CORE
Metadata, citation and similar papers at core.ac.uk
Provided by Queen Mary Research Online

Queen Mary
University of London

Ensembles related to the rich-club coefficient for non-evolving networks

Mondragon, RJ; Zhou, S

https://arxiv.org/abs/0808.4062

For additional information about this publication click this link.
http://qmro.qmul.ac.uk/xmlui/handle/123456789/15773

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk arXiv:0808.4062v2 [physics.soc-ph] 16 Dec 2008

Ensembles of reference networks based on the rich–club structure for non–evolving networks

Raul J. Mondragón[1] and Shi Zhou[2]

[1]School of Electronic Engineering and Computer Sciences,
Queen Mary University of London
Mile End Road, London, E1 4NS, United Kingdom

[2]University College London, Adastral Park Campus
Ross Building, Adastral Park,
Ipswich, IP5 3RE, United Kingdom

Abstract

In networks the rich nodes are the subset of nodes with large numbers of links, or high degrees. The rich nodes and the connectivity between themselves (*rich–club connectivity*) tend to dominate the organisation of network structure. Recently there has been a considerable effort to characterise and model the rich–club connectivity in a variety of complex networks. In this paper we firstly clarify a number of terms: the rich-club *coefficient* quantifies the density of connectivity between a subset of rich nodes; the rich-club *structure* is the rich-club coefficient measured across the hierarchies of nodes; the rich-club *phenomenon* refers to the dynamic behaviour responsible for the formation of the rich–club connectivity in evolving networks; and the rich-club *ordering* discerns whether the connectivity between rich nodes in a non-evolving network (i.e. closed network or network snapshot) is higher than a reference network obtained by network randomisation. We then evaluate a recently proposed null model which is based on an ensemble of reference networks conserving the degree distribution of the original network. We remark that one should not confuse the rich-club structure of a network with the rich-club ordering detected by the null model. We also demonstrate that the null model cannot identify the dynamical mechanism that generates the rich–club connectivity. The main contribution of the paper is that we introduce two new ensembles of reference networks based on the rich-club structure for non–evolving networks. The first ensemble preserves the rich-club coefficient (as a function of the rank of a node) of the original network. Members of the ensemble exhibit similar degree distribution as the original network, and for assortative networks, similar assortative mixing as well. We propose that this ensemble can be used to study networks where assortativeness is a fundamental property, e.g. to detect the community structure in social networks. Analysis on the ensemble also provides a different way to interpret and model the evolution of social networks. The second ensemble preserves both the degree distribution and the rich–club coefficient (as a function of node degree). The reference networks in this ensemble have a similar structure as the original network. We use them to quantify the correlation profile between the rich nodes and pinpoint which links between the nodes are the backbone of network structure.

1 Introduction

Many social, economic, biological and technological networks contain a small set of nodes which have large numbers of links, the so–called rich nodes. In some networks the rich nodes are tightly interconnected between themselves, forming a rich–club [31]. The rich–club is an oligarchy in that it dominates the organisation of the whole network. In scale–free networks [1] the connectivity between the rich nodes plays an important role in network functionality, for example in the transmission of rumours in social networks [15] or the efficient delivery of data packets in the Internet [30]. The density of connections between the rich nodes is quantified by the rich–club coefficient [31]. The rich–club coefficient and its generalisations [8, 16, 28, 33, 23, 25] have been proved to be a useful measure for studying complex networks.

Many of the complex networks studied in the literature are single networks in the sense that each network's structure is not one of several but unique, for example the Internet. In this case we do not have the equivalent to a physical law to verify whether a statistical measure obtained from a single network is expected or extraordinary. Instead, a common technique to assess the significance of a property of a single network is to use the statistical randomisation method [12] to create a null model. The procedure consists of using the observed network to generate an ensemble [5] of reference networks via randomisation. The null model is then generated from this ensemble.

In this paper we study three ensembles of reference networks which are generated by different network randomisation processes. The first ensemble has been widely studied. In one recent study it was used to assess whether the connectivity between the rich nodes in a network is due to chance or an unknown mechanism. We discuss the results and clarify the confusions arising from the study. We then introduce the other two ensembles, which are based on the rich–club structure for non-evolving networks. We analyse their properties and propose useful applications for the statistical physics study on complex networks. Our work not only advances the understanding of the rich-club structure in networks but also provides new methods for studying other properties related to the rich-club structure.

2 Definitions

2.1 The rich–club coefficient and the rich-club structure

Degree of a node, k, the number of links a node has. The rank r of a node is its position in the list of decreasing order of node degrees, i.e. the best-connected node is ranked as $r = 1$, the second best-connected node is $r = 2$ and so on. Rich nodes can be defined as nodes with large degrees or small ranks. The *density* of connections between the r richest nodes is quantified by the rich–club coefficient [31]

$$\Phi(r) = \frac{2E_{\leqslant r}}{r(r-1)}, \tag{1}$$

where $E_{\leqslant r}$ is the number of links between the r nodes and $r(r-1)/2$ is the maximum number of links that these nodes can share. If $\Phi(r) = 0$ the nodes do not share any link at all, if $\Phi(r) = 1$ the nodes form a fully connected sub–graph, a clique. As a function of degree, the rich–club coefficient can also be given as [8]

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

where $N_{\geqslant k}$ is the number of nodes with degrees greater or equal to k and $E_{\geqslant k}$ is the number of links between the $N_{\geqslant k}$ nodes.

The rich–club coefficients $\Phi(r)$ and $\phi(k)$ are related but they are not the same. The rank gives a unique label to each node, and the degree can be used to group nodes into subsets. If r^*_k is the node with degree k such that $r^*_k + 1$ is the rank of the node with degree $k-1$, then $\phi(k) = \Phi(r^*_k)$. This is to say, $\{\phi(k)|k = 0, \ldots, k_{\max}\}$ is a subset of $\{\Phi(r)|r = 1, \ldots, N\}$, where $k_{\max}$ is the maximum degree in the network and N is the total number of nodes. Two networks can have the same $\phi(k)$ and the same degree distribution $P(k)$ for all k, but different $\Phi(r)$.

Originally the term rich–club was defined as the set of the richest nodes that are tightly interconnected. This definition is vague as 'tight' is a relative concept. Recently, Valverde and Solé [27] proposed a criteria to define the rich–club and hence the rich nodes. The rich–club is defined by the existence of a crossover at k_c in $\phi(k)$ (or r_c in $\Phi(r)$) and this crossover characterise the rich nodes.

The rich–club *structure* refers to the density of connections across the hierarchies of nodes. It is given by the rich–club coefficient $\phi(k)$ for all k (or $\Phi(r)$ for all r). The rich–club structure can be fully defined by the degree–degree distribution, $P(k,\,k')$, the probability that an arbitrary link connects a node of degree k with a node of degree k' [8],

$$\phi(k) = \frac{N\langle k\rangle \sum_{k'=k}^{k_{\max}} \sum_{k''=k}^{k_{\max}} P(k', k'')}{(N\sum_{k'=k}^{k_{\max}} P(k'))(N\sum_{k'=k}^{k_{\max}} P(k') - 1)}, \tag{3}$$

where $\langle k\rangle$ is the average degree. Conversely, we show in the following that fixing the rich–club structure will constrain the degree–degree distribution [32, 11]. From the definition of the rich–club coefficient $\phi(k)$, the number of links that have at one end a node with degree k and at the other end a node with degree at least k is

$$E_k = E_{\geqslant k} - E_{\geqslant k+1} = \phi(k)\frac{N_{\geqslant k}(N_{\geqslant k}-1)}{2} - \phi(k+1)\frac{N_{\geqslant k+1}(N_{\geqslant k+1}+1)}{2}. \tag{4}$$

In terms of the conditional probability $P(k'|k)$ that a node with degree k has a neighbour with degree k',

$$E_k = 2NP(k)\left(\sum_{k'=k}^{k_{\max}} P(k'|k) - \frac{1}{2}P(k+1|k)\right), \tag{5}$$

where is $NP(k)$ is the number of nodes with degree k, and the term $P(k'|k) - \frac{1}{2}P(k+1|k)$ gives the proportion of these nodes that are connected to a node with degree at least k. We know that $P(k,\,k')$ and $P(k'|k)$ are related as $P(k'|k) = \langle k\rangle P(k,k')/(kP(k))$ where $\sum_{k'} P(k'|k) = 1$. Hence, fixing $\phi(k)$ for all k constrains the possible values of $P(k|k')$ or equivalently $P(k,k')$.

2.2 The rich–club phenomenon and the rich–club ordering

The rich–club phenomenon [31] refers to the *dynamic* behaviour in some evolving networks where, if a node becomes rich, it will tend to connect with other rich nodes forming a rich–club or join an existing rich–club. An evolving network model can introduce such dynamic mechanisms to reproduce the rich-club phenomenon and therefore generate a network with similar rich–club structure as the real network [7, 3, 31, 29]. One such mechanism was introduced in 1942 by Simon [26]. Simon's model was based on the addition of new nodes and the addition of new links between nodes that belong to the same class, where a class is the set of nodes with the same degree. Bornholdt and Ebel [7] have pointed out that this growth mechanism allows different growth rate for different classes of nodes and hence it can create a rich–club clique. Recently Krapivsky and Krioukov [11] showed how the inclusion of the rich–club phenomenon in evolving networks drastically constraints their structure.

The term rich–club phenomenon has been mixed with the term *rich–club ordering* in the recent study [8] which compared the rich-club coefficient of the original network against a reference network. We suggest to use the term rich–club ordering to refer to the increase of rich-club coefficient in comparison with the reference network. It is a *static* property obtained for non–evolving networks (closed networks) or a snapshot of evolving networks.

3 The relative rich–club reference network

A widely studied ensemble of reference networks are the maximal random networks generated by randomly reshuffling link–pairs of the network under study [13, 14]. The intrinsic structure of the original network is taken into account by imposing the restriction that the reshuffling process should not change the degree distribution.

Recently Colizza *et al.* [8] used this ensemble as a null model to discern whether the connections between the rich nodes in a network is due to chance or due to an "organisational principle". To do so, they compared the original rich–club coefficient $\phi(k)$ with the randomised rich–club coefficient $\phi_{\mathrm{ran}}(k)$ obtained from the maximal random networks in the ensemble. Colizza *et al.* suggested that the "normalised" rich–club coefficient $\rho_{\mathrm{ran}}(k) = \phi(k)/\phi_{\mathrm{ran}}(k)$ discounts the structural correlation imposed by the finite–size effects [6]. If $\rho_{\mathrm{ran}}(k) > 1$, it means there is an organisational principle that leads to an increase in the density of connections between rich nodes in a more pronounced way than in the null model, i.e. the rich–club ordering.

We prefer to call $\rho_{\mathrm{ran}}(k)$ the *relative* rich–club coefficient. Since the number of nodes with degree k does not change by the randomisation procedure, $\rho_{\mathrm{ran}}(k) = E_{\geqslant k}/E_{\mathrm{ran},\geqslant k}$ is the ratio of original–network links to the reference–network links. Hence $\rho_{\mathrm{ran}}(k)$ does not give information about the *density* of connections between rich nodes (which is measured by the rich–club coefficient). This simple observation is relevant because there has been some confusion in the literature on what the rich–club coefficient and $\rho_{\mathrm{ran}}(k)$ are measuring. For example Fig. 1 shows the Internet network at the Autonomous System level (AS–Internet) [9, 24] and for the scientific collaborations network in the area of condensed matter physics (Collaborations–A) [17, 18]. Figs. 1(a) and (c) show the two networks' original $\phi(k)$ in green colour and values of the randomised $\phi_{\mathrm{ran}}(k)$ obtained from 10^3 maximal random networks, where the frequency of a particular value of $\phi_{\mathrm{ran}}(k)$ is labelled with different colours, from seldom (0.1, red) to often (1.0, blue). The null model $\langle\rho_{\mathrm{ran}}(k)\rangle$ is obtained by averaging over all the maximal random networks and hence corresponds to the 'bluest' dots. For the AS Internet, the observation that $\langle\phi_{\mathrm{ran}}(k)\rangle > \phi(k)$, i.e. $\langle\rho_{\mathrm{ran}}(k)\rangle < 1$, for almost all values of k, suggests that the Internet does not have a rich–club ordering. This has created the misinterpretation as stated in [8, 2] that the rich nodes in the AS Internet were not tightly interconnected with each other. However, Fig. 1(b) shows that the 20 best-connected nodes in the AS Internet are tightly interconnected between themselves. Even more, the top seven best-connected nodes form a fully connected clique. For the Collaborations–A network, the property $\langle\phi_{\mathrm{ran}}(k)\rangle < \phi(k)$ insinuates that the top scientists form tighter collaborations compared to the reference networks and this has been interpreted in [8, 2] as the rich nodes are tightly interconnected. However, the fact is that the top 20 best-connected nodes are sparsely interconnected in both the original network and the reference networks.

We remark that the rich-club ordering of a network is a relative property which is based on the comparison with reference networks generated from the network itself. It should not be confused with the rich-club structure which is measured and compared between different networks by the rich-club coefficient. Furthermore, as pointed out in [10], for both the AS–Internet and Collaborations network, the range of $\phi_{\mathrm{ran}}(k)$ increases as k increases, hence any assertion based upon the relationship between $\langle \phi_{\mathrm{ran}}(k) \rangle$ and $\phi(k)$ should be statistically tested.

Here we show that the null model based on the maximal random networks cannot detect the rich-club phenomenon, i.e. whether there is a dynamic mechanism behind the formation of a rich–club. Consider the preferential attachment mechanism introduced by Barabási–Albert (BA) [4]. The preferential attachment correlates the age of a node with its connectivity, i.e. 'rich–gets–richer'. As new nodes join the network the old nodes become richer. However if two old nodes do not already share a link, they will never acquire a new one during the network growth. This is to say, the BA growth mechanism is irrelevant to the formation of a rich–club. Figure 2 shows that if a BA network grows from a fully connected seed, i.e. a clique, it will contain a fully connected rich–club; if it grows from a poorly connected seed, e.g. a ring, it will have a poorly connected rich–club. The null model $\langle \phi_{\mathrm{ran}}(k) \rangle$ will produce contradictory results for the two networks generated by the same BA growth mechanism. The null model in this case reflects the connectivity of the initial seeds not the growth mechanism.

4 The assortative reference network

We can measure whether the maximal random networks used in the above null model discount the degree-degree correlation in the original network by examining the nearest-neighbours average degree of k-degree nodes [24], $k_{\mathrm{nn}}(k) = \sum_{k'=1}^{k_{max}} k' P(k'|k)$. If $k_{\mathrm{nn}}(k)$ is an increasing function of k, a network is assortative [19]; and if $k_{\mathrm{nn}}(k)$ is a decreasing function of k, a network is disassortative. If $k_{\mathrm{nn}}(k)$ is neither an increasing nor an decreasing function of k then a network is uncorrelated, more specifically, the degree of a node is independent of its neighbours' degree. If the maximal random networks are less correlated than the original network, the slope of $k_{\mathrm{nn}}(k)$ should be less pronounced than the original network. Figure 3 compares $k_{\mathrm{nn}}(k)$ of the original network and the maximal random networks for four real networks: the AS Internet, the protein interaction network of the yeast *Saccharomyces cerevisiae* [13], the Collaborations–A network, and the giant component of the scientific collaboration network in the area of network theory [20] (Collaborations–B). Fig. 3 shows that the AS–Internet and the Protein networks are disassortative and their null models are also disassortative, whereas the two collaborations networks are assortative and their null models are uncorrelated networks.

The assortative mixing is the inherent property of social networks [22]. Their maximal random networks do not reflect this property. In the following we define a new ensemble of reference networks, called assortative reference networks, which respect both the degree distribution and the assortative mixing of social networks. We obtain such networks by conserving[1] the original network's rich–club structure measured by the rich-club coefficient $\Phi(r)$ as a function of rank r. We study the Collaborations–A network as a typical example of assortative social networks.

4.1 Progressive rewiring

For a given network, we start with a random network having the same number of nodes and links as the original network. A link is selected and rewired at random[2]. We then evaluate the square

[1] Note that conserving $\Phi(r)$ does not imply that the degree distribution $P(k)$ and the rich-club coefficient $\phi(k)$ are the same as in the original network.

[2] We avoid creating self-loop, duplicate link, or isolated node.

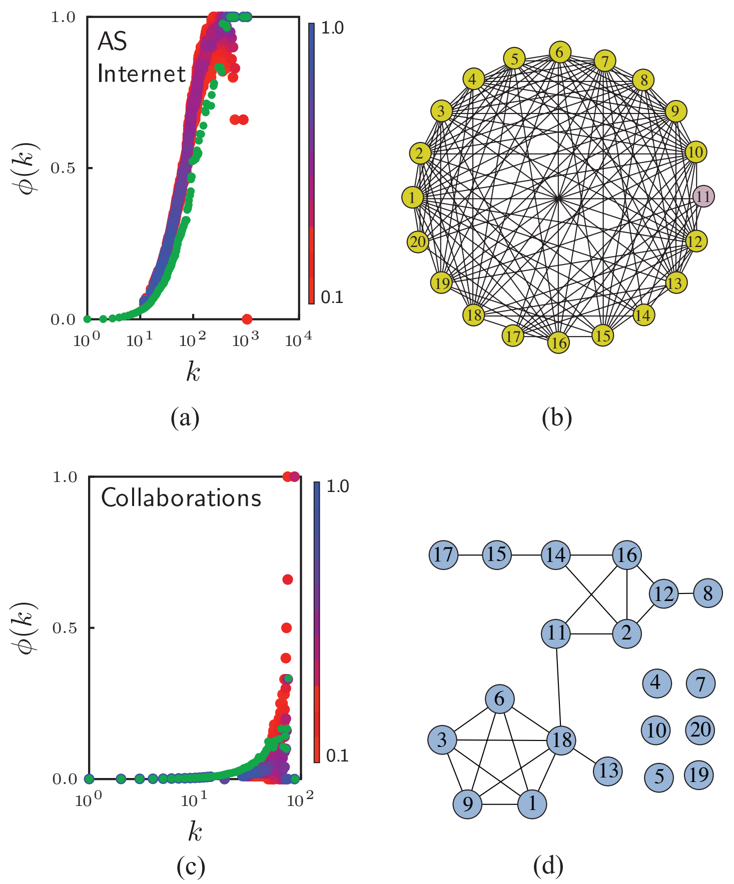
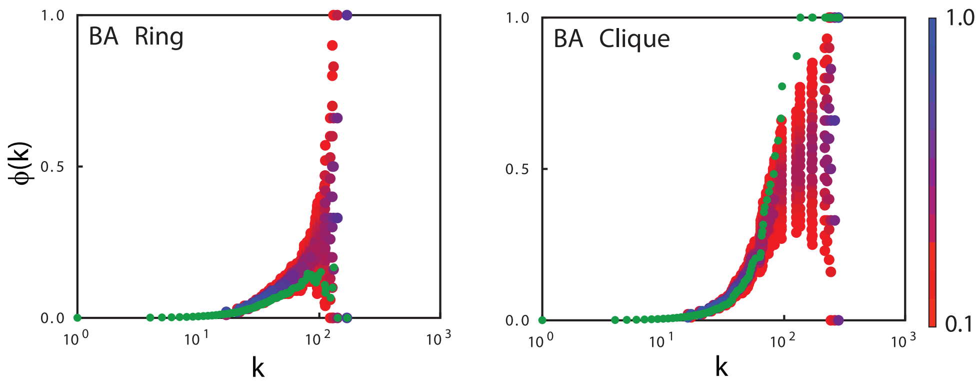
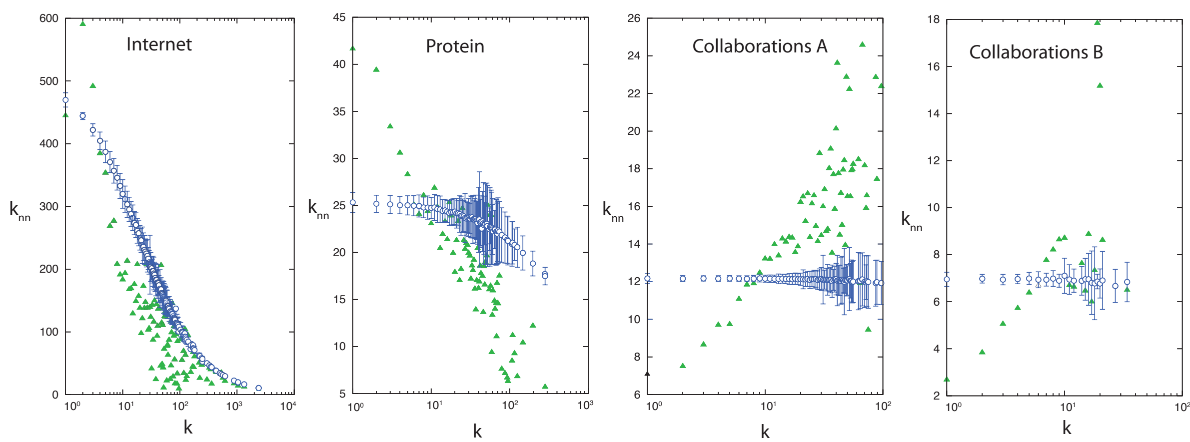


Figure 1: The rich–club coefficient $\phi(k)$ (green) and $\phi_{\text{ran}}(k)$ (red to blue) for the (a) AS–Internet and (c) the Collaborations network. For a given k, the range of values of $\phi_{\text{ran}}(k)$, obtained from 10^3 maximal random networks, are divided into 200 bins such that both the dispersion and the frequency of the values can be displayed in the same graph. The frequency scale ranges from seldom (0.1, red) to often (1.0, blue). The null model $\langle \rho_{\text{ran}}(k) \rangle$ corresponds to the 'bluest' dots. (b) and (d) show the interconnections between the 20 best-connected nodes in the two networks.



(a) (b)

Figure 2: The rich–club coefficient for two Barabási–Albert networks. Both networks have 10^4 nodes and are grown from network seeds of 10 nodes: one seed is a ring (BA–Ring) and the other is a fully connected clique (BA–Clique). The colour scheme is the same as in Fig. 1.



(a) (b) (c) (d)

Figure 3: Nearest-neighbours average degree of k-degree nodes, $k_{\mathrm{nn}}(k)$, for the original network (green) in comparison with the average and standard deviation of $k_{\mathrm{nn}}(k)$ for 10^3 maximal random networks (blue). (a) The AS–Internet, (b) the Protein, (c) the Collaborations–A and (d) the Collaborations–B networks.

error $\Delta = \sum_{r=1}^{N}[\Phi(r) - \Phi^*(r)]^2$, where $\Phi(r)$ is the original rich-club coefficient and $\Phi^*(r)$ is the rich-club coefficient of the rewired random network. If the rewiring decreases the value of Δ, then the rewiring is accepted; otherwise it is rejected. This procedure continues until Δ is small.

The assortative reference network obtained using this method not only conserve the rich-club structure, but also resembles the degree distribution and the assortative mixing of the original network (see Figures 4(a) and (b)). This is due to the structural constrain between the rich-club structure and the degree-degree correlation (see Eqs. 4 and 5). Basically, preserving $\Phi(r)$ of the r richest nodes means preserving the density of connections between nodes with the r highest degrees. Community detection [21] is based on the comparison between the density of connections of the original network and reference networks. The assortative reference network could be useful for detecting community structures of social networks, which are inherently assortative.

4.2 Analytical solution

Here we give the analytical solution for the assortative reference network. From the definition of the rich–club coefficient $\Phi(r)$ (see Eq. 1), the number of links that the node with r shares with the $r-1$ nodes of smaller ranks is

$$E_r = E_{\leqslant r} - E_{\leqslant r-1} = \Phi(r)\frac{r(r-1)}{2} - \Phi(r-1)\frac{(r-1)(r-2)}{2}. \tag{6}$$

Assuming the E_r links are randomly distributed between the $r-1$ nodes, the probability that node r has a link with node r', where $r' < r$, is

$$\mathcal{P}(r) = \frac{E_r}{r-1}. \tag{7}$$

Thus the probability that there is a link between two nodes with ranks i and j is

$$p_{ij} = \begin{cases} \mathcal{P}(i), & \text{if } i > j \\ \mathcal{P}(j), & \text{if } i < j, \end{cases} \tag{8}$$

The above equation satisfies the property that $\sum_{i=1}^{N}\sum_{j=1}^{N} p_{ij} = 2L$, where L is the total number of links of the original network. For a random network which has the same number of nodes N and the same rich-club structure $\Phi(r)$ as the original network, the degree of node i can be approximated by

$$k_i \approx \sum_{j=1}^{N} p_{ij}, \tag{9}$$

and the nearest-neighbours average degree of k-degree nodes is approximated by

$$k_{\rm nn}(k) \approx \frac{1}{N_k}\sum_{i;k_i=k}\left(\frac{1}{k}\sum_{j=1}^{N} p_{ij}k_j\right), \tag{10}$$

where the sum with index i adds the nodes that have degree equal to k and N_k is the number of k-degree nodes.

Figures 4(c) and (d) confirms that the analytical solution of the assortative reference network has similar $P(k)$ and $k_{\rm nn}(k)$ (obtained from Eqs. 9 and 10) as the original network. If the original

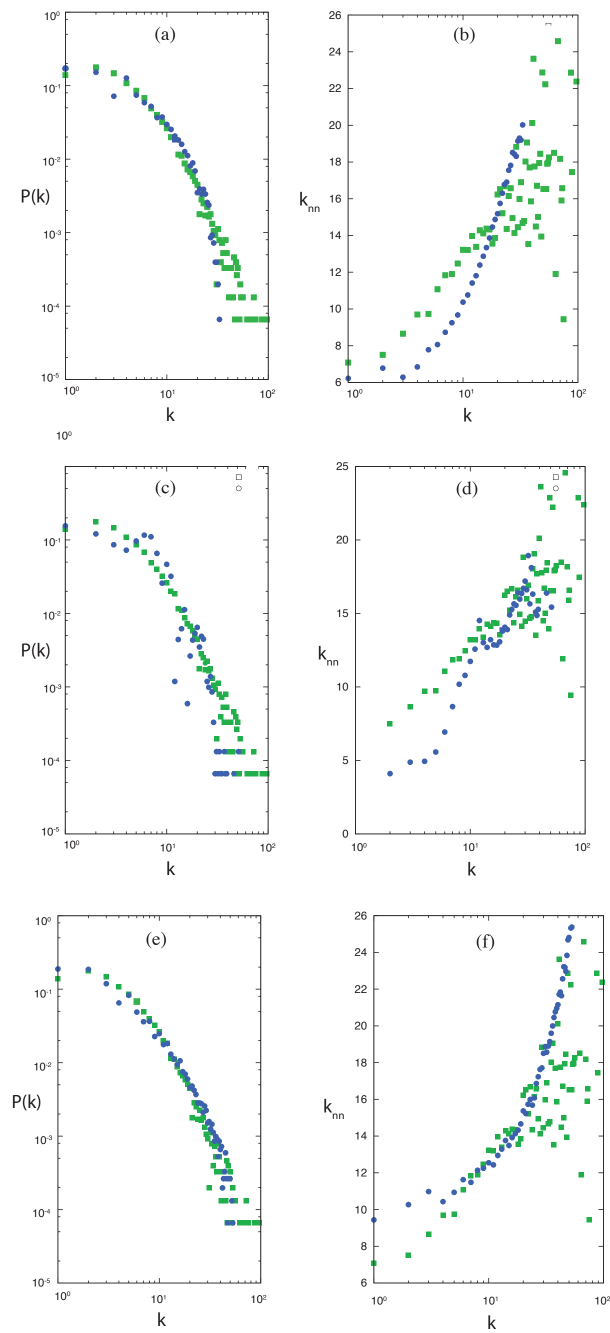


Figure 4: Assortative reference networks for the Collaborations–A network. (a) (c) and (e) show the degree distribution $P(k)$; and (b) (d) and (f) show the nearest-neighbours average degree of k-degree nodes, $k_{\mathrm{nn}}(k)$. (a) and (b) are obtained from the progressive rewiring; (c) and (d) are from the analytical solution; and (e) and (f) are from the rank-based preferential attachment. The original network is shown in green and the reference network is in blue.

network is assortative, i.e. high-degree nodes (with small r) tend to connect with high-degree nodes, we can expect from Eq. (7) that the assortative reference network will also be assortative. However if the original network is disassortative, i.e. high-degree nodes tend to connect to low-degree nodes (with large r), Eq. (7) will not capture this property as it does not favour the connectivity between high-degree nodes and low-degree nodes.

Rank–based preferential attachment

Eq. (8) assumes a uniform probability for node r to attach to any node r' with $r' < r$. Here we modify Eq. (8) to include a preferential attachment mechanism and therefore favour the attachment to richer nodes with smaller r,

$$p_{ij} = \begin{cases} 2j\mathcal{P}(i)/(i+1), & \text{if } i > j \\ 2i\mathcal{P}(j)/(j+1), & \text{if } i < j. \end{cases} \tag{11}$$

This expression was obtained by using $\sum_{i=1}^{N} i = N(N+1)/2$ such that $2P(r)(r+1)\sum_{i=1}^{r} i = E_r$. This is a preferential attachment based on the relative rank of nodes. Figures 4(e) and (f) show that $P(k)$ and $k_{\rm nn}(k)$ estimated from this rank–based preferential attachment are very close to the original network.

This rank–based preferential attachment puts across a different way to study the collaborations network in the context of its reference network. The original network is a member of an ensemble which is defined by the density of connections between a referent group (rich nodes) and the total number of nodes. If the probability of connection between two nodes is related to their rank difference, then the collaborations network looks like a typical member of the ensemble. This provides a different way to conjecture *how* collaborations between scientists arise. In general, a scientist will prefer to work with a scientist of the same or higher status, where status is a relative concept[3].

5 The reference network preserving $P(k)$ and $\phi(k)$

Recently we introduced another ensemble of reference networks which are obtained by randomly reshuffling link–pairs of the original network with the restriction that both $P(k)$ and $\phi(k)$ are preserved for all k [32]. Consider a pair of links with end nodes n_1, n_2, n_3 and n_4 with degrees $k_1 < k_2 < k_3 < k_4$ respectively. If node n_1 is linked to n_2 and n_3 is linked to n_4, we call they are assortatively wired. If a randomly chosen pair of links are assortatively wired they are discarded and a new pair is considered; otherwise the four end nodes of the links are reshuffled at random. The procedure is repeated for a large number of times. Reshuffling a pair of links which are not assortatively wired does not changes the number of links between nodes with degrees equal or larger than k, hence $\phi(k)$ is preserved. Notice that this randomisation procedure also conserves the degree distribution $P(k)$. Fig. 5 shows that the original network and the reference networks obtained from this method show similar degree-degree correlation, and therefore have similar network structure.

[3]This does not explain *why* the higher–status scientist will agree to collaborate, perhaps to carry on more work, or to keep his/her high status.

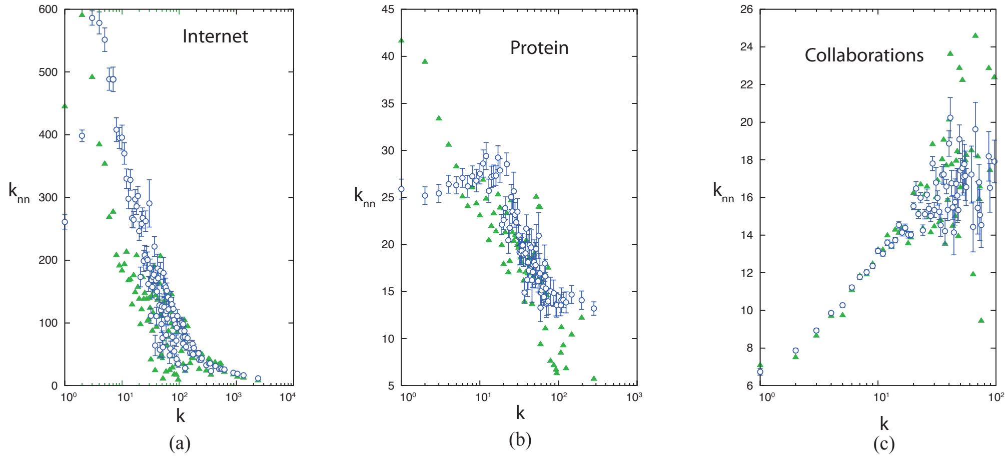


Figure 5: The reference network preserving $P(k)$ and $\phi(k)$ for (a) the Internet, (b) the Protein and (c) the Collaborations–A networks. Nearest-neighbours average degree $k_{\text{nn}}(k)$ of the original network is in green and that of the reference networks is in blue. For each network, 10^3 reference networks are generated and averaged. The error bars give the standard deviation for each value.

5.1 Correlation profile between rich nodes

The above ensemble of reference networks allows us to assess the correlation between rich nodes in a network in relative to the randomised version of the network having similar network structure. The correlation measures how important a connection between two rich nodes is in terms of the organisation of the rich–club and the global structure.

Consider a_{ij} as the ij-th entry of the adjacency matrix describing the original network, and $a'_{ij}(m)$ is the ij-th entry of the adjacency matrix of the m-th reference network obtained by the above randomisation process, where M is the total number of the reference networks. The frequency probability that there is a connection between nodes i and j in the M reference networks is

$$q_{ij} = \frac{1}{M} \sum_{m=1}^{M} a'_{ij}(m). \tag{12}$$

The correlation profile is obtained by evaluating $b_{ij} = a_{ij} - q_{ij}$. The case $b_{ij} = 0$ happens if the connectivity of the original network and the reference networks is the same. This case could happen if (1) the original network and the reference networks never have a link between nodes i and j; or (2) the original network and the reference networks always have a link between nodes i and j. The case $b_{ij} = 1$ means the original network has a link between nodes i and j but this link never appears in the reference networks. The case $b_{ij} = -1$ means that in the original network there is not a link between nodes i and j but this link always appears in the reference networks.

Figure 6 shows the correlation between the rich nodes in three real networks. The nodes are labelled by their rank. As the case $b_{ij} = 0$ can represent two different situations we label the profile using two colour codes. The blue and white squares ($b_{ij} = 0$) are the links that define the backbone of the network. The existence, or not, of these links is fundamental for the network structure and perhaps network functionality [5].

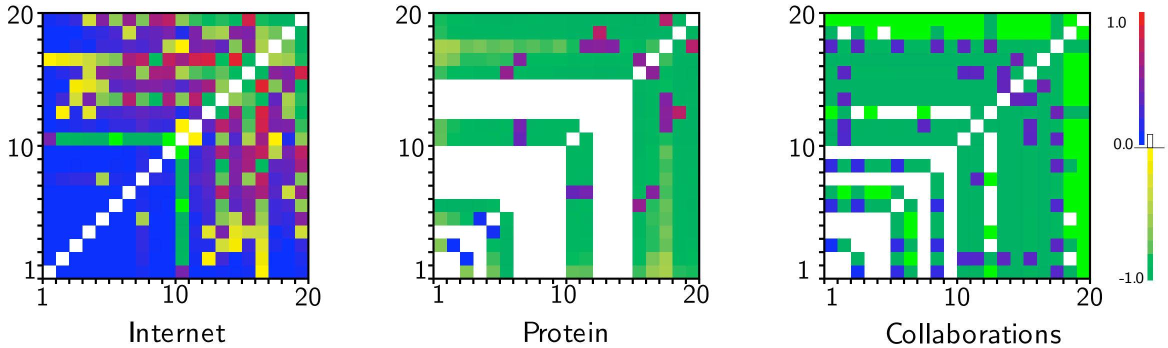


Figure 6: Correlation profiles for the top 20 richest nodes for the Internet, Protein and Collaborations–A networks. For each network, the profile is obtained from 10^3 reference networks preserving $P(k)$ and $\phi(k)$. The colour codes are: from blue to red ($b = 0 \rightarrow b = 1$) labels the chance that a link exist in the original network but not in the reference networks; from white to green ($b = 0 \rightarrow b = -1$) labels the chance that a link does not exist in the original network but exist in the reference networks.

In the Internet profile, the 7 richest nodes are always fully interconnected with each other in the original network and in each of the reference networks (blue). Such a rich–club clique is an important structure for the Internet because it provides a large number of shortcuts for the delivery of data traffic and makes the average shortest path between any two nodes as small as just over 3 hops[4]. We conjecture that the existence of such the rich–club clique is a fundamental property if the ensemble represents networks that deliver traffic efficiently. Another interesting behaviour in the Internet's correlation profile is that there is a link between nodes with ranks 15 and 17 that is present in the original data but appears very rarely in its randomised version (bright red square). Whether this reflects erroneous or incomplete measurements, or for some reason these nodes interact with each other against the odds, makes this interaction of particular interest.

The profile of the Protein network shows a completely different behaviour. The rich nodes in the ensemble tend not to connect with each other. The white vertical and horizontal bands show there are interactions between proteins that never occur in the original network or the reference networks. To decide if this structure is reflecting impossible protein interactions will require a more specific analysis.

The profile of the Collaborations–A network shows that some researchers tend to always collaborate with each other (blue) and others never collaborate (white), possibly reflecting the friendship and rivalries between researchers. What is never present is a collaboration between two researchers that it should not happen (no red squares in the profile), i.e. hindsight cannot be detected via the reference networks.

[4]The AS-Internet contains more than 9 thousand nodes

6 Conclusions

The rich nodes of a network tend to dominate the organisation of network structure so it is relevant to understand if their interconnectivity is due to chance or to an organisational principle. A technique to analyse this issue is to compare the rich–club coefficient against the equivalent structure evaluated from an ensemble of reference networks. The ensemble is generated via the restricted randomisation of the original network.

We evaluate a recently proposed null model which is based on the ensemble of maximal random networks conserving the degree distribution of the original network. We remark that one should not confuse the rich-club structure of a network with the rich-club ordering detected by the null model.

We presented a method to generate, from an assortative network, an ensemble of assortative reference network with the condition that the rich-club coefficient $\Phi(r)$ is conserved. The assortative reference network could be used for detecting the community structure of social networks which are inherently assortative. The assortative reference networks also provides a different way to explain the evolution of social collaborations.

We also presented a method to generate reference networks which preserve both the degree distribution and the rich-club coefficient $\phi(k)$. The ensemble of such reference networks have a similar structure as the original network. We use them to study the correlations between the rich nodes and pinpoint which connections between them form the backbone of the network.

Finally we'd like to remark that in the all cases of ensembles, the reference networks are generated from the original network. Properties of the ensemble strongly depend on the properties of the original network. Any deduction based on these ensembles should take into consideration this dependance.

Acknowledgments

RJM would like to thank the EPSRC UK (EP/C520246/1) for support. SZ is supported by The Royal Academy of Engineering and the UK Engineering and Physical Sciences Research Council (EPSRC) under grant no. 10216/70.

References

[1] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:49–97, 2002.

[2] L. A. Nunes Amaral and R. Guimerà. Lies, damned lies and statistics. *Nature Physics*, 2:75–76, 2006.

[3] S. Bar, M. Gonen, and A. Wool. An incremental superlinear preferential internet topology model. In *Passive and Active Network Measurement: 5th International Workshop, PAM '04,*, Proceedings (Lecture Notes in Computer Science), Antibes Juan-les-Pins, France, 2004. Springer–Verlag.

[4] A. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509, 1999.

[5] G. Bianconi. Entropy of randomized network ensembles. *Europhysics Lett.*, 81:28005, 2008.

[6] M. Boguñá, R. Pastor-Sartorras, and A. Vespignani. Cut–offs and finite size effects in sacle–free networks. *Eur. Phys. J. B*, 38:205–210, 2004.

[7] S. Bornholdt and H. Ebel. World–wide web scaling exponent from Simon's 1955 model. *Phys. Rev. E*, 64:035104, 2001.

[8] V. Colizza, A. Flammini, M. A. Serrano, and A. Vespignani. Detecting rich–club ordering in complex networks. *Nature Physics*, 2:110–115, 2006.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power–law relationships of the Internet topology. *Comput. Commun. Rev.*, 29:251–262, 1999.

[10] Z-Q. Jiang and W-X. Zhou. Statistical significance of the rich-club phenomenon in complex networks. *New J. Phys.*, 10:043002, 2008.

[11] P. Krapivsky and D. Krioukov. Scale-free networks as pre-asymptotic regimes of super-linear preferential attachment. *Phys. Rev. E*, 78:026114, 2008.

[12] B.F.J. Manly. *Randomization, Bootstrap and Monte Carlo Methdos in Biology.* Chapman & Hall, 2nd edition, 1997.

[13] S. Maslov and K. Sneppen. Specificity and stability in topology of protein networks. *Science*, 296:910–913, 2002.

[14] S. Maslov, K. Sneppen, and A. Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the Internet. *Physica A*, 333:529–540, 2004.

[15] N. Masuda and N. Konno. VIP–club phenomenon: emergence of elites and masterminds in social networks. *Social Networks*, 28:297, 2006.

[16] Julian J. McAuley, Luciano da Fontoura Costa, and Tiberio S. Caetano. The rich-club phenomenon across complex network hierarchies. 2007.

[17] M. E. J. Newman. Scientific collaboration networks. i. network construction and fundamental results. *Physical Review E*, 64(016131):016131, 2001.

[18] M. E. J. Newman. Scientific collaboration networks. ii. shortest paths, weighted networks,and centrality. *Physical Review E*, 64(016132):016132, 2001.

[19] M. E. J Newman. Assortative mixing in networks. *Phys. Rev. Lett.*, 89(208701):208701, 2002.

[20] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, 2006.

[21] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, 2004.

[22] M. E. J. Newman and J. Park. Why social networks are different from other types of networks. *Physical Review E*, 68:036122, 2003.

[23] T. Opsahl, V. Colizza, P. Panzarasa, and J. J. Ramasco. Weighted rich–club effect: exploring the strength among top nodes. arXiv:0804.0417, 2008.

[24] R. Pastor-Satorras, A. Vázquez, and A. Vespignani. Dynamical and correlation properties of the Internet. *Phys. Rev. Lett.*, 87(258701):258701, 2001.

[25] M. A. Serrano. Rich-club vs rich-multipolarization phenomena in weighted networks. *Phys. Rev. E*, 78:026101, 2008.

[26] H. A. Simon. On a class of skew distribution functions. *Biometrika*, 42(3/4):425–440, 1955.

[27] S. Valverde and R.V. Solé. Self–organization versus hierarchy in open–source social networks. *Phys. Rev. E*, 76:046118, 2007.

[28] S. Wuchty. Rich–club phenomenon in the interactome of *p. falciparum*– artifact or signature of a parasitic life style? *PLoSone*, 2, 2007.

[29] S. Zhou and R. J. Mondragón. Accurately modeling the Internet topology. *Phys. Rev. E*, 70:066108, 2004.

[30] S. Zhou and R. J. Mondragón. Redundancy and robustness of the AS–level Internet topology and its models. *IEE Electronic Letters*, 2:151–152, 2004.

[31] S. Zhou and R. J. Mondragón. The rich-club phenomenon in the internet topology. *IEEE Communications Letters*, 8(3):180–182, March 2004.

[32] S. Zhou and R. J. Mondragón. Structural constraints in complex networks. *New Journal of Physics*, 9, 2007.

[33] V. Zlatic, G. Bianconi, A. Díaz-Guilera, D. Garlaschelli, F. Rao, and G. Caldarelli. On the rich–club effect in dense and weighted networks. arXiv:087.0893v1, 2008.