Local vs global rich-club effects in complex networks

Author: Marc Flo Ponsà
Facultat de Física, Universitat de Barcelona (UB), Av. Diagonal 645, 08028 Barcelona, Spain.

Advisor: M. Ángeles Serrano

Abstract: Rich-club effect is a phenomenon where nodes of higher degree in a network tend to be highly well-connected between them. In this work we aim to quantify it, analyze its presence in a wide range of different real networks, propose a new expression to calculate rich-club effect and discuss its utility to quantify this phenomemon.

I. INTRODUCTION

Complex networks describe a large number of systems, whose nature may be very different, from molecular interactions in the cell to social networks [1]. The structure of these systems is typically modeled as random graphs following complex connectivity patterns [2], far from completely ordered lattices or fully random topologies [2].

In complex networks, the elements of the system are represented as nodes and interactions between them are called links. The number of links that a node shares with others is the degree of the node [2]. In a network, the nodes with high degree much above the average are called hubs and these play an important role in this work [1], because rich-club effect is related to the connectivity between hubs, as will be explained below.

All networks display similar topological features, independently of their domain. Networks describing real systems usually present fat-tailed degree distributions described by a power-law functional form with negative exponent [3]. Another typical characteristic of real networks is that they are sparse, meaning that the total number of links in the network represents a low fraction of the total possible, that is $N(N-1)/2$ where the number of nodes is N [1]. They share with random graphs the small-world property, meaning that all pair of nodes are separated, on average, by a small number of intermediate links (the topological distance between nodes is small)[1]. Another feature that differentiates a real from a random network is the clustering coefficient $\langle C \rangle$ [2], which measures the tendency of the neighbors of a given node to share links between them [2]. Real networks typically display a high clustering coefficient, while clustering is small in random networks.

Finally, one important property of real networks is the rich-club effect, a phenomenon implying that high degree nodes tend to be well-connected between them [4]. In 2004 was introduced an expression to measure it, which is obtained by calculating the density of links over all possible pairs [4]. Two years later, in 2006, this expression was normalized to take into account the structural properties of each system [5]. The study of this phenomenon is important because the detection or not of its presence gives information about hubs behaviour, allowing to analyze in the context of each network the reasons why nodes of higher degree tend to associate between them or not [5].

In this manuscript, we characterize the rich club effect in model and real networks. We also compare results for the standard metric based on global density of links with an alternative based on local connectivity, and discuss similarities and differences.

II. MATERIALS AND METHODS

A. Models of complex networks

To provide a reference for real networks, in this paper we will consider two different network models.

The Erdös-Rényi (ER) model is the simplest random network model. ER graphs are described by two parameters, the number of nodes N and the probability p that two nodes share a link between them [6]. Since the maximum number of node pairs is $N(N-1)/2$, and each pair is created with a probability p, the resulting average degree is given by the expression $p = \frac{<k>}{N-1}$ [6]. Degree distribution for this model is binomial and as $n \to \infty$ it becomes a Poisson distribution. Erdös-Rényi model presents low clustering unlike many real networks. [6]

The Barabási-Albert (BA) model is a growing network model. It generates graphs whose structure is closer to the observed in real networks because it incorporates a preferential attachment mechanism, so that new nodes added to the graph prefer to link with more connected nodes[7]. To obtain a network following this model one has to fix two parameters, the number of nodes N at the end of the process and the number of links m. The process starts by creating $N_0 = m$ nodes and connecting them. At any step until the system reaches N nodes, a node t is added that will share a link with m nodes i of the network, choosing each node with a probability $p_i = \frac{k_i}{\sum_j k_j}$. As a consequence, nodes with higher degree will have more probability to be chosen, obtaining thus the condition of preferential attachment[7].

B. Network data

We analyze the rich club effect in some real networks from different domains. Some of their statistics are shown in Table 1.

- **Roads**. The road network compiles information about E-road network. Each node represents a city and each link denotes that they are connected by an E-road [8].(From KONECT database, April 2017).

Networks					
Type	N	L	$\langle k \rangle$	k_{max}	$\langle C \rangle$
Erdös-Rényi	1000	3000	6	17	0.01
Barabási-Albert	1000	2994	6	97	0.03
Roads	1174	1417	2.41	10	0.02
Power grid	4961	6594	2.67	19	0.11
Air transport	500	2980	11.94	145	0.72
Internet	6444	12194	3.78	1457	0.38
Emails	1133	5451	9.62	71	0.25
Co-authorship	5241	28968	11.05	162	0.88
Metabolic	829	1975	4.76	202	0.19
Proteins	4100	13356	6.52	313	0.09

Table I: This table shows the values of the most important magnitudes for the synthetic and real networks considered in this work. N is the number of nodes, L represents the number of links, $\langle k \rangle$ is the average degree of the network, k_{max} is the maximum degree and $\langle C \rangle$is the average clustering coefficient.

- **Power grid**. Power grid network compiles information about power grid of the Western States of the United States of America. Each node represents neither a generator, a transformator or a substation, each link represents a power supply line [9].(From KONECT database , April 2017).
- **Air transportation**. Air transportation network compiles information about the 500 busiest commercial airports in United States. Each node represents an airport and each link a flight between them [10].(From Index of Complex Networks database).
- **Internet**. Internet network compiles information about Autonomous systems in Internet. Each node represents an autonomous system and each link represents BGP traffic between them [11].(From Stanford database SNAP)
- **Emails**. Emails network compiles information about the email communication network at the University Rovira i Virgili. Each node represents a user and each link that at least one email between them was sent [12].((From KONECT database, April 2017).
- **Scientific collaborations**. Collaborations network compiles information about collaborations between authors in the e-print arXiv. Each node represents an author and each link a coauthorship between them [13].(From Stanford database SNAP)
- **Metabolic reactions**. Metabolic network compiles information about enzymatic reactions. Each node is neither a reactant or a product and each link represents a reactant-product pair.
- **Proteins**. Proteins network compiles information about interactions between proteins. Each node represents a protein and each link a connection between them [15].

All these networks are undirected and unweighted. It means there is no direction in the connection between nodes and links have no weight assigned to them [1, 16] .

C. Quantifying the rich-club coefficient

S.Zhou and R.J.Mondragon(2004) discussed for first time about rich-club effect [4]. Moreover, they gave an analytic expression to calculate it, the rich-club coefficient $\phi(k)$. It is expressed as:

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

For a given k, this expression allows to calculate the fraction of links in a network among the $N(N-1)/2$ possible pairs [4]. Then, if this coefficient is calculated and it is observed that increases with the value of k, means higher degree nodes tend to connect between them. That is the definition of rich-club effect, this coefficient gives an accurate numerical expression of this phenomena.

This behaviour means that hubs are well-connected between them, so the connections between nodes are not formed randomly, otherwise following some kind of process that ends up in this rich-club effect being observed. Despite that, calculating and plotting this coefficient for different networks, from pure random networks generated by models to real systems represented with data, it is observed that in every case this coefficient follows the same behaviour of increasing with k even for the ER model that produces random graphs with trivial structure where all pairs of nodes are connected with the same probability [5]. This indicates that Eq. (1) is not good for measuring the rich-club phenomenon.

In [5], it is proposed a normalized version of this coefficient that takes into account the contribution due to the high random generation of connections between high degree nodes, the expression of this coefficient is:

$$\rho(k) = \frac{\phi(k)}{\phi_{rand}(k)} \tag{2}$$

$\phi_{rand}(k)$ is Eq. (1) calculated for a network that preserves the degree distribution of the original, but has been randomized by a process of L^2 steps, where in each step two different links between nodes are selected and two of their ends are changed, so they connect different nodes. Eq. (2) can be simplified as $\frac{L_{>k}}{L_{>k_{rand}}}$, taking into account that as degree distribution does not change, $N_{>k}$ is constant for each k, while $L_{>k_{rand}}$ may be different every time that two ends of links are changed. This process is done in such a big amount of steps before calculating $L_{>k_{rand}}$ to make sure that the resulting network is completely random. This value allows to discount some contributions due to structural effects of the system. Then, if $\rho(k) > 1$ the network presents rich-club effect [5].

D. Local rich-club coefficient

In this manuscript, we also evaluate a local version of the rich-club coefficient on real networks that focuses on local connectivity instead of on density of links. Its written as:

$$\phi_{loc} = \frac{\langle k \rangle_{>k}}{\langle k \rangle} \tag{3}$$

where numerator is average degree for the subgraph obtained with nodes of degree higher than k. We aim to compare this expression to Eq. (1) and analyze if it gives some kind of information about the network. The normalized version of this coefficient can be calculated as $\rho_{loc} = \frac{\phi_{loc}}{\phi_{loc,rand}}$, that is simplified into $\frac{L_{>k}}{L_{>k_{rand}}}$ as for the global coefficient.

III. RESULTS

A. Model networks

Next, we evaluate the plain and normalized rich-club effect in the ER and BA network models as defined in Eq. (1) and (2). To calculate Eq. (2), $L_{>k_{rand}}$ has been averaged by calculating this magnitude 100 times, each time the process of L^2 steps being done.

Erdös-Rényi Model. Results for the rich-club effect in ER networks with $p = 6 \cdot 10^{-2}$ and $\langle k \rangle = 6$ are shown in the top row of Fig. 1. These values of p and $\langle k \rangle$ have been chosen so ER networks are sparse as real systems. These graphs have been simulated with $N = 1000$. Coefficient $\phi(k)$ shows an increasing behaviour with k, as expected. Moreover, $\rho(k) = 1$ for all k, due to the fact that ER networks are completely random.

As shown in FIG. 2, fluctuations in the normalized rich-club coefficient become larger for larger values of k due to the decreasing number of nodes with larger k. These fluctuations decrease as N becomes larger.

Barabási-Albert Model. The process to obtain the results for $\phi(k)$ and $\rho(k)$ has been the same than for Erdös-Rényi Model, with $\langle k \rangle = 6$ and $m = 3$. However, with $N = 1000$ we were obtaining values far from expected ones from [5], to improve results we have calculated $\phi(k)$ and $\rho(k)$ for $N = 10000$, but averaging over 30 networks. Thus achieving better compromise between computing capabilities and network size. The values are shown in the bottom row of Fig. 1. The results show $\rho(k) < 1$, so Barabási-Albert nerworks with $N = 10000$ lack of rich-club effect.

B. Real networks

Following the same process as in the models section, for each real network $\phi(k)$ and $\rho(k)$ have been calculated. The results will be shown below to discuss what kind of networks present rich-club effect.

In Fig. 3, we can see that the increasing behaviour of Eq. (1) is shared by all kind of networks, like random models. Hubs have a large amount of links in comparison with low-degree nodes, so they have higher probability of being connected with other hubs. We can not see the presence of any organization patterns, global rich-club coefficient gives no information about the system.

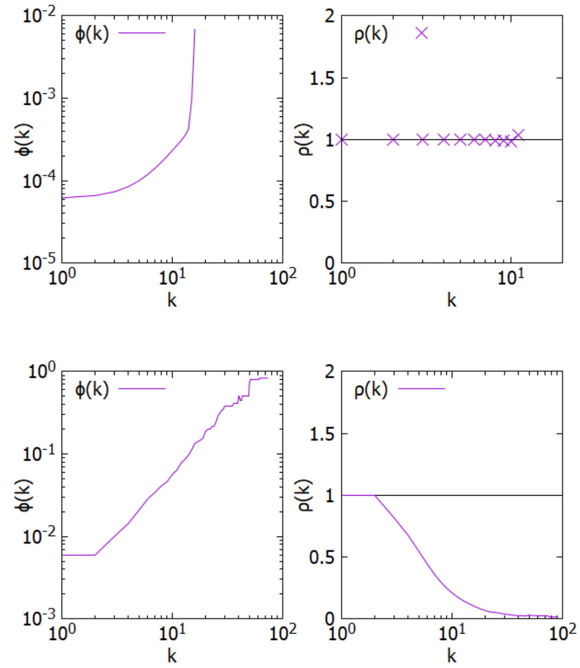

Figure 1: Top row: $\phi(k)$(left) and $\rho(k)$(right) averaged over a set of 100 Erdös-Rényi networks with N=1000. Bottom row: $\phi(k)$(left) and $\rho(k)$(right) averaged over a set of 30 Barabási-Albert networks with N=10000.

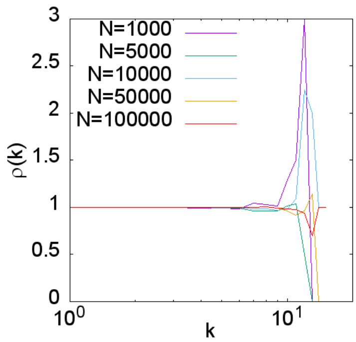

Figure 2: $\rho(k)$ for different values of N for a network generated with Erdös-Rényi model.

Fig. 4 exposes the results obtained for Eq. (2), being possible to determine which networks present rich-club effect. Structural networks show a strong presence of rich-club effect, which implies that this kind of hubs tend to associate strongly between them. Social graphs present the same behaviour for $\rho(k)$. Biological systems display $\rho(k) < 1$, their more important components do not use to interact between them. Internet network shows a weak presence of rich-club effect, denoting that this phenomenon takes place but less than in social and structural networks.

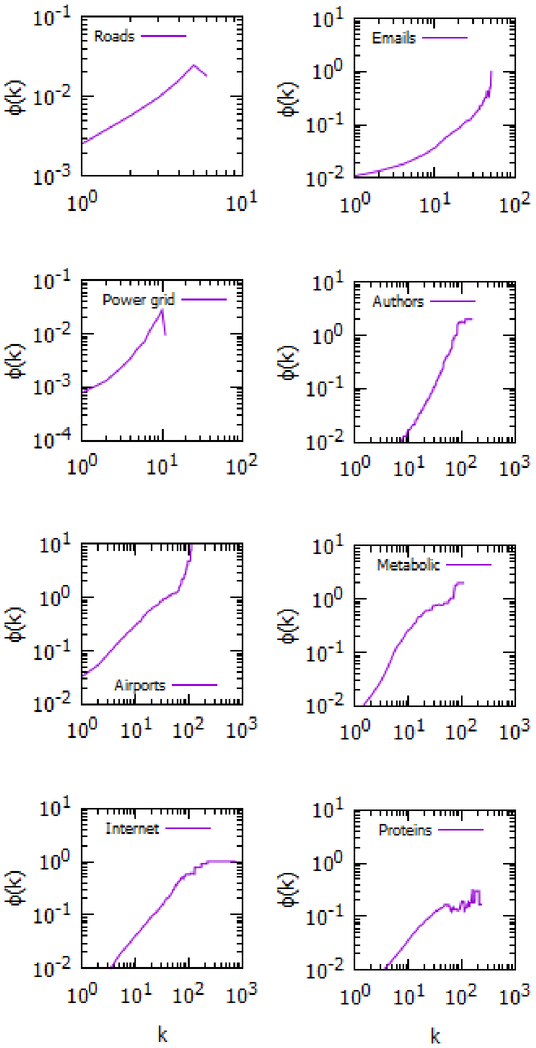

Figure 3: Each graph corresponds to the results obtained for $\phi(k)$. Its behaviour can be observed for each real network.

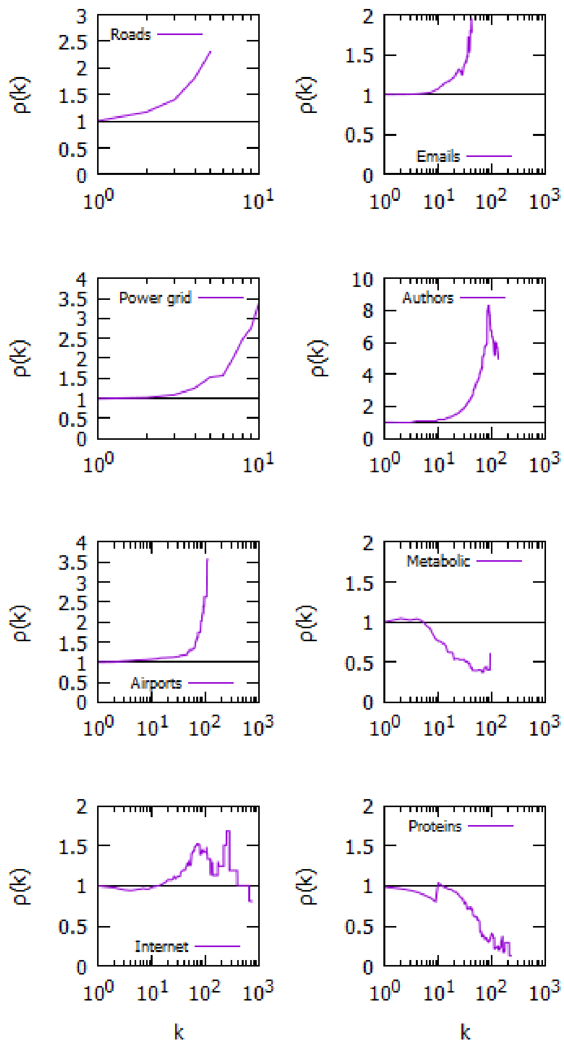

Figure 4: Each graph corresponds to the results obtained for $\rho(k)$. Its behaviour can be observed for each real network.

We can see a strong rich-club effect in social and structural systems. Hubs connectivity in air transportation plays an important role in the study of spreading diseases in global world. This identical result for social networks provides support to the idea that more famous people are likely to know each other. Internet systems with $\rho(k) > 1$ have well-connected hubs, implying that they are highly-sensitive to cyber-attacks. Results in biological graphs may mean that more important components of biological networks tend to participate in diverse functions.

These results give important information about these systems and, furthermore, can be applied to any kind of network to analyze and discuss its hubs properties. Then, if is possible to modelize a system as a graph, hubs behaviour can be studied and lately discussed by calculating its normalized rich-club coefficient.

Results for the local version of the rich-club coefficient are shown in Fig. 5. Due to finite size effects, all networks present a decreasing behaviour of this magnitude for large k. Two kinds of graphs can be observed by analyzing these results: ones where ϕ_{loc} increase with k until reaching a peak and then decrease due to finite-size effects, others where ϕ_{loc} decrease since the beginning. This second kind of behaviour is observed only on networks that have a geographical component, roads and power grid graphs, where nodes distribution in space plays an important role, while the presence of the peak is shared by very different networks.

Without normalizing local rich-club coefficient it tells us if the system organization principles are influenced by geographical distribution. It is important because the global version does not allow to distinguish even between random or real networks. Once they are normalized, both coefficients are equal and allow to discuss the presence or not of rich-club effect, so Eq. (3) is more useful as provides extra information.

IV. CONCLUSIONS

When not normalized, the global rich-club coefficient is not useful to assess whether a network presents the rich-club effect, as it shows the same qualitative behaviour

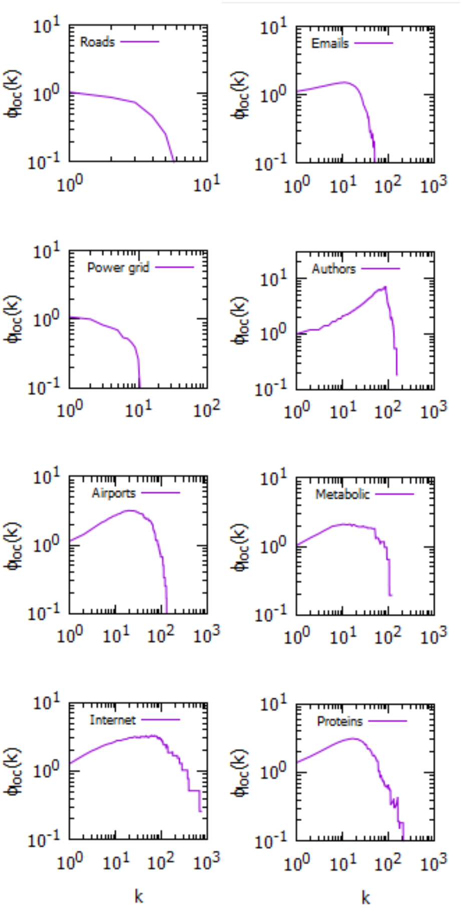

Figure 5: Each graph corresponds to the results obtained for $\phi_{loc}(k)$. Its behaviour can be observed for each real network.

for all networks. Once normalized, the coefficient allows to observe the presence of rich-club effect in model and real networks. As discussed above, knowing if network hubs are well-connected has huge importance in the study of complex systems represented as networks, and can have implications for their function.

Moreover, we have also evaluated a local version of the rich-club coefficient that takes into account the contribution of each node by its degree. The results obtained have shown that the normalized local rich-club coefficient is identical to the corresponding global metric. In contrast, the local measure without normalization is better than the global one because it provides us more information about the analyzed networks. We observed that networks with a geographical component present a characteristic decreasing pattern of the non-normalized rich club as compared to other networks. The calculus of this coefficient may be used to detect geographical importance in networks, specially each ones where it is not clear. It would be important to know the reasons behind those differences in geographical networks and if there is another information that we can obtain analyzing this magnitude. Despite this, it exceeds the aim of this work but could be analyzed in incoming studies.

The networks in our study share two properties, they are undirected and unweighted. To analyze weighted or directed graphs the equations used in this paper should be rewritten taking into account these characteristics, but this is left for future works.

Acknowledgements

I would like to thank M.Ángeles Serrano for her guidance and efforts in this work. Also, i would like to appreciate the support from my family and friends. Finally, thanks to Isaac Roca and Mar Tura for the suggestions and the time dedicated in order to improve this work.

[1] M.E.J.Newman, *Networks:An introduction*, Oxford University press,11-18(2010)
[2] R.Albert,A.-L.Barabasi, *Statistical mechanics of complex networks*, Rev. Mod. Phys. 74, 1-3(2016)
[3] S.N.Dorogovtsev,J.F.F.Mendes, *The shortest path to complex networks*, Arxiv,5-6(2014)
[4] S.Zhou,R.J.Mondragon, *The Rich-Club Phenomenon In The Internet Topology*, IEEE Communications Letters. 8 (3), 180–182(2004)
[5] V.Colizza,A.Flammini,M.A.Serrano,A.Vespignani, *Detecting rich-club ordering in complex networks*, Indiana University,110-114(2006)
[6] P.Erdös,A.Rényi, *On random graphs*, Publ. Math. 6, 290–297(1959)
[7] R.Albert,A.-L.Barabasi, *Emergence of scaling in random networks*,Science 286, 509–512(1999)
[8] L.Šubelj,M.Bajec., *Robust network community detection using balanced propagation*, Eur. Phys. J. B, 81(3), 353-362(2011)
[9] D.J.Watts,M.Bajec., *Collective dynamics of 'small-world' networks*, Nature, 393(1), 440-442(1998)
[10] V.Colizza,R.Pastor-Satorras,A.Vespignani, *Reaction-diffusion processes and metapopulation models in heterogeneous networks* Nature Physics 3, 276-282(2007)
[11] J. Leskovec,J.Kleinberg,C.Faloutsos, *Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations* ACM SIGKDD International Conference on Knowledge Discovery and Data Mining(2005)
[12] R.Guimerà,L.Danon,A.Díaz-Guilera,F.Giralt,A.Arenas, *Self-similar community structure in a network of human interactions* Phys. Rev. E, 68(6), 065103(2003)
[13] J. Leskovec,J. Kleinberg,C. Faloutsos, *Graph Evolution: Densification and Shrinking Diameters*,CM Transactions on Knowledge Discovery from Data (ACM TKDD), 1(1)(2007)
[14] M.Huss,P.Holme, *Currency and commodity metabolites: their identification and relation to the modularity of metabolic networks*,ET Syst. Biol. 1(5), 280-285(2007)
[15] T. Rolland et al.,*A proteome-scale map of the human interactome network*, Cell 159, 1212–1226(2014)
[16] R.Barrat,M.Barthelemy,R.Pastor-Satorras,A.Vespignani, *The architecture of complex weighted networks*, Proceedings of the National Academy of Sciences. 101 (11), 3747–3752(2004)