Rich-club phenomenon across complex network hierarchies

Julian J. McAuley[a)]
Statistical Machine Learning Program, NICTA, Canberra, Australian Capital Territory 0200, Australia and RSISE, Australian National University, Canberra, Australian Capital Territory 0200, Australia

Luciano da Fontoura Costa
Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos SP 13560-970, Brazil

Tibério S. Caetano
Statistical Machine Learning Program, NICTA, Canberra, Australian Capital Territory 0200, Australia and RSISE, Australian National University, Canberra, Australian Capital Territory 0200, Australia

(Received 17 February 2007; accepted 31 July 2007; published online 23 August 2007)

The "rich-club phenomenon" in complex networks is characterized when nodes of higher degree are more interconnected than nodes with lower degree. The presence of this phenomenon may indicate several interesting high-level network properties, such as tolerance to hub failures. Here, the authors investigate the existence of this phenomenon across the hierarchies of several real-world networks. Their simulations reveal that the presence or absence of this phenomenon in a network does not imply its presence or absence in the network's successive hierarchies, and that this behavior is even nonmonotonic in some cases. © *2007 American Institute of Physics*. [DOI: 10.1063/1.2773951]

The so-called *rich-club phenomenon* in complex networks is characterized when the hubs (i.e., nodes with high degrees) are on average more intensely interconnected than the nodes with smaller degrees. More precisely, it happens when the nodes with degree larger than k tend to be more densely connected among themselves than the nodes with degree smaller than k, for some significant range of degrees in the network.[1] This is quantified by computing the so-called *rich-club coefficient* across a range of k values. The name "rich club" arises from the analogy that hubs are "rich" because they have high degrees, and when the phenomenon is present, they form "clubs" because they are well connected among themselves.

The relevance of the rich-club phenomenon is that its presence or absence typically reveals important high-level semantic aspects of a complex network. For example, its presence in the scientific collaboration network of a given research area reveals that the particularly famous and influential scientists in that field are frequently coauthors with many other influential scientists in the same field. Similarly, the *absence* of the rich-club phenomenon in a protein-protein interaction dataset possibly reveals that proteins with large connectivity are presiding over different functions and are thus possibly coordinating distinct and specific functional modules.[2] The presence of the phenomenon in a power-grid network may indicate the robustness or stability of the network against blackouts, since several neighboring hubs would be available to aid a faulty hub in case of an emergency.

Given a specific network node i, it is possible to define its successive neighborhoods, i.e., the set of nodes which are at shortest distances of 1, 2, and so forth from the reference node i (e.g., Refs. 3–8). The set of nodes at length h from node i is then said to constitute the hth *hierarchical level* of node i. Such nodes can be understood as being linked to the reference node i through *virtual links*.[5] The number of nodes at the hth hierarchical level shall henceforth be called the hth *degree* of node i. We are now able to define successive degrees (or *hierarchies*) of entire networks.[7] That is, each node in the degree h network is connected to all nodes at its hth hierarchical level. The successive degrees of a small network are presented in Fig. 1.

Because of the finite size and diameter of the network, the hth degree tends to increase up to a peak and then decrease as the network is progressively encompassed by the higher hierarchies. Therefore, the maximum hierarchical level which can be considered for the hth degree is equal to the network diameter, i.e., the longest length of the shortest path among any two nodes in the network. The hth degree provides a natural means for gradually expressing more global aspects of the connectivity around each node. In other words, while the traditional node degree is an exclusively local measurement, the degree at successive levels provides information also about the medium to global scales of the network.

In this letter we investigate the behavior of the rich-club coefficient across different hierarchies of a complex network as the means to obtain more global extensions of that coefficient. We study, in particular, a power-grid network, a scientific collaboration network, and a protein-protein interac-

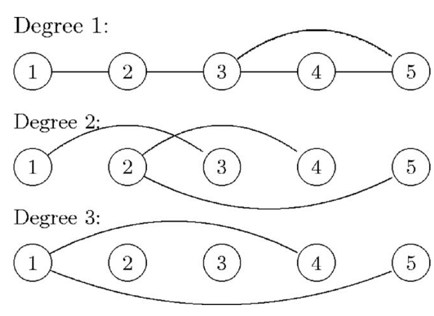

FIG. 1. Small network presented at degrees 1, 2, and 3. The hth degree of a node is now simply its degree in the corresponding hierarchy. That is, edges in the second and third degree networks simply connect those nodes whose distances in the first degree network is 2 and 3, respectively. It should be noted that each of the ten possible edges appears in exactly one of the three hierarchies.

[a)]Electronic mail: j.mcauley@student.unsw.edu.au

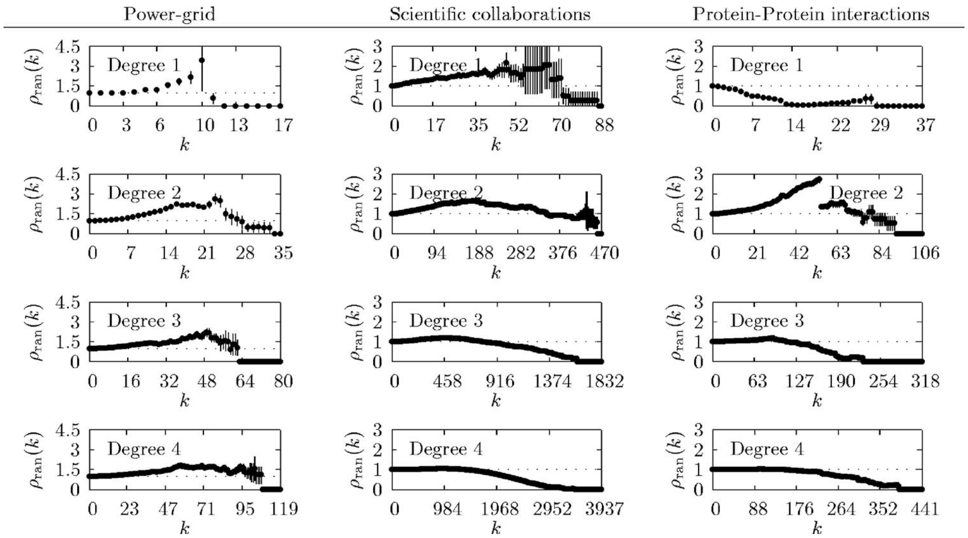

FIG. 2. Plots of the normalized rich-club coefficient for three different networks, up to degree $h=4$. Each plot shows the normalized rich-club coefficient $[\rho_{\mathrm{ran}}(k)]$, plotted against each value of the hth degree (k). The error bars reflect the mean and standard deviation of 50 experiments (the standard deviation is approximately zero in many graphs).

tion network. Our results reveal a variety of different behaviors for the rich-club phenomenon. The presence of the phenomenon may depend on the hierarchy, and we even report a nonmonotonic behavior for one of the networks in which the phenomenon appears and disappears as we progress over the hierarchies.

The rich-club phenomenon may be described as follows: consider a graph $G=(V,E)$ representing a complex network (here we restrict ourselves to *simple* graphs, i.e., unweighted graphs with no multiple edges or loops). Let $V_{>k}$ be the set of vertices with degree larger than k, $N_{>k}$ be the number of such vertices, and $E_{>k}$ be the number of edges among such vertices. The so-called rich-club coefficient is given by

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}, \qquad (1)$$

i.e., the fraction between the *actual* and the *potential* number of edges among $V_{>k}$.[1]

This measure clearly reflects how densely connected the vertices $V_{>k}$ are. One could at first think that the rich-club phenomenon would apply if $\phi(k)$ were an increasing function of k, i.e., if vertices with large degree were more densely connected among themselves than vertices with low degree. This was indeed assumed in Ref. 1, where the increasing dependency of $\phi(k)$ on k was called the "rich-club phenomenon."

However, one must notice that vertices with higher degree will be naturally more likely to be more densely connected than vertices with smaller degree, simply due to the fact that they have more incident edges. As a result, for a proper evaluation of this phenomenon we must normalize out this factor. This point was raised in Ref. 2, which derived an analytical expression for the rich-club coefficient of uncorrelated large-size networks at high degrees,

$$\phi_{\mathrm{unc}}(k) \underset{k,k_{\max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N}, \qquad (2)$$

where $k_{\max}$ is the maximum degree in the network, and claimed that it should be used to find a normalized rich-club coefficient, $\rho_{\mathrm{unc}}(k)=\phi(k)/\phi_{\mathrm{unc}}(k)$. $\phi_{\mathrm{unc}}(k)$ is, however, not properly defined in some cases, such as for heavy-tailed distributions.[2] In practice then the normalization factor is obtained by generating a randomized version of the network with the same degree distribution. A simple algorithm[9] to achieve this consists in flipping the endpoints of two random edges and iterating: at each iteration the degrees of the four nodes involved will remain the same but the edge structure will change. If sufficiently many iterations are carried out, the final network will be in some sense a random network but with the same degree distribution as the initial network. We then compute the rich-club coefficient for the resulting "maximally random network," $\phi_{\mathrm{ran}}(k)$, and use it for finding the normalized rich-club coefficient, $\rho_{\mathrm{ran}}(k)=\phi(k)/\phi_{\mathrm{ran}}(k)$. As a result, while $\rho_{\mathrm{unc}}(k)$ gives the rich-club coefficient with respect to an ideal uncorrelated graph, $\rho_{\mathrm{ran}}(k)$ is a realistic normalized measure that takes into account the structure and finiteness of the network. In our simulations we compute $\rho_{\mathrm{ran}}(k)$ for real-world complex networks across a range of values of k but also across the *hierarchy* of networks derived from the original one.[5,7]

We have set up a series of experiments on several complex network datasets. The first is related to the power grid of the western states of the United States of America.[10] We also investigated a scientific collaboration network from the great area of condensed matter physics[2] and a protein-protein interaction network of the yeast *Saccharomyces cerevisiae*[11] (these data sets are available in Refs. 12–14). We have computed the normalized rich-club coefficient across the hth degrees of the network for the first four hierarchies. Figure 2 shows the results we obtained. In each graph, the vertical axis corresponds to the (normalized) rich-club coefficient, while the horizontal axis corresponds to the hth degree (plotted up to the degree of the largest hub in the corresponding hierarchy). Here, the random network was obtained by performing $|E|/2$ edge swaps, repeated 50 times. The rich-club coefficient was computed after each of these 50 steps. The error bars in the plot reflect the mean and standard deviation of these 50 experiments.

The rich-club phenomenon is characterized by an increasing dependency of the normalized rich-club coefficient on the degree of the network. For the power-grid network, the phenomenon is present with significant strength for all hierarchies. For the scientific collaboration network, the phenomenon appears for the first degree network and progressively attenuates along further levels. Finally, the protein-protein interaction network reveals a particularly interesting behavior: the phenomenon is absent for the first degree, appears with strength in the second degree, and disappears again along the higher degrees. This nonmonotonic behavior of the rich-club phenomenon across hierarchies is a nontrivial fact that can provide valuable information about the overall structure of the network.

For the power-grid network, the presence of the rich-club phenomenon reveals that hubs are highly connected and thus presumably, there is more stability in the sense that the duties of faulty hubs may be more easily taken over by neighboring hubs (since there are many of them). The presence of the phenomenon across all hierarchies might reveal the fact that such stability is verified across a range of scales of the network, suggesting higher resilience. For example, connections among neighborhoods, cities, and counties may all exhibit a certain degree of stability. In the scientific collaboration network, the phenomenon is present for the first degree as expected, indicating that renowned scientists in a given field are likely to have been coauthors in at least one paper. However, as we move across hierarchies, the strength of the phenomenon is progressively dissipated. This may be interpreted as follows: for higher hierarchies, progressively different scientific subcommunities are being considered, and in this case, it is unlikely that great scientists from different subareas have been coauthors in at least one paper. Finally, we have the results for the protein-protein interaction network. The absence of the phenomenon for a given hierarchy of this network might indicate that at this hierarchy, key proteins are specialized and preside over different groups of proteins. The malfunction of a protein will then in general be critical. On the other hand, the presence of the phenomenon may indicate that key proteins act in concert, which suggests a certain degree of stability in the activities for which they are responsible. The nonmonotonicity observed then implies that different patterns of specialization are characteristic of specific hierarchies instead of being a progressive feature over hierarchies. For this network, the first degree reveals a high degree of specialization of the proteins, the second degree reveals much less specialization, and the higher degrees suggest a more neutral regime. This is a particularly interesting finding because it reveals that patterns of stability or specialization may alternate as the scale from which an organism is observed is varied. An interesting question to be further pursued would then be the investigation of whether such varying patterns of signatures of specialization or stability/resilience would correlate with data or prior knowledge of, say, subsystems of the human body which present varying degrees of resilience to malfunction or disease. Our results possibly suggest that overspecialization or perhaps even instability of subsystems of an organism does not necessarily imply instability of the organism in a global scale.

One of the authors (L.d.F.C.) is grateful to CNPq (Nos. 308231/03-1 and 301303/06-1) and FAPESP No. (05/00587-5) for financial support. NICTA is funded through the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

[1] S. Zhou and R. J. Mondragon, IEEE Commun. Lett. **8**, 180 (2004).
[2] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Nat. Phys. **2**, 110 (2006).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, Proceedings of ACM SIGCOMM, 1999 (unpublished).
[4] R. Cohen, S. Havlin, S. Mokryn, D. Dolev, T. Kalisky, and Y. Shavitt, e-print arXiv:cond-mat/0305582.
[5] L. da F. Costa, Phys. Rev. Lett. **93**, 098702 (2004); e-print arXiv:cond-mat/0312646.
[6] L. da F. Costa and L. H. C. da Silva, Eur. Phys. J. B **50**, 237 (2006); arXiv:cond-mat/0408076.
[7] R. F. Andrade, J. G. Miranda, and T. Petit Lobão, Phys. Rev. E **73**, 046101 (2006).
[8] L. da F. Costa and F. N. Silva, J. Stat. Phys. (in press); e-print arXiv:cond-mat/0412761.
[9] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon, e-print arXiv:cond-mat/0312028.
[10] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).
[11] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, Nature (London) **411**, 41 (2001).
[12] D. J. Watts and S. H. Strogatz (http://cdg.columbia.edu/cdg/datasets).
[13] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani (http://cxnets.googlepages.com/research2).
[14] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai (http://www.nd.edu/networks/resources.htm).