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

R. J. Mondragón and S. Zhou

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

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

Abstract

In complex networks the rich nodes are the subset of nodes with high degree. These well connected nodes tend to dominate the organisation of the network’s structure. In non–evolving networks, a reference network has been used to detect if the connectivity between the rich nodes is due to chance or caused by an unknown mechanism. Chance is represented as a reference network obtained from an ensemble of networks. When compared with the original network the reference network discounts suggests the existence of a well connected rich club beyond structural constraints. Here we revise some of the properties of the ensemble obtained by conserving only the degree distribution and introduce two new reference networks to study the importance of the rich nodes as organisers of the network structure. The first reference network is obtained from an ensemble of networks where all the members of the ensemble have the same rich–club coefficient. The reference network obtained from the ensemble is assortative. We propose that this reference network can be used to study networks where assortativness is a fundamental property, a common case in many social networks. The second reference network is obtained from an ensemble where the members of the ensemble all have the same probability degree distribution and rich–club coefficient. The reference network obtained from this ensemble has a very similar structure to the original network. This ensemble can be used to quantify correlations between the rich nodes and pinpoint which links are the backbone of the network’s structure.

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 tend to be tightly interconnected between themselves, forming a rich–club [27]. The rich–club is an oligarchy in that it dominates the organisation of the whole network. In scale–free networks [1] the connectivity of the rich–club plays an important role in the functionality of the network, for example in the transmission of rumours in social networks [13] or the efficient delivery of information in the Internet [26]. If the connectivity of the rich nodes is so relevant to the structure of a network, could this connectivity have been caused by chance. The answer to this question depends on how chance is defined.

Many of the complex networks studied in the literature are single networks in the sense that their structure is unique not one of several. In general, we do not have the equivalent to a physical law to verify if the expected statistical measures obtained from a single network are expected or extraordinary. Instead, a common technique to analyse the properties of a single network is to use statistical randomisation methods [11] to create a reference network which is used for comparison.
purposes. The procedure consist of using the observed network to generate an ensemble of networks via randomisation. The reference network, or null model, is generated from this ensemble and it is used to assess the significance of a property of the network. In its simplest form, the randomisation assume that the data are independent hence the shuffling of the connections between the nodes are equally likely. However, this assumption is not valid in many networks as it does not take into account the intrinsic structure of the network. As the first step in describing and discriminating between different networks is to measure the degree distribution $P(k)$, the fraction of nodes in the network with degree $k$, this distribution of degrees is taken into account by using a restricted randomisation procedure. The restriction is to reshuffle the connections of the nodes without changing the degree distribution of the nodes $P(k)$ \cite{12}.

The density of connections between the rich nodes is measured by the rich–club coefficient \cite{27}. The rich–club coefficient and its generalisations \cite{7,14,25,30,20,22} has proved to be a useful measure when studying complex networks. For the density of connections between rich nodes, a null model obtained from a randomisation procedure \cite{12}, has been proposed as the reference point to decide if the connectivity of the rich nodes is due to chance or due to an “organisational principle” \cite{7}. The purpose of this article is to clarify some of the properties of the ensemble produced by this randomisation procedure. In particular, any deduction based on the ensemble should not disregard the dependence of the ensemble with the original network. Also we introduce two other ensembles based on the connectivity of the rich nodes. The first ensemble is the set of networks that all have the same density of connections between rich nodes. The networks of the ensemble are assortative \cite{17} so the reference network obtained from this ensemble could be useful in the analysis of social network, which tend to be assortative. The second ensemble is the set of networks that have the same degree distribution and density of connections between rich nodes. The networks of the ensemble have similar degree–degree correlation. The reference network obtained from the ensemble pinpoints the importance of individual links between rich nodes as organisers of the network’s structure.

**Rich–club coefficient**

As a function of the node’s rank, where the nodes are labelled in decreasing order of their degree, the density of connections between the $r$ best connected nodes is given by the rich–club coefficient \cite{27}

\[
\Phi(r) = \frac{2E_r}{r(r-1)},
\]

where $E_r$ is the number of links among the 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 the degree the density of connections between nodes with degree greater or equal to $k$ is \cite{7}

\[
\phi(k) = \frac{2E_{\geq k}}{N_{\geq k}(N_{\geq k} - 1)},
\]

where $N_{\geq k}$ is the number of nodes with degrees at least $k$ and $E_{\geq k}$ is the number of links among these $N_{\geq k}$ 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, their degree is used to group the nodes together into subsets. If $r^*_k$ is the node with degree $k$ such that $r^*_k + 1$ is the rank of the node with degree
has a neighbour with degree $k$ is a subset of $\{\Phi(r); r = 1, \ldots, N\}$, where $k_{max}$ is the maximum degree in the network and $N$ is the number of nodes. Two networks can have the same $\phi(k)$ and the same probability degree distribution $P(k)$ for all $k$, but different $\Phi(r)$. Originally the rich–club was defined as the set of nodes that are tightly interconnected. This definition is vague as ‘tight’ is a relative concept. Recently, Valverde and Solé [24] 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 $k_c$ (or $r_c$) in $\phi(k)$ (or $\Phi(r)$), this crossover characterise the rich nodes.

The network structure can be described 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 $k'$. In terms of the rich–club coefficient, the rich–club structure refers to the density of connections across the hierarchies of nodes. The rich–club structure is fully defined by the degree–degree correlation [7]

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

where $\langle k \rangle$ is the average degree. Conversely, fixing the rich–club structure will constrain the network’s degree–degree correlation [28, 10]. From the definition of the rich–club coefficient 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

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

This quantity can be related to $P(k, k')$ by using the conditional probability that a node with degree $k$ has a neighbour with degree $k'$, i.e. $P(k'|k) = \langle k \rangle P(k, k')/(kP(k))$, where $\sum_k P(k'|k) = 1$. In terms of the conditional probability

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

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 $k$. Hence, fixing $\phi(k)$ for all $k$, constrains the possible values of $P(k'|k)$ or equivalently of $P(k,k')$.

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

The rich–club phenomenon [27] refers to the 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 joining an existing rich–club. If an evolving network model is used to reproduce the connectivity of a real network, the phenomenon is introduced into the model to generate a network with similar rich–club structure as the real network [6, 27, 29]. The dynamics to create a rich–club were first introduced in 1942 by Simon [23]. Simon’s model was based on the addition of new nodes an the addition of new links between nodes that belonged to the same class, where a class is the set of nodes with the same degree. Bornholdt and Ebel [6] have pointed out that this growth mechanism allows different growth rate for different classes of nodes and hence it can create a rich–club core. Recently Krapivsky and Krioukov [10] showed how the inclusion of the rich–club phenomenon in evolving networks drastically constraints their structure.
The term rich–club phenomenon has been used when analysing non–evolving (closed) networks. By comparing the original network with a reference frame network, several authors have concluded the existence (or not) of a rich–club phenomenon in the network under study. For closed networks we would like to use the term rich–club ordering instead of rich–club phenomenon. Where the rich–club ordering measures an organisational principle that leads to an increase in the density of connections between rich nodes in a more pronounced way than in the reference network.

The relative rich–club reference network

Colizza et al. addressed whether the presence of a rich–club in a network is due to an organisational principle or their connectivity is just due to chance. To do so, they compared the original rich–club coefficient \( \phi(k) \) with the randomised rich–club coefficient \( \phi_{\text{ran}}(k) \). The coefficient \( \phi_{\text{ran}}(k) \) is obtained from an ensemble of maximal random networks. The ensemble is generated by randomly reshuffling link–pairs of the network under study. 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 \( P(k) \). Colizza et al. showed that the “normalised rich–club coefficient” \( \rho_{\text{ran}}(k) = \phi(k)/\phi_{\text{ran}}(k) \) discounts the structural correlation imposed by the finite–size effects and that, \( \rho_{\text{ran}}(k) \) can be use to discern if there is an organisational principle behind the formation of a rich–club. If \( \rho_{\text{ran}}(k) > 1 \) then 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.

We prefer to call \( \rho_{\text{ran}}(k) \) the relative rich–club coefficient instead of the normalised rich–club coefficient. The reason is that, as the number of nodes with degree \( k \) does not changed by the randomisation procedure, \( \rho_{\text{ran}}(k) = E_{\geq k}/E_{\text{ran},\geq k} \) is the ratio of links between the original network and the reference network. Hence \( \rho_{\text{ran}}(k) \) does not give information about the density of connections of the original network (which is measured by the rich–club coefficient). This simple observation is relevant because there has been some confusion in the literature what does the rich–club coefficient and the relative rich–club are measuring. For example fig. 1 shows the rich–club structure for the Internet network at the Autonomous System (AS–Internet) level and for the scientific collaboration network (Collaborations) in the area of condensed matter Physics. Figs. 1(a) (Internet) and 1(c) (Collaborations) show in green \( \phi(k) \) and, from red to blue, the range of values for the randomised \( \phi_{\text{ran}}(k) \) obtained from \( 10^4 \) realisations. The frequency of a particular value of \( \phi_{\text{ran}}(k) \) is labelled with different colours. From seldom (0.1, red) to often (1.0, blue). The null model \( \langle \rho_{\text{ran}}(k) \rangle \) is obtained by averaging over all the realisation of the random process and corresponds to the ‘bluest’ dots. Figs. 1(b) and 1(d) show the connectivity of the 20 best connected nodes for both networks. For the AS network from the observation that \( \langle \phi_{\text{ran}}(k) \rangle > \phi(k) \), i.e. \( \rho_{\text{ran}}(k) < 1 \), for almost all values of \( k \), it was stated that the Internet does not have a rich–club ordering creating the misinterpretation that the rich nodes were not tightly connected. On the contrary, fig. 1(b) shows the connectivity of the 20 highest degree nodes, and it is evident that this nodes are densely interconnected. Even more, the seven highest degree nodes form a clique. For the Collaborations network the property \( \langle \phi_{\text{ran}}(k) \rangle < \phi(k) \) insinuates that the top scientists form tighter collaborations compared to the reference network, this has been interpreted as the rich nodes are tightly interconnected. However in this case for both, the original network and its reference network, the top 20 rich nodes are sparsely connected (fig. 1(d)). Notice that for both, the AS–Internet and Collaboration networks, the range of \( \phi_{\text{ran}}(k) \) increases as \( k \) increases, hence any
Figure 1: (a) The rich–club coefficient $\phi(k)$ (green) for the AS–Internet and $\phi_{\text{ran}}(k)$ (red to blue). For a given $k$, the range of values of $\phi_{\text{ran}}(k)$, obtained from $10^3$ realisations, are divided into 200 bins such that the dispersion and frequency of $\phi_{\text{ran}}(k)$ can be displayed in the same graph. The frequency scale ranges from seldom (0.1, red) to often (1.0, blue). (b) Connectivity of the rich–club formed by the 20 nodes with the highest degree.

The reference network obtained from the maximally randomised method will not disclose the mechanism behind the formation of the rich–club of the original network. Consider the preferential attachment mechanism introduced by Barabási–Albert (BA) \cite{9}. 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 share a link, they will never acquire one. The BA growth mechanism is irrelevant to the formation of a rich–club. Fig. 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. In the figure, the behaviour of $\phi_{\text{ran}}(k)$ reflects the connectivity of the initial conditions not the dynamics of the growth mechanism. Fig. 2 also shows that the rich–club coefficient of the ensemble strongly depends on the connectivity of the rich nodes. If the ensemble is used to gen-
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 original $\phi(k)$ is shown in green. The null model $\langle \rho_{ran}(k) \rangle$ corresponds to the ‘bluest’ dots.

In order to characterise the global behaviour of the network then, we have to be aware that a specific minor set of links (the ones between rich nodes) could change this global characterisation.

The rich–club density reference network

The method of reshuffling link–pairs used to generate the maximally random network discounts the structural correlations due to the finite size of the network. From the average degree of nearest neighbours [21] $k_{nn}(k) = \sum_{k'=1}^{k_{max}} k' P(k'|k)$, it is possible to measure if the reshuffling method decreases the degree–degree correlation of the null model relative to the original network. If $k_{nn}(k)$ is an increasing function of $k$ the network is assortative and if $k_{nn}(k)$ is a decreasing function of $k$ the network is disassortative. If $k_{nn}(k)$ is not an increasing or decreasing function of $k$ then the network is uncorrelated, more specifically, the degree of a node is independent of its neighbours’ degree.

Figure 3 shows $k_{nn}(k)$ of the original network and $\langle k_{nn}(k) \rangle$, the average of the degree of nearest neighbours for the maximally random networks. The four real networks shown in the figure are the Internet network at the autonomous system level [8, 21], the protein interaction network of the yeast Saccharomyces cerevisiae [12], the scientific collaboration network in the area of condensed matter Physics [15, 16] (referred as Collaborations A) and the giant component of the network of collaborators of scientists working on network theory [18] (referred as Collaborations B). Fig. 3(a-b) shows that the AS–Internet and the Protein networks, which are disassortative, their null models are also disassortative. For the two networks of collaboration, which are assortative, their null models are uncorrelated networks (fig 3(c-d)).

The maximal randomisation method assumes that the reshuffling of links between nodes are equally likely and therefore can be exchange freely. If assortativeness is the property that characterise the collaboration networks, then their null models do not reflect the correlation between the rich
Figure 3: Average degree of nearest neighbours $k_{nn}(k)$ of the original network (green) and average and standard deviation of $k_{nn}(k)$ for the maximally random networks obtained from $10^3$ realisations (blue). (a) The AS–Internet, (b) the Protein, (c) the Collaborations A and (d) the Collaborations B networks.

nodes. The connectivity of the rich nodes is not independent of the rearrangement of the links so the reshuffling does not capture the inherent structure of the collaboration networks.

To generate an assortative null model from the original network we define the ensemble as the set of networks that have the same number of nodes $N$ and the same rich–club structure $\Phi(\phi)$. Recalling that the nodes are ranked in decreasing order of their degree, conserving $\Phi(\phi)$ means conserving the density of connections of the rich nodes. Note that conserving $\Phi(\phi)$ does not imply that the rich club $\phi(k)$ or the degree distribution $P(k)$ of the reference network are the same as in the original network.

An approximation to the reference network

From the definition of the rich–club coefficient the number of links than node $r$ shares with nodes of smaller rank is

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

The reference network is defined by considering that the $E_r$ links are equally distributed between the $r$ nodes that have rank less than $r$, that is the probability that node $r$ has a link with a node $r' < r$ is

$$P(r) = \frac{E_r}{r}. \quad (7)$$

The probability that there is a link between node with rank $i$ and node with rank $j$ is

$$P_{ij} = \begin{cases} P(i), & \text{if } i > j \\ P(j), & \text{if } i < j. \end{cases} \quad (8)$$
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. The degree of node \( i \) is approximately \( k_i \approx \lceil \sum_{j=1}^{N} p_{ij} \rceil \). The average degree of nearest neighbours is approximated by

\[
\langle k_{nn} \rangle \approx \frac{1}{N_k} \sum_{i,k_i=1}^{k_{max}} \left( \frac{1}{k} \sum_{j=1}^{N} p_{ij} k_j \right).
\]

(9)

The sum with index \( i \) adds the nodes that have degree equal to \( k \) and \( N_k \) is the total number of these nodes. If the original network is assortative, that is high degree nodes tend to connect with high degree nodes, the reference network will also be assortative. This behaviour is expected from Eq. (7). However if the original network is disassortative, i.e. high degree nodes tend to connect to low degree nodes, eq. (7) will not capture this property as it does not favours the connectivity between low degree nodes with high degree nodes.

Figure 4(a) shows the degree distribution \( P(k) \) for the condensed–matter collaborations networks compared with the degree distribution of its reference network. The degree distribution of the reference network and the original network are similar, but the reference network has a truncated tail. Fig. 4(b) compares the average degree of nearest neighbours for the original networks and its reference network.

We verified that the assumptions that eqns. (6)–(8) are a good approximation to the reference network by constructing from the original network a randomised network with very similar \( \Phi(r) \). For a given network with rich-club coefficient \( \Phi(r) \), we considered a random network having the same number of nodes and links as the original network and rich-club coefficient \( \Phi^*(r) \). A link of the random network is rewired at random and the square error \( \Delta = \sum_{r=1}^{N} [\Phi(r) - \Phi^*(r)]^2 \) is evaluated. If the rewiring decreases the value of \( \Delta \), then the rewiring is accepted; otherwise it is rejected. This procedure continues until \( \Delta \) is small. Figure 4(c-d) show that the random network obtained using this method have similar \( P(k) \) and \( k_{nn}(k) \) to the cases shown in fig. 4(a-b) confirming the validity of the approximation.

**Rank preferential attachment**

Eqn. (8) assumes that a node could attach to any other node of higher degree without any preference. It is easy to modify eqn. (8) to include a preferential attachment mechanism and favour the connectivity to higher degree nodes. For example

\[
p_{ij} = \begin{cases} 
2jP(i)/(i+1), & \text{if } i > j \\
2iP(j)/(j+1), & \text{if } i < j.
\end{cases}
\]

(10)

increases the probability of connections between two nodes relative to their rank. This expression was obtained by using \( \sum_{i=1}^{N} i = N(N+1)/2 \) such that \( 2P(r)(r+1) \sum_{i=1}^{r} i = E_r \) (see eq. (7)). This is a preferential attachment based on the relative rank of the nodes. Fig. 4(e-f) shows that using this linear weight, the tail of \( P(k) \) for the reference network is closer to the original network and it increases its assortativness. This rank–preferential attachment puts across a different way to study the original network in terms 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 (richer nodes) and the total number of nodes. If the probability of connection between two nodes is related to their rank difference, then the collaboration network looks like a typical member of the ensemble.
Figure 4: (a) Probability degree distribution $P(k)$ and (b) average degree distribution $k_{nn}(k)$ for the condense–matter collaborations (green) compared with the reference network (blue). (c) The probability degree distribution and (d) the average degree of the neighbours of a randomised network that has similar rich–club coefficient as the original network. (e) $P(k)$ and (f) $k_{nn}(k)$ using a preferential attachment based on the rank.
The rank–preferential attachment suggest a different way to conjecture how interactions between collaborators could arise. In general, a scientist will prefer to work with a scientist of the same or higher status, where status is a relative concept. However this would not explain why the higher status scientist will agree to the collaboration, perhaps to carry on more work, or to keep his/her high status.

**The correlated rich–club reference network**

Recently Bianconi [4] considered how to characterise the complexity of an ensemble of networks where any member of the ensemble would perform the same task equally well as the real network. The possible tasks that a network can perform are constraint by the global functionality of the network. As structure and functionality are closely linked, the complexity of the ensemble is related to the number of different networks that belong to the ensemble. High complexity corresponds to a small number of networks in the ensemble. Eqns. (4) and (5) show that if the ensemble was defined by having the same degree distribution $P(k)$ and rich–club structure $\phi(k)$ then the networks of the ensemble will have similar degree–degree distribution, i.e. structure hence functionality. As the connectivity of the rich nodes has a large impact on the global structure of the network, the correlation between rich nodes relative to a reference network, will measure how important are the interactions between these nodes in the organisation of the rich–club, and therefore the global structure. We introduce a new null model obtained by randomly reshuffling link–pairs of the original network with the restriction that both $P(k)$ and $\phi(k)$ are preserved [28].

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 called this an assortative wiring. The randomisation procedure to generate the ensemble is: if a randomly chosen pair of links are assortative wired they are discarded and a new pair is considered; otherwise the four end nodes are reshuffled at random. The procedure is repeated for a large number of times. Discarding the pair of links with assortative wiring does not changes the number of links between nodes with degrees 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 randomised networks have similar structure, as measured from the average neighbours degree.

Consider $a_{ij}$ as the $ij$-th entry of the adjacency matrix describing the original network, and $q_{ij}(m)$ is the $ij$-th entry of the adjacency matrix of the $m$-th realisation of the randomisation process, where $M$ is the total number of realisations. For the reference network the frequency that node $i$ is linked to node $j$ is

$$p_{ij} = \frac{1}{M} \sum_{m=1}^{M} q_{ij}(m). \tag{11}$$

The correlation profile is obtained by evaluating $b_{ij} = a_{ij} - p_{ij}$. The case $b_{ij} = 0$ happens if the connectivity of the original network and the randomised networks is the same. This case could happen if; the original network and the randomised networks did not have a link between nodes $i$ and $j$, or, the original network and the randomised 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 randomised networks. The case $b_{ij} = -1$ means that in the original network there is no link between nodes $i$ and $j$ but this link always appears in the randomised networks.

Figure 6 shows the correlation profiles between the rich nodes for the Internet, Protein and
Figure 5: Average degree of nearest neighbours $k_{nn}(k)$ of the original network (green) and average and standard deviation of $k_{nn}(k)$ for the maximally random networks obtained from conserving $P(k)$ and $\phi(k)$ (blue). (a) The AS–Internet, (b) the Protein and (c) the Condensed–matter Collaborations. For the maximally random networks the degree–degree correlation is the average of $10^3$ realisations. The error bars show the standard deviation.

Figure 6: Correlation profiles for the top 20 rich nodes for the Internet, Protein and Collaborations networks. The profile was obtained from $10^3$ realisations of the randomisation procedure. The colour code is: from blue to red ($b = 0 \rightarrow b = 1$) labels the chance that a link exist in the original network but not in the randomised 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 randomised networks.
condensed–matter collaborations networks. The nodes are labelled by their rank, where lower rank means higher degree. As the case $b_{ij} = 0$ can represent two different situations we labelled the profile using two colour scales. From blue to red labels the chance that a link exist in the original network but not in the randomised networks. From white to green labels the chance that a link does not exist in the original network but exist in the randomised networks. The blue and white squares ($b_{ij} = 0$) are the links that define the backbone of the ensemble. The existence, or not, of these links is fundamental for the structure of the networks and probably to their functionality.

In the Internet profile the first 7 nodes are always connected (blue), this clique appears always in the ensemble. In the Internet the rich–club is an important structure because provides a shortcut for the delivery of traffic. The rich–club makes the average shortest path of the Internet very small. We conjecture that the existence of the well connected rich–club is a fundamental property if the ensemble represents networks that deliver traffic efficiently. The Internet’s correlation profile also shows an interesting behaviour, there is a link between nodes 15 and 17 that is present in the original data but appears very rarely in its randomised version (bright red square). If a link exist on the real network but not in the reference network will this reflect erroneous or incomplete measurements, or is it the case that for some reason these nodes interact with each other against the odds, making this interaction of particular interest.

If the profile of the Internet was characterised by the existence of a rich–club, the profile of the Protein shows a completely different behaviour. The rich nodes of the ensemble tend not to connect with each other. The white vertical and diagonal bands, show that there are interactions between the nodes that will never occur. To decide if this structure is reflecting impossible protein interactions will require a more specific analysis.

Finally the Collaborations network profile shows that some researchers tend to always collaborate (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). Hindsight cannot be detected via the reference model.

Conclusions

The rich nodes of a network tend to dominate the organisation of the whole network so it is relevant to understand if their connectivity is due to chance or to an organisational principle. A technique to analyse the connectivity of the rich nodes is to compare its structure with the equivalent structure of a reference network. The reference network is evaluated from an ensemble of networks. The ensemble is generated via the restricted randomisation of the original network.

We observed that the ensemble based on the conservation of the degree distribution $P(k)$ \cite{12} and used to measure the rich–club ordering \cite{7} cannot be used to identify the dynamical mechanism that generated the connectivity of the rich nodes. Due to this result, we propose that the term rich–club phenomenon refers to the dynamical behaviour of evolving networks where a rich club is formed or augmented as the network evolves. In closed networks the term rich–club ordering refers to the property, measured using a reference network, to discern if the density of connections between rich nodes is greater than the random case.

We presented a method to generate, from an assortative network, an assortative reference network. The ensemble is defined by the conditions that the density of connections $\Phi(r)$ and the number of nodes $N$ is conserved. Hence the members of the ensemble have different degree distribution $P(k)$ but equal rich–club coefficient $\Phi(r)$. Community detection \cite{19} is based on the
comparison between the density of connections of the original network and a null model. As the
rich–nodes have such a strong influence in the global structure of the network and the null model
conserves their density and is assortative, this null model can be useful for detecting community
structures of assortative networks, for example social networks.

We presented also how to generate a reference network to study the correlations between the
rich nodes. In this case the restriction that \( P(k) \) and \( \phi(k) \) are the same for all the members of the
ensemble, creates a null model with similar degree–degree distribution as the original network. If
the complexity of a network is related to its functionality and hence to its structure, this ensemble
can be used to pinpoint which connections between the rich nodes form the backbone of the network.

Finally in all these cases, the ensemble defining the reference network is generated from the
original network. The properties of the ensemble strongly depend on the properties of the rich–
club connectivity 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] 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.

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

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

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

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

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

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

[9] Z-Q. Jiang and W-X. Zhou. Statistical significance of the rich-club phenomenon in complex
networks. New J. Phys., 10:043002, 2008.
[10] P. Krapivsky and D. Krioukov. Scale-free networks as pre-asymptotic regimes of super-linear preferential attachment. arXiv.org:0804.1366, 2008.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[26] 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.

[27] S. Zhou and R. J. Mondragón. The rich-club phenomenon in the internet topology. IEEE Communications Letters, 8(3):180–182, March 2004.
[28] S. Zhou and R. J. Mondragón. Structural constraints in complex networks. *New Journal of Physics*, 9, 2007.

[29] S. Zhou and R.J. Mondragón. The positive-feedback preference model of the AS–level internet topology. In *Communications, 2005. ICC 2005. 2005 IEEE International Conference on*, volume 1, pages 163–167Vol.1, 16-20 May 2005.

[30] 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.