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,*
[2]*Nonlinear Science Center and Department of Modern Physics,*
University of Science and Technology of China,
Hefei Anhui, 230026, PR China

In this letter, we propose a new 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 largest degree are very susceptible to traffic congestion, an effective way to improve traffic and control congestion, as our new strategy, can be as redistributing traffic load in central nodes to other non-central 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.

PACS numbers: 02.50.Le, 05.65.+b, 87.23.Ge, 87.23.Kg

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 on complex networks have recently attracted a tremendous amount of interest and devotion from physics community. The increasing importance of large communication networks such as the Internet, 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[3]. However, many realistic networks like the Internet 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[4]. This formalism can be used to optimal networks 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[5]. Since the load of a certain vertex/edge is defined at its betweenness[6], there is also a latent assumption that the routing protocols involved global knowledge of networks. By using global and dynamical searching algorithm aiming at shortest paths, Zhao et al provide the theoretical estimates of the communication capacity [7]. Since the global information
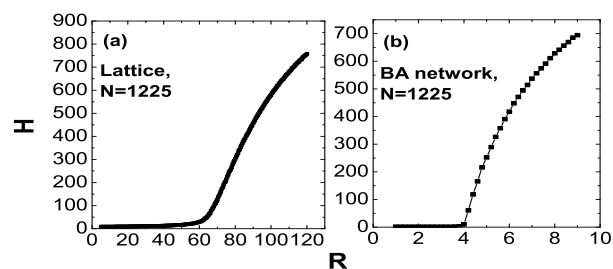

FIG. 1: The order parameter H versus R for two-dimensional lattice (a) and BA networks (b) with the same size $N = 1225$. The routing algorithm at the shortest path yields $R_{Lattice} \approx 60$ and $R_{BA} \approx 4.0$.

is usually unavailable in large-scale networks, Tadić et al investigate the traffic dynamics on WWW network model[8] based on local knowledge, providing insights into the relationship of global statistical properties and microscopic density fluctuations[9]. The routing strategies for Internet[10] and disordered networks[11] are also studied. Another interesting issue is the interplay of traffic dynamics and network structures, which suggests a new scenario of network evolutionary[12].

In this context, the information processor are routers that have the same function in the postal serves. For simplicity, we treat all the nodes as both hosts and routers [4]. 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 following 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
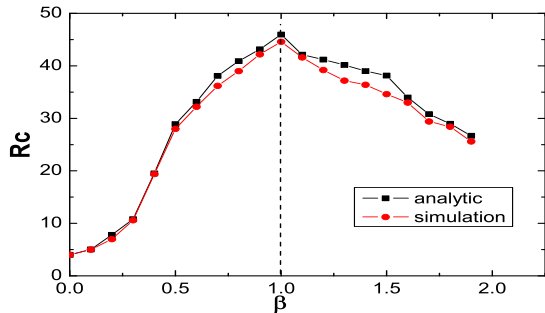
*Electronic address: zhutou@ustc.edu

FIG. 2: (Color online). The critical R_c versus 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.

any pair of source and destination, there may be several shortest paths between them. We can 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 mid- or small-systems [13]. It is because that the fixed routing method has obvious advantages in economical and technical costs, compared with dynamical routing algorithm and information feed-back 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, once reaching its destination, is removed from the system. We are most interested in the critical value R_c of information generation (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 for that the capacities of nodes for delivering packets are limited. We use the order parameter to characterize the phase transition

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

where $\Delta W = W(t + \Delta t) - W(t)$ with $\langle \cdots \rangle$ indicates average over time windows of width Δt, and W(t) is the total number of packets in the network at time t. Fig. 1 shows the order parameter H versus R for (a) the two-dimensional lattice with periodical boundary condition
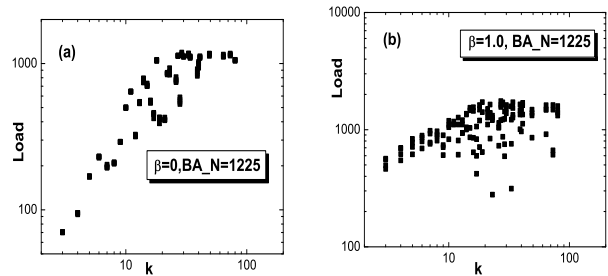

FIG. 3: The load distribution when congestion occurs for BA network with size $N = 1225$. (a) The case of $\beta = 0$ where $R_c = 4.0$ and we set $R = 10$. (b) The case of $\beta = 1$ where $R_c = 45$ and we set $R = 60$.

and (b) the scale-free BA network with average degree $\langle k \rangle = 4$ [1], given all the packets follows their shortest paths. We see that the critical point R_c in lattice is much larger than that in scale-free network. This phenomena can be simply explained by their different betweenness centralities (BC) [6]. The BC of a node v is defined as

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

where σ_{st} is the number of shortest paths going from s to t and $\sigma_{st}(v)$ is the number of shortest paths going from s to t and passing through v. This definition means that central nodes are part of more shortest paths within the network than peripheral nodes. Moreover, BC gives in transport networks an estimate of the traffic handled by the vertices, assuming that the number of shortest paths is a zero-th 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

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

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 scales with k as $g(k) \sim k^\mu$ where μ depends on the network. Hence in the scale-free networks, the betweenness distribution also obeys a power-law form. In comparison, the BC in lattice will behave 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 just the key of traffic congestion. To alleviate the congestion, a feasible and effective way is to bypass such high-degree nodes in traffic routing design. This leads us to question the commonly used shortest-path routing mechanism.
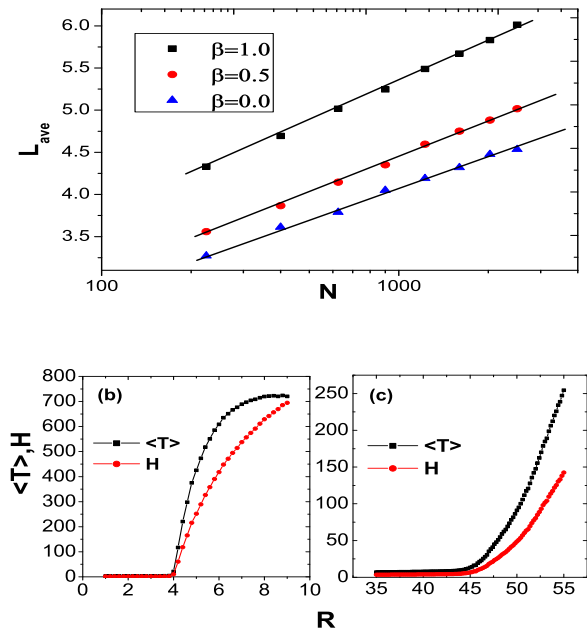
FIG. 4: (Color online). (a) The average actual path length L_{ave} versus the network size N under various values of β, by using the efficient path routing. (b) and (c) show $\langle T \rangle$ and H versus R for $\beta = 0.0$ and $\beta = 1.0$ respectively.

Actually, the path with 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 smallest number of links). Obviously, nodes with larger connections are more likely to bear traffic congestion. Intuitively, a packet will by average spends more time to travel through a node with higher BC (for waiting time). 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 & j as $P(i \to j) := i \equiv x_0, x_1, \cdots x_{n-1}, x_n \equiv j$, we define

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

For any given β, 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 letter. In the following, the fixed routing table is designed on the basis of efficient path. If there are several efficient paths between two nodes, the 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 betweennes centralities(EBC) of a node v as,

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

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. Since congestion occurs at the node with the largest betweenness, R_c can be estimated as[4, 7],

$$R_c = \frac{N(N-1)}{g_{max}} \quad (6)$$

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

$$R_c(\beta) = \frac{N(N-1)}{g_{max}^\beta} \quad (7)$$

where g_{max}^β 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 corresponds 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 us the effectiveness of the routing strategy by our efficient path length. Fig. 3 shows the optimized behavior of (b) our efficient path routing in load distribution when congestion just occurs, in comparison with that of (a) the shortest path routing mechanism. Clearly, the heavy load on central nodes (with highest connectivity) is strongly redistributed to those nodes with less degree by using 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 $L_{ave} \sim \ln N$ is still kept. 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 Fig.4(b) and Fig.4(c) show that $\langle T \rangle$ and H indicate the same critical value R_c.

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 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 [7]. 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 can reach 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 reflect the system capability in processing information under certain routing strategy. An effective way found 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. In addition, some previous models aiming at communication networks are closer to reality than BA networks, such as directed scale-free model for World-Wide-Web [8] and positive-feedback preference model for Internet [14]. To investigate the present traffic model and routing strategy onto these network models are significant in practice. These will be among the 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 evidences indicating there may exist some common features between network traffic and synchronization in dynamical level, thus the present method may be also useful for enhancing the network synchronizability[15].

[1] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[2] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[3] H. Li and M. Maresca, IEEE Trans. Comput. **38**, 1345 (1989); W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, Comput. Commun. Rev. **23**, 283 (1993); M. S. Taqqu, W. Willinger, and R. Sherman, Comput. Commun. Rev. **27**, 5 (1997); M. E. Crovella and A. Bestavros, IEEE/ACM Trans. Netw. **5**, 835 (1997); A. Arenas, A. Díaz-Guilera, and R. Guimerá, Phys. Rev. Lett. **86**, 3196 (2001).

[4] R. Guimerá, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).

[5] P. Holme and B. J. Kim, Phys. Rev. E **65**, 066109 (2002); P. Holme, Phys. Rev. E **66**, 036119 (2002).

[6] K.-I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001); M. E. J. Newman, Phys. Rev. E **64**, 016131 (2001); M. Barthélemy, Eur. Phys. J. B **38**, 163 (2003).

[7] L. Zhao, Y.-C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).

[8] B. Tadić, Physica A **293**, 273(2001).

[9] B. Tadić, S. Thurner and G. J. Rodgers, Phys. Rev. E **69**, 036102 (2004); B. Tadić, *Modeling Traffic of Information Packets on Graphs with Complex Topology*, edited by P. Sloot et. al., Lecture Notes in Computer Science Vol. 2657, P.136 (Springer, Berlin, 2003); B. Tadić, and G. J. Rodgers, Adv. in Comp. Sys. **5**, 445 (2002).

[10] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E **70**, 056105 (2004).

[11] L. A. Braunstein, S. V. Buldyrev, R. Cohen, S. Havlin and H. E. Stanley, Phys. Rev. Lett. **91**, 168701 (2003).

[12] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004); A. Barrat, M. Barthélemy, and A. Vespignani, Phys. Rev. Lett. **92**, 228701 (2004); Y. Moreno, R. Pastor-Satorras, A. Vazquez and A. Vespignani, Europhys. Lett. **62**, 292 (2003); W.-X. Wang, B.-H. Wang, B. Hu, G. Yan, and Q. Ou, Phys. Rev. Lett. **94**, 188702 (2005).

[13] A. S. Tanenbaum, *Computer Networks* (Prentice Hall Press, 1996).

[14] S. Zhou, and R. J. Mondragón, Phys. Rev. E **70**, 066108 (2004).

[15] A. E. Motter, C. Zhou, and J. Kurths, Phys. Rev. E **71**, 016116(2005); M. Chavez, D. -U. Huang, A. Amann, H. G. E. Hentschel, and S. Boccaletti, Phys. Rev. Lett. **94**, 218701(2005); M. Zhao, T. Zhou, B. -H. Wang, and W. -X. Wang, arXiv: cond-mat/0507626.

(b)

BA Network, N=1225