# The rich-club phenomenon across complex network hierarchies

Julian J. McAuley[1,2], Luciano da Fontoura Costa[3] and Tibério S. Caetano[1,2]

[1] *National ICT Australia, Canberra ACT 0200, Australia*
[2] *RSISE, Australian National University, Canberra ACT 0200, Australia*
[3] *Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos SP 13560-970, Brazil*

The so-called rich-club phenomenon in a complex network is characterized when nodes of higher degree (hubs) are better connected among themselves than are nodes with smaller degree. The presence of the rich-club phenomenon may be an indicator of several interesting high-level network properties, such as tolerance to hub failures. Here we investigate the existence of the rich-club phenomenon across the hierarchical degrees of a number of real-world networks. Our simulations reveal that the phenomenon may appear in some hierarchies but not in others and, moreover, that it may appear and disappear as we move across hierarchies. This reveals the interesting possibility of non-monotonic behavior of the phenomenon; the possible implications of our findings are discussed.

## I. INTRODUCTION

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 co-authors 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 the 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 distance of 1, 2, and so forth, from the reference node $i$ (e.g. [3, 4, 5, 6]). Recently, a series of hierarchical measurements have been proposed and investigated for the characterization of the structure of a complex network [6, 7, 8]. These involve the definition of the hierarchical degree, expressing the connectivity between the successive hierarchical neighborhoods centered at each network node. Such a formalism is useful since it not only progressively extends the locality of the node degree but also has the ability to reveal patterns associated with the *indirect* relations in a network, i.e. the so-called *virtual links* among nodes [6].

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 interaction 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 non-monotonic behavior for one of the networks, in which the phenomenon appears and disappears as we progress over the hierarchies.

## II. THE RICH-CLUB PHENOMENON

Consider a graph $G = (V, E)$ representing a complex network. 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 [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 [2], who derived an analytical expression for the rich-club coefficient of uncorrelated large-size networks at high degrees

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

and claimed that it should be used to find a normalized rich-club coefficient, $\rho_{\text{unc}}(k) = \phi(k)/\phi_{\text{unc}}(k)$. $\phi_{\text{unc}}(k)$ 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", $\phi_{\text{ran}}(k)$, and use it for finding the normalized rich-club coefficient, $\rho_{\text{ran}}(k) = \phi(k)/\phi_{\text{ran}}(k)$. As a result, while $\rho_{\text{unc}}(k)$ gives the rich-club coefficient with respect to an ideal uncorrelated graph, $\rho_{\text{ran}}(k)$ is a realistic normalized measure that takes into account the structure and finiteness of the network. In our simulations we compute $\rho_{\text{ran}}(k)$ for real-world complex networks across a range of values of $k$ but also across the *hierarchy* of networks derived from the original one [6].

### III. COMPLEX NETWORK HIERARCHIES

Given a node $i$, the other nodes which are at shortest path of length $h$ from $i$ constitute the $h^{th}$ hierarchical level of that node. For a specific hierarchical level $h$ defined by a node $i$, the number of nodes between this level and the next level (i.e. the hierarchical level $h + 1$) is defined as the *hierarchical degree* of node $i$ [6, 7, 8]. Because of the finite size and diameter of the network, the hierarchical node 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 hierarchical node degree is equal to the network diameter, i.e. the longest length of the shortest path among any two nodes in the network. The hierarchical node 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 hierarchical degree at successive levels provides information also about the medium to global scales of the network.

### IV. EXPERIMENTS

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 at [12, 13, 14], respectively). We have computed the normalized rich-club coefficient across the hierarchical degrees of the network for the first 4 hierarchies. Figure 1 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 hierarchical degree (plotted up to the degree of the largest hub in the corresponding hierarchy). 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 order and progressively attenuates along further levels. Finally, the protein-protein interaction network reveals a particularly interesting behavior: the phenomenon is absent for the first order, appears with strength in the second order and disappears again along the higher orders. This non-monotonic behavior of the rich-club phenomenon across hierarchies is a nontrivial fact that can provide valuable information about the overall structure of the network.

### V. DISCUSSION

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 order as expected, indicating that renowned scientists in a given field are likely to have been co-authors 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 sub-communities are being considered and in this case it is unlikely that great scientists from different sub-areas have been co-authors 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

Power-grid



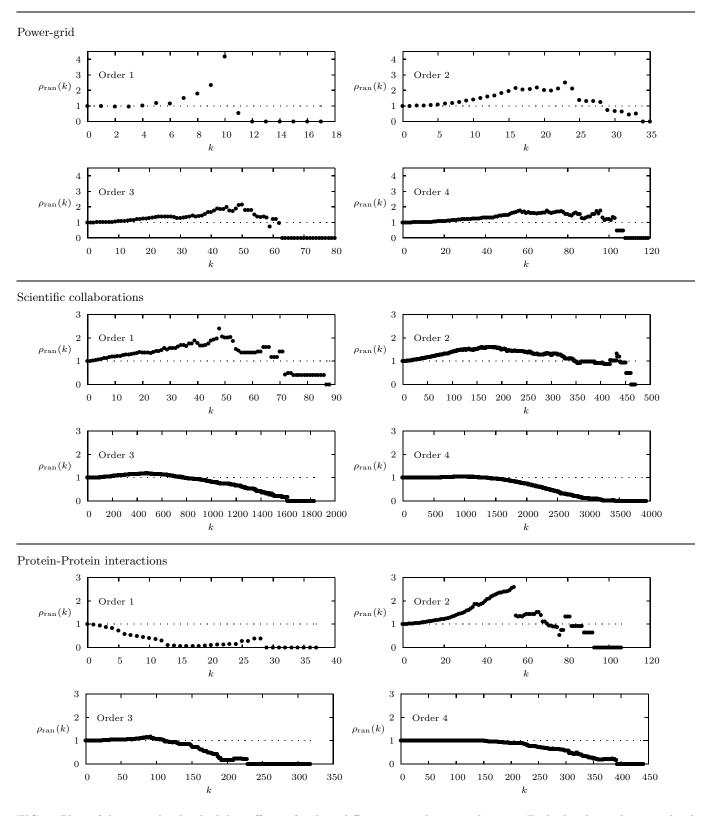Scientific collaborations

Protein-Protein interactions

FIG. 1: Plots of the normalized rich-club coefficient for three different networks, up to degree 4. Each plot shows the normalized rich-club coefficient ($\rho_{\mathrm{ran}}(k)$), plotted against each value of the degree ($k$).

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, what suggests a certain degree of stability in the activities for which they are responsible. The non-monotonicity 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 order reveals a high degree of specialization of the proteins, the second order reveals much less specialization, and the higher orders 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, sub-systems of the human body which present varying degrees of resilience to malfunction or disease. Our results possibly suggest that over-specialization or perhaps even instability of sub-systems of an organism does not necessarily imply instability of the organism in a global scale.

[1] S. Zhou and R. J. Mondragon, IEEE Commun. Lett. **8**, 180 (2004).
[2] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Nature Phys. **2**, 110 (2006).
[3] M. Faloutsos, P. Faloutsos, and C. Faloutsos, SIAM Review **29**, 251 (1999).
[4] M. E. Newman (2001), URL cond-mat/0111070.
[5] R. Cohen, S. Havlin, S. Mokryn, D. Dolev, T. Kalisky, and Y. Shavitt (2003), URL cond-mat/0305582.
[6] L. da F. Costa, Phys. Rev. Letts. **93**, 098702 (2004), URL cond-mat/0312646.
[7] L. da F. Costa and L. H. C. da Silva, Eur. Phys. J. B **50**, 237 (2006), URL cond-mat/0408076.
[8] L. da F. Costa and F. N. Silva, J. Stat. Phys. (in press) (2007), URL cond-mat/0412761.
[9] R. Milo, N. Kashtan, S. Itzkovitz, M. E. J. Newman, and U. Alon (2006), cond-mat/0312028.
[10] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998), ISSN 0028-0836, URL http://dx.doi.org/10.1038/30918.
[11] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai, Nature **411**, 41 (2001), URL http://dx.doi.org/10.1038/411041a0.
[12] D. J. Watts and S. H. Strogatz (1998), URL http://cdg.columbia.edu/cdg/datasets.
[13] H. Jeong, S. P. Mason, A. L. Barabási, and Z. N. Oltvai (2001), URL http://www.nd.edu/~networks/resources.htm.
[14] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani (2006), URL http://cxnets.googlepages.com/research2.