Effects of network structure and routing strategy on network capacity

Zhen Yi Chen* and Xiao Fan Wang†

Department of Automation, Shanghai Jiao Tong University, Dongchuan Road 800, Shanghai 200240, People's Republic of China

(Received 16 September 2005; revised manuscript received 20 December 2005; published 3 March 2006)

The capacity of maximum end-to-end traffic flow the network is able to handle without overloading is an important index for network performance in real communication systems. In this paper, we estimate the variations of network capacity under different routing strategies for three different topologies. Simulation results reveal that the capacity depends on the underlying network structure and the capacity increases as the network becomes more homogeneous. It is also observed that the network capacity is greatly enhanced when the new traffic awareness routing strategy is adopted in each network structure.

DOI: 10.1103/PhysRevE.73.036107 PACS number(s): 89.75.Hc, 89.20.−a, 89.20.Hh

I. INTRODUCTION

The research on the efficiency of transport of information in social, biological, and electronic communication systems, etc., is significantly important in different aspects of natural science and more generally in its practical application. One of the main focuses is how to make the flow of the network most efficient, which means to maximize the network capacity at the same time minimizing the delivery time and loss of information. There are two ingredients, the local or global topological properties of the network and the microscopic dynamical process involved in the information transportation procedure, which are believed to affect those complex dynamical processes. Also, the efficient performance of these systems is affected by the ability of the system to avoid congestion. It is thus of great interest to study the effect of network topology on traffic flow and to find out optimal strategies for traffic routing on certain network structure.

A great number of works in the subject of dynamics of traffic transitions has been carried out for regular and random graphs [1–8]. However, recent works reveal that many realistic networks, as diverse as the Internet, the WWW, the social and biological networks are complex with scale-free and small-world features, which are far from being completely regular or completely random [9–14]. The well established model by Barabási and Albert (BA), which characterizes the scale-free nature of many networks, pointed out that the probability a given node has k connections with other nodes follows a power law distribution $P_k \sim k^{-\gamma}$. Based on the BA model and its extensions [15–17], the dynamic processes taken place on these complex topologies have been investigated extensively, such as the search or congestion process in networks [18–22].

So far, the analysis of network structure and dynamics has been mostly separated from each other. Only very recently, a first coupling of these issues has become a topic of investigation [23–29]. It is not surprising of the result that the effect of the same mechanism for traffic flow varies from case to case, depending on the underlying network structure. The topological transitions in network from random to scale-free affects the load distribution and the network performance distinctly.

The new insight along the way of optimal network structure for network efficiency promotes a more practical research on complex networks. Investigation of optimal network topology as for effective packet transportation has become an important problem in recent network research, which is a useful guide to design computer or traffic networks since the nature of the optimal network topology [30–32]. Nevertheless, it will be costly or even impossible to change the real network topology arbitrarily. For example, it is out of the question to change the Internet topology into a starlike or homogeneous-isotropic topology so as to optimize search processes [30].

In contrast, it is comparatively easy to adapt the routing protocols in real communication networks. The performance of the communication systems can be upgraded by implementing the more appropriate routing protocols without changing the underlying network structure, which is more realizable in practice [33–35].

In this paper, we analyze the effects of network structures and routing protocols on network performance. Concerning the data-traffic performance, the estimate on the throughput, i.e., the capacity of maximum end-to-end traffic flow the network is able to handle without overloading is given under different network topologies. It reveals that capacity depends on the underlying network structure, which increases as the network becomes more homogeneous. Since the shortest-path routing strategy is not aware of the local traffic of the network, we design and implement a new routing strategy by incorporating local traffic information into the basic shortest path routing policy. We find that the network capacity is greatly enhanced when the new traffic awareness routing strategy is adopted.

The structure of the paper is as follows: The network model is described in Sec. II. In Sec. III, the traffic awareness routing strategy is introduced with the detailed dynamic process of packet transportation. Simulations and explanations are presented in Sec. IV. The conclusion is given in Sec. V.

*Electronic address: chen_zhenyi@sjtu.edu.cn

†Electronic address: xfwang@sjtu.edu.cn

II. NETWORK MODEL

It would be natural to regard the real network as a random graph at first sight. However, real networks show statistical properties that are far from being completely random. A large number of networks, including the World Wide Web, the Internet or the metabolic networks, have typically power law degree distribution with exponents between 2 and 3 [12]. Though some other networks display an exponential tail, the degree distribution deviates significantly from the Poisson distribution expected for a completely random graph. In this paper, we grow three different network topologies from the more heterogeneous one to the completely random graph for comparison.

The completely random graph is constructed according to the well known model proposed by Erdos and Renyi [36]. A network with N labeled vertices is connected by M edges, which are chosen randomly from $C^{M}_{[N(N-1)/2]}$ possible edges. Also, we grow a scale-free model introduced by Barabási and Albert where a fixed number of vertices are added at each time and are linked to the growing graph with a linear attachment probability, which shows the power-law degree distribution $P(k)\sim k^{-\gamma}$ with $\gamma=3$ [10]. A more heterogeneous network is constructed according to the model of Goh [16]. In this model, N vertices are labeled by an integer i $(i=1,\ldots,N)$ and each vertex is assigned the weight $p_i=i^{-\alpha}$, where α is a control parameter in $[0,1)$. The two different vertices i,j are selected with probabilities equal to the normalized weight $p_i/\Sigma_k p_k$ and $p_j/\Sigma_k p_k$, respectively, and add an edge between them unless one exists already. This process is repeated until mN edges are added in the network. The result network shows a degree distribution $P(k)\sim k^{-\gamma}$ with $\gamma=(1+\alpha)/\alpha$. Thus, we can obtain various exponent γ in the range $2<\gamma<\infty$ by varying α in $[0,1)$. This model can be viewed as an extension of the standard ER model with prescribed degree distribution but completely random with respect to all the other features. In our paper, we grow the model with exponent $\gamma=2.1$.

Though the networks we consider have different degree distributions, they are characterized by the same number of available resources, which means the same number of vertices and edges.

III. DYNAMICAL PROCESS AND ROUTING STRATEGIES

Real communication systems or traffic systems usually involve finite queue length and limited process rate of each node. Thus packets will be hampered in the routers' queue while going from origin to destination, causing time delays. At each time step, the probability for node i to generate a packet is λ_i. We assume $\lambda_1=\lambda_2=\cdots=\lambda_N=\lambda$ for simplicity. Also, it is assumed that each node has different capabilities in delivering and handling information packets according to its degree, that is, at each time step, each node can deliver at most k_i packets one step toward their destinations. Once a packet is generated, it is placed at the end of the node queue, which contains the undelivered packets created at current time steps or transmitted from the other nodes. At each time step, the node processes k_i packets in its queue based on the First-In-First-Out rule and selects the next routing node for the packets according to the given routing strategy. This procedure applies to every node in the network at the same time. Once the packet reaches the destination, it is removed from the network.

A basic and widely used routing strategy is the shortest path routing. Each node in the network delivers a packet by forwarding it to one of the node's neighbors that is closest to the destination of the packet. This strategy is simple but has its limitation in that it takes no account of the node state it delivers to. Even if the selected node is overloaded and packets will wait a long time to be processed at that node, this routing policy makes no change.

In order to make the routing policy be aware of the traffic of the network, we design a traffic awareness routing strategy as follows. Let us assume that node s holds a packet that should be delivered to node t. We first compute the weight H_i of a neighbor node i of s. This weight, which can be viewed as the cost of each packet to pass through node i, is defined as

$$H_i=\alpha\frac{E_i}{\sum\limits_{j\in g(i)}E_j}+(1-\alpha)\frac{L_i}{\sum\limits_{j\in g(i)}L_j}\quad 0\leqslant\alpha\leqslant 1, \tag{1}$$

where $g(i)$ is the set of all the neighbor nodes of i, L_i is the shortest path length from node i to target t, $E_i=c_i/k_i$ is the estimated waiting time at node i. The queued packet information c_i is changed dynamically in each time step according to the local traffic dynamics. However, the process rate k_i, which is defined as the maximum number of packets that node i can deliver in each time step, is assigned appropriately according to the node degree and remains unchanged during the simulation. In real communication networks, different nodes certainly have different ability to forward packets and it is reasonable to believe that the highly connected nodes will have high probability to have large process rate k_i. However, it has been assumed in Ref. [33,34] that each node has the same capability of delivering packets. Since the less congested node is likely to be less efficient in processing packets, the packet will have more chance to be trapped in such node if it is delivered only by the knowledge of queued packets number c_i as adopted in Refs. [33,34]. In the special case of $k_i=k_j$ for all $i,j=1,\ldots,N$, our definition reduces to that in Refs. [33,34]. After computing the weight of each neighbor node of node s, we select the next router node among the neighbors which has the minimum weight. If there is more than one node with the minimum weight, we select one of them randomly. The rest of the algorithm remains the same as before, i.e., at each time step, the weight H_i will be calculated dynamically according to the current traffic information in the network and the minimum cost node is selected as the next router.

Given a specific network model, the critical traffic load $\lambda_{\rm crit}$ is defined as the maximum packet creation rate λ, where on average, the flux of newly created packets is equal to the flux of the delivered packets. When $\lambda<\lambda_{\rm crit}$, the network is in the subcritical phase, where the newly created packets is less than or equal to the delivered packets. The number of packets $W(t)$ in the network is balanced, leading to a steady free traffic flow. When $\lambda > \lambda_{crit}$, the number of packets $W(t)$ in the network is increased with time and will lead to traffic congestion. This per-node quantity λ_{crit} is related to the overall-network capacity $T=\lambda_{crit}N$, which is denoted as the end-to-end capacity of the network. It describes the maximum number of end-to-end traffic flows that can be completed per time step without network overloading. Since we fix the network size $N=1000$, we simply use the value of the critical traffic load λ_{crit} to characterize the overall-network capacity T.

With all the generic traffic simulations now at hand, the natural questions are: How to characterize the critical traffic load? How does the network capacity depend on the network topology and how does it vary with the interplay between routing strategy and network structure?

IV. SIMULATIONS AND EXPLANATIONS

We use the order parameter ξ to characterize the transition

$$\xi = \lim_{t\to\infty} \frac{1}{\lambda N} \frac{W(t)}{t}, \tag{2}$$

where $W(t)$ is the total number of packets in the network at time t. In particular, for $\lambda < \lambda_{crit}$, the packets in the network are balanced, leading to a steady free traffic flow. For $\lambda > \lambda_{crit}$, the number of packets grows linearly in time. Figures 1(a), 1(c), and 1(e) plot the total number of active packets as a function of time steps corresponding to three different topologies, i.e., the static scale-free network with exponent $\gamma=2.1$, the BA scale-free network with exponent $\gamma=3$, and the *E-R* random network, respectively. In each figure, three curves are plotted, each representing a different congestion phase that is subcritical, around-critical and supercritical. It can be seen from all the figures that $W(t)$ remains nearly unchanged when $\lambda < \lambda_{crit}$, however, it grows continuously with time t when $\lambda > \lambda_{crit}$. With the increasing of λ above the critical value, the curve of $W(t)$ becomes steeper as time goes on.

Figures 1(b), 1(d), and 1(f) summarize the end-to-end capacity results by the phase transition of the traffic load λ. For each network model, ξ represents an average over an ensemble of 10 network realizations. The simulation time of 5000 time steps for each realization is sufficient for the determination of λ_{crit}. For clarity, we amplify k times of ξ and only use the data of the last 1000 time steps when the network traffic flow can be viewed as stable. Firstly, by adopting the shortest path routing strategy, it is found that the critical traffic load $\lambda_{crit} \approx 0.022$ for the static scale-free network with exponent $\gamma=2.1$. The critical value is slightly increased to $\lambda_{crit} \approx 0.026$ when network structure becomes BA scale-free network with exponent $\gamma=3$. When the network topology evolves to the most homogeneous *E-R* random graph, the critical value is increased to $\lambda_{crit} \approx 0.16$. Though the first two networks can both be regarded as scale-free, they have different degree distribution exponents. There are some hub nodes, each connecting to almost 30% of the whole network vertices, that exist in the static scale-free network with exponent $\gamma=2.1$. However, in the BA scale-free network, the hub nodes connect to only 15% of the whole network vertices that are relatively less than the first one, which reflects the less heterogeneous of the BA scale-free network than the first topology. Thus, it is manifested that the detail variations of the network structure affect the value of λ_{crit}. Due to the intrinsic difference of the random and the scale-free topology, the value of λ_{crit} varies more clearly. This outcome evidently shows that network structure has a prominent influence on the network capacity with respect to data traffic. In particular, random networks seem to have larger network capacity than scale-free networks under the shortest path routing algorithm, which is similar with the results shown in Ref. [20]. Based on the shortest path routing strategy, packets are forwarded to the node along the shortest path with no awareness of the queue length or process rate information of the node. Since the scale-free network has wider distribution of vertex betweenness, which is the measure of the number of shortest path pass through the vertex, the packets will have greater chance to be forwarded to the most connected node, leading to the accumulation of packets in such nodes and to the network congestion. However, in the homogeneous random network, packets will be distributed more dispersedly due to the absence of the hub nodes, which ameliorates the local jam in such node and prevents the packet accumulation in the whole network, leading to the larger network capacity. Thus, it is natural to ask if we could

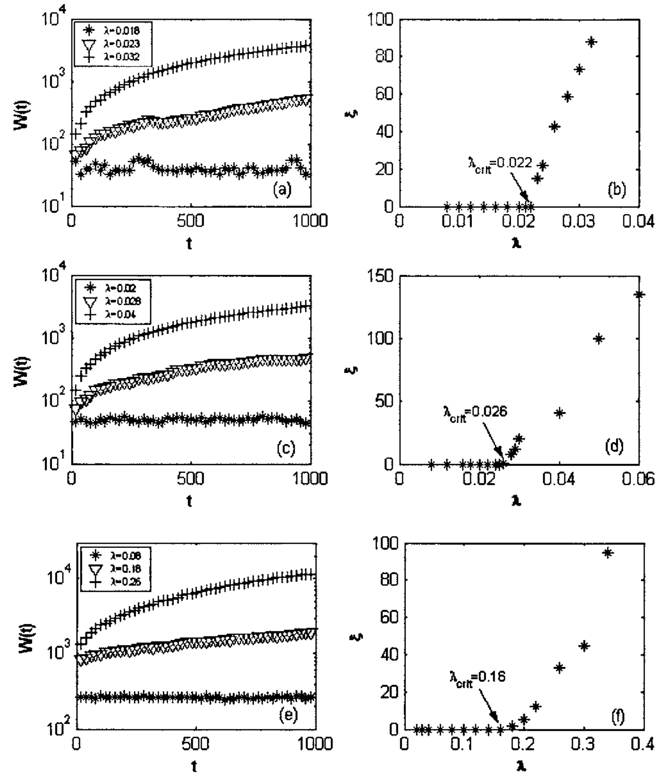

FIG. 1. (a),(c),(e) The total number of active packets as a function of time steps with different packet creation rate. (b),(d),(f) The phase transition of the traffic load λ with the shortest path routing policy. The network model is the static scale-free network with exponent $\gamma=2.1$, the BA scale-free network with exponent $\gamma=3$, and *E-R* random network, respectively, from top to bottom.

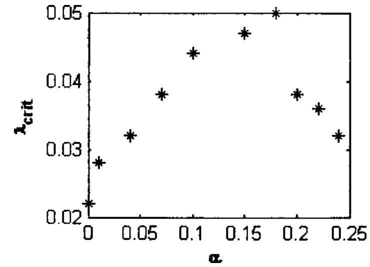

FIG. 2. The value of the critical traffic load λ_{crit} with different α set in the traffic awareness routing policy under the static scale-free network with exponent γ=2.1. α=0 is corresponding to the shortest path routing under the same network structure.

design a more effective routing protocol to argument the network capacity without changing the network structure.

Now we consider the traffic awareness routing strategy introduced in the last section. It is easy to see that when α=0, the traffic awareness routing strategy equals the shortest path routing. Furthermore, by tuning the value of α, the weight of E_i and L_i will vary and the effectiveness of this routing strategy will also change [37]. With each network structure, we adopt a set of values of α in the traffic awareness routing and investigate its effect on the critical traffic load λ_{crit}. From Figs. 2 to 4, the values of the critical traffic load λ_{crit} with different α set in the traffic awareness routing policy in each network structure are plotted. It can be seen that in all the figures, the critical traffic load λ_{crit} is gradually increased until the optimal parameter α is reached, while at the same time λ_{crit} also reaches its maximal. After the optimal α is reached, λ_{crit} is decreasing because of the less efficient of the routing strategy compared with the optimal α being set. In Fig. 2, it can be seen that the critical traffic load λ_{crit} is increased from 0.022 to the maximum 0.05 for the static scale-free network with exponent γ=2.1 under the traffic awareness routing strategy, which is almost three times larger than that of the shortest path routing strategy. For the BA scale-free network, the augment is much more distinct in Fig. 3, which is increased from 0.026 to the maximum 0.16, nearly a six times enhancement. Though the homogeneous nature of the random graph makes it have a larger network capacity under the shortest path routing strategy, its capacity can still be upgraded under the new routing policy, with an increment of λ_{crit} from 0.16 to the maximum 0.39 as shown in Fig. 4. Since the traffic awareness routing strategy achieves better tradeoff between the waiting time and the path length information of the next route node, it will forward the packets to the node that is most efficient in transmission at each time step. Thus, packets will have more chance to be delivered to the destination through the neighbors of the jamming nodes that are usually the most connected ones. By avoiding the packets jamming locally, this routing strategy successfully overcomes some side effects caused by the essential of the network structure and enhances the network capacity for each network topology. In addition, we want to mention that, although it seems that the network capacity under the traffic awareness routing strategy with different parameter α are all larger than it is under the shortest path routing from Figs. 2 to 4, the tendency of the better performance of the traffic awareness routing strategy will not hold when α approaches 1. Since the packets will travel such longer paths that do not compensate the time they would lose waiting on the congested nodes, the decrease of the network capacity is straightforward [37].

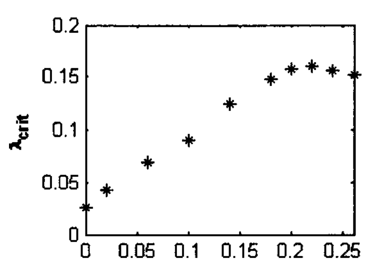

FIG. 3. The value of the critical traffic load λ_{crit} with different α set in the traffic awareness routing policy under the BA scale-free network with exponent γ=3. α=0 is corresponding to the shortest path routing under the same network structure.

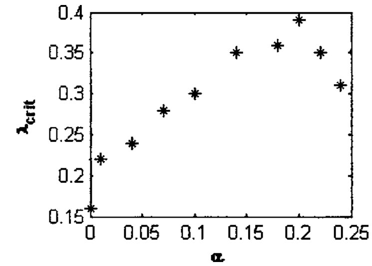

FIG. 4. The value of the critical traffic load λ_{crit} with different α set in the traffic awareness routing policy under the *E-R* random network. α=0 is corresponding to the shortest path routing under the same network structure.

It is also observed that the maximum enhancement of the network capacity, which is computed as the difference of the critical traffic load λ_{crit} between the shortest path routing and the traffic awareness routing with optimal α being set, varies with the network structure. It seems that the BA scale-free network achieves almost two times increment of the capacity compared with the other two topologies. Since when network becomes most heterogeneous, the local jam induced by the hub nodes can be so severe that the packets cannot be redirected to the neighbor nodes efficiently even by the traffic awareness routing strategy, consequently inhibits the further enhancement of the network capacity, which is manifested by the less improvement factor of the static scale-free network. Also, for the most homogeneous random graph, the neighbors of the jamming node are very likely to be jammed too, thus the packets will also get trapped in the neighbors, leading to less efficiency of the routing algorithm compared to the BA scale-free network, which is shown by less enhancement of the network capacity in the same way. The more effectiveness of the traffic awareness routing strategy in BA scale-free network exhibits here is consistent with the results mentioned in [23,24], in which it is pointed out that BA scale-free network ensures efficient transportation process by some local optimization. Nevertheless, we believe that an optimal topology may exist between the most heterogeneous and the most homogeneous network structure, which achieves the largest upgrade of its capacity under the traffic awareness routing strategy.

V. CONCLUSION

In this paper, we estimate the variations of network capacity under two different routing strategies for three different network topologies. It is shown that the capacity of each network depends on the underlying network structure with the maximum capacity achieved in the most homogeneous random graph. By adapting to a traffic awareness routing strategy, the network capacity is greatly improved in each topology compared with the basic shortest path routing policy. Considering the feasibility and the cost of changing the real network topology to enhance the network performance, it is comparatively easy to adjust the routing protocols in real communication systems.

ACKNOWLEDGMENTS

This work is supported by the National Science Fund for Distinguished Young Scholars of the People's Republic of China under Grant No. 60225013 and the National Science Foundation of the People's Republic of China under Grant Nos. 70271072 and 90412004.

[1] T. Ohira and R. Sawatari, Phys. Rev. E **58**, 193 (1998).
[2] R. V. Solé and S. Valverde, Physica A **289**, 595 (2001).
[3] A. Arenas, A. Díaz-Guilera, and R. Guimerà, Phys. Rev. Lett. **86**, 3196 (2001).
[4] R. Guimerà, A. Arenas, and A. Díaz-Guilera, Physica A **299**, 247 (2001).
[5] R. Guimerà, A. Arenas, A. Díaz-Guilera, and F. Giralt, Phys. Rev. E **66**, 026704 (2002).
[6] S. Valverde and R. V. Solé, Physica A **312**, 636 (2002).
[7] T. M. Janaki and N. Gupte, Phys. Rev. E **67**, 021503 (2003).
[8] D. K. Arrowsmith, R. J. Mondragn, J. M. Pittsy, and M. Woolf, Report 08, Institut Mittag-Leffler, Stockholm (2004).
[9] D. Watts and S. Strogatz, Nature (London) **393**, 440 (1998).
[10] A. L. Barabási and R. Albert, Science **286**, 509 (1999).
[11] S. Strogatz, Nature (London) **410**, 268 (2001).
[12] R. Albert and A. L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[13] M. E. J. Newman, SIAM Rev. **45**, 67 (2003).
[14] X. F. Wang and G. R. Chen, IEEE Trans. Circuits Syst., I: Fundam. Theory Appl. **3**, 6 (2003).
[15] W. Aiello, F. Chung, and L. Lu, in *Proceedings of the 32th Annual ACM Symposium on Theory of Computing* (2000), pp. 171–180.
[16] K. I. Goh, B. Kahng, and D. Kim, Phys. Rev. Lett. **87**, 278701 (2001).
[17] P. Holme and B. J. Kim, Phys. Rev. E **65**, 026107 (2002).
[18] L. A. Adamic, R. M. Lukose, A. R. Puniyani, and B. A. Huberman, Phys. Rev. E **64**, 046135 (2001).
[19] L. A. Adamic, R. M. Lukose, and B. A. Huberman, in *Handbook of Graphs and Networks* (Wiley-VCH, Berlin, 2003).
[20] A. Arenas, A. Cabrales, A. Díaz-Guilera, R. Guimerà, and F. Vega-Redondo, Lect. Notes Phys., Statistical mechanics of complex networks, XVIII Sitges Conference on Statistical Mechanics (Springer, Sitges, 2003), pp. 175–194 (2003).
[21] A. Arenas, L. Danon, Díaz-Guilera, and R. Guimerà, Lect. Notes Comput. Sci. **3038**, 1078 (2004).
[22] M. Rosvall, A. Gronlund, and P. Minnhagen, e-print cond-mat/0505400 (2005).
[23] Z. Toroczkai and K. E. Bassler, Nature (London) **428**, 716 (2004).
[24] Z. Toroczkai, B. Kozma, K. E. Bassler, N. W. Hengartner, and G. Korniss, e-print cond-mat/0408262.
[25] L. Zhao, Y. C. Lai, K. Park, and N. Ye, Phys. Rev. E **71**, 026125 (2005).
[26] K. Park, Y. C. Lai, L. Zhao, and N. Ye, Phys. Rev. E **71**, 065105(R) (2005).
[27] D. K. Arrowsmith, M. di Bernardo, and F. Sorentino, in IEEE International Symposium on Circuits and Systems, Kobe, Japan (2005).
[28] W. Krause, I. Glauche, R. Sollacher, and M. Greiner, Physica A **338**, 633 (2004).
[29] I. Glauche, W. Krause, R. Sollacher, and M. Greiner, Physica A **341**, 677 (2004).
[30] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. **89**, 248701 (2002).
[31] J. Ohkubo and T. Horiguchi, Physica A **353**, 649 (2005).
[32] W. Krause, M. Greiner, and M. Greiner, e-print cs.NI/0503010 (2005).
[33] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E **70**, 056105 (2004).
[34] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, e-print cond-mat/0412053 (2004).
[35] G. Yan, T. Zhou, B. Hu, Z. Q. Fu, and B. H. Wang, e-print cond-mat/0505366 (2005).
[36] P. Erdos and A. Renyi, Publ. Math. Inst. Hung. Acad. **5**, 17 (1960).
[37] Z. Y. Chen and X. F. Wang, Physica A (to be published).