Optimizing controllability of complex networks by minimum structural perturbations

Wen-Xu Wang, Xuan Ni, Ying-Cheng Lai, and Celso Grebogi

Phys. Rev. E **85**, 026115 — Published 22 February 2012

DOI: 10.1103/PhysRevE.85.026115

Optimizing controllability of complex networks by minimum structural perturbations

Wen-Xu Wang,[1,2] Xuan Ni,[2] Ying-Cheng Lai,[2,3] and Celso Grebogi[3]

[1]Department of Systems Science, School of Management and Center for Complexity Research, Beijing Normal University, Beijing 100875, China
[2]School of Electrical, Computer, and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA
[3]Institute for Complex Systems and Mathematical Biology, King's College, University of Aberdeen, Aberdeen AB24 3UE, UK

(Dated: January 19, 2012)

To drive a large, complex, networked dynamical system toward some desired state using as few external signals as possible is a fundamental issue in the emerging field of *controlling complex networks*. Optimal control is referred to the situation where such a network can be fully controlled using only one driving signal. We propose a general approach to optimizing the controllability of complex networks by judiciously perturbing the network structure. The principle of our perturbation method is validated theoretically and demonstrated numerically for homogeneous and heterogeneous random networks and for different types of real networks as well. The applicability of our method is discussed in terms of the relative costs of establishing links and imposing external controllers. Besides the practical usage of our approach, its implementation elucidates, interestingly, the intricate relationship between certain structural properties of the network and its controllability.

PACS numbers: 05.45.-a,89.75.Hc,05.45.Xt,02.30.Yy

The ability to control complex networks is utter-mostly important to many critical problems in science, engineering and medicine, and has the potential to generate great technological breakthroughs as well. Indeed, because of the ubiquity of complex networks in natural, technological, social, and economical systems, it is highly desirable to be able to apply proper control to guide the network dynamics toward states with the best performance and, at the same time, to avoid undesired or deleterious states. While actual control of complex networks has not been achieved at the present, a necessary stepping stone is to understand the *controllability* of complex networks, which has become a topic of active pursuit [1–8]. Specifically, given a complex-networked dynamical system, one wishes to assess whether it would be possible to apply certain number of control signals at an arbitrary set of nodes so as to drive the system toward some desirable state. The number of control signals, N_D, is thus a key quantity of interest [6–8] as, qualitatively, it characterizes the cost to bring the system under control. If N_D is the same as N (the total number of nodes in the network) so that each node receives one control signal, the likelihood to achieve control will be high but the associated cost will be high, too. To search for ways to reduce N_D thus becomes an issue of significant practical interest in network control, and one naturally asks whether a networked system can be harnessed by using only one control signal. This can indeed be achieved for specific network configurations [6–8].

In this paper, we ask the following question: given an arbitrary network that requires a certain number of signals to be controlled, can one slightly perturb the network so as to achieve the optimal controllability characterized by $N_D = 1$? The theoretical framework under which this question may be addressed is the minimum-input theory developed recently [6] to characterize the controllability of networks with linear dynamics, which is based on the classical control and graph theories [9–11]. The basic goal of the minimum-input theory is to determine the minimum number of nodes to be driven externally to bring the whole network under control. According to this theory [6], only topological changes can alter the network controllability. To be illustrative, we shall explore structural perturbation via adding links to the network to enhance its controllability. It is practically important to develop a paradigm which minimizes the number of added links to achieve $N_D = 1$; for otherwise optimal controllability can be achieved trivially by keeping adding links to the network until it becomes fully connected, which according to the minimum-input theory is fully controllable with a single input. Guided by this general consideration, we shall articulate a strategy to perturb the network by providing a minimum number of additional links at suitable locations determined by certain criterion (to be discussed below). The performance of our perturbation scheme will be compared with that in the case where links are randomly added to the network. Our optimization strategy bridges the network topology and controllability by providing useful insights into the effect of the former on the latter.

To motivate our structural perturbation strategy to optimize network controllability, we briefly describe the minimum-input theory. According to Kalman's controllability rank condition [9, 11], a canonical, linear, and time-invariant dynamical system $\dot{\mathbf{x}}(t) = \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t)$ can be controlled from any initial state to any desired state in finite time, if and only if the $N \times NM$ controllability matrix $\mathbf{C}$ has full rank, i.e.,

$$\text{rank}(\mathbf{C}) \equiv \text{rank}\big[\mathbf{B}, \mathbf{AB}, \mathbf{A}^2\mathbf{B}, \cdots, \mathbf{A}^{N-1}\mathbf{B}\big] = N \qquad (1)$$

where $\mathbf{x} \in R^N$, $\mathbf{B}$ is the $N \times M$ input matrix, M is the number of driver nodes, and $\mathbf{u}(t)$ is the time-dependent input control vector. As pointed out in an earlier work [10], the full-rank condition (1) is appropriate for characterizing the controllability of network systems if A is the transpose of the adjacency matrix and N is the number of nodes. Of particular importance to our perturbation strategy is the concept of structural controllability [6], which can be used to identify the minimum number N_D of driver nodes required for the system to satisfy the full-rank condition (1). However, it is practically difficult to check this condition for large complex networks, as the number of input combinations grows exponentially with the number of nodes ($\sim 2^N$). To overcome this difficulty, Liu et al. [6] proposed the concept of *maximum-matching set* to assess and quantify structural controllability. A particularly useful result is $N_D = 1$ if the network is perfectly matched; otherwise $N_D = N - N_M$, where N_M is the size of the maximum-matching set, i.e., the maximum set of links that do not share starting or ending nodes (Details of maximum matching can be seen in Supplemental Materials). As demonstrated [6], many real-world networks are far from being perfectly matched. Consequently, in order to fully control such a network, a large number of input signals applied to an equally large number of nodes are necessary, which motivates us to ask whether optimal control $N_D = 1$ is achievable by making deliberate, small structural perturbations to the network. In the following, we shall detail our strategy and demonstrate that, given *any* network, a minimum number of links can indeed be added so that all nodes except one are matched. That is, under only one input control signal the perturbed network will meet the full-rank condition.

To be concrete, we shall formulate our strategy to optimize network controllability by adding minimum number of additional links for both directional and bidirectional networks. To explain our strategy, we introduce the concept of "matching path," a subset of links in the set of maximum matching (or "isolated" nodes), which can be (i) starting from an unmatched node and ending at a matched node without outgoing link belonging to the set of maximum matching, (ii) starting from an arbitrary node in a directed loop and ending at the "superior" node that points at the starting node, or (iii) an "isolated" node without any link belonging to the set of maximum matching. Here, case (ii) defines a "close matching path."

If one controller can control multiple drivers simultaneously, our optimization process involves three steps: (1) finding the minimum number of independent matching paths, except close matching paths (details of finding independent matching paths based on the maximum matching algorithm can be found in Supplemental Material [12]); (2) randomly ordering all found matching paths; (3) linking the ending points of each matching path to the starting nodes of the matching paths next to it in order, as illustrated in Fig. 1. The minimum number of independent matching paths, except close matching paths, is equal to one less than the number N_D of unmatched nodes. Applying such structural perturbations, the maximum fraction m_{max} of added links (m is the ratio of the number of added links to the number N_l of links in the original network) to achieve $N_D = 1$ is

$$m_{max} = \frac{N_D - 1}{N_l}. \quad (2)$$

The network is fully controllable with a single controller imposed at the starting node of the first matching path and any one node in each of other close matching pathes simultaneously. We can prove, according to Lin's structural controllability theory [10], that the optimal network resulted from the above structural perturbations satisfies the full-rank condition with a single input (see Supplemental Material [12]). The value of N_D can always be reduced to 1 by adding a minimum number of links.

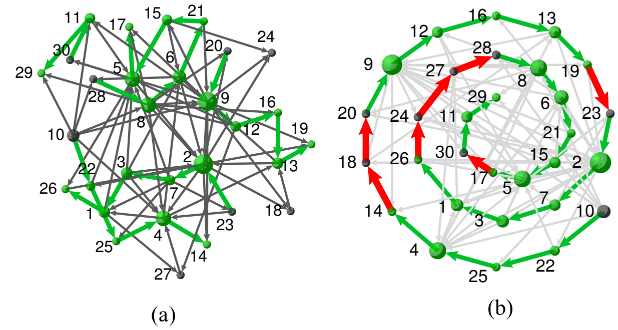

FIG. 1: (Color online.) (a) A network of 30 nodes with heterogeneous degree distribution, generated according to the preferential attachment mechanism [14]. (b) All matching paths in order, starting from node 10 outside and ending at node 29 inside. The links of the set of maximum matching and the matched nodes are marked by green (gray). Structural perturbations are represented by the added links connecting the tail of a matching path in higher order to the head (black color) of matching path in lower order, which are marked by red (dark gray). Other links are marked by light gray. The configuration of added links is not unique, but their minimum number is.

To numerically demonstrate our perturbation strategy, we use the Erdős-Rényi (ER) random [13] and scale-free (SF) networks [14] and calculate n_D, the density of unmatched nodes for the two cases: (1) adding optimal links determined by our optimization strategy and (2) adding random links. As shown in Fig. 2, our method leads to a much faster reduction in n_D toward the minimal value $1/N$ than merely adding random links for any average node degrees of the network. Moreover, our strategy requires only a minimum number of additional links to make the network fully controllable under only one external controller.

The minimality of the number of additional links can be justified in terms of the minimum input theory and the definition of maximum matching. According to the minimum input theory, a node can be fully controlled either it has an independent 'superior' node pointing at it or it is controlled by an external controller. In the maximum matching set, all nodes have their own superior (no two nodes share the same superior) so that they can be fully controlled. Each of the other N_D unmatched nodes (drivers) has to be controlled by an external controller. The minimality of the set of unmatched nodes is guaranteed by the maximum matching algorithm. A way to reduce the number of required external controllers is to make unmatched nodes to be matched. This can be implemented by linking unmatched nodes to unused superiors. Based on the definition of matching, each additional link can at most reduce one unmatched node, since a link can only point at one unmatched node. In this sense, at least $N_D - 1$ additional links are required to reduce N_D to 1. If $N_D - 1$

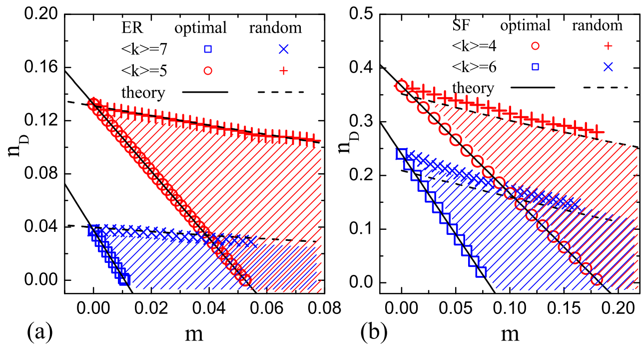

FIG. 2: (Color online.) Density n_D of unmatched nodes as a function of the fraction m of added links using our optimization and random strategies for (a) ER random networks and (b) SF networks with different average node degree. The network size is 5000 and the solid and dashed straight lines are theoretical predictions obtained from Eqs. (3), (4), and (6). The shadowed regions are controllability-enhanced regions. In the (m, n_D) plot, any strategy to enhance network controllability must fall into these regions.

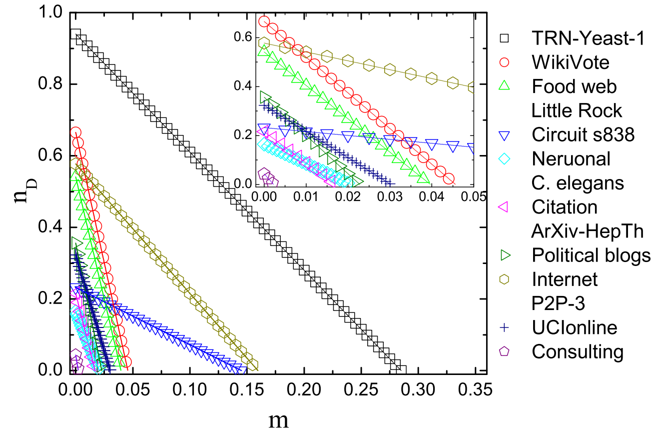

FIG. 3: (Color online) Density n_D as a function of the fraction m resulted from application of our optimization strategy to 10 real-world networks, which include transcriptional regulatory network of S. cerevisiae (TRN-Yeast-1) [18], who-vote-whom network of Wikipedia users (WikiVote) [19], food web in Little Rock lake (Food Web Little Rock) [20], electronic circuit (Circuit s838) [21], neural network of C. elegans (Neruonal C. elegans) [16], citation network in HEP-TH of ArXiv (Citation ArXiv-HepTh) [22], hyperlinks between weblogs on US politics (Political blogs) [23], Gnutella peer-to-peer file sharing network (Internet P2P-3) [24], online message network of students at UC, Irvine (UCIonline) [25] and Social network from a consulting company (Consulting) [26]. The inset zooms in the region of small values of m. The lines are the theoretical predictions from Eq. (3).

unmatched nodes are assigned to unused superior nodes, single external controller is sufficient to accomplish full control. Our method indeed achieves the optimal controllability with this minimum number of additional links. Figure 2 also indicates that a region can be identified (the shadowed region), into which any method designed to enhance the network controllability via structural changes must fall. In this sense, the optimal (steeper) lines in Fig. 2 represent the best strategy.

We now provide a theory to establish the optimality of our perturbation strategy. As has been confirmed, adding one link can at most decrease the number N_D of driver nodes by one. The optimal lines (the steeper lines in Fig. 2) as the result of applying our perturbation algorithm can be obtained through the relation $n_D(m) \cdot N = n_D^0 \cdot N - m \cdot N_l^0$, which gives

$$n_D(m) = n_D^0 - \frac{mN_l^0}{N}, \qquad (3)$$

where n_D^0 is the original density value in the absence of additional links and N_l^0 is the number of links in the original network. For the case of adding random links, one can use the cavity method to obtain n_D [6, 15]. In particular, for directed network with similar in- and out-degree distribution $P(k)$, where k is the corresponding incoming or outgoing degree, the density of driver nodes is

$$n_D = G(w_2) + G(1-w_1) - 1 + \langle k\rangle w_1(1-w_2), \qquad (4)$$

where $G(x)$ is the generating function given by $G(x) = \sum_{k=0}^{\infty} P(k)x^k$. The quantities w_1 and w_2 in Eq. (4) can be obtained by the following self-consistent equations:

$$\begin{aligned} w_1 &= H[1 - H(1-w_1)], \\ w_2 &= 1 - H[1 - H(w_2)], \end{aligned} \qquad (5)$$

where $H(x) = \sum_0^{\infty} Q(k+1)x^k$ is a generating function and $Q(k) = kP(k)/\langle k\rangle$. Equation (5) is valid for general networks in the absence of degree-degree correlations.

For ER random networks, $P(k)$ follows the Poisson distribution $e^{-\langle k\rangle}\langle k\rangle^k/k!$. Since randomly adding links into the network will not affect the degree distribution $P(k)$, we have $G(x) = H(x) = \exp[-\langle k\rangle^0(1+m)(1-x)]$, and

$$n_D = w_1 - w_2 + \langle k\rangle^0(1+m)w_1(1-w_2), \qquad (6)$$

where $\langle k\rangle^0$ is the average in- or out-degree in the original network, $w_1 = H(w_2) = \exp[-\langle k\rangle^0(1+m)(1-w_2)]$ and $w_2 = 1 - H(1-w_1) = 1 - \exp[-\langle k\rangle^0(1+m)w_1]$. For $k \gg 1$, we have

$$n_D \sim \exp[-\langle k\rangle^0(1+m)]. \qquad (7)$$

For SF networks, since m is small, we can assume that $P(k)$ is fixed, which leads to

$$n_D \sim \exp\left[-\langle k\rangle^0(1+m)\left(1 - \frac{1}{\gamma^0 - 1}\right)\right], \qquad (8)$$

where for the original SF network, $P(k) \sim k^{\gamma^0}$. These analytical results agree well with numerical simulations, as shown in Fig. 2. We have verified that, after adding all key links according to our optimization method, there exist directed paths from the single driver node (the starting node of the first matching path) to all other nodes, so that the network satisfies the full-rank condition.

We have also applied our structural perturbation strategy to a number of real networks from nature and society, as shown in Fig. 3. For every case examined, the perturbed network can be fully controlled by a single controller via adding a minimum number of links as determined by our method. For most real networks, about 5% of the additional links are sufficient to optimize their controllability (inset of Fig. 3), demonstrating that our structural perturbations are quite effective with low cost while maintaining the topology of the original networks. However, there are three networks (TRN-Yeast-1, Circuit 838 and Internet P2P-3) for which many more additional links are needed to achieve full control. This different behavior is mainly caused by the low average degrees of these networks (see Supplemental Material [12]), which induces many unmatched nodes with large ratio of the number of added links to the original number of links.

In general, our method is applicable to networks for which establishing a link costs less than imposing a time-variant controller at a node, such as many technological and social networks. However, there are networks in the real world for which the opposite is true, such as gene regulatory networks, where to establish a new regulatory connection between genes may be more difficult than exogenously altering the expression of a gene. For such networks, our optimization method is not meaningful and alternative ways to enhance the network controllability must be explored. In addition, the issue of trade-off between network robustness in response to failures/attacks and lower control cost with less controllers may be interesting.

Our optimization strategy, besides its practical usage to enhance network controllability, can surprisingly reveal the intricate relationship between certain structural properties of a complex network and its controllability. To illustrate this, we study how a number of fundamental quantities characterizing various structural properties of the network depend on the number of optimally added links, as follows: (1) the clustering coefficient (CC) [16] defined as the average of $2\bigtriangleup_i/[k_i(k_i-1)]$ over all nodes in the network, where $\bigtriangleup_i$ is the number of triangles that node i belongs to and k_i is node degree of i; (2) the degree-degree correlation (DDC) defined as [17]

$$\frac{N_l^{-1}\sum_i l_i k_i - [N_l^{-1}\sum_i \frac{1}{2}(l_i+k_i)]^2}{N_l^{-1}\sum_i \frac{1}{2}(l_i^2+k_i^2) - [N_l^{-1}\sum_i \frac{1}{2}(l_i+k_i)]^2}, \tag{9}$$

where l_i and k_i are the degrees of the two nodes at the ends of the ith link; (3) the inverse of average distance (IAD), where the distance between nodes is the length of the directed shortest path connecting them; and (4) heterogeneity (H) defined as $[\sum_i\sum_j |k_i-k_j|P(k_i)P(k_j)]/\langle k\rangle$ [6]. The values of all four quantities resulting from our optimization method are compared with those from the process of simply adding random links. For each quantity s, it is convenient to define the following relative value $S(m)=[\bar{s}_I(m)-\bar{s}_R(m)]/[\bar{s}_I(m)+\bar{s}_R(m)]$, where s stands for the value of either CC, DDC, IAD, or H, $s_I(m)$ and $s_R(m)$ are the corresponding values from our optimization procedure and from adding random links, respectively, $\bar{s}_I(m)=s_I(m)-s_0$, $\bar{s}_R(m)=s_R(m)-s_0$, and s_0 is the value of s for the unperturbed network. Computations

TABLE I: Four regions of the measure S of network structural properties as determined by $\bar{s}_I$ and $\bar{s}_R$.

S	$\bar{s}_I\bar{s}_R>0$	$\bar{s}_I\bar{s}_R<0$
$\lvert\bar{s}_I\rvert>\lvert\bar{s}_R\rvert$	(0, 1)	(1, ∞)
$\lvert\bar{s}_I\rvert<\lvert\bar{s}_R\rvert$	(-1, 0)	(-∞, -1)

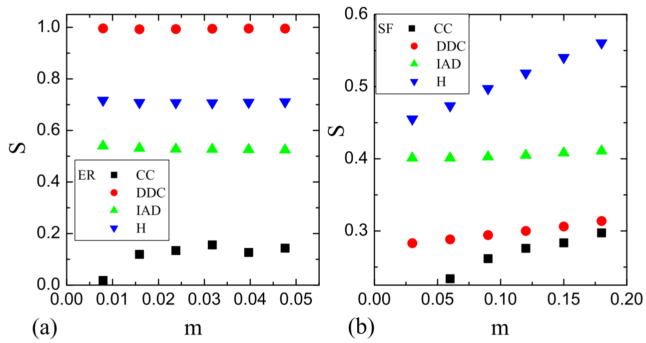

FIG. 4: (Color online.) Four measures of network structural properties as a function of m for (a) ER and (b) SF networks, which are CC, DDC, IAD, and H (see text for definitions). The network size is 5000 and the average node degrees of the ER and SF networks are 5 and 4, respectively. Each data point is obtained by averaging over 500 independent network realizations.

reveal four regions of S, as shown in Table I, depending on the variations in $\bar{s}_I(m)$ and $\bar{s}_R(m)$. Figure 4 shows $S(m)$ for both ER and SF networks, where we observe that all values of S fall in the interval $(0,1)$, indicating that $\bar{s}_I(m)$ and $\bar{s}_R(m)$ have the same variational trend but $|\bar{s}_I(m)|>|\bar{s}_R(m)|$. For the ER random network [Fig. 4(a)], we have $S_{DDC}\approx 1$, followed by S_H, S_{IAD}, and S_{CC}, which indicates that the DDC measure is the most pertinent quantity to the network controllability. In contrast, for the SF network [Fig. 4(b)], the heterogeneity measure H is the most relevant in shaping the network controllability, followed in order by the IAD, DDC, and CC measures. For both ER and SF networks, the CC measure has little effect on the controllability.

We can provide a heuristic explanation for the structural effects in terms of the node degree. According to the principle of maximum matching, nodes with larger degrees usually have relatively higher probability to find their own 'superior' and 'inferior' nodes. In contrast, smaller-degree nodes are more likely to be unmatched node (without superior) or/and the ending node of a matching path (without inferior). Therefore, more links are needed to be added to such nodes. This phenomenon can be used to understand the strong correlation between controllability and the quantities DDC and H in random and scale-free networks, respectively. In particular, in random networks, since node degrees are homogeneous, intentionally connecting smaller-degree nodes may not affect the value of H much as compared with adding random links, but the same act can cause DDC to become more positive. As a result, DDC is the most pertinent structural characteristic for random networks. However, for scale-free networks, H is considerably affected by intentional connections among small-degree nodes due to their heterogeneous nature, so its correlation with controllability is much stronger than those associated with other structural parameters. A complete understanding of the effects of structural properties on network controllability in complex networks is still challenging at the present, partly due to the fact that the effects of different structural properties cannot be separated from each other in a straightforward manner.

In conclusion, we have presented a perturbation approach to optimizing the controllability of complex networks. By adding a minimum number of links at judiciously chosen locations in the network, the full-rank condition can be guaranteed so that the perturbed network can be fully controlled using a single input signal. The control regions in the parameter space have been predicted analytically. An additional feature of our optimization framework is that it identifies, quantitatively, certain structural properties of the network that are key to its controllability. The field of controlling complex networks has gained momentum recently, and the principle presented here can be useful to guide the control of large complex networks at low cost.

We thank Dr. Y. Liu for valuable discussion and an anonymous referee for insightful suggestions. This work was supported by AFOSR under Grant No. FA9550-10-1-0083. WXW was also supported by NSFC under Grand No. 11105011.

[1] X. F. Wang and G. Chen, Physica A **310**, 521 (2002);X. Li, X. Wang and G. Chen, IEEE Trans. Circ. Syst.-I: **51**, 2074 (2004).

[2] F. Sorrentino, M. di Bernardo, F. Garofalo and G. Chen, Phys. Rev. E **75**, 046103 (2007).

[3] A. Lombardi and M. Hörnquist, Phys. Rev. E **75**, 56110 (2007).

[4] B. Liu, T. Chu, L. Wang and G. Xie, IEEE Trans. Automat. Contr. **53**, 1009 (2008).

[5] A. Rahmani, M. Ji, M. Mesbahi and M. Egerstedt, SIAM J. Contr. Optim. **48**, 162 (2009).

[6] Y.-Y. Liu, J.-J. Slotine and A.-L. Barabási, Nature (London), 473, **167** (2011).

[7] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg, C. T. Bergstrom, preprint arxiv:1106.2573.

[8] W.-X. Wang, Y.-C. Lai, J. Ren, B. Li, C. Grebogi, preprint arxiv:1107.2177.

[9] R. E. Kalman, J. Soc. Indus. Appl. Math. Ser. A **1**, 152 (1963).

[10] C.-T. Lin, IEEE Trans. Automat. Contr. **19**, 201 (1974).

[11] D. G. Luenberger, *Introduction to Dynamic Systems: Theory, Models, and Applications* (Wiley, 1979).

[12] See Supplemental Material at [URL will be inserted by publisher] for [give brief description of material].

[13] P. Erdős and A. Rényi, Publ. Math. Debrecen **6**, 290 (1959).

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

[15] H. Zhou and Z.-C. Ou-Yang, arxiv.org/abs/cond-mat/0309348.

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

[17] M. E. J. Newman, Phys. Rev. Lett. **89**, 208701 (2002).

[18] S. Balaji, M. MadanBabu, L. Iyer, N. Luscombe, and L. Aravind, J. Mol. Biol. **360**, 213 (2006).

[19] J. Leskovec, K. Lang, A. Dasgupta, and M. Mahoney, arXiv.org: 0810.1355 (2008).

[20] N. Martinez, Econological Monographs **61**,367(1991).

[21] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon, Science **298**, 824 (2002).

[22] J. K. J. Leskovec and C. Faloutsos, Graphs over time: Densification laws, shrinking diameters and possible explanations, ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD) (2005).

[23] L. A. Adamic and N. Glance, The political blogosphere and the 2004 US election, Proceedings of the WWW-2005 Workshop on the Weblogging Ecosystem (2005).

[24] J. K. J. Leskovec and C. Faloutsos, Graph evolution: Densification and shrinking diameters, ACM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1), (2007).

[25] T. Opsahl and P. Panzarasa, Social Networks **31**, 155 (2009).

[26] R. Cross and A. Parker, *The Hidden Power of Social Networks* (Harvard Business School Press, Boston, MA, 2004).