PHYSICAL REVIEW E **73**, 046108 (2006)

Efficient routing on complex networks

Gang Yan,[1] Tao Zhou,[1,2,*] Bo Hu,[2] Zhong-Qian Fu,[1] and Bing-Hong Wang[2]

[1]*Department of Electronic Science and Technology, University of Science and Technology of China, Hefei Anhui, 230026, People's Republic of China*

[2]*Nonlinear Science Center and Department of Modern Physics, University of Science and Technology of China, Hefei Anhui, 230026, People's Republic of China*

(Received 15 May 2005; published 7 April 2006)

We propose a routing strategy to improve the transportation efficiency on complex networks. Instead of using the routing strategy for shortest path, we give a generalized routing algorithm to find the so-called *efficient path*, which considers the possible congestion in the nodes along actual paths. Since the nodes with the largest degree are very susceptible to traffic congestion, an effective way to improve traffic and control congestion, as our strategy, can be redistributing traffic load in central nodes to other noncentral nodes. Simulation results indicate that the network capability in processing traffic is improved more than 10 times by optimizing the efficient path, which is in good agreement with the analysis.

DOI: 10.1103/PhysRevE.73.046108 PACS number(s): 89.75.Hc

Since the seminal work on scale-free networks by Barabási and Albert (BA model) [1] and on the small-world phenomenon by Watts and Strogatz [2], the structure and dynamics of complex networks have recently attracted a tremendous amount of interest and attention from the physics community (see the review papers [3–5] and references therein). The increasing importance of large communication networks such as the Internet [6], upon which our society survives, calls for the need for high efficiency in handling and delivering information. In this light, to find optimal strategies for traffic routing is one of the important issues we have to address. There have been many previous studies to understand and control traffic congestion on networks, with a basic assumption that the network has a homogeneous structure [7–11]. However, many real networks display both scale-free and small-world features, and thus it is of great interest to study the effect of network topology on traffic flow and the effect of traffic on network evolution. Guimerá *et al.* present a formalism that can cope simultaneously with the searching and traffic dynamics in parallel transportation systems [12]. This formalism can be used to optimize network structure under a local search algorithm, while to obtain the formalism one should know the global information of the whole networks. Holme and Kim provide an in-depth analysis on the vertex/edge overload cascading breakdowns based on evolving networks, and suggest a method to avoid such avalanches [13,14]. By using a global and dynamical searching algorithm aimed at the shortest paths, Zhao *et al.* provide the theoretical estimates of the communication capacity [15]. Since global information is usually unavailable in large-scale networks, Tadić *et al.* investigate the traffic dynamics on the WWW network model [16] based on local knowledge, providing insight into the relationship of global statistical properties and microscopic density fluctuations [17–19]. The routing strategies for the Internet [20] and disordered networks [21] are also studied. Another interesting issue is the interplay of traffic dynamics and network structures, which suggests a new scenario of network evolutionary [22–26].

In this context, for simplicity, we treat all the nodes as both hosts and routers [12,27]. In communication networks, routers deliver data packets by ensuring that all converge to a best estimate of the path leading to each destination address. In other words, the routing process takes place according to the criterion of the shortest available path from a given source to its destination. When the network size N is not too large, it is possible to calculate all the shortest paths between any nodes, and thus the traffic system can use a fixed routing table to process information. As for any pair of source and destination, there may be several shortest paths between them. We randomly choose one of them and put it into the fixed routing table, which is followed by all the information packets. Though it becomes impractical in huge communication systems, the fixed routing algorithm is widely used in medium-sized or small systems [28,29]. This is because the fixed routing method has obvious advantages in economical and technical costs, compared with the dynamical routing algorithm and information feedback mechanism. The model is described as follows: at each time step, there are R packets generated in the system, with randomly chosen sources and destinations. It is assumed that all the routers have the same capabilities in delivering and handling information packets, that is, at each time step all the nodes can deliver at most C packets one step toward their destinations according to the fixed routing table. We set $C=1$ for simplicity. A packet, upon reaching its destination, is removed from the system. We are most interested in the critical value R_c (as measured by the number of packets created within the network per unit time) where a phase transition takes place from free flow to congested traffic. This critical value can best reflect the maximum capability of a system handling its traffic. In particular, for $R<R_c$, the numbers of created and delivered packets are balanced, leading to a steady free traffic flow. For $R>R_c$, traffic congestion occurs as the number of accumulated packets increases with time, simply because the capacities of the nodes for delivering packets are limited. We use

*Electronic address: zhutou@ustc.edu

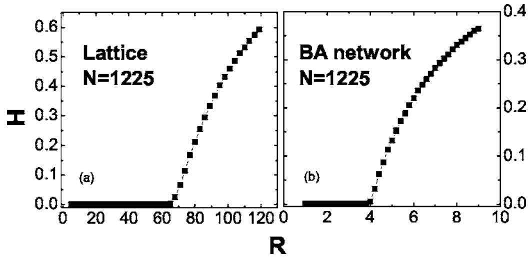

FIG. 1. The order parameter H vs R for two-dimensional lattice (a) and BA networks (b) with the same size $N=1225$. The routing algorithm at the shortest path yields $R_{\text{lattice}} \approx 65$ and $R_{\text{BA}} \approx 4.0$.

the order parameter to characterize the phase transition,

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

where $\Delta W = W(t+\Delta t) - W(t)$, with $\langle \cdots \rangle$ indicating average over time windows of width Δt, and $W(t)$ is the total number of packets in the network at time t. Figure 1 shows the order parameter H versus R for (a) the two-dimensional lattice with periodical boundary condition and (b) the scale-free BA network with average degree $\langle k \rangle = 4$ [1], given all the packets following their shortest paths. The critical point R_c in the lattice is much larger than that in the scale-free network, which can be simply explained by their different betweenness centralities (BC) distributions [30–33]. The BC of a node v is defined as

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

where σ_{st} is the number of shortest paths going from s to t and $\sigma_{st}(v)$ is the number of shortest paths going from s to t and passing through v. Moreover, BC gives in transport networks an estimate of the traffic handled by the vertices, assuming that the number of shortest paths is a zeroth-order approximation to the frequency of use of a given node. It is generally useful to represent the average BC for vertices of the same degree,

$$g(k) = \frac{1}{N_k} \sum_{v, k_v = k} g(v), \tag{3}$$

where N_k denotes the number of nodes with degree k. For most networks, $g(k)$ is strongly correlated with k. In general, the larger the degree, the larger the centrality. For scale-free networks it has been shown that the centrality approximately scales as $g(k) \sim k^{\mu}$. In comparison, the BC in the lattice will behave as a homogeneous distribution. Noticeably, in scale-free networks, traffic congestion generally occurs at nodes with the largest degree (or BC), and immediately spreads over all the nodes. When all the packets follow their shortest paths, it will easily lead to the overload of the heavily linked router, which is precisely the cause of traffic congestion. To alleviate the congestion, a feasible and effective way is to bypass such high-degree nodes in the traffic-routing design. This leads us to question the commonly used shortest-path routing mechanism.

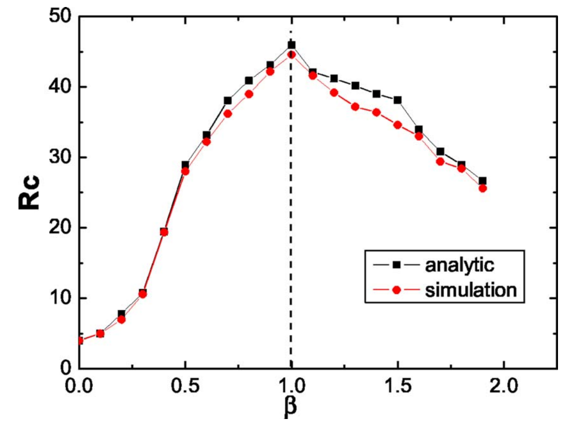

FIG. 2. (Color online). The critical R_c vs β for scale-free networks with size $N=1225$. Both Simulation and analysis indicate that the maximum of R_c corresponds to $\beta \approx 1.0$. The data shown here are the average over 10 independent runs.

Actually, the path with the shortest length is not necessarily the quickest way, considering the presence of possible traffic congestion and waiting time along the shortest path (by "shortest" we mean the path with the smallest number of links). Obviously, nodes with larger connections are more likely to bear traffic congestion, thus a packet will by average spend more waiting time to pass through a high-BC node. All too often, bypassing those high-BC nodes, a packet may reach its destination quicker than taking the shortest path. In order to find the optimal routing strategy, we define the "efficient path." For any path between nodes i and j as $P(i \to j) := i \equiv x_0, x_1, \ldots, x_{n-1}, x_n \equiv j$, denote

$$L(P(i \to j):\beta) = \sum_{i=0}^{n-1} k(x_i)^{\beta}. \tag{4}$$

The efficient path between i and j is corresponding to the route that makes the sum $L(P(i \to j):\beta)$ minimum. Obviously, $L_{\min}(\beta=0)$ recovers the traditionally shortest path length. We expect that the system behaves better under the routing rule with $\beta > 0$ than it does traditionally, and we aim to find the optimal β in this paper. In the following, the fixed routing table is designed on the basis of efficient path. If there are several efficient paths between two nodes, one is chosen at random. We are now interested in determining the phase-transition point R_c under various β, in order to address which kind of routing strategy is more flexible to traffic congestion, and therefore find the optimal β.

Aiming to estimate the value of R_c for different β, we define the efficient betweenness centralities (EBC) of a node v as

$$g^{\beta}(v) = \sum_{s \neq t} \frac{\sigma_{st}^{\beta}(v)}{\sigma_{st}^{\beta}}, \tag{5}$$

where σ_{st}^{β} is the number of efficient paths for a given β going from s to t and $\sigma_{st}^{\beta}(v)$ is the number of efficient paths for a given β going from s to t and passing through v. It is well known that for low values of R the system reaches a steady state in which $W(t)$ fluctuates around a finite value. As R

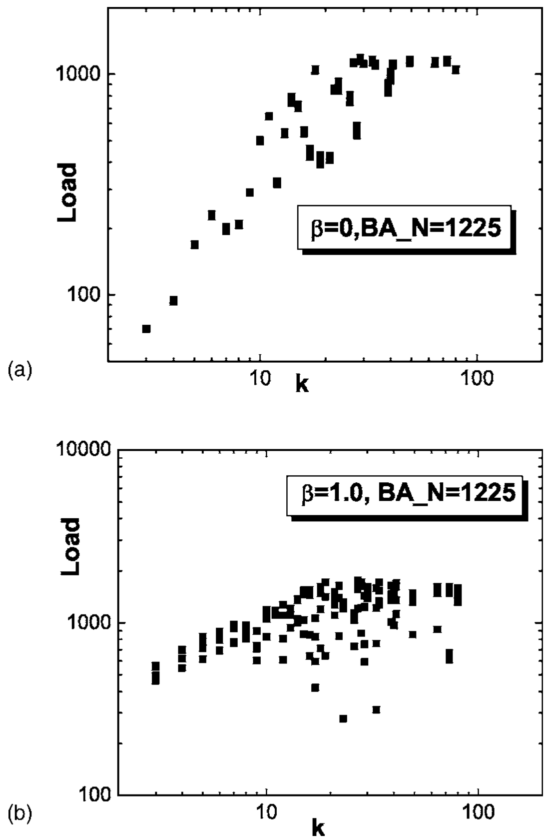

FIG. 3. The load distribution when congestion occurs for a BA network with size $N=1225$. (a) The case of $\beta=0$ where $R_c=4.0$ and we set $R=10$. (b) The case of $\beta=1$ where $R_c=45$ and we set $R=60$.

increases, the system undergoes a continuous phase transition to a congested phase. Below the critical value R_c, there is no accumulation at any node in the network and the number of packets that arrive at node u is, on average, $Rg_u/N(N-1)$. Therefore, a particular node will collapse when $Rg_u/N(N-1)>C_u$, where g_u is the betweenness and C_u is the transferring capacity of node u. Considering the transferring capacity of each node is fixed to 1 in this paper and congestion occurs at the node with the largest betweenness, R_c can be estimated as [12,15]

$$R_c = N(N-1)/g_{\max}, \tag{6}$$

where $g_{\max}$ is the largest BC of the network. Similarly, for different β, we can estimate $R_c(\beta)$ as

$$R_c(\beta) = N(N-1)/g^{\beta}_{\max}, \tag{7}$$

where $g^{\beta}_{\max}$ is the largest EBC for a given β.

In Fig. 2, we report the simulation results for the critical value R_c as a function of β on BA networks, which is in good agreement with the analysis. As one can see, R_c first increases with β and then decreases, with the maximum of R_c corresponding to $\beta \approx 1.0$. In comparison with the shortest path routing case (i.e., $\beta=0$), the capability of the network in freely handling information is greatly improved, from $R_c \approx 4.0$ when $\beta=0$ to $R_c \approx 45$ when $\beta=1.0$, more than ten times. This result suggests the effectiveness of the routing strategy by our efficient path length. Figure 3 shows the optimized behavior of our efficient path routing in load distribution when congestion just occurs (b), in comparison with that of the shortest path routing mechanism (a). Clearly, the heavy load on central nodes (with highest connectivity) is strongly redistributed to those nodes with a lower degree by using an efficient path routing table. We also report in Fig. 4(a) the average actual path length L_{ave} versus the network size N under various values of β. As one can see, although L_{ave} increases with β, the small-world property $L_{\text{ave}} \sim \ln N$ is still maintained. The system capability in processing information is considerably enhanced at the cost of increasing the average routing path length. Such a sacrifice may be worthwhile when a system requires large R_c. Moreover, we investigate the average transporting time $\langle T \rangle$ of packets. The results in Figs. 4(b) and 4(c) show that $\langle T \rangle$ and H indicate the same critical value R_c.

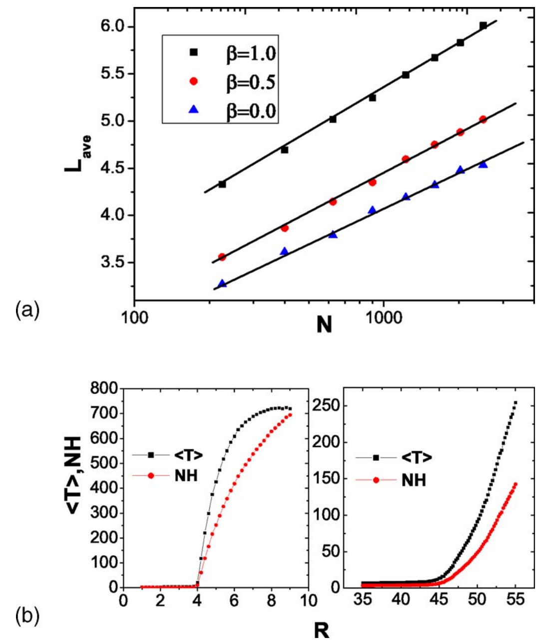

FIG. 4. (Color online) (a) The average actual path length L_{av} vs the network size N under various values of β, by using the efficient path routing. (b) and (c) show $\langle T \rangle$ and NH vs R for $\beta=0.0$ and 1.0, respectively, where $N=1225$.

To realize the routing strategy we have studied, each router must have the complete knowledge of the network topology, which is often difficult for large-scale systems. Anyway, it is possible to divide one large system into several autonomous subsystems in which every router has its local topological knowledge. Thus, the hierarchical structure of the network will make possible the implementation of our routing strategy. This paper has mainly discussed how to effectively design a routing algorithm when the capabilities of processing information are the same for all the nodes. To account for the network topology, one can assume that the capabilities for processing information are different for different nodes, depending on the numbers of links or the number of the shortest paths passing through them [15]. In addition, the shortest path is shortest just in a topological sense;

in practice, it is not necessarily the best. As for a single packet, its best routing as we have argued is not absolutely the shortest path. From the systematic view, the total information load that a communication network can freely handle without congestion depends on all the packets reaching their destinations in a systematically optimal time. We use R_c to denote the upper limit of the total information load that a communication system can handle without congestion. This parameter reflects the system capability in processing information under a certain routing strategy. An effective way to alleviate traffic congestion for scale-free networks is to make the heavily linked nodes as powerful and efficient as possible for processing information. This is further supported by examining the effect of enhancing the capabilities of these nodes. Moreover, we have checked the efficient routing on scale-free networks with $\gamma=2.0$ and 2.5 (obtained by the extensional BA model [34,35]), where γ is the exponent of power-law distribution $p(k) \propto k^{-\gamma}$. We obtained the same optimal value as $\beta=1.0$ on these different scale-free networks. In average, the capability R_c increases 7.5 times and 9.3 times for the above two cases, respectively. In addition, some models aiming at communication networks, such as the models of the World-Wide-Web[16] and the Internet [36,37], are closer to reality than BA networks. To investigate the present traffic model and routing strategy for these network models is significant in practice. This will be done in future works.

While our model is based on computer networks, we expect it to be relevant to other practical networks in general. Our studies may be useful for designing communication protocols for complex networks, considering there appears no increase in its algorithmic complexity. The optimized routing strategy studied in this paper can be easily implemented in practice.

Many previous works focus on the relationship between the distribution of BC and the capability of communication networks, with a latent assumption that the information packets go along the shortest paths from source to destination. Therefore, the BC is always considered as a static topological measure of networks. Here we argue that this quantity is determined both by the routing algorithm and network topology, thus one should pay more attention to the design of routing strategies. We believe this work may enlighten readers on this subject and be helpful for understanding the intrinsic mechanism of network traffic. Finally, it is worthwhile to emphasize that we have found some evidence indicating there may exist some common features between network traffic and synchronization on a dynamical level, thus the present method may also be useful for enhancing the network synchronizability [38–42].

The authors wish to thank Dr. Wen-Xu Wang for discussions. This work has been partially supported by the National Natural Science Foundation of China under Grants No. 70471033, No. 10472116, No. 10532060, No. 70571074, and No. 10547004, the Specialized Research Fund (SRFDP No. 20020358009), the Special Research Founds for Theoretical Physics Frontier Problems under Grant No. A0524701, and Specialized Program under President Funding of the Chinese Academy of Science.

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[2] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[3] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[4] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Phys. Rep. **424**, 175 (2006).
[5] T. Zhou, Z.-Q. Fu, and B.-H. Wang, e-print physics/0508096; , Prog. Natl. Sci. (to be published).
[6] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, 2004).
[7] H. Li and M. Maresca, IEEE Trans. Comput. **38**, 1345 (1989).
[8] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, Comput. Commun. Rev. **23**, 283 (1993).
[9] M. S. Taqqu, W. Willinger, and R. Sherman, Comput. Commun. Rev. **27**, 5 (1997).
[10] M. E. Crovella and A. Bestavros, IEEE/ACM Trans. Netw. **5**, 835 (1997).
[11] A. Arenas, A. Díaz-Guilera, and R. Guimerá, Phys. Rev. Lett. **86**, 3196 (2001).
[12] R. Guimerá, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).
[13] P. Holme and B. J. Kim, Phys. Rev. E **65**, 066109 (2002).
[14] P. Holme, Phys. Rev. E **66**, 036119 (2002).
[15] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[16] B. Tadić, Physica A **293**, 273 (2001).
[17] B. Tadić, S. Thurner, and G. J. Rodgers, Phys. Rev. E **69**, 036102 (2004).
[18] B. Tadić, Lect. Notes Comput. Sci. **2657**, 136 (2003).
[19] B. Tadić and G. J. Rodgers, Adv. Complex Syst. **5**, 445 (2002).
[20] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E **70**, 056105 (2004).
[21] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin, and H. E. Stanley, Phys. Rev. Lett. **91**, 168701 (2003).
[22] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[23] A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. Lett. **92**, 228701 (2004).
[24] Y. Moreno, R. Pastor-Satorras, A. Vazquez, and A. Vespignani, Europhys. Lett. **62**, 292 (2003).
[25] W.-X. Wang, B.-H. Wang, B. Hu, G. Yan, and Q. Ou, Phys. Rev. Lett. **94**, 188702 (2005).
[26] W.-X. Wang, B. Hu, T. Zhou, B.-H. Wang, and Y.-B. Xie, Phys. Rev. E **72**, 046140 (2005).
[27] T. Zhou, G. Yan, B.-H. Wang, Z.-Q. Fu, B. Hu, C.-P. Zhu, and W.-X. Wang, Dynamics of Continuous, Discrete and Impulsive Systems Series B (to be published).
[28] A. S. Tanenbaum, *Computer Networks* (Prentice Hall, Engle-

wood Cliffs, NJ, 1996).
[29] C. Huitema, *Routing in the Internet* (Prentice Hall, Upper Saddle River, NJ, 2000).
[30] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
[31] M. E. J. Newman and M. Girvan, Phys. Rev. E **69**, 026113 (2004).
[32] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[33] M. Barthélemy, Eur. Phys. J. B **38**, 163 (2003).
[34] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, Phys. Rev. Lett. **85**, 4633 (2000).
[35] P. L. Krapivsky and S. Redner, Phys. Rev. E **63**, 066123 (2001).
[36] S. Zhou and R. J. Mondragón, Phys. Rev. E **70**, 066108 (2004).
[37] J.-G. Liu, Y.-Z. Dang, Z.-T. Wang, and T. Zhou, e-print physics/0510064.
[38] A. E. Motter, C. Zhou, and J. Kurths, Phys. Rev. E **71**, 016116 (2005).
[39] M. Chavez, D.-U. Hwang, A. Amann, H. G. E. Hentschel, and S. Boccaletti, Phys. Rev. Lett. **94**, 218701 (2005).
[40] M. Zhao, T. Zhou, B.-H. Wang, and W.-X. Wang, Phys. Rev. E **72**, 057102 (2005).
[41] T. Zhou, M. Zhao, and B.-H. Wang, Phys. Rev. E **73**, 037101 (2006).
[42] C.-Y. Yin, B.-H. Wang, W.-X. Wang, T. Zhou, and H.-J. Yang, Phys. Lett. A **351**, 220 (2006).