Pramana – J. Phys. (2016) 87: 84
DOI 10.1007/s12043-016-1290-y

Linear analysis of degree correlations in complex networks

JU XIANG[1,2,3,*], TAO HU[4], YAN ZHANG[5,*], KE HU[6], YAN-NI TANG[3], YUAN-YUAN GAO[3] and KE DENG[7]

[1]Neuroscience Research Center, Changsha Medical University, Changsha 410219, China
[2]Department of Anatomy, Histology and Embryology, Changsha Medical University, Changsha 410219, China
[3]Department of Basic Medical Sciences, Changsha Medical University, Changsha 410219, China
[4]College of Science, Qi Lu University of Technology, Jinan 250353, Shandong, China
[5]Department of Computer Science, Changsha Medical University, Changsha 410219, China
[6]Department of Physics, Xiangtan University, Xiangtan 411105, Hunan, China
[7]Department of Physics, Jishou University, Jishou 416000, Hunan, China
*Corresponding authors. E-mail: xiang.ju@foxmail.com; zhangyancsmu@formail.com

MS received 7 July 2015; revised 3 February 2016; accepted 7 March 2016; published online 2 November 2016

Abstract. Many real-world networks such as the protein–protein interaction networks and metabolic networks often display nontrivial correlations between degrees of vertices connected by edges. Here, we analyse the statistical methods used usually to describe the degree correlation in the networks, and analytically give linear relation in the degree correlation. It provides a simple and interesting perspective on the analysis of the degree correlation in networks, which is usefully complementary to the existing methods for degree correlation in networks. Especially, the slope in the linear relation corresponds exactly to the degree correlation coefficient in networks, meaning that it can not only characterize the level of degree correlation in networks, but also reflects the speed that the average nearest neighbours' degree varies with the vertex degree. Finally, we applied our results to several real-world networks, validating the conclusions of the linear analysis of degree correlation. We hope that the work in this paper can be helpful for further understanding the degree correlation in complex networks.

Keywords. Complex network; degree correlation; Pearson coefficient.

PACS Nos 89.75.−k; 89.75.Fb

1. Introduction

The study of complex networks have attracted much attention from physics community and other interdisciplinary fields [1]. Complex systems from various fields can be investigated with the help of the representations of networks – sets of vertices connected by edges. Examples include the Internet, the WWW, the protein–protein interactions, the collaborative relationship between scientists and actors [2–9] etc. Many statistical quantities such as the degree distribution, clustering coefficient etc. were used to characterize complex networks, by which many common topological properties in the complex networks, such as the small-world property, the transitivity (or clustering), the scale-free property and the degree correlation, have been revealed.

Here, we focus on the property of degree correlation in complex networks. In recent years, many studies have shown the existence of nontrivial correlation between degrees of vertices connected by edges in real networks [10–12]. Most social networks show assortative mixing or positive correlation of degrees, implying that high- or low-degree vertices preferably connect to other nodes with similar degrees, while technological and biological networks show disassortative mixing or negative correlation of degrees, implying high-degree vertices preferably connect to low-degree vertices in the networks. The correlations in degree have important impact on the structures and functions of networks, such as the stability [13], the robustness of networks against attacks [14], the controllability of networks [15], the traffic dynamics on networks [16], the synchronization [17,18], the spreading of information or infections and other dynamic processes [19–24].

In order to understand the correlations of degree in complex networks, some statistical measures and models have been applied [11,12,25–28]. The degree correlation coefficient (DCC), a scalar quantity, is one of the widely used measures. Another widely used measure is ANND, average nearest-neighbours' degree, on vertex degree which gives a curve that varies with vertex degree. It can be characterized by suitable fitting functions. For example, Maslov and Sneppen [29] showed a power-law dependence of ANND in protein–protein interaction network and the Internet map. Ma and Szeta [25] gave a linear analysis of the total degrees of the nearest neighbours as a function of vertex degree by extending the Aboav–Wearie law to complex networks. The studies provide alternative ways to analyse the degree correlation in the networks, but it is difficult to explicitly give analytical relations between DCC and the fitting parameters.

Generally, the linear analysis in the study of complex systems should be more appreciated, thanks to its simplicity. In this paper, we shall analytically reveal the linear relation in the degree correlation, by the linear analysis of ANND as a function of vertex degree. Here, the slope and the intercept in the linear relation can be expressed easily by using the degree correlation coefficient and the first and second moments of the degree distribution of the networks. The fitting parameters are therefore more meaningful. Especially, the slope corresponds exactly to the degree correlation coefficient. So the linear analysis provides a simple and interesting perspective on the analysis of degree correlation in complex networks, which is complementary to the existing methods.

This paper is organized as follows: In §2, we introduce the concept and the statistical measures for the degree correlation in networks. In §3, we analyse the linear relation in the degree correlation. In §4, the results are applied to real-world networks. Finally, we come to our conclusion.

2. Degree correlation in networks

Consider an undirected, unweighted network of N vertices and M edges, with degree distribution p_k. The degree distribution is the probability that a randomly chosen vertex in the network will have degree k (the degree of a vertex in a network is the number of other vertices that it connects to). Many real networks show power-law degree distribution, i.e., the well-known scale-free property. If we pick up a vertex by following a randomly chosen edge in a network, the probability that the vertex will have degree k, denoted by q_k, is proportional to kp_k, rather than p_k, because of a clear preference for the edges to attach to vertices with high degree. The normalized distribution of q_k is $q_k = kp_k/\sum_j jp_j$. If the probability that a randomly selected edge in a network has two ends with degree j and k is proportional to $q_j \cdot q_k$. The networks of this type are usually considered to have no correlation of degrees, while most of real-world networks display a clear deviation from this value, implying the existence of nontrivial correlation of degrees in the networks [10–12].

Degree–degree joint probability: The correlation between degrees of vertices connected by edges can be naturally described by the degree–degree joint probability e_{jk}, the probability that one of the two ends of a randomly selected edge in a network will have a vertex of degree j and another will have a vertex of degree k. This quantity is a symmetric matrix in undirected network ($e_{jk} = e_{kj}$), $\sum_{jk} e_{jk} = 1$ and $\sum_j e_{jk} = q_k$. If e_{jk} takes the value of $q_j q_k$ in a network, then the network does not exhibit correlation between vertex degrees, called as uncorrelated networks. When there occurs (dis-) assortative correlation in the network, the value of e_{jk} will differ from $q_j q_k$. To show their difference, one can naturally define a correlation function,

$$p(j,k) = \frac{e_{jk}}{q_j q_k}. \tag{1}$$

$p(j,k) \equiv 1$ means that the degree correlation is absent in the network topology and any other value other than 1 indicates the nontrivial correlations of degree.

Average nearest-neighbours' degree (ANND): It is still a rather difficult task to directly identify the tendency of degree correlation in the whole network by e_{jk} or $p(j,k)$, due to the statistical fluctuation. A more clear indication of degree correlation is the average nearest-neighbours' degree of vertices with degree k (ANND) [12],

$$\begin{aligned}\overline{k_{nn}(k)} &= \sum_j j p_c(j|k) = \sum_j j e_{jk} / \sum_j e_{jk} \\ &= \sum_j j e_{jk}/q_k, \end{aligned} \tag{2}$$

where $p_c(j|k) = e_{jk}/\sum_j e_{jk} = e_{jk}/q_k$ is the conditional probability that an edge belonging to a vertex with degree k will connect to a vertex with degree j. When $\overline{k_{nn}(k)} = C$, ANND is independent of k, indicating the absence of degree correlation in networks under study. In general, $\overline{k_{nn}(k)}$ will increase with k in assortative mixing networks, while it will decrease with k in disassortative mixing networks. So, one can classify networks by ANND. Of course, it can give a clear but only qualitative description of the degree correlation.

Degree correlation coefficient (DCC): A more coarse-grained description of the degree correlation can be given by the degree correlation coefficient, also called the Pearson correlation coefficient [11],

$$r = \frac{1}{\sigma_q^2} \sum_{jk} jk(e_{jk} - q_j q_k), \tag{3}$$

where $\sigma_q^2 = \sum_{jk} k^2 q_k - (\sum_k k q_k)^2$ and $-1 \leq r \leq 1$. One can determine the type of correlation by the sign of the coefficient. It can also estimate the strength of correlation by the value of the coefficient. Therefore, it is more convenient to compare correlation of different networks by the degree correlation coefficient.

3. Linear analysis of degree correlation

The statistical measures for describing the degree correlation in networks are closely related. As we see, both the definitions of ANND and DCC are based on the degree–degree joint probability. Considering the definition of ANND, DCC can be rewritten as

$$r = \frac{\sum_k q_k k \overline{k_{nn}(k)} - \langle k \rangle^2}{\langle k^2 \rangle - \langle k \rangle^2}, \tag{4}$$

where $q_k = kp(k)/\sum_j jp(j)$, $\langle \cdots \rangle$ denotes the statistical average of q_k. In general, $\overline{k_{nn}(k)}$ is a monotonically increasing or decreasing function of k. In assortative networks, $\overline{k_{nn}(k)}$ is a monotonically increasing function of k. In this case, $r > 0$, and so the assortative mixing of degree is also called as *positive* correlation, especially when $\overline{k_{nn}(k)} \propto k$, $r = 1$. It corresponds to a network with perfect assortative mixing of degree. In disassortative networks, $\overline{k_{nn}(k)}$ is a monotonically decreasing function of k. In this case, $r < 0$, and so the disassortative mixing of degree is also called negative correlation. When $\overline{k_{nn}(k)} \equiv C$, $r = 0$, the degree correlation is absent. Networks of this kind are called the uncorrelated networks.

Clearly, the two measures for describing the degree correlation are consistent. However, ANND can only give a qualitative description of the degree correlation, and so the representation above only gives qualitatively the relation between DCC and ANND, not quantitatively. In literatures, ANND has been characterized by various fitting functions. Similar to the fit of degree distribution in scale-free networks, Maslov and Sneppen [29] showed a power-law fit of ANND, i.e. $\overline{k_{nn}(k)} \sim k^{\upsilon}$ in protein–protein interaction network and the Internet map. When $\upsilon > 0$, $r > 0$; when $\upsilon < 0$, $r < 0$; when $\upsilon = 0$, $r = 0$. Similar to DCC, the exponent υ of the power-law fit can also show the strength of degree correlation in networks and the tendency of the mixing of degrees (assortative or disassortative), but it is difficult to give quantitative relation between DCC and υ. Ma and Szeta [25], by extending the Aboav–Wearie law for two-dimensional cellular patterns such as soap froth, to complex networks, investigated the linear analysis of the total degrees of the nearest neighbours as a function of vertex degree. The extension of the Aboav–Wearie law provides a different perspective on the analysis of degree correlations from the power-law fit of the data, while the analytical relation between DCC and the fitting parameters was not explicitly given at least in the reference.

In the study of complex systems, the simplicity and clarity of the linear analysis are often more appreciated. And we find that for some real-world networks, such as the airline transportation network of China, the linear fit of ANND is clearly more suitable than the above methods. Specifically, we analyse the data of ANND in the networks by using weighted linear least square fitting of q_k, because each data point in ANND is often an average of more than one sample (the number of sample vertices for each point is proportional to q_k). The least square fitting method has been extensively applied to the data analysis in various fields. Given the fitting function $f(k) = Ak + B$ and the fitting weight q_k, the residual function of the linear fit of ANND can be written as

$$G = \sum_k q_k [f(k) - \overline{k_{nn}(k)}]^2 = \sum_k q_k [Ak + B - \overline{k_{nn}(k)}]^2. \tag{5}$$

According to the least square fitting method, to find the minimum of the G function, we need to solve the following equations:

$$\begin{cases} \dfrac{\partial G}{\partial A} = 2 \sum_k q_k k [Ak + B - \overline{k_{nn}(k)}] = 0 \\ \dfrac{\partial G}{\partial B} = 2 \sum_k q_k [Ak + B - \overline{k_{nn}(k)}] = 0 \end{cases} \tag{6}$$

and because of the topological constraint in network,

$$\sum_k Np(k) k \overline{k_{nn}(k)} = \sum_k Np(k) k^2 = N\overline{k^2}, \tag{7}$$

where the sum of $k\overline{k_{nn}(k)}$ over all the vertices in the network is equivalent to summing up degree k of each vertex by the times of k, or say, it is equal to the sum of the degrees of ends of all edges in the network. Further, as $q_k = kp_k/\bar{k}$, we obtain

$$\sum_k q_k\overline{k_{nn}(k)} = \sum_k q_k k = \frac{\overline{k^2}}{\bar{k}} = \langle k\rangle. \tag{8}$$

Combining eqs (6) and (8), one can obtain the regular equations

$$\begin{cases} A\langle k^2\rangle + B\langle k\rangle - \sum_k q_k k\bar{k}_{nn}(k) = 0 \\ A\langle k\rangle + B - \langle k\rangle = 0 \end{cases}. \tag{9}$$

Solving the above equations, one can obtain the expressions of the parameters A and B,

$$\begin{cases} A = \dfrac{\sum_k q_k k\overline{k_{nn}(k)} - \langle k\rangle^2}{\langle k^2\rangle - \langle k\rangle^2} \\ B = \langle k\rangle\,(1 - A) \end{cases}. \tag{10}$$

Comparing eq. (4), one can obtain

$$\begin{cases} A = r \\ B = \langle k\rangle\,(1 - r) \end{cases}. \tag{11}$$

We give the fitting parameters of ANND analytically, which can be expressed easily by using the DCC and the first and second moments of the degree distribution of the networks ($\langle k\rangle = \overline{k^2}/\bar{k}$). Especially, the slope A in the linear fit corresponds exactly to the degree correlation coefficient r, which provides an alternative way to classify networks based on the property of assortative mixing. Obviously, DCC (or the slope A) is meaningful. It is not only a number that characterizes the level of degree correlation in networks, but also denotes the slope of ANND's linear fit, reflecting the speed that the curve of ANND varies with k.

4. Application to real-world networks

Here, we apply the above results to several real-world networks.

(1) Airline network of China (ANC) [30,31]: It describes the air routes among all cities with operating airports in mainland China (excluding Taiwan, Hong Kong and Macao) from October 28, 2007 to March 29, 2008 (data from the Civil Aviation Administration of China (CAAC) (2009)). The cities are denoted by the vertices and the air routes are denoted by the links in the network.
(2) Protein interaction network in budding yeast [32]: The vertices are the proteins in budding yeast and the links denote the physical interactions discovered between proteins.
(3) Political blogs [33]: It is a network of hyperlinks between weblogs on US politics, recorded in 2005 by Adamic and Glance.
(4) High-energy theory collaborations (HEP Collaborations) [34]: A network of co-authorships between scientists posting preprints on the High-Energy Theory E-Print Archive between January 1, 1995 and December 31, 1999. The vertices denote the authors and the connections are the co-authorship in scientific papers.
(5) Geom Collaborations [35]: A collaboration network in computational geometry.
(6) DIC28 [35]: A word network from Pajek datasets.

We consider the real-world networks as undirected and unweighted networks in our study and analyse the degree correlation in the giant component of the networks (see table 1). As an example, figure 1 shows the degree–degree joint probability e_{jk} in the airline network of China (ANC) and the protein interaction

Table 1. Degree correlation in various networks. Here, we consider the giant component of the networks. N denotes the number of vertices in the networks. $\bar{k}$ is the mean degree of vertices in networks. $\overline{k^2}$ is the second moment of the degree distribution in networks. A and B are the slope and intercept of the linear fitting function of ANND.

Network	N	$\bar{k}$	$\overline{k^2}$	A	B	Correlation
Airline network of China	144	14.15	543.06	−0.43	54.68	Disassortative
Protein interaction	2224	5.94	98.99	−0.11	18.41	Disassortative
Political blogs	1222	27.36	2222.98	−0.22	99.25	Disassortative
Geom Collaborations	3621	5.23	96.23	0.17	15.32	Assortative
HEP Collaborations	5835	4.74	43.19	0.19	7.43	Assortative
DIC28	24831	5.72	54.40	0.61	3.74	Assortative

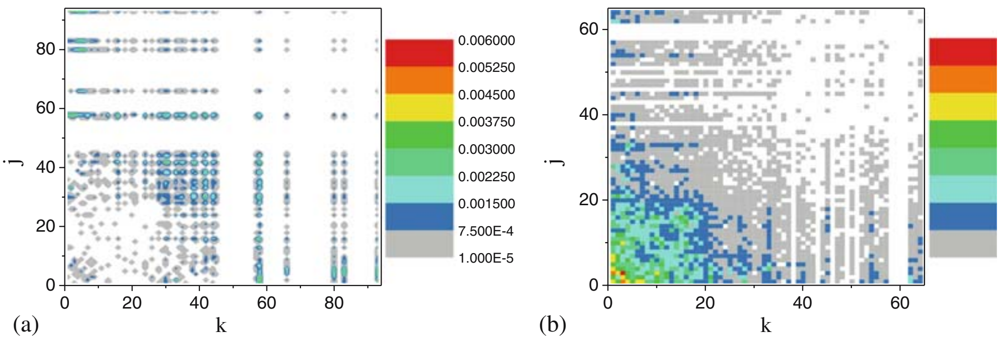

Figure 1. The degree–degree joint probability e_{jk} in (**a**) the airline network of China (ANC) and (**b**) the protein interaction network of budding yeast (PIN).

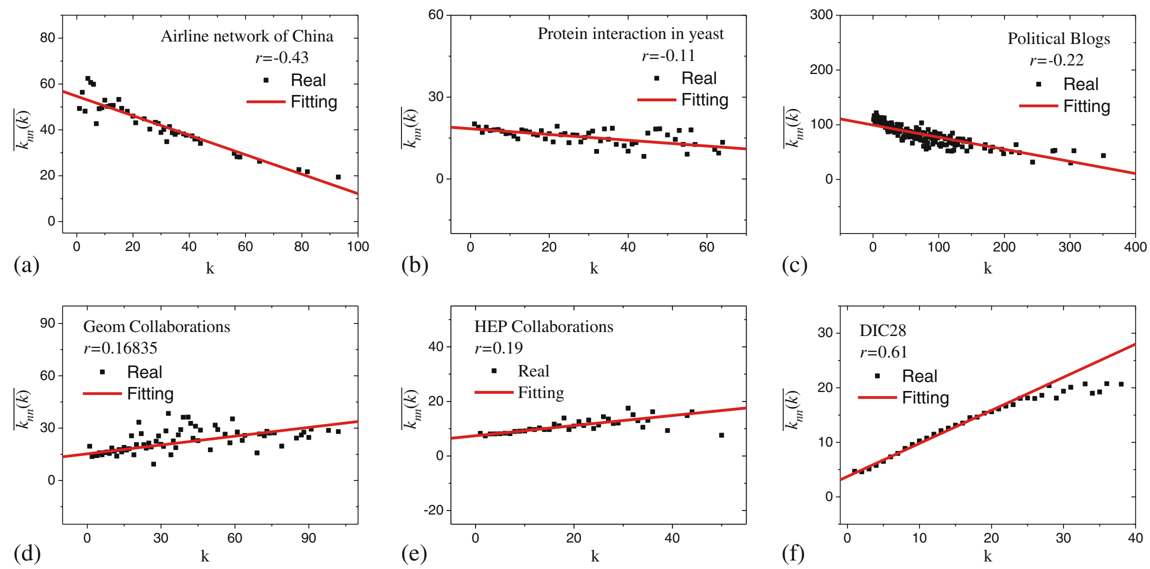

Figure 2. The data of $\overline{k_{nn}(k)}$ (ANND) as a function of k and the degree correlation coefficient for various networks. The solid lines are the linear fit of ANND in the corresponding networks. The result of the KS-test about the linear fit is 0 (i.e. acceptance) at the 5% significance level, except for Political Blogs.

network of budding yeast. Clearly, this measure cannot easily indicate the extent of degree correlation in the networks. In figure 2, we show $\overline{k_{nn}(k)}$ (ANND) as a function of k as well as the degree correlation coefficient (DCC) for various networks. We can easily identify the tendency of degree correlation in the networks by ANND or DCC. The results show that the ANND's description for degree correlation is consistent with DCCs. If ANND is an increasing function of k, DCC is positive; if ANND is a decreasing function of k, DCC is negative. Here, we fit the data of $\overline{k_{nn}(k)}$ by a linear function ($y = Ax + B$). To provide visual perception for the linear analysis of the degree correlation, we show straight lines of fitting ANND in various networks. In table 1, we record the values of the parameters A and B as well as the trend of degree correlation in the networks. As we see, the slope of the linear fit exactly corresponds to the degree correlation coefficient in the networks, perfectly showing the quantitative relation between ANND and DCC. The slope in the linear relation can reflect the speed of ANND varying with k, and can estimate the strength of degree correlation in the networks.

5. Discussion

In ref. [25], the linear analysis of the total degrees of nearest neighbours as a function of vertex degree was investigated by extending the Aboav–Wearie law to complex networks, which provides a different perspective on the degree correlations from the power-law fit. Here, we revisit the linear analysis by using weighted linear least square fitting of $k\overline{k_{nn}(k)}$ with weight p_k. We write the residual function of the linear analysis as $G = \sum_k p_k[k\overline{k_{nn}(k)} - (ak + b)]^2$, and give the analytical expressions of the fitting parameters explicitly,

$$a = \langle k \rangle - b \quad \text{and} \quad b = -r \cdot \delta_q^2 \cdot \overline{k}/(\langle k \rangle - \overline{k}), \tag{12}$$

where $\overline{k} = \sum_k kp_k$, $\overline{k^2} = \sum_k k^2 p_k$ and $\langle k \rangle = \overline{k^2}/\overline{k}$. When $r > 0$, $b < 0$; $\overline{k_{nn}(k)} = a + b/k$ will increase with k, corresponding to the assortative mixing networks. When $r < 0$, $b > 0$; $\overline{k_{nn}(k)} = a + b/k$ will decrease with k, corresponding to the disassortative networks. When $r = 0, b = 0$, $\overline{k_{nn}(k)}$ is independent of k, indicating the absence of degree correlation in networks. Clearly, the above presentations are consistent with those in ref. [25].

6. Conclusion

In this paper, we revisit the statistical measures used usually to describe the degree correlation in networks. We find that for some real-world networks, such as the airline transportation network of China, the linear fit of ANND is more suitable than other methods, and it can give some interesting results, where the parameters can be interpreted better and more meaningfully.

We give analytically the parameters of the linear fit of ANND in networks. The fitting parameters can be expressed easily by using the DCC and the first and second moments of the degree distribution of the networks. Especially, the slope in the linear fit can correspond exactly to the degree correlation coefficient in the networks, providing an alternative way to the classification of networks based on the property of degree correlation. DCC can not only characterize the level of degree correlation in networks, but also denote the slope of ANND's linear fit, reflecting the speed that the curve of ANND varies with k. The linear analysis of the degree correlation provides a simple and interesting perspective on the analysis of the degree correlation in complex networks, which is usefully complementary to the existing methods. Finally, we applied our results to analyse the degree correlation in several real-world networks, showing the consistency of several descriptions for degree correlation in networks. We hope that the work in this paper can be helpful for further analysis and understanding of the degree correlation in complex networks.

Acknowledgement

The authors are grateful to the reviewer for the comments as well as the suggestion about the KS-test. This work has been supported by the construct program of the key discipline in Hunan Province, the Scientific Research Fund of Education Department of Hunan Province (Grant Nos 14C0126, 11B128, 14C0127, 14C0112, 12C0505 and 14B024), the Project of Changsha Medical University (Grant No. KY201517), the Department of Education of Hunan Province (Grant No. 15A023), the Hunan Provincial Natural Science Foundation of China (Grant No. 2015JJ6010), the Hunan Provincial Natural Science Foundation of China (Grant No. 13JJ4045), the National Natural Science Foundation of China (Grant No. 11404178) and the National Natural Science Foundation of China (Grant No. 81300231), and partly by the Doctor Startup Project of Xiangtan University (Grant No. 10QDZ20).

References

[1] S Boccaletti, V Latora, Y Moreno, M Chavez and D U Hwang, *Phys. Rep.* **424**, 175 (2006)
[2] S Mukherjee and N Gupte, *Pramana – J. Phys.* **70**, 1109 (2008)
[3] Q Zhao, X Kong and Z Hou, *Pramana – J. Phys.* **73**, 955 (2009)
[4] L Zemanová, G Zamora-López, C Zhou and J Kurths, *Pramana – J. Phys.* **70**, 1087 (2008)
[5] L Ying and D W Ding, *Pramana – J. Phys.* **80**, 337 (2013)
[6] R Pan and S Sinha, *Pramana – J. Phys.* **71**, 331 (2008)
[7] P Chen and S Redner, *J. Inform.* **4**, 278 (2010)
[8] R Guimera and L A Nunes Amaral, *Nature* **433**, 895 (2005)
[9] S-H Zhang, X-M Ning, C Ding and X-S Zhang, *BMC Syst. Biol.* **4**, 1 (2010)
[10] M E J Newman, *Phys. Rev. E* **67**, 026126 (2003)
[11] M E J Newman, *Phys. Rev. Lett.* **89**, 208701 (2002)
[12] R Pastor-Satorras, A Vázquez and A Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001)
[13] C Friedel and R Zimmer, *BMC Bioinform.* **8**, 297 (2007)
[14] T Tanizawa, S Havlin and H E Stanley, *Phys. Rev. E* **85**, 046109 (2012)
[15] Q Miao, Z Rong, Y Tang and J Fang, *Physica A* **387**, 6225 (2008)
[16] J-T Sun, S-J Wang, Z-G Huang and Y-H Wang, *Physica A* **388**, 3244 (2009)
[17] C E La Rocca, L A Braunstein and P A Macri, *Physica A* **390**, 2840 (2011)
[18] M Chavez, D U Hwang, J Martinerie and S Boccaletti, *Phys. Rev. E* **74**, 066107 (2006)
[19] A L Pastore y Piontti, L A Braunstein and P A Macri, *Phys. Lett. A* **374**, 4658 (2010)
[20] L D Valdez, C Buono, L A Braunstein and P A Macri, *Europhys. Lett.* **96**, 38001 (2011)
[21] F Sorrentino, M di Bernardo, G H Cuéllar and S Boccaletti, *Physica D* **224**, 123 (2006)
[22] J Menche, A Valleriani and R Lipowsky, *Europhys. Lett.* **89**, 18002 (2010)

[23] J L Payne, P S Dodds and M J Eppstein, *Phys. Rev. E* **80**, 026125 (2009)
[24] M Schläpfer and L Buzna, *Phys. Rev. E* **85**, 015101 (2012)
[25] C W Ma and K Y Szeto, *Phys. Rev. E* **73**, 047101 (2006)
[26] J L Garcia-Domingo, D Juher and J Saldaña, *Physica D* **237**, 640 (2008)
[27] A Pusch, S Weber and M Porto, *Phys. Rev. E* **77**, 017101 (2008)
[28] M Ramos and C Anteneodo, *Random degree–degree correlated networks*, arXiv:1206.6266v1 (2012)
[29] S Maslov and K Sneppen, *Science* **296**, 910 (2002)
[30] http://www.caac.gov.cn/C1/200807/t20080723_17530.html, in: Civil Aviation Administration of China (CAAC) (2009)
[31] J Wang, H Mo, F Wang and F Jin, *J. Transport Geography* **19**, 712 (2011)
[32] D Bu, Y Zhao, L Cai, H Xue, X Zhu, H Lu, J Zhang, S Sun, L Ling, N Zhang, G Li and R Chen, *Nucl. Acid. Res.* **31**, 2443 (2003)
[33] L A Adamic and N Glance, The political blogosphere and the 2004 U.S. election: Divided they blog, in: *Proceedings of the 3rd International Workshop on Link Discovery* (ACM, Chicago, Illinois, 2005) pp. 36–43
[34] M E J Newman, *Proc. Natl Acad. Sci. USA* **98**, 404 (2001)
[35] B Vladimir and M Andrej, in: *Pajek datasets*, http://vlado.fmf.uni-lj.si/pub/networks/data/