Rich-club vs rich-multipolarization phenomena in weighted networks

M. Ángeles Serrano

IFISC, Instituto de Física Interdisciplinar y Sistemas Complejos (CSIC-UIB), Campus Universitat Illes Balears,
E-07122 Palma de Mallorca, Spain

Large-scale hierarchies characterize complex networks in different domains. Elements at the top, usually the most central or influential, may show multipolarization or tend to club together, forming tightly interconnected communities. The rich-club phenomenon quantified this tendency based on unweighted network representations. Here, we define this metric for weighted networks and discuss the appropriate normalization which preserves the nodes' strengths and discounts structural strength-strength correlations if present. We find that in some real networks the results given by the weighted rich-club coefficient can be in sharp contrast to the ones in the unweighted approach. We also discuss the ability of the scanning of weighted subgraphs formed by the high-strength hubs to unveil features in contrast to the average: the formation of local alliances in multipolarized environments, or a lack of cohesion even in the presence of rich-club ordering. Beyond structure, this analysis matters for correct understanding of functionalities and dynamical processes relying on hub interconnectedness.

I. INTRODUCTION

A common feature of many real systems is a strongly hierarchical organization which arises at the large scale as a consequence of their microscopic dynamics. As a result, many structural properties describing these systems are far from uniform and show an extreme dispersion, which marks a few of their elements with the highest values as dominant. It is common to extrapolate their prevalence beyond structure to recognize them as the most central, influential, or primal in general terms. In what manner these top elements relate to each other, in particular, whether they are polarized or, on the contrary, show a tendency to club-forming elites or backbones, is an open question that matters for understanding the makeup and performance of the whole system.

In the context of complex network science [1], the quantitative discussion of this issue is known as the rich-club phenomenon. In the network conceptualization of a real system, the most fundamental statistic associated with the elements represented as nodes is the number of neighbors they are connected to, the degree k. In a vast majority of real networks the degree distribution $P(k)$ is very broad and defines a topological hierarchy with rich nodes, those with a high degree, at the top. To detect if they aggregate in a well-interconnected core, a first uniparametric measure, the rich-club coefficient, was proposed as the fraction of edges actually connecting nodes with degree larger than a certain threshold k_T out of the maximum number of connections if they formed a perfect clique [2]. Later on, the original metric was redesigned in order to discount structural effects forcing hubs to be connected without the intervention of special ordering principles [3]. In this way, rich clubs have been found in scientific collaboration networks and in critical infrastructures such as the world air transportation system [3], or in the protein interaction map of the human malaria parasite [4].

However, the first approximation of taking the interactions between pairs of elements i and j as binary, just present or absent as described by the adjacency matrix a_{ij}, turns out to be an oversimplification that in many analyses can distort the interpretation of the results. An interaction can more exactly be quantified by its intensity or weight w_{ij}, and a node can better be characterized by its strength s_i, giving the actual intensity of the interactions it handles and defined as the sum of the weights on the links attached to it.

Here, we explain how to generalize the concept of rich-club ordering to evaluate weighted networks [5,6]. More specifically, our work makes the following contributions. (1) We define the rich-club coefficient in weighted networks as a function of the total weight of links connecting nodes of strength larger than a certain value as compared to the appropriate normalization (Sec. II A). (2) To compute the normalization, we introduce a null model for weighted networks that preserves the nodes' strengths but otherwise produces maximally random networks. This null model can provide a reference value in weighted networks beyond the rich-club coefficient (Sec. II A). (3) We compute analytically the strength-based rich-club coefficient in the uncorrelated limit (Sec. II B). (4) We explore the role of structural strength-strength correlations by introducing two further metrics: the average nearest neighbor strength and the weighted average nearest neighbor strength (Sec. II C). (5) We apply the methodology to three different real networks and find that the consideration of weights can bring results in sharp contrast to those obtained from the degree-based measure (Sec. III). (6) We also discuss the fact that the measurement of the rich-club coefficient is not enough for a complete assessment of the rich-club phenomenon. The averaging character of the metric hides the particulars about the internal organization of the subsets gathering the richest nodes. Their direct inspection, comparing the real weight in each inner link with the reference value given by the null model, uncovers features sometimes contrary to what the overall measure is implying: very fragmented subsets in the presence of rich-club ordering or local alliances when the dominant trend is rivalry or multipolarization (avoidance of connection to each other in a group formed by several) (Sec. IV). (7) Finally, conclusions can be found in Sec. V.

II. DETECTING RICH-CLUB ORDERING IN WEIGHTED NETWORKS

We begin by noticing that the degree-based rich-club coefficient $\rho(k_T)$ of a given graph is computed in two recursive steps. In the first, a simple degree thresholding procedure is applied to produce a hierarchy of nested subgraphs formed of nodes with degrees larger than an increasing threshold k_T.[1] In the second, the number of connections within each subgraph $k > k_T$ is evaluated and compared to the corresponding value in the randomized version of the graph that preserves the degree distribution $P(k)$ [8]. Hence, the rich-club coefficient can be written as [3]

$$\rho(k_T) = \frac{E_{k>k_T}}{E_{k>k_T}^{\mathrm{ran}}}. \tag{1}$$

A ratio larger than 1 indicates the presence of a rich-club phenomenon, high-degree nodes being intertwined with one another more tightly than expected from randomness. In contrast, a ratio less than 1 is a signature of an opposite organizing principle that leads to a lack of interconnectivity among high-degree nodes.

A. Null model that preserves the strength distribution and definition of the rich-club coefficient

The computational procedure for weighted networks needs to redefine the subgraphs and the appropriate null model. There is not a unique choice, since one could, for instance, be interested in a topology where weights and links are randomized, keeping the nodes' strengths and degrees constant [9]. Here we are, however, interested in avoiding the constraint of the actual topology of the real network in order to detect departures from the random counterpart not only in terms of intensities but even in the presence or absence of interactions. A link between hubs predicted by the null model that is actually missing in the real network clearly indicates a tendency contrary to club formation.

Hence, we focus exclusively on strengths and assume that the rich nodes are those with the highest values. The appropriate thresholding procedure applied to the weighted networks generates then a hierarchy of nested subgraphs of nodes with strengths larger than an increasing threshold s_T, and the sum of the weights on the links within each subgraph, $W_{s>s_T}$, is considered. Regarding the normalization, we propose to compare to a randomized version of the graph that preserves the strength distribution $P(s)$. This null model can be achieved by approximating the weights in the network by integers, so that they could be considered as multiple connections formed by decoupled links. Then, the usual randomization based on rewirings [8] can be done, avoiding self-connections but not multiple ones. In this way, the nodes maintain their strength but the weights in the links (or the degrees) can change in the process. To avoid inducing corre-

lations, notice that each decoupled link should be selected independently with the same probability. This ensures that in the steady state the weights agree with the expected values in a strength-preserving but otherwise maximally random conformation. Formally, the rich-club coefficient in the weighted approach can be written as

$$\rho(s_T) = \frac{W_{s>s_T}}{W_{s>s_T}^{\mathrm{ran}}}. \tag{2}$$

B. Analytical computation of the uncorrelated limit in the weighted approach

Strengths are less prone than degrees to be affected by structural constraints [10,11]. This fact makes meaningful the consideration of the uncorrelated limit in the weighted approach,

$$\rho^{\mathrm{unc}}(s_T) = W_{s>s_T}/W_{s>s_T}^{\mathrm{unc}}, \tag{3}$$

in which the strengths of attached nodes are independent. The normalization $W_{s>s_T}^{\mathrm{unc}}$ can be computed analytically. In the same spirit of the original measure of Zhou and Mondragon [2], it is given by the sum of the uncorrelated weights in fully connected subsets, which can be calculated just by taking into account that, on average, they must be proportional to the product of the strengths of the nodes i and j they are associated with [11]. If loops are not allowed,

$$W_{s>s_T}^{\mathrm{unc}} = \sum_{\nu_i} \sum_{\nu_{j\neq i}} w_{ij}^{\mathrm{unc}} = \langle s \rangle \frac{\displaystyle\sum_{\nu_i}\sum_{\nu_{j\neq i}} s_i s_j}{N\langle s\rangle^2 - \langle s^2\rangle}, \tag{4}$$

where ν_i designates the subsets of nodes such that $s_i > s_T$, N is the total number of nodes in the network, and $\langle s \rangle$ and $\langle s^2 \rangle$ are the first and second moments of the strength distribution.

C. Structural strength-strength correlations

As happens for the degrees [10], in some networks closure conditions enforce the presence of correlations between the strengths of connected pairs that cannot be avoided even in maximally random configurations. Equation (2) discounts these structural strength-strength correlations and other higher-order effects which are not removed by the random procedure. However, the uncorrelated approximation assumes a total absence of dependencies between strengths and it is not perfectly valid if structural strength-strength correlations are present. To help to discern in which networks these are important, we define—from the formalism in [12] and in analogy to the average nearest neighbor degree [13]—the average nearest neighbor strength and the weighted average nearest neighbor strength, both as functions of the strength,

$$\bar{s}_{nn}(s) = \frac{1}{N_s}\sum_{i\in s,j}\frac{a_{ij}}{k_i}s_j, \quad \bar{s}_{nn}^w(s) = \frac{1}{N_s}\sum_{i\in s,j}\frac{w_{ij}}{s}s_j, \tag{5}$$

N_s being the number of nodes with strength s. The average $\bar{s}_{nn}(s)$ is coupled to the underlying degree structure and com-

---

[1]This nested hierarchy of subgraphs turns out to have self-similarity properties for some real scale-free networks such as the Internet at the autonomous system level; see [7].
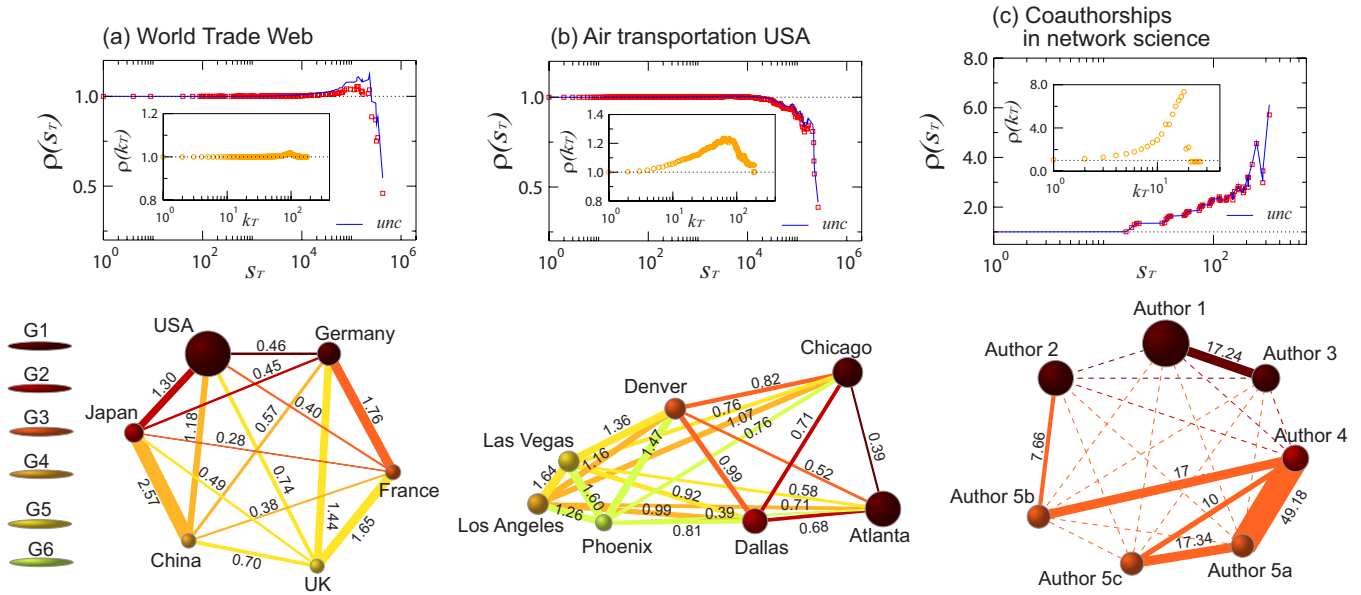
FIG. 1. (Color online) Weighted rich-club phenomenon in the real networks WTW, USAN, and SCN. AT the top, graphs for the strength-based rich-club coefficient. Dotted curves correspond to $\rho(s_T)$ normalized and averaged over 100 randomizations (rewirings) of the original network. Weights have been discretized in all cases: rounded off to the nearest integer in the WTW, coarse grained to hundreds and rounded off to the nearest integer in the USAN, and divided by the minimum weight and rounded off to the nearest integer in the SCN. Solid lines in the plots represent the analytical curves for the uncorrelated approximation $\rho^{\mathrm{unc}}(s_T)$. The insets show the degree-based rich-club $\rho(k_T)$ of the unweighted representation. At the bottom, sketches showing greater detail of how the hubs interact among themselves. Darker colors represent inner subsets in the nested hierarchy as defined by the threshold strength. Within a plot, the sizes of the nodes are proportional to their strengths. The numerical values labeling the links represent the ratio of the actual weight of the tie to its average value in the randomized versions.

putes strength-strength correlations that are structural and cannot be destroyed by the randomization. In contrast, $\bar{s}_{nn}^{w}(s)$ is disentangled from degrees and expected to be perfectly flat in the maximally random case. Both measures combined enable the discrimination of structural and nonstructural strength-strength correlations in weighted networks, setting up the validity of the uncorrelated approximation.

III. REAL NETWORKS

In the graphs of Fig. 1 (top), we report the behavior of $\rho(s_T)$ and $\rho^{\mathrm{unc}}(s_T)$ in three different real networks, and for comparison we also provide the curves for the degree-based definition $\rho(k_T)$ (insets). These examples correspond to (a) the world trade web (WTW) of commercial relationships between states [14] in 2000, where the weights give the annual merchandise exchanges in millions of current-year U.S. dollars ([15]), (b) the domestic segment of the U.S. airport network (USAN) for the year 2006,[16] where the weights are given by numbers of passengers [6], and (c) an extract of the actual scientific collaboration network (SCN) of researchers in the area of complex networks [17], where the weights represent the intensity of the collaborative ties depending on the number of coauthored papers and the number of authors in each. To validate our methodology, we also generated a maximally random network at the weighted level with a size of 2×10^4 nodes and a strength distribution $P(s) \sim s^{-1.85}$, making use of the weighted configuration model (WCM); see [11] and references therein. By construction and as expected, its strength-based rich-club coefficient does not detect any ordering but has a value of 1 in the whole domain (plot omitted for brevity).

Surprisingly, very different behaviors of $\rho(k_T)$ and $\rho(s_T)$ can be detected in networks where the degree and strength are not trivially related. Two instances are the WTW and the USAN, which have a neutral (meaning that the unweighted representation is dominated by structural connectivity effects) and a mild degree-based rich-club ordering, respectively, but exhibit a decreasing strength-based coefficient, providing evidence of a clear rich-multipolarization phenomenon. Their oligarchies of rich nodes are on average loosely interconnected in terms of weight as compared with the random null model counterpart, in contrast to the rich-club situation. On the other hand, the degree-based and strength-based spectra of networks with weights uncorrelated with degrees are expected to be qualitatively similar. This is what happens in the SCN case. The presence of a strong rich-club ordering in both the unweighted and the weighted representations seems to provide support to the idea that the more collaborative (and, by extrapolation, maybe the more influential) researchers in complex network science tend to club following an expected tendency in social systems [as we will discuss below, maybe excluding the very top hubs as suggested by the sharp decay of $\rho(k_T)$ for very high degrees].

Turning now to evaluate the goodness of fit between the uncorrelated approximation $\rho^{\mathrm{unc}}(s_T)$ and $\rho(s_T)$, these measures match almost perfectly in the USAN and the SCN cases. The agreement is slightly worse for the WTW and in the case of the simulated WCM the approximation is clearly
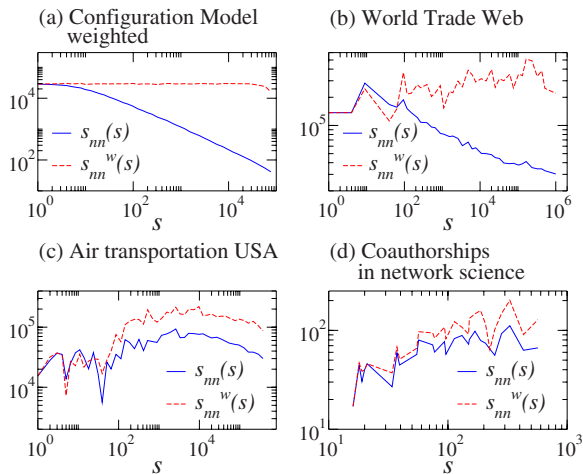
FIG. 2. (Color online) Weighted strength-strength correlation structure of the real and simulated networks. $\bar{s}_{nn}(s)$ for the WTW and WCM are power laws with negative exponents, larger in the latter case, which indicates disassortative strength mixing.

bad (plot omitted for brevity). The explanation for this divergence can be found in the presence of structural strength-strength correlations. In Fig. 2, we report these functions for the real and simulated networks. Disassortative [18] structural strength correlations measured by $\bar{s}_{nn}(s)$ are important for the WTW but much more so for the WCM, limiting the validity of the uncorrelated approximation in these cases.

IV. ANALYSIS OF THE INNER SUBGRAPHS

We have now a methodology to detect rich-club ordering in weighted networks which provides valuable information about the average tendency in the interactions between hubs. But averages can hide local regularities in contrast to the overall behavior, and a thorough assessment of the property requires greater detail. A direct inspection of the subgraphs formed by the hubs, comparing the actual weight in each link with the reference value given by the null model, is necessary. If, for a given link, the ratio of the two values is above 1, the two nodes connected by that link show a tendency to interact, while a ratio below 1 denotes the opposite tendency, to avoid interaction. In addition, stronger polarization is manifested when a link between two nodes is absent in the real network but has a certain expected weight according to the null model.

The results for the real networks under study are summarized in the sketches of Fig. 1 (bottom). The WTW shows, for instance, a very clear and sharp rich-multipolarization phenomenon. The function $\rho(s_T)$ runs at the value 1 just until the tail end, where a group of rich nodes (the United Kingdom, China, France, Japan, Germany, and the United States, in increasing order of strength) form subsets with $\rho(s_T) < 1$. Notice the overlap of five members—all except China—with the club of the seven largest industrialized and richest countries in the world, the G7, which also includes Italy and Canada (the next two countries in the strength hierarchy of the WTW in 2000). The decreasing tail indicates that these hub countries share on average less weight among them than

expected in the random situation. So they seem to have an aversion to connecting to each other, as anticipated between powers in a strongly competitive economic system. However, they interact anyway, forced by structural constraints, and form fully connected subsets. What is more interesting, the biggest world economies seem to be polarized into two connected blocks in direct competition: the United States and its Asian allies Japan and China, against Europe (France, Germany, and the United Kingdom). Each block is tightly connected, shown by trading volumes larger than random predictions, and competition between the two is exposed by the reduced exchange of merchandise as compared to the null model.

The situation for the USAN is similar in the sense that it presents a distinct rich-multipolarization phenomenon [although the transition from $\rho(s_T) = 1$ to $\rho(s_T) < 1$ is smoother] and that the hub airports form fully connected subgraphs. In the sketch we show the seven largest airports in numbers of domestic passengers. Again, although the overall tendency is a reduced interaction reflecting competition, examination of the subsets shows tight interconnections among the big west coast and central airports, Los Angeles, Phoenix, Las Vegas, and Denver, and a neutral interaction with Dallas, while the interactions with Chicago and Atlanta in the east are weak. Chicago and Atlanta are the biggest hubs with a weak tie to each other and to the rest. An explanation can be found in the geographical layout. While Chicago and Atlanta are far away and surrounded by nonoverlapping basins of attraction of smaller airports and passengers, Los Angeles, Phoenix, and Las Vegas are close enough to be forced to share approximatively the area of influence so that they interact with each other more than expected. Surely, other aspects such as the provision of connections to international flights or operational constraints of the biggest air companies can also have a role.

Finally, the SCN illustrates the opposite situation. This is a network with a strong rich-club ordering in both the degree-based and strength-based approaches, which seems to suggest that researchers tend to form collaborative groups. However, this does not necessarily mean a high degree of cohesiveness among all the most collaborative scientists. Indeed, the inner subgraphs of the network are actually very sparse, even though they are expected to be fully connected according to the random null model. This points to a certain level of rivalry between the hubs. They form ties with just very few other hubs, but in exchange the interactions are very strong and enough to determine the averages.

We would like to emphasize that, in all three cases, these results are corroborated and in good agreement with other sources of information about the system: commercial agreements in the case of the WTW, geographical layout in the case of the airport network, and strongly competitive character of collaborations in the field of complex networks science.

V. CONCLUSIONS

These findings have important consequences. At the theoretical level, the detection of rich-club ordering is depen-

dent on whether the intensities of the interactions among elements are taken into account in systems where intensities and number of interactions are related in a nontrivial form. In addition, an exhaustive assessment of the property requires greater detail than can be achieved by the averaging coefficient. The scanning of the subgraphs formed by the hubs as compared to the appropriate null model is also relevant, and can uncover the formation of local alliances in multipolarization environments or a lack of cohesion even in the presence of rich-club ordering. Beyond structure, this analysis matters for understanding functionalities and dynamical processes relying on hub interconnectedness and, in a broader context, may help explain how primary forces such as competition and cooperation influence collective form and performance.

Note added. Recently, other researchers proposed an alternative null model in order to detect rich-club ordering in weighted complex networks; see [19].

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

[2] S. Zhou and R. J. Mondragon, IEEE Commun. Lett. **8**, 180 (2004).

[3] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Nat. Phys. **2**, 110 (2006).

[4] S. Wuchty, PLoS One **2**, e335 (2007).

[5] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).

[6] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).

[7] M. A. Serrano, D. Krioukov, and M. Boguñá, Phys. Rev. Lett. **100**, 078701 (2008).

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

[9] K. Bhattacharya, G. Mukherjee, J. Saramäki, K. Kaski, and S. S. Manna, J. Stat. Mech.: Theory Exp. (2008), P02002.

[10] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Eur. Phys. J. B **38**, 205 (2004).

[11] M. A. Serrano and M. Boguñá, in *Science of Complex Networks: from Biology to the Internet and WWW; CNET 2004*, edited by J. F. F. Mendes, S. N. Dorogovtsev, A. Povolotsky, F. V. Abreu, and J. G. Oliveira, AIP Conf. Proc. No. 776 (AIP, Melville, NY, 2005), p. 101.

[12] M. A. Serrano, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E **74**, 055101(R) (2006).

[13] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[14] M. A. Serrano and M. Boguñá, Phys. Rev. E **68**, 015101(R) (2003).

[15] K. S. Gleditsch, J. Conflict Resolut. **46**, 712 (2002); http://weber.ucsd.edu/ kgledits/exptradegdp.html

[16] http://transtats.bts.gov

[17] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

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

[19] T. Opsahl, V. Colizza, P. Panzarasa, and J. J. Ramasco, e-print arXiv:0804.0417.