

Linear relation on the correlation in complex networks

C. W. Ma and K. Y. Szeto*

Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Hong Kong SAR, China
(Received 17 December 2005; published 3 April 2006)

Correlation in complex networks follows a linear relation between the degree of a node and the total degrees of its neighbors for six different classes of real networks. This general linear relation is an extension of the Aboav-Weaire law in two-dimensional cellular structures and provides a simple and different perspective on the correlation in complex networks, which is complementary to an existing description using Pearson correlation coefficients and a power law fit. Analytical expression for this linear relation for three standard models of complex networks: the Erdos-Renyi, Watts-Strogatz, and Barabasi-Albert networks is provided. The slope and intercept of this linear relation are described by a single parameter a together with the first and second moment of the degree distribution of the network. The assortivity of the network can be related to the sign of the intercept.

DOI: 10.1103/PhysRevE.73.047101

PACS number(s): 89.75.Fb, 89.75.Da, 89.75.Kd

Complex networks are a convenient model for studying the topological and structural property of complex systems [1,2]. It consists of nodes which interact among themselves via their connections. Mathematical quantities such as the degree distribution, cluster coefficient, and the average shortest path length are the standard properties in a preliminary characterization of networks. In this paper, we focus on the neighbor connectivity that relates to the degree correlation among the nodes. By extending the Aboav-Weaire Law [3,4], which was well studied for two-dimensional cellular patterns such as soap froth [5,6], we find that it is also a good measure to describe the correlation in a wide variety of real and artificial networks in higher dimensions. The complex networks that we have checked are the neural networks [7], food webs [8], word co-occurrence [9,10], scientist collaboration [11], internet [12], and yeast protein interaction [13]. We also report here the Aboav's parameters for several standard models: the Erdos-Renyi [14], Watts-Strogatz [15], and Barabasi-Albert networks [16]. We find that the results on the assortivity of the network using the Aboav-Weaire law are consistent with conclusions based on the analysis of the Pearson correlation coefficient.

Originally, the Aboav-Weaire law was discovered from the empirical analysis of two-dimensional cellular structures in metal grains and later extended to a variety of cellular structures in two and three dimensions. The findings of Aboav provide a linear relation between the total degrees of nearest neighbor $nM(n)$ with the degree of the cell $nM(n) = 5n + 8$, where $M(n)$ is the mean of number of edges of neighboring cells surrounding a cell with n edges [3]. Weaire generalizes Aboav's observation and restates this observation in terms of the variance σ^2 of the degree distribution of the cellular network [4]:

$$M(n) = A + \frac{B}{n};$$

with

*Email address: phszeto@ust.hk

$$A = 6 - a' + \frac{b\sigma^2}{6},$$

$$B = 6a' + (1 - b)\sigma^2. \quad (1)$$

This form for the expression of $M(n)$ is usually tested empirically by a plot of $nM(n)$ vs n , which should be linear with slope A and intercept B . As the Aboav-Weaire law can be understood as a statement on the topological correlation of the cellular network, we attempt to generalize this empirical law in two-dimensional networks to high dimensional complex network, with numerical testing of its validity as well as analytical results on some common complex network such as Erdos-Renyi, small world, and scale free networks.

Before we analyze the linearity of $nM(n)$ vs n , we would like to mention other more commonly used descriptions of correlation. The first one is the Pearson correlation coefficient r , which was applied by Newman [17] for networks

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

with

$$\sigma_q^2 = \sum_k k^2 q_k - \left[\sum_k k q_k \right]^2. \quad (2)$$

Here e_{jk} is the joint probability distribution of an edge connecting two nodes which have degree j and k , respectively, while the remaining degree $q_k = \sum_j e_{jk}$ is the degree distribution of a node connected by an edge of another node of degree k . Note that $-1 < r < 1$. Many networks in nature show "assortative mixing" [18–20], the term assortative mixing means that there is a preference in network that a high-degree node connecting to another high-degree node. In contrast, there is another sort of degree correlation in a network call "Disassortative mixing," i.e., a high-degree node prefers to attach to low-degree node. If $r > 0$, the network exhibits a assortative mixing phenomenon. Examples of an assortative mixing network can be found in a coauthorship network, film actor collaborations, etc. For $r < 0$, we have disassortative

mixing, such as the internet network, protein network, food web, etc.

Another commonly used description [18] is to consider the mean degree of the nearest neighbor of a node with degree n , denoted by $k_{nn}(n)$, which is exactly the quantity $M(n)$ we consider in Aboav's empirical law. From the study of the network of interacting proteins and the internet, Maslov and Sneppen [18] observed that the mean degree of nearest neighbors obeys a power law $k_{nn}(n) \sim n^{-l}$, with $l=0.6$ and $l=0.5$, respectively, for protein networks and the internet. These works provide an alternate way to describe the correlation between nearest neighbors in a complex network.

Here, we will start our data analysis by checking the validity of the Aboav-Weaire law, $k_{nn}(n)=M(n)=A+(B/n)$. What we find is that the data analysis using the AW law to fit $M(n)$ is equally good and often better than the fitting of $M(n)$ by a power law. There are three basic reasons for our advocate of fitting $nM(n)$ vs n by a linear relation. The first one is that the slope A and intercept B can all be expressed in terms of the Aboav parameter a and the mean and variance of the degree distribution of the network. This single parameter description for the correlation is similar in its simplicity when compared to the Pearson correlation parameter, which is harder for the experimental analysis. The second reason, which is more important theoretically, is that for the standard networks commonly discussed in the literature, such as the Erdos-Renyi, Watts-Strogatz, and Barabasi-Albert networks, we have succeeded in computing the Aboav parameter analytically. Thus, we have provided a theoretical framework for further discussion of the modeling of complex networks from the point of view of the correlation. The third reason is that the classification of a network being assortative mixing or disassortative mixing can be simply inspected from the intercept B in our data analysis. This is an alternative way to the classification of the assortative mixing property of networks by the Pearson coefficient. After all, we feel that the simplicity in the linearity of $nM(n)$ vs n should be more appreciated in the network literature.

First we show that there is a relation between A and B , due to the topological constraint of the network: Let us sum $nM(n)$ for all nodes in the network. This is equivalent to sum up the degree of each node (of degree n) by the times of that node being connected (which is also n), so that this sum also gives the second moment of $NP(n)$

$$\int nM(n)NP(n)dn = \int n^2NP(n)dn = N\langle n^2 \rangle.$$

This result gives the relation between A and B when we assume the Aboav-Weaire law of $M(n)$

$$\int nM(n)P(n)dn = \int n\left(A + \frac{B}{n}\right)P(n)dn = \langle n^2 \rangle \Rightarrow A\langle n \rangle + B = \langle n^2 \rangle \quad (3)$$

To relate to the conventional notation in the Aboav Weaire law, we define $a=\langle n \rangle - A$, where the parameter a can be related to the parameter a' in Eq. (2) as $a=a' - b\sigma^2/6$. We then deduce that $B=\sigma^2+a\langle n \rangle$. When $B>0$, or $a>\sigma^2/\langle n \rangle$, then

$M(n)=A+(B/n)$ will decrease with n , implying that high-degree nodes like to connect to low-degree nodes, and vice versa. This corresponds to the disassortative mixing property. On the other hand, if $B<0$, or $a<\sigma^2/\langle n \rangle$, the higher-degree nodes will prefer connecting to other high-degree nodes, corresponding to the assortative mixing property. When $B=0$, or $a=\sigma^2/\langle n \rangle$, we have the special case where the nearest neighbor connectivity is independent with the degree of the nodes, implying no correlation among the degrees of nodes and its neighbors. We call such networks exhibiting no assortative property.

After explaining the basic topological relation, we investigate both numerically and analytically the validity of the Aboav-Weaire law in several well-studied artificial network models. Our analysis include the Erdos and Renyi random network (ER model), the Watts and Strogatz small world network model (WS model), and the scale free network of Barabasi and Albert (BA model). The analytical results are listed in Table I, in which numerically we have verified the validity of the Aboav-Weaire Law for these models. One should note that for the BA model, there are two sets of nodes, the set of offspring and the set of the ancestors. There is no predefined mechanism of generating the set of ancient nodes in the original BA model. Here we assume that there are m_0 ancient nodes which are fully interconnected with an average degree $\langle n \rangle = m_0 - 1$. As the BA network grows from the ancient set, a new node with a degree m ($1 \leq m \leq m_0$) is

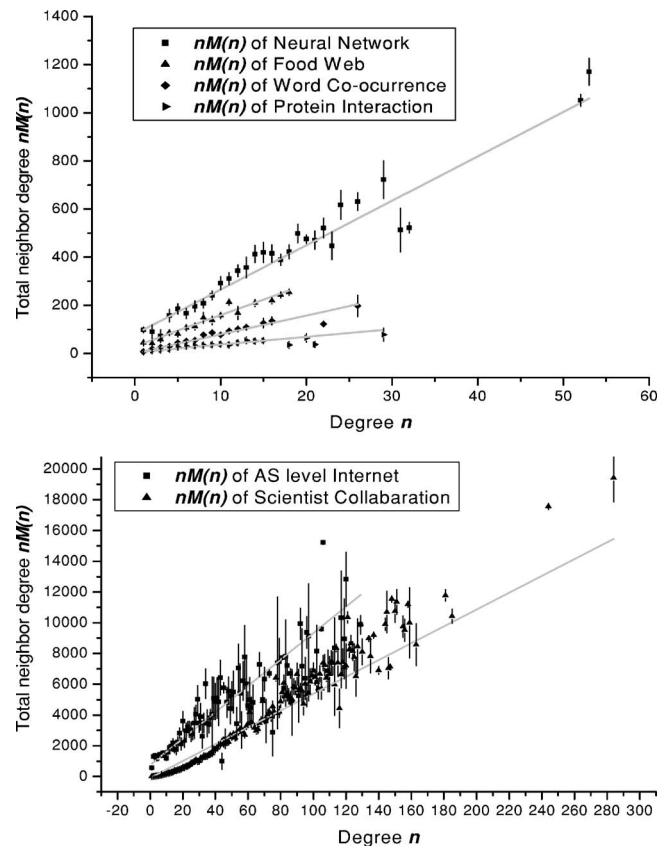


FIG. 1. The linear fit of $nM(n)$ vs n for (a) neural network, food web, word co-occurrence, and protein interaction network. (b) AS level internet and scientist collaboration.

TABLE I. Correlation in different artificial networks.

Artificial network	A	B	a	Mixing Property
ER model	$\langle n \rangle + 1$	0	-1	No Assortative
WS model	$\langle n \rangle$	σ^2	0	Disassortative
BA model (offspring)	$\frac{\langle n \rangle}{4} \ln\left(\frac{N}{\langle n \rangle + 2}\right) + \langle n \rangle$	0	$-\left\{\frac{\langle n \rangle}{4} \ln\left(\frac{N}{\langle n \rangle + 2}\right)\right\}$	No Assortative

added into the network at each time step. By choice we choose $m = \frac{1}{2}\langle n \rangle = \frac{1}{2}(m_0 - 1)$ so that the average degree $\langle n \rangle$ is always $2m$. In Table I, we have listed the analytical result for the offspring nodes only, which should be a good approximation for the entire set if the total number of nodes N is large so that the set of ancestors only create small corrections. However, we must mention that the sum rule result on the relation between A and B in the case of the offspring in the BA model does not apply, since the relation between A and B in the Aboav law rely on a sum rule that includes all N nodes in the network.

Now, it is of great interest to see if the Aboav-Weaire law also applies in real world networks. We have investigated six different networks.

(1) Food web [8]: It describes the relation of predators and preys of the Ythan river in Newburgh. The food web consists of 92 species. Based on a field observation, 409 links are made.

(2) Neural network of *C. Elegans* [7]: The nodes are the neurons of *C. Elegans* and the links represent the connection of synapse among neurons.

(3) Word co-occurrence [9,10]: A language network established from the English version article of European human

rights. Each word that appeared in the paper is treated as a node in network. The connection among the node is based on the co-occurrence of the words of the sentences.

(4) Yeast's protein interaction [13]: The proteins in yeast *S. Cerevisiae* are nodes connected by identified undirected physical interactions

(5) Autonomous system level of internet [11] (AS level internet): The internet can be decomposed into separated subsets in different levels. In the autonomous level, each domain is represented by a node of networks and each link is an interdomain connection.

(6) Author collaboration in scientific papers [12]: The nodes are authors and the connections refer to the joint authorship in a scientific paper.

These real networks are considered as undirected networks in our analysis of the Aboav-Weaire law, though in reality a more accurate description should take into account the direction of the links. The results are shown in Table II and Fig. 1. In Table II, the parameters A and B are obtained by weighted linear least square fitting. Since each data point in $nM(n)$ usually contains more than one sample node, each data for a given n will have its variance taken into account in our linear fit. One can also show that the Aboav parameters

TABLE II. Correlation in different real networks. Here N is the number of nodes in the network, $\langle n \rangle$ is the mean degree, σ is the standard deviation of degree distribution; A is the slope and B the intercept in the linear fit of $nM(n)$ vs n . The generalized Aboav parameter for complex networks is defined by $a \equiv \langle n \rangle - A$ and the Pearson correlation coefficient r is also tabulated for reference.

Network	Size N	Average degree $\langle n \rangle$	Standard deviation of degree σ	Slope A	Intercept B	Aboav parameter a	Pearson correlation r
Food web in Ythan river	85	7.78	7.13	12.9±0.6	31±3	-5.1±0.6	-0.28
Neural network	219	15.16	9.35	18.5±0.7	81±13	-3.3±0.7	-0.16
Word co- occurrence network	444	4.63	11.64	7.8±0.5	2±2	-3.2±0.5	-0.05
Yeast's protein-protein interaction	2115	2.12	9.29	3.2±0.3	5.3±0.6	-1.1±0.3	-0.51
AS Internet	11405	3.15	5.13	86.0±1.0	753±87	-83.0±1.0	-0.16
Scientist collaboration	52909	9.27	15.37	55.0±1.0	-77±7	-45.0±1.0	0.36

can be related to the Pearson coefficient in the following relation:

$$r = \frac{1}{\sigma_q^2} \left(\frac{A\langle n^2 \rangle + B\langle n \rangle}{\langle n \rangle} - \left[\sum_j j q_j \right]^2 \right). \quad (4)$$

In summary, we have shown the validity of the Aboav-Wearie description of the neighbor correlation in a wide class of real networks, as well as in three standard models of complex networks, the ER, WS, and BA networks. The linearity of the total degrees of the neighbors [$nM(n)$] of a node with degree n can be summarized by the single parameter a , which is shown to be related to the Pearson correlation coefficients. The simplicity of this law and its generality in

complex networks in high dimensions make it a very convenient tool for the topological classification of networks, here specifically to the question of the assortivity of the network. This has been shown to be related to the intercept B . Our extension of the Aboav empirical analysis for two-dimensional cellular patterns to complex networks provides a different perspective on correlations from the power law fitting of the data.

ACKNOWLEDGMENTS

K.Y. Szeto acknowledges the support of CERG Grant No. 603203 and DAG Grant No. 04-05/SC32.

-
- [1] S. Bornholdt and H. G. Schuster, *Handbook of Graphs and Networks* (Wiley-VCH, Weinheim, 2003).
- [2] R. Albert and A. L. Barabasi, *Rev. Mod. Phys.* **74**, 47 (2002).
- [3] D. A. Aboav, *Metallography* **3**, 383 (1970); **13**, 43 (1980).
- [4] D. Weaire, *Metallography* **7**, 157 (1974).
- [5] K. Y. Szeto, Xiujun Fu, and W. Y. Tam, *Phys. Rev. Lett.* **88**, 138302 (2002).
- [6] K. Y. Szeto, T. Aste, and W. Y. Tam, *Phys. Rev. E* **58** 2656 (1998).
- [7] J. G. White, J. N. Thompson, and S. Brenner, *Philos. Trans. R. Soc. London* **314**, 1 (1986).
- [8] S. J. Hall and D. Raffaelli, *J. Anim. Ecol.* **60**, 823 (1991).
- [9] C. W. Ma, MPhil. thesis, Department of Physics, HKUST library, 2005 (unpublished).
- [10] R. F. I. Cancho and R. V. Sole, *Proc. R. Soc. London, Ser. B* **268**, 2261 (2001); S. N. Dorogovtsev and J. F. F. Mendes, *ibid.* **268**, 2603 (2001).
- [11] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, *The Origin of Power Laws in Internet Topologies Revisited*, Proceedings of the 21st Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Computer Society, Los Alamitos, CA (2002).
- [12] M. E. J. Newman, *Phys. Rev. E* **64**, 016132 (2001).
- [13] H. Jeong, S. P. Mason, A.-L. Barabasi, and Z. N. Oltvai, *Nature (London)* **411**, 41 (2001).
- [14] P. Erdos and A. Renyi, *Publ. Math. (Debrecen)* **6**, 290 (1959).
- [15] D. J. Watts and S. H. Strogatz, *Nature (London)* **393**, 440 (1998); S. H. Strogatz and D. J. Watts, *Science* **296**, 1302 (2002).
- [16] A. L. Barabási and R. Albert, *Science* **286**, 509 (1999).
- [17] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [18] S. Maslov and K. Sneppen, *Science* **296**, 910 (2002).
- [19] I. Garfinkel and D. A. Gleit, *Journal of Population Economics* **15**, 417 (2002).
- [20] M. E. J. Newman, *Phys. Rev. E* **67**, 026126 (2003).