



Rich-club ordering and the dyadic effect: Two interrelated phenomena



Matteo Cinelli ^{*}, Giovanna Ferraro, Antonio Iovanella

Department of Enterprise Engineering, University of Rome "Tor Vergata", Via del Politecnico, 1 - 00133 Rome, Italy

HIGHLIGHTS

- We prove how rich-club ordering and dyadic effect are strongly interrelated.
- Rich-club ordering can be studied using the dyadic effect measures.
- We improve the formulation of the rich-club phenomenon by means of the dyadic effect.
- We provide a baseline for the evaluation of the rich-club size.

ARTICLE INFO

Article history:

Received 9 March 2017

Received in revised form 19 June 2017

Keywords:

Rich-club phenomenon
Dyadic effect
Complex networks

ABSTRACT

Rich-club ordering and the dyadic effect are two phenomena observed in complex networks that are based on the presence of certain substructures composed of specific nodes. Rich-club ordering represents the tendency of highly connected and important elements to form tight communities with other central elements. The dyadic effect denotes the tendency of nodes that share a common property to be much more interconnected than expected. In this study, we consider the interrelation between these two phenomena, which until now have always been studied separately. We contribute with a new formulation of the rich-club measures in terms of the dyadic effect. Moreover, we introduce certain measures related to the analysis of the dyadic effect, which are useful in that they confirm the presence and relevance of rich-clubs in complex networks and provide certain insights and a baseline for the evaluation of the rich-club size. In addition, certain computational experiences show the usefulness of the introduced quantities with regard to different classes of real networks.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Complex networks are a valid framework of representation and way to analyze a wide variety of real world phenomena. Thanks to a rapid surge of interest in the field, a set of measures able to capture and quantify the peculiarities of various heterogeneous networks have been subsequently introduced, in order to verify empirical observations at different levels of detail. The majority of these measures focus on particular tendencies (dynamics) in nodes linkage and the related emerging phenomena such as assortative mixing [1–3], homophily [4], and the rich-club phenomenon [5,6].

The *rich-club phenomenon* has been defined as the tendency of nodes with a high centrality (usually degree) to form highly interconnected communities. It was initially introduced in [6] in order to analyze the Internet topology at the Autonomous

^{*} Corresponding author.

E-mail addresses: matteo.cinelli@uniroma2.it (M. Cinelli), giovanna.ferraro@uniroma2.it (G. Ferraro), antonio.iovanella@uniroma2.it (A. Iovanella).

Systems (AS) level and to provide a reasonable explanation as to why such kinds of networks include tightly interconnected hubs.

Commonly recognized as a power-law network [7], the Internet highlights one of the main properties of such degree distributed networks; i.e. the presence of a small amount of nodes having a surprisingly large number of links, the *rich nodes*. Defining the rich-club coefficient $\phi(k)$, Zhou and Mondragon [6] revealed that the Internet topology displays the rich-club phenomenon, and that such tendency cannot be completely expressed by regular models that generate power-law degree distribution such as the BA-model [7], the fitness model [8], and the Inet-3.0 Model [9]. Furthermore, they noted that well known network measures, like degree assortativity, cannot detect this phenomenon (for instance some degree disassortative networks show the rich-club phenomenon). For this reason, studies that aim to recognize the connection patterns among rich nodes are crucial, in order to provide important insights into a number of network structures ranging from computer to chemical and social sciences [6,10].

This topic gained additional interest in the complex networks literature [11–14] being applied in the analysis of other structural properties, as well as in more recent studies concerning the structural connectivity in the brains of both animals and human beings [15–18]. Among all the contributions, the work of Colizza et al. [5] is particularly noteworthy, as it utilized null models to analyze the implications of such a phenomenon from a more appropriate quantitative point of view. The introduction of null models, obtained through a process of network rewiring which preserves the degree sequence [19,20], was very important in order to overcome a bias when considering the rich-club coefficient $\phi(k)$ alone. Indeed, they proved that all networks, even random ones, share a monotonic increasing behavior of the rich-club coefficient with increasing value of degree. This problem was attributed to an intrinsic feature of every network structure in which *hubs*, i.e. high degree nodes, show a higher probability to be more interconnected than low degree ones. Essentially, the introduction of null models normalizes the rich-club coefficient $\phi(k)$ dividing it by another coefficient $\phi(k)^{ann}$, which is computed in the same way of $\phi(k)$ but results in an average over a proper set of random networks.

When considering the rich-clubs in networks, one can state that they show, in a broad sense, assortative mixing among club members and, at the same time, disassortative mixing since the hubs are connected even to a certain amount of lower degree nodes. In such trend, the rich-club coefficient may be seen as a measure able to quantify the level of assortative mixing either to degree or to a discrete characteristic, i.e. to be part of the club or not. However, we ignore *a priori* the size of the rich-club and, as mentioned before, the study of assortative mixing using standard measures [2] is not sufficient in this case. For this reason, in order to make an association between the two phenomena, and thus meet the recent needs of network science [21] by gaining new insights into rich-club ordering from an assortative mixing perspective, other approaches are required.

Within this context, a particularly useful and relevant method was initially introduced by Park and Barabási [22], who investigated assortative mixing using the so-called *dyadic effect*, which can be observed through a binary characterization of the nodes in a network. The dyadic effect is present when the number of nodes sharing a common characteristic is larger than expected if the characteristic were distributed randomly on the network. This effect, quantified by proper measures, has been widely appreciated and adopted, with a number of extensions throughout many different contexts [23–26].

In this paper we consider the rich-club phenomenon and the rich-club coefficient as introduced in [6], studying them with the solid theoretical basis provided by [5]. We contribute with a novel formulation of $\phi(k)$ in terms of dyads and dyadic effect. Using this formulation, which holds if we interpret the excess of a certain degree threshold as a node characteristic, we reveal certain related measures that provide alternative and additional tools to confirm and test the presence of rich-club ordering in complex networks.

In more detail, we refine the denominator of the original rich-club coefficient, using a theoretical formulation of the upper bounds to the occurrence of dyads types introduced in [27]. In addition, we quantify the improvement in the denominator of the rich-club coefficient through a measure that we call $\delta(k)$. The cut-off point in the curve related to $\delta(k)$ could be considered as an indicator of the rich-club size. We then prove how this new formulation maps into the coefficient $\rho(k)$ introduced by [5], which needs to be considered when the significance of the rich-club coefficient is tested.

Moreover, we compute a novel index called $\bar{\phi}(k)$, complementary to $\phi(k)$, which is able to report the extent and way in which nodes of the club tend to connect to the outside, i.e. to non-club nodes. Using $\bar{\phi}(k)$, we are able to understand whether nodes within the club preferentially connect to other high degree nodes, or to those with lower degree. Additionally, with the procedure shown in [5], we compute the normalized index $\bar{\rho}(k)$, which will be useful for confirming and strengthening the presence of rich-club ordering from a different perspective.

As this approach shows how the rich-club phenomenon and the dyadic effect map into one another, we are able to quantify their magnitude using degree sequence, i.e. combinatorial, based arguments that take into account the different kinds of connections among network nodes.

The paper is organized as follows: Section 2 provides the theoretical background; Section 3 shows the reformulation and normalization of the rich-club coefficient; Section 4 contains the computational analysis and Section 5 presents the conclusions.

2. Problem settings

2.1. Theoretical background

A network can be represented as a graph G with N nodes and M edges and herein we consider each graph as undirected, unweighted, connected and simple, i.e. loops and multiple edges are not allowed.

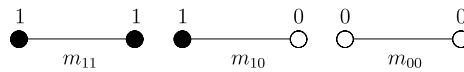


Fig. 1. Types of dyads.

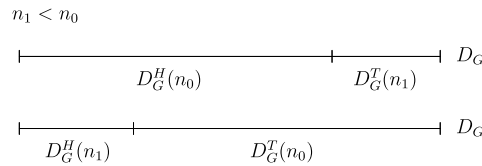


Fig. 2. Two different partitions when $n_1 < n_0$.

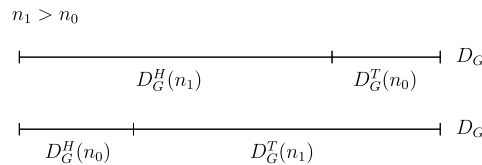


Fig. 3. Two different partitions when $n_1 > n_0$.

The degree d_i of a node i is defined as the number of edges incident to i . The list of nodes degrees in non-increasing order is called *degree sequence* D_G and for every connected graph holds the Degree-Sum Formula or *Handshaking Lemma*, $\sum_{i=1}^N d_i = 2M$. A *graphic sequence* is defined as a list of nonnegative numbers, which is the degree sequence of certain simple graphs, and a graph G with degree sequence D_G is called a *realization* of D_G . A generic list L of nonnegative numbers is not necessarily a graphic sequence. Indeed, a necessary and sufficient condition is that $\sum_{i=1}^N d_i$ is even and $\sum_{i=1}^N d_i \leq k(k-1) + \sum_{i=k+1}^N \min\{k, d_i\}$, $1 \leq k \leq N$ [28]. The problem of determining whether L is a graphic sequence is referred to as the *Graph Realization Problem* [29,30].

Given an integer $n \leq N$, we consider the subsequence of the first n elements of D_G whereby $D_G^H(n) \subseteq D_G$ is the *head* of D_G , and the subsequence of the last n elements of D_G whereby $D_G^T(n) \subseteq D_G$ is the *tail* of D_G .

A *clique* K_n is a complete subgraph of G of dimension n , i.e. a subgraph of n mutual interconnected nodes. A *star* S_n is a subgraph of G of dimension n showing one node with degree $n - 1$ and the others $n - 1$ having degree 1.

2.2. The rich-club coefficient

For our purpose, in this paper we consider the rich-club coefficient expressed in terms of nodes degree as in [5] that can be written as:

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

where $E_{>k}$ is the number of edges among the $N_{>k}$ nodes having degree d_i higher than a given value k and $\frac{N_{>k}(N_{>k}-1)}{2}$ is the maximum possible number of edges among the $N_{>k}$ nodes. Therefore $\phi(k)$ measures the fraction of edges connecting the $N_{>k}$ nodes out of the maximum number of edges they might possibly share.

2.3. Nodes characteristics and the dyadic effect

Herein, we refer to a given characteristic c_i , which can assume the values 0 or 1, for each $i \in N$. Consequently, N can be divided into two subsets: the set of n_1 nodes with characteristic $c_i = 1$, the set of n_0 nodes with characteristic $c_i = 0$; thus, $N = n_1 + n_0$. We distinguish three kinds of *dyads*, i.e. edges and their two end nodes, in the network: $(1 - 1)$, $(1 - 0)$, and $(0 - 0)$ as depicted in Fig. 1.

We label the number of each dyad in the graph as m_{11} , m_{10} , m_{00} , respectively. Hence, $M = m_{11} + m_{10} + m_{00}$. We consider m_{11} and m_{10} as independent parameters that represent the dyads containing nodes with characteristic 1.

Let D_G be the degree sequence of G , we can use n_1 and n_0 to define its head $D_G^H(n_1)$ or $D_G^H(n_0)$ and its tail $D_G^T(n_1)$ or $D_G^T(n_0)$ such that $D_G = D_G^H(n_1) \cup D_G^T(n_0)$ or $D_G = D_G^H(n_0) \cup D_G^T(n_1)$. These partitions of the degree sequence are given arbitrarily, assigning the characteristic $c_i = 1$ to the n_1 nodes with the highest degree or to the n_1 nodes with the lowest degree or vice versa. Such partitions are reported in Figs. 2 and 3, distinguishing the case in which $n_1 < n_0$ or $n_1 > n_0$.

When nodes in a network fit within two distinct groups according to their characteristics, two different parameters are required to determine the existence of the relations between the network topology and the nodes features [22]. Whether the number of edges between nodes sharing a common characteristic is larger than expected if the characteristics are distributed randomly on the graph we observe a phenomenon called the *dyadic effect* [22]. Considering the N nodes and the case in which any node possesses an equal chance of having the characteristic 1, given n_1 , the expected values of m_{11} and m_{10} are [22]:

$$\bar{m}_{11} = \binom{n_1}{2} \delta = \frac{n_1(n_1 - 1)}{2} \delta \tag{2}$$

$$\bar{m}_{10} = \binom{n_1}{1} \binom{n_0}{1} \delta = n_1(N - n_1) \delta \tag{3}$$

where δ is the density and is equal to $\delta = 2M/N(N - 1)$. The relevant deviations of m_{11} and m_{10} from the expected values \bar{m}_{11} and \bar{m}_{10} denote that the characteristic 1 is not randomly distributed [22,31]. Such deviations can be calculated through the ratios called *dyadicity* D and *heterophilicity* H defined as:

$$D = \frac{m_{11}}{\bar{m}_{11}} \tag{4}$$

$$H = \frac{m_{10}}{\bar{m}_{10}} \tag{5}$$

In [22], it is established that m_{11} and m_{10} cannot assume arbitrary values, as there are indirect constraints due to the network structure. Indeed, m_{11} cannot exceed

$$UBm_{11} = \min(M, \binom{n_1}{2}) \tag{6}$$

and m_{10} cannot be larger than

$$UBm_{10} = \min(M, n_1 n_0) \tag{7}$$

where UB stands for upper bound. These upper bounds are improved in [27], which explicitly outlines the structural constraints regarding the m_{11} and m_{10} links. The new upper bounds, which are herein used throughout the paper, are written as:

$$UBm_{11} = \min\left(M, \binom{n_1}{2}, \left[\sum_{i \in D_G^H(n_1)} \frac{\min(d_i, n_1 - 1)}{2} \right] \right) \tag{8}$$

$$UBm_{10} = \min\left(M, n_1 n_0, \min\left(\sum_{i \in D_G^H(n_1)} \min(d_i, n_0), \sum_{i \in D_G^H(n_0)} \min(d_i, n_1) \right) \right) \tag{9}$$

The first upper bound is based on the fact that large cliques may be rare substructures in networks [32,33]; therefore, using D_G we check whether G can actually contain a complete subgraph of size n_1 . If not, we take into account the densest hypothetical substructure that could be realized using the degree sequence of G . In the second upper bound we check if G can contain n_1 stars of size n_0 or n_0 stars of size n_1 looking at the degree sequence D_G . If not, we consider a set of stars made of the first n_1 elements of D_G if $n_1 < n_0$ or made of the first n_0 elements of D_G if $n_0 < n_1$. The exhaustive explanations and proofs are described in [27].

3. Reformulation and normalization of the rich-club coefficient

In this section we provide a reformulation of the rich-club coefficient involving most of the measures used in the study of the dyadic effect.

3.1. Reformulation of the rich-club coefficient

Let us write:

$$c_i = \begin{cases} 1 & \text{if } d_i > k \\ 0 & \text{if } d_i \leq k \end{cases}$$

we immediately obtain $N_{>k} = n_1$ and $N_{\leq k} = n_0$. Since $E_{>k}$ is defined as the number of edges between nodes $N_{>k} = n_1$, we can consequently write $E_{>k} = m_{11}$ while the maximum possible number of links between nodes N_k can be written as $\frac{N_{>k}(N_{>k}-1)}{2} = \binom{n_1}{2}$.

Let us also define $\bar{E}_{>k} = M - E_{>k} - E_{\leq k}$ as the number of edges among the $N_{>k}$ and $N_{\leq k}$ nodes having degree d_i greater and less equal than a given value k . Note that $N_{>k}N_{\leq k}$ is the maximum possible number of edges among the $N_{>k}$ and $N_{\leq k}$ nodes

Table 1

Correspondences between rich-club measures and their counterparts in terms of dyadic effect.

Rich-club notation	$N_{>k}$	$N_{\leq k}$	$E_{>k}$	$\bar{E}_{>k}$	$E_{\leq k}$
Dyadic effect notation	n_1	n_0	m_{11}	m_{10}	m_{00}

and, given that we have a nodal bipartition, thus $N = N_{>k} + N_{\leq k} = n_1 + n_0$. Therefore, we can write $N_{>k}N_{\leq k} = n_1n_0$ and, using the equality $\bar{E}_{>k} = M - E_{>k} - E_{\leq k} = M - m_{11} - m_{00} = m_{10}$ thus $\bar{E}_{>k} = m_{10}$. A summary of the discussed measures is reported in Table 1.

Under these circumstances we can reformulate $\phi(k)$ as:

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k} - 1)} = \frac{m_{11}}{\binom{n_1}{2}} \quad (10)$$

As previously outlined, by knowing the degree sequence D_G , we understand that the maximum number of links between the n_1 nodes is bounded by UBm_{11} and, as shown in [27], it can be much lower than $\binom{n_1}{2}$. Through this upper bound, we observe if the considered graph, as well as all the other ones having the same D_G , can contain a complete subgraph of order n_1 . If such case does not occur, D_G cannot generate a graph containing a clique with n_1 nodes and we can compute the rich-club coefficient obtaining a more realistic value, tailored on the degree sequence of the graph under consideration.

This approach may lead to higher values of $\phi(k)$, and thus a better understanding and quantification of the phenomenon, since we are taking into account a proper measure of dyadic effect. Indeed, in this case, the denominator of the coefficient embeds an informative content which derives directly from a structural property of the network, i.e. the degree sequence D_G . Finally we can reformulate an improved version of $\phi(k)$ as:

$$\phi(k)^{new} = \frac{m_{11}}{UBm_{11}} \quad (11)$$

3.2. The complementary rich-club coefficient

While computing m_{11} in order to have $\phi(k)$ or $\phi(k)^{new}$ we can easily compute m_{10} , i.e. the number of links between the nodes of the rich-club (n_1) and the nodes outside the club (n_0). Using this approach, we are able to quantitatively determine to which extent the nodes of the rich-club are connected to others in the network by making the following comparison: which portion of the total degree of the club members realizes internal connections, i.e. m_{11} , versus which portion realizes external connections, i.e. m_{10} . Under these circumstances, it is useful to exploit a classification of network links that has been provided (for instance, in [34]) in order to identify different classes of links within the human brain in case of rich-club ordering. In the article, the different kinds of links are called *rich-club connections* if they link rich-club nodes, *feeder connections* if they link rich-club nodes to nonrich-club nodes and *local connections* if they link nonrich-club nodes to each other. In our notation the rich-club connections correspond to m_{11} , the feeder connections correspond to m_{10} and the local connections correspond to m_{00} .

Following the same rationale as before, we can quantify the interaction among club and non-club members through the index $\bar{\phi}(k) = \frac{\bar{E}_{>k}}{N_{>k}N_{\leq k}}$ and *ipso facto* reformulate and improve it with certain considerations similar to those of Section 3.1.

Summarizing we can initially write $\bar{\phi}(k) = \frac{m_{10}}{n_1n_0}$ and, using the formulation of UBm_{10} as mentioned in Section 2.1, express $\bar{\phi}(k)$ as:

$$\bar{\phi}(k) = \frac{m_{10}}{UBm_{10}} \quad (12)$$

The index $\bar{\phi}(k)$ represents the number of feeder connections divided by the maximum possible number of such connections.

3.3. On the normalization of the rich-club coefficient

The use of null models and of the normalization process of topological and structural measures in complex networks represents a practice widely used to comprehend the real magnitude of certain phenomena. For this reason, the normalization of the rich-club coefficient, suggested in [5] and adopted in many further studies [11,12,14–18], is a necessary procedure that has to be adopted in order to take into account the significance of this index. The procedure involves an ensemble of random networks, sometimes called *maximally random networks* [5,12], which have the same degree distribution and sequence as the one under investigation that, if generated in a sufficiently large number, provide a null distribution of the rich-club coefficient. The rewiring procedure itself is simple since it chooses two arbitrary edges at each step ((a, b) and (c, d) for instance) and changes their endpoints (such that we obtain (a, d) and (c, b)); in case one or both of these new links already exist in the network, this step is aborted and a new pair of edges is selected. The described procedure has been widely

adopted since it preserves an important network parameter represented by nodes degree though other procedures aiming to preserve other parameters may be adopted [19,35,36].

Using this method, the normalized rich-club coefficient is defined [5] as:

$$\rho(k) = \frac{\phi(k)}{\phi(k)_{ran}} \tag{13}$$

where $\phi(k)_{ran}$ is the average rich-club coefficient across the random network ensemble. We observe rich-club ordering when $\rho(k) > 1$. As previously mentioned, the process generates random networks with prescribed degree sequence D_G , thus the number of nodes with degree higher than k , $N_{>k}$, will be the same for each k and for each network with the same D_G of the original one, implying that the denominator of $\phi(k)$, i.e. $\frac{N_{>k}(N_{>k}-1)}{2}$, will be the same for either $\phi(k)$ or $\phi(k)_{ran}$. We can consequently write:

$$\rho(k) = \frac{\frac{2E_{>k}}{N_{>k}(N_{>k}-1)}}{\frac{2E_{>k|ran}}{N_{>k}(N_{>k}-1)}} = \frac{E_{>k}}{E_{>k|ran}} \tag{14}$$

or using the dyadic effect notation:

$$\rho(k) = \frac{m_{11}}{m_{11|ran}} \tag{15}$$

where $m_{11|ran} = E_{>k|ran}$ is the average value of m_{11} across the random networks ensemble. Similar considerations hold even when we take into account $\phi(k)^{new}$ and they are motivated by the nature of its denominator, UBm_{11} , that depends uniquely on the degree sequence D_G and includes the quantity $\binom{n_1}{2} = \frac{N_{>k}(N_{>k}-1)}{2}$, as shown in [27]. Therefore, for any value of k , UBm_{11} will be the same for each network, having the same degree sequence D_G , leading to:

$$\rho(k)^{new} = \frac{\frac{m_{11}}{UBm_{11}}}{\frac{m_{11|ran}}{UBm_{11}}} = \frac{m_{11}}{m_{11|ran}} \tag{16}$$

and subsequently to the equality $\rho(k)^{new} = \rho(k)$.

This implies that our approach improves the rich-club coefficient $\phi(k)$ itself, and that the evaluation of the presence of such ordering maps into the well known approach of [5]. So when we observe the presence of the rich-club ordering, i.e. $\rho(k) > 1$ and $\phi(k) \neq 0$, we can evaluate its magnitude by observing the value of $\phi(k)^{new}$ or more precisely compare it with the value of $\phi(k)$ considering the coefficient $\delta(k)$ defined as:

$$\delta(k) = \frac{\phi(k)^{new} - \phi(k)}{\phi(k)^{new}} \tag{17}$$

Since $\phi(k)^{new} \geq \phi(k)$ for each k , $\delta(k) \in [0, 1]$.

The coefficient $\delta(k)$ may be helpful in further testing the significance of rich-club ordering in networks. It is important to underline that for high values of k , the number of nodes within the rich-club decreases, as does the denominator of $\phi(k)$. In this case, the value UBm_{11} tends to equal the number of links within the clique, leading to $\delta(k)$ being valued at zero. The index $\delta(k)$ underlines the fact that very large cliques are uncommon in real networks [32] and that the presence of dense substructures should be conducted using measures, such as UBm_{11} , that are more related to the network under investigation.

When $\delta(k) > 0$ then $\phi(k) \leq \phi(k)^{new}$ because $\binom{n_1}{2} \geq UBm_{11}$. Indeed, UBm_{11} exploits combinatorial arguments and captures the densest realizable substructure when a clique of n_1 nodes is not realizable within the given D_G . Thus, $\phi(k)^{new}$ is more likely to identify the best (densest) rich-club not as a clique, but as a sparser subgraph of the same size.

Another relevant information derives from the analysis of the relationships between club and non-club members, i.e. from the so-called feeder connections, using the aforementioned coefficient $\bar{\phi}(k)$. It compares the number of feeder connections in the network with the number of feeder connections across the random network ensemble, quantifying the portion of edges that have start point in club nodes and end point in non-club nodes. Even in this case, we are able to adopt the normalization process to obtain a second coefficient: $\bar{\rho}(k)$. This can be helpful in capturing and evaluating the interplay among different classes of nodes, quantifying for each club member the portion of its degree invested in the outside and computing the same quantity across all the club members. Therefore, we can write

$$\bar{\rho}(k) = \frac{\bar{\phi}(k)}{\phi(k)_{ran}} \tag{18}$$

where $\bar{\phi}(k)_{ran}$ is the average across the random network ensemble.

For networks showing rich-club ordering we expect that, compatibly with the size of the club, most of the degree of club members will be invested in links with other club members, meaning $\bar{\phi}(k) \leq \phi(k)_{ran}$. In other words, when rich-club ordering is present, we expect members showing $d_i > k$ (i.e. n_1) to be more connected among themselves that to other with $d_i < k$ (i.e. n_0) especially in comparison to randomized networks.

Thus, with the presence of rich-club ordering, it is expected that the curve of $\bar{\rho}(k)$ approaches the value $\bar{\rho}(k) = 1$ from below while k increases, meaning that we have fewer feeder connections than in random networks having the same degree

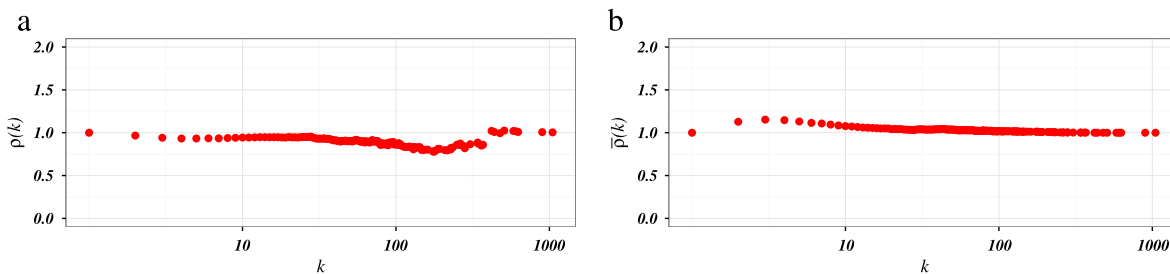


Fig. 4. Curves related to $\rho(k)$ and $\bar{\rho}(k)$ for the AS-level Internet network.

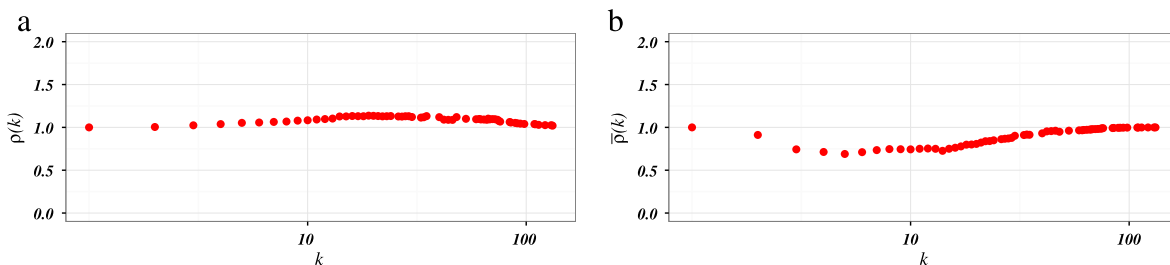


Fig. 5. Curves related to $\rho(k)$ and $\bar{\rho}(k)$ for the top 500 US airports network.

sequence as that of the original. Indeed, $m_{10} \leq m_{10|ran}$ means that high degree nodes tend to connect more densely to one another than to low degree nodes.

It is also important to point out that the trend of the curve $\bar{\rho}(k)$ reaching the value $\bar{\rho}(k) = 1$ (either from below or from above) is not surprising, especially for high values of k , since for such values the clique size could be very small. For instance, let us consider a certain network showing rich-club ordering. If we imagine that the club is made up of the three highest degree nodes of the network, a triangle would be the biggest clique, thus allowing the rich-club to have three links at most. In this case, the three nodes will necessarily connect to other ones showing a lower degree, thereby increasing the chance of having $\bar{\phi}(k) = \bar{\phi}(k)_{ran}$. For these reasons, the values assumed by $\bar{\rho}(k)$ have to be considered collectively via a curve that is able to represent the tendency of high degree nodes to connect mainly to other high degree nodes; i.e. a curve able to unveil the tendency of nodes to form connected cores of important members. If the curve $\bar{\rho}(k)$ approaches the value $\bar{\rho}(k) = 1$ from below, this seems to further confirm the presence of rich-club ordering.

4. Computational results

Herein, we show computational results for three different networks: the AS-level Internet topology as in [6], the top 500 US airports network as in [37] and the Scientific Collaboration Network as in [38]. In order to compute $\rho(k)$ for the three case studies we consider a set of 1000 random graphs generated by randomizing the connections of the adjacency matrix while preserving the original degree sequence such as, for instance, in [5,39,40]. The data processing, the network analysis and all simulations were performed using the software R [41] with the *igraph* library [42].

In the case of the Internet topology and referring to $\rho(k)$ in Fig. 4(a), we obtained results corresponding to those of [5] that are in some way strengthened by the fact that $\bar{\rho}(k)$ is approaching the value $\bar{\rho}(k) = 1$ from above, as shown in Fig. 4(b). This means that the feeder connections are distributed as in random networks with the same degree sequence of the original one. This shows the low tendency of hubs to connect mainly to other hubs, thus confirming the absence of a strong rich-club ordering in such a structure. It is also worth mentioning at this point that the study of the rich-club ordering on the Internet topology is still a debated issue [11,13].

In the case of the top 500 US airports network, we observe rich-club ordering as shown in Fig. 5(a), which is still confirmed by the distribution of the feeder connections approaching the value $\bar{\rho}(k) = 1$ from below as depicted in Fig. 5(b). Indeed, in this network, $\bar{\rho}(k)$ is always less than 1, thus the nodes within the club tend to be more densely connected among themselves than observed in random networks.

In the third case of Citation network, shown in Fig. 6, we also have rich-club ordering, once again confirmed by the trend of $\bar{\rho}(k)$ approaching the value $\bar{\rho}(k) = 1$ from below. These three examples seem to confirm the discriminating power of $\bar{\rho}(k)$.

Moreover, similarly as in [11], we call:

$$\tilde{p} = \frac{\#[\phi(k) \leq \phi(k)_{ran}]}{n}$$

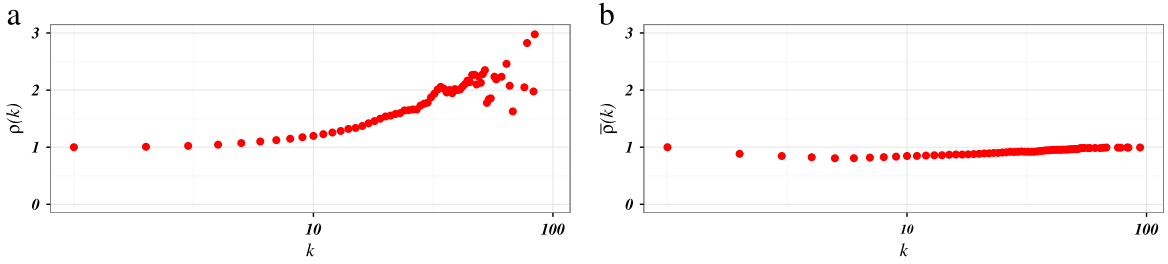


Fig. 6. Curves related to $\rho(k)$ and $\tilde{\rho}(k)$ for the Scientific Collaboration network.

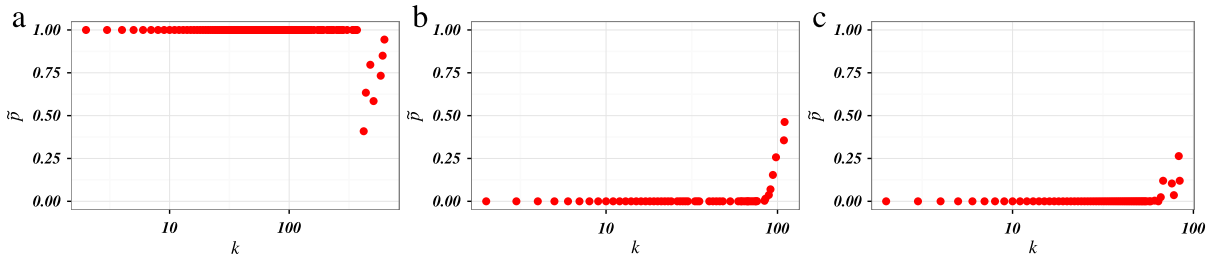


Fig. 7. Curves related to \tilde{p} , for the AS-level Internet network (a), the top 500 US airports network (b) and the Scientific Collaboration network (c).

where $\phi(k)$ is the rich-club coefficient of the actual network, $\phi(k)|_{ran}$ is the set of rich-club coefficients of the rewired networks and n is the amount of rewired networks that we generated. Thus, \tilde{p} can be considered as the likelihood that $\phi(k) \leq \phi(k)|_{ran}$, i.e. the likelihood of finding a random network with an equal or tighter structural organization of high degree nodes.

As \tilde{p} decreases, so too does the expectation of finding a value of $\phi(k)|_{ran}$ greater than $\phi(k)$, thus indicating stronger evidence towards the presence of rich-club ordering. In Fig. 7 we observe certain changes in the curves of \tilde{p} while k increases. In more detail, the US airports and the Scientific Collaboration networks show an increasing trend of \tilde{p} because hubs possess a higher probability (proportional to their degree) of being interconnected when we consider randomized networks. In the case of the Internet, we observe a peculiar arrangement of nodes interconnections (lower values of \tilde{p}) only if they are hubs. In any case, the curve of \tilde{p} continues to increase after a first discontinuity, displaying a behavior that is similar to $\rho(k)$, which indeed overcomes the threshold value $\rho(k) = 1$ only for very high values of k .

Figs. 8–10(a) display the curves of $\phi(k)$ and $\phi(k)^{new}$, which were obtained by computing the rich-club coefficient in its original reformulation [6] and the reformulated rich-club coefficient.

We observe that, up to a certain point, our reformulation outperforms the older one because of the impossibility to realize a clique of size n_1 using the degree sequence of the considered network. Conversely, the two coefficients become equal when a clique is realizable and thus the network structure allows the presence of a rich-club. It is worth mentioning that the situations in which a rich-club is likely to be present are those of interest in the study of real networks as the rich-club structural properties are able to provide insights about local and global aspects of the whole system [13,43–46].

Through the analysis of the curves related to $\phi(k)$ and $\phi(k)^{new}$, we can also discuss the monotonic increase of the original rich-club coefficient reported in [5]. Indeed, our coefficient, which is an extension of the previous one, is not monotonically increasing at all and thus overcomes the bias reported in [5] for any value of k . It is worth noting that the monotonic increase of $\phi(k)$ was also discussed in [36], where it is shown that the coefficient may assume a bell shape for higher values of k .

Figs. 8–10(b) display the curves related to the coefficient $\delta(k)$, which has been introduced in order to synthetically evaluate the gain that derives from the adoption of $\phi(k)^{new}$. The trend of the three curves mainly depends on the shape of the degree sequence D_G thus, on the degree distribution of the network under examination, since the new upper bound (the denominator of $\phi(k)^{new}$) is computed using that element. The decreasing behavior of the curve is not surprising; indeed while the degree k grows, the number of elements with degree higher than k ($N_k = n_1$) decreases, implying a consequent reduction in the number of links within the clique of size $N_{>k} = n_1$ that becomes more likely to be present. It is therefore interesting to discuss the cut-off points of the curves related to $\delta(k)$ that correspond to the points where the curves $\phi(k)$ and $\phi(k)^{new}$ encounter each other. The cut-off points occur in correspondence to the degree value k , which is necessary to overcome in order to guarantee a certain likelihood of finding a complete subgraph of size n_1 . As such, the cut-off points can thus be interpreted as a first indicator of the rich-club size. In the Internet network ($N = 11\,461$, $M = 32\,730$), we observe the cut-off point when $k = 83$ and in such a network the number of nodes with degree higher than k is $N_{>k} = n_1 = 84$. Therefore, if we consider the cut-off point as a rich-club size indicator, we would obtain a subgraph with $n_1 = 84$ nodes, i.e. a subgraph whose size is $0.7\% (\frac{n_1}{N} = 0.0073)$ with respect to the whole network. In the case of the US airports network

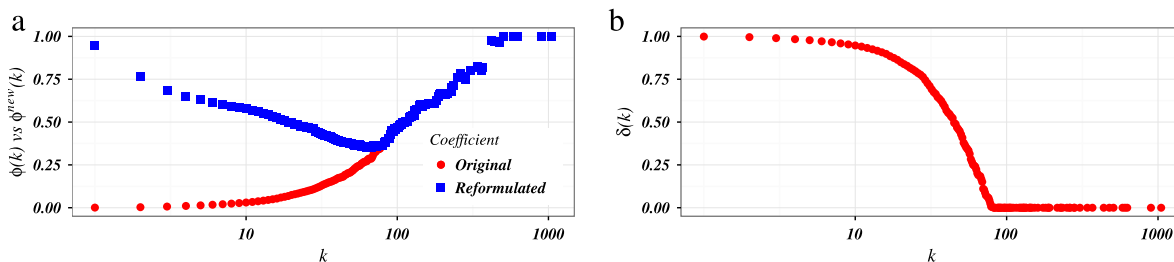


Fig. 8. Curves related to $\phi(k)$ vs $\phi(k)^{new}$ and $\delta(k)$ for the AS-level Internet network.

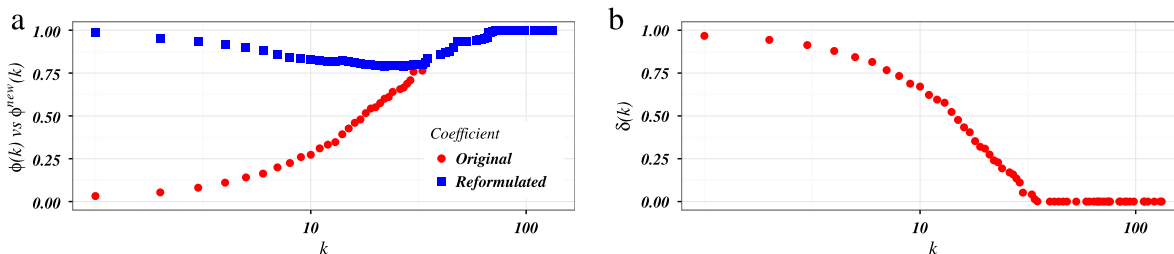


Fig. 9. Curves related to $\phi(k)$ vs $\phi(k)^{new}$ and $\delta(k)$ for the top 500 US airports network.

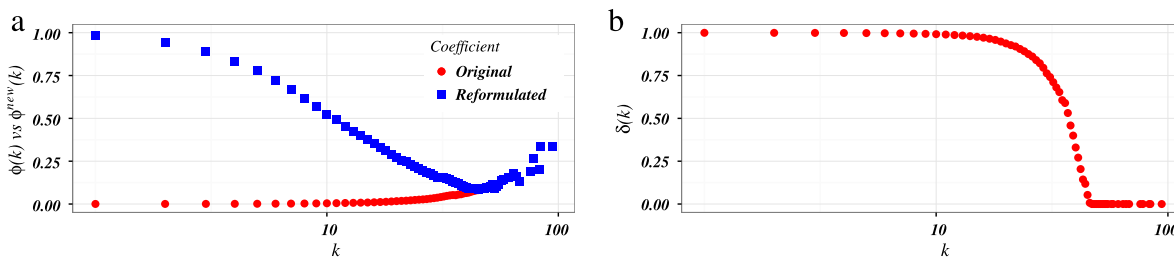


Fig. 10. Curves related to $\phi(k)$ vs $\phi(k)^{new}$ and $\delta(k)$ for the Scientific Collaboration network.

($N = 500, M = 2980$), the cut-off occurs when $k = 83$ and, in this case, $N_{>k} = n_1 = 39$, thus the rich-club size is estimated to be around 8% ($\frac{n_1}{N} = 0.078$) of the whole network. In the case of the Scientific Collaboration network ($N = 13\,861, M = 44\,619$), the cut-off occurs when $k = 47$ and, in this case, $N_{>k} = n_1 = 48$, thus the rich-club size would be close to 0.3% ($\frac{n_1}{N} = 0.0034$) of the whole network.

The greater size of the rich-club estimated in the case of the US airports network may be due to the density of such a network (0.023) while the densities of the Internet and the Scientific Collaboration networks are much lower (around 0.0005 and 0.0004 respectively). Thus the size of the rich-club may be related to network density and the cut-off point in the curve of $\delta(k)$ can be considered as a first indicator (or upper bound) to the rich-club size. Additionally, this quantity that cannot be computed a priori [47] and is conventionally assumed to be a certain low percentage [6,13] of the nodes with the highest degree without analytical evidence.

5. Conclusions

In this paper, we investigated the organization of rich nodes in networks from a slightly different perspective: the dyadic effect. We proved how rich-club ordering and the dyadic effect have much in common, showing how one can map into another and vice versa. Exploiting certain properties concerning the study of dyads, we provided additional instruments in order to investigate and confirm the presence of rich-club ordering. As previously mentioned, the issue of knowing whether a certain network possesses a rich-club is important for many reasons, including factors relating to either its functional or topological role. Indeed, the presence of this peculiar substructure is relevant in problems such as resilience and information flow in networks, mainly because of its influence on crucial properties like assortativity and transitivity.

From the computational point of view, the analysis of rich-club in terms of the dyadic effect needs the same resources of its original formulation since it provides simultaneously (i.e from a single network observation) the values of m_{11} and m_{10}

that are necessary in order to obtain $\phi(k)$ and $\bar{\phi}(k)$. The last observation holds even for what concerns the computation of the set of rewired networks that is the same for both $\rho(k)$ and $\bar{\rho}(k)$. Moreover, the analysis of the cut-off point of the curve related to $\delta(k)$ gives the possibility of providing an upper bound to the rich-club size, which is independent of the ensemble of randomized networks due to the fact that it is based only on structural constraints. The consistency of our approach with the literature, in term of normalization processes, represents a further advantage because it includes a more detailed information content related to the presence of both rich-club connections and feeder connections.

As further investigation, it may be interesting to study the organization of non-rich nodes and their connection patterns. For instance, a coefficient $\tilde{\phi}(k) = \frac{m_{00}}{UBm_{00}}$ (where UBm_{00} corresponds to UBm_{11} when considering nodes with $c_i = 0$) may be introduced to test the connectance among nodes linked by local connections. However, the analysis of such an aspect could represent a difficult issue to deal with because of the interpretation given to the presence of highly interconnected nodes with relatively low degree.

Another important aspect may concern the relationship between rich-club ordering and other centrality measures beyond degree. It would be interesting to study dyadic interactions among nodes possessing values of centrality over a certain richness threshold. Lastly, the extension of the dyadic effect to weighted and directed networks could provide room for new developments linked to results in the existing literature that has focused on the investigation of particular networks, such as the World Trade Web and the brain networks in mammals.

Acknowledgments

We thank Shi Zhou for useful suggestions and comments. We also thank the three anonymous reviewers.

References

- [1] David Lusseau, Mark E.J. Newman, Identifying the role that animals play in their social networks, *Proc. R. Soc. Lond. [Biol]* 271 (Suppl. 6) (2004) S477–S481.
- [2] Mark E.J. Newman, Mixing patterns in networks, *Phys. Rev. E* 67 (2) (2003) 026126.
- [3] Rogier Noldus, Piet Van Mieghem, Assortativity in complex networks, *J. Complex Netw.* 3 (4) (2015) 507.
- [4] Miller McPherson, Lynn Smith-Lovin, James M. Cook, Birds of a feather: Homophily in social networks, *Annu. Rev. Sociol.* 27 (1) (2001) 415–444.
- [5] Vittoria Colizza, Alessandro Flammini, M. Angeles Serrano, Alessandro Vespignani, Detecting rich-club ordering in complex networks, *Nat. Phys.* 2 (2) (2006) 110–115.
- [6] Shi Zhou, Raúl J. Mondragón, The rich-club phenomenon in the internet topology, *IEEE Commun. Lett.* 8 (3) (2004) 180–182.
- [7] Albert-László Barabási, Reka Albert, Emergence of scaling in random networks, *Science* 286 (5439) (1999) 509–512.
- [8] Ginestra Bianconi, Albert-László Barabási, Competition and multiscaling in evolving networks, *Europhys. Lett.* 54 (4) (2001) 436.
- [9] Jared Winick, Sugih Jamin, Inet-30: Internet Topology Generator. Technical Report, Technical Report CSE-TR-456-02, University of Michigan, 2002.
- [10] Sergi Valverde, Ricard V. Solé, Self-organization versus hierarchy in open-source social networks, *Phys. Rev. E* 76 (2007) 046118.
- [11] Zhi-Qiang Jiang, Wei-Xing Zhou, Statistical significance of the rich-club phenomenon in complex networks, *New J. Phys.* 10 (4) (2008) 043002.
- [12] Julian J. McAuley, Luciano da Fontoura Costa, Tibério S. Caetano, Rich-club phenomenon across complex network hierarchies, *Appl. Phys. Lett.* 91 (8) (2007) 084103.
- [13] Xiao-Ke Xu, Jie Zhang, Michael Small, Rich-club connectivity dominates assortativity and transitivity of complex networks, *Phys. Rev. E* 82 (4) (2010) 046117.
- [14] Vinko Zlatić, Ginestra Bianconi, Albert Díaz-Guilera, Diego Garlaschelli, Francesco Rao, Guido Caldarelli, On the rich-club effect in dense and weighted networks, *Eur. Phys. J. B* 67 (3) (2009) 271–275.
- [15] Ed Bullmore, Olaf Sporns, Complex brain networks: graph theoretical analysis of structural and functional systems, *Nat. Rev. Neurosci.* 10 (3) (2009) 186–198.
- [16] Logan Harriger, Martijn P. van den Heuvel, Olaf Sporns, Rich club organization of macaque cerebral cortex and its role in network communication, *PLoS One* 7 (9) (2012) 1–13.
- [17] Olaf Sporns, Dante R. Chialvo, Marcus Kaiser, Claus C. Hilgetag, Organization, development and function of complex brain networks, *Trends Cogn. Sci.* 8 (9) (2004) 418–425.
- [18] Gorka Zamora-López, Changsong Zhou, Jürgen Kurths, Exploring brain function from anatomical connectivity, *Front. Neurosci.* 5 (2011) 83.
- [19] Shweta Bansal, Shashank Khandelwal, Lauren Ancel Meyers, Exploring biological network structure with clustered random networks, *BMC Bioinformatics* 10 (1) (2009) 1.
- [20] Sergei Maslov, Kim Sneppen, Specificity and stability in topology of protein networks, *Science* 296 (5569) (2002) 910–913.
- [21] Duncan J. Watts, Should social science be more solution-oriented?, *Nat. Hum. Behav.* 1 (0015) (2017).
- [22] Juyoung Park, Albert-László Barabási, Distribution of node characteristics in complex networks, *Proc. Natl. Acad. Sci.* 104 (46) (2007) 17916–17920.
- [23] Luisa Di Paola, Micol De Ruvo, Paola Paci, Daniele Santoni, Alessandro Giuliani, Protein contact networks: an emerging paradigm in chemistry, *Chem. Rev.* 113 (3) (2012) 1598–1613.
- [24] Giovanna Ferraro, Antonio Iovanella, Gianluca Pratesi, On the influence of nodes' characteristic in inter-organisational innovation networks structure, *Int. J. Comput. Econ. Econometrics* 6 (3) (2016) 239–257.
- [25] Ting Hu, Angeline S. Andrew, Margaret R. Karagas, Jason H. Moore, Functional dyadicity and heterophilicity of gene-gene interactions in statistical epistasis networks, *BioData Min.* 8 (1) (2015) 1.
- [26] Jeffrey Q. Jiang, A.W. Dress, Ming Chen, Towards prediction and prioritization of disease genes by the modularity of human phenome-genome assembled network, *J. Integr. Bioinform.* 7 (2) (2010) 149.
- [27] Matteo Cinelli, Giovanna Ferraro, Antonio Iovanella, Structural bounds on the dyadic effect, *J. Complex Netw.* (2017) <http://dx.doi.org/10.1093/comnet/cnx002>.
- [28] Paul Erdős, Tibor Gallai, Graphs with prescribed degrees of vertices (in hungarian), *Mat. Lapok* 11 (1960) 265–274.
- [29] S. Louis Hakimi, On realizability of a set of integers as degrees of the vertices of a linear graph i, *J. Soc. Ind. Appl. Math.* 10 (3) (1962) 496–506.
- [30] Václav Havel, A remark on the existence of finite graphs, *Casopis Pest. Mat.* 80 (477–480) (1955) 1253.
- [31] L. Maurício de Almeida, A. Gabriel Mendes, G. Madras Viswanathan, R. Luciano da Silva, Scale-free homophilic network, *Eur. Phys. J. B* 86 (2) (2013) 1–6.

- [32] Ginestra Bianconi, Matteo Marsili, Emergence of large cliques in random scale-free networks, *Europhys. Lett.* 74 (4) (2006) 740.
- [33] Béla Bollobás, *Random Graphs*, in: *Modern Graph Theory*, Springer, 1998, pp. 215–252.
- [34] Martijn P. van den Heuvel, René S. Kahn, Joaquín Goñi, Olaf Sporns, High-cost, high-capacity backbone for global brain communication, *Proc. Natl. Acad. Sci.* 109 (28) (2012) 11372–11377.
- [35] R.J. Mondragón, S. Zhou, Random networks with given rich-club coefficient, *Eur. Phys. J. B* 85 (9) (2012) 328.
- [36] S. Zhou, R.J. Mondragón, Structural constraints in complex networks, *New J. Phys.* 9 (6) (2007) 173.
- [37] Vittoria Colizza, Romualdo Pastor-Satorras, Alessandro Vespignani, Reaction–diffusion processes and metapopulation models in heterogeneous networks, *Nat. Phys.* 3 (4) (2007) 276–282.
- [38] Mark E.J. Newman, The structure of scientific collaboration networks, *Proc. Natl. Acad. Sci.* 98 (2) (2001) 404–409.
- [39] Guusje Collin, Olaf Sporns, René C.W. Mandl, Martijn P. van den Heuvel, Structural and functional aspects relating to cost and benefit of rich club organization in the human cerebral cortex, *Cerebral Cortex* 24 (9) (2014) 2258–2267.
- [40] Martijn P. Van Den Heuvel, Olaf Sporns, Rich-club organization of the human connectome, *J. Neurosci.* 31 (44) (2011) 15775–15786.
- [41] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, ISBN: 3-900051-07-0, 2008.
- [42] Gabor Csardi, Tamas Nepusz, The igraph software package for complex network research, *Int. J. Complex Syst.* 1695 (2006).
- [43] Gareth Ball, Paul Aljabar, Sally Zebari, Nora Tusor, Tomoki Arichi, Nazakat Merchant, Emma C. Robinson, Enitan Ogundipe, Daniel Rueckert, A. David Edwards, Serena J. Counsell, Rich-club organization of the newborn human brain, *Proc. Natl. Acad. Sci.* 111 (20) (2014) 7456–7461.
- [44] David S. Grayson, Siddharth Ray, Samuel Carpenter, Swathi Iyer, Taciana G. Costa Dias, Corinne Stevens, Joel T. Nigg, Damien A. Fair, Structural and functional rich club organization of the brain in children and adults, *PLoS One* 9 (2) (2014) 1–13.
- [45] Naoki Masuda, Norio Konno, Vip-club phenomenon: Emergence of elites and masterminds in social networks, *Social Networks* 28 (4) (2006) 297–309.
- [46] Shi Zhou, Raúl J. Mondragón, Accurately modeling the internet topology, *Phys. Rev. E* 70 (2004) 066108.
- [47] Athen Ma, Raúl J. Mondragón, Rich-cores in networks, *PLoS One* 10 (3) (2015) 1–13.