care more? Previous studies concentrate on the cost and expenses. However we argue the traffic that the network can carry is even more important. It is because not only high efficiency will bring much benefit but also an inefficient network will cause even more additional expense or loss. Inefficiency of the transportation network will cause inestimable loss since the infrastructure plays an extremely important role on the development of a country. Now consider $n$ nodes, every two nodes have their demand for some information exchange. What we care is that which of these demands are the most exigent, or in other words, which of the expected traffic among these nodes is the largest. Such information can be obtained in advance by Eq(3). Thus the network can be constructed efficiently by preferentially investing those node pairs with large expected traffic. If there is no other restriction, a fully connected graph is obtained. But the real-world networks are usually sparse because a fully connected network requires too large cost. Thus the number of links is limited. In the present paper the number of edges is defined as the budget of constructing the network. One may argue that the budget should rely more on the spatial distance because the cost of different links might be different and for airline networks very long links might be infeasible. We argue here that this point does not conflict with our model because the spatial ingredient in our work have been considered in the expected traffic(see Eq(3)) and the simplification of the budget only indicates that compared with the expected traffic the cost is a minor ingredient. For a pair of distant nodes, their $w_{ij}$ is not large enough to be connected under the fixed budget. However if the fitness is large enough to eliminate the effect of long geographical distance, the two nodes still have chance to link with each other. This can help us understand the phenomenon observed in the airline networks where small airports usually connect to the nearby hubs while the large airports can connect with each other despite of their long spatial distances.

Now suppose there are $n$ nodes distributed on a two-dimension plane. The fitness and coordinates of each node are known. By Eq(3), the expected traffic $w_{ij}$ of any two nodes can be calculated. We connected preferentially those node pairs with larger $w_{ij}$ and complete such process when the links come to the value we preset. Such process can be described as an optimal model:

$$Max \quad W_{exp} = \sum_{i<j} w_{ij}\eta_{ij} = \sum_{i<j} K \frac{M_i^{\alpha} M_j^{\alpha}}{D_{ij}^{\gamma}} \eta_{ij} \ . \tag{4}$$

$$s.t. \quad \sum_{i<j} \eta_{ij} = \epsilon \ . \tag{5}$$

where $\eta_{ij}$ is the adjacency matrix element of the network. $\epsilon$ is the number of edges we prescribe and $W_{exp}$ is the whole expectant traffic of network. However, the above-mentioned process may cause some isolated nodes. Two more restrictions are introduced to ensure each node is connected:

$$\sum_i \eta_{ij} \geq 1 (j = 1, 2, 3, ..., n) \ . \sum_j \eta_{ij} \geq 1 (i = 1, 2, 3, ..., n) \ . \tag{6}$$

Following this method, a network of thirty nodes is simulated. Set the budget $\epsilon$=39 and the coefficient $K = 1$ (actually $K$ makes no difference to the model). By varying the value of $\alpha$, $\gamma$, we got four networks with different topology as is seen in Fig. 1.

As to Fig. 1(a), the value of $\gamma = 0$ makes the network only rely on the fitness of nodes, which causes the topology to be dominated by two hubs since the nodes of good fitness is easy to magnetize others. Whereas with $\gamma$ increasing, more effect of the geographic factor makes the node tend to connect to the closer ones and weaken the hub-and-spoke effect. When $\alpha$=1,$\gamma$=2(Fig. 1(c)),the network exhibits some features similar to the airline network. When $\alpha$=0(Fig. 1(d)), the topology is entirely constrained by the geography, which forms a two-dimensional network strongly reminiscent of roads.

Our model has simple realistic and physical significance. Since high efficiency can bring high profit, from operator's view, the significance of our model is that it provides a possible way for the operators to gain the highest profit. Moreover the idea of the expected traffic indicates that link between nodes may

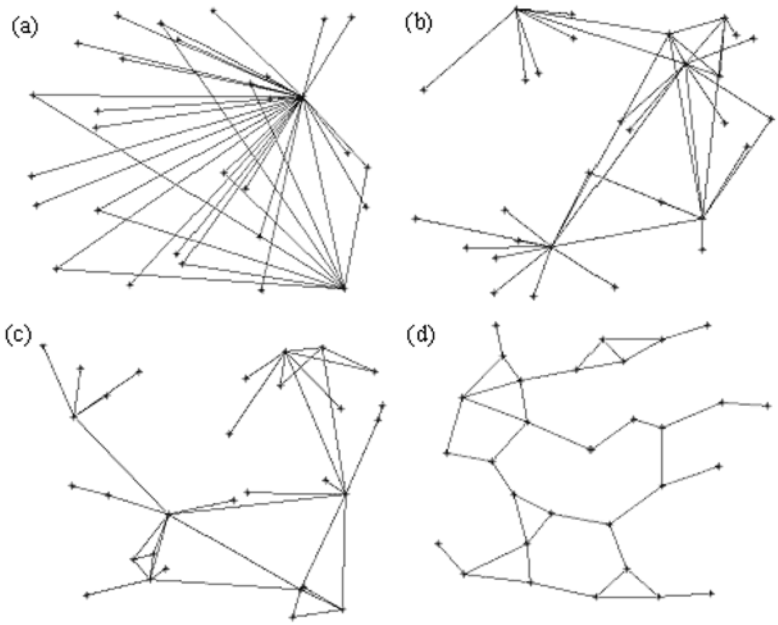


**Fig. 1.** Four networks with different topology are shown which is controlled by the parameters $\alpha$,$\gamma$. (a)$\alpha$ takes any value,$\gamma$=0. The network is dominated by two large hubs when $\gamma$=0;(b)$\alpha = 1$,$\gamma = 1$. With $\gamma$ increasing, the hubs become much smaller than figure (a); (c)$\alpha$=1,$\gamma$=2;(d)$\alpha$=0,$\gamma$takes any value. The topologies are strongly reminiscent of airlines and roads respectively.

depend on their potential dynamical strength. If there is no traffic or information exchange demand between two nodes, the link is unwanted even though the cost establishing the link is small. In a circuit, for example, a lead equals to disconnection if its current is zero. On the other hand if the potential traffic is large enough, link will be constructed even though it costs much because large traffic can bring high profit.

## 4 Simulation for the Chinese City Airline Network

To take the idea of maximizing expected traffic into application, we use our gravity model to simulate the Chinese airline network and make comparison with the real data [24]. In the following simulation, the nodes represent cities and the edges represent the airlines. We set the number of nodes $n = 121$ and number of links $\epsilon = 689$. The fitness $M$ and the distance $D_{ij}$ are respectively defined as the population of the city and the Euclidean distance of city $i$ and $j$. Selecting $\alpha$=1,$\gamma \in$[1,2] [25], they are to describe the interactions of cities. Here we set $\gamma = 1.5$.

Fig. 2 shows the simulated network. Obviously, the hubs in the real network such as Beijing, Shanghai, Guangzhou, Harbin, Urumchi exhibit the similar

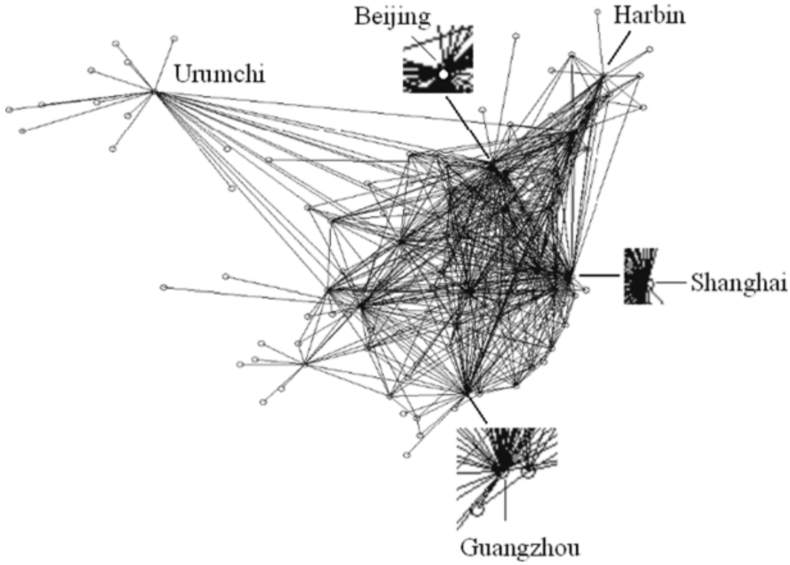


**Fig. 2.** The simulation for the Chinese airline network. Each link of two nodes reflects an airline between them. There is totally 689 links. The cities, such as Beijing, Shanghai, Guangzhou, Harbin, Urumchi, exhibit hub-and-spoke phenomenon in our simulation just as they do in the real network.

hub-and-spoke phenomenon in our simulation. In real condition, Beijing, Shanghai and Guangzhou are the three cities with the highest degree while in our model they are respectively Shanghai, Beijing, Wuhan (Guangzhou is the fourth). The reason for this difference may be that using the population to denote the fitness of city is intuitive but not exact because the economy and the administration factors are also important indexes for the grade of city. In spite of this difference, we still succeed in reproducing the every hub and their hub-and-spoke phenomenon existing in the real network.

We calculate the average shortest-path length $L$[1] and the Pearson correlation coefficient $r$[26] of the model network. The average shortest-path length of a network with $N$ nodes is the average number of edges that has to be crossed on the shortest path from any one node to another. The Pearson correlation coefficient describe that a network is assortative or disassortative. In our simulation $L = 2.302$, $r = -0.401$ while in the real network $L = 2.263$, $r = -0.408$. Fig. 3 shows the clustering-degree distribution. The clustering-degree distribution is the correlation of the degree $k$ and the average clustering coefficient of all nodes with degree $k$. It meets $C(k) \sim k^{-1}$ which indicates the model network exhibits the same hierarchy [27] as the real network does.

Fig. 4 is the degree distribution of the model network.It satisfies the two-regime power-law distribution. It satisfies the two-regime power-law distribution with the exponent $\gamma_1 = -0.46$ for the first power laws and $\gamma_2 = -2.3$ for the

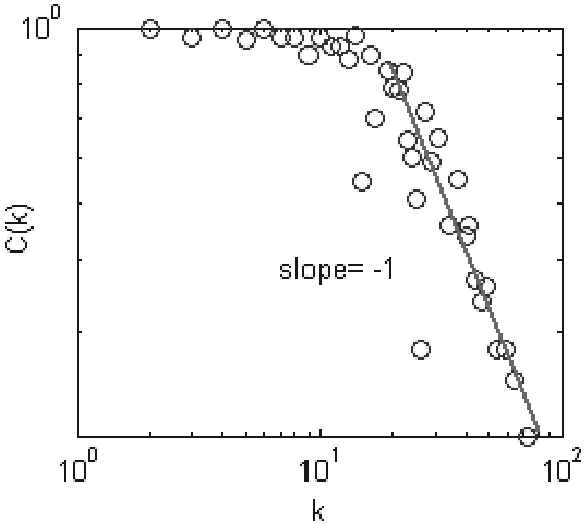


**Fig. 3.** The clustering-degree distribution of the simulation. The distribution satisfies a linear decreasing feature with slope -1 in log-log coordinate, which indicates that the result of simulation reproduces the hierarchy.

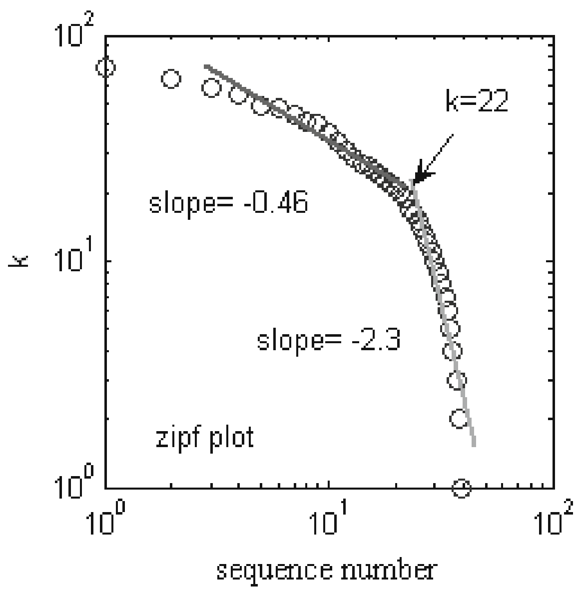


**Fig. 4.** The degree distribution of the simulation. The rank of the node is the sequence according to its degree. The node with the maximum degree has the rank number one and the second has rank number two, the rest may be deduced by analogy. The result presents a behavior of two-regime power-law degree distribution with the exponent$\gamma_1$=-0.46 for the first power laws and $\gamma_2$=-2.3 for the second. The turning point happens at degree k=22. In comparison, $\gamma_1$=-0.53,$\gamma_2$=-2.05 and the turning point at degree k=20 in the Chinese City Airline network.

second. The turning point happens at degree $k = 22$. Both the exponents and the turning point fit the real network well.

## 5 Expected Traffic and Real Traffic

Since the expected traffic plays an important role in our network design, it is meaningful to study its relationship with the real traffic. As is defined in section 2, expected traffic $w_{ij}$, calculated by Eq(3), is a good prediction of the traffic produced by the corresponding nodes. It does well in describing the direct interaction of two nodes. However in the transportation networks a so-called transfer mechanism may causes the difference between expected traffic and real weight. Transfer mechanism widely exists in the technical network. Most commonly the basic role of the routers in Internet is to transfer the packet from the original to the destination. By the transfer mechanism link carries not only the traffic produced by the directly connected nodes but also the traffic of other pairs that travel through it. Thus the real traffic of a link is the sum of the expected traffic of the directly connected nodes and other additional transferred traffic. To make clear this phenomenon the process is visualized in Fig. 5.

Fig. 5 shows that the expected can be different from the real one in a network with transfer mechanism, but it doesn't mean our model based on the prediction of the traffic is uncorrect. Because when a link is to be established, what we care more is the direct interaction between two nodes. As long as the expected traffic is large enough, it is necessary and profitable to construct the link. Besides it is well-known that a self-organized system usually evolves by the local information,

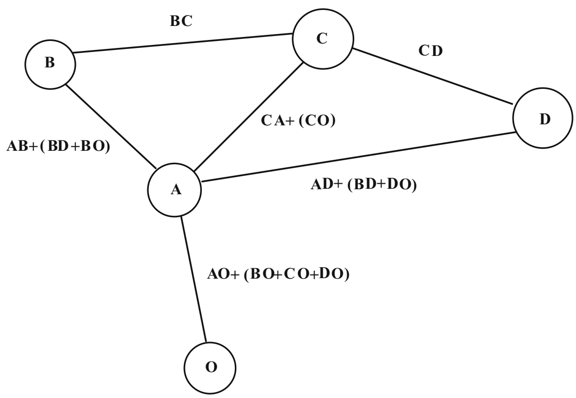


**Fig. 5.** Relationship of the expectant traffic and the real traffic. As an example, we only focus on the link(A,O). AO,BO,CO,DO in the figure represents the expectant traffic between the node O and the others respectively. However since there is no direct link to node B,C,D, the traffic to these nodes have to be distributed to the link (A,O). In the figure, traffic in the bracket are the transferred ones and out of the bracket are the expectant traffic of the directly connected nodes. The sum of the both represents the real traffic.

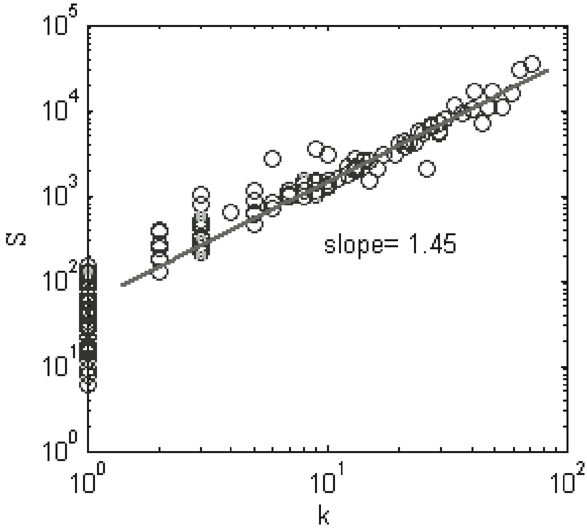


**Fig. 6.** The correlation of the degree and node strength. The correlation follows $S(k) \sim k^{1.45}$. Compared with $S(k) \sim k^{1.37}$ for real network, our simulation was succeeded in reproducing the non-linearly correlation of degree and node strength.

but the traffic distribution caused by the transfer mechanism depends on the global topology information. So it is difficult and impossible to consider the corresponding effect in the self-organized network design.

The transfer mechanism indicates that although the expected traffic can be predicted in some way and play a role on network construction, once the topology is determined, the real traffic relies greatly on the topology information. And when a new-born node is connected to the network, it affects the weight not only on the nodes it connects directly but also others. In other words traffic within many old links may be modified. Similar idea has been proposed by BBV model [28] which considers the weight update process happening only among the neighbors of the node that the new one connects to. However gravity model provides us a better understanding to this process and the real traffic within each link can be calculated.

First we suppose that the traffic transmitting from the original to the destination always go through the shortest path. If there is more than one such choice, the actual path is randomly chosen among them. Now we define matrix $A_N^{mn}$ to describe one of the shortest path from node $m$ to $n$. $A_N^{mn}$ is a matrix with $N \times N$ elements, where $N$ is the number of the nodes. The element of $A_N^{mn}$, denoted by $a_{ij}^{mn}$, takes one if the shortest path from node $m$ to $n$ goes through the link$(i, j)$, otherwise it takes zero. Then the real traffic of each link can be written as:

$$\Omega_{real} = \sum_{(m,n)} w_{mn} A_N^{mn} \ . \tag{7}$$

where $w_{mn}$ is the expected traffic between node $m$ and $n$. $\Omega_{real}$ is a matrix with $N \times N$ elements and its element denoted by $\omega_{ij}$ represents the real traffic within link$(i, j)$. The node strength reads as

$$s_i = \sum_i \omega_{ij} \ . \tag{8}$$

Following the above method, we calculate the strength-degree correlation of the simulated Chinese airline network(see Fig. 6). As is shown in Fig. 6, the node strength increases with the degree, but does quicker than linearly, as 1.45 power, namely satisfies $S(k) \sim k^{1.45}$ while in the real airline network the correlation follows $S(k) \sim k^{1.37}$.

## 6 Conclusion

In contrast to the previous studied spatial graph, we consider the traffic is the most important factor in network design and the traffic can be predicted by the gravity equation. Based on this idea a simple model for transportation network is proposed whose aim is to maximize the expected traffic of networks. Our model has its realistic significance that it provides a possible way to construct network efficiently. The gravity model can generate different kinds of topology by controlling two parameters such as $\alpha$ and $\gamma$ . With $\alpha$ decreasing and $\gamma$ increasing the topology changes from a star-like network to a decentralized road network. The agreement of our simulation with the properties found in real airline network suggest the idea proposed may play a key role in the network topology.

The expected traffic may be different from the real one in networks with transfer mechanism since it only contributes to the direct interaction of two nodes. The relationship between expected and real traffic is obtained by Eq(7) which indicates how the topology information influences the traffic in the network. More studies and demonstration to this question is essential to be done in future. Besides it is interesting to note that transfer mechanism is uncommon in social networks such as citation network and movie actor collaboration network. This essential difference may be a key factor resulting in the difference behavior of weight in all kinds of networks.

This work was partially supported by Shanghai Development Foundation for Science and Technology under Grant Numbers 06JC14082 and 05XD14021.

## References

1. Watts, D.J., Strogatz, S.H.: Collective dynamics of "small-world" networks. Nature (London) 393, 440 (1998)
2. Barabsi, A.-L., Albert, R.: Emergence of scaling in random networks. Science 286, 509 (1999); Barabsi, A.-L., Albert, R., Jeong, H.: Mean-field theory for scale-free random networks. Physica A 272, 173 (1999)

3. Han, D.-D., Liu, J.-G., Ma, Y.-G., Cai, X.-Z., Shen, W.-Q.: Scale-Free Download Network for Publication. Chinese Phy. Lett. 21(9), 1855 (2004); Han, D.-D., Liu, J.-G., Ma, Y.-G.: Fluctuation of the Download Network. Chinese Phy. Lett. 25(2), 765 (2008); Han, D.-D., Qian, J.-h., Liu, J.-G.: Network Topology and Correlation Feature Affiliated with the European Airline Companies. Physica A (in press) (2008) doi:10.1016/j.physa.2008.09.021
4. Pitts, F.: The Profess. Geograph. 17, 15 (2004)
5. Guimer, R., Mossa, S., Turtschi, A., Amaral, L.A.N.: The worldwide air ransportation network: Anomalous centrality, community structure, and cities global roles. Proc. Natl. Acad. Sci. USA 102, 7794 (2005); Guimer, R., Amaral, L.A.N.: Modeling the world-wide airport network. Eur. Phys. J. B 38, 381 (2004)
6. Barrat, A., Barthlemy, M., Pastor-Satorras, R., Vespignani, A.: Proc. Natl. Acad. Sci. USA 101, 3747 (2004)
7. Smith, D.A., Timberlake, M.: Urban Stud. 32, 287 (1995)
8. Crucitti, P., Latora, V., Porta, S.: Preprint physics/0504163
9. Latora, V., Marchiori, M.: Is the Boston subway a small-world network. Physica A 314, 109 (2002)
10. Carlson, J.M., Doyle, J.: Phys. Rev. E 60, 1412 (1999)
11. Gastner, M.T., Newman, M.E.J.: The spatial structure of networks. Eur. Phys. J. B 49, 247–252 (2006)
12. Gastner, M.T., Newman, M.E.J.: Optimal design of spatial distribution networks. Phys. Rev. E 74, 016117 (2006)
13. Kaiser, M., Hilgetag, C.C.: Spatial growth of real-world networks. Phys. Rev. E 69, 036103 (2004)
14. Waxman, B.: Routing of multipoint connections. IEEE J. Selec. Areas Commun. 6, 1617 (1988)
15. Yook, S.-H., Jeong, H., Barabsi, A.-L.: Proc. Natl. Acad. Sci. USA 99, 13382 (2002)
16. Xie, Y.-B., Zhou, T., Bai, W.-j., Chen, G., Xiao, W.-K., Wang, B.-H.: Geographical networks evolving with optimal policy. Phys. Rev. E 75, 36106 (2007)
17. Bianconi, G., Barabási, A.-L.: Competition and multiscaling in evolving networks. Europhys. Lett. 54, 436 (2001)
18. Bianconi, G., Barabási, A.-L.: Bose-Einstein consideration in complex networks. Phy. Rev. Lett. 86, 5632 (2001)
19. Masuda, N., Miwa, H., Konno, N.: Geographical threshold graphs with small-world and scale-free properties. Phys. Rev. E 71, 036108 (2005)
20. Xulvi-Brunet, R., Sokolov, I.M.: Growing networks under geographical constraints. Phys. Rev. E 75, 046117 (2007)
21. Wu, Z., Braunstein, L.A., Colizza, V., Cohen, R., Havlin, S., Eugene Stanley, H.: Phys. Rev. E 74, 056104 (2006)
22. Macdonald, P.J., Almaas, E., Barabsi, A.-L.: Minimum spanning trees of weighted scale-free networks. Europhys. Lett. 72, 308 (2005)
23. Jung, W.-S., Wang, F., Eugene Stanley, H.: Gravity model in the Korean highway. Europhys. Lett. 81, 48005 (2008)
24. Liu, H.-K., Zhou, T.: Empirical study of Chine se city airline network. Acta Physica Sinica 56, 0106 (2007)
25. Liu, J.S., Chen, Y.G.: Scintia Geographica Sinica 20, 0528 (2000)
26. Newman, M.E.J.: Mixing patterns in networks. Phys. Rev. E 67, 026126 (2003)
27. Ravasz, E., Barabási, A.-L.: Hierarchical organization in complex networks. Phys. Rev. E 67, 026112 (2003)
28. Barrat, A., Barthélemy, M., Vespignani, A.: Weighted evolving networks: coupling topology and weight dynamics. Phy. Rev. Lett. 92, 228701 (2004)