Computational Social Networks
a SpringerOpen Journal

RESEARCH   **Open Access**

Long-range degree correlations in complex networks

Michael Mayo[1*], Ahmed Abdelzaher[2] and Preetam Ghosh[2]

*Correspondence:
Michael.L.Mayo@usace.army.mil
[1] Environmental Laboratory, US
Army Engineer Research and
Development Center, 3909 Halls
Ferry Rd, 39183 Vicksburg, MS, USA
Full list of author information is
available at the end of the article

Abstract

Social networks are often degree correlated between nearest neighbors, an effect termed homophily, wherein individuals connect to nearest neighbors of similar connectivity. Whether friendships or other associations are so correlated beyond the first-neighbors, and whether such correlations are an inherent property of the network or are dependent on other specifics of social interactions, remains unclear. Here we address these problems by examining long-range degree correlations in three undirected online social and three undirected nonsocial (airport, transcriptional-regulatory) networks. Degree correlations were measured using Pearson correlation scores and by calculating the average neighbor degrees for nodes separated by up to 5 sequential links. We found that the online social networks exhibited primarily weak anticorrelation at the first-neighbor level, and tended more strongly towards disassortativity as separation distances increased. In contrast, the nonsocial networks were disassortative among first-neighbors, but exhibited assortativity at longer separation distances. In addition, the average degrees of the separated neighbors approached the average network connectivity after approximately 3-4 steps. Finally, we observed that two algorithms designed to grow networks on a node-by-node basis failed to reproduce all the correlative features representative of the social networks studied here.

Keywords: Social networks; Pearson correlation; Assortativity; Degree correlation

1 Introduction

A complex network is said to be degree correlated if the degrees of nodes at the end of links occur together in a nonrandom manner. The tendency of nodes to connect with others of similar degree is termed assortativity [1], or homophily when referenced specifically to social networks [2]. Conversely, nodes connected to others of dissimilar degree are said to be disassortatively mixed [1]. In a social network the nodes represent individuals, and the links between them conceptualize friendships or other social associations. In this setting, an assortative network emphasizes the surprising result that "...your friends have more friends than you do" [3]. Although results from the literature mostly involve degree mixing among nearest neighbors, little else has been reported regarding degree correlations extending beyond the first neighbors. Can a degree "correlation length" be defined for complex networks? If so, how far do degree correlations extend into a network based from any node?

Some insight into these questions has come from the social sciences. For example, the probability that an individual (termed the "ego") and his/her acquaintance (termed the "alter") are jointly obese decreases with geodesic distance (i.e., the number of sequential steps that link two nodes). However, this assortative effect is nearly independent of geographic distance [4], and is therefore a network property. Similar results hold for other health-related outcomes, such as smoking [5]. For a wide variety of social outcomes, such as happiness, divorce, depression, sleep length, marijuana use, Christakis and Fowler [6] reported assortative correlations to up to 4 and 6 steps from the ego.

Several explanations were proposed to explain these nonrandom effects [6]. For example, individuals could choose to associate with others of similar traits (homophily); individuals could associate with others exposed to similar environments; or traits could spread their influence through "conduction," like a contagion. However, such hypotheses cannot explain all assortative correlations beyond the first step (nearest neighbors), because similar effects have been observed in networks of more "autonomous" agents, such as food webs [7] (see also the commentary in [8]). Nevertheless, Christakis and Fowler conclude that, for social networks, traits extend to, on average, 3 steps beyond the ego [6]. Because node degrees are an elementary feature of networks and tend to correlate assortatively, we could ask: Does this 3-step observation hold for degree-degree correlations in general, across many different types of networks? If so, is there a defining mechanism for the effect?

Here we address these questions by comparing degree correlations for several large social networks to exemplary nonsocial ones, including an airline transportation network and two well-annotated transcriptional regulatory networks. We developed and executed an algorithm to evaluate degree correlations between nodes separated by more than one step, which is general enough to be applicable to nearly any undirected network. These methods could also be used to evaluate correlations between properties or features of the nodes beyond those associated directly with the network topology.
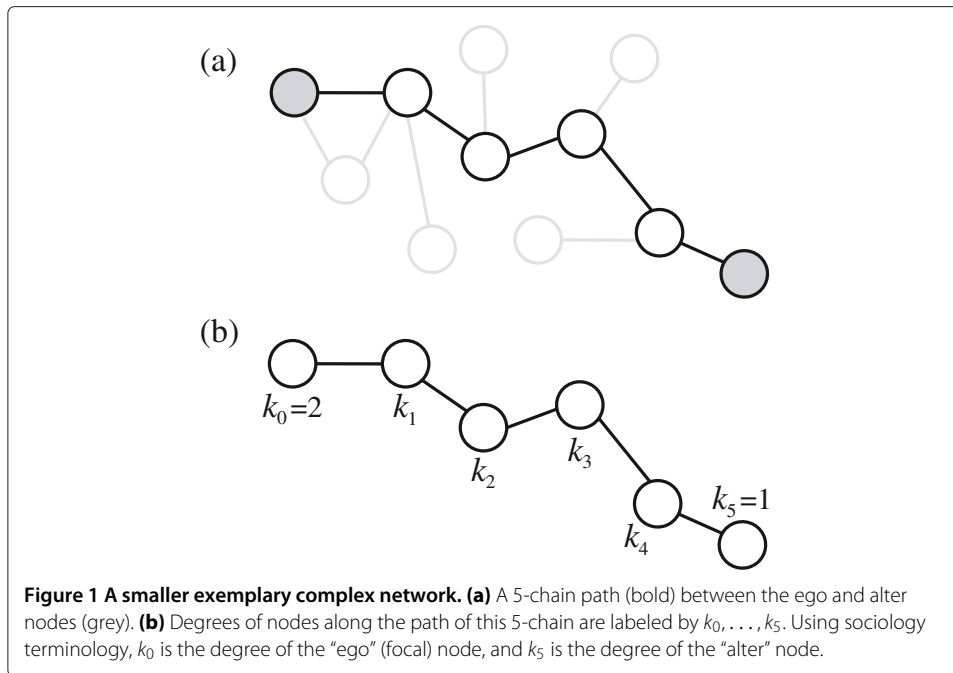
2 Methods

2.1 Measuring degree correlations

The degree of a node measures the number of links to its nearest neighbors, and is a critical property of the network topology because it accounts for coupling of each node to the greater network. It is therefore of great interest to examine the distribution of and correlation in network degrees. For simplicity, we will assume that all edges of the networks we examine are undirected (Section 2.3, below).

2.1.1 Average neighbor degree

The quantity $\langle k_1 k_0 \rangle$ may be calculated, which is the correlation function between k_0, the degree of the "ego" or focal node, and k_1, the degree of the "alter" node connected to the ego by 1 link (Figure 1). This two-point correlation function can be expressed as: $\langle k_1 k_0 \rangle = \sum_{k_0 k_1} k_1 k_0 p(k_1, k_0)$, wherein $p(k_1, k_0)$ is the joint probability that nodes of degree k_1 and k_0 appear together at the ends of a link [1]. Here, the sum spans $k_1, k_0 = 1, \ldots, L$, with L the number of links in the network. The joint probability can be expressed as $p(k_1, k_0) = p(k_1|k_0)p(k_0)$, so that the correlation between degrees is contained within the conditional probability, $p(k_1|k_0)$.

Figure 1 A smaller exemplary complex network. (a) A 5-chain path (bold) between the ego and alter nodes (grey). **(b)** Degrees of nodes along the path of this 5-chain are labeled by $k_0, \ldots, k_5$. Using sociology terminology, k_0 is the degree of the "ego" (focal) node, and k_5 is the degree of the "alter" node.

One way to determine whether any degree correlation exists is to measure the average nearest neighbor degree as a function of node degree, $\langle k_1^{\mathrm{nn}} \rangle (k_0)$. This quantity is directly linked to the conditional probability [9]:

$$\langle k_1^{\mathrm{nn}} \rangle (k_0) = \sum_{k_1} k_1 p(k_1 | k_0). \tag{1}$$

If the conditional probability is uncorrelated, then $p(k_1 | k_0) = p(k_1)$, and Eq. 1 can be evaluated to give $\langle k_1^{\mathrm{nn}} \rangle (k_0) = \langle k_0^2 \rangle / \langle k_0 \rangle$, which is a constant of the network. Here, $\langle k_1^{\mathrm{nn}} \rangle > \langle k_0 \rangle$ for nonzero variance, which quantifies the notion that "...your friends have more friends than you do" [3]. Thus, any observed dependence of $\langle k_1^{\mathrm{nn}} \rangle$ on k_0 indicates the presence of degree correlations in the network.

Do degree correlations extend beyond direct neighbors? To address this question, we extend Eq. 1 to nodes separated by long chains of m-many links. By "long chains", we mean the number of links separating one node (the ego) from another (the alter) that can be reached by following successive links, generation-by-generation, out from the ego without any back-tracking. We will refer to these paths as "m-chains", which constitute the shortest paths between two nodes of the network, and are identical to the geodesic distance between ego and alter nodes [6]. The basic idea is conceptualized in Figure 1, wherein m denotes the number of sequential links that compose the path between the ego and alter.

We label the joint probability that a node of degree k_0 is connected by an m-chain to another of degree k_m by $p_m(k_m, k_0)$. For $m = 1$, this quantity represents the one described in the previous section for nearest neighbors, and we drop the subscript: $p_1(k_1, k_0) = p(k_1, k_0)$. In a similar way that Eq. 1 links the average nearest-neighbor degree to the node degree, we have:

$$\langle k_m^{\mathrm{nn}} \rangle (k_0) = \sum_{k_m} k_m p_m(k_m | k_0). \tag{2}$$

Thus, any dependence of $\langle k_m^{\text{nn}} \rangle$ on k_0 signifies degree correlations between nodes at the ends of an m-chain.

2.1.2 Pearson correlation

The Pearson correlation, r, is the ratio of the covariance of fluctuations, $\langle (k_0 - \langle k_0 \rangle)(k_m - \langle k_m \rangle) \rangle$, to the variance in degree, $\langle k_0^2 \rangle - \langle k_0 \rangle^2$ [1]:

$$r = \frac{\langle (k_0 - \langle k_0 \rangle)(k_m - \langle k_m \rangle) \rangle}{\langle k_0^2 \rangle - \langle k_0 \rangle^2}, \tag{3}$$

wherein the average, $\langle \cdots \rangle$, is taken over the nodes connected by m-chains. For practical purposes, Eq. 3 can be implemented as a sum over m-chains:

$$r = \frac{L_m^{-1} \sum_i k_0^i k_m^i - \left[L_m^{-1} \sum_i \frac{1}{2} \left(k_0^i + k_m^i \right) \right]^2}{L_m^{-1} \sum_i \frac{1}{2} \left(k_0^{i\,2} + k_m^{i\,2} \right) - \left[L_m^{-1} \sum_i \frac{1}{2} \left(k_0^i + k_m^i \right) \right]^2}, \tag{4}$$

wherein L_m is the number of m-chains. Note that $L_1 = L$ is the number of links in the network, and that $L_m \geq L$.
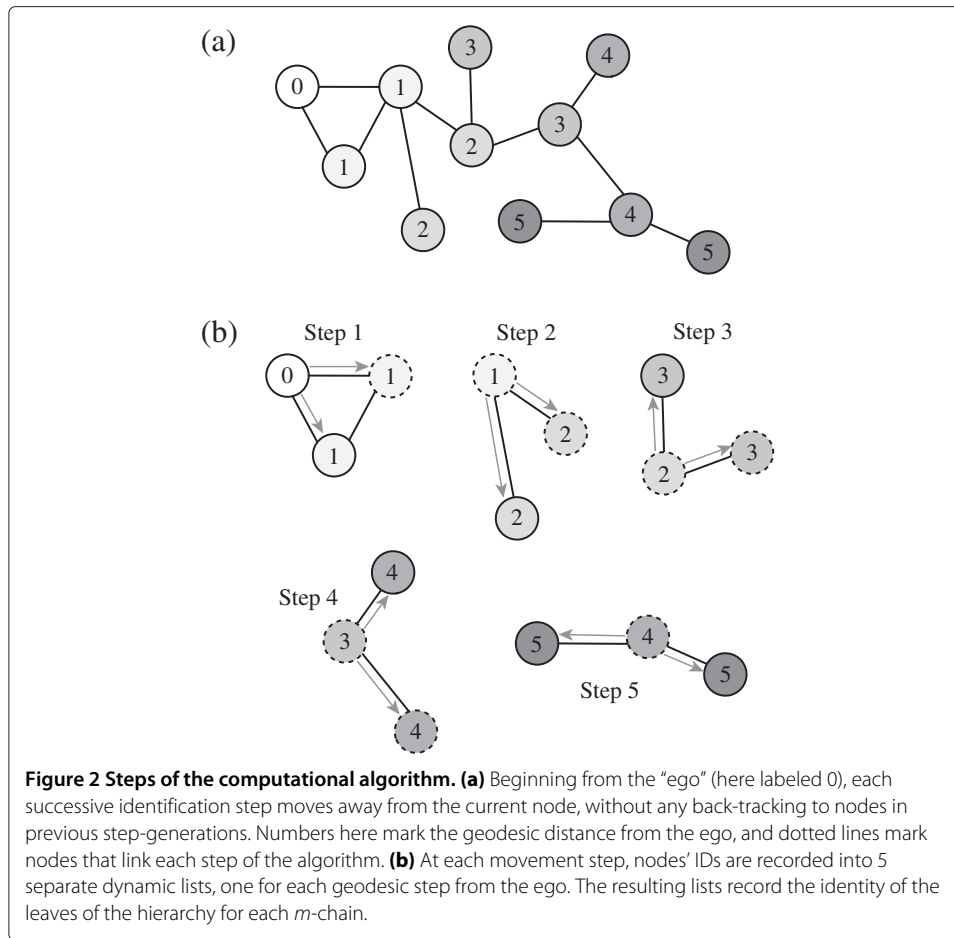
The Pearson correlation is often used to measure the assortativity of nodes connected by links (i.e., $m = 1$), because the variance in k_0 is equal to the value of $\langle k_0 k_m \rangle - \langle k_0 \rangle \langle k_m \rangle$ for a maximally assortative network [1]. Therefore, r is bounded on $[-1, 1]$; $r = -1$ corresponds to a purely disassortative network, while $r = 1$ marks a network as purely assortative. However, this measure obscures the k_0-dependence of $p_m(k_m|k_0)$ [10].

2.2 Algorithm to identify m-chain neighbors

We evaluated Eqs. 1 and 2 using a computational algorithm to determine nodes connected by m-chains, for $m = 1, \ldots, 5$, which is conceptualized in Figure 2. We chose a maximum geodesic distance of $m = 5$ to balance computational resources with the reports that such correlations nearly vanish for $m > 3$ [6]. Referring to Figure 2 with the understanding that the networks are undirected, the steps of the algorithm can be outlined as follows.

    I.   Choose an ego node (node 0, Figure 2(a));

    II.   Follow all links from the ego to its neighboring nodes, and append the IDs of these neighbor nodes into one of five dynamic lists, one list for each geodesic distance, m;

   III. Now, follow all links from these neighbor nodes to their neighbors (e.g., from node generations 1 to 2 in Figure 2(b)), excluding nodes already identified in a list from a previous generation/geodesic distance;

   IV. Continue this process until $m = 5$, then return to step I.

We followed a method outlined in Ref. [11] to evaluate Eqs. 1 and 2, but using nodes connected by m-chains obtained from the above algorithm. For each node i with degree k_0^i in the network (the ego), we identified a set of neighbor nodes found at the end of each m-chain, n_m^j (the alters), each of which has degree k_m^j. To each ego node, we associated an averaged m-chain neighbor degree: $\langle k_m \rangle \left(k_0^i \right) = \left| \left\{ n_m^j \right\} \right|^{-1} \sum_{\{n_m^j\}} k_m^j$. Finally, we took the arithmetic average of all instances of a given degree value, $k_0^i = k$, to give [11]: $\langle k_m \rangle = |\{\langle k_m \rangle\}|^{-1} \sum_{k_0^i = k} \langle k_m \rangle \left(k^i \right)$.

Mayo *et al. Computational Social Networks* (2015) 2:4

Page 5 of 13

Figure 2 Steps of the computational algorithm. (a) Beginning from the "ego" (here labeled 0), each successive identification step moves away from the current node, without any back-tracking to nodes in previous step-generations. Numbers here mark the geodesic distance from the ego, and dotted lines mark nodes that link each step of the algorithm. **(b)** At each movement step, nodes' IDs are recorded into 5 separate dynamic lists, one for each geodesic step from the ego. The resulting lists record the identity of the leaves of the hierarchy for each *m*-chain.
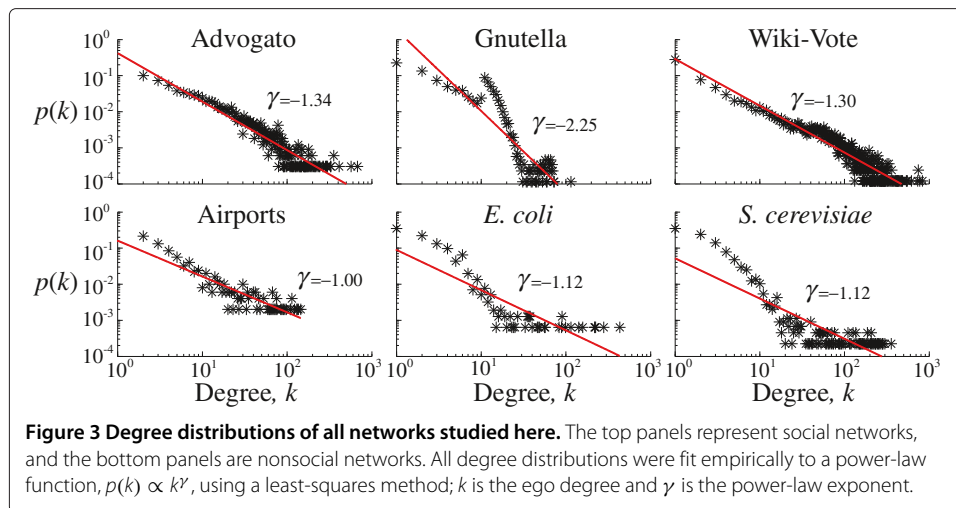
2.3 Network datasets

We studied several online social network datasets, and compared their results to those obtained from a transportation network and two transcriptional regulatory networks. By social network, we mean a network wherein the nodes represent individuals and the links between them signify social associations. All of these networks were manifestly directed, but for simplicity we studied them as undirected networks by examining their total degree, which is the sum of in- and out-degrees for each node, and ignoring link-directions. Although many nodes could therefore support multiples links, we found that all of the considered networks, both social and nonsocial, closely followed a "scale-free" degree distribution, $p(k) \propto k^{-\gamma}$ ($k_0 = k$ for notational convenience), as shown in Figure 3.

2.3.1 Online social networks

We evaluated a dataset from the Advogato online social network, wherein users can express the level of "trust" between themselves and another [12]. As mentioned above, we are only interested in the structure of the links between all individuals, and therefore ignored any weights assigned to them. The Advogato network is composed of 3,302 nodes/users linked together by 32,954 links.

A snapshot of the decentralized Gnutella peer-to-peer file-sharing network was captured on 6 August, 2002 [13]. In this dataset, the 8,717 nodes represent the hosts, and each of the 31,525 links signify connections established between them.

Mayo *et al. Computational Social Networks* (2015) 2:4

Page 6 of 13

Figure 3 Degree distributions of all networks studied here. The top panels represent social networks, and the bottom panels are nonsocial networks. All degree distributions were fit empirically to a power-law function, $p(k) \propto k^{\gamma}$, using a least-squares method; k is the ego degree and γ is the power-law exponent.

The Wiki-Vote network was derived from a complete dump of the Wikipedia page-edit history (3 January, 2008) [14,15]. Wikipedia users may be promoted to administrators, who enjoy additional technical and maintenance capabilities of the website, which requires a public vote among its users. In this network, all 8,297 nodes represent individual users, and each of the 103,689 links indicate that one person voted for the other.
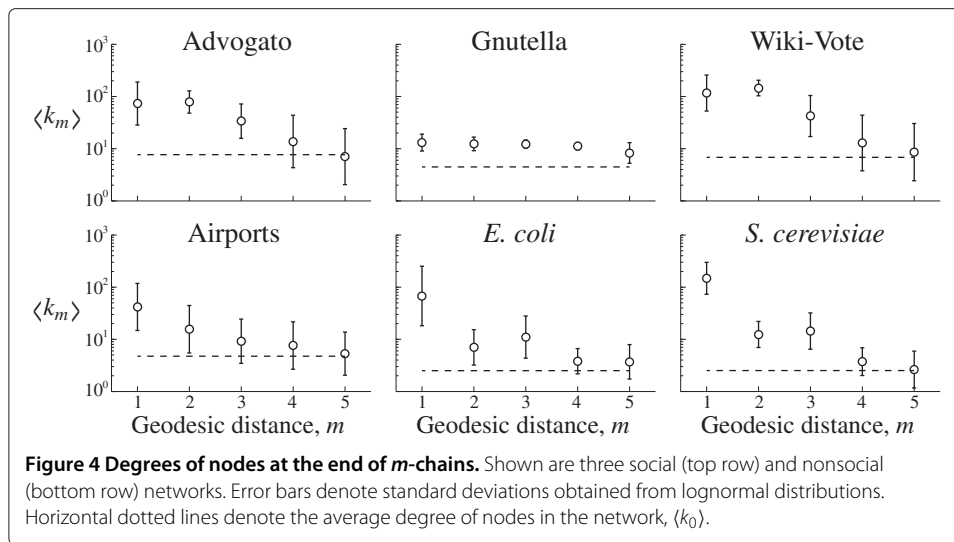
2.3.2 Nonsocial networks

As examples of nonsocial networks, we chose a physical transportation network, labeled "Airports". This network maps flights scheduled between the 500 busiest airports in the United States (US) in 2002 [16]. In this dataset, a node represents one of 500 US airports, while each of its 2980 links denote whether a flight was scheduled from one airport to another. While this network is manifestly undirected, it is weighted. We therefore ignored the weights in favor of the network topology alone.

We compared this transportation network with two transcriptional regulatory networks, which relate the expression of genes (nodes) that interact by producing proteins, termed transcription factors, that may alter the expression level of other genes. We employed two experimentally validated datasets from the literature, obtained using the GeneNetWeaver software package [17]; one for the model bacterium *Escherichia coli* (*E. coli*), and the common baker's yeast *Saccharomyces cerevisiae* (*S. cerevisiae*). The *E. coli* network consisted of 1565 nodes and 3758 (directed) links, whereas the *S. cerevisiae* network supported 4441 nodes and 12873 links. While the degree distribution of these networks generally follows a power-law (Figure 3), its structure differs substantially from a social network in that it is primarily hierarchical [18], with a few apical "master regulator" proteins that control the expression of a great many genes.

3 Results

3.1 Above-average *m*-chain neighbor degrees in social networks

Figure 4 shows the average degrees of nodes found at the end of all *m*-chains, $\langle k_m \rangle$, independent of the starting point. The long-distance behavior of this metric should be intuitive: as we move step-by-step away from a node, the average degree of nodes found

Mayo *et al. Computational Social Networks* (2015) 2:4

Page 7 of 13

Figure 4 Degrees of nodes at the end of *m*-chains. Shown are three social (top row) and nonsocial (bottom row) networks. Error bars denote standard deviations obtained from lognormal distributions. Horizontal dotted lines denote the average degree of nodes in the network, $\langle k_0 \rangle$.

at the end of the chain should approach the average connectivity of the graph, $\langle k_0 \rangle$ (dotted lines, Figure 4). To estimate $\langle k_m \rangle$, we observed that *m*-chain degree neighbor distributions appeared lognormal, from which we estimated the mean (circles) and standard deviation (error bars); however, the degree distribution of the nodes themselves, $\langle k_0 \rangle$, were power-law distributed (Figure 3).
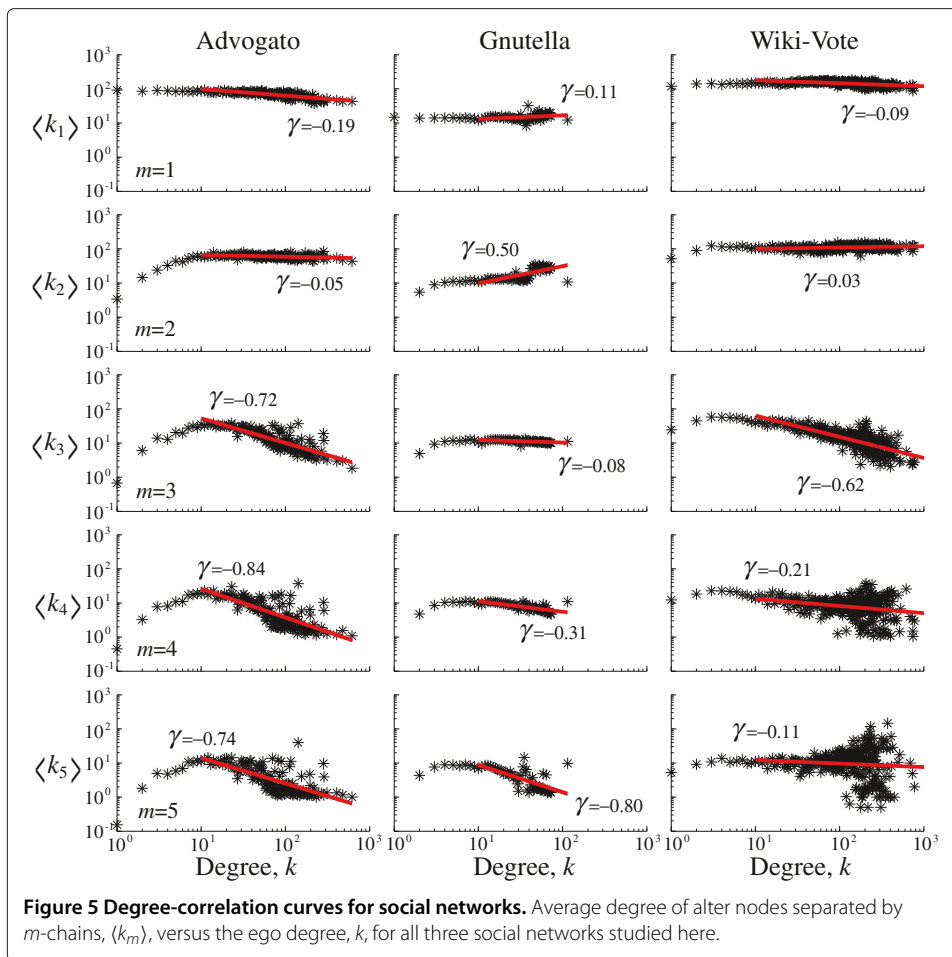
For the nonsocial networks (bottom row of Figure 4), the condition $\langle k_m \rangle = \langle k_0 \rangle$ occurs at approximately $m = 3$ or $m = 4$, while for the social networks we find $m \geq 4$. Additionally, the quantity $\langle k_m \rangle$ ($m > 0$) remains elevated over the identical geodesic length of the nonsocial networks. In other words, not only do your friends have more friends than you do, but so do your friends' friends' friends' friends.

One potential explanation for this effect may come from the tendency for social networks to form larger clumps of highly-connected nodes that, together, are only sparsely connected [19]. If nodes that are connected through *m*-chains can often be found within a highly-connected community, or if a node within a community can be easily reached through an *m*-chain, then $\langle k_m \rangle$ will be biased toward the connectivity of the community.

3.2 Assortative mixing beyond the nearest neighbors in social networks

Figure 5 illustrates how the average *m*-chain neighbor degree, $\langle k_m \rangle$, varies with ego degree, k, for the three social networks; Figure 6 illustrates this relationship for the three additional nonsocial networks. It has been noted previously that some networks exhibit non-monotone degree correlation [10], with a cross-over point near $k = 10$, which has been observed before in models of random networks [20]. We therefore used a power-law function, $\langle k_m \rangle(k) \sim k^{\gamma}$, wherein $k_0 = k$ labels ego degrees, to empirically model the tail of the *m*-chain neighbor degrees. This feature is not clearly present in the nonsocial networks; so, we fit a power-law function across the whole domain of its degree.

We can make several observations by comparing results from the social networks (Figure 5) to results from nonsocial networks (Figure 6). First, as geodesic distance increases, $\langle k_m \rangle$ for all social networks exhibits disassortative tendency. Park and Newman have argued [21] that social networks are different from other networks in that they are substantially assortative in nearest neighbor degree correlations. While $\langle k_1 \rangle$ for the social
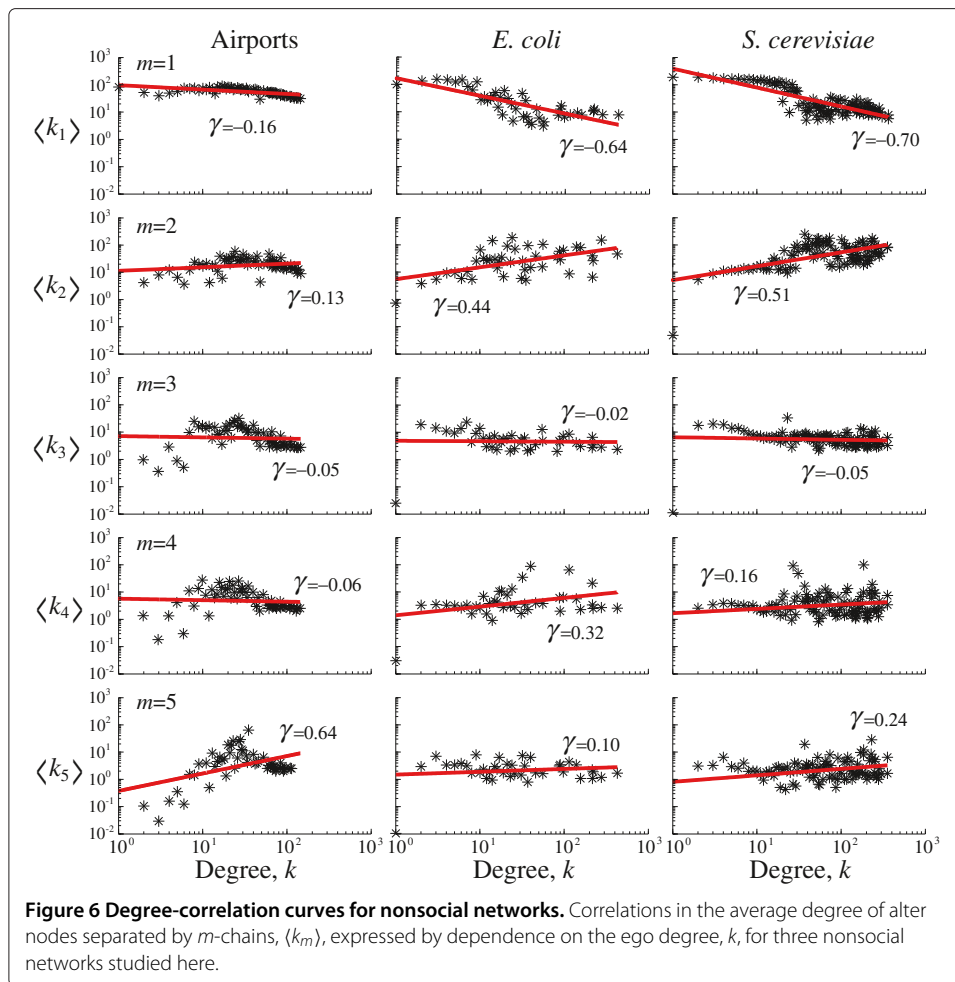
Mayo *et al. Computational Social Networks* (2015) 2:4

Page 8 of 13

Figure 5 Degree-correlation curves for social networks. Average degree of alter nodes separated by *m*-chains, $\langle k_m \rangle$, versus the ego degree, *k*, for all three social networks studied here.

networks of Figure 5 exhibit nearly flat correlation, the nonsocial networks of Figure 6 appear disassortative in $\langle k_1 \rangle$. In light of the argument made by Park and Newman [21], the nearly flat behavior of $\langle k_1 \rangle$ seen in Figure 5 could result from positive correlative trends.

Another observation we can make by comparing Figures 5 and 6, is that the nonsocial networks, specifically the transcriptional networks, exhibit opposite correlations between 1- and 2-chain neighbors, $\langle k_1 \rangle$ and $\langle k_2 \rangle$, respectively. Additionally, the extended correlations ($m > 2$) in the nonsocial networks are consistently positive (Figures 6 and Tables 1 and 2), which should be contrasted against the consistently disassortative correlations ($\gamma < 0$ and $r < 0$) exhibited by the social networks (Figure 5).

3.3 Network growth models cannot fully explain long-range social network correlations

Do long-range disassortative correlations observed for social networks in Figure 5, occur in networks created using random mechanisms? To address this problem, we used two node-by-node network-growing algorithms. The first is a modified version of the well-known Barabàsi-Albert model [22] which reproduces scale-free degree distributions. Networks grown using this algorithm are known to generate degree correlations at the nearest-neighbor level due to the preference of older nodes to acquire more links [23]. We have implemented this model with the addition of incorporating a selection method that

Mayo *et al. Computational Social Networks* (2015) 2:4

Page 9 of 13

Figure 6 Degree-correlation curves for nonsocial networks. Correlations in the average degree of alter nodes separated by m-chains, $\langle k_m \rangle$, expressed by dependence on the ego degree, k, for three nonsocial networks studied here.

Table 1 Correlation metrics for online social networks

	Correlation metric	
Geodesic distance, *m*	**γ**	**r**
	Advogato	
1	-0.1892 [-0.2156, -0.1637]	-0.0996
2	-0.0505 [-0.0735, -0.0275]	-0.0559
3	-0.7166 [-0.7969, -0.6363]	-0.2308
4	-0.8408 [-0.9624, -0.7191]	-0.1687
5	-0.7448 [-0.8722, -0.6174]	-0.0783
	Gnutella	
1	0.1066 [0.0110, 0.2021]	0.0516
2	0.5027 [0.3675, 0.6379]	0.2904
3	-0.0776 [-0.1140, -0.0411]	-0.0914
4	-0.3106 [-0.3920, -0.2292]	-0.1609
5	-0.8001 [-1.0349, -0.5654]	-0.2012
	Wiki-Vote	
1	-0.0889 [-0.1042, -0.0737]	-0.0721
2	0.0371 [0.0195, 0.0547]	-0.0665
3	-0.6191 [-0.6754, -0.5628]	-0.2600
4	-0.2129 [-0.3050, -0.1208]	-0.1022
5	-0.1051 [-0.2475, 0.0374]	-0.0438

Power-law exponents from Figure 5 reported with 95% confidence intervals, and Pearson correlation scores calculated using Eq. 4.
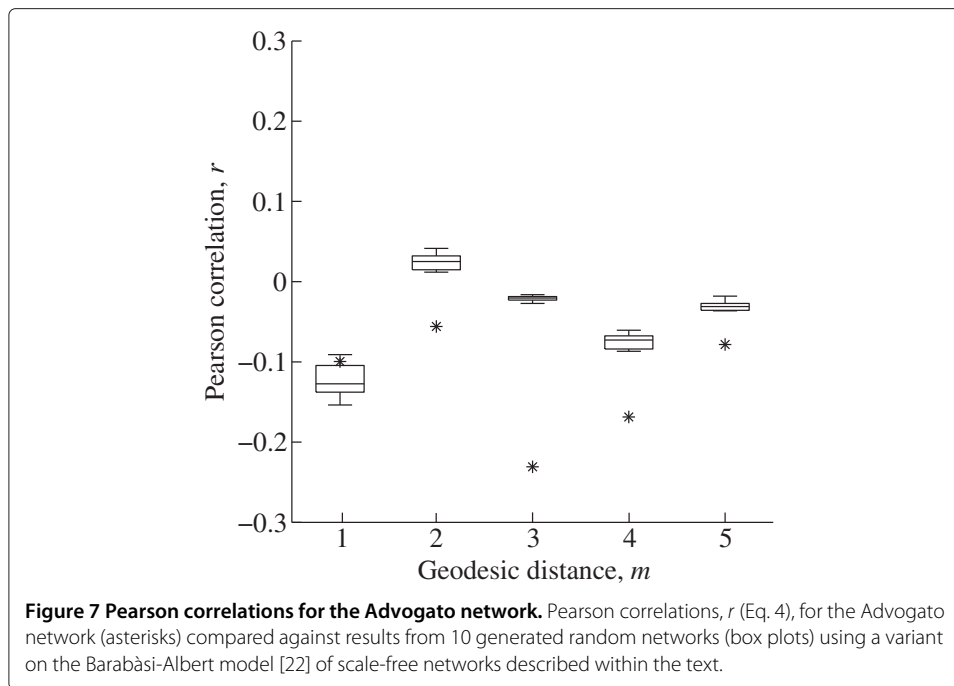
Mayo *et al. Computational Social Networks* (2015) 2:4

Page 10 of 13

Table 2 Correlation metrics for nonsocial networks

Geodesic distance, m	Correlation metric	
	γ	r
	Airports	
1	-0.1612 [-0.2160, -0.1065]	-0.2679
2	0.1332 [-0.0134, 0.2798]	-0.6291
3	-0.0467 [-0.2775, 0.1840]	-0.5655
4	-0.0553 [-0.2916, 0.1809]	-0.6014
5	0.6351 [0.2824, 0.9877]	-0.5974
	E. coli	
1	-0.6441 [-0.8127, -0.4756]	-0.3274
2	0.4348 [0.2096, 0.6599]	0.2695
3	-0.0173 [-0.2507, 0.2160]	-0.0304
4	0.3207 [0.0586, 0.5828]	-0.0076
5	0.1040 [-0.1330, 0.3411]	-0.0235
	S. cerevisiae	
1	-0.6961 [-0.7920, -0.6002]	-0.5967
2	0.5103 [0.3732, 0.6474]	0.2197
3	-0.0466 [-0.1680, 0.0749]	-0.0608
4	0.1584 [-0.0033, 0.3200]	-0.0055
5	0.2376 [0.0758, 0.3994]	-0.0041

Power-law exponents from Figure 6 reported with 95% confidence intervals, and Pearson correlation scores calculated using Eq. 4.

allows for a variable number of links to be drawn at each attachment step. More specifically, we choose to attach l-many links at each attachment step by rounding $NP(x \leq k)$ up to the nearest whole number l, wherein $P(x \leq k)$ is the cumulative degree distribution and N is the sum-total of nodes, both evaluated at the current attachment step. Because all random networks are grown on a node-by-node basis, wherein the number of links are determined by the step-wise attachment algorithm, we "grew" each network to the size of a chosen representative social network, the Advogato network, which hosts a total of 3302 nodes.

The other node-by-node attachment mechanism was reported by Vàzquez [20], and termed the "random walk" model. Here, nodes are attached following the linear attachment kernel of the Barabàsi-Albert model as stated above, but an additional step is added: a neighbor node is chosen at random with uniform probability, and with probability q_e, a link is drawn from the candidate node (the one just attached) to the neighbor node. If this secondary link attachment is successful, then this "random walk" procedure continues until the check of each new q_e fails. A primary feature of networks grown using this model is that their degree correlations are "mixed"; that is, lower-degree nodes exhibit positive correlations, while the higher degree nodes exhibit disassortative tendencies. We have previously observed such behavior in a wide variety of directed, real-world networks [10], but this behavior can also be seen in the behavior of $\langle k_m \rangle(k)$ for the social networks illustrated in Figure 5.

Figure 7 shows how the altered version of the Barabàsi-Albert model performs in terms of Pearson correlation scores (box plots), compared to a representative social network, the Advogato online social network (asterisks). While we can see that the slope of the power-law tail for the Advogato network indicates higher levels of disassortativity at

Mayo *et al. Computational Social Networks* (2015) 2:4

Page 11 of 13

Figure 7 Pearson correlations for the Advogato network. Pearson correlations, *r* (Eq. 4), for the Advogato network (asterisks) compared against results from 10 generated random networks (box plots) using a variant on the Barabàsi-Albert model [22] of scale-free networks described within the text.

higher degree nodes (Figure 5), the Pearson scores show a weaker overall correlation at long geodesic distances; however, the random network models show nearly no correlation except among first-neighbors ($m = 1$, Figure 7).

This can be contrasted against results from the random walk model of Vàzquez [20], illustrated for various values of q_e in Figure 8. These random networks generally show assortative degree correlations in first-neighbors for all values of q_e, but mostly disassortative degree mixing among nodes at longer geodesic distances. This result is generally consistent with the trends observed for the social networks (see Figure 5, and asterisks in
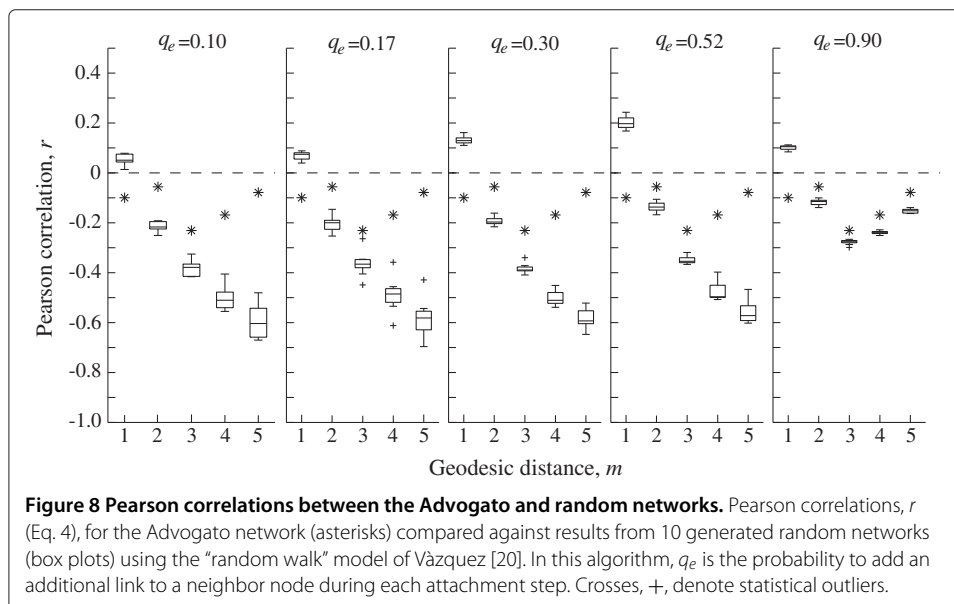

Figure 8 Pearson correlations between the Advogato and random networks. Pearson correlations, *r* (Eq. 4), for the Advogato network (asterisks) compared against results from 10 generated random networks (box plots) using the "random walk" model of Vàzquez [20]. In this algorithm, q_e is the probability to add an additional link to a neighbor node during each attachment step. Crosses, +, denote statistical outliers.

Mayo *et al. Computational Social Networks* (2015) 2:4

Page 12 of 13

Figure 8). Nevertheless, close matching of the Pearson scores only occurs for $q_e = 0.9$. Such a high value of q_e guarantees many successful sequential attachment rounds in the random walk procedure, and thus increases the overall number of links. Whether the closer matching of Pearson scores at high q_e is the result of the increased number of links, or their approximate placement, remains an open question.

4 Summary and Conclusions

In this paper we have studied three online social networks, and compared their long-range degree correlation behavior to those of three nonsocial networks by measuring both the average number of neighbors or calculating the Pearson correlation score. We found that the number of friendships/associations in the social networks remained above the background level for at least $m = 4$ "degrees of separation". In contrast, the nonsocial networks reached the background level after approximately $m = 3$ steps from each node.

We also examined the conditional probability that a node degree is connected to one separated by at least one link, $p(k_m|k_0)$, by measuring the average number of nearest neighbors, $\langle k_m \rangle (k_0)$. We found that the social networks generally exhibited a power-law tail with exponent $\gamma < 1$, for the longer-range interactions ($m \geq 3$). We did not observe this phenomenon in the nonsocial networks, which appeared nearly uncorrelated at this geodesic distance.

Finally, we considered the Advogato network as a prototypical social network, and examined whether two network-growing algorithms known to generate degree correlations could reproduce the long-range correlations observed in the social network as measured by the Pearson correlation. While we observed that the "random walk" algorithm [20] and a variant of the celebrated Barabàsi-Albert (preferential attachment) model [22] showed similar uncorrelated results at the farthest separation ($m = 5$), correlations in the Advogato network deviated substantially from the random models for $m < 5$. We conclude that these random node-attachment mechanisms cannot fully explain how social networks gain new users, but could not entirely reject this possibility. Further investigations are therefore required to definitively answer this question.

Author details
[1]Environmental Laboratory, US Army Engineer Research and Development Center, 3909 Halls Ferry Rd, 39183 Vicksburg, MS, USA. [2]Department of Computer Science, Virginia Commonwealth University, 401 West Main Street, 23284 Richmond, VA, USA.

Mayo *et al. Computational Social Networks*   (2015) 2:4

Page 13 of 13

References

1. Newman, MEJ: Assortative mixing in networks. Phys. Rev. Lett. **89**, 208701 (2002)
2. McPherson, M, Smith-Lovin, L, Cook, JM: Birds of a feather: Homophily in social networks. Annu. Rev. Sociol. **27**, 415–444 (2001)
3. Feld, SL: Why your friends have more friends than you do. Am. J. Sociol. **96**, 1464–1477 (1991)
4. Christakis, NA, Fowler, JH: The spread of obesity in a large social network over 32 years. New Engl. J. Med. **357**, 370 (2007)
5. Christakis, NA, Fowler, JH: The collective dynamics of smoking in a large social network. New Engl. J. Med. **358**, 2249 (2008)
6. Christakis, NA, Fowler, JH: Social contagion theory: examining dynamic social networks and human behavior. Stat. Med. **32**, 556 (2011)
7. Dawah, HA, Hawkins, BA, Claridge, MF: Structure of the parasitoid communities of grass-feeding chalcid wasps. J. Anim. Ecol. **64**, 708 (1995)
8. Redner, S: Teasing out the missing links. Nature. **453**, 47 (2008)
9. Pastor-Satorras, R, Vàzquez, A, Vespignani, A: Dynamical and correlation properties of the internet. Phys. Rev. Lett. **87**, 258701 (2001)
10. Mayo, M, Abdelzaher, A, Ghosh, P: Mixed degree-degree correlations in directed social networks. In: Zhang, Z, Wu, L, Xu, W, Du, D-Z (eds.) Combinatorial Optimization and Applications. Lecture Notes in Computer Science, pp. 571–580. Springer, (2014)
11. Serrano, M, Maguitman, A, Boguñá, M, Fortunato, S, Vespignani, A: Decoding the structure of the www: A comparative analysis of web crawls. ACM Trans. Web. **1**, 10 (2007)
12. Massa, P, Salvetti, M, Tomasoni, D: Bowling alone and trust decline in social network sites. In: Proceedings of the Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing, DASC'09, pp. 658–663. IEEE, (2009)
13. Ripeanu, M, Foster, I, Iamnitchi, A: Mapping the gnutella network: Properties of large-scale peer-to-peer systems and implications for system design. IEEE Internet. Comput. **6**, 50–57 (2002)
14. Leskovec, J, Huttenlocher, D, Kleinberg, J: Signed networks in social media. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1361–1370. ACM, (2010)
15. Leskovec, J, Huttenlocher, D, Kleinberg, J: Predicting positive and negative links in online social networks. In: Proceedings of the 19th International Conference on World Wide Web, pp. 641–650. ACM, (2010)
16. Colizza, V, Pastor-Satorras, R, Vespignani, A: Reaction-diffusion processes and metapopulation models in heterogeneous networks. Nature Physics. **3**, 276–282 (2007)
17. Schaffter, T, Marbach, D, Floreano, D: Genenetweaver: in silico benchmark generation and performance profiling of network inference methods. Bioinformatics. **27**(16), 2263–2270 (2011). doi:10.1093/bioinformatics/btr373
18. Ma, H-W, Buer, J, Zeng, A-P: Hierarchical structure and modules in the escherichia coli transcriptional regulatory network revealed by a new top-down approach. BMC Bioinf. **5**, 199 (2004)
19. Girvan, M, Newman, MEJ: Community structure in social and biological networks. P. Natl. Acad. Sci. USA. **99**, 7821–7826 (2002)
20. Vàzquez, A: Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev. E. **67**, 056104 (2003)
21. Park, J, Newman, MEJ: Origin of degree correlations in the internet and other networks. Phys. Rev. E. **68**, 026112 (2003)
22. Barabàsi, A-L, Albert, R: Emergence of scaling in random networks. Science. **286**, 509–512 (1999)
23. Krapivsky, PL, Redner, S: Organization of growing random networks. Phys. Rev. E. **63**, 066123 (2001)