Analysis and perturbation of degree correlation in complex networks

Ju Xiang¹, Ke Hu²*, Tao Hu³, Yan Zhang⁴

¹Department of Basic Medical Sciences, Changsha Medical University, Changsha 410219, Hunan, China
²Department of Physics, Xiangtan University, Xiangtan 411105, Hunan, China
³College of Science, QiLu University of Technology, Jinan 250353, Shandong, China
⁴Department of Computer Science, Changsha Medical University, Changsha 410219, Hunan, China

Abstract – Degree correlation is an important topological property common to many real-world networks. In this paper, the statistical measures for characterizing the degree correlation in networks are investigated analytically. We give an exact proof of the consistency for the statistical measures, reveal the general linear relation in the degree correlation, which provide a simple and interesting perspective on the analysis of the degree correlation in complex networks. By using the general linear analysis, we investigate the perturbation of the degree correlation in complex networks caused by the addition of few nodes and the “rich club”. The results show that the assortativity of homogeneous networks such as the Erdős-Rényi graphs is easily to be affected strongly by the simple structural changes, while it has only slight variation for heterogeneous networks with broad degree distribution such as the scale-free networks. Clearly, the homogeneous networks are more sensitive for the perturbation than the heterogeneous networks.

PACS: 89.75.Hc - Networks and genealogical trees
PACS: 89.75.Fb - Structures and organization in complex systems

* Corresponding author. Email: huke1998@aliyun.com, xiang.ju@foxmail.com
Introduction. - Complex networks provide a useful tool for investigating the topological structure and statistical properties of complex systems with networked structures, which have recently attracted much attention from physics community and other interdisciplinary fields [1]. Many complex systems in real world have been investigated by the way of networks—sets of nodes connected by edges, examples including the Internet, the WWW, the protein-protein interactions, as well as the collaboration relationship of scientists as well as actors. These networked systems are found to possess many common topological properties, such as the small world property, the transitivity or clustering, the scale-free property and so on. Here, we will study on another network property of importance—the correlation of degree in the networks.

Many real-world networks exhibit the existence of the nontrivial correlation between degrees of nodes connected by edges [2-8]. Empirical studies show that almost all social networks display the property that high- or low-degree nodes tend to connect to other nodes with similar degrees. The degree correlation of this type is often referred to as “assortative mixing” or positive correlation of degrees. While, in technological and biological networks, high-degree nodes often preferably connect to low-degree nodes. In this case, the correlation is referred to as “dissassortative mixing” or negative correlation of degrees. The correlations on degree have important influence on the topological properties of networks and may impact related problems on networks such as stability [9], the robustness of networks against attacks [10], the network controllability [11], the traffic dynamics on networks [12-13], the network synchronization [14-16], the spreading of infections and other dynamic processes [10, 16-27]. In order to characterize and understand such preference of connections in complex networks, many statistical measures and network models have been introduced and investigated [3, 4, 28-39].

In this paper, we will study analytically the statistical measures for characterizing the degree correlation in complex networks, and give some interesting results, including the exact proof of the consistency of the measures and the general linear relation in the degree correlation, which provides a simple and interesting perspective on the analysis of the degree correlation in complex networks. In some cases, the assortativity of networks can easily be affected by small structural changes. Here, we analyze the perturbation of the degree correlation in networks by the addition of few nodes and the “rich club”, in terms of the general linear relation in the degree correlation. We show that the slope of the average nearest neighbors’ degree in homogeneous networks is affected strongly by such simple structural changes, while it has only slight variation for heterogeneous networks. Naturally, the degree correlation in homogeneous networks such as Erős-Rényi (ER) graphs is clearly sensitive much more than in heterogeneous networks with broad degree distribution such as BA scale-free networks.

Analysis of degree correlation in complex networks. - The degree correlation between two nodes connected by edges can be naturally characterized 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 node of degree $j$ and another will has node of degree $k$. This quantity is a symmetric matrix in undirected network ($e_{jk} = e_{kj}$), and have $\sum_j e_{jk} = 1$ and $\sum_k e_{jk} = q_k$, where $q_k = k p_k / \sum_j k p_j$ is the probability that the end of a randomly chosen edge in a network has a node of degree $k$, while $p_k$ is the degree distribution of the network—the probability that a randomly chosen node in the network has degree $k$ (the degree of a node is the number of other nodes to which it connects in the network). If $e_{jk}$ takes the value of $q_j q_k$ in a network, the network is usually considered to have no correlation of degrees. While most of real-world networks always exhibit an obvious deviation from the value, meaning the existence of degree correlation in the networks. However, the degree-degree joint probability is easily affected by statistical fluctuations in finite networks, and it is rather difficult to identify the tendency of degree correlation in the network by the quantity. Therefore, other statistical measures may be more convenient and efficient.

Statistical measures for degree correlation. One of the widely used measures for characterizing degree correlation is the average nearest neighbors’ degree of nodes with degree $k$ (ANND) [4],

$$
\overline{k_{nn}}(k) = \sum_j j p_j(j | k),
$$

where $p_j(j | k) = e_{jk} / \sum_i e_{ik} = e_{jk} / q_k$ is the conditional probability that an edge belonging to a node with degree $k$ will connect to a node with degree $j$. This measure considers the average degree of the neighbors of a node as a function of its degree $k$, and it can provide a clear indication for the presence or absence of degree correlation in networks. When it is independent of $k$, meaning that networks under study have no clear correlation of degree. In homogeneous and uncorrelated networks $\overline{k_{nn}}(k) = k$, while it will increase with the increase of the heterogeneity of networks. In fact, $\overline{k_{nn}}(k) = k^2 / \overline{k}$ ($k = \sum j k p_j$ and $\overline{k} = \sum j k^2 p_j$) in general uncorrelated networks, which was generally used to characterize the level of heterogeneity of networks. In general correlated networks, $\overline{k_{nn}}(k)$ will increases with $k$ for assortative mixing, meaning that nodes preferentially attach to other nodes with similar degrees, while it will decrease with $k$ for dissassortative mixing, meaning that high-degree nodes preferentially attach to other low-degree nodes, so one can classify networks by the quantity. The representation above provides a plain interpretation for the origin of degree correlations. However, this quantity can give a clear but only qualitative characterization for the degree correlation in networks.

A more coarse-grained and quantitative characterization for degree correlation in networks can be given by the degree correlation coefficient (DCC) [3],

$$
r = \frac{1}{\sigma^2} \sum_{jk} j k (e_{jk} - q_j q_k),
$$

where $\sigma^2 = \sum j k^2 q_j - (\sum j k q_j)^2$. It can describe the level of degree correlation in networks by a scalar quantity. The coefficient is also called the Pearson correlation coefficient of the degrees at the two ends of edges. According to the degree-degree joint probability $e_{jk}$, it can be rewritten as
where \( < \cdots > \) indicates an average over all edges. The value of the correlation coefficient lies in the range \(-1 \leq r \leq 1\).

The results suggest that the correlation coefficient can not only describe the level of the degree correlation in networks, but also reflect the speed that the mean degree of the nearest neighbors varies with \( k \). Moreover, for a perfect assortative network, the degree correlation coefficient is equal to 1, while ANND will be a perfect linear function \( k_{\text{md}}(k) = k \). This means that only nodes with the same degree are connected with each other. For disassortative networks, \( k_{\text{md}}(k) \) is still a decreasing function of \( k \). Particularly, for perfect negative correlation networks, we have \( r=1 \) and \( k_{\text{md}}(k) = -k + 2k^2/\bar{d} \). In uncorrelated networks, \( r=0 \), while \( k_{\text{md}}(k) = k^2/\bar{d} \).

### Linear relation in degree correlation

As we know, ANND as a function of \( k \) gives a curve that varies with \( k \). It can be characterized by suitable fitting functions. For example, researchers showed a power-law dependence of ANND on degree [4, 5]. Ma and Szeto extended the Aboav-Weaire law to the analysis of degree correlation in complex networks [28]. In the study of complex systems, the linear analysis is often more appreciated, due to the simplicity and clarity of it. Here, we show the results of interest can be obtained by using the linear fitting function \( (ak+b) \). Substituting \( k_{\text{md}}(k) \) in the equation (4) by the linear function, and using the constraint equation (6), we can obtain,

\[
a=r \quad \text{and} \quad b=\frac{k^2}{\bar{d}}(1-r),
\]

where \( \bar{d} = \sum k p_k \) and \( \bar{k} = \sum k^2 p_k \). Clearly, the slope \( a \) in the linear relation corresponds to the correlation coefficient \( r \).

Application to the airline network of China. There exist some real-world networks whose ANND may be more suitably characterized by the linear relation. In Fig. 1 (a), we show the airline network of China (ANC) [40, 41] where the cities are denoted by nodes and the air routes are denoted by links, and the cumulative degree distribution of the network. In the network, a few busy cities have a large number of air routes, dominating the transportation system, the number of routes of
Assortative, while Net B becomes assortative, the distribution \( P(k) \) of the number of nodes and the degree distribution \( K_{nn}(k) \) becomes an increasing function of \( k \) on the whole. Naturally, \( r < 0 \) for Net A, while \( r > 0 \) for Net B, according to the linear relation between \( r \) and ANND. Moreover, we also notice that both the statistical measures for degree correlation seem to be too sensitive in such (quasi-) homogeneous networks.

Effect of structural variation on degree correlation. – Here, we investigate the perturbation of the degree correlation in complex networks caused by the addition of few nodes and the “rich club”.

Effect of addition of node and edge. In terms of the linear relation between ANND and DCC, DCC can characterize the level of degree correlation in whole networks, but can also reflect the monotonicity of ANND and the speed that ANND varies with \( k \) as a whole. Inversely, by ANND, one can also learn about the type of the correlations and the strength of such correlations. In Ref [42], Estrada shows that the assortativity of networks can be affected by simple structural changes, and gives an interesting explanation for the phenomena. In Fig. 2, Net A is disassortative, while Net B becomes assortative, though the two networks have only simple structural difference caused by the addition of one node and one edge. We can see that ANND in Net A is a decreasing function of \( k \), while ANND in Net B becomes an increasing function of \( k \) on the whole.

FIG. 1: (Color Online) (a) The airline network of China (ANC) and the cumulative degree distribution of the network. The (undirected and un-weighted) network 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)) [40]. The cities are denoted by nodes and the air routes are denoted by links in the network. (b) The data of \( K_{nn}(k) \) (ANND) as a function of \( k \) and the degree correlation coefficient in the network. The blue solid line is the linear fitting of ANND in the network.

FIG. 2: (Left) Network (Net A) constructed by linking a regular graph and a single node, and \( K_{nn}(k) \) (ANND) and the degree distribution \( P(k) \) of the network. This network is disassortative in terms of \( r \) and/or ANND. (Right) Network (Net B) which has more nodes and edges than Net A by one node and one edge, and \( K_{nn}(k) \) (ANND) and the degree distribution \( P(k) \) of the network. This network is assortative in terms of \( r \) and/or ANND. The nodes in the two networks have similar degrees.

FIG. 3: (Left) For Erdös-Rényi (ER) network with 5000 nodes and 30000 edges [43], we show the ANND \( K_{nn}(k) \) and the degree distribution \( P(k) \) of the homogeneous network with “rich club” (0.5% nodes), compared to the results of the network without “rich club”. The original ER network is uncorrelated in degree in terms of \( r \) and/or ANND. (Right) For BA scale-free network where the number of nodes and the mean degree are the same as in the ER networks [44], we also show the ANND and the degree distribution of the heterogeneous network with “rich club” (0.5% nodes), compared to the results of the network without “rich club”.

4
Effect of Rich club. Here, we choose the top 0.5% of nodes with the highest degrees as the set of rich nodes in a network and manipulate the connections among the rich nodes: (1) remove the edges among the rich nodes, so there is no rich-club in the network; (2) make the rich nodes fully connected to each other, so they form a fully connected sub-graph. The main topological structures in the networks are the same for the two types of networks above except for the connection pattern among the rich nodes. The rich club can control the assortativity in some networks [45], but what determines the contrarility of the rich club for the assortativity in the networks? Fig. 3 shows two typical examples of the rich club affecting the degree correlation. The rich club can change more strongly ANND in the ER network, leading to the large change of the slope (i.e. DCC) of the linear fit of it. From the absence to presence of the “rich club”, DCC in the ER network changes from 0.00 to 0.53, while, for the BA scale-free network, it only changes from -0.08 to 0.04. Clearly, the “rich club” can affect more strongly the assortativity of the ER network.

In the above networks, there is the same number of nodes as well as the same mean degree, then what does lead to the difference? In fact, the two networks above have very different topological features, especially the degree distribution. The “rich club” can more strongly affect the degree distribution of the ER network (see Fig. 3). The degree distribution is very narrow for the ER network, while it is very broad for the BA scale-free network, following the power-law distribution. The range of the degrees in a network is closely related to the contrarility of the rich club for the degree correlation, because, in terms of the linear analysis of ANND, the larger the range of degrees in the networks is, the more difficulty the rich club changes the slope of ANND (i.e. DCC). This also suggest that the (quasi-) homogenous networks (with narrow degree distribution) such as ER are very sensitive to “rich club”, while the heterogeneous networks (with broad degree distribution) such as BA are insensitive.

In Table I, we further list the results of 9 different un-directed networks: 5 real-world networks and 4 model networks, arranged with Δr increasing. $k_{max}$ denotes the upper bounds of degrees in the networks, which can largely reflect the range of degrees in the networks. As expected, on the whole, the smaller $k_{max}$, the larger Δr. Based on the values of Δr or $k_{max}$, the networks can be classified into three different groups. In the first group (SW, PG and ER) with small values of $k_{max}$ (<100), the rich club strongly changes DCC of the networks. In the second group (EPA, Cond and BA) where the values of $k_{max}$ are between 100 and 1000, the rich club has slight effect on DCC. In the third group (AS, PFP and BOOK) with large values of $k_{max}$ (>1000), the rich club hardly changes DCC of the networks.

The rich club manipulates the connections of a small proportion of rich nodes, affecting the mean degrees of the rich nodes. The degrees of the rich nodes are also crucial to the changes of DCC. On the whole, the larger the mean degree of the rich nodes is, the smaller the changes of DCC is (see Table I and Fig. 4). In terms of the relation between the linear analysis of ANND and DCC, we can give a simple expectation for the increase of DCC caused by the rich club. The mean degrees of the nearest neighbors of most nodes and the rich nodes in the networks can be estimated by

\[
\overline{k_{m}(k)} = b + r \cdot \overline{k}.
\]

and,

\[
\overline{k_{m}(k_R)} = \frac{(b + r \cdot k_e)k_e + (N_r - 1)k_R}{k_R},
\]

where $b$ is the intercept in equation (7), $r$ is the degree correlation in the original network, $\overline{k}$ is the mean degree of the whole networks $N_r$ is the number of the rich nodes, while $k_e$ and $k_R$ are the mean degrees of the rich nodes in the networks without rich club and with rich club ($k_R \approx k_e + N_r - 1$). Then the increase of DCC caused by the rich club can be written as

\[
\Delta r = \frac{\overline{k_{m}(k_R)} - \overline{k_{m}(\overline{k})}}{\overline{k_R} - \overline{k}} - r.
\]

Fig. 4 shows that the overall trend of the expected values of $\Delta r$ is consistent with the real values of it, though this is only a simple linear evaluation for the degree correlation.
nodes in the networks with rich-club.

| Network | N  | $\bar{k}$ | $k_{max}$ | $r$   | $r'$   | $\Delta r$ | $k_k$ |
|---------|----|----------|----------|-------|-------|------------|------|
| SW      | 5000 | 6.0     | 15.5     | 0.00  | 0.65  | 0.65       | 34.1 |
| PG      | 4941 | 2.7     | 19.0     | -0.01 | 0.60  | 0.61       | 34.2 |
| ER      | 5000 | 6.0     | 16.0     | 0.00  | 0.53  | 0.53       | 38.7 |
| Cond    | 16726 | 5.7    | 107.0    | 0.17  | 0.32  | 0.15       | 94.7 |
| EPA     | 4772 | 3.7     | 175.0    | -0.31 | -0.15 | 0.16       | 128.4|
| BA      | 5000 | 6.0     | 218.6    | -0.08 | 0.04  | 0.12       | 150.4|
| AS      | 5375 | 3.9     | 1193     | -0.19 | -0.19 | 0.00       | 216.6|
| PFP     | 5000 | 6       | 1258.8   | -0.25 | -0.24 | 0.01       | 354.0|
| BOOK    | 7724 | 11.4    | 2568     | -0.24 | -0.24 | 0.00       | 628.5|

**Conclusion and discussion.** - We analyzed the statistical measures for characterizing the degree correlation in networks, gave the exact proof of the consistency of the measures, and then exhibited the general linear relation in the degree correlation, which provides a simple and interesting perspective on the analysis of the degree correlation in complex networks. We showed that the degree correlation coefficient corresponds exactly to the slope of the linear fit of ANND, meaning that the degree correlation coefficient can not only characterize the level of the degree correlation in network, but can also reflect the speed that ANND varies with $k$. And then, as an exemplification for the results above, we analyzed the linear degree correlation in the airline network of China.

In some cases, the assortativity of networks can easily be affected by small structural changes. Here, we analyze the perturbation of the degree correlation in networks caused by the addition of few nodes and the “rich club”, in terms of the general linear relation in the degree correlation. The slope of the average nearest neighbors’ degree in homogeneous networks such as Erdős-Rényi (ER) graphs is affected strongly by the simple structural changes, while it has only a relatively slight variation for heterogeneous networks such as BA scale-free networks. According to the linear relation between $r$ and ANND, the degree correlation in homogeneous networks is naturally sensitive much more than in heterogeneous networks with broad degree distribution. But the two statistical measures seem too sensitive in homogeneous networks.

To our knowledge, the concept of degree correlation is used to reveal the statistical feature in heterogeneous networks. Whether it should be applied to (quasi-)homogeneous networks (which have narrow degree distribution). How to understand the results that such statistical quantities as DCC and ANND generate in the networks. The stability of the two statistical quantities is related to heterogeneity of networks. Particularly, as we see, simple structural changes in homogeneous networks can strongly influence the correlation strength in the networks of this type indicated by DCC and/or ANND. When one applies these statistical quantities to the analysis of the degree correlation in complex networks, maybe the heterogeneity of networks should also be an important factor to be considered. We will discuss this topic in depth in future work. Finally, we hope that the work can help to further understand the property of the degree correlation in networks.

**Acknowledgement.** - We sincerely appreciate the careful and professional review of the referees as well as the valuable comments for the improvement of the paper. This work has been supported by the Scientific Research Fund of Education Department of Hunan Province (Grant No. 14C0126), the Scientific Research Fund of Education Department of Hunan Province (Grant No. 11B128), the Scientific Research Fund of Education Department of Hunan Province (Grant No. 14C0127) and Natural Science Foundation of Hunan Province (Grant No. 13JJ4045), and partly by the Doctor Startup Project of Xiangtan University (Grant No. 10QD2Z0).

**REFERENCES**

[1] S. Boccaletti et al., Phys. Rep. 424, 175 (2006).
[2] M. E. J. Newman, Phys. Rev. E 67, 026126 (2003).
[3] M. E. J. Newman, Phys. Rev. Lett. 89, 208701 (2002).
[4] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. 87, 258701 (2001).
[5] S. Maslov and K. Sneppen, Science 296, 910 (2002).
[6] S. Abe and N. Suzuki, Europhys. Lett. 99, 39001 (2012).
[7] J. Li and J. Zhou, Physica A 380, 629 (2007).
[8] S. Abe and N. Suzuki, Phys. Rev. E 74, 026113 (2006).
[9] C. Friedel and R. Zimmer, BMC Bioinformatics 8, 297 (2007).
[10] T. Tanizawa, S. Havlin, and H. E. Stanley, Phys. Rev. E 85, 046109 (2012).
[11] Q. Miao et al., Physica A 387, 6225 (2008).
[12] J.-T. Sun et al., Physica A 388, 3244.
[13] Y.-h. Xue et al., Phys. Rev. E 81, 037101 (2010).
[14] C. E. La Rocca, L. A. Braunstein, and P. A. Macri, Physica A 390, 2840 (2011).
[15] M. Chavez et al., Phys. Rev. E 74, 066107 (2006).
[16] F. Sorrentino et al., Physica D: Nonlinear Phenomena 224, 123 (2006).
[17] A. L. Pastore y Piontti, L. A. Braunstein, and P. A. Macri, Physics Letters A 374, 4658 (2010).
[18] L. D. Valdez et al., Europhys. Lett. 96, 38001 (2011).
[19] J. Menche, A. Valleriani, and R. Lipowsky, Europhys. Lett. 89, 18002 (2010).
[20] J. L. Payne, P. S. Dodds, and M. J. Eppstein, Phys. Rev. E 80, 026125 (2009).
[21] M. Schläpfer and L. Buzna, Phys. Rev. E 85, 015101 (2012).
[22] V. Nicosia, F. Bagnoli, and V. Latora, Europhys. Lett. 94, 68009 (2011).
[23] Y. Yang, X. Li, and Z. Rong, Europhys. Lett. 89, 18006 (2010).
[24] Y. Yin et al., Physica Scripta 76, 606 (2007).
[25] S. Bradde and G. Bianconi, J. Stat. Mech. 2009, P07028 (2009).
[26] A. Vazquez, Phys. Rev. E 74, 056101 (2006).
[27] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. Lett. 90, 028701 (2003).
[28] C. W. Ma and K. Y. Szeto, Phys. Rev. E 73, 047101 (2006).
[29] M. Ramos, and C. Anteneodo, arXiv:1206.6266v1 (2012).
[30] M. H. Li, S. G. Guan, and C.-H. Lai, Europhys. Lett. 95, 58004 (2011).
[31] M. Catanzaro, G. Caldarelli, and L. Pietronero, Phys. Rev. E 73, 026125 (2006).
[32] C. J. Paley, S. N. Taraskin, and S. R. Elliott, Phys. Rev. E 81, 041912 (2010).
[33] M. Small et al., Phys. Rev. E 77, 066112 (2008).
[34] W.-X. Wang et al., Phys. Rev. E 73, 016133 (2006).
[35] R. Xulvi-Brunet, and I. M. Sokolov, Phys. Rev. E 70, 066102 (2004).
[36] S. Johnson et al., Phys. Rev. Lett. 104, 108702 (2010).
[37] J. L. Garcia-Domingo, D. Juher, and J. Saldaña, Physica D: Nonlinear Phenomena 237, 640 (2008).
[38] A. Pusch, S. Weber, and M. Porto, Phys. Rev. E 77, 017101 (2008).
[39] A. Srivastava et al., Phys. Rev. E 86, 036106 (2012).
[40] (Civil Aviation Administration of China (CAAC), 2009.
[41] J. Wang et al., Journal of Transport Geography 19, 712 (2011).
[42] E. Estrada, Phys. Rev. E 84, 047101 (2011).
[43] P. Erdős, and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci., Ser. A 5, 17 (1960).
[44] A.-L. Barabási, and R. Albert, Science 286, 509 (1999).
[45] X.-K. Xu, J. Zhang, and M. Small, Phys. Rev. E 82, 046117 (2010).