◆◆ Scientific
◆◆ Research

An Analytical Approach for Degree Correlations in Complex Network

Kosuke Takagi

Uwado-shinmachi 5-4, Kawagoe-shi, Saitamaken 3500817, Japan
Email: coutakagi@mse.biglobe.ne.jp

ABSTRACT

We investigate correlations between neighbor degrees in the scale-free network. According to the empirical studies, it is known that the degree correlations exhibit nontrivial statistical behaviors. With using an analytical approach, we show that the scale-freeness and one of statistical laws for degree correlations can be reproduced consistently in a unified framework. Our result would have its importance in understanding the mechanisms which generate the complex network.

Keywords: Scale-Free Network; Degree Distribution; Degree Correlations; Power Law

1. Introduction

It is known that a diversity of complex networks including sociological, technological, and biological ones exhibit the scale-freeness [1-13]. These results pose us a problem about the origin of this feature and about mechanisms which produce such organized behavior in complex networks. Naturally it is considered that the complex networks are generated through processes in which nodes are correlated to each other. The experimental data which exhibits the organized and hierarchical structures enhances the importance of the node correlations in the complex networks [14-18]. Indeed the model based studies have shown that, in order to reproduce the network structure in the real world, additional ingredients other than the simple rule such as the preferential attachment are required in the simulation [11-18].

Recent empirical studies have revealed that there exist ordered structures of node correlations in real world complex networks. One example of these structures would be given by fractality which characterize geometrical structures of various complex systems, in which they show the self-similarity on all length scales [14,15]. On the other hand, more primitive relation between nodes would be represented by a joint probability $P(k_1, k_2)$ for two neighbor nodes of degree k_1 and k_2 connected by an edge. In this paper, we investigate this basic statistics, the degree-degree correlation in the complex network. One characteristic feature of $P(k_1, k_2)$ can be quantified by $\langle k_2 \rangle$ for each fixed k_1, the average of the neighbor degrees for a given value of k_1. It has been reported that the $\langle k_2 \rangle$ profile is fitted with a power law

$$\langle k_2 \rangle \sim k_1^{-\gamma_p} \tag{1}$$

with a constant γ_p for the interaction and regulatory networks of proteins [16]. The same tendency is also confirmed with the Internet, another typical example of the complex networks [17].

The ubiquity of scale-free networks in the real world is one of the fundamental issues in the complex network studies. It would suggest that there exist common mechanisms which underlie complex networks. Then one of our final goals is to obtain a theory which can describe various complex networks and their statistical behaviors in a unified framework. For this aim, we have introduced in the recent study an analytical approach in which conditions required for the scale-free degree distribution are considered [19]. Due to the ubiquity of the scale-freeness, it is expected that the analytical conditions are given by those which are independent to specific systems and common to a wide variety of networks. Indeed, it has been shown that the power law distribution can be obtained without introducing conditions except for general ones. In this paper we extend the framework given in this previous study and show that it gives the degree correlation which agrees to the experimental measurement represented by Equation (1).

2. Framework for $P(k_1, k_2)$

With using a framework introduced in the recent study [19], we investigate the degree correlation $P(k_1, k_2)$. At first we normalize the pair of degrees (k_1, k_2) and in-

troduce variables (X, Y) which take their values respectively on the finite interval $X, Y \in [0,1]$. For example, relations between X and k_1 are given by

$$\left(k_1 - k_{1,\min}\right) \big/ \left(k_{1,\max} - k_{1,\min}\right) = X, k_{1,\min} / k_1 = X \quad (2)$$

where the former example is taken for cases such as $k_1 \in \left[k_{1,\min}, k_{1,\max}\right]$ and the latter for $k_{1,\max} = \infty$. These normalizations are summarized by the expression

$$\left(\left(\alpha_1 + k_i\right) / \alpha_2\right)^{\alpha_3} \quad (3)$$

with constants α_1, α_2, and α_3. Under the transforms between (k_1, k_2) and (X, Y) given by the expression (3), the probability $P(k_1, k_2)$ is represented by $P(X, Y)$.

In this approach we take an analytical expression of $P(X, Y)$ in the expanded form and consider the condition required for this function. Then, for variables (X, Y) and the probability $P(X, Y)$, we require conditions that X and Y are continuous and that $P(X, Y)$ is given by the smooth function with respect to (X, Y). Also in order to investigate the scaling behavior of $P(X, Y)$, we take (x, y)

$$X \equiv e^{-x}, Y \equiv e^{-y} \quad (4)$$

where $x, y \in [0, \infty)$. Because $P(X, Y)$ is a positive function which takes its value in the finite interval $[0, 1]$, the analytical representation of $P(X, Y)$ as the function of (x, y) is given in the form

$$P(x, y) = e^{-\psi(x, y)} \quad (5)$$

with the scaling function $\psi(x, y)$

$$\psi(x, y) = \sum_{n_1, n_2} \psi_{n_1, n_2} \cdot x^{n_1} y^{n_2} \quad (6)$$

where ψ_{n_1, n_2} are constant coefficients and $\psi(x, y)$ is taken to satisfy the normalization

$$1 = \int_0^\infty dx \int_0^\infty dy P(x, y) \quad (7)$$

For the single variable case, $P(k_1)$ is transformed to $P(x) = e^{-\psi(x)}$ with $\psi(x)$, the expansion of x. If $P(X)$ is scale-free and given in the power law $P(X) \sim X^{-\gamma}$ with a constant γ, $\psi(x)$ is given by the first order expansion which satisfies

$$\partial_x^2 \psi(x) = 0 \quad (8)$$

Although, according to the result given in the recent study [19], the converse fact has been shown that the condition (8) is derived without introducing special conditions except for the continuous conditions for x and $P(x)$. The point of this result is that this condition (8) is obtained with using the identical relations which the continuous distribution function satisfies generally. Then this condition is required for arbitrary variables such as X and Y.

3. The Representation of $P(k_1, k_2)$

Extending the analysis with the single variable given in the reference [19], we can show that $P(x, y)$ representation is uniquely determined in our framework. At first we take a conditional probability $P_x(y)$, the function with respect to y for each fixed value of x, defined as

$$P_x(y) = P(x, y) \Big/ \int_0^\infty dy P(x, y) \quad (9)$$

It apparently satisfies $1 = \int_0^\infty dy P_x(y)$.

For convenience of calculation, we introduce the cumulative distribution of $P_x(y)$ by

$$P_{x,0}(y) = \int_y^\infty dy P_x(y) \quad (10)$$

Then we can represent it as

$$P_{x,0}(y) = e^{-\phi_x(y)} \quad (11)$$

with an expansion

$$\phi_x(y) = \sum_{n \geq 1} \phi_{x,n}(x) y^n \quad (12)$$

where $\phi_{x,n}(x)$ is given by the other expansion of x. According to these definitions, $P_x(y)$ is given by

$$P_x(y) = -\partial_y P_{x,0}(y) = \partial_y \phi_x(y) e^{-\phi_x(y)} \quad (13)$$

Applying the condition (8), it is required that

$$\partial_y^2 \phi_x(y) = 0 \quad (14)$$

and

$$P_x(y) = \phi_{x,1}(x) e^{-\phi_{x,1}(x) y} \quad (15)$$

For $P(x, y) = e^{-\psi(x, y)}$, we can show that the same condition (11) requires that

$$\partial_x^2 \psi(x, y) = \partial_y^2 \psi(x, y) = 0 \quad (16)$$

If we take the conditional probability $P_y(x)$ with respect to x, then from the condition (11) it is required to have the form

$$P_y(x) = \phi_{y,1}(y) e^{-\phi_{y,1}(y) x} \quad (17)$$

Introducing $N_y(y)$ and $N_x(x)$ by

$$1/N_y(y) = \int_0^\infty dx P(x, y), 1/N_x(x) = \int_0^\infty dy P(x, y) \quad (18)$$

$P(x, y)$ is represented by the equivalent two forms

$$P(x, y) = \frac{\phi_{y,1}(y)}{N_y(y)} e^{-\phi_{y,1}(y) x} = \frac{\phi_{x,1}(x)}{N_x(x)} e^{-\phi_{x,1}(x) y} \quad (19)$$

and we obtain the identical relation

$$\ln\left(\phi_{y,1}(y)/N_y(y)\right)+\phi_{y,1}(y)x$$
$$= \ln\left(\phi_{x,1}(x)/N_x(x)\right)+\phi_{x,1}(c)y \quad (20)$$

Because $N_y(y)$ and $N_x(x)$ are independent to x and y respectively, the condition (16) for $\psi(x,y)$ is given by comparing each side of Equation (20). Thus we obtain the $\psi(x,y)$ representation

$$-\ln P(x,y)=\psi(x,y)=\psi_0+\psi_x x+\psi_y y+\psi_c xy \quad (21)$$

with constants ψ_0, ψ_x, ψ_y and ψ_c.

4. Degree Correlations in Real World Networks

In order to confirm our result in the previous section, we give a comparison to the experimental measurement of the real world networks. For the degree-degree correlations $P(k_1,k_2)$ given from Equation (21), we calculate $\langle k_2 \rangle$ for each fixed value of k_1 and compare it to the experimental representation (1).

At first, with using the expression (3) for the normalization of k_1, the correspondence between k_1 and x is given by

$$\ln(k_1+\alpha_1)=\gamma_{1,0}+\gamma_{1,1}x \quad (22)$$

with constants α_1, $\gamma_{1,0}$, and $\gamma_{1,1}$. While, from the representation (21), the transform between k_2 and (x,y) is generally given by

$$\ln(k_2+\alpha_2)=\gamma_{2,0}(x)+\gamma_{2,1}(x)y \quad (23)$$

where $\gamma_{2,0}(x)$ and $\gamma_{2,1}(x)$ are given by linear equations of x and α_2 is a constant.

Then $\langle k_2+\alpha_2 \rangle$ is given by

$$\langle k_2+\alpha_2 \rangle=\int_0^\infty dy \frac{\partial(k_2+\alpha_2)}{\partial y}\cdot(k_2+\alpha_2)\cdot P_x(y) \quad (24)$$

Using the representation of $P_x(y)$ given by Equation (9) and Equation (21), the average is given by

$$\langle k_2+\alpha_2 \rangle$$
$$= \int_0^\infty dy\gamma_{2,1}(x)e^{-2(\gamma_{2,0}(x)+\gamma_{2,1}(x)y)}\cdot(\psi_y+\psi_c x)e^{-(\psi_y+\psi_c x)y} \quad (25)$$

with constants ψ_y and ψ_c and this is calculated as

$$\frac{(\psi_y+\psi_c x)\gamma_{2,1}(x)}{(\psi_y+\psi_c x)+2\gamma_{2,1}(x)}e^{-2\gamma_{2,0}(x)} \quad (26)$$

Because $\gamma_{2,0}(x)$ and $\gamma_{2,1}(x)$ are represented by the linear equations of x, the first term in the above equation is estimated as

$$\frac{(\psi_y+\psi_c x)\gamma_{2,1}(x)}{(\psi_y+\psi_c x)+2\gamma_{2,1}(x)} \sim x \sim \ln k_1 \quad (27)$$

for $k_1 \gg 1$. Furthermore, using the approximation

$\ln k_1 + \ln(\ln k_1) \sim \ln k_1$, Equation (26) is approximated by a power law

$$\langle k_2 \rangle \sim k_1^{-\gamma} \quad (28)$$

with a constant γ for a large k_1. Then behavior of $\langle k_2 \rangle$ tail for a large k_1 agrees to the experimental representation (1).

5. Discussions

As we have mentioned in the introduction, our final goal is to give a description of the complex network in a unified framework. For this aim, it is required to obtain a theory which explains the organized structure of the complex network and allows to deal with different networks. In this final section, we discuss this issue.

In this paper we have shown that the extension of the framework introduced in the recent study [19] consistently produces the degree correlation in the form which agrees to the experimental data. Applying the analytical condition (8) for the single variable distribution to the joint probability, we obtain a unique representation for $P(k_1,k_2)$, the distribution of the neighbor degrees. We should notice that some properties of the complex network, the scale-free distribution and the degree correlation represented by Equation (1), can be derived from the same condition (8). It would suggest that the rules which generate scale-free networks and their correlated structures can be described in a unified framework.

Also we should notice that our framework which gives the condition (8) does not depend on the specific system. Then we can apply our results straightforwardly to various complex networks such as the protein networks and the Internet. Then our result would provide a clue to understand the mechanism which underlies various types of complex networks.

Although, for our final goal, further investigations should be required. At first we should take into account some exceptional cases for which we can not apply our representation. An example is given by a random network, in which distributions take forms such as the Gaussian or the Poisson distributions. Then it would be required for us to describe explicitly the difference between our framework and these random systems. Also we should notice that the condition (8) is derived under the assumption that the variable is continuous. However some variables such as the degree take only the discrete number. We will deal with issues such as above in the future works.

REFERENCES

[1] A.-L. Barabási and R. Albert, "Emergence of Scaling in Random Networks," *Science*, Vol. 286, No. 5439, 1999, pp. 509-512. doi:10.1126/science.286.5439.509

[2] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins and J. Wiener, "Graph Structure in the Web," *Computer Networks*, Vol. 33, No. 1-6, 2000, pp. 309-320. doi:10.1016/S1389-1286(00)00083-9

[3] M. Faloutsos, P. Faloutsos and C. Faloutsos, "On Power-Law Relationships of the Internet Topology," *ACM SIGCOMM Computer Communication Review*, Vol. 29, No. 4, 1999, pp. 251-262. doi:10.1145/316194.316229

[4] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of Networks," *Advances in Physics*, Vol. 51 No. 4, 2002, pp. 1079-1187. doi:10.1080/00018730110112519

[5] R. Albert and A.-L. Barabási, "Statistical Mechanics of Complex Networks," *Reviews of Modern Physics*, Vol. 74, No. 1, 2002, pp. 47-97. doi:10.1103/RevModPhys.74.47

[6] M. E. J. Newman, "The Structure and Function of Complex Networks," *SIAM Review*, Vol. 45, No. 2, 2003, pp. 167-256 doi:10.1137/S003614450342480

[7] R. Albert, H. Jeong and A.-L. Barabási, "Internet: Diameter of the World-Wide Web," *Nature*, Vol. 401, No. 6749, 1999, pp. 130-131. doi:10.1038/43601

[8] M. E. J. Newman, "Scientific Collaboration Networks. I. Network Construction and Fundamental Results," *Physical Review E*, Vol. 64, No. 1, 2001, pp. 1-8. doi:10.1103/PhysRevE.64.016131

[9] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai and A.-L. Barabási, "The Large-Scale Organization of Metabolic Networks," *Nature*, Vol. 407, No. 6804, 2000, pp. 651-654. doi:10.1038/35036627

[10] H. Jeong, S. Mason, A.-L. Barabási and Z. N. Oltvai, "Lethality and Centrality in Protein Networks," *Nature*, Vol. 411, No. 6833, 2001, pp. 41-42. doi:10.1038/35075138

[11] A.-L. Barab´asi, R. Albert and H. Jeong, "Mean-Field Theory for Scale-Free Random Networks," *Physica A*, Vol. 272, No. 1-2, 1999, pp. 173-187. doi:10.1016/S0378-4371(99)00291-5

[12] P. L. Krapivsky, S. Redner and F. Leyvraz, "Connectivity of Growing Random Networks," *Physical Review Letters*, Vol. 85, No. 21, 2000, pp. 4629-4632. doi:10.1103/PhysRevLett.85.4629

[13] S. N. Dorogovtsev, J. F. F. Mendes and A. N. Samukhin, "Structure of Growing Networks with Preferential Linking," Physical Review Letters, Vol. 85, No. 21, 2000, pp. 4633-4636. doi:10.1103/PhysRevLett.85.4633

[14] C. Song, S. Havlin and H. A. Makse, "Self-Similarity of Complex Networks," *Nature*, Vol. 433, No. 7024, 2005, pp. 392-395. doi:10.1038/nature03248

[15] C. Song, S. Havlin and H. A. Makse, "Origins of Fractality in the Growth of Complex Networks," *Nature Physics*, Vol. 2, 2006, pp. 275-281. doi:10.1038/nphys266

[16] S. Maslov and K. Sneppen, "Specificity and Stability in Topology of Protein Networks," Science, Vol. 296 no. 5569 2002, pp. 910-913. doi:10.1126/science.1065103

[17] R. Pastor-Satorras A. Vázquez and A. Vespignani, "Dynamical and Correlation Properties of the Internet," *Physical Review Letters*, Vol. 87, No. 25, 2001, p. 258701. doi:10.1103/PhysRevLett.87.258701

[18] M. E. J. Newman, "Assortative Mixing in Networks," *Physical Review Letters*, Vol. 89, No. 20, 2002, p. 208701. doi:10.1103/PhysRevLett.89.208701

[19] K. Takagi, "Scale Free Distribution in an Analytical Approach," *Physica A*, Vol. 389, No. 10, 2010, pp. 2143-2146. doi:10.1016/j.physa.2010.01.034