Prominence and Control: The Weighted Rich-Club Effect

Tore Opsahl,[1] Vittoria Colizza,[2] Pietro Panzarasa,[1] and José J. Ramasco[2]

[1]*School of Business and Management, Queen Mary College, University of London, London, United Kingdom*
[2]*Complex Systems Lagrange Laboratory, Complex Networks, ISI Foundation, Turin, Italy*

Complex systems are often characterized by large-scale hierarchical organizations. Whether the prominent elements, at the top of the hierarchy, share and control resources or avoid one another lies at the heart of a system's global organization and functioning. Inspired by network perspectives, we propose a new general framework for studying the tendency of prominent elements to form clubs with exclusive control over the majority of a system's resources. We explore associations between prominence and control in the fields of transportation, scientific collaboration, and online communication.

Research has long documented the abundance of systems characterized by heterogeneous distribution of resources among their elements [1,2]. Back in 1897, Pareto noticed the social and economic disparity among people in different societies and countries [1]. This empirical regularity inspired the 80-20 rule of thumb stating that only a select minority (20%) of elements in many real-world settings are responsible for the vast majority (80%) of the observed outcome. While recent studies have examined the tendency of prominent elements to establish connections among themselves [3], how they leverage on their connections to gain and maintain control over resources circulating in a system still remains largely unexplored. In particular, do they collude and choose to exchange a disproportionately large amount of resources among themselves rather than with others? Or does competition prevent them from deepening the connections they have with one another? To answer these questions, we need to test for the tendency of prominent elements to engage in stronger or weaker interactions among themselves than expected by chance. We call this tendency the weighted rich-club effect. In this Letter, we adopt the framework of network theory—where the elements of the system are seen as nodes and the links among the elements represent interactions [4–8]—and provide a novel method to properly assess this tendency.

Previous work has focused on highly connected nodes and the degree to which they preferentially interact among themselves [3]. This feature is known as the rich-club phenomenon [3,9], a metaphor that alludes to the tendency of the highly connected nodes (i.e., the rich nodes) to establish more links among themselves than randomly expected. Evidence of the phenomenon has been reported for scientific collaboration networks [3], transportation networks [3], and interbank networks [10]. Conversely, research has shown that highly connected routers on the Internet tend not to be connected with one another [3], whereas the pattern of interactions among proteins has been found to depend on the particular organism under consideration [3,11]. Although they uncover interesting

structural aspects of the systems, these studies are limited in that they only detect whether or not links among prominent nodes are present. In so doing, they neglect a crucial piece of information encoded in the weight of links, which is a measure of their intensity, capacity, duration, intimacy, or exchange of services [12,13]. A full understanding of how systems are organized requires a shift towards a new paradigm that allows us to evaluate whether nodes that rise to network prominence also tend to exchange among themselves the majority of the resources flowing within the network.

To this end, we rank all nodes of a system in terms of a richness parameter r. For each value of r, we select the group (the club) of all nodes whose richness is larger than r. We thus obtain a series of increasingly selective clubs. For each of these clubs, we count the number $E_{>r}$ of links connecting the members, and measure the sum $W_{>r}$ of the weights attached to these links [Fig. 1(a)]. We then measure the ratio $\phi^w(r)$ between $W_{>r}$ and the sum of the weights attached to the $E_{>r}$ strongest links within the whole network [Fig. 1(b)]. Formally, we have

$$\phi^w(r) = \frac{W_{>r}}{\sum_{l=1}^{E_{>r}} w_l^{\mathrm{rank}}}, \tag{1}$$

where $w_l^{\mathrm{rank}} \geq w_{l+1}^{\mathrm{rank}}$ with $l = 1, 2, \ldots, E$ are the ranked weights on the links of the network, and E is the total number of links. Equation (1) thus measures the fraction of weights shared by the rich nodes compared with the total amount they could share if they were connected through the strongest links of the network. Other measures can be introduced that depend on the local network structure surrounding the rich nodes [3,14–16]. Here we aim at investigating the extent to which the prominent nodes control the flow of resources over the whole system.

In analogy with the topological rich-club measure [3,17], a high value of $\phi^w(r)$, however, is not in itself sufficient to account for an actual tendency of the rich nodes to preside over the strongest links. This is due to the fact that even networks where links are randomly established could display a nonzero value of $\phi^w(r)$. To
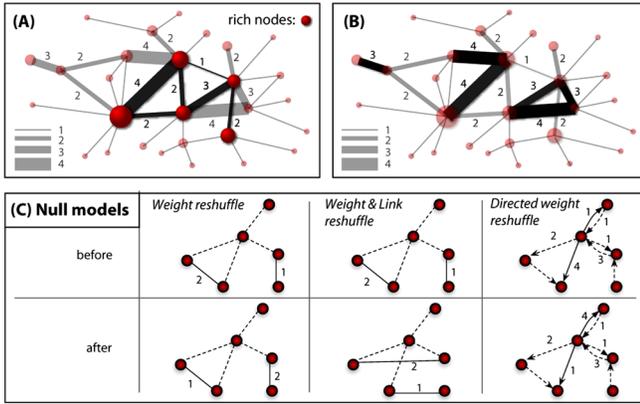
FIG. 1 (color online). (a),(b) Schematic representation of a weighted network, with size of nodes proportional to their richness, and width of links to their weight indicated by the corresponding numbers. Several definitions of richness can be considered. (a) The nodes and links in the rich club are highlighted, giving $E_{>r} = 6$ links and $W_{>r} = 4 + 2 + 2 + 3 + 1 + 2 = 14$. (b) The strongest $E_{>r} = 6$ links of the network are highlighted, yielding the following value for the denominator of Eq. (1): $\sum_{l=1}^{E_{>r}} w_l^{\text{rank}} = 4 + 4 + 4 + 3 + 3 + 3 = 21$. We thus obtain $\phi^w(r) = 14/21$. (c) Null models. Solid lines refer to the links reshuffled; the numbers refer to their weight.

assess the actual presence of the weighted rich-club phenomenon, discounted of random expectations, $\phi^w(r)$ must be compared with an appropriate benchmark. To this end, we introduce a null model that is random but at the same time comparable to the real network. In particular, this model should break the associations between weights and links while preserving some crucial features of the network encoded in its degree distribution $P(k)$ (i.e., the probability that a given node is connected to k neighbors) and weight distribution $P(w)$ (i.e., the probability that a given link has weight w). In addition, the nodes in the rich club must be the same as in the real network, which also preserves the richness distribution $P(r)$ (i.e., the probability that a given node has richness r).

In what follows, we introduce three procedures for constructing null models [see Fig. 1(c)] that correspond to different ways of preserving $P(r)$, depending on the choice of the richness parameter r. In this Letter, we explore three possible definitions of r: the degree k, the strength s (i.e., the sum of the weights attached to the links originating from a node) [12], and the average weight $\bar{w}$ (i.e., the ratio between s and k) [18]. If the richness of a node is given by its degree, we adopt the following two randomization procedures. First, the weight reshuffle procedure consists simply in reshuffling the weights globally in the network, while keeping the topology intact. Second, the weight and link reshuffle procedure, which introduces a higher degree of randomization, consists in reshuffling also the topology, while preserving the original degree distribution $P(k)$ [6,19]. Weights are automatically redistributed by remaining attached to the reshuffled links. Both randomization

procedures can be easily generalized to directed networks. The weight and link reshuffle procedure, mixing the signal coming from the topology with that generated by the location of weights, is considered here to assess the effects of higher degrees of randomization on the results, as well as for the sake of comparison with the topological rich-club coefficient [3].

Inevitably, since weights are reshuffled globally, both procedures produce null models in which the nodes do not maintain the same strength s as in the real network. When node richness is defined in terms of s, we need to introduce a third procedure that preserves this quantity. We construct a null model based on the randomization of directed networks [20] that preserves not only the topology and $P(w)$, but also the strength distribution $P(s)$ (i.e., the probability that a given node has strength s) of the real network. To this end, we reshuffle weights locally for each node across its outgoing links (directed weight reshuffle procedure). In so doing, we also obtain null models where the average weight $\bar{w}$ of outgoing links is kept invariant. We extend this procedure to the undirected case by duplicating an undirected link into two directed links, one in each direction.

For a given definition of the richness r, the weighted rich-club effect can be detected by measuring the ratio

$$\rho^w(r) = \frac{\phi^w(r)}{\phi^w_{\text{null}}(r)},\qquad(2)$$

where $\phi^w_{\text{null}}(r)$ refers to the weighted rich-club effect assessed on the appropriate null model. When ρ^w is larger than 1, the original network has a positive weighted rich-club effect, with rich nodes concentrating most of their efforts towards other rich nodes compared with what happens in the random null model. Conversely, if it is smaller than 1, the links among club members are weaker than randomly expected.

In order to examine the applicability of our method, we study three real-world networks drawn from different domains. (i) The U.S. airport network, obtained from the U.S. Department of Transportation [21], composed of 676 commercial airports and 3523 routes connecting them. Each weight corresponds to the average number of seats per day available on the flights connecting two airports [12,22]. (ii) The scientific collaboration network [23], extracted from the arXiv [24] electronic database in the area of condensed matter physics, from 1995 to 1999. Nodes represent scientists and a link exists between two scientists if they have coauthored at least one paper. Link weight reflects the authors' contribution in their collaboration [23]—the larger the number of authors collaborating on a paper, the weaker their interaction. (iii) The online social network [25] comprising 59 835 directed online messages exchanged among 1899 college students at the University of California, Irvine, from April to October 2004. Link weight is the number of messages sent from one student to another.

We begin by defining network prominence in terms of node degree. In this case, $r = k$. We examine whether the highly connected nodes control the exchange of resources. For the three networks, Fig. 2 (left column) reports the weighted rich-club ratio and its topological counterpart (inset). With only a mild topological effect [3], the airport network shows a strong weighted rich-club effect, as can be identified from the remarkable growth of ρ^w as a function of the degree of the airports. This finding agrees with previous studies that reported the presence of non-trivial correlations between weight of the links and degrees of the nodes [12,22,26]. Connections among hub airports, with flights to many destinations, are characterized by large travel fluxes. Different results are found for the scientific collaboration network: while there is evidence of a strong positive topological rich-club effect, the network does not display a weighted one. As shown in Fig. 2, ρ^w remains flat around 1 for almost the whole range of k. The authors that have many collaborators tend to work together. However, the intensity of their collaboration does not differ from what is randomly expected, thus providing additional support to the observed lack of correlations between collaboration intensity and number of collaborators [27,28]. Finally, the weighted and topological rich-club effects display strikingly different trends for

the online social network. Very gregarious individuals, with a large number of contacts, poorly communicate with one another. However, when they do, they choose to forge stronger links than randomly expected.

To investigate how different definitions of prominence might affect the results, we restricted our attention to a subset of the arXiv collaboration network based on the publications on network science [29]. In Figs. 3(a) and 3(b), we mapped the interaction patterns within the clubs obtained by defining r in terms of the degree k (number of coauthors) and the strength s (number of published papers), respectively. In this network, each paper corresponds to a fully connected group of collaborators. When a paper is cowritten by a large number of authors, these authors acquire a high degree and thus increase their chances of becoming members of the club based on k. Large collaborations tend to secure club membership, yet generate weaker links than smaller ones [23]. Experimental papers on biological networks are authored by a large number of scientists, and therefore only few such papers may suffice to substantially increase the topological rich-club effect [see the very large clique in Fig. 3(a)]. By contrast, they bring about weaker links than small-scale collaborations, thereby reducing their contribution to the weighted rich-club effect. However, when network prominence is defined in terms of s, club members as well as their interaction
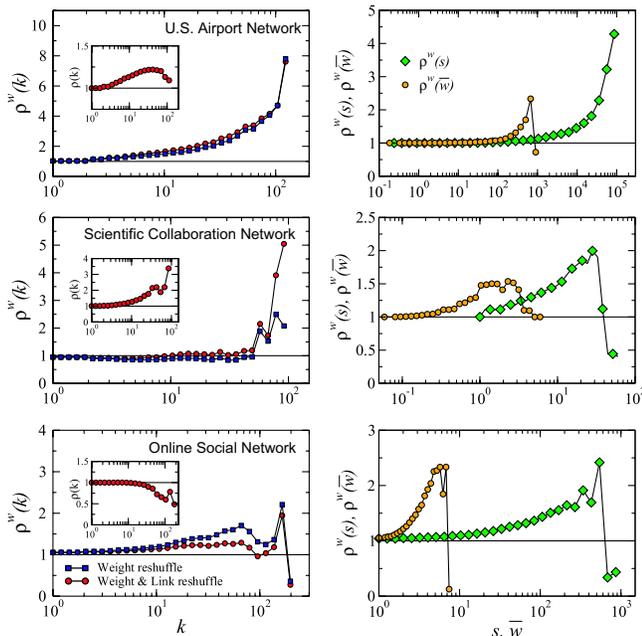

FIG. 2 (color online). Weighted rich-club effect in the U.S. airport network (top), the scientific collaboration network (center), and the online social network (bottom). Left column: $r = k$. The insets refer to the topological rich-club coefficient $\rho(k)$ [3], defined as the ratio between $\phi(k)$ (i.e., the fraction of links connecting rich nodes, out of the maximum possible number of links among them) [9] and $\phi_{null}(k)$ [i.e., $\phi(k)$ measured on the corresponding weight and link reshuffle null model]. Right column: $r = s$ (diamonds) and $r = \bar{w}$ (circles).
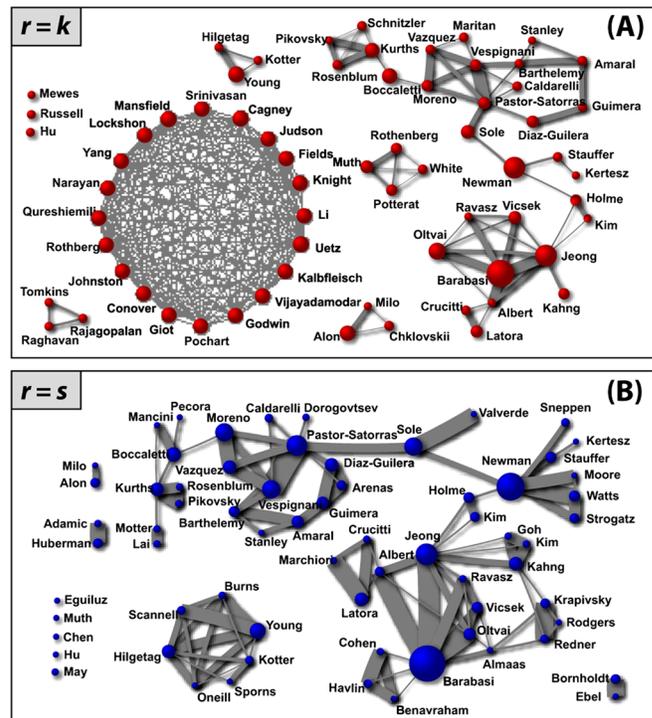

FIG. 3 (color online). Subsets of the rich nodes in the network science collaboration network [29] based on degree (a) and strength (b). Only links among the rich nodes are shown. The size of the nodes is proportional to their richness; the width of the links is proportional to their weight.

patterns substantially change with respect to the case in which $r = k$ [Fig. 3(b)].

The next step is thus to define network prominence in terms of node strength s and shift our attention from the most connected to the most involved nodes in the system's activity. Our findings show that active nodes preferentially direct their efforts towards one another, and this tendency becomes more pronounced as the involvement of nodes in network activity increases (see Fig. 2, right column). Not only do busy airports direct routes to one another, but they also secure control over travel fluxes by channeling on those routes a larger proportion of their passengers than randomly expected. This behavior is in sharp contrast with what was found using a different null model [14], a pointed reminder of the crucial role played by such models in assessing the rich-club effect. When scientists are chosen on the basis of their scientific productivity s, exclusive clubs emerge in which scientists tend to collaborate with one another to a greater extent than randomly expected, unlike what was found within the club of the most connected scientists. In the online social network, highly active users tend to communicate with one another more frequently than would be the case if contacts were chosen at random.

While node strength gives a general indication of how involved a node is in the activity of a network, it does not allow us to discriminate between nodes with a large number of weak links and nodes with a small number of strong links. To address this issue, we define the richness parameter in terms of the average weight $\bar{w}$. We find positive effects in all networks (see Fig. 2, right column). Airports that optimize the traffic per link tend to direct their busy routes to one another. Scientists that show the ability to maximize their resources per collaboration tend to intensively collaborate with one another. Online communication tends to occur among users that maximize the attention directed to each contact.

By shifting focus from the network topology to the weight of links, we have proposed a new general framework for studying the tendency of prominent nodes to attract and exchange among themselves the majority of the resources available in a system. Unlike a merely topological assessment of the network, our method allows us to uncover organizing principles that would otherwise remain undetected. In addition, by varying the definition of prominence, we found evidence of different organizing principles and paved the way towards a deeper understanding of the multiple layers of a system's organization. Our method is widely applicable, in that it enables us to study the control benefits of prominent elements in a variety of empirical settings, by decoupling prominence from strictly network properties. To the extent that the components of a system can be sorted into a hierarchy according to a given property, our framework suggests several new ideas for future research, including the impact of performance, cen-trality, status, age, and size on the ability of elements to control flows of resources. In this respect, our study helps shed new light on the global organization of complex systems.

[1] V. Pareto, *Cours d'Economie Politique* (MacMillan, Paris, 1897).
[2] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).
[3] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani, Nature Phys. **2**, 110 (2006).
[4] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).
[5] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks: From Biological Nets to the Internet and WWW* (Oxford University, New York, 2003).
[6] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).
[7] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, England, 2004).
[8] G. Caldarelli, *Scale-Free Networks: Complex Webs in Nature and Technology* (Oxford University, New York, 2007).
[9] S. Zhou and R. J. Mondragon, IEEE Commun. Lett. **8**, 180 (2004).
[10] G. De Masi, G. Iori, and G. Caldarelli, Phys. Rev. E **74**, 066112 (2006).
[11] S. Wuchty, PLoS ONE **2**, e355 (2007).
[12] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, Proc. Natl. Acad. Sci. U.S.A. **101**, 3747 (2004).
[13] M. Granovetter, Am. J. Sociology **78**, 1360 (1973).
[14] M. A. Serrano, arXiv:0802.3122.
[15] S. Valverde and R. V. Solé, Phys. Rev. E **76**, 046118 (2007).
[16] V. Zlatic *et al.*, arXiv:0807.0793.
[17] L. A. N. Amaral and R. Guimerà, Nature Phys. **2**, 75 (2006).
[18] J. J. Ramasco and S. Morris, Phys. Rev. E **73**, 016122 (2006).
[19] M. Molloy and B. Reed, Random Struct. Algorithms **6**, 161 (1995).
[20] M. A. Serrano, M. Boguñá, and A. Vespignani, J. Econ. Interact. Coord. **2**, 111 (2007).
[21] http://www.transtats.bts.gov/.
[22] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proc. Natl. Acad. Sci. U.S.A. **102**, 7794 (2005).
[23] M. E. J. Newman, Phys. Rev. E **64**, 016132 (2001).
[24] http://www.arxiv.org/.
[25] P. Panzarasa, T. Opsahl, in *Proceedings of the XXVII International Sunbelt Social Network Conference, Corfu Island, Greece, 2007* (INSNA, Buffalo, NY, 2007).
[26] Z. Wu *et al.*, Phys. Rev. E **74**, 056104 (2006).
[27] J. J. Ramasco and B. Gonçalves, Phys. Rev. E **76**, 066106 (2007).
[28] J. J. Ramasco, Eur. Phys. J. Special Topics **143**, 47 (2007).
[29] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).