Vol 16 No 2, February 2007
1009-1963/2007/16(02)/0282-05
Chinese Physics
© 2007 Chin. Phys. Soc.
and IOP Publishing Ltd

A novel configuration model for random graphs with given degree sequence*

Xu Xin-Ping(徐新平)† and Liu Feng(刘 峰)

Institute of Particle Physics, Huazhong Normal University, Wuhan 430079, *China*

(Received 16 March 2006; revised manuscript received 4 April 2006)

Recently, random graphs in which vertices are characterized by hidden variables controlling the establishment of edges between pairs of vertices have attracted much attention. This paper presents a specific realization of a class of random network models in which the connection probability between two vertices (i, j) is a specific function of degrees ki and kj. In the framework of the configuration model of random graphs, we find the analytical expressions for the degree correlation and clustering as a function of the variance of the desired degree distribution. The obtained expressions are checked by means of numerical simulations. Possible applications of our model are discussed.

Keywords: random graphs, configuration model, correlations
PACC: 0250, 0500, 0520, 8000

1. Introduction

Many social, biological, and communication systems can be properly described as complex networks with nodes representing individuals or organizations and edges mimicking the interactions among them.[1] One of the ultimate goals of researches on complex networks is to understand how the structure of complex networks affects the dynamical process taking place on them, such as traffic flow,[2] epidemic spread,[3] cascading behaviour[4] and so on. Recently, synchronization has been observed in a variety of natural, social, physical, and biological systems,[5] and it is important to understand how the network structure affects the synchronizability of the network.[6,7]

Since the network structure plays an important role in the dynamical process taking place, proposing network models with different structure is of great interest in the physics community. In fact, most real networks share three prominent structural features: small average path length (APL), high clustering and scale-free (SF) degree distribution. Current attention has focused on proposing network models incorporating these features. To explain the small-world phenomenon (i.e. small APL), the idea of shortcuts has been proposed by Watts and Strogatz.[8] To reveal the emergence of SF degree distributions in real networks, the concept of evolving networks based on preferential attachment has been introduced by Barabási and Albert (BA).[9] A large number of models based on the variations of these two models have been proposed.

Despite the diversity of network structure, the models proposed to embody the fundamental characteristics can be roughly divided into two classes: static and evolving. The second class of causal networks encompasses, in particular, the famous BA model, whereas the configuration model[10−12] and the large group of networks with hidden variables[13−15] belong to the first class of static networks.

In this paper, we introduce a novel configuration model for random graphs by modifying the vertex pair connection probability of the Chung Lu model. We present the parallel work with this specific implementation of connection function in the framework of configuration model with given degree sequences. As we know, the connection function of the classical random graphs is constant while the influential model of BA possesses a connection function as the model of Chung and Lu does, and much effort has been devoted to the solvable properties for random graphs with the two types of connection functions. Since much previous work has been concentrated on static random graphs with the two types of connection functions, our specific connection function may enrich the class of random graphs models. The rest of the text is organized

*Project supported by the National Natural Science Foundation of China (Grant Nos 10375025 and 10275027) and the Cultivation Fund of the Key Scientific and Technical Innovation Project, Ministry of Education of China (Grant No 704035).
†Corresponding author. E-mail: xuxp@ihep.ac.cn

as follows. Section 2 gives a general description for our model. In Section 3, we find analytical expressions for the degree correlation and clustering of our model and compare these properties with those of the previous random graph models. In Section 4, we compare the simulation result with the analytical result. Conclusions and discussion are given in the last part, Section 5.

2. The model

Random graph model in which vertices are characterized by hidden variables controlling the establishment of edges between pairs of vertices has been recently proposed by Boguñá and Pastor–Satorras.[14] This general model considers graphs in which each vertex is assigned to a hidden variable, randomly drawn from a fixed probability distribution. Edges are assigned to pairs of vertices with a given connection probability, depending on the values of the hidden variables at the edge end points. Another approach to generating random graphs is the configuration model,[16] and it is a construction model which has the degree distribution as an input but is random in all other respects. Very recently, Fronczak *et al*[17] have shown that the configuration model with expected degree sequence is equivalent to random graph model with hidden variables.[17] Here, we adopt the methodology of configuration model with given degree sequence. In the configuration model, the total number of nodes is fixed to N, and degrees of all nodes $i = 1, 2, \ldots, N$ create a specific degree sequence $\{k_i\}$. Until now, nothing has been said about the connections between nodes. As a rule, random graphs with a given degree sequence are constructed in the following way: first, attach to each node i a number k_i of stubs (ends of edges); next, choose pairs of these stubs uniformly at random and join them together to make complete edges. The configuration model is determined to be an ensemble of graphs with desired degree sequences and each has equal weight.

A typical example of the configuration model has recently been proposed by Chung and Lu.[18] In their model, each vertex i is assigned to a desired degree k_i chosen from a certain distribution of interest, and then $m = \frac{1}{2}\sum_i k_i$ edges are placed between vertex pairs (i, j) with probability $p_{ij} = k_i k_j / 2m$.[16] This model has the features that the expected degree sequence of the ensemble of networks is equal to the desired degree sequence while degree sequence of the single network is not in general exactly equal to the desired degree sequence. Recently, Xu *et al*[19] have found that the model of Chung and Lu is equivalent to maximum-entropy networks in the case of sparse limit. More features of the Chung Lu model have been widely studied by a number of authors.[20,21]

Inspired by the model of Chung and Lu, we propose a configuration model for random graphs, in which the connection probability between vertex pairs (i, j) is provided as a function of degrees k_i and k_j in the form

$$p_{ij} = \frac{k_i + k_j}{N} - p = \frac{k_i + k_j - z}{N}\,, \qquad (1)$$

where N is the network size, p is the average connection probability and z is the average degree.

Let us do some analyses of the connection probability of our model. Once the desired degree sequence $\{k_i\}$ is given, the network size N, total number of edges $l = \frac{1}{2}\sum_i k_i$ and average edge connection probability $p = 2l/N^2$ are fixed. After a simple algebra, one can find that the expected degree of vertex i is

$$\bar{k}_i = \sum_j p_{ij} = k_i + \frac{\sum_j k_j}{N} - \frac{\sum_j k_j}{N} = k_i\,. \qquad (2)$$

The expected degree is equal to the desired degree (or given degree). Therefore we can study the properties of our model derived by requiring the expected degree sequence of a graph ensemble to match a given set of degree sequence of a real-world network.

3. Degree correlation and clustering of our model

Let us use the connection probability Eq.(1) to calculate the ensemble mean of the degree correlation and clustering coefficient. Since the expected degree is equal to the desired degree (or given degree), we use the expected degrees to implement the calculations. The degree correlation can be quantified by calculating the mean degree of the network neighbours of a vertex as a function of the degree k of that vertex.[1] The mean sum of the degrees of the neighbours of a vertex i, which is denoted by $\bar{K}_i^{nn}$, can be given by

$$\begin{aligned}\bar{K}_i^{nn} &= \sum_j p_{ij}\bar{k}_j = \sum_j (\frac{\bar{k}_i + \bar{k}_j}{N} - p)\bar{k}_j \\ &= \bar{k}_i \frac{\sum_j \bar{k}_j}{N} + \frac{\sum_j \bar{k}_j^2}{N} - p\sum_j \bar{k}_j \\ &= pN\bar{k}_i + \frac{N\sum_j \bar{k}_j^2 - (\sum_j \bar{k}_j)^2}{N^2}\end{aligned}$$

$$= pN\bar{k}_i + \langle \bar{k}^2 \rangle - \langle \bar{k} \rangle^2 . \quad (3)$$

Thus the mean degree of a neighbour of vertex i is equal to

$$\bar{k}_i^{nn} = \frac{\bar{K}_i^{nn}}{\bar{k}_i} = pN + \frac{Q}{\bar{k}_i} = z + \frac{Q}{\bar{k}_i} , \quad (4)$$

where

$$z = \frac{\sum_i \bar{k}_i}{N} = pN$$

is the average expected degree and Q is variance of the expected degree distribution. Throughout this paper, we denote the expected values or ensemble mean by an overbar (e.g. $\bar{k}$), and the average over network vertices by an additional bracket (e.g. $\langle \bar{k} \rangle$). Equation (4) is composed of an uncorrelated item z and a correlated item $Q/\bar{k}_i$, indicating that $(\bar{k}_i^{nn} - z)$ versus $\bar{k}_i$ follows a power law with an exponent -1. Thus our model belongs to the correlated random graph model, in contrast to uncorrelated classical random graph model and Chung–Lu model.

Incidentally, the degree correlation coefficient determined by Newman is given by [22]

$$\begin{aligned} r &= \frac{\sum_i \bar{k}_i \bar{K}_i^{nn} - (2\bar{l})^{-1} (\sum_i \bar{k}_i^2)^2}{\sum_i \bar{k}_i^3 - (2\bar{l})^{-1} (\sum_i \bar{k}_i^2)^2} \\ &= -\frac{(\langle \bar{k}^2 \rangle - \langle \bar{k} \rangle^2)^2}{\langle \bar{k} \rangle \langle \bar{k}^3 \rangle - \langle \bar{k}^2 \rangle^2} \\ &= -\frac{Q^2}{\langle \bar{k} \rangle \langle \bar{k}^3 \rangle - \langle \bar{k}^2 \rangle^2} . \end{aligned} \quad (5)$$

Thus our model has negative correlation and shows a disassortative mixing structure. Since a large number of real networks seem to exhibit negative degree correlation and there is no correlation in the classical random graph model and Chung–Lu model, our model may be closer to real networks.

Next, we derive an analytical formula for the clustering coefficient. Before we obtain the clustering coefficient of a certain vertex i, we need to calculate the ensemble mean of edges among its nearest neighbours. Consider two arbitrary vertices m and n. The probability at which these two vertices are connected to vertex i simultaneously can be given by $p_{im} p_{in}^{im}$, where p_{in}^{im} is the connection probability between i and n on condition that i and m are already connected. In the same way, we use $p_{mn}^{im,in}$ to denote the connection probability between m and n, provided that they have a common vertex i. The average connections $\bar{E}_i$ between the nearest neighbours of vertex i, can be formally computed as the probability at which vertex i is connected to vertices m and n, and also these two vertices are at the same time joined by an edge, which is the summation over all the possible vertex pairs m and n, and divided by 2. Thus, we can write the clustering coefficient of vertex i as

$$\begin{aligned} \bar{C}_i &= \frac{\bar{E}_i}{\bar{k}_i(\bar{k}_i - 1)/2} = \frac{2\bar{E}_i}{\bar{k}_i(\bar{k}_i - 1)} \\ &= \frac{\sum_{m,n} p_{im} p_m^{im} p_{mn}^{im,in}}{\bar{k}_i(\bar{k}_i - 1)} . \end{aligned} \quad (6)$$

In our ensemble model, each link is present or absent, independent of all others. Therefore the conditional probability can be simplified into $p_{in}^{im} = p_{in}$, $p_{mn}^{im,in} = p_{mn}$. Substituting these formulae into Eq.(6), we have

$$\bar{C}_i \approx \frac{\sum_{m,n} p_{im} p_{in} p_{mn}}{\bar{k}_i^2} = p + \frac{2Q}{N\bar{k}_i} . \quad (7)$$

Equation (7) gives the relation between $\bar{C}_i$ and $\bar{k}_i$, which we call the clustering spectrum or degree-dependent clustering.[23–25] Equation (7) indicates that $(\bar{C}_i - p)$ versus $\bar{k}_i$ also follows a power law with an exponent -1. Considering that the classical random graph and Chung–Lu model provide an insufficient description for the nontrivial clustering of real networks, we believe that our model is more realistic. Consequently, the average clustering coefficient of the whole network is

$$C = p + \frac{2Q}{N^2} \sum_i \frac{1}{\bar{k}_i} = p + \frac{2Q}{N} \Big\langle \frac{1}{\bar{k}} \Big\rangle . \quad (8)$$

Combining Eqs.(8) and (11), we obtain

$$\bar{C}_i = \frac{2\bar{k}_i^{nn}}{N} - p . \quad (9)$$

Therefore, the clustering coefficient $\bar{C}_i$ increases linearly with $\bar{k}_i^{nn}$, which implies that degree–degree correlation leads to degree-dependent local clustering. Previously, this feature was observed in a number of real networks.[26–28] Here we demonstrate that this dependence is a direct consequence of degree–degree correlation.

4. Comparison with numerical simulation

In order to test the analytical predictions, we carry out computer simulation in the following steps. First, we choose an advisable degree sequence from a certain distribution of interest as an input of the desired degree sequence. Second, we use the connection probability Eq.(1) to generate edges and finally

the whole network is taken as a realization in the ensemble. Third, repeat step 2 to create an ensemble of single networks. Fourth, calculate the corresponding variables including degree correlation and clustering for each single network, and obtain the average values of the variables in the ensemble networks. In order to minimize the statistical fluctuations, we average the physical quantities over a large number of realizations. Considering that there is one link, at most, between any vertex pairs, we find that the connection probability should satisfy $0 \leq P_{ij} \leq 1$ for any vertex pair (i, j), and the case of $p_{ij} < 0$ or $p_{ij} > 1$ can be avoided.

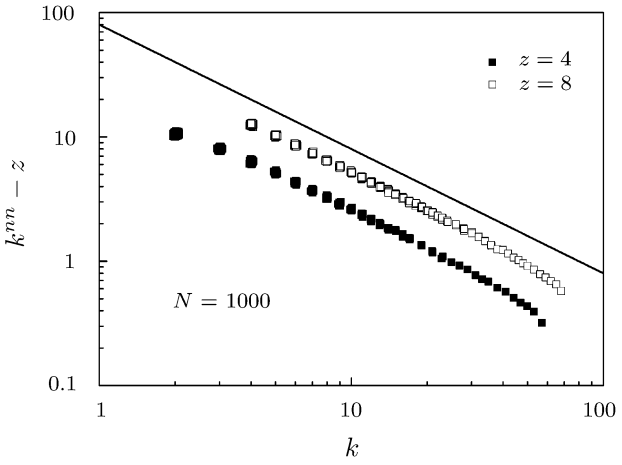


Fig.1. $\bar{k}^{nn} - z$ versus $\bar{k}$ in our model in the cases of $z = 4$ and $z = 8$. The scatters are the numerical results which are averaged over 5000 realizations for $N = 1000$. The solid line corresponds to the analytical dependency $(k^{nn} - z) \sim k^{-1}$.

Since for most real networks, the degree distributions follow power-law $P(k) = ck^{-r}$ with $2 < \gamma < 3$, we choose desired degree sequence from a power law with $\gamma = 2.5$ to perform our numerical simulation. We expect that other types of degree sequences should give qualitatively the same behaviour. The main results are shown in Figs.1–3. Consider first Fig.1, which is a plot of the mean degree $\bar{k}^{nn} - z$ as a function of $\bar{k}$. As the figure shows, $\bar{k}^{nn} - z$ versus $\bar{k}$ follows a power law with an exponent -1, as expected. The same behaviour is evident in Fig.2 also, which shows $\bar{C} - p$ as a function of the mean degree $\bar{k}$ of that vertex. The straight lines in Figs.1 and 2 correspond to a strict power law with an exponent -1, as predicted by Eqs.(4) and (7). In Fig.3, we plot the degree-dependent local clustering $\bar{C}$ versus the average nearest neighbour degree k^{nn}, which indicates a linear dependency between them. We have tried to alter the parameters of desired degree sequences, and similar behaviours or results are obtained.

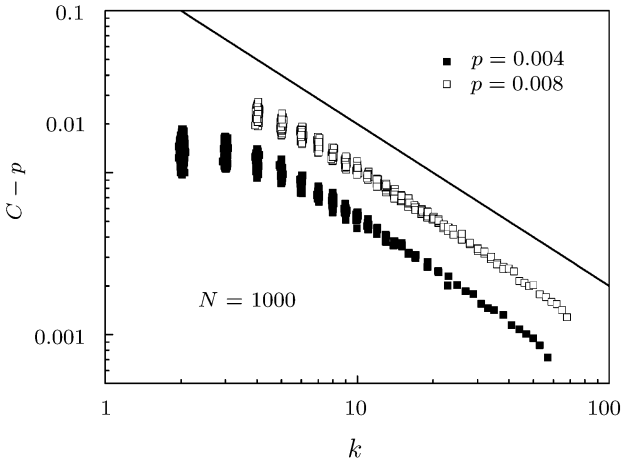


Fig.2. $\bar{C} - p$ versus $\bar{k}$ in our model in the cases of $p = 0.004$ and $p = 0.008$. The scatters are the numerical results which are averaged over 5000 realizations for $N = 1000$. The solid line corresponds to the analytical dependency $(C - p) \sim k^{-1}$.

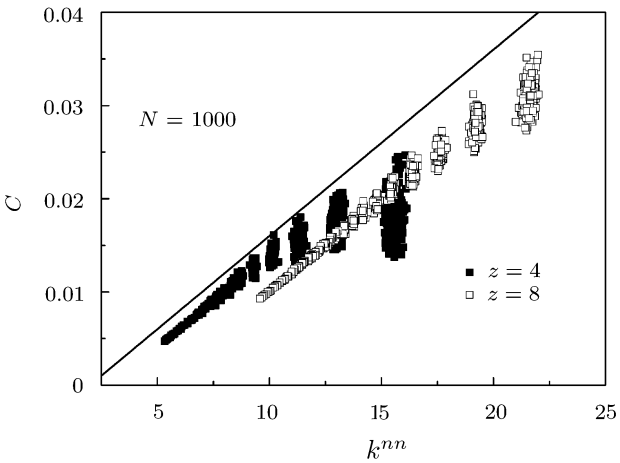


Fig.3. $\bar{C}$ versus $\bar{k}^{nn}$ in our model in the cases of $z = 4$ and $z = 8$. The scatters are the numerical results which are averaged over 5000 realizations for $N = 1000$. The solid line has a slope 0.002, as predicted by Eq.(9).

5. Conclusions and discussion

In this paper, we have presented a specific realization of a class of random network models in which the connection probability between two vertices (i, j) is a specific function of degrees k_i and k_j. We have derived the exact analytic formulas for the degree correlation and clustering of our model. It is found that the degree correlation and clustering are dependent on the variance of the given degree sequence. The analytical expressions are tested by numerical simulation.

Particularly, for the Poisson degree distribution, the second item of Eq.(8) is neglectable, and the clustering coefficient of our model is $C \approx p$, Therefore, we can conclude that our model is close to classical random graph for the Poisson degree sequence. Furthermore, while the classical random graph model and Chung–Lu model give a trivial description for the degree correlation, our model may provide a recommendable tool to unveil the origin of degree correlations observed in real networks. Many other properties of our model also require further study, such as the position of the phase transition at which a giant component first forms, the mean component size, the size of the giant component if there is one, and the average vertex-vertex distance within a graph, etc. Ultimately, we expect that our model can shed much light on features or behaviours which are consistent with real networks, and may take a step forward towards the study of random graphs.

References

[1] Mendes J F F, Dorogovtsev S N and Ioffe A F 2003 *Evolution of networks: From Biological Nets to the Internet and WWW* (Oxford University Press, Oxford).

[2] Tadi B, Thurner S and Rodgers G J 2004 *Phys. Rev.* E **69** 036102

[3] Pastor-Satorras R and Vespignani A 2001 *Phys. Rev. Lett.* **86** 3200

[4] Goh K-I, Lee D-S, Kahng B and Kim D 2003 *Phys.Rev. Lett.* **91** 148701

[5] Glass L *Nature* 2001 **410** 277

[6] Lu J and Chen G 2005 *IEEE Trans. Auto. Contr.* **50** 841

[7] Lu J, Yu X, Chen G and Cheng D 2004 *IEEE Trans. Cir. Sys. I* **51** 787

[8] Watts D J and Strogatz S H 1998 *Nature* **393** 440

[9] Barab'asi A-L and Albert R 1999 *Science* **286** 509

[10] Molloy M and Reed B 1995 *Random Struct. Algorithms* **6** 161

[11] Newman M E J, Strogatz S H and Watts D J 2001 *Phys. Rev.* E **64** 026118

[12] Fronczak A, Fronczak P and Holyst J A 2005 *e-print cond-mat*/0502663

[13] Söderberg B 2002 *Phys. Rev.* E **66** 066121

[14] Boguñá M and Pastor-Satorras R 2003 *Phys. Rev.* E **68** 036112

[15] Fronczak A and Fronczak P 2005 *e-print cond-mat*/0503069

[16] Newman M E J 2003 *SIAM Rev.* **45** 167

[17] Fronczak A, Fronczak P and Holyst J A 2006 *Phys. Rev.* E **73** 016108

[18] Chung F and Lu L 2002 *Proc. Natl. Acad. Sci. USA* **99** 15879

[19] Xu X P, Liu F and Liu L S 2005 *Physica* A **356** 662

[20] Bekessy A, Bekessy P and Komlos J 1972 *Stud. Sci. Math. Hun-gar* **7** 343

[21] Bender E A and Canfield E R 1978 *Journal of Combinatorial Theory* A **24** 296

[22] Newman M E J 2002 *Phys. Rev. Lett.* **89** 208701

[23] Huang Z X *et al* 2004 *Chin. Phys.* **13** 273

[24] Dai S and Guo Y J 2004 *Chin. Phys.* **13** 423

[25] Li Y *et al* 2005 *Chin. Phys.* **14** 2153

[26] Dorogovtsev S N, Goltsev A V and Mendes J F F 2002 *Phys. Rev.* E **65** 066122

[27] Szabo G, Alava M and Kertesz J 2003 *Phys. Rev.* E **67** 056102

[28] Ravasz E and Baraba'si A-L 2003 *Phys. Rev.* E **67** 026112