Structural bottlenecks for communication in networks

Sameet Sreenivasan,[1,2] Reuven Cohen,[3] Eduardo López,[2] Zoltán Toroczkai,[4,2] and H. Eugene Stanley[1]
[1]*Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA*
[2]*Center for Nonlinear Studies, Theoretical Division, Los Alamos National Laboratory, MS B258, Los Alamos, New Mexico 87545, USA*
[3]*Laboratory of Networking and Information Systems and Department of Electrical and Computer Engineering, Boston University, Boston, Massachusetts 02215, USA*
[4]*Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA*
(Received 4 April 2006; published 5 March 2007)

We consider the effect of network topology on the optimality of packet routing which is quantified by γ_c, the rate of packet insertion beyond which congestion and queue growth occurs. We show that for any network, there exists an absolute upper bound, expressed in terms of vertex separators, for the scaling of γ_c with network size N, irrespective of the static routing protocol used. We then derive an estimate to this upper bound for scale-free networks and introduce a static routing protocol, the "hub avoidance protocol," which, for large packet insertion rates, is superior to the shortest path routing protocol.

DOI: 10.1103/PhysRevE.75.036105 PACS number(s): 89.75.Hc, 89.20.Hh, 89.75.Da

The advent of the Internet has brought with it the possibility of exponentially increasing information transfer through its infrastructure. In the current paradigm of packet-switched communication, there can be delays caused in packet delivery due to device latency—i.e., the time taken to process and forward a single packet by a router. If more packets arrive at nodes (routers) than they are able to process per unit time, queues will build up on the nodes, leading to delays in packet delivery. Motivated by these facts, we address two questions of broad significance for communication networks. (i) How can one characterize a packet switched communication network's ultimate carrying capacity? (ii) What routing algorithms will achieve this ultimate capacity?

In this paper we demonstrate the existence of an upper bound γ_T for the congestion threshold γ_c [1] determined solely by the network topology; γ_c is the packet insertion rate above which queuing and congestion appears in the network. It has been argued that the degree distribution of the internet is a power law or "scale free" [2–5]. Moreover, the scale-free network topology has been proposed as a suitable candidate for the structure of sensor networks [6,7]. Motivated by these reasons, we focus on scale-free networks. We will restrict our discussion to the configuration model (CM) [8], which is one of the simplest models to generate a scale-free network. While this model may not capture all structural features of the Internet [9], it nevertheless provides a suitable starting point for the demonstration of our approach which is applicable to arbitrary graph structures. Without loss of generality, we assume all routers to have a latency of unity. We also assume that routers have infinite storage capacity.

Denote by G the physical substrate graph (network) for communication, which we assume to be singly connected. Once a packet entering node s reaches its destination node d, it disappears from the system. The sequence of nodes and edges that the packet visits constitutes the route for that source-destination pair. For a network of size N, the routing problem consists of finding an assignment of routes for *all* $N(N-1)/2$ pairs of nodes. We shall call such an assignment set a "static routing protocol" (SRP). We consider a model [1,10,11] where packet transmission is described by a discrete time-parallel update algorithm. At time t and at every node, a packet enters with probability $0 \leq \gamma \leq 1$. The packet has a destination, chosen uniformly at random from the remaining $N-1$ nodes. Every node i maintains a set of all packets that were sent to it by its neighbors in the previous step, eliminates from this set the packets whose destination is i, and adds to the set the freshly injected packet, created with probability γ. The first packet in the queue is then sent to a neighbor on G following the SRP. Above the congestion threshold γ_c of packet creation there is an onset of congestion, when packets start accumulating on the network [10]. In Fig. 3(a), below, we show the rate of steady-state packet-growth [10], $\theta(\gamma) \equiv \lim_{t\to\infty}[n(t+\Delta t)-n(t)]/(N\gamma\Delta t)$, as function of γ for both the shortest-path protocol (SPP) and the protocol proposed in this paper. Here $n(t)$ is the number of packets on the entire network at time t.

We define the betweenness b for a node as the number of SRP routes passing through that node. The largest of the N betweenness values resulting from the SRP is the "maximal node betweenness" B. The threshold γ_c can be expressed in terms of B for a given SRP [11]. For a given SRP route between a source s and destination d the average packet current incurred from the source at s is $\gamma/(N-1)$. For a node with betweenness b, the average packet inflow current will be given by $b\gamma/(N-1)$. Since the outflow of packets occurs at unit latency, we will have queuing and congestion at the node for which $\gamma/(N-1)$ reaches unity for the first time: namely, at the node with $b=B$. Thus

$$\gamma_c = \frac{N-1}{B}. \tag{1}$$

For the SPP [1,10], the node betweenness becomes identical to the shortest-path betweenness B_{SPP} [12]. From Eq. (1) it follows that for a given routing protocol, the dependence of the congestion threshold γ_c on N is determined by the scaling with N of B. Therefore, the best routing protocol from the point of view of router congestion avoidance would be one

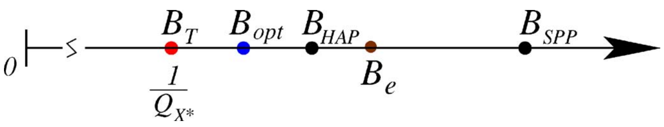

FIG. 1. (Color online) The relative sizes of the betweenness values introduced in the text.

for which B exhibits the *slowest* growth with N. Although there have been prescribed *ad hoc* "adaptive routing protocols" [13,14] that increase γ_c, the issue of finding a bound on γ_c has not been systematically addressed.

Next, we show that there is a lower bound $B_T \le B$ (and thus $\gamma_c \le \gamma_T$) induced only by the *topology* of the network G and it is independent of the routing protocol used; i.e., no SRP can give rise to a congestion threshold greater than γ_T. Thus, by dictating an upper bound on the congestion threshold, B_T quantifies the intrinsic bottleneck present in the graph G. Among the set $\mathcal{P}$ of all possible SRP's, let B_{opt} be the smallest maximal betweenness value: namely, $B_{opt} = \min_{SRP \in \mathcal{P}} B^{SRP}$, so $B_T \le B_{opt}$ (Fig. 1). It is an open question whether the topological bound can be achieved by a routing protocol. Similar considerations have been made in the context of edge betweenness in Refs. [15,16]. Here we focus on the scaling of the bound B_T as a function of N.

We next discuss B_T using graph partitioning arguments. Given an arbitrary network G, partition the set of all nodes V into three nonempty sets denoted A, X, and A'. Since G is singly connected, there will be edges running between at least two pairs of the three sets. Unless G is the complete graph, we can choose X such that there are no edges running directly between A and A', in which case X is called a *vertex separator*. For any SRP we must designate a route for all pairs of nodes, therefore also for those pairs for which one node is in A and the other in A'. Since X is a separator set, all routes from A to A' must go through the nodes in X. Therefore, there are at least $|A||A'|$ routes passing through X for any SRP. Since the maximum is always larger or equal than the average, the maximum betweenness incurred on the nodes in X can be no less than $(|A||A'|)/|X|$. We define the *sparsity* [17] of the separator X the quantity $Q_X \equiv |X|/(|A||A'|)$. Thus, associated with every vertex separator X there is a quantity $B_X = 1/Q_X$ providing a lower bound to the maximal betweenness on nodes in X. Denote by $\mathcal{M}$ the set of all possible vertex separators in G. If we systematically consider all possible choices of vertex separators $X \in \mathcal{M}$, we can find (at least) one separator X^* for which $B_X = 1/Q_X$ achieves its maximal value defined as B_T. Thus, the *topology* of the graph constrains the maximal betweenness to be no less than B_T, and for arbitrary routing, $B \ge B_T = 1/Q_{X^*} = 1/\min_{X \in \mathcal{M}} Q_X$.

Finding minimal sparsity vertex separators is an NP-hard problem [18], and we shall not deal with it here. Due to the analytical and the computational difficulty in determining B_T, we focus on obtaining an *analytical estimate* B_e to B_T and derive its scaling with N for random, uncorrelated, scale-free networks. While possibly being greater than the true topological bound B_T, this estimate B_e nevertheless provides a comparative value dependent only on the network topology and allows us to quantify the performance of the SPP.

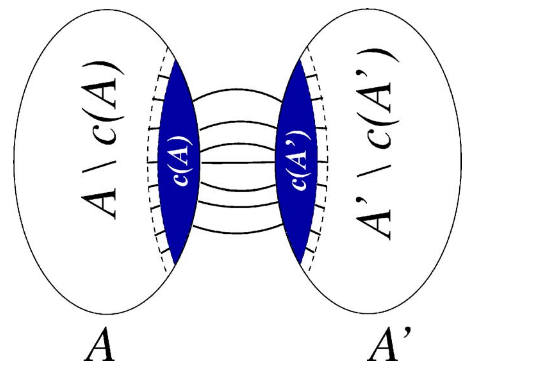

FIG. 2. (Color online) Bipartitioning G into A and A' such as to obtain two vertex separators $c(A)$ and $c(A')$; see text.

We start by systematically considering every possible vertex separator in the graph as follows. First, bipartition the graph as shown in Fig. 2 into sets A and A' with $|A| \le |A'|$. Let $c(A)$ be the subset of nodes in A which are adjacent to at least one node in A' and let $c(A')$ be the subset of nodes in A' which are adjacent to at least one node in A. We can now obtain a vertex separator $c(A)$ which separates sets $A \setminus c(A)$ [19] and A' or, similarly, a vertex separator $c(A')$ which separates sets $A' \setminus c(A')$ and A. Thus, going through all possible bipartitions of the graph with $|A| \le N/2$ ensures that we have considered all possible vertex separators of the graph.

If $c(A)$ is chosen as the separator, then the sparsity is $Q_{c(A)} = |c(A)|/[|A - c(A)||A'|] \ge |c(A)|/(|A||A'|)$. We obtain a similar expression for $Q_{c(A')}$ if $c(A')$ is chosen as the vertex separator. Therefore

$$Q_{c(A)} \ge \frac{1}{|A'|}\frac{|c(A)|}{|A|}, \quad Q_{c(A')} \ge \frac{1}{|A'|}\frac{|c(A')|}{|A|}. \tag{2}$$

Since $|A| \le N/2$, $|A'| \equiv O(N)$ and a lower bound for the sparsity Q_{X^*} is determined by

$$Q_{X^*} \ge \frac{1}{O(N)} \min_{A \subset V, A \le N/2} \left\{ \min\left(\frac{|c(A)|}{|A|}, \frac{|c(A')|}{|A|} \right) \right\}. \tag{3}$$

Next we use the notion of *edge expansion*, χ_e, defined below. For a bipartition of the graph G into sets A and A', denote the number of edges simultaneously adjacent to a node in A and A' as $c_e(A,A')$. Then

$$\chi_e = \min_{A \subset V, A \le N/2} \frac{|c_e(A,A')|}{|A|}, \tag{4}$$

and an *edge expander* graph has $\chi_e = O(1)$. Next consider a bipartition of the graph into A and A', and let $|A| = cN^\alpha$ where c is a constant and $0 < \alpha \le 1$. From the edge expansion property of scale-free graphs with minimum degree $k_{min} \ge 3$ [15], the number of cut edges between A and A' is at least $\chi_e c N^\alpha = O(N^\alpha)$. We can bound from below both $|c(A)|$ and $|c(A')|$, as needed by (3), by the minimal size m of the set of nodes that can contribute $\chi_e c N^\alpha$ cut edges. The size m is obtained by taking all nodes with degree higher than $\hat{k}$, such that $N \int_{\hat{k}}^{\infty} k P(k) \mathrm{d}k = \chi_e c N^\alpha$, where $P(k) = Ak^{-\lambda}$ is the degree distribution of the graph. This yields $\hat{k} \sim N^{(1-\alpha)/(\lambda-2)}$. Therefore the minimal size of the set of nodes that can contribute $\chi_e c N^{\alpha}$ edges is $m = N\int_{\hat{k}}^{\infty} P(k)dk \sim NN^{(1-\lambda)(1-\alpha)/(\lambda-2)}$ and, therefore,

$$|c(A)|,|c(A')| \geq m = O(NN^{(1-\lambda)(1-\alpha)/(\lambda-2)}). \quad (5)$$

The quantity m is bounded below by $O(1)$. For a given λ we see that when $\alpha=1$ or in other words A and A' in the bipartition are both $O(N)$, we get $m \equiv O(N)$. For all other values of α, we get $m < O(N)$. As α decreases from 1, m also decreases until it becomes $O(1)$ and this occurs for the first time when $\alpha = 1/(\lambda-1)$. Thus, from (5) and (3) we get $Q_{X^*} \geq O(N^{-\lambda/(\lambda-1)})$ and so

$$B_T \leq B_e \equiv O(N^{\lambda/(\lambda-1)}). \quad (6)$$

From (6) we see that when $\lambda \to 2$, we obtain the worst possible scaling of $B_e = O(N^2)$, which can be understood from the fact that the graph becomes increasingly star like, and for such a graph the central node trivially has $B = O(N^2)$ [20]. On the other hand, when $\lambda \to \infty$, $B_e \to O(N)$. In this case the graph approaches a random regular graph and random regular graphs are good *vertex expanders* [21]. This implies that for any bipartition into A and A', there exists a constant μ such that $|c(A')| \geq \mu |A|$. Thus $|c(A)| \geq (\mu |A|)(1+\mu)$, so $|c(A)|$ and $|c(A')|$ are linear in $|A|$ and hence $B_e = O(N)$.

When $2 < \lambda < 3$, for the networks generated by the configuration model to be uncorrelated requires that the maximum degree in the network $K_{max} \sim N^{1/2}$ [22]. Incorporating this upper cutoff in the arguments made above, we obtain $Q_{X^*} \geq O(N^{3/2})$ and hence $B_T \leq B_e \equiv O(N^{3/2})$ [same as for $\lambda = 3$ in (6)]. From Fig. 3(b), we see that $B_{SPP} \sim N^{1.80}$. This is much worse than the scaling of B_e and therefore suggests that an SRP for which the maximal betweenness scales like B_e would provide better performance from the point of view of congestion than the SPP. The question arises whether B_e can be achieved by *any* static routing protocol. We answer this question affirmatively by presenting next an SRP for which the scaling of the maximal betweenness surpasses that of B_e and is therefore significantly better than the scaling of B_{SPP}.

While B_e imposes an upper bound on B_T, a nontrivial lower bound can be imposed on B_T using the following consideration: Assume we take as a vertex separator all nodes of degree $k > K^*$ for some K^*. The total number of links emanating from the nodes of this separator is $N_\ell = N\int_{K^*}^{\infty} ckk^{-\lambda}dk = O(N(K^*)^{2-\lambda})$. The total number of nodes in this separator is $n = N\int_{K^*}^{\infty} ck^{-\lambda}dk = O(N(K^*)^{1-\lambda})$. The probability that a node of degree k_{min} has all its neighbors in the vertex separator is order $(N_\ell/N)^{k_{min}}$, decaying exponentially with k. Since there are $O(N)$ nodes of degree k_{min}, at least $O(N(N_\ell/N)^{k_{min}})$nodes must communicate through the separator. This yields

$$B_T > O\left(\frac{NN(N_\ell/N)^{k_{min}}}{n}\right) \equiv O(N(K^*)^{(2-\lambda)k_{min}-(1-\lambda)}). \quad (7)$$

For $k_{min}=3$ and $\lambda < 2.5$ this gives a nontrivial lower bound on B_T.

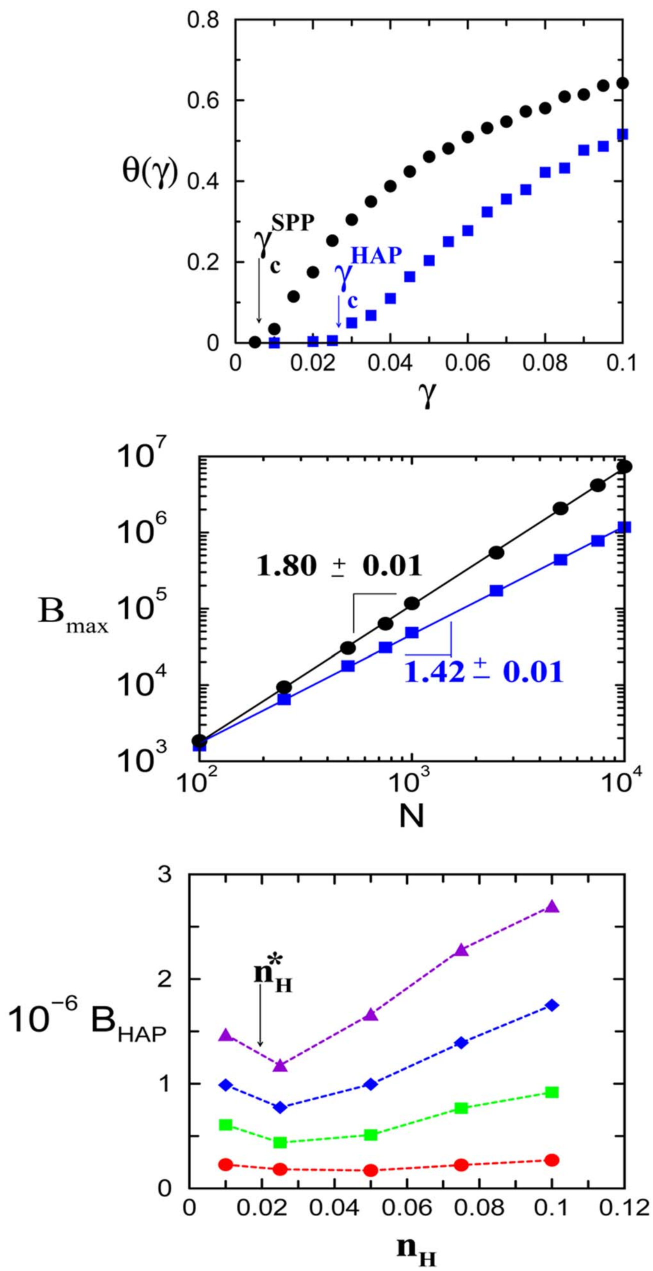

FIG. 3. (Color online) (a) Performance of SPP (solid circles) and hub avoidance HAP (squares) on a scale-free graph of size $N=10^3$ and $\lambda=2.5$. The number of hubs removed is $n_H = n_H^* = 0.05N$, which in our simulations yields the minimal value of B_{HAP} for $N=1000$. The congestion threshold γ_c beyond which packet growth occurs $[\theta(\gamma) > 0]$ is higher for the HAP as compared to the SPP. (b) Log-log plot of the scaling of the maximum betweenness versus system size N for the two protocols. B_{HAP} has scaling exponent of 1.42 ± 0.01, consistent with the worst case estimate for the topological bound on the maximal betweenness, $B_e \sim N^{3/2}$. For each value of N, we use the minimal value of B_{HAP} obtained in simulations by varying n_H. B_{SPP} grows much faster, $B_{SPP} \sim N^{1.80 \pm 0.01}$. The errors in the slopes are the standard errors in the regression coefficient. (c) Nonmonotonic behavior of B_{HAP} as the number of hubs removed in the HAP, n_H, is varied. We show results for $N=2500$, 5000, 7500, and 10000 from bottom to top, respectively. We see that B_{HAP} achieves a minimal value at a certain value of $n_H = n_H^*$.

The performance of the SPP can be approximated by considering the role of the high degree nodes with $k=O(\sqrt{N})$. Since a sharp cutoff is induced on the degrees, the number of such nodes is $N\int_{\sqrt{N}}^{\infty}ck^{-\lambda}dk=O(N^{(3-\lambda)/2})$. These nodes are connected in an approximate clique with a finite fraction of the shortest paths going through them (since the probability of connecting to one of these high degree nodes is larger than to any other node and the path between them is short), and due to their similar degree, these nodes may be assumed to have similar betweenness values. Thus, the estimate on the betweenness of these nodes is $B_{SPP}\approx O(N^2/N^{(3-\lambda)/2})=O(N^{(\lambda+1)/2})$. For the case $\lambda=2.5$, from Fig. 3(b), this yields $B_{SPP}\approx O(N^{1.75})$ in agreement with the measured 1.8. Note that if a packet can be split to access multiple shortest paths between source and destination, the definition of betweenness and hence the scaling of B_{SPP} is altered [23].

Our derivation of B_e suggests that the sparsity is smallest when obtained from a bipartition where the smaller set is of size of the order of the maximal degree. This suggests that, topologically, the betweenness for hubs is high and using the SPP increases this betweenness since shorter paths tend to use hubs. Moreover, using the SPP leaves a large number of alternate paths unused for routing. Exploiting these observations, we obtain a novel SRP, which we call the *hub avoidance protocol* (HAP), as follows: (i) Remove a number n_H of the highest degree nodes. The network could now consist of several disconnected clusters. (ii) In every disconnected cluster, assign a routing path for every pair of nodes using SPP. (iii) Place back the removed nodes with their edges. (iv) For every pair of nodes which have not been assigned a routing path in step (i), assign one using the SPP. For our simulations [Fig. 3(b)] we have chosen for a given N that value of n_H which gives the optimal performance—i.e., the lowest value for the maximal betweenness B^{HAP}. In general, varying the value of n_H gives rise to a nonmonotonic behavior, with a minimum at a certain value of $n_H=n_H^*(N)$ [Fig. 3(c)]. Thus, the HAP is an SRP for which the scaling of the maximal betweenness not only achieves, but surpasses the scaling of the topological estimate B_e, and therefore is a significant improvement over the SPP. This improvement comes from utilizing available alternate paths which, while not significantly longer than the shortest path, considerably reduces the betweenness of the hubs. Figure 3(a) shows the improvement in performance of our HAP as reflected by the increase in the value of the γ_c and the decrease in the number of accumulating packets at a given packet creation rate γ compared to the SPP.

We argue for the attainability of the scaling $B_e\sim N^{3/2}$ using the HAP. Assume that the HAP proceeds by initially removing nodes with degree $k>K^*$ and connecting remaining nodes using the SPP. Then, gradually start restoring nodes of degree $K>K^*$. At a given stage j, assume that we restore all nodes with $k=K_j$. For all j, we choose K_j sufficiently large so that the network is still above the percolation threshold and therefore has a connected component of $O(N)$. The network at stage j can therefore be considered to consist of three sets of nodes: (i) the set X of newly restored nodes with $k=K_j$, (ii) the set A' which forms the connected component of $O(N)$, and (iii) the set A of nodes that are detached from A' by the removal of X. Moreover, the size of the set A can be no greater than $O(|X|K_j)$. Hence, nodes in X can connect $O(|X|K_j)$ nodes in A with the $O(N)$ nodes in A'. Assuming that nodes of the set X share equally the betweenness arising from paths between A and A' (which is reasonable since the degrees of the constituent nodes are the same), the betweenness on each node in set X is $[O(|X|K_jN)]/[O(|X|)]=O(NK_j)$. Since K_j can be at most $O(\sqrt{N})$, it follows that the maximal betweenness obtained by the HAP is at most $O(N^{3/2})$.

In summary, we identified a bound to communication arising purely from the network topology and we used this bound to show that there exist better SRPs than the SPP for routing on scale-free networks.

S.S. and H.E.S. thank the ONR, E.L. and Z.T. thank DOE Contract No. W-7405-ENG-36, and R.C. thanks ISF and DysoNet.

[1] H. Fuks and A. T. Lawniczak, Math. Comput. Simul. **51**, 101 (1999).
[2] H. Tangmunarunkit *et al.* (unpublished).
[3] A. Vazquez *et al.*, e-print cond-mat/0206084.
[4] http://www.caida.org/tools/measurement/skitter/router_topology
[5] http://www.netdimes.org
[6] S. Dixit *et al.*, IEEE Commun. Mag. **43**, 86 (2005).
[7] F. Saffre *et al.*, Bell Syst. Tech. J. **22**, 200 (2004).
[8] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[9] S. Carmi *et al.*, e-print cond-mat/0601240.
[10] A. Arenas *et al.*, Phys. Rev. Lett. **86**, 3196 (2001).
[11] R. Guimera *et al.*, Phys. Rev. Lett. **89**, 248701 (2002).
[12] L. C. Freeman, Sociometry **40**, 35 (1977).
[13] P. Echenique *et al.*, Phys. Rev. E **70**, 056105 (2004).
[14] G. Yan *et al.*, Phys. Rev. E **73**, 046108 (2006).
[15] C. Gkantsidis *et al.* (unpublished).
[16] A. Akella *et al.*, *ACM Principles of Distributed Computing* (ACM, New York, 2003).
[17] V. Vazirani, *Approximation Algorithms* (Springer-Verlag, Berlin, 2001).
[18] T. N. Bui and C. Jones, Inf. Process. Lett. **42**, 153 (1992).
[19] $A\backslash c(A)$ denotes the set obtained by the removal of vertices in $c(A)$ from the set A.
[20] N. Alon, Lect. Notes Comput. Sci. **1380**, 206 (1998).
[21] G. Davidoff *et al.*, *Elementary Number Theory, Group Theory and Ramanujan Graphs* (Cambridge University Press, Cambridge, England, 2003).
[22] M. Boguña *et al.*, Eur. Phys. J. B **38**, 205 (2004).
[23] B. Danila *et al.*, e-print cond-mat/0607017.