Enhancing the transmission efficiency by edge deletion in scale-free networks

Guo-Qing Zhang,[1,2,*] Di Wang,[1,2] and Guo-Jie Li[1]
[1]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing, 100080, People's Republic of China*
[2]*Graduate University of Chinese Academy of Sciences, Beijing, 100080, People's Republic of China*
(Received 25 January 2007; published 13 July 2007)

How to improve the transmission efficiency of Internet-like packet switching networks is one of the most important problems in complex networks as well as for the Internet research community. In this paper we propose a convenient method to enhance the transmission efficiency of scale-free networks dramatically by kicking out the edges linking to nodes with large betweenness, which we called the "black sheep." The advantages of our method are of facility and practical importance. Since the black sheep edges are very costly due to their large bandwidth, our method could decrease the cost as well as gain higher throughput of networks. Moreover, we analyze the curve of the largest betweenness on deleting more and more black sheep edges and find that there is a sharp transition at the critical point where the average degree of the nodes $\langle k\rangle \to 2$.

DOI: 10.1103/PhysRevE.76.017101 PACS number(s): 89.75.Fb, 89.75.Hc, 89.40.$-$a

I. INTRODUCTION

In recent years, network structure and dynamics have attracted tremendous interest [1–5]. One of the most important motivations is that large communication networks, such as the Internet, wireless networks, and so on, have become the most influential media in modern society. The explosive growth of the number of users and hence the amount of traffic has posed a number of problems, which are not only important in practice, e.g., avoiding undesired congestion, but also of theoretical interest as an interdisciplinary topic [6,7]. Prompted by other disciplines like statistical physics, biology, and sociology, such interest has flourished in the broader framework of network science [8].

In transportation systems, e.g., the Internet, the underlying networks play a significant role in the traffic dynamics taking place upon them [9–12]. To the best of our knowledge, recent evidence has shown that the autonomous system (AS) level topology of the Internet is scale-free, i.e., the degree distribution follows a power law [13,14]. Thus, how to improve the efficiency of the systems, taking account of the topological structures, is a crucial problem. Some recent studies in this direction can be roughly classified into two categories. One involves optimizing the underlying networks [15,16] and the other developing better routing strategies [17–21]. The former is maybe useful to P2P (peer-to-peer) networks [22] as its logical structure is demonstrated artificially by the controlling software, while the latter is practical in packet transmission for some applications on the Internet, of which the topological structure is not expected to be changed. However, compared with greatly redesigning or modifying the topology of the Internet, it is actually easy to delete some edges. In this paper, we propose a method to remarkably improve the transmission efficiency of Internet-like networks by deleting a few edges linking to nodes with relatively large betweenness [23–26].

The paper is organized as follows. In Sec. II, the prototype traffic model, which was used in many former studies, is introduced. In Sec. III, we present our method, the simulation results, and the corresponding discussions. The conclusion is given in Sec. IV.

II. THE MODEL AND NOTATION

Several models have been proposed to simulate packet traffic dynamics on complex networks by introducing random generation of packets in each time step and various routing strategies. We here adopt one widely used before.

In each time step, each node generates R/N packets, where N is the size of the network and R is a parameter tuning the generation rate, i.e., there are R packets generated in the network at each step. When R/N is not an integer, we create $\mathrm{Int}(R/N)$ packets determinately and create a packet simultaneously with probability $p=R/N-\mathrm{Int}(R/N)$, where $\mathrm{Int}(R/N)$ is the integral part of R/N and thus p is the fractional part. The packets are initialized with random destinations. Moreover, we set a transmission capacity C_i to the node i, $i=1,2,\ldots,N$, which means that, at each time step, the maximal number of packets transferred by the node i to the next node according to the routing table is C_i. When the node cannot transfer all the packets accumulated in its queue, it deals with them following the first-in-first-out rule. Hence, the routing strategy also plays an important role in the traffic dynamics. We see that, in the Internet, the routing strategy of within domains is the shortest path algorithm and, between domains, i.e., for the AS level, the border gateway protocol causes the packets to be transmitted along almost the shortest path [27]. Therefore, in the paper, we adopt the shortest path routing strategy. When a packet reaches its destination through the shortest path routing, the packet will be deleted from the system.

In order to analyze the transition from free flow to a congested state, we use the order parameter presented in previous studies [28],

$$\mu(R) = \lim_{t\to\infty} \frac{\langle \Delta W\rangle}{R\Delta t}, \tag{1}$$

where $W(t)$ is defined as the number of packets on the network at time step t, and $\Delta W = W(t+\Delta t) - W(t)$, with $\langle\cdots\rangle$ in-

*Electronic address: gqzhang@ict.ac.cn

dicating averaging over time windows of width Δt. In other words, the order parameter μ represents the ratio between the existing flow and the inflow of packets calculated through a long enough time period. Obviously, in the free flow state, i.e., where there is no congestion in the network, the system can deal with the generated packets and thus the existing flow is close to zero. Otherwise, in the congested state, the number of generated packets is too large to be transmitted and the flow existing in the network will also be large, which causes the order parameter to approach 1.

As mentioned above, we adopt the shortest path routing in this paper, and hence another characteristic quantity, the betweenness of a node, is of utmost importance in traffic dynamics. The betweenness of node v is defined as [23]

$$g_v = \sum_{s \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}, \qquad (2)$$

where σ_{st} is the number of shortest paths going from node s to node t and $\sigma_{st}(v)$ is the number of shortest paths going from s to t and passing through the node v. In general, the nodes with large betweenness (hub nodes) are part of more shortest paths than other peripheral nodes. Obviously, for a traffic system networked according to the shortest path routing, these hub nodes will be congested more easily than others. Therefore we can improve the transmission efficiency of the traffic network by controlling or detouring properly around the nodes with large betweenness. Our method described below follows this idea.

III. METHOD, RESULTS, AND DISCUSSIONS

In this section we will give our method to improve the transmission efficiency of the traffic network, show the simulation results, and then discuss them extensively.

As analyzed before, the nodes with large betweenness will be congested more easily than the others. When a few hub nodes are congested, the others can still deal with packets successfully. Because of the edges connecting the hub nodes, the hub nodes will become more and more congested. In contrast, if there are fewer edges linking the hub nodes, the packets will make a detour round the hub nodes, and hence the capacity of the networked system will be enhanced.

First, we calculate the betweenness of each node according to the definition mentioned above. Next the edge adjoining two nodes i and j is given a weight $w_{ij}=g_i g_j$, where g_i and g_j are the betweennesses of the two nodes, respectively, and the value of w_{ij} will not be changed in the following deletion process. Then we sort the edges with their weights from large to small. Finally, a fraction f_d of the edges ranked first are deleted. However, we should keep the connectedness of the network which means that, if deleting an edge will cause some nodes to be disconnected, we will not delete it, and go to deal with the edge ranked next.

In order to test our method we should generate Internet-like topological networks. There exist several models of Internet topology (see, e.g., the positive-feedback preference () model [32] and the locality-driven (LD) model [33]). Here, for simplicity, we select the Barabasi-Albert (BA) model [2]

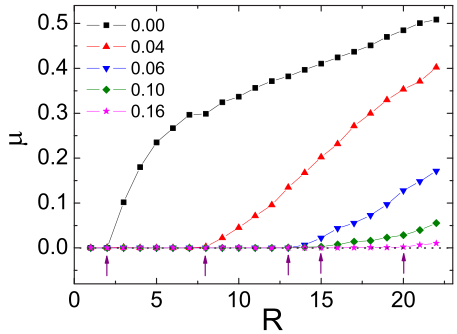

FIG. 1. (Color online) Order parameter μ versus rate R of generating packets with different values of f_d: 0.0, 0.04, 0.06, 0.10, 0.16 and $C=1$. The network parameters are $N=1225$, $\langle k \rangle=4$. The arrows point to critical values R_c for different f_d. The dashed line is $\mu=0$ as a reference.

to generate the topological network, the degree distribution of which is a power law $p(k) \propto k^{-3.0}$ where $p(k)$ is the ratio of the nodes with degree k to the number of all nodes in the network. The model could represent the heterogeneous node degree of many real-world networks, including the Internet AS level topology, the logical topology of unstructured P2P distributed systems (e.g., Gnutella [29]), and so on. In the simulation, we set the capacity of each node mentioned in Sec. II to a constant, i.e., $C_i=C$, $i=1,2,\ldots,N$ [30]. All the simulations in this paper are performed with $C=1$ and the mean degree of the network $\langle k \rangle=4$. It is worth pointing out that these values of C and $\langle k \rangle$ involve no loss of generality and do not change the analysis and conclusion of this paper.

Figure 1 displays the order parameter μ versus the packet generation rate R for different values of the fraction f_d of edges deleted. One can see that there is a critical value R_c of the generation rate, above which the order parameter is not zero and grows with increasing R, while below which μ is always equal to zero. Thus the critical value R_c can be used to represent the transmission capacity of the traffic network system. That is to say, the larger the value of R_c, the better the efficiency of the traffic network system. It is shown in this figure that the value of R_c increases with the growth of f_d. For example, $R_c=2$ when $f_d=0$, which means that scale-free networks have poor efficiency of transmission, while for $f_d=0.04$, 0.06, 0.1, and 0.16, the values of R_c are 8, 13, 15, and 20, respectively. But this is not the whole story. In Fig. 2(c) we show the value of R_c and the average shortest path length L versus the deleted fraction f_d. The curve of R_c presents an inverse U shape, and the curve of L goes up very slowly when $f_d \lesssim 0.45$ and rises sharply when $f_d \to 0.5$. This shows that the efficiency of the network will become very much higher on expurgating a small fraction of edges with large weight, whereas the efficiency will be lower when deleting many edges. Figure 2(a) displays the degree distribution of networks with various values of f_d. Here we emphasize the remarkable increase of R_c after deleting just a few edges, while the topology structure of the network is still nearly scale-free. For example, when $f_d=0.08$ the value of R_c can be improved from 2 to 13 [see Fig. 2(c)] while the degree distribution is nearly a power law [see Fig. 2(a)]. Moreover, the average shortest path length L increases slowly in

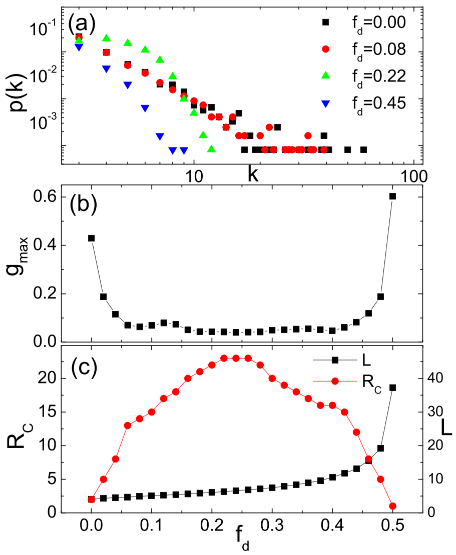

FIG. 2. (Color online) (a) Degree distribution of networks with f_d=0.00,0.08,0.22,0.45. (b) Maximal betweenness in the network versus f_d. (c) R_c and the average shortest path length L versus f_d. The initialized simulation parameters are the same as the previous ones.

the range f_d=0.25–0.4 in which the efficiency decreases very fast. We recall previous studies of the relation between R_c and the largest betweenness g_{max} of the nodes in a network [15],

$$R_c \propto \frac{1}{g_{max}}. \tag{3}$$

This indicates that the value of R_c is approximately inversely proportional to the value of the maximal betweenness g_{max}. We display the curve of g_{max} versus f_d in Fig. 2(b). It is shown that the value of g_{max} goes up very quickly when f_d approaches 0.5. This can be explained as follows. Since the mean degree $\langle k\rangle$=4, when $f_d \to 0.5$ the number of edges is almost equal to the number of nodes. This means that, if we delete a few more edges, the network will be disconnected. When f_d approaches the critical point f_d=0.5, the network will be modified to a treelike topology that causes a few crossroad nodes; hence the large betweenness of such nodes. The network topology with three typical values of f_d =0.0,0.24,0.50, is displayed in Fig. 3, and confirms our analysis. It is worthy of note that the remaining treelike topology, after many links are deleted, looks like but is different from the gradient networks [31] studied by Toroczkai *et al.* In [31], the authors studied traffic congestion in directed gradient networks and have shown the interesting result that the jamming is limited in scale-free systems, i.e., even though the network is congested, some nodes can also process some packets. In contrast, we study in this paper the traffic congestion in undirected networks and the remaining treelike topology has low efficiency of transmission.

In addition, we should point out that R_c=2, in Fig. 1, is true just for scale-free networks generated by the BA model with average degree $\langle k\rangle$=4 and no links removed. When $\langle k\rangle$ increases, the average shortest path length decreases, and thus the maximal betweenness decreases. According to Eq. (3), R_c will increase. However, our method will still be effective.

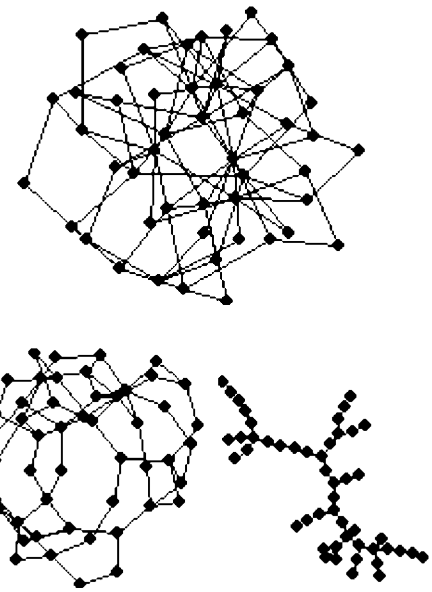

FIG. 3. Network topologies with three typical values of f_d=0.00,0.24,0.50 (from top to bottom, left to right). The initial network parameters are N=50 and $\langle k\rangle$=4.

As shown above one can see that deleting just a few edges with large weights can improve remarkably the transmission efficiency and simultaneously reduce the cost of network construction and maintenance, while the average shortest path length increases very slowly. Thus we call those edges "black sheep" and kick them out. It is noteworthy that the method in Ref. [21] is different from ours. In the former, the authors proposed an interesting and effective routing scheme inspired by the extreme optimization method. In contrast, our method in this paper is to adjust the topology structure to enhance the throughput of scale-free communication networks.

IV. CONCLUSION

To summarize, we have proposed a convenient method, kicking out the edges linking to nodes with large betweenness, which we called the black sheep, to improve the efficiency of scale-free traffic network systems. Compared with previous studies in this direction, including redesigning the topology and designing better routing strategies, its advantages are of convenience and practical importance. In this paper we set the weight of the edge adjoining node i and j as $w_{ij}=g_i g_j$, where g_i and g_j are the betweennesses of nodes i and j, respectively. The calculation of the betweenness needs overall information about the network. An alternative choice is to set $w_{ij}=k_i k_j$, where k_i and k_j are the degrees of nodes i and j. Actually, we have implemented the alternative method and found that its performance is just a little worse. Thus, for very large networks, one can choose the alternative method. In addition, we have performed our method on the more realistic PFP, LDPFP model [32,33] and the actual AS level Chinese Internet topology [14]; The conclusion is almost the same as on the BA model.

Furthermore, it is well known that the black sheep edges are costly lines with large bandwidth in practice. Therefore our method can improve the efficiency of communication network systems as well as decrease the cost. In short, our method solves two problems at one time.

ACKNOWLEDGMENTS

The authors thank Gang Yan for useful discussions. This work is partly supported by the National Natural Science Foundation of China under Grant No. 60673168, the Hi-Tech Research and Development Program of China under Grant No. 2006AA01Z207, and the ICT innovation fund under Grant No. 20066033.

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[4] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[5] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[6] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University Press, New York, 2003).
[7] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, U.K., 2004).
[8] M. E. J. Newman, A.-L. Barabási, and D. J. Watts, *The Structure and Dynamics of Complex Networks* (Princeton University Press, Princeton, NJ, 2005).
[9] B. Tadić and G. J. Rodgers, Adv. Complex Syst. **5**, 445 (2002).
[10] B. Tadić, in *Modeling Traffic of Information Packets on Graphs with Complex Topology*, edited by P. Sloot *et al.*, Lecture Notes in Computer Science Vol. 2657 (Springer, Berlin, 2003).
[11] B. Tadić, S. Thurner, and G. J. Rodgers, Phys. Rev. E **69**, 036102 (2004).
[12] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[13] M. Faloutsos, P. Faloutsos, and C. Faloutsos, in *ACM SIGCOMM Comput. Commun. Rev.*, (ACM Press, New York 1999), Vol. 29, p. 251.
[14] S. Zhou, G.-Q. Zhang, and G.-Q. Zhang, IET Commun. **1**, 209 (2007).
[15] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).
[16] V. Cholvi, V. Laderas, L. López, and A. Fernández, Phys. Rev. E **71**, 035103(R) (2005).
[17] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E **70**, 056105 (2004).
[18] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, Phys. Rev. E **73**, 046108 (2006).
[19] W.-X. Wang, C.-Y. Yin, G. Yan, and B.-H. Wang, Phys. Rev. E **74**, 016101 (2006).
[20] S. Sreenivasan, R. Cohen, E. López, Z. Toroczkai, and H. E. Stanley, Phys. Rev. E **75**, 036105 (2007).
[21] B. Danila, Y. Yu, J. A. Marsh, and K. E. Bassler, Phys. Rev. E **74**, 046106 (2006).
[22] A. Oram, *Peer-to-Peer: Harnessing the Power of Disruptive Technologies* (O'Reilly Associates, Sebastopol, CA, 2001).
[23] L. Freeman, Sociometry **40**, 35 (1977).
[24] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[25] M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001).
[26] M. Barthélemy, Eur. Phys. J. B **38**, 163 (2003).
[27] http://www.ietf.org/rfc/rfc1771.txt
[28] A. Arenas, A. Díaz-Guilera, and R. Guimerà, Phys. Rev. Lett. **86**, 3196 (2001).
[29] http://www.gnutella.com
[30] We test other capacity distribution schemes, e.g., that the capacity of a node is proportinal to its degree. The results (not displayed here) show that our method could increase R_c from 236 to 359 while f_d=0.16 and the curve is similar to that displayed in Fig. 2(c).
[31] Z. Toroczkai and K. E. Bassler, Nature (London) **428**, 716 (2004).
[32] S. Zhou and R. J. Mondragón, Phys. Rev. E **70**, 066108 (2004).
[33] Guoqiang Zhang and G. Q. Zhang, in Proceedings of the IEEE International Conference on Communications 2007, (IEEE, Glasgow, Scotland).