Rich-club and page-club coefficients for directed graphs

Daniel Smilkov
Macedonian Academy for Sciences and Arts, Skopje, Macedonia

Ljupco Kocarev
Macedonian Academy for Sciences and Arts, Skopje, Macedonia
BioCircuits Institute, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093-0402

(Dated: November 10, 2009)

Rich-club and page-club coefficients and their null models are introduced for directed graphs. Null models allow for a quantitative discussion of the rich-club and page-club phenomena. These coefficients are computed for four directed real-world networks: Arxiv High Energy Physics paper citation network, Web network (released from Google), Citation network among US Patents, and Email network from a EU research institution. The results show a high correlation between rich-club and page-club ordering. For journal paper citation network, we identify both rich-club and page-club ordering, showing that "elite" papers are cited by other "elite" papers. Google web network shows partial rich-club and page-club ordering up to some point and then a narrow declining of the corresponding normalized coefficients, indicating the lack of rich-club ordering and the lack of page-club ordering, i.e. high in-degree (PageRank) pages purposely avoid sharing links with other high in-degree (PageRank) pages. For UC patents citation network, we identify page-club and rich-club ordering providing a conclusion that "elite" patents are cited by other "elite" patents. Finally, for e-mail communication network we show lack of both rich-club and page-club ordering. We construct an example of synthetic network showing page-club ordering and the lack of rich-club ordering.

PACS numbers: 05.40.Fb, 02.50.Ga, 02.50.Cw

I. INTRODUCTION

The study of complex systems pervades through almost all the sciences, from cell biology to ecology, from computer science to meteorology, to name just a few. A paradigm of a complex system is a network, described usually as a graph, where complexity may come from different sources: topological structure, network evolution, connection and node diversity, and/or dynamical evolution. Perhaps the most widely known graph property is the *node degree distribution* $P(k)$, which specifies the probability of nodes having degree k in a graph. The unexpected findings that degree distributions of some real-world network topologies closely follow power laws stimulated further interest in network research [1].

However, node degree distribution does not describe the interconnectivity of nodes with given degrees, that is, it does not provide any information on the total number $m(k_1, k_2)$ of links between nodes of degree k_1 and k_2. *Joint degree distribution* is defined as $P(k_1, k_2) = m(k_1, k_2)\mu(k_1, k_2)/(2m)$, where $\mu(k_1, k_2)$ is 2 if $k_1 = k_2$ and 1 otherwise, and m is the number of links in the graph. Clearly joint degree distribution contains more information about connectivity in a graph than degree distribution: it provides information about 1-hop neighborhoods around a node. Given $P(k_1, k_2)$, we can calculate $P(k) = (\bar{k}/k)\sum_{k'} P(k, k')$, but not vice versa, where $\bar{k} = \sum_k kP(k)$.

Although looking into the high order distributions is a complex task, reminding us there is a price to pay, a well chosen set of metrics can give us a simple partial view of high-order distributions. Several such graph metrics that exploit joint degree distribution are:

- *Assortativity coefficient*:

$$r \sim \sum_{k_1,k_2}^{k_{max}} k_1 k_2 \left[P(k_1, k_2) - \frac{k_1 k_2 P(k_1) P(k_2)}{\bar{k}^2} \right]$$

- *Average neighbor connectivity*:

$$k_{nn}(k) = \sum_{k'}^{k_{max}} k' P(k'|k)$$

- *Local clustering*:

$$C(k) = 2m_{nn}(k)/[k(k-1)],$$

  where $m_{nn}(k)$ is the number of links between the neighbors of k-degree nodes

- *Rich-club coefficient*:

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

  where $E_{>k}$ is the number of edges among the $N_{>k}$ nodes having degree higher than a given value k.

In this paper we define several new metrics for directed graphs. These metrics can give us a deeper level of understanding the complex networks as they represent different projections of the joint degree distribution. This is the

outline of the paper. Section II overviews two graph metrics, namely rich-club coefficient and average degree of nearest neighbors. In section III several new metrics for directed graphs are introduced. In section IV these new metrics are computed for 4 networks: (A) journal papers citation network, (B) web graph (released from Google), (C) UC patents citation network, and (D) e-mail communication network. Our conclusions are presented in section V.

II. PRELIMINARIES

In this section we overview two graph metrics: rich-club coefficient and average degree of nearest neighbors. Graphs considered here are undirected and unweighted simple graphs.

A. Rich-club coefficient

The rich-club coefficient, introduced by Zhou and Mondragon in the context of the Internet [2], refers to the tendency of high degree nodes, the hubs of the network, to be very well connected to each other. Denoting by $E_{>k}$ the number of edges among the $N_{>k}$ nodes having degree higher than a given value k, the rich-club coefficient is expressed as:

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

After some basic analytical analysis of the rich-club coefficient [3], we can see that it can be expressed as a function of the joint degree distribution

$$\phi(k) = \frac{N\langle k\rangle \sum_{k'=k+1}^{k_{max}} \sum_{k''=k+1}^{k_{max}} P(k',k'')}{N_{>k}(N_{>k}-1)} \tag{2}$$

giving us a partial view of the high-order degree distribution which is far more economical to compute.

In [2], the rich-club coefficient ϕ is defined in terms of nodes with rank less than r_{max} where nodes are sorted by decreasing degree values and the node rank r denotes the position of a node on this ordered list normalized by the total number of nodes. Several networks are compared and a threshold value of 1%, i.e. the value of $\phi(1\%)$ was used to differentiate the networks and provide evidence of the rich-club phenomenon. However, a monotonic increase of $\phi(k)$ does not necessarily imply the presence of the rich-club phenomenon. Indeed, even in the case of the ER graph – a completely random network – has an increasing rich-club coefficient. This implies that the increase of $\phi(k)$ is a natural consequence of the fact that vertices with large degree have a larger probability of sharing edges than low degree vertices. This feature is therefore imposed by construction and does not represent a signature of any particular organizing principle or structure, as is clear in the ER case [3]. The simple inspection of the $\phi(k)$ trend is therefore potentially misleading in the discrimination of the rich-club phenomenon, it can only be used as a simple statistical property to differentiate several networks in their complex structure.

Therefore, in order to detect rich-club phenomenon several null models were proposed that normalize the basic rich-club coefficient. A null model was presented in [3] where the rich-club is normalized by the expression $\rho(k) = \phi(k)/\phi_{ran}(k)$ where

$$\phi_{ran}(k) = \frac{1}{N\langle k\rangle}\left[\frac{\sum_{k'=k+1}^{k_{max}} k'P(k')}{\sum_{k'=k+1}^{k_{max}} P(k')}\right]^2 \underset{k,\,k_{max}\to\infty}{\sim} \frac{k^2}{\langle k\rangle N} \tag{3}$$

is the rich-club coefficient of the maximally random network (uncorrelated network) with the same degree distribution $P(k)$ as the network under study. Operatively, the maximally random network can be thought of as the stationary ensemble of networks visited by a process that, at any time step, randomly selects a couple of links of the original network and exchange two of their ending points (automatically preserving the degree distribution). An actual rich-club ordering is denoted by a ratio $\rho(k) > 1$. Note that in sufficiently large networks and large k, $\phi_{ran}(k)$ becomes clearly dependent of k.

B. Average degree of nearest neighbors

Another graph metric exploiting the joint degree probability distribution is the average degree of the nearest neighbors, $k_{nn}(k)$, for vertices of degree k

$$k_{nn}(k) = \frac{1}{N_k}\sum_i k_{nn,i}, \tag{4}$$

where the sum runs over all nodes of degree k and where N_k is the number of nodes of degree k and $k_{nn,i}$ denotes the average nearest neighbors degree of vertex i, i.e.

$$k_{nn,i} = \frac{1}{k_i}\sum_{j\in V(i)} k_j,$$

where the sum is over the nearest neighbors vertices of i. This quantity is related to the correlations between the degrees of connected vertices since on average it can be expressed as

$$k_{nn}(k) = \sum_{k'} k'P(k'|k). \tag{5}$$

III. NOVEL METRICS FOR DIRECTED GRAPHS

In this section we consider directed graphs. A directed graph (or digraph) is a pair $G = (V, E)$ of a set V, whose

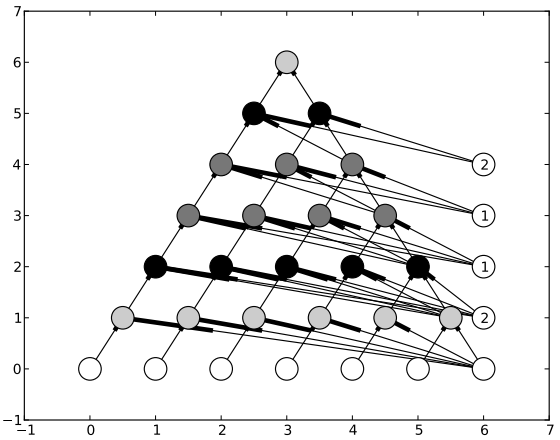


Figure 1: $d = 6$. Nodes are colored gradually according to their in-degree with white to black color denoting lowest to highest in-degree respectively. The additional nodes, i.e. nodes of the set S_i are aggregated in one node with label $|S_i|$ for simplicity.

elements are called vertices or nodes, and a set E of ordered pairs of vertices, called arcs or directed edges. We suggest several new graph metrics.

A. In-degree rich-club coefficient

Having directed networks in mind, rich-club coefficient can be defined in two ways, in terms of in-degree and out-degree. The one that we are interested in is the in-degree rich-club coefficient defined, in a very similar way to (1), as

$$\phi^{in}(k) = \frac{E^{in}_{>k}}{N^{in}_{>k}(N^{in}_{>k} - 1)}, \tag{6}$$

where $E^{in}_{>k}$ is the number of directed edges among the $N^{in}_{>k}$ nodes having in-degree higher than a given value k. Note the number 2 missing in the numerator since in directed full-mesh graph the number of edges is twice than that in the undirected graph.

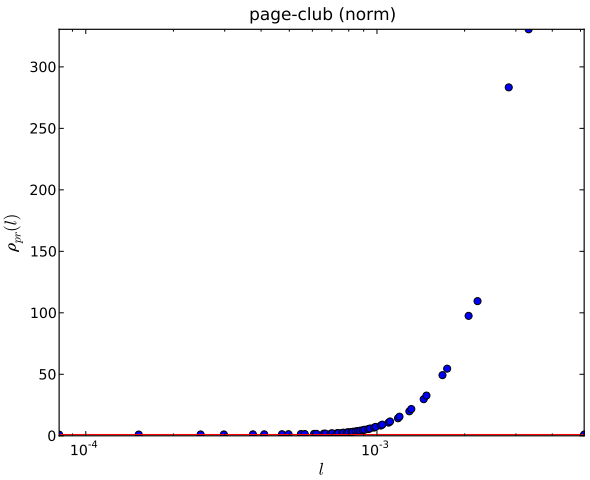
We can express the numerator in (6) as

$$E^{in}_{>k} = \sum_{k'=k+1}^{k^{max}_{in}} \sum_{k''=k+1}^{k^{max}_{in}} E^{in}_{k'\to k''}, \tag{7}$$

where k^{max}_{in} is the maximum node in-degree of the network and $E^{in}_{k'\to k''}$ denotes the number of edges pointing from a node of in-degree k' to a node of in-degree k''. Only in the case of random uncorrelated networks, $E^{in}_{k'\to k''}$ takes the simple form

$$E^{in}_{k'\to k''} = \frac{N P_{in}(k'') k'' \left\langle k^{k_{in}=k'}_{out} \right\rangle P_{in}(k')}{\langle k_{in} \rangle}, \tag{8}$$

(a)



(b)

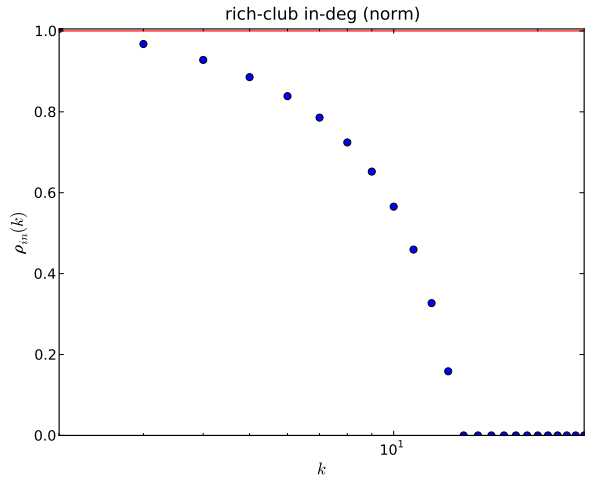


Figure 2: Synthetic network: (a) normalized page-club coefficient $\phi^{PR}(l)/\phi^{PR}_{ran}(l)$ versus page rank (b) normalized rich-club coefficient $\phi^{in}(k)/\phi^{in}_{ran}(k)$ versus in-degree. The network shows the lack of rich-club ordering, but strong page-club ordering.

where $\left\langle k^{k_{in}=k'}_{out} \right\rangle$ denotes the out-degree averaged over all nodes of in-degree k' and $P_{in}(k)$ denotes the probability of a node having in-degree k. At first sight $\left\langle k^{k_{in}=k'}_{out} \right\rangle$ may seem constant in the case of large networks representing web graphs, but having in mind the power-law distribution of in-degree, the number of nodes belonging to the same in-degree class for high in-degree becomes considerably small and is insufficient for converging $\left\langle k^{k_{in}=k'}_{out} \right\rangle$ to the general $\langle k_{out} \rangle$. By inserting (8) and (7) into (6) we obtain the null model $\phi^{in}_{ran}(k)$ for uncorrelated directed networks as

$$\phi^{in}_{ran}(k) = \frac{N \sum_{k+1}^{k^{max}_{in}} k'' P_{in}(k'') \sum_{k+1}^{k^{max}_{in}} \left\langle k^{k_{in}=k'}_{out} \right\rangle P_{in}(k')}{\langle k_{in} \rangle N^{in}_{>k}(N^{in}_{>k} - 1)}$$

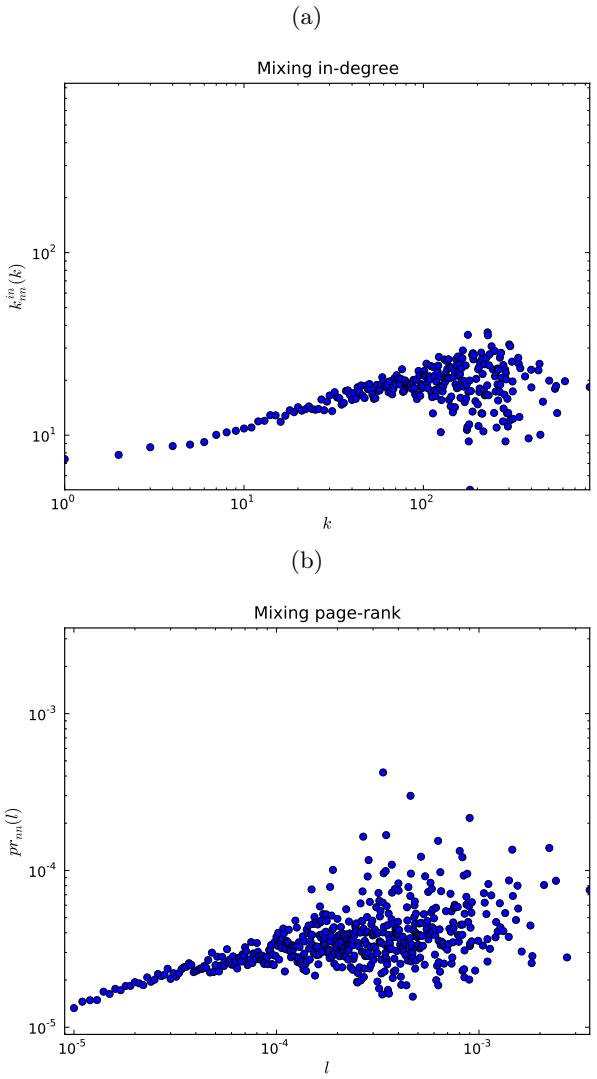


Figure 3: Journal papers citation network: (a) average in-degree of the nearest neighbors, and (b) average nearest neighbors PageRank.

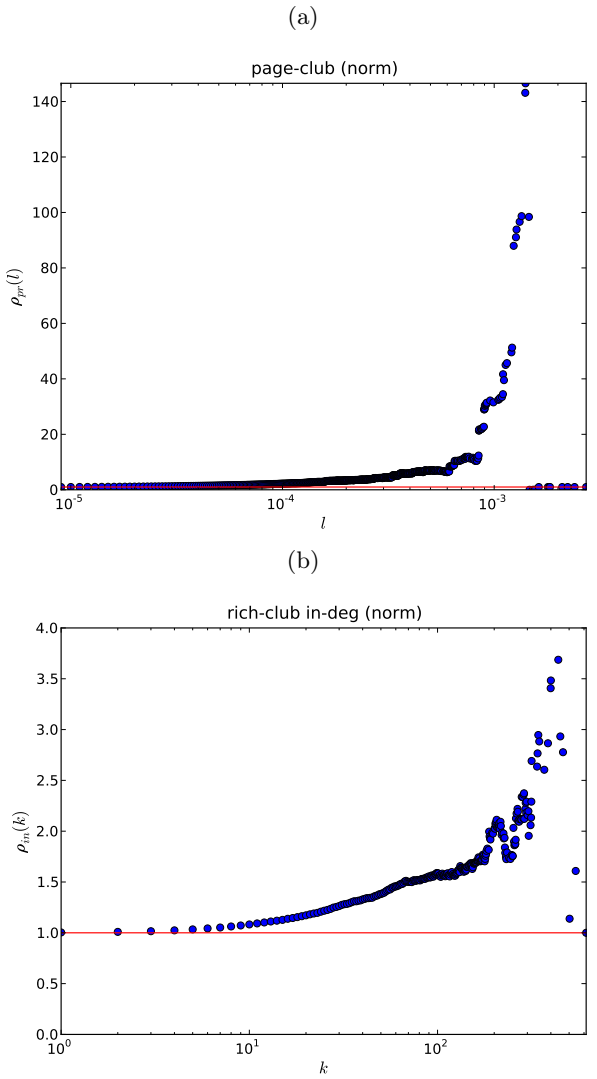


Figure 4: Journal papers citation network: (a) normalized page-club coefficient $\phi^{PR}(l)/\phi^{PR}_{ran}(l)$ versus page rank (b) normalized rich-club coefficient $\phi^{in}(k)/\phi^{in}_{ran}(k)$ versus in-degree. The network shows both page-club and rich-club ordering.

B. Page-club coefficient

Analogously to rich-club coefficient we define a new graph metric called page-club coefficient which refers to the tendency of high PageRank nodes, i.e. most popular pages in the web, to be highly interconnected. Consider a modified random walker whose behavior is defined by the following two rules: (a) with probability $1-q$, the walker follows any outgoing link of i, chosen with equal probability, and (b) with probability q it moves to a generic node of the network (including i), chosen with equal probability. Therefore,

$$pr(i) = \frac{q}{N} + (1-q) \sum_{j \in V; j \to i} \frac{pr(j)}{k_{out}(j)}$$

We assume that each node has at least one outgoing link, and therefore the last equation is well defined. Thus, for each i, stationary probability $pr(i)$, called also PageRank, is well defined and $pr(i) > 0$. The probability q is referred as damping factor; the damping factor adopted in real applications is generally small ($q \approx 0.15$).

Denoting by $E_{PR>l}$ the number of directed edges among the $N_{PR>l}$ nodes having PageRank value higher than a given value l, the page-club coefficient is expressed as:

$$\phi^{PR}(l) = \frac{E_{PR>l}}{N_{PR>l}(N_{PR>l}-1)} \qquad (9)$$

Please note that in the uncorrelated networks this metric converges to classical in-degree rich-club, since in uncorrelated networks the average PageRank of nodes of the same in-degree class becomes linearly dependent of the in-degree [4], i.e.

$$\overline{pr}(k_{in}) = \frac{q}{N} + \frac{1-q}{N}\frac{k_{in}}{\langle k_{in}\rangle}$$

where $\overline{pr}(k_{in})$ is the average PageRank of nodes with in-degree k_{in}. Also, it is important to mention that relative fluctuations of PageRank within the same class decrease as the in-degree increases. Analogous to (9), an appropriate null model for page-club can be defined

$$\phi^{PR}_{ran}(l) = \frac{N}{\langle k_{in}\rangle N_{PR>l}(N_{PR>l}-1)} \times \sum_{l+1}^{l_{max}} \left\langle k_{in}^{pr=l''}\right\rangle P_{pr}(l'') \sum_{l+1}^{l_{max}} \left\langle k_{out}^{pr=l'}\right\rangle P_{pr}(l')$$

where $P_{pr}(l)$ denotes the probability of a node having PageRank l and where $\left\langle k_{in}^{pr=l}\right\rangle$ and $\left\langle k_{out}^{pr=l}\right\rangle$ denote the in-degree and out-degree averaged over all nodes of PageRank class l respectively.

C. Mixing properties for directed networks

The mixing properties of networks reveal information of the network join degree distribution. In the case of directed networks, similarly to (4), we define the average in-degree of the nearest neighbors, $k^{in}_{nn}(k)$, for vertices of in-degree k as

$$k^{in}_{nn}(k) = \frac{1}{N^{in}_k}\sum_i k^{in}_{nn,i}, \tag{10}$$

where the sum runs over all vertices of in-degree k and where N^{in}_k is the number of nodes of in-degree k and $k^{in}_{nn,i}$ denotes the average nearest neighbors in-degree of vertex i, i.e.

$$k^{in}_{nn,i} = \frac{1}{k^{in}_i}\sum_{j\in V(i)} k^{in}_j,$$

where the sum is over the nearest neighbors vertices of i. This metric enables distinction between assortative networks, where large in-degree nodes preferentially attach to large in-degree nodes, and disassortative networks, showing the opposite tendency. It is worth stressing that the rich-club phenomenon is not trivially related to the mixing properties of networks. Indeed, the rich-club phenomenon is not necessarily associated to assortative mixing.

Analogously to the in-degree, we can as well define the average PageRank of the nearest neighbors, $pr_{nn}(l)$ for vertices belonging to PageRank class l

$$pr_{nn}(l) = \frac{1}{N^{pr}_l}\sum_i pr_{nn,i}, \tag{11}$$

where the sum runs over all vertices of PageRank class l and where N^{pr}_l is the number of nodes of PageRank l and $pr_{nn,i}$ denotes the average nearest neighbors PageRank of vertex i, i.e.

$$pr_{nn,i} = \frac{1}{k^{in}_i}\sum_{j\in V(i)} pr(j),$$

where the sum is over the nearest neighbors vertices of i and k^{in}_i denotes the in-degree of node i. In general, this metric may be important because it can indirectly reveal the correlation between PageRank and in-degree. If $pr_{nn}(l)$ is increasing function of l, it will denote very low correlation between in-degree and PageRank because important nodes will generally get their vote from an important voters, hence a low number of these voters will be enough to achieve this "importance". Further in this case, we should expect page-club and rich-club to give different results, but not necessarily opposite. In fact, in none of the evaluated real networks an opposite behavior was observed. In the other case, when $pr_{nn}(l)$ is a constant (decreasing) function of l , a high-correlation between PageRank and in-degree should emerge because nodes having a high PageRank should have a high number (high in-degree) of medium important voters, whereas low PageRank nodes should have a low number (low in-degree) of medium important voters.

D. Synthetic network

Generally, all the results of the networks used in this paper show a high correlation between page-club and rich-club coefficients, see section IV, but one should not derive a general conclusion from this observation. In this section we generate a synthetic network showing increase of page-club coefficient and decrease of rich-club coefficient. Consider the directed tree graph $G = (V, E)$ where V represents the node set, and E, the edge set. Let this tree be with depth d, where the depth of a node i is the length of the path from the root to the node, and the depth of the tree is the maximal length of all such paths. Further, let G be a labeled graph, where each node can have one of several labels (depending on its depth in the tree), i.e. $V = V_0 \cup V_2 \ldots \cup V_d$. We denote by $|V_i|$ the number of depth i nodes. Assume that $|V_i| = i + 1, i = 0, \ldots, d$, i.e. the number of nodes increments as we move to a larger depth in the tree. We compose the edge set of the ordered pairs of vertices $\{(V_{ik}, V_{i-1,k}) \mid k \in (1, |V_{i-1}|), i \in (0, d)\} \cup \{V_{i,|V_i|} \times V_{i-1} \mid i \in (0, d)\}$ where $V_{i,k}$ denotes the k-th node in V_i. In other words, depth i nodes propagate their PageRank score to depth $i-1$ nodes, and furthermore, all nodes in the same class (depth) have equal PageRank values. By this construction, all the nodes, except the leaves, have in-degree 2. To change this property, we add additional set of nodes S_i connecting to the set V_i with the edge set $E_i = S_i \times V_i$, i.e. we increment the in-degree of the

nodes in the set V_i by $|S_i|$. Note that the in-degree of these additional nodes is 0. So, we tweak the in-degree of the nodes in the class V_i, $k^{in}_{V_i}$, by the following rule: beside the leaf nodes, we start with a $k^{in}_{V_0} = 2$ and increment the in-degree by one in every even depth, whereas for odd depths, we start with a high in-degree in lower depths and decrement the in-degree as we go in higher depths. A more formal definition would be

$$k^{in}_{V_i} = \begin{cases} 0, & i = d \\ \frac{i}{2} + 2, & i = 2n, n \in \mathbb{N} \\ \left\lfloor \frac{d-i}{2} \right\rfloor + 2, & i = 2n+1, n \in \mathbb{N} \end{cases} \tag{12}$$

Such graph with depth $d = 6$ is shown in Fig. 1. What we want to achieve is the lack of rich-club ordering where nodes with high in-degree connect to nodes with low in-degree and vise versa. Further, a direct consequence of the tree structure is that nodes with high page-rank will propagate their score to the successor nodes and therefore positive page-club ordering should arise. The results are shown in Fig. 2 for a generated graph of depth 50 with 1926 nodes. We observe the lack of rich-club ordering, but strong page-club ordering, as expected. Also, we stress that the top 612 in-degree nodes are not sharing any links. Thus, for such networks, the results of the analysis of the inter-connectivity of nodes, would clearly depend on the definition of the "rich" nodes (in-degree or PageRank).

IV. REAL NETWORKS: RESULTS AND DISCUSSIONS

The network data used in this paper consists of four networks [5]:

- Arxiv High Energy Physics paper citation network (cit-HepPh): Directed, Temporal, Labeled network with 34,546 nodes and 421,578 edges;
- Web graph from Google (web-Google): Directed network with 875,713 nodes and 5,105,039 edges;
- Citation network among US Patents (cit-Patents): Directed, Temporal, Labeled network with 3,774,768 nodes and 16,518,948 edges; and
- Email network from a EU research institution (email-EuAll): Directed network with 265,214 nodes and 420,045 edges.

For each network we compute normalized rich-club and page-club coefficients $\rho_{in}(k) = \phi^{in}(k)/\phi^{in}_{ran}(k)$, $\rho_{PR}(l) = \phi^{PR}(l)/\phi^{PR}_{ran}(l)$, as well as the mixing properties, average nearest neighbors in-degree and average nearest neighbors PageRank, $k^{in}_{nn}(k)$ and $pr_{nn}(l)$, respectively.

We stress that $\rho_{in}(k)$ and $\rho_{PR}(l)$ may, in some cases, be undefined. In the following, we discuss only the quantity $\rho_{in}(k)$ since the discussion for $\rho_{PR}(l)$ is exactly same.

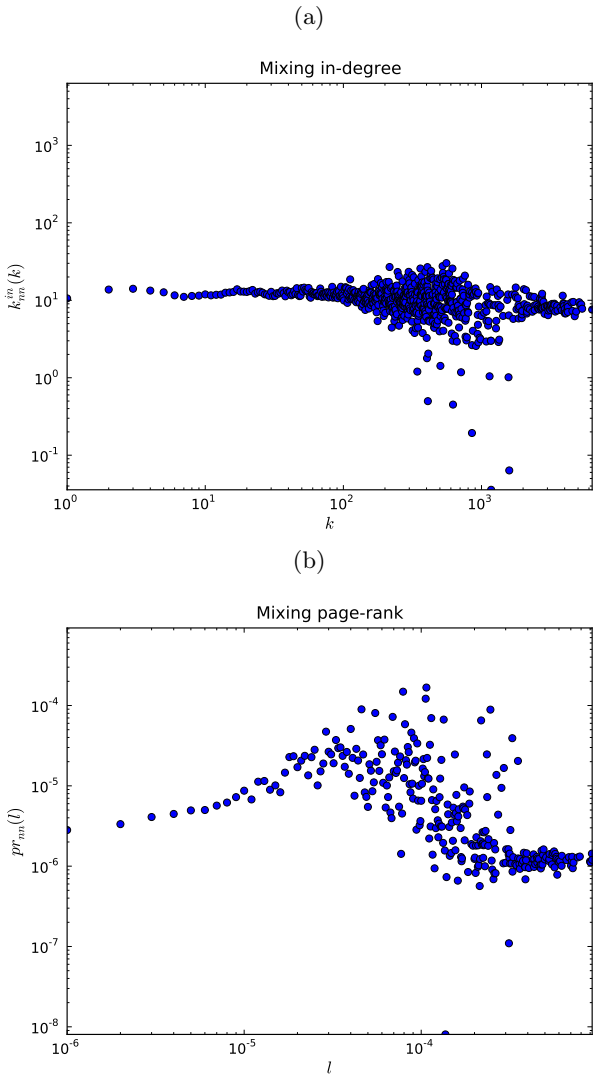


Figure 5: Web graph (released in 2002 by Google): (a) average in-degree of the nearest neighbors, and (b) average nearest neighbors PageRank.

$\rho_{in}(k)$ is undefined when its denominator is equal to zero, $\phi^{in}_{ran}(k) = 0$. We rewrite $\phi^{in}_{ran}(k)$ from Eq. (9) as

$$\phi^{in}_{ran}(k) = \frac{inlinks_{>k} outlinks_{>k}}{|E|\, N^{in}_{>k}(N^{in}_{>k} - 1)} \tag{13}$$

where $inlinks_{>k}$ ($outlinks_{>k}$) denotes all the *in-links* (*out-links*) arriving to (departing from) nodes that have in-degree greater than k and $|E|$ denotes the number of directed edges in the network.

We consider several cases:

- When $N^{in}_{>k} = 1$ or $N^{in}_{>k} = 0$, i.e. when we have a single node or no nodes in the "club".
- When $inlinks_{>k} = 0$. Note that this case should not happen in practice since in the two special cases of $\phi^{in}_{ran}(0)$ and $\phi^{in}_{ran}(k^{max}_{in})$ we generally have a positive number of *in-links*.

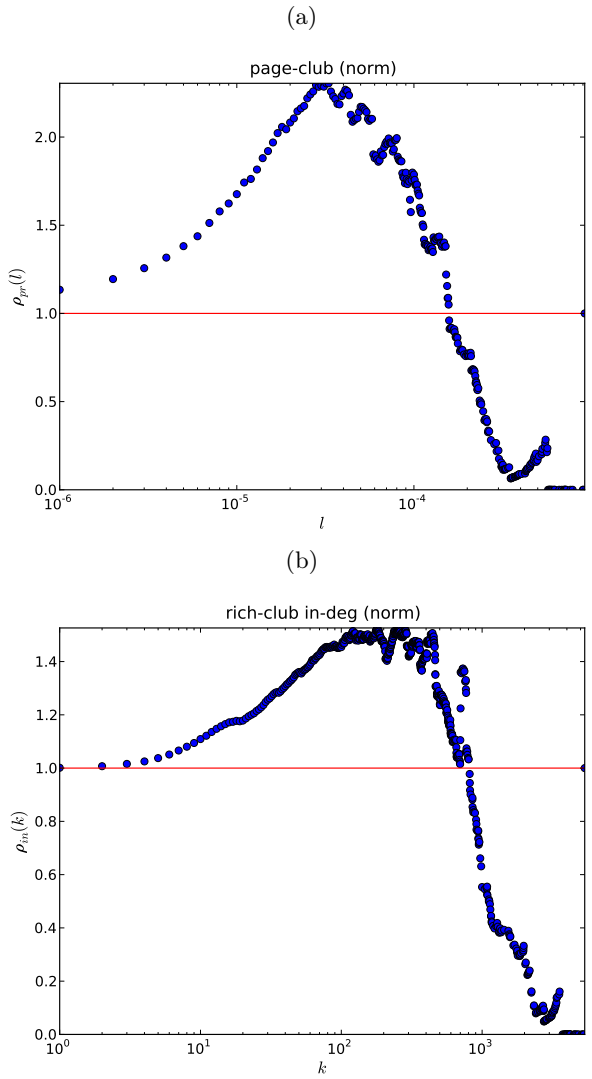


Figure 6: Web graph (released in 2002 by Google): (a) normalized page-club coefficient $\phi^{PR}(l)/\phi^{PR}_{ran}(l)$ versus page rank, and (b) normalized rich-club coefficient $\phi^{in}(k)/\phi^{in}_{ran}(k)$ versus in-degree. The page-club and rich-club ordering are observed only to some point, then the network shows the lack of page-club ordering and the lack of rich-club ordering.

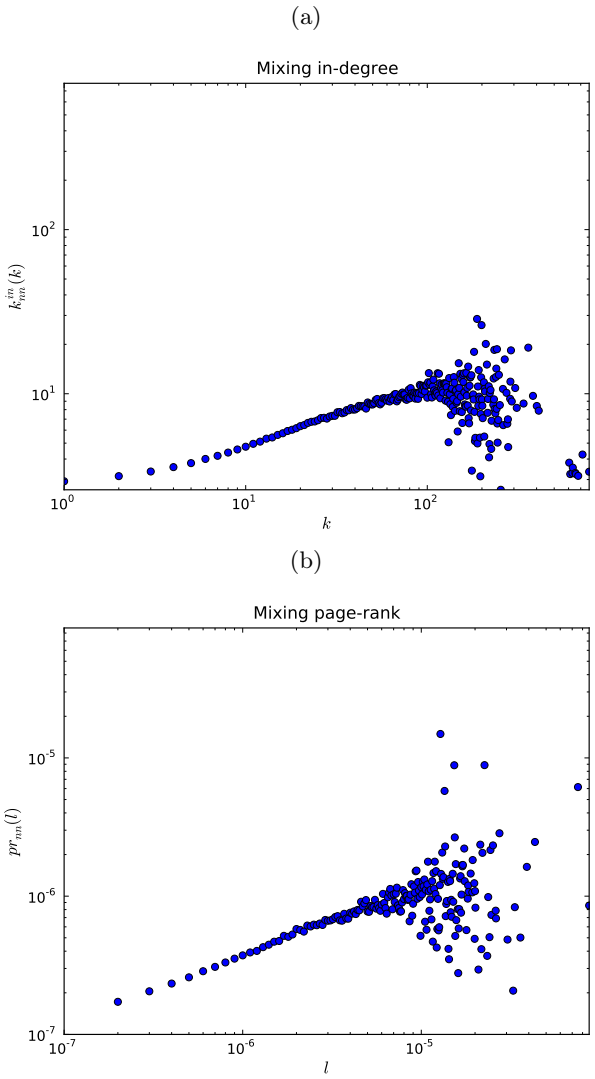


Figure 7: US Patents citation network: (a) average in-degree of the nearest neighbors, and (b) average nearest neighbors PageRank.

- When $outlinks_{>k} = 0$. This case can happen very often in tree graphs, such as citation networks, where the top in-degree nodes (roots) have no *out-links*.

We handle all these cases by assigning $\rho_{in}(k)$ a value 1. It is important to stress that $\rho_{in}(k)$ can have a value zero, i.e. its denominator can be well defined, thus having a positive number of *out-links* departing the "club", but no links are shared within the "club".

We also stress that for the PageRank computation we used the damping factor $q = 0.15$ for all the networks. We also used $q = 0.5$ for the journal citation network as proposed by [6] but no significant changes are observed, so these results are omitted.

A. Papers citation network

In Fig. 3 and Fig. 4 we show the results for the Arxiv High Energy Physics paper citation network. The network is formed from the e-print arXiv dataset and covers all the citations within a dataset of 34,546 papers with 421,578 edges. If a paper i cites paper j, the graph contains a directed edge from i to j. If a paper cites, or is cited by, a paper outside the dataset, the graph does not contain any information about this. The data covers papers in the period from January 1993 to April 2003 (124 months). It begins within a few months of the inception of the arXiv, and thus represents essentially the complete history of its HEP-PH section. The graph has an exponential degree distribution and a tree structure.

Fig. 3a shows the average in-degree of nearest neighbors $k^{in}_{nn}(k)$; clearly high in-degree pages generally get

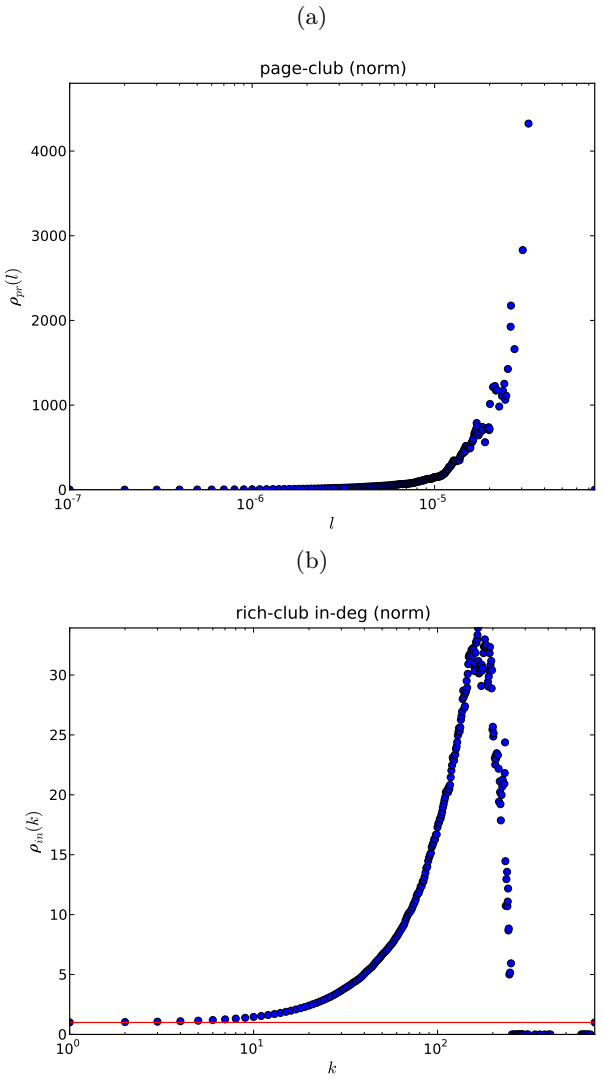


Figure 8: US Patents citation network: (a) normalized page-club coefficient $\phi^{PR}(l)/\phi^{PR}_{ran}(l)$ versus page rank (b) normalized rich-club coefficient $\phi^{in}(k)/\phi^{in}_{ran}(k)$ versus in-degree. The network shows both page-club and rich-club ordering.

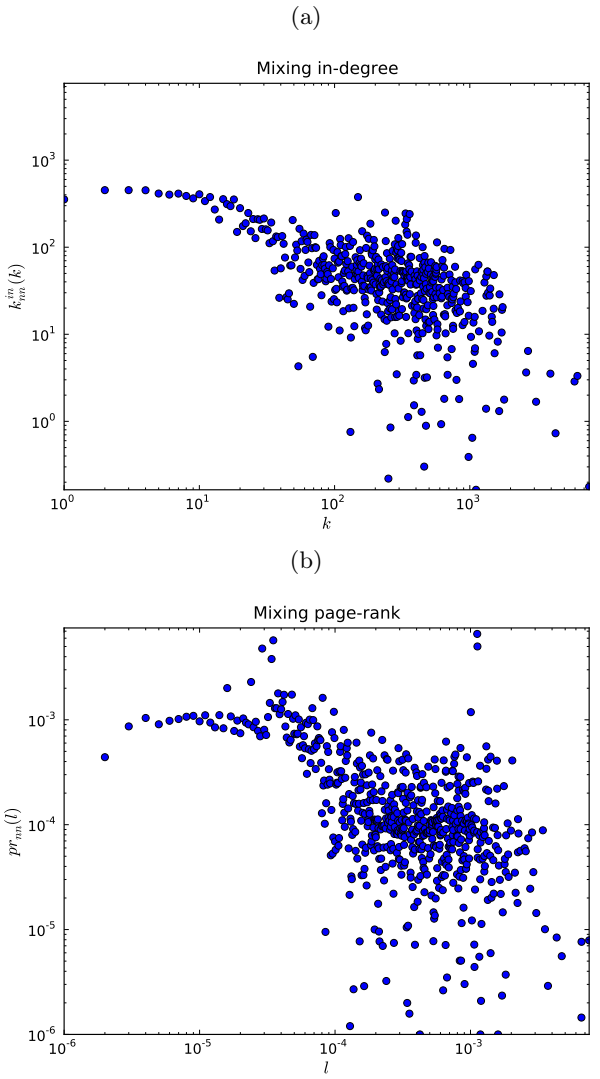


Figure 9: Email communication network: (a) average in-degree of the nearest neighbors, and (b) average nearest neighbors PageRank.

their in-degree from high in-degree pages. Note the large fluctuations in larger values of k. In Fig. 3b we show the average nearest neighbors PageRank $pr_{nn}(l)$ and identify that high PageRank nodes generally get their votes from high PageRank nodes. Note that here the fluctuations are much larger than in $k^{in}_{nn}(k)$.

In Fig. 4 we show the normalized rich-club and page-club coefficients and identify both rich-club and page-club ordering providing a conclusion that "elite" papers are cited by other "elite" papers. If we say that "elite" papers are written by "elite" scientists and those scientists decide to reference "elite" papers, i.e. papers written by other "elite" scientists than this result coincides with previous findings [3] i.e. it indicates existence of an "oligarchy" of highly influential and mutually communicating scientists. Note the difference between page-club and rich-club. The page-club ordering is much more stronger than the rich-club which can be explained by having in mind the tree structure in citation network, i.e. older papers which are higher in the hierarchy have higher PageRank retrieved from all the successors independently of the number of their direct successors, i.e., their in-degree. We also point that the top 8 PageRank nodes (roots) have no out-links, thus having a page-club value of one, whereas the top in-degree nodes have a positive rich-club ordering.

B. Google Web graph

In Fig. 5 and Fig. 6 we show the results of the analysis of the web graph released in 2002 by Google as a part of Google Programming Contest. Nodes represent

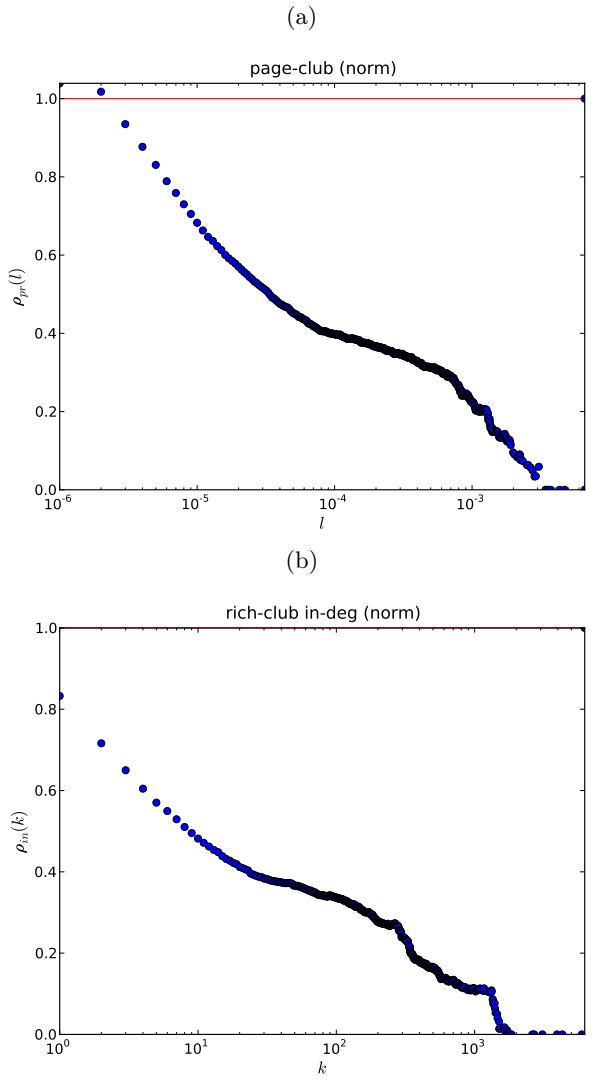


Figure 10: Email communication network: (a) normalized page-club coefficient $\phi^{PR}(l)/\phi^{PR}_{ran}(l)$ versus page rank (b) normalized rich-club coefficient $\phi^{in}(k)/\phi^{in}_{ran}(k)$ versus in-degree. The lack of page-club ordering and the lack of rich-club ordering are observed.

web pages and directed edges represent hyperlinks between them. The graph consists of nearly one million (1,000,000) nodes and over five million (5,000,000) edges following a power-law degree distribution.

Figure 5 indicates that $k^{in}_{nn}(k)$ and $pr_{nn}(l)$ are generally constant functions, revealing the uncorrelated property of the network. Note in $pr_{nn}(l)$ the large fluctuation appearing in the middle layer. In Fig. 6 we show the normalized rich-club and page-club coefficients where we identify partial rich-club (page-club) ordering up to some point (the middle layer) and then a narrow declining of the coefficients, which show the lack of rich-club ordering and the lack of the page-club ordering. In other words, in this networks high-degree (PageRank) pages purposely avoid sharing links with other high-degree (PageRank) pages. Some sort of competitiveness among strong pages could be a possible explanation of this phenomenon. Also note the uncorrelated property of the network, therefore explaining the high similarity between page-club and rich-club coefficient. We stress that the top 24 (20) in-degree (PageRank) nodes are not sharing any links between them beside their positive number of out-links, therefore the zero values of the rich-club and page-club coefficients.

C. US Patents citation network

In Fig. 7 and Fig. 8 we show the results of a U.S. patent dataset maintained by the National Bureau of Economic Research. The dataset spans 37 years (January 1, 1963 to December 30, 1999), and includes all the utility patents granted during that period, totaling about four million (4,000,000) patents. The citation graph includes all citations made by patents granted between 1975 and 1999, totaling 16,522,438 citations. For the patents dataset there are 1,803,511 nodes for which we have no information about their citations (we only have the in-links).

Figure 7 identifies a strong assortative mixing of $k^{in}_{nn}(k)$ and $pr_{nn}(l)$. Also note the large fluctuation appearing in higher values of k (l). In Fig. 8b we show the normalized rich-club coefficient: the rich-club ordering is observed up to some point, and after that point, the ordering quickly decreases to zero, where the top 27 in-degree nodes are not sharing any links between them. Fig. 8a shows the normalized page-club coefficient: one could identify page-club ordering providing the conclusion that top PageRank patents are cited by top PageRank patents. Note the much stronger page-club than rich-club ordering, generally, because of the tree structure.

D. Email communication network

In Fig. 9 and Fig. 10 we show the results of the analysis the network of email communication of a large European research institution. This network contains all incoming and outgoing email of the research institution for the period of October 2003 to May 2005 (18 months). Given a set of email messages, each node corresponds to an email address. A directed edge between nodes i and j was created if i sent at least one message to j. The network consists of 265214 nodes and 420045 edges.

In Fig. 9 (a) and (b) we show that $k^{in}_{nn}(k)$ ($pr_{nn}(l)$) is somewhat decreasing function of k (l) with very large fluctuations. Given its non-increasing property, see section III C, we should expect high correlation between PageRank and in-degree, thus, a high correlation between page-club and rich-club coefficients. Fig. 10 shows the normalized rich-club and page-club coefficients. The lack of the page-club ordering and the lack of the rich-club ordering for this network could be explained by observing that scientists are working in research groups where

each group has one to few "elite" scientists managing the group, where communication between the "elite" scientists from different groups is reduced to a minimum. The top 9 (6) in-degree (PageRank) nodes are not sharing any links between them beside their positive number of out-links, therefore the zero values of the rich-club and page-club coefficients.

V. CONCLUSION

In this paper several new metrics for directed graphs are introduced. Two such metrics are normalized rich-club coefficient and normalized page-club coefficient. For different directed graphs these two coefficients are computed as well as the average nearest neighbors in-degree and the average nearest neighbors PageRank. The results have indicated a high correlation between page-club and rich-club coefficients except for the synthetic network, for which the coefficients have opposite behavior. In general, beside the high correlation observed in several real networks, these metrics are not same. Detecting rich-club phenomenon often used to indicate the dominance of an "oligarchy" of "rich" and mutually communicating entities. However, this analysis clearly depends of the definition of "rich" nodes. The page-club coefficient annotates nodes with high PageRank as the "popular" nodes, thus, in networks where PageRank emerges as a natural metric for distinguishing between popular and unpopular nodes, one should use page-club to indicate the emergence of an "oligarchy" formed by "elite" nodes.

[1] A-L. Barabasi, "Linked: How Everything Is Connected to Everything Else", 2002; T. G. Lewis, "Network Science: Theory and Applications", Wiley, New York, April 2009.

[2] S. Zhou and R. J. Mondragon, "The rich-club phenomenon in the Internet topology", IEEE Communications Letters 8(3), 180–182 (2004).

[3] V. Colizza, A. Flammini, M. A. Serrano and A. Vespignani, "Detecting rich-club ordering in complex networks", Nature Physics 2, 110–115 (2006).

[4] S. Fortunato and A. Flammini, "Random Walks on Directed Networks: the Case of PageRank", International Journal of Bifurcation and Chaos 17(7), 2343–2353 (2007).

[5] http://snap.stanford.edu/data/index.html

[6] S. Maslov and S. Redner, "Promise and Pitfalls of Extending Google's PageRank Algorithm to Citation Networks," Journal of Neuroscience 28, 11103 (2008).