Available online at www.sciencedirect.com
ScienceDirect
Physica A 376 (2007) 673–678
PHYSICA A
www.elsevier.com/locate/physa

A growth model of community graph with a degree distribution consisting of two distinct parts

Kenichi Nakazato[a,*], Takaya Arita[b]

[a]Graduate School of Human Informatics, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan
[b]Graduate School of Information Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya 464-8601, Japan

Received 29 June 2006; received in revised form 4 August 2006
Available online 15 November 2006

Abstract

Many complex systems can be represented as complex networks. Among them there has been much interest on community structure recently, and many studies focus attention on it, in particular community detection. While community detection can provide us much information, the community structure implies another feature, hierarchy of the system. Coarse-graining of complex networks can lead us to the definition of community graph. The empirical degree distribution of community graph has a unique nature, where it consists of two distinct parts, exponential and power law distribution. In this paper, we propose a modified model of community graph [Pollner et al. Europhys. Lett. 73 (2006) 478.] that mimics the empirical features of it. The growth mechanism of the model is a combination of preferential and non-preferential attachment in a higher level. We show that the model can reproduce the unique degree distribution by theoretical and numerical analysis. While such features might stem from some other reasons, we would expect to provide a unique aspect of community graph.

Keywords: Complex networks; Power law; Community structure

1. Introduction

Many complex phenomena have both dynamical aspects and structural ones. Representing complex systems as networks leads to a quite fruitful approach for investigating their structural or dynamical properties individually or both of them simultaneously. Such complex networks have been studied extensively in the last decade, in particular, WWW, co-authorship networks, metabolic networks, protein interaction networks, social networks, and so on. Most class of networks exhibit striking topological features in common, e.g., the small world effect or scale-free degree distribution [1–4], which is characterized by a dynamical growth mechanism known as *preferential attachment* [2–4]. In those properties, the main focus is centered on the macroscopic average or the local topological aspects such as vertex degree, clustering coefficient, average path length and those distributions. Another issue which attracts much attention in recent years is about modular

*Corresponding author.
E-mail address: nakazato@alife.cs.is.nagoya-u.ac.jp (K. Nakazato).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2006.10.083

structure or sub-units in networks known as *community structure* [5–7], which means the contrasting density of edges in networks or the distinguishing parts with many edges in comparison with sparse areas. Such contrast in edge density is practically used for detection of community structure buried in complex networks and the definitions of some modularity measures [7–17]. Additionally the existence of such sub-units can be indicative of hierarchical structure, fractality, and consequently other topological properties at coarse-grained levels [5,18,19].

The community structure can be widely detected among many types of networks [7–13,15,16] and provide significant information about not only structural properties but also functional ones. For instance, community structure detected in biological networks like metabolic pathway or protein interaction networks is thought to be relevant with functional units [9–11]. As another example, the community in social networks like e-mail networks of organizations or collaboration networks of musicians could correspond to social groups and represent its social role [12,13]. Those communities thereby represent collective units, which we can regard as meta-nodes. As a next step, the inter-community structure or the positioning of a community in the whole system becomes the focus for us. In general, the functionality or the role of the parts is not independent of its contexts, in other words, other communities in the network. Each part in networks plays its roll in collaboration for relevant communities. In practice, communities detected by the clique percolation method in which a community is defined as the k-clique percolation cluster in networks can overlap, that is to say, multiple communities can share some nodes at the same time [17]. Furthermore it is suggested that complex networks cannot always be partitioned into distinct communities without overlapping in a deterministic manner [14]. The shared nodes connecting their affiliated communities can characterize the relationships between communities. Thus, *community graph* or *overlapping community structure* can be defined as the network with the set of nodes corresponding to communities and the set of links which represents overlaps between pairs of communities [17,20].

As for such community graphs, some significant topological features are reported [17]. The first one is the scale-free distribution of community size [15,16,21]. The second is about the degree distribution of community graph. The degree distribution can be divided into two distinct regions: an exponential decay in low degree region and a power law in the high-degree region. Moreover, the exponent of the power-law region is the same as the one in the community size distribution. The scale-free nature of a community graph suggests the presence of the growth mechanism based on the *preferential attachment* rule [2,3,22]. The empirical data on the growing co-authorship network on the Los Alamos e-print archive indeed supports the existence of preferential growth behind the evolving mechanism [20]. However, why does the degree distribution of community graph have two distinct parts, especially exponential decay in low degree area?

In this paper we focus on the features of community graph. We extend the growth model of overlapping community structure [20], where we adopt a growth mechanism with a combination of preferential and non-preferential attachment in a higher level, and demonstrate the model can reproduce the observed features by theoretical analysis and numerical simulation.

2. Model description

In this section, we construct a growth model of a community graph as an extension of the model proposed by Pollner et al. [20]. In their model, the connections between the elementary nodes are not considered. Instead, the graph structure representing connections between communities are their focus. For this reason, the property which characterizes the elementary units does not include the information about the connections between them, but instead consists of their affiliation to communities. The growth process is realized by joining new nodes to each community selected with a probability which is proportional to the community size. In such a preferential growing process, the distribution of the size develops into a power law asymptotically [22,23]. On the occasion of incorporation, the number of communities where each new node participates is determined by a Poissonian distribution, which creates a new overlap between those selected communities. If the number of communities to participate in equals zero, the new node will remain outside the communities and result in constituting a new community at the time when the ratio of such independent nodes increase to a predefined value. It has been shown numerically that the community size distribution and the community degree distribution follow a scale-free distribution with the same exponent.

Although the fat tailed distribution can be reproduced by the model, the other aspect of the community graph, i.e., the feature in the low degree region before the cross-over to the power-law decay is not focused on by them. We propose a modified version of the model that mimics the empirical distribution which has two distinct parts. The main idea regarding modification is as follows: when a new node is incorporated into a selected community, the new node does not necessarily come from the outside of the network, in other words, it is not necessarily a novice. Considering the team assembly process [24], for instance, it is plausible that a new team is composed of a mixture of persons with various degrees of experience. Thus, as for the community growth in our model, veteran nodes can be recruited from other communities selected randomly. The rules of the modified version are as follows:

(1) At each time step a new community with m initial nodes is made with a creation probability α.
(2) Some existing communities are selected in proportion to their own community size, where the number of selected communities β is drawn from a Poissonian distribution with an average value β_m.
(3) With a non-novice probability δ each selected community gets a new member from an existing community selected randomly in a non-preferential manner, otherwise with a probability $1-\delta$ from the outside of the system.

In the former case in (3), the preferentially-selected community and the randomly-selected community come to share the member, and this overlap corresponds to the link between the two communities. As the initial condition, we assume some communities with m nodes at time step $t=0$. The rules described above are limited to a minimal set for simplicity, while we could take some more mechanisms into the rules such as recruitment of a new member from a preferentially-selected community.

3. Theoretical analysis

3.1. The size of community

Using the continuum theory approach [2,25,26], the size n_i of community i grows according to the following equation:

$$\frac{\partial n_i}{\partial t} = \beta \frac{n_i}{\sum_k n_k}, \tag{1}$$

where $\beta \approx \beta_m$ is the number of communities which are selected to grow at each time step. This equation corresponds to the preferential growth process of each community. By solving Eq. (1) with the initial condition that $n_i = m$ at the time $t = t_i$, we obtain

$$n_i = m\left(\frac{t}{t_i}\right)^{\beta/(\beta+\alpha m)}, \tag{2}$$

where t_i is the time step when the community i is created. Thus the probability distribution function of the community size can be described as follows:

$$P(n) \sim n^{-\gamma}, \tag{3}$$

where

$$\gamma = \frac{2\beta + \alpha m}{\beta}. \tag{4}$$

This shows that the rules we adopted in this paper can generate power-law behavior on community size as each community grows in proportion to its size regardless of whether the new node is a novice or a veteran.

3.2. The degree distribution of community graph

As for community graph, the links correspond to the overlappings between pairs of communities. These overlappings are formed only when recruiting new nodes from other existing communities in this model. The recruitment process begins with random selection of a community, in which the selector is the community selected in a preferential manner in terms of community size at each time step. As the recruitment of a veteran node occurs with probability δ, the evolution equation for community degree k_i for community i can be written as

$$\frac{\partial k_i}{\partial t} = \beta\delta \frac{1}{c(t)} + \beta\delta \frac{n_i}{\sum_k n_k}, \tag{5}$$

where $c(t)$ is the number of the existing communities at time step t and given by $c(t) \approx \alpha t$. The first term on the right-hand side reflects random selection by the other community, which makes the community i grow, while the second term represents the case the community i selects a new member from other communities. Since the size n_i is given by Eq. (2), we can integrate this equation. By using the initial condition $k_i = 0$ when $t = t_i$, we obtain

$$k_i = \frac{\beta\delta}{\alpha} \ln\left(\frac{t}{t_i}\right) + \delta m \left(\frac{t}{t_i}\right)^{\beta/(\beta+\alpha m)} - \delta m. \tag{6}$$

We can divide the degree k_i into two parts: $k_i = k_{exp} + k_{sf}$, where they correspond to the first term and the remaining terms on the right side of this equation, respectively:

$$k_{exp} = \frac{\beta\delta}{\alpha} \ln\left(\frac{t}{t_i}\right), \tag{7}$$

$$k_{sf} = \delta m \left(\frac{t}{t_i}\right)^{\beta/(\beta+\alpha m)} - \delta m. \tag{8}$$

Provided that these two parts, k_{exp} and k_{sf}, are independent, the distribution of the community degree, which is the main result in this section, thereby has the form:

$$P(k) = P(k_{exp} + k_{sf}) \sim P(k_{exp}) + P(k_{sf}), \tag{9}$$

where

$$P(k_{exp}) \sim \exp\left(\frac{-\alpha k_{exp}}{\beta\delta}\right), \tag{10}$$

$$P(k_{sf}) \sim k_{sf}^{-\gamma}. \tag{11}$$

As a matter of fact, there is a positive correlation between the two types of community degree, i.e., the older the community is, the larger both types of community degrees. The empirical unique distribution consisting of two distinct parts however can be derived in the presence of such correlation, considering the dominance of an exponential distribution in the low-degree region and a power law in the high-degree region. Thus the community degree distribution in this model also has the form of an exponential decay followed by a fat-tailed distribution with the same exponent as the one in the community size distribution.

4. Numerical simulation

Analytic results described in the previous section are compared with numerical simulations. We present the numerical data in Fig. 1. In accordance with the analytic results, the community size distribution follows the power law. Also, the exponent of the distribution is very close to γ predicted by Eq. (3).

Fig. 2 shows the numerical results on the community degree distribution for some cases, $\delta = 0.05{-}0.125$. We see that the numerical data is approximately consistent with the prediction from Eqs. (10) and (11), that is an exponential decay followed by a power-law tail. In particular, they show good agreement when δ is low,

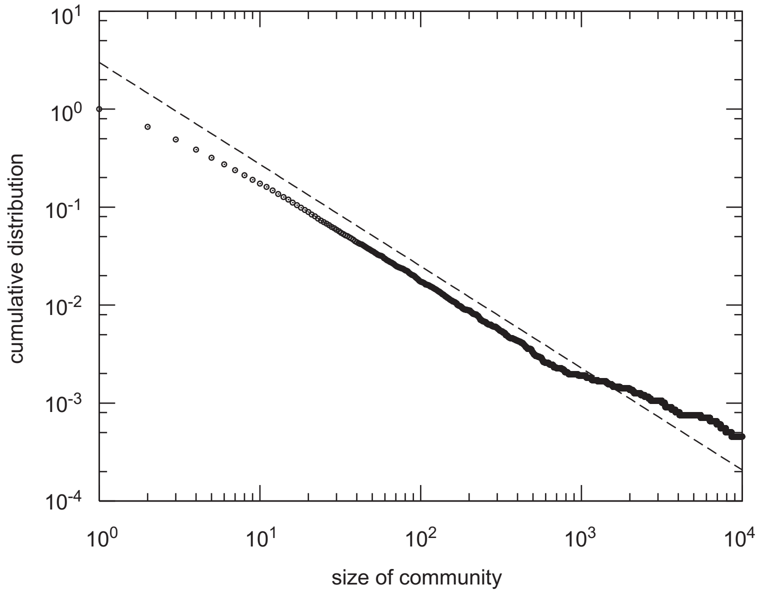

Fig. 1. Double logarithmic plot of the cumulative probability distribution of the community size $P(n)$. The result is consistent with the predicted power-law distribution $P(n){\sim}n^{-\gamma}$. In the numerical simulation, we used $t = 200\,000, m = 2, \alpha = 0.1, \beta_m = 5.0$, and $\delta = 0.05$. The dashed line corresponds to the theoretical prediction, $\gamma = 1.04$.

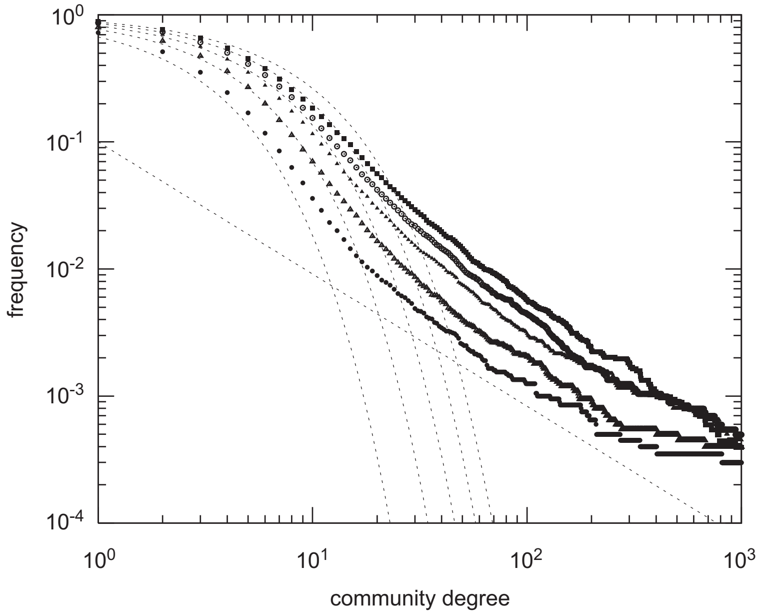

Fig. 2. Double logarithmic plot of the cumulative probability distributions of the community degree $P(k)$. Each result is approximately consistent with the predicted distribution which starts with an exponential curve and then crosses over to the power law. The dashed curves correspond to the theoretically predicted exponential, $\exp(-(\alpha/\beta\delta)k)$. The dashed line corresponds to the predicted power-law with the exponent $\gamma = 1.04$. We used $t = 200\,000, m = 2, \alpha = 0.1, \beta_m = 5.0$, and $\delta = 0.05, 0.075, 0.1, 0.125, 0.15$.

while the higher δ is, the more vague the distinctness between these two regimes. This tendency stems from the increase of connectivity as it consequently moves the crossover between the exponential and the power-law regions back to the lower-degree.

5. Summary and discussion

In this paper, we have studied the growing process of community structure using a model which is a modified version of the model of the overlapping community structure [20]. Since the focus here is on the structure of community graph, the connections between elementary nodes are not considered so that the model can be sufficiently simple to represent the features of community graph. The model has two growth mechanisms besides the creation of a new community: a preferential growth of community and a non-preferential growth making overlaps in coarse-grained (community) level. We have conducted both theoretical analysis and numerical simulations, which showed that this model can mimic the empirical distribution unique to the overlapping community structure.

It should be noted that such a combination of random and preferential growth rules for community graph could reproduce the community degree distribution divided into two distinct parts, while it has been shown in contrast that the growth mechanisms with the same type of combination for growing networks with elementary nodes and edges between them show not such a unique degree distribution but either a simple scale-free distribution or alternatively an exponential one [26]. Therefore, the feature captured by this model can be considered unique to community graph, though the empirical distribution might stem from some other reasons [17]. We hope that this model can provide a good example of such dissimilar features emerging in different levels even with the similar rules in contrast with the similarities among the diverse levels of hierarchies like fractal phenomena in complex networks [18,19].

References

[1] D. Watts, S. Strogatz, Nature 393 (1998) 440.
[2] A.-L. Barabási, R. Albert, Science 286 (1999) 509.
[3] R. Albert, A.-L. Barabási, Rev. Mod. Phys. 74 (2002) 47.
[4] J.F.F. Mendes, S.N. Dorogovtsev, Evolution of Networks, Oxford University Press, Oxford, 2003.
[5] E. Ravasz, A.L. Somera, D.A. Mongru, Z. Oltvai, A.-L. Barabási, Science 297 (2002) 1551.
[6] D. Watts, P.S. Dodds, M.E.J. Newman, Science 296 (2002) 1302.
[7] M. Girvan, M.E.J. Newman, Proc. Natl. Acad. Sci. USA 99 (2002) 7821.
[8] F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, D. Parisi, Proc. Natl. Acad. Sci. USA 101 (2004) 2658.
[9] R. Guimerà, L.A.N. Amaral, Nature 433 (2005) 895.
[10] V. Spirin, L.A. Mirny, Proc. Natl. Acad. Sci. USA 100 (2003) 12123.
[11] P. Holme, M. Huss, H. Jeong, Bioinformatics 19 (2003) 532.
[12] R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, A. Arenas, Phys. Rev. E 68 (2003) 065103.
[13] P. Gleiser, L. Danon, Adv. Complex Syst. 6 (2003) 565.
[14] J. Reichardt, S. Bornholdt, cond-mat/0603718.
[15] M.E.J. Newman, Phys. Rev. E 69 (2004) 066133.
[16] L. Danon, A. Díaz-Guilera, J. Duch, A. Arenas, J. Stat. Mech. (2005) P09008.
[17] G. Palla, I. Derényi, I. Farkas, T. Vicsek, Nature 435 (2005) 814.
[18] C. Song, S. Havlin, H.A. Makse, Nature 433 (2005) 392.
[19] K.-I. Goh, G. Salvi, B. Kahng, D. Kim, Phys. Rev. Lett. 96 (2006) 018701.
[20] P. Pollner, G. Palla, T. Vicsek, Europhys. Lett. 73 (2006) 478.
[21] J.D. Noh, H.-C. Jeong, Y.-Y. Ahn, H. Jeong, Phys. Rev. E 71 (2005) 036131.
[22] H.A. Simon, Biometrika 42 (1955) 425.
[23] D.H. Zanette, S.C. Manrubia, Physica A 295 (2001) 1.
[24] R. Guimerà, B. Uzzi, J. Spiro, L.A.N. Amaral, Science 308 (2005) 697.
[25] R. Albert, A.-L. Barabási, Physica A 272 (1999) 173.
[26] R. Albert, A.-L. Barabási, Phys. Rev. Lett. 85 (2000) 5234.