

PAPER

## Analyzing complex networks through correlations in centrality measurements

To cite this article: José Ricardo Furlan Ronqui and Gonzalo Travieso *J. Stat. Mech.* (2015) P05030

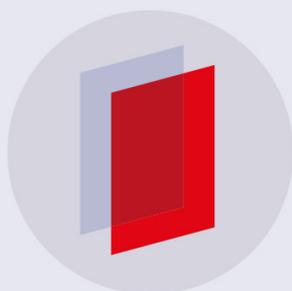
View the [article online](#) for updates and enhancements.

### Related content

- [Finding and evaluating the hierarchical structure in complex networks](#)  
Fei Chen, Zengqiang Chen, Zhongxin Liu et al.
- [Forman curvature for complex networks](#)  
R P Sreejith, Karthikeyan Mohanraj, Jürgen Jost et al.
- [Iterative resource allocation for ranking spreaders in complex networks](#)  
Zhuo-Ming Ren, An Zeng, Duan-Bing Chen et al.

### Recent citations

- [Character networks and book genre classification](#)  
Adriano J. Holanda *et al*



**IOP | ebooks™**

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

# Analyzing complex networks through correlations in centrality measurements

**José Ricardo Furlan Ronqui and Gonzalo Travieso**

Instituto de Física de São Carlos, Universidade de São Paulo, São Paulo—SP,  
05508-070, Brazil

E-mail: [jose.ronqui@usp.br](mailto:jose.ronqui@usp.br) and [gonzalo@ifsc.usp.br](mailto:gonzalo@ifsc.usp.br)

Received 5 March 2015

Accepted for publication 19 April 2015

Published 26 May 2015

Online at [stacks.iop.org/JSTAT/2015/P05030](http://stacks.iop.org/JSTAT/2015/P05030)

[doi:10.1088/1742-5468/2015/05/P05030](https://doi.org/10.1088/1742-5468/2015/05/P05030)



**Abstract.** Many real world systems can be expressed as complex networks of interconnected nodes. It is frequently important to be able to quantify the relative importance of the various nodes in the network, a task accomplished by defining some centrality measures, with different centrality definitions stressing different aspects of the network. It is interesting to know to what extent these different centrality definitions are related for different networks. In this work, we study the correlation between pairs of a set of centrality measures for different real world networks and two network models. We show that the centralities are in general correlated, but with stronger correlations for network models than for real networks. We also show that the strength of the correlation of each pair of centralities varies from network to network. Taking this fact into account, we propose the use of a *centrality correlation profile*, consisting of the values of the correlation coefficients between all pairs of centralities of interest, as a way to characterize networks. Using the yeast protein interaction network as an example we show also that the centrality correlation profile can be used to assess the adequacy of a network model as a representation of a given real network.

**Keywords:** random graphs, networks

**Contents**

<b>1. Introduction</b>	<b>2</b>
<b>2. Basic concepts and datasets</b>	<b>3</b>
2.1. Node centralities	3
Degree centrality	3
Eigenvector centrality	4
Closeness centrality	4
Betweenness centrality	4
Current flow betweenness centrality	5
Current flow closeness	5
Subgraph centrality	5
2.2. Networks and network models	6
<b>3. Results and discussion</b>	<b>6</b>
3.1. Correlations	6
3.2. Correlation profile of networks	9
3.3. Comparison with random rewiring	10
3.4. Evaluating models with the correlation profile	11
<b>4. Conclusion</b>	<b>11</b>
<b>References</b>	<b>13</b>

**1. Introduction**

Important aspects of many systems can be represented by complex networks [1–5]. As the nodes frequently differ with respect to their essentiality, since the beginning of the study of networks the quantification of the importance of their nodes has been receiving attention, as can be seen, e.g. in [6–13]. There are different ways to quantify the importance, or centrality, of a node and therefore a large number of measures used for this purpose, with new centrality measures being constantly proposed for use in new applications or to achieve better results in old ones, see e.g. [14–33].

Although based on different definitions, the various node centralities are, in real networks, correlated: important nodes using one of the definitions are frequently also important using others. For example, nodes with high degree have also high closeness centrality [7]. Some papers already analysed those correlations, while others do correlation analysis when proposing new centralities [7, 29, 30, 32–36]. Nonetheless, there are nodes with high value for one centrality and low value for another and the correlations are not the same for all networks, as will be shown below.

In this work we systematically study the correlations between all pairs of a set of centralities using some real-world networks and two network models. We find that the correlations are generally strong, but there are marked differences between correlations in real network and models and also among different real networks. We therefore suggest the use of such correlations as a way to characterize networks.

This paper is organized as follows. In section 2 we present the considered network centralities and the real networks and models used. In section 3 we present scatter plots of the centrality values for some pairs of centralities (section 3.1) and show that the relations are mostly power law; taking into account this power-law behaviour, we present Pearson correlations for the *logarithms* of the centralities, which will give us a measure of how closely the centralities are related by a power law. Next (section 3.2) we present the concept of *centrality correlation profile* to characterize networks. This is followed by showing (section 3.3) that this centrality correlation profile can be used to distinguish the real networks from their randomly rewired counterparts, as well as from the network models. Finally (section 3.4) we propose the use of this profile to evaluate models for a given real network, using as an example a model for protein–protein interaction networks.

## 2. Basic concepts and datasets

We are considering only undirected, unweighted networks without multiple edges or self connections. In this case, the network can be represented by a symmetric *adjacency matrix*  $\mathbf{A}$  whose  $N \times N$  elements (where  $N$  is the number of nodes)  $A_{ij}$  are 1 if nodes  $i$  and  $j$  are connected and 0 otherwise. In the following, we use  $E$  to represent the number of edges (connections) in the network. Some networks are in fact weighted, as described below, but we disregard the weights. We also drop self and multiple connections when present. When a network has more than one connected component, we consider only the nodes in the largest component.

An important concept is that of *shortest paths*. A path is a sequence of nodes where each two subsequent nodes are directly connected and no node is repeated in the path. A shortest path between nodes  $i$  and  $j$  is a path starting at node  $i$ , ending at node  $j$  and with the smallest possible number of intermediate nodes in the path.

### 2.1. Node centralities

There is a large number of centrality measures in the literature. For brevity, we will work with some of them, including the most used ones, instead of trying to be comprehensive. Although other centralities can be important in many applications, the methodology employed here could be, if needed, easily extended to include other centralities. Furthermore, we use PCA (see section 3.2) to automatically compensate for possible redundancies among the centralities and the results show that using other centralities would not contribute significantly for the considered networks (as enough discrimination is already achieved). It is plausible that other centralities could be necessary for a different dataset.

*Degree centrality.* This centrality quantifies the importance of a node counting its number of connections. Using the adjacency matrix, the degree of node  $i$ , represented as  $k_i$  is

computed as

$$k_i = \sum_{j=1}^N A_{ij}. \quad (1)$$

Here we use a normalized degree centrality, given by dividing the degree by the maximum possible degree:

$$\tilde{k}_i = \frac{k_i}{N-1}. \quad (2)$$

Degree centrality has a low complexity of  $\mathcal{O}(E)$  when using sparse network representations.

*Eigenvector centrality.* Just counting the number of connections, as done in the degree centrality, can give a distorted view of the importance of a node, because it does not quantify the importance of its neighbors. In principle, the importance of the neighbors should be considered when assessing the importance of a node. If  $v_i$  is the importance of node  $i$ , we can compute it in a self-consistent way through

$$v_i = \frac{1}{\lambda} \sum_j A_{ij} v_j. \quad (3)$$

where  $\lambda$  must be chosen appropriately. In vector form we have:

$$\lambda \mathbf{v} = \mathbf{A} \mathbf{v}, \quad (4)$$

which tells us that  $\mathbf{v}$  is an eigenvector of the adjacency matrix and  $\lambda$  the corresponding eigenvalue. In fact, we use the eigenvector associated with the *largest* eigenvalue of the adjacency matrix and the *eigenvector centrality* of node  $i$  is the  $i$ -th entry in this eigenvector. Computation of this centrality takes  $\mathcal{O}(N+E)$  operations for sparse networks.

*Closeness centrality.* It is also possible to take the word ‘centrality’ more literally and search for nodes that are central in the sense of being in average closer to the other nodes. If  $d_{ij}$  is the shortest path distance between nodes  $i$  and  $j$  we can compute the closeness centrality of node  $i$  as

$$c_i = \frac{1}{\sum_j d_{ij}}. \quad (5)$$

This computation is  $\mathcal{O}(NE)$ .

*Betweenness centrality.* Assuming pairs of nodes in the network must interact, if they are not directly connected the interaction must go through intermediary nodes. A node is important in the betweenness centrality sense if it must be used as an intermediary for many pairs of nodes (under the assumption that the interactions always follow a shortest path, that is, a path with minimum number of intermediaries).

The betweenness centrality of node  $i$ , represented as  $b_i$ , is computed by the expression:

$$b_i = \sum_{j,k} \frac{n(j,i,k)}{n(j,k)}, \quad (6)$$

where  $j \neq i \neq k$ ,  $n(j,k)$  is the number of shortest paths from  $j$  to  $k$  and  $n(j,i,k)$  is the number of shortest paths from  $j$  to  $k$  that pass through  $i$ . This centrality can be computed in  $\mathcal{O}(NE)$  using the algorithm of Brandes [37].

*Current flow betweenness centrality.* Betweenness centrality takes into account only the shortest paths from a node  $j$  to another node  $k$ . It is possible for the nodes to interact through other paths. This is taken into account by the centrality measure based on computing the current flow through the network elements supposing that each link is a resistor (with a value of 1 for unweighted networks) considering all possible sources and drains for the current. This is equivalent to counting the number of times a random walk from a node  $j$  to node  $k$  passes through a given node  $i$ , for all pair  $(j, k)$  (but canceling back-and-forth movements of the walker that do not contribute to a net movement toward the target) [18, 38] and is therefore also called random walk betweenness centrality. This centrality can be computed as the average of the current  $I_i^{(s,t)}$  flowing through the node  $i$  over all pairs of sources and drains  $(s, t)$ ,

$$\beta_i = \frac{2}{N(N-1)} \sum_{s < t} I_i^{(s,t)}. \quad (7)$$

The currents are given by

$$I_i^{(s,t)} = \begin{cases} 1, & i = s, t \\ \frac{1}{2} \sum_j A_{ij} |V_i^{(s,t)} - V_j^{(s,t)}|, & i \neq s, t \end{cases} \quad (8)$$

where  $V_i^{(s,t)}$  is the potential on node  $i$  when source is in  $s$  and drain in  $t$  and can be computed using the Laplacian of the graph (see [18]). This computation has complexity  $\mathcal{O}((N+E)N^2)$ .

*Current flow closeness.* Also known as information centrality [38], this measure first proposed in [15], is a generalization of the closeness centrality in the same lines than the current flow (or random walk) betweenness centrality is a generalization of the shortest-path betweenness centrality: by considering alternate paths from a node to other nodes instead of just the shortest path. It is evaluated by the expression

$$\gamma_i = \frac{N}{\sum_j R_{ij}}, \quad (9)$$

where  $R_{ij} = V_i^{(i,j)} - V_j^{(i,j)}$  is the equivalent resistance from  $i$  to  $j$  when the source is in  $i$  and the drain in  $j$ . Computation takes time  $\mathcal{O}(N^3)$ .

*Subgraph centrality.* This measure takes into account the participation of a node in subgraphs, given larger weight for smaller subgraphs [19]. Closed walks starting and ending in a node  $i$  are counted and weighted with the inverse factorial of their size. With the chosen weighting, the values can be efficiently computed using the spectral decomposition of the adjacency matrix. If  $\lambda_j$  are the eigenvalues and  $\mathbf{v}_j$  are the respective eigenvectors, the subgraph centrality  $s_i$  of node  $i$  can be computed in  $\mathcal{O}(N^3)$  using the expression

$$s_i = \sum_{j=1}^N (v_j^i)^2 e^{\lambda_j}, \quad (10)$$

where  $v_j^i$  is the  $i$ -th element of eigenvector  $\mathbf{v}_j$ .

**Table 1.** Topological measurements for the (largest components of the) networks: karate club (KT), dolphins (DP), high-energy physics collaboration (HP), network science collaboration (NS), political books (PB) and power grid (PW).

Measures	KT	DP	HP	NS	PB	PW
$N$	34	62	5835	379	105	4941
$m$	78	159	13815	914	441	6594
$\langle k \rangle$	4.59	5.13	4.74	4.82	8.4	2.67
$\langle k^2 \rangle$	35.65	34.90	43.19	38.69	100.25	10.33
$\Delta_{\text{local}}$	0.57	0.26	0.51	0.74	0.49	0.08
$\Delta_{\text{transitivity}}$	0.26	0.31	0.28	0.43	0.35	0.10
$\ell$	2.34	3.30	7.03	6.03	3.05	18.99
$E$	0.49	0.38	0.16	0.20	0.40	0.06
$r$	-0.48	-0.04	0.19	-0.08	-0.13	0.00

*Note:*  $N$ : number of nodes,  $m$ : number of edges,  $\langle k \rangle$ : average degree,  $\langle k^2 \rangle$ : variance of the degree distribution,  $\Delta_{\text{local}}$ : average local clustering coefficient,  $\Delta_{\text{transitivity}}$ : global clustering coefficient (transitivity),  $\ell$ : average shortest path length,  $E$ : efficiency,  $r$ : assortativity.

## 2.2. Networks and network models

We work here with the following available network datasets: Zachary's karate club (represents friendship between 34 members of a karate club) [39]; dolphin social network (frequent association between 62 dolphins) [40]; high-energy theory collaboration (coauthorship in preprints on the hep-th section in arXiv.org) [41–43]; network science collaborations (coauthorship in network science papers) [44]; books about US politics published around 2004 and sold on Amazon.com (edges show frequent co-purchase) [45]; power grid (topology of the power grid of the Western States of the USA) [46]. Table 1 shows some measurements for the networks.

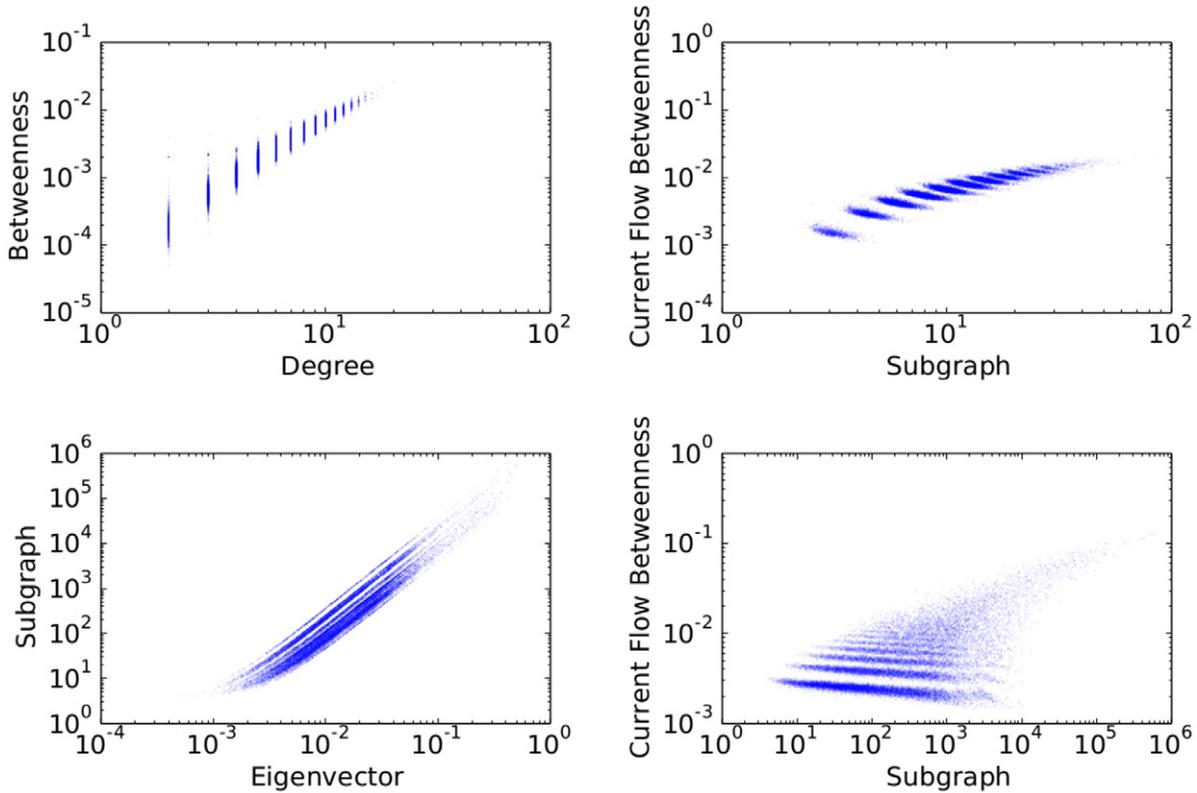
Our emphasis is showing results for real networks. We therefore include only two simple models for comparison, the Erdős–Rényi (ER) random graphs [47] and the Barabási–Albert (BA) scale-free networks [48]. The method used here could be applied for other models (as done in section 3.4), if appropriate.

## 3. Results and discussion

Given a network, we compute the centralities of each of its nodes and search for correlations between pairs of centralities, with each node in the network corresponding to a data point.

### 3.1. Correlations

Figure 1 shows scatterplots for some of the pairs of centralities for the network models, while the plots for the real networks are presented in figure 2 (best correlations) and figure 3 (worst correlations). We do not show closeness or current flow closeness centralities results for best cases as these measurements span a limited range, which limits the



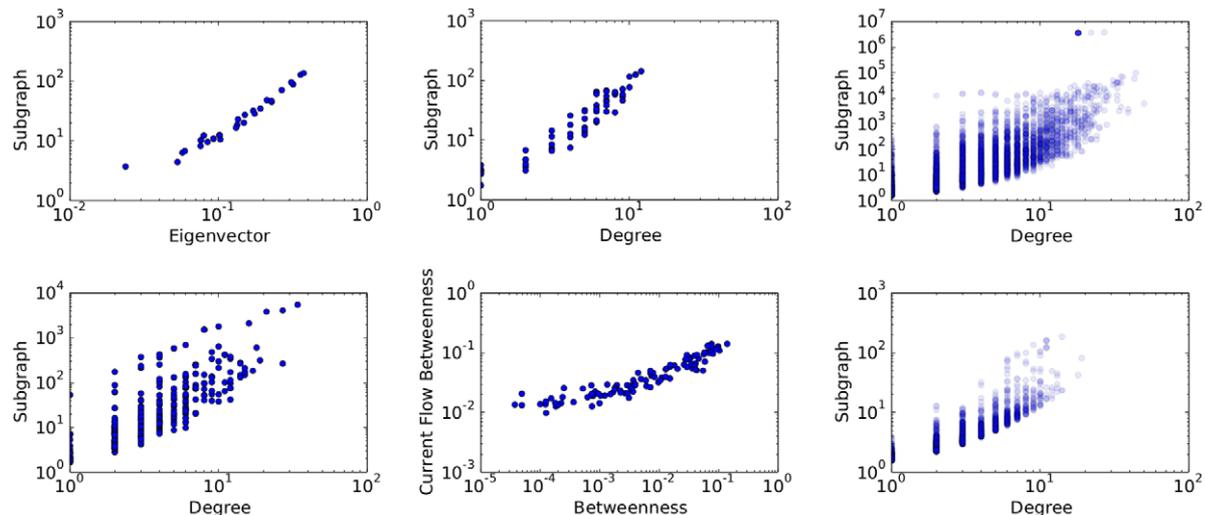
**Figure 1.** Scatter plots for some pairs of centralities for the Erdős–Rényi (top) and Barabási–Albert (bottom) network models. Left: best correlations (excluding closeness related cases, see text); right Worst correlations.

significance of a good correlation in a log log plot. Due to space limitations, we show only two of the pairs with largest and smallest correlation values, respectively, for each network. These plots suggest that the centralities are correlated, with visible correlations even in the weakest cases for some networks and that the correlations are close to a power law, specially for high values of centralities.

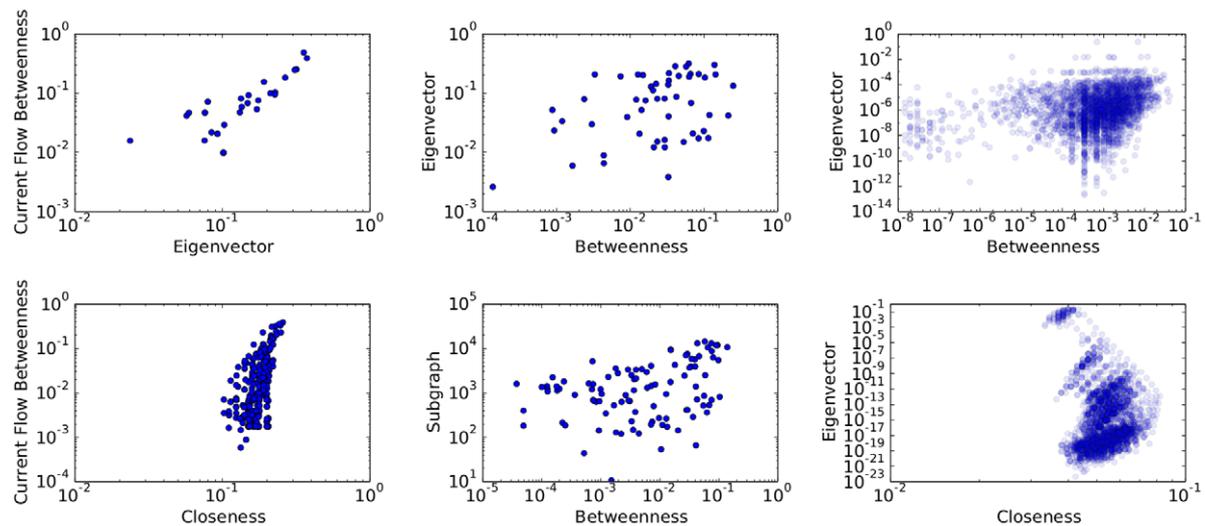
To quantify how closely two measurements are related by a power law, we use log–log plots and compute the Pearson correlation coefficient between the logarithms of the values of the centralities. Similar results were also found when using the Pearson and Spearman correlations of the measurements (without the logarithms). The use of the logarithms here is to emphasize possible power-laws, as observed in figures 1 and 3.

Table 2 shows the values of the Pearson coefficients for all pairs of logarithms of the centralities in the networks studied. For the ER and BA models, 50 networks (for each model) of 1000 vertices and average degree 6 where used. We can see that the centralities have, in general, large values of the Pearson coefficient, which in our case implies proximity to a power law relation. With the exception of a small negative coefficient between closeness and eigenvector centralities for the power grid network, all coefficients are positive. This means that nodes that are important with respect to one definition are, in general, also important according to other definitions. It can be seen that the network models studied present larger coefficients than the real networks (with the exception of

Analyzing complex networks through correlations in centrality measurements



**Figure 2.** Scatter plots for the best correlations between pairs of centralities for the real networks (excluding closeness related cases). Top, left to right: karate, dolphins and high-energy physics; bottom, left to right: network science, political books and power grid.



**Figure 3.** Scatter plots for the worst correlations between pairs of centralities for the real networks. Top, left to right: karate, dolphins and high-energy physics; bottom, left to right: network science, political books and power grid.

the karate club network, which also has large coefficients<sup>1</sup>). Especially noticeable is the coefficient of closeness with eigenvector centrality, which is almost perfect for the network models (0.99 for ER and 0.98 for BA), but non-existent for the power grid network. It is, therefore, important to be careful when generalizing conclusions from results using such simplified models to real networks. The coefficients of degree with random walk closeness and subgraph centrality are large for all networks, with the exception of the power grid

<sup>1</sup> The karate club is a small network, with a few dominating nodes and many peripheral nodes. This could explain the large values.

**Table 2.** Pearson coefficients for the networks: karate club (KT), dolphins (DP), high-energy physics collaboration (HP), network science collaboration (NS), political books (PB), power grid (PW), Erdős–Rényi model (ER) and Barabási–Albert model (BA).

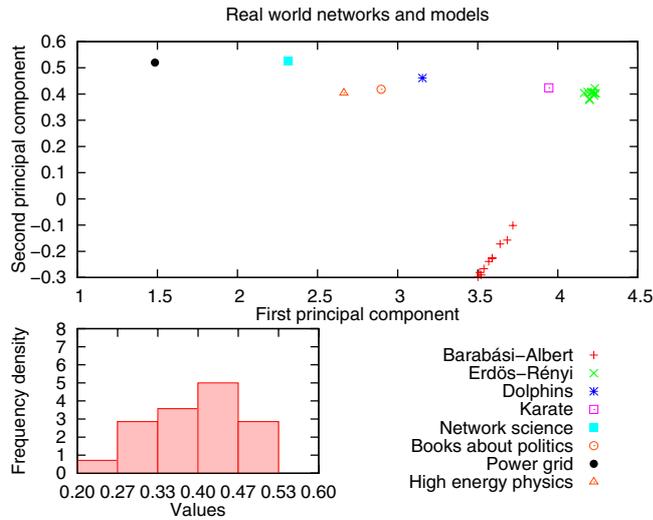
Measurements	KT	DP	HP	NS	PB	PW	ER	BA
Degree/Closeness	0.80	0.75	0.54	0.26	0.61	0.22	0.92	0.67
Degree/Betweenness	0.84	0.73	0.52	0.58	0.65	0.47	<b>0.98</b>	0.92
Degree/Eigenvector	0.82	0.63	0.49	0.21	0.69	0.16	0.91	0.62
Degree/Subgraph	0.91	<b>0.94</b>	0.79	<b>0.79</b>	0.83	<b>0.89</b>	<b>0.97</b>	0.62
Degree/CF betweenness	<b>0.95</b>	0.70	0.59	0.51	0.79	0.46	0.51	<b>0.94</b>
Degree/CF closeness	<b>0.95</b>	<b>0.96</b>	0.77	0.59	<b>0.95</b>	0.35	<b>0.98</b>	<b>0.97</b>
Closeness/Betweenness	<i>0.78</i>	0.71	0.35	0.40	0.75	0.40	0.90	0.77
Closeness/Eigenvector	<b>0.92</b>	0.83	<b>0.91</b>	0.58	0.45	<i>-0.04</i>	<b>0.99</b>	<b>0.98</b>
Closeness/Subgraph	0.89	0.72	0.71	0.34	0.47	0.16	<b>0.96</b>	<b>0.97</b>
Closeness/CF betweenness	0.82	0.55	0.48	<i>0.14</i>	0.84	0.50	<i>0.43</i>	<i>0.54</i>
Closeness/CF closeness	0.89	0.87	0.85	<b>0.76</b>	0.69	0.75	0.93	0.70
Betweenness/Eigenvector	<i>0.78</i>	<i>0.39</i>	<i>0.24</i>	0.34	<i>0.35</i>	<i>0.05</i>	0.89	0.72
Betweenness/Subgraph	0.82	0.60	0.30	0.33	<i>0.31</i>	0.35	<b>0.95</b>	0.70
Betweenness/CF betweenness	0.90	<b>0.93</b>	0.69	0.71	<b>0.93</b>	0.55	<b>0.98</b>	0.91
Betweenness/CF closeness	0.85	0.71	<i>0.29</i>	0.39	0.62	0.34	<b>0.96</b>	<b>0.93</b>
Eigenvector/Subgraph	<b>0.97</b>	0.70	0.71	0.46	0.82	0.32	<b>0.95</b>	<b>0.96</b>
Eigenvector/CF betweenness	<i>0.74</i>	0.61	0.40	<i>0.15</i>	0.45	0.10	<i>0.44</i>	<i>0.49</i>
Eigenvector/CF closeness	<b>0.92</b>	0.76	0.78	0.65	0.72	<i>0.06</i>	0.92	0.66
Subgraph/CF betweenness	0.82	0.55	0.41	0.34	0.46	0.37	<i>0.40</i>	<i>0.49</i>
Subgraph/CF closeness	<b>0.96</b>	0.90	0.77	0.71	0.87	0.29	<b>0.94</b>	0.65
CF betweenness/CF Closeness	0.91	0.76	0.57	0.39	0.77	0.70	0.58	<b>0.93</b>

*Note:* The largest (absolute) value of correlation for each network and all values within a 0.05 inclusive range are marked in bold; idem for the smallest (absolute) values, marked in italics. (CF is an abbreviation for ‘current flow’).

network. To a lesser extent, the same is true for other pairs involving degree, betweenness and current flow betweenness. Other pairs have large coefficients in some networks, but small coefficients in others. For instance, betweenness and subgraph centralities have large coefficients for the network models, the dolphins and karate networks, but small values for the other networks. Most interesting is the case of betweenness and eigenvector centralities: they have small coefficients for all real networks (with the exception of the karate club network, where it is large, but smaller than for other pairs), but large coefficients for the network models. This suggests that they complement each other when analysing real world networks and reinforces our previous observation of inadequacy of generalizing conclusions based on simple models.

### 3.2. Correlation profile of networks

These results suggest that each network or network model has a specific profile of correlations between centrality measurements. We call this the *centrality correlation profile* of the network. To show that this profile can be used to characterize the networks, figure 4 plots a two-dimensional projection of the real world networks from the space defined by the



**Figure 4.** PCA projection of the centrality correlation profile space for the networks used in this work. A total of 26 networks are included (6 real networks and 10 for each model, BA and ER), each one characterized by 21 correlation values. Also shown is a histogram of the values of the contributions of each correlation to the two principal components.

centrality correlation profile using principal component analysis (PCA) [49]. Each network is a point in a 21-dimensional space defined by the values of the Pearson correlation between the logarithms of the seven considered measurements. The points are projected to the two principal components for visualization. Note how the networks generated by the same model are clustered in small regions, while the different real networks or models are spread through the graph. The only exception is the small karate club network, which is close to the ER model cluster. To evaluate if there are some correlations that are dominant or negligible for the discrimination of the networks in the projection, we follow the method employed in [50] and compute the contribution of each correlation to the PCA by summing up the magnitudes of their respective dimensions in the principal components (see further details in the reference). A value of contribution close to zero would imply that the corresponding correlation could have been discarded. A histogram of these values is included in the figure. It is clear that all contributions are of the same order and no correlation could have been ignored without reduction of the discriminating power.

### 3.3. Comparison with random rewiring

In our next experiment we generate, for each real network, 100 random networks with the same degree sequence through link rewiring [51]. In this method, a pair of edges is randomly chosen, the original edges are removed and substituted by two new edges among the same vertices; the process is repeated a certain number of times. Each rewired network is generated by a number of random rewirings equal to 100 times the number of edges in the original network. We also include, for comparison, 100 networks each for the ER and BA models with the same number of nodes and similar average degrees. We compute the centrality correlation profiles of all networks and generate a two-dimensional PCA projection. The results are shown in figure 5. With the exception of the karate

club network, where the real network is inside the region of the rewired networks, we can distinguish the networks from their random rewirings and the two other models. This stresses the fact that, although there is generally a strong correlation between the various centralities, there is also important information in the specific wiring pattern of real networks, resulting in distinct correlation profiles. It is also interesting to note that the randomly rewired networks are sometimes closer to the ER, sometimes to the BA networks, but always closer to the models than the corresponding real network, supporting the assertion that the centrality correlation profile is characteristic of the specific network. In all the shown projections, the contributions of all correlations are important, as demonstrated by the attached histograms.

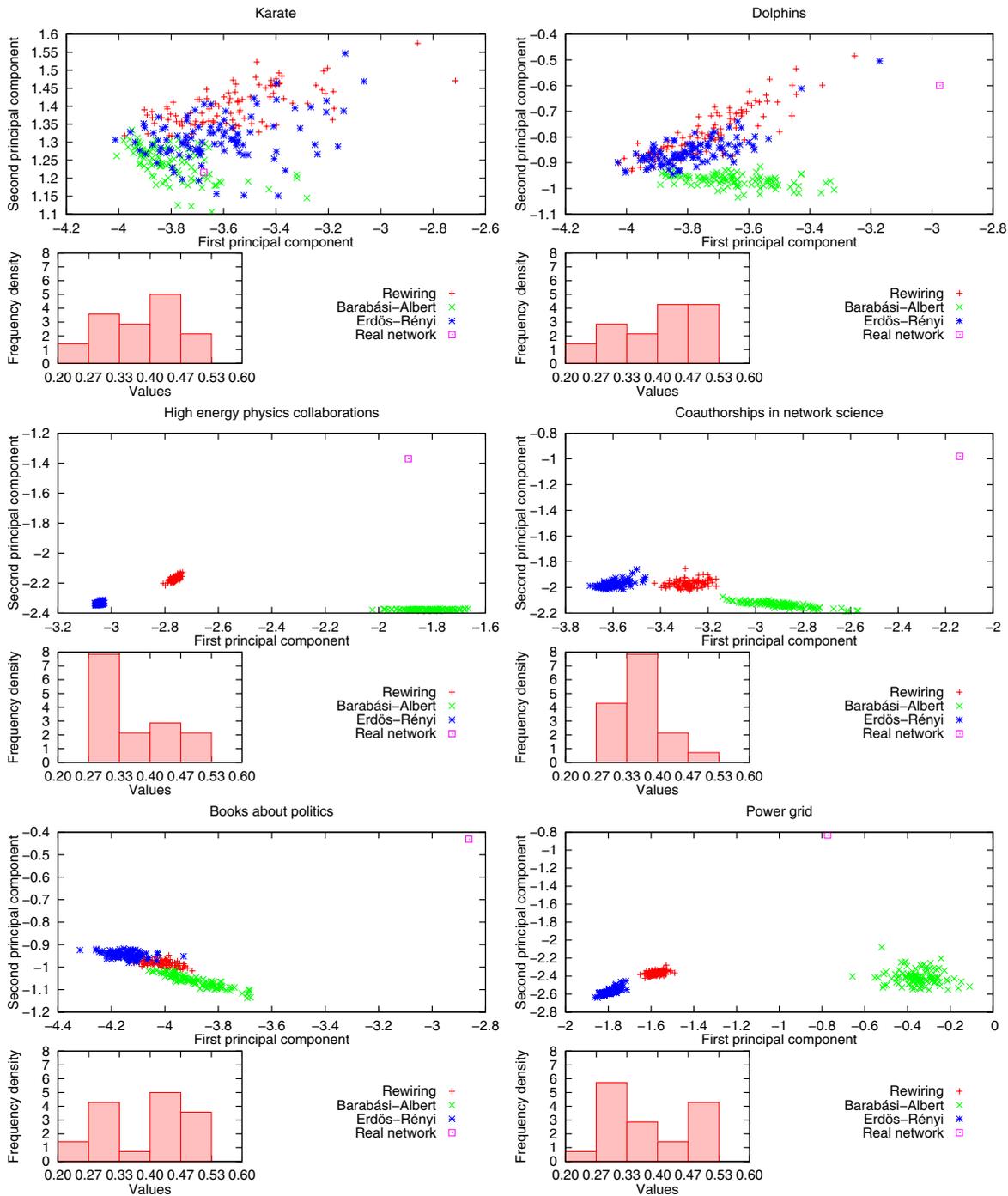
### 3.4. Evaluating models with the correlation profile

Considering the previously presented results, we suggest that the centrality correlation profile can be used as a tool to test the adequacy of a network model developed to study a given real network. If the real network can be considered typical, with respect to the correlation profile, in comparison to networks generated using the proposed model, the model is appropriate. In an ideal case, we would know the distribution of points representing the generated networks in the correlation profile space and use standard statistical methods to evaluate the probability of the real network being generated by the model. In practice, when the correlation profile of the model is not known, we can use PCA projections of the real network and a large number of generated networks to achieve an informal confirmation of the model. To demonstrate this procedure, we use the yeast protein interaction network from [52] and compare it with Barabási–Albert networks and the model developed by Pastor-Satorras *et al* [53] specifically for protein interaction networks. Figure 6 shows a PCA projection of the centrality correlation profile of the network and 30 random networks generated by each model. The real network is much closer to the networks generated by the Pastor-Satorras *et al* model than to the ones generated the Barabási–Albert model. But the yeast network cannot be considered a typical network from the Pastor-Satorras *et al* model, as it lies outside of the region of correlation profile space spanned by the random networks generated according to the model, demonstrating that there are still important structural details in the real network not accounted for by the model.

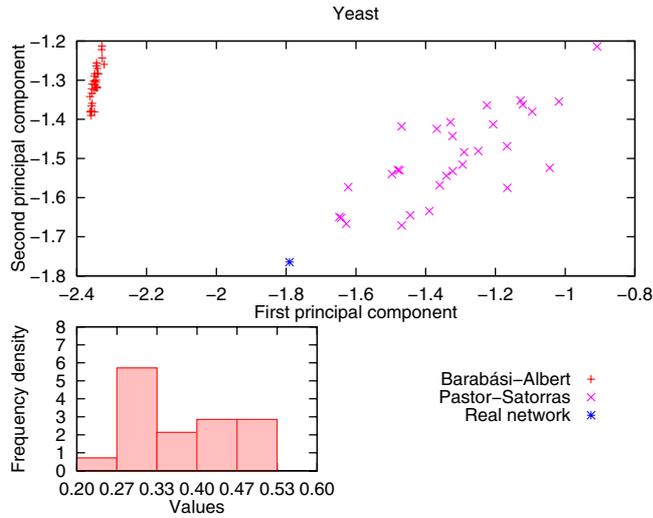
## 4. Conclusion

Various centrality measurements are commonly used to discriminate important nodes in complex networks. The different measurements correspond to different definitions of the importance of the nodes, but our results have shown that they are in general strongly correlated for real networks and even more for the two network models studied. We considered the following measurements: degree, closeness, betweenness, eigenvector, subgraph, current flow closeness and current flow betweenness centralities. For most pairs of centralities, their Pearson correlation coefficients are above 0.5 for most networks, with some pair showing coefficients above 0.95 for some networks, specially the network models. The log–log scatter plots show that the correlations are specially strong for high centrality

Analyzing complex networks through correlations in centrality measurements



**Figure 5.** PCA projections comparing the real networks with rewired random version with the same degree sequence, Barabási–Albert and Erdős–Rényi with the same number of nodes and similar average degree. Each graph include a real network, 100 rewired networks and 100 of each model, BA and ER), with each network characterized by 21 correlation values. From top left to bottom right: karate, dolphins, high-energy physics, network science, political books and power grid. Also shown are histograms of the contributions of each correlation to the projections.



**Figure 6.** PCA projections of the positions in the centrality correlation profile space of the yeast network and 30 networks generated by the Barabási-Albert model and 30 networks generated by the Pastor-Satorras *et al* model, each network characterized by 21 correlation values. The histogram of contributions show that all correlations are relevant.

nodes, where they follow a power law. But the correlation values vary strongly from one network to another. For example, while the Pearson correlation coefficient between closeness and eigenvector centrality is 0.99 for the ER model and 0.92 for the karate club network, it is almost zero for the power grid network. We proposed therefore the use of the *centrality correlation profile*, consisting of the values of the correlation coefficient for all pairs of centralities studied, to characterize a network. Our results show that the networks can be distinguished using this profile and all studied correlations have a significant contribution to the discriminating power. We have also shown, using the example of the yeast protein interaction network, how the centrality correlation profile can be used to verify to what extent a model (in our example the Pastor-Satorras *et al* model) is adequate to explain a given network.

Interesting open questions suggested by this work include: why are the correlation coefficients for the network models so strong for almost all pairs of centralities? Are the power laws seen for high centrality values due to specific topological features of the considered networks or do they result from the definitions of the measurements? Why are correlations in real networks consistently smaller than in the models? Do the results hold for other models and real networks? Would the inclusion of other centrality measurements increase the discriminating power of the method? What kind of topological features makes some correlations smaller and other larger for a given network? An answer to the last question would help us design more adequate models for some network and therefore understand them better.

## References

- [1] Réka A and Barab A-L 2002 Statistical mechanics of complex networks *Rev. Mod. Phys.* **74** 47–97
- [2] Dorogovtsev S N and Mendes J F F 2002 Evolution of networks *Adv. Phys.* **51** 1079–87

- [3] Newman M 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–256
- [4] Newman M 2010 *Networks: An Introduction* (Oxford: Oxford University Press)
- [5] da F Costa L, Oliveira O N, Travieso G, Rodrigues F A, Villas P R, Antiqueira L, Viana M P and Correa L E 2011 Analyzing and modeling real-world phenomena with complex networks: a survey of applications *Adv. Phys.* **60** 329–412
- [6] Freeman L C 1978 Centrality in social networks conceptual clarification *Soc. Netw.* **1** 215–39
- [7] Bolland J M 1988 Sorting out centrality: an analysis of the performance of four centrality models in real and simulated networks *Soc. Netw.* **10** 233–53
- [8] Friedkin N E 1991 Theoretical foundations for centrality measures *Am. J. Sociol.* **96** 1478–504
- [9] Faust K 1997 Centrality in affiliation networks *Soc. Netw.* **19** 157–91
- [10] Jeong H, Mason S P, Barabasi A L and Oltvai Z N 2001 Lethality and centrality in protein networks *Nature* **411** 41–2
- [11] Borgatti S P and Everett M G 2006 A graph-theoretic perspective on centrality *Soc. Netw.* **28** 466–84
- [12] Manimaran P, Hegde S R and Mande S C 2009 Prediction of conditional gene essentiality through graph theoretical analysis of genome-wide functional linkages *Mol. BioSyst.* **5** 1936–42
- [13] Wang J, Mo H, Wang F and Jin F 2011 Exploring the network structure and nodal centrality of China’s air transport network: a complex network approach *J. Transp. Geogr.* **19** 712–21
- [14] Bonacich P 1987 Power and centrality: a family of measures *Am. J. Sociol.* **92** 1170–82
- [15] Stephenson K and Zelen M 1989 Rethinking centrality: methods and examples *Soc. Netw.* **11** 1–37
- [16] Rothenberg R B, Potterat J J, Woodhouse D E, Darrow W W, Muth S Q and Klovdahl A S 1995 Choosing a centrality measure: epidemiologic correlates in the Colorado Springs study of social networks *Soc. Netw.* **17** 273–97
- [17] Valente T W and Foreman R K 1998 Integration and radiality: measuring the extent of an individual’s connectedness and reachability in a network *Soc. Netw.* **20** 89–105
- [18] Newman M E J 2005 A measure of betweenness centrality based on random walks *Soc. Netw.* **27** 39–54
- [19] Estrada E and Rodríguez-Velázquez J A 2005 Subgraph centrality in complex networks *Phys. Rev. E* **71** 056103
- [20] Borgatti S P 2005 Centrality and network flow *Soc. Netw.* **27** 55–71
- [21] da F Costa L, Rodrigues F A, Travieso G and Villas P R 2007 Characterization of complex networks: a survey of measurements *Adv. Phys.* **56** 167–242
- [22] Estrada E and Hatano N 2008 Communicability in complex networks *Phys. Rev. E* **77** 036111
- [23] Guimerà R, Mossa S, Turtschi A and Amaral L A N 2005 The worldwide air transportation network: anomalous centrality, community structure and cities’ global roles *Proc. Natl Acad. Sci.* **102** 7794–9
- [24] Koschützki D, Lehmann K A, Peeters L, Richter S, Tenfelde-Podehl D and Zlotowski O 2005 Centrality Indices *Network Analysis (Lect. Notes Comput. Sci. vol 3418)* ed U Brandes and T Erlebach (Berlin: Springer) pp 16–61
- [25] Koschützki D and Schreiber F 2008 Centrality analysis methods for biological networks and their application to gene regulatory networks *Gene Regul. Syst. Biol.* **2** 193–201 PMID: [PMC2733090](https://pubmed.ncbi.nlm.nih.gov/1733090/)
- [26] Kitsak M, Gallos L K, Havlin S, Liljeros F, Muchnik L, Stanley H E and Makse H A 2010 Identification of influential spreaders in complex networks *Nat. Phys.* **6** 888–93
- [27] Pei S and Makse H A 2013 Spreading dynamics in complex networks *J. Stat. Mech.* **P12002**
- [28] Pei S, Muchnik L, Andrade J S Jr, Zheng Z and Makse H A 2014 Searching for superspreaders of information in real-world social media *Sci. Rep.* **4** 5547
- [29] Pauls S D and Remondini D 2012 Measures of centrality based on the spectrum of the Laplacian *Phys. Rev. E* **85** 066127
- [30] Avrachenkov K, Litvak N, Medyanikov V and Sokol M 2013 Alpha current flow betweenness centrality *Algorithms and Models for the Web Graph (Lect. Notes Comput. Sci. vol 8305)* ed A Bonato, M Mitzenmacher and P Prakash (Berlin: Springer) pp 106–117
- [31] Molinero X, Riquelme F and Serna M 2014 Power indices of influence games and new centrality measures for social networks *Ambient Intelligence—Software and Applications (Advances in Intelligent Systems and Computing vol 291)* ed C Ramos, P Novais and C E Nihan and J M Corchado Rodríguez (Berlin: Springer) pp 23–30
- [32] Campiteli M G, Holanda A J, Soares L D H, Soles P R C and Kinouchi O 2013 Lobby index as a network centrality measure *Physica A* **392** 5511–5
- [33] Benzi M and Klymko Total C 2013 Total communicability as a centrality measure *J. Complex Netw.* **1** 124–49
- [34] Valente T W, Coronges K, Lakon C and Costenbader E 2008 How correlated are network centrality measures? *Connections* **28** 16–26 PMID: [PMC2875682](https://pubmed.ncbi.nlm.nih.gov/1733090/)
- [35] Ochab J K 2012 Maximal-entropy random walk unifies centrality measures *Phys. Rev. E* **86** 066109

- [36] Iyer S, Killingback T, Sundaram B and Wang Z 2013 Attack robustness and centrality of complex networks *PLoS One* **8** E59613
- [37] Brandes U 2001 A faster algorithm for betweenness centrality *J. Math. Sociol.* **25** 163–77
- [38] Brandes U and Fleischer D 2005 Centrality measures based on current flow *STACS (Lect. Notes Comput. Sci.* vol 3404) ed U Brandes and T Erlebach (Berlin: Springer) pp 533–44
- [39] Zachary W W 1977 An information flow model for conflict and fission in small groups *J. Anthropol. Res.* **33** 452–73
- [40] Lusseau D, Schneider K, Boisseau O J, Haase P, Slooten E and Dawson S M 2003 The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations *Behav. Ecol. Sociobiol.* **54** 396–405
- [41] Newman M E J 2001 The structure of scientific collaboration networks *Proc. Natl Acad. Sci.* **98** 404–9
- [42] Newman M E J 2001 Scientific collaboration networks. I. Network construction and fundamental results *Phys. Rev. E* **64** 016131
- [43] Newman M E J 2001 Scientific collaboration networks. II. Shortest paths, weighted networks and centrality *Phys. Rev. E* **64** 016132
- [44] Newman M E J 2006 Finding community structure in networks using the eigenvectors of matrices *Phys. Rev. E* **74** 036104
- [45] Krebs V 2014 [www.orgnet.com](http://www.orgnet.com)
- [46] Watts D J and Strogatz S H 1998 Collective dynamics of ‘small-world’ networks *Nature* **393** 440–2
- [47] Erdős P and Rényi A 1959 On random graphs *Publicationes Math.* **6** 290–7
- [48] Barab A-L and Albert R 1997 Emergence of scaling in random networks *Science* **286** 509–12
- [49] Abdi H and L J Williams 2010 *Principal Component Analysis Wiley Interdisciplinary Rev.: Comput. Stat.* **2** 433–59
- [50] Song F, Guo Z and Mei D 2010 Feature selection using principal component analysis *IEEE Computer Society Int. Conf. on System Science, Engineering Design and Manufacturing Informatization* pp 27–30
- [51] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 Network motifs: simple building blocks of complex networks *Science* **298** 824–7
- [52] Bu D *et al* 2003 Topological structure analysis of the protein–protein interaction network in budding yeast *Nucl. Acids Res.* **31** 2443–50
- [53] Pastor-Satorras R, Smith E and Solé R V 2003 Evolving protein interaction networks through gene duplication *J. Theor. Biol.* **222** 199–210