Detecting rich-club ordering in complex networks

Vittoria Colizza
vcolizza@indiana.edu

Alessandro Flammini
aflammin@indiana.edu

M. Angeles Serrano
mdserran@indiana.edu

Alessandro Vespignani
alexv@indiana.edu

School of Informatics and Center for Biocomplexity
Indiana University
Bloomington 47406 IN

ABSTRACT

Uncovering the hidden regularities and organizational principles of networks arising in physical systems ranging from the molecular level to the scale of large communication infrastructures is the key issue for the understanding of their fabric and dynamical properties [1, 2, 3, 4, 5]. The "rich-club" phenomenon refers to the tendency of nodes with high centrality, the dominant elements of the system, to form tightly interconnected communities and it is one of the crucial properties accounting for the formation of dominant communities in both computer and social sciences [4, 5, 6]. Here we provide the analytical expression and the correct null models which allow for a quantitative discussion of the rich-club phenomenon. The presented analysis enables the measurement of the rich-club ordering and its relation with the function and dynamics of networks in examples drawn from the biological, social and technological domains.

Keywords

Complex networks, analysis, measurement, hierarchies

Recently, the informatics revolution has made possible the analysis of a wide range of large scale, rapidly evolving networks such as transportation, technological, social and biological networks [1, 2, 3, 4, 5]. While these networks are extremely different from each other in their function and attributes, the analysis of their fabric provided evidence of several shared regularities, suggesting general and common self-organizing principles beyond the specific details of the individual systems. In this context, the statistical physics approach has been exploited as a very convenient strategy because of its deep connection with statistical graph theory and because of its power to quantitatively characterize macroscopic phenomena in terms of the microscopic dynamics of the various systems [1, 2, 3, 4]. As an initial discriminant of structural ordering, attention has been focused on the networks' degree distribution; i.e., the probability $P(k)$ that any given node in the network shares an edge with k neighboring nodes. This function is, however, only one of the many statistics characterizing the structural and hierarchical ordering of a network; a full account of the connectivity pattern calls for the detailed study of the multi-point degree correlation functions and/or opportune combination of these.

.

In this work, we tackle a main structural property of complex networks, the so-called "rich-club" phenomenon. This property has been discussed in several instances in both social and computer sciences and refers to the tendency of high degree nodes, the hubs of the network, to be very well connected to each other. Essentially, nodes with a large number of links - usually referred to as *rich nodes* - are much more likely to form tight and well interconnected subgraphs (*clubs*) than low degree nodes. A first quantitative definition of the rich-club phenomenon is given by the rich-club coefficient ϕ, introduced by Zhou and Mondragon in the context of the Internet [6]. Denoting by $E_{>k}$ the number of edges among the $N_{>k}$ nodes having degree higher than a given value k, the rich-club coefficient is expressed as:

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

where $N_{>k}(N_{>k}-1)/2$ represents the maximum possible number of edges among the $N_{>k}$ nodes. Therefore, $\phi(k)$ measures the fraction of edges actually connecting those nodes out of the maximum number of edges they might possibly share. The rich club coefficient is a novel probe for the topological correlations in a complex network, and it yields important information about its underlying architecture. Structural properties, in turn, have immediate consequences on network's features and tasks, such as e.g. robustness, performance of biological functions, or selection of traffic backbones, depending on the system at hand. In a social context, for example, a strong rich-club phenomenon indicates the dominance of an "oligarchy" of highly connected and mutually communicating individuals, as opposed to a structure comprised of many loosely connected and relatively independent sub-communities. In the Internet, such a feature would point to an architecture in which important hubs are much more densely interconnected than peripheral nodes in order to provide the transit backbone of the network [6]. It is also worth stressing that the rich club phenomenon is not trivially related to the mixing properties of networks, which enable the distinction between assortative networks, where large degree nodes preferentially attach to large degree nodes, and disassortative networks, showing the opposite tendency [4, 7, 8]. Indeed, the rich club phenomenon and the mixing properties express different features that are not trivially related or derived one from each other.

We analyze the behavior of the rich club coefficient as a function of the degree in a variety of real world networks drawn from the biological, social and technological world. In particular, we study: (1) the *Protein Interaction Network* [8, 9] of the yeast *Saccharomyces Cerevisiae*; (2) the *Scientific Collaboration Network* [10]; (3) the network of *Worldwide Air Transportation* [11]; (4) the *Internet* network at the Autonomous System level [4, 12, 13]. We also consider three standard network models: the Erdös-Rényi (ER) graph [14], the generalized random network having a heavy-tailed degree distribution obtained with the Molloy-Reed (MR) algorithm [15], and the Barabasi-Albert (BA) model [16].

All analyzed datasets display a monotonic increasing behavior of $\phi(k)$. This feature is claimed to provide evidence of the rich-club phenomenon since $\phi(k)$ progressively increases in vertices with increasing degree (*e.g.*, see Ref. [6] for the Internet case). However, a monotonic increase of $\phi(k)$ does not necessarily implies the presence of the rich-club phenomenon. Indeed, even in the case of the ER graph - a completely random network - we find an increasing rich-club coefficient. This implies that the increase of $\phi(k)$ is a natural consequence of the fact that vertices with large degree have a larger probability of sharing edges than low degree vertices. This feature is therefore imposed by construction and does not represent a signature of any particular organizing principle or structure, as is clear in the ER case. The simple inspection of the $\phi(k)$ trend is therefore potentially misleading in the discrimination of the rich-club phenomenon.

Appropriate baselines have to be identified in order to be able to detect the rich-club phenomenon. From a theoretical analysis of $\phi(k)$, we derive the expressions for two normalized measures which provide the discrimination of the actual presence of the rich club-phenomenon by pointing out unavoidable structural correlations and ordering principles shaping the network.

Results show a strong rich-club ordering in the Scientific Collaboration Network, providing support to the idea that the elite formed by more influential scientists tends to form collaborative groups within specific domains. A clearly opposite result is found for the Protein Interaction Network, where the lack of rich-club ordering indicates that proteins with large number of interactions are presiding over different functions and thus, in general, are coordinating specific functional modules. This different kind of structural organization is shown in Figure 1 which displays portions of the Protein Interaction Network and the Scientific Collaboration Network including the club of $N_{>k}$ nodes and the connections among them. The network representations clearly show the presence of a rich-club phenomenon in the Scientific Collaboration Network, where the majority of rich nodes are highly interconnected forming tight subgraphs, in contrast with the Protein Interaction Network case, where only few links appear to connect rich nodes, the rest linking to lower degree vertices.

In summary, this analysis provides the baseline functions for the detection of the rich-club phenomenon and its effect on the structure of large scale networks. This allows the measurement of this effect in a wide range of systems, finally enabling a quantitative discussion of various claims such as “high centrality” backbones in technological networks and “elitarian” clubs in social systems.

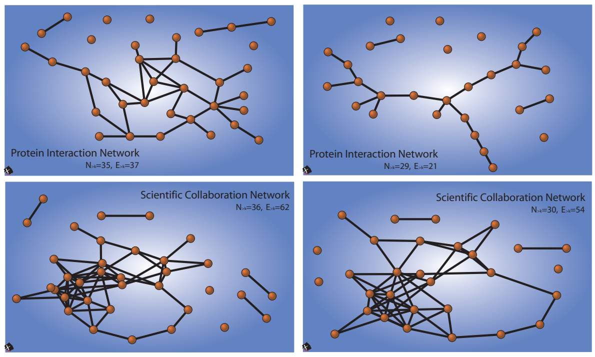

Figure 1: Graph representations of the rich-clubs. Progressively smaller clubs of $N_{>k}$ rich nodes in the Protein Interaction Network -top- and in the Scientific Collaboration Network -bottom- are shown together with the $E_{>k}$ connections among them. The graphs have been produced with the Pajek software (http://vlado.fmf.uni-lj.si/pub/networks/pajek/).

1. REFERENCES

[1] Albert, R. & Barabási, A.-L. *Rev. Mod. Phys.* **74**, 47–97 (2002).

[2] Newman, M.E.J. *SIAM Review* **45**, 167–256 (2003).

[3] Dorogovtsev, S.N. & Mendes, J.F.F., *Evolution of networks: From Biological nets to the Internet and WWW*, Oxford Univ. Press, Oxford (2003).

[4] Pastor-Satorras, R. & Vespignani, A. *Evolution and Structure of the Internet: A statistical physics approach*, Cambridge Univ. Press, Cambridge (2004).

[5] Wasserman, S. & Faust, K. *Social Network Analysis*, Cambridge Univ. Press, Cambridge (1994).

[6] Zhou, S. & Mondragon, R.J. *IEEE Comm. Lett.* **8**, 180–182 (2004).

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

[8] Maslov, S. & Sneppen, K. *Science* **296**, 910–913 (2002).

[9] Colizza, V., Flammini, A., Maritan, A. & Vespignani, A. *Phys. A* **352**, 1–27 (2005).

[10] Newman, M.E.J. *Phys. Rev. E* **64**, 016131 (2001).

[11] Barrat, A., Barthélemy, M., Pastor-Satorras, R. & Vespignani, A. *Proc. Natl. Acad. Sci. USA* **101**, 3747–3752 (2004).

[12] Faloutsos, M., Faloutsos, P.& Faloutsos, C. *Comput. Commun. Rev.* **29**, 251–263 (1999).

[13] Vázquez, A., Pastor-Satorras, R. & Vespignani, A. *Phys. Rev. E* **65**, 066130 (2002).

[14] Erdös, P. & Rényi, A. *Publicationes Mathematicae* **6**, 290–297 (1959).

[15] Molloy, M. & Reed, B. *Random Structures Algorithms* **6**, 161–179 (1995).

[16] Barabási, A. -L. & Albert, R. *Science* **286**, pp. 509–512 (1999).