Growing network model for community with group structure

Jae Dong Noh,[1] Hyeong-Chai Jeong,[2] Yong-Yeol Ahn,[3] and Hawoong Jeong[3]

[1]*Department of Physics, Chungnam National University, Daejeon 305-764, Korea*
[2]*Department of Physics and IFP, Sejong University, Seoul 143-747, Korea*
[3]*Department of Physics, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea*

We propose a growing network model for a community with a group structure. The community consists of individual members and groups, gatherings of members. The community grows as a new member is introduced by an existing member at each time step. The new member then creates a new group or joins one of the groups of the introducer. We investigate the emerging community structure analytically and numerically. The group size distribution shows a power-law distribution for a variety of growth rules, while the activity distribution follows an exponential or a power law depending on the details of the growth rule. We also present an analysis of empirical data from online communities the "Groups" in http://www.yahoo.com and the "Cafe" in http://www.daum.net, which show a power-law distribution for a wide range of group sizes.

I. INTRODUCTION

Emergent properties of artificial or natural complex systems have attracted growing interest recently. Some of them are conveniently modeled with a network, where constituent ingredients and interactions are represented with vertices and links, respectively. Watts and Strogatz demonstrated that real-world networks display the small-world effect and the clustering property, which cannot be explained with the regular and random networks [1]. Later on, in the study of the WWW network, Albert *et al.* found that the degree (i.e., the number of attached links) of each vertex follows a power-law distribution [2]. Those works have triggered a burst of research on the structure and the organization principle of complex networks (see Refs. [3–5] for reviews).

Many real-world networks, e.g., in biological, social, and technological systems, are found to obey the power-law degree distribution [3]. A network with the power-law distribution is called a scale-free network. One of the possible mechanisms for the power law is explained with the Barabási-Albert (BA) model [6]. The model assumes that a network is growing and that the rate acquiring a new link for an existing vertex is proportional to a popularity measured by its degree. The popularity-based growth appears very natural since, e.g., in creating a new web site, one would link it preferentially to popular sites having many links. With the BA and related network models, structural and dynamical properties of networks have been explored extensively.

On the other hand, there exists another class of networks which have a group structure. Consider, for example, online communities such as the "Groups" operated by Yahoo [7] and the "Cafes" operated by the Korean portal site Daum [8]. They consist of individual members and groups, gatherings of members with a common interest, and growth of the community is driven not only by members but also by groups. A community evolves as an individual registers as a new member. The newcomers can create new groups with existing members or join existing groups. The online community is a rapidly growing social network [9]. The emerging structure would be distinct from that observed in networks without the group structure. In this paper, we propose a growing network model for the community with the group structure. We model the community with a bipartite network consisting of two distinct kinds of vertices representing members and groups, respectively. A link may exist only between a member vertex and a group vertex, which represents a membership relation.

The bipartite network [10] has been considered in the study of the movie actor network [1] consisting of actors and movies, the scientific collaboration network [10,11] of scientists and articles, and the company director network [10] of directors and boards of directors. Usually those networks are treated as unipartite by projecting out one kind of vertex of less interest [12,13]. Some biological and social networks are known to have a modular structure [14,15], where vertices in a common module are densely connected, while vertices in different modules are sparsely connected. The modular structure is coded implicitly in the connectivity between vertices. Unipartite network models with a modular structure were also studied in Refs. [15–21], where vertices form modules which in turn form bigger modules hierarchically [15–17] or the modular structure emerges dynamically as a result of social interactions [18–21]. In Ref. [21], each vertex is assigned to a Potts-spin-like variable pointing to its module [21]. These studies on the group structures of networks have focused mainly on the groups with a finite number of members. However, there are groups in the real-world online community which keep growing as the community evolves.

Reflecting growing dynamics of the real-world online community, our model takes account of the group structure explicitly with a bipartite network consisting of member and group vertices. Upon growing, both the member and group vertices evolve in time. We study the dynamics of the size of groups and the activity of the members. The size of a group is defined as the number of members in the group and the activity of a member is the number of groups in which the member participates. When the community grows large enough, the group size distribution shows a power-law distribution unlike the network models studied previously [16,21]. To test our model, we analyze the empirical data from online communities, the Yahoo "Groups" [7] and the

Daum "Cafe" [8], and show that both communities indeed show power-law group size distributions for wide ranges of group sizes.

This paper is organized as follows. In Sec. II, we introduce the growing network model. Depending on the choice of detailed dynamic rules, one may consider a few variants of the model. Characteristics such as the group size distribution, the member activity distribution, and the growth of the number of groups are studied analytically in a mean field theory and numerically in Sec. III. Those characteristics are also calculated for the real-world online communities and compared with the model results. We conclude the paper with a summary in Sec. IV.

II. MODEL

We introduce a model for a growing community with a group structure. The community grows by adding a new member at a time, who may open a new group or join an existing group [22]. The following notations are adopted: A member entering the community at time step i is denoted by I_i. The activity, the number of participating groups, of I_i is denoted by A_i. As members enter the community, new groups are created or existing groups expand. The αth group is denoted by G_α, its creation time by τ_α, and its size by S_α. The total numbers of members and groups are denoted by N and M, respectively.

Initially, at time $t=0$, the community is assumed to be inaugurated by m_0 members, denoted by $I_{-(m_0-1)}, \ldots, I_0$, belonging to an initial group G_1. That is, we have that $N(t=0)=m_0$, $M(t=0)=1$, $A_j(t=0)=1$ for $j=-(m_0-1), \ldots, 0$, $\tau_1=0$, and $S_1(t=0)=m_0$. At time t, a new individual I_t is introduced into the community and becomes a member by repeating the following procedures until its activity reaches m:

• **Selection:** It selects a partner I_j among existing members $\{I_{k<t}\}$ with a selection probability P_j^S.

• **Creation or Joining:** With a creation probability P_j^C, it creates a new group G_{M+1} with the partner I_j. Otherwise, it selects randomly one of the groups of I_j with an equal probability and joins it. If I_t is already a member of the selected group, then the procedure is canceled.

A specific feature of the model varies with the choice of probabilities P^S and P^C. Regarding the selection, the simplest is the random choice among existing members with equal probability $P_j^S=1/(m_0+t-1)$. Note that the selection may be regarded as an invitation of a new member by existing members. It may then be natural to assume that active members invite more newcomers. Such a case is modeled with a preferential selection probability $P_j^S=A_j/(\Sigma_{k<t} A_k)$. After selecting a partner I_j, the newcomer may create a new group or join one of I_j's groups having an equal probability. In that case the creation probability is variable as $P_j^C=1/(A_j+1)$. In the other case, it may create a new group with a fixed probability $P_j^C=\omega$. Combining the strategies in the two procedures, we consider the four possible different growth models denoted by RV, RF, PV, and PF. Here, R (P) stands for the random (preferential) selection, and V (F) for the group cre-

TABLE I. Model description and mean field results for the group size distribution exponent γ. Here, Θ_{RV} and Θ_{PV} are the group number growth rate given in Eqs. (8) and (17), respectively. The activity distribution follows a power law only for the PF model with the exponent $\lambda=2+1/\omega$.

	R $[P_j^S=1/(m_0+t-1)]$	P $[P_j^S=A_j/\Sigma_{k<t}A_k]$
V $[P_j^C=1/(A_j+1)]$	$1+\Theta_{RV}^{-1}$	$1+\Theta_{PV}^{-1}$
F $(P_j^C=\omega)$	$2/(1-\omega)$	$2/(1-\omega)$

ation with the variable (fixed) probability. For example, the RF model has the selection probability $P_j^S=1/(m_0+t-1)$ and the creation probability, $P_j^C=1/(A_j+1)$. The growth rules are summarized in Table I.

The whole structure of the community is conveniently represented with a bipartite network of two kinds of vertices: one for the group and the other for the member. A link exists only from a member vertex to a group vertex to which it belongs. The member activity and the group size correspond to the degree of the corresponding vertex. Figure 1 shows a typical network configuration for the RV model with $m_0=m=1$. To help readers understand the growth dynamics, we add the indices for members I_i and groups G_α in the figure. It is easily read off that I_1 selects I_0 and becomes a member of G_1 at $t=1$ and that I_2 opens a new group G_2 with I_0 at $t=2$, and so on. Figure 2 shows a configuration of a RV network with $m=m_0=1$ grown up to $N=1000$ members with $M=452$ groups. It is noteworthy that there appear hub groups having a lot of members. The emerging structure of the network will be studied in the next section.

III. NETWORK STRUCTURE

The number of groups $M(t)$, the activity of each member $A_i(t)$, and the size of each group $S_\alpha(t)$ increase as the network grows. With those quantities, we characterize the growth dynamics and the network structure. In the following, we study the dynamics of those quantities averaged over network realizations. For simplicity's sake, we make use of the same notations for the averaged quantities. The network dynamics implies that they evolve in time as follows:
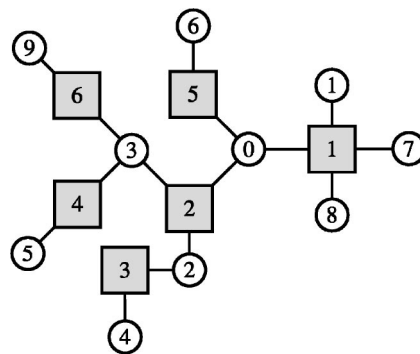

FIG. 1. A network for the RV model with $m_0=m=1$ and $N=10$ with six groups. The symbols i (circles) and α (squares) represent a member I_i and a group G_α, respectively.
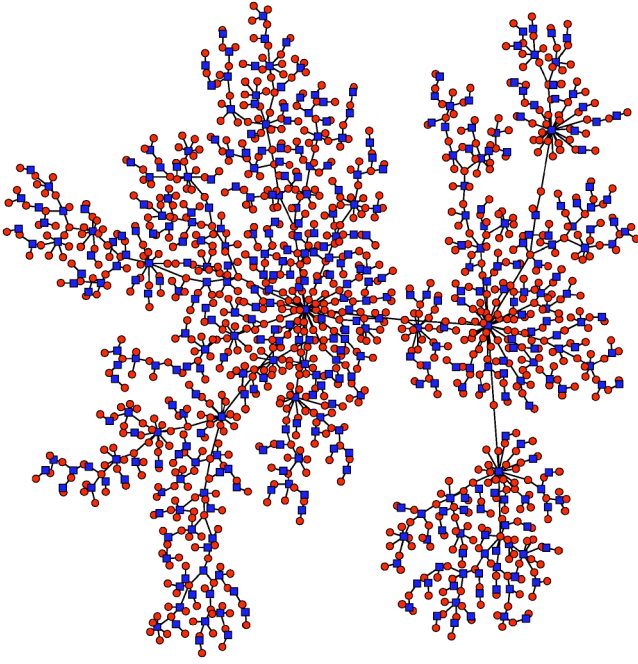
FIG. 2. (color online) A network for the RV model with $m_0 = m = 1$ and $N = 1000$. A square (circle) symbol stands for a group (member).

$$A_i(t+1) = A_i(t) + mP_i^S P_i^C, \qquad (1)$$

$$M(t+1) = M(t) + m\sum_{j \leqslant t} P_j^S P_j^C, \qquad (2)$$

$$S_\alpha(t+1) = S_\alpha(t) + m\sum_{j \leqslant t} P_j^S \chi_{j\alpha}(1 - P_j^C)/A_j, \qquad (3)$$

where $\chi_{j\alpha} = 1$ if I_j belongs to G_α, or 0 otherwise. The initial conditions are given by $A_i(t=i) = m$, $M(t=0) = 1$, and $S_\alpha(t = \tau_\alpha) = 2$ with τ_α the creation time of G_α. We analyze the equations in a continuum limit and in a mean field scheme, neglecting any correlation among dynamic variables.

Firstly we consider the RV model. Using the corresponding P^C and P^S in Table I, Eqs. (1)–(3) become

$$\frac{dA_i}{dt} = \frac{m}{(A_i + 1)(m_0 + t)}, \qquad (4)$$

$$\frac{dM}{dt} = \frac{1}{(m_0 + t)}\sum_{j \leqslant t} \frac{m}{(A_j + 1)}, \qquad (5)$$

$$\frac{dS_\alpha}{dt} = \left(\frac{1}{m_0 + t}\right)\left(\frac{S_\alpha}{m_0 + t}\right)\sum_{j \leqslant t} \frac{m}{(A_j + 1)}, \qquad (6)$$

where we approximate $\chi_{j\alpha}$ in Eq. (3) with $S_\alpha/(m_0+t)$, the fraction of members of G_α among all members. The solution for $A_i(t)$ is given by

$$A_i(t) = -1 + \sqrt{(m+1)^2 + 2m \ln\left[\frac{m_0 + t}{m_0 + i}\right]}. \qquad (7)$$

This shows that an older member with smaller i has a larger activity and that the activity grows very slowly in time. With the solution for A, one can easily show that $\sum_{j \leqslant t} m/(A_j+1) \simeq \Theta_{RV}(m_0+t)$ for large t with

$$\Theta_{RV} = \int_0^1 du \frac{m}{\sqrt{(m+1)^2 - 2m \ln u}}. \qquad (8)$$

Hence, the average number of groups increases linearly in time as $M(t) \simeq \Theta_{RV} t$ with the group number growth rate Θ_{RV}. The group size increases algebraically as

$$S_\alpha(t) \simeq 2\left(\frac{m_0 + t}{m_0 + \tau_\alpha}\right)^{\Theta_{RV}}. \qquad (9)$$

We have obtained the activity of each member and the size of each group, which allow us to derive the distribution functions $P_a(A)$ and $P_s(S)$ for the activity and the group size, respectively. The activity distribution function is given by the relation $P_a(A) = P_{in}(i)|di/dA|$ with the uniform individual distribution, $P_{in}(i) = 1/(m_0+t)$. The differentiation can be done through Eq. (7), which yields that the activity distribution is bounded as $P_a(A) = (A+1)\exp\{-((A+1)^2 - (m+1)^2)/(2m)\}/m$. Similarly, the group size distribution is given by $P_s(S) = P_a(\tau)|d\tau/dS|$ with the group creation time distribution $P_a(\tau)$. We assume that the group creation time is distributed uniformly, which is justified with the linear growth of $M \simeq \Theta_{RV}(m_0+t)$. Then the group size distribution follows a power law $P_s(S) \sim S^{-\gamma_{RV}}$ with the exponent

$$\gamma_{RV} = 1 + \Theta_{RV}^{-1}. \qquad (10)$$

Note that the distribution exponent is determined by the group number growth rate Θ_{RV}.

We now turn to the PF model. With the selection and creation probabilities, Eqs. (1)–(3) are written as

$$\frac{dA_i}{dt} = \frac{m\omega A_i}{\sum_{j \leqslant t} A_j}, \qquad (11)$$

$$\frac{dM}{dt} = m\omega, \qquad (12)$$

$$\frac{dS_\alpha}{dt} = (1 - \omega)S_\alpha \frac{m}{\sum_{j \leqslant t} A_j}. \qquad (13)$$

We also took the approximation $\chi_{i\alpha} = S_\alpha/(m_0+t)$ in Eq. (3). Trivially, we find that the group number grows in time as $M(t) = m\omega t + 1$. For A_i and S_α, one need evaluate the quantity $\sum_{j \leqslant t} A_j$. Summing over all i both sides of Eq. (11), one obtains that $\sum_{i \leqslant t}(dA_i/dt) = m\omega$. Note that $d(\sum_{i \leqslant t} A_i)/dt = \sum_{i \leqslant t}(dA_i/dt) + m = (1+\omega)m$, which yields that $(\sum_{j \leqslant t} A_j) = m(1+\omega)t + m_0$. Hence we obtain the algebraic growth of the activity and the group size as

$$A_i(t) = m \left(\frac{m(1+\omega)t + m_0}{m(1+\omega)i + m_0} \right)^{\omega/(1+\omega)}, \qquad (14)$$

$$S_\alpha(t) = 2 \left(\frac{m(1+\omega)t + m_0}{m(1+\omega)t_\alpha + m_0} \right)^{(1-\omega)/(1+\omega)}. \qquad (15)$$

These results allow us to find the distribution functions $P_a(A)$ and $P_s(S)$. They follow the power distributions $P_a(A) \sim A^{-\lambda_{PF}}$ and $P_s(S) \sim S^{-\gamma_{PF}}$ with the exponents

$$\lambda_{PF} = 2 + 1/\omega \quad \text{and} \quad \gamma_{PF} = 2/(1-\omega). \qquad (16)$$

Here we also assumed the uniform distribution of τ_α in Eq. (15), which is supported from the linear growth of $M(t) \sim m\omega t$. In contrast to the RV model, both distributions follow the power law. The exponents do not depend on the parameter m, but only on the group creation probability ω.

For the PV and the RF models, the following can be shown easily: The PV model behaves similarly to the RV model. The group number increases linearly in time as $M(t) \simeq \Theta_{PV} t$ with the group number growth rate Θ_{PV}. Unfortunately, we could not obtain a closed form expression for it. However, if we adopt the assumption that the selection probability P_i^S is proportional to $A_i + 1$ instead of A_i, it can be evaluated analytically as

$$\Theta_{PV} \simeq \left[\sqrt{m^2 + 6m + 1} - (m+1) \right]/2. \qquad (17)$$

The approximation would become better for larger values of m. The group size grows algebraically as in Eq. (9) with Θ_{PV} instead of Θ_{RV}. Therefore, the group size distribution follows the power law with the exponent γ_{PV} presented in Table I. The RF model also displays the power-law group size distribution. The distribution exponent γ_{RF} is given in Table I. Note that γ_{RF} and γ_{PF} are the same. On the other hand, the activity distribution follows an exponential distribution in the RF and the PV models.

The origin for the power-law distribution of the group size is easily understood. In all models considered, the size of a group increases when one of its members invites a new member. The larger a group is, the more chance to invite new members it has. Therefore, there exists a preferential growth in the group size, which is known to lead to the power-law distribution [6].

The activity of a member increases when a newcomer selects it and creates a new group. When the random selection probability is adopted, such a process does not occur preferentially for members with higher activity. It results in an exponential-type activity distribution in the RV and RF models. In the PV model, although the selection probability is proportional to the activity, the creation probability is inversely proportional to the activity. Hence, it does not have a preferential growth mechanism in the member activity either. Only in the PF model, the activity growth rate is proportional to the activity of each member. Therefore, the activity distribution follows the power law only in the PF model.

The analytic mean field results are compared with numerical simulations. In simulations, we chose $m_0 = m$ and all data were obtained after the average over at least 10 000 samples. We present the numerical data in Fig. 3. In accordance with
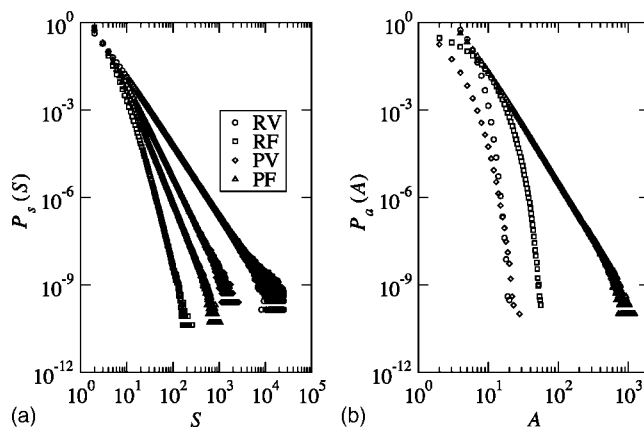

FIG. 3. (a) The group size distribution and (b) the activity distribution. The model parameters are $m = 4, 1$ for the RV and the PV models, respectively. For the RF model, $m = 4$ and $\omega = 0.6$, for the PF model, $m = 4$ and $\omega = 0.5$. The community has grown up to $N = 10^6$ and the distributions are averaged over 10^4 samples.

the mean field results, the group size distribution follows the power law in all cases. The activity distribution also shows the expected behavior: the power-law distribution for the PF model and exponential-type distributions for the other models. We summarize the distribution exponents in Fig. 4. The measured values of the distribution exponents are in good agreement with the analytic results.

Our network models display distinct behaviors from those bipartite networks such as the movie actor network, the scientific collaboration networks, and the director board network, which have been studied previously. For the first two examples, their growth is driven only by the member vertices, the actors, and the scientists, respectively. The activity of members may increase in time. However, the group vertices, the movies and the papers, respectively, are frozen dynamically and their sizes are bounded practically. For the last example, both the members (directors) and the groups
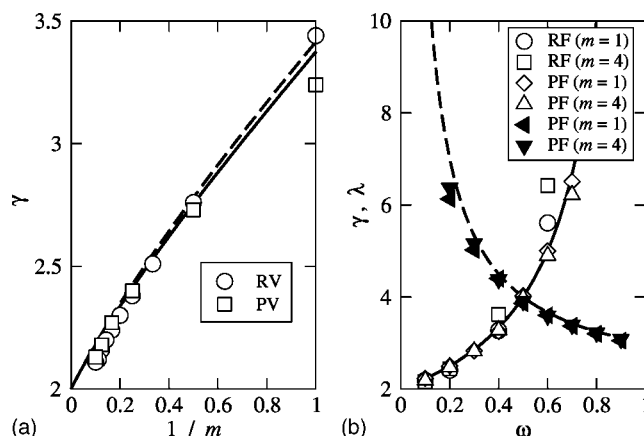

FIG. 4. (a) Numerical results for γ for the RV and the PV models. The solid (dashed) curve represents the analytic mean field results for the RV (PV) model. (b) Numerical results for γ (open symbols) of the RF and the PF models, and for λ (filled symbols) of the PF model. The solid (dashed) curve represents the analytic results for $\gamma(\lambda)$ in Table I.
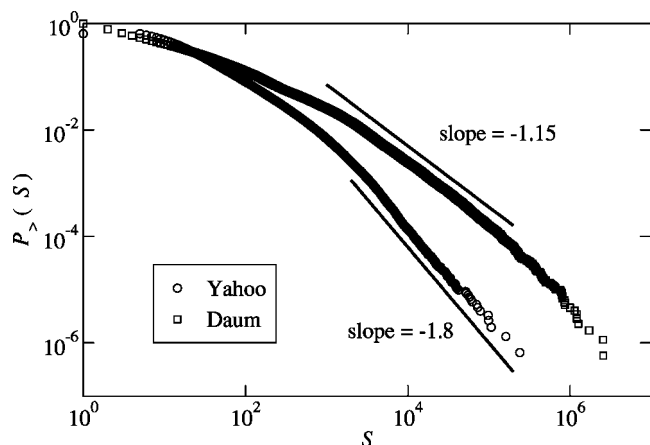
FIG. 5. Cumulative group size distribution of the online communities in Yahoo and Daum.

(boards) may evolve in time. However, it was shown that the group size distribution is also bounded [10].

Our model is applicable to evolving networks with the group structure wherein the size of a group may increase unlimitedly. The online community is a good example of such networks. To test the possibility, we study the empirical data obtained from the Groups and the Cafe operated by Yahoo and Daum, respectively. It is found in August, 2004, that there are 1 516 750 (1 743 130) groups (cafes) with 76 587 494 (351 565 837) cumulative members in the Yahoo (Daum) site. The numbers of members of the groups are available via the web sites. Figure 5 presents the cumulative distribution $P_>(S) = \Sigma_{S'>S} P_s(S')$ of the group size. The distribution has a fat tail [24]. Although the distribution function in the log-log scale show a non-negligible curvature in the entire range, it can still be fitted reasonably well into the power law for a range over two decades (see the straight lines drawn in Fig. 5). From the fitting, we obtain the group size distribution exponents $\gamma_{\text{Yahoo}} \simeq 2.8$ and $\gamma_{\text{Daum}} \simeq 2.15$. The power-law scaling suggests that the online community may be described by our network model. Unfortunately, in-

formation on the activity distribution is not available publicly. Thus, we could not compare the activity distribution of the communities with the model results. We would like to add the following remark: A real-world online community evolves in time as new members are introduced and new groups are created. At the same time, it also evolves as members leave it and groups are closed. Those processes are not incorporated into the model. Our model is a minimal model for the online community, wherein the effects of leaving members and closed groups are neglected.

IV. SUMMARY

We have introduced the bipartite network model for a growing community with the group structure. The community consists of members and groups, gatherings of members. Those ingredients are represented with distinct kinds of vertices. A membership relation is represented with a link between a member and a group. Upon growing, a group increases its size when one of its members introduces a new member. Hence, a larger group grows preferentially faster than a smaller group. With the analytic mean field approaches and the computer simulations, we have shown that the preferential growth leads to the power-law distribution of the group size. On the other hand, the activity distribution follows the power law only for the PF model with the preferential selection probability and the fixed creation probability (see Table I). We have also studied the empirical data obtained from the online communities, the Groups of Yahoo and the Cafe of Daum. Both communities display a power-law distribution of the group size. It suggests our network model be useful in studying their structure.

ACKNOWLEDGMENTS

[1] D. J. Watts and S. H. Strogatz, Nature (London) **393**, 440 (1998).

[2] R. Albert, H. Jeong, and A.-L. Barabási, Nature (London) **401**, 130 (1999).

[3] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[4] S. N. Dorogovtsev and J. F. F.Mendes, Adv. Phys. **51**, 1079 (2002).

[5] M. E. J. Newman, SIAM Rev. **45**, 167 (2003).

[6] A.-L. Barabási and R. Albert, Science **286**, 509 (1999); A.-L. Barabási, R. Albert, and H. Jeong, Physica A **272**, 173 (1999).

[7] http://www.yahoo.com

[8] http://www.daum.net

[9] For instance, the Daum Cafe has grown up with more than 3 $\times 10^6$ of groups and 28×10^6 of members since it was first launched in 1999.

[10] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev.

E **64**, 026118 (2001).

[11] M. L. Goldstein, S. A. Morris, and G. G. Yen, cond-mat/0409205.

[12] M. E. J. Newman, Proc. Natl. Acad. Sci. U.S.A. **98**, 404 (2001); Phys. Rev. E **64**, 016131 (2001); *ibid.* **64**, 016132 (2001).

[13] M. E. J.Newman, Phys. Rev. E **68**, 026121 (2003).

[14] M. Girvan and M. E.J.Newman, Proc. Natl. Acad. Sci. U.S.A. **99**, 7821 (2002).

[15] E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Baraási, Science **297**, 1551 (2002); E. Ravasz and A.-L. Barabási, Phys. Rev. E **67**, 026112 (2003).

[16] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science **296**, 1302 (2002).

[17] A. E. Motter, T. Nishikawa, and Y.-C. Lai, Phys. Rev. E **68**, 036105 (2003).

[18] B. Skyrms and R. Pemantle, Proc. Natl. Acad. Sci. U.S.A. **97**, 9340 (2000).

[19] E. M. Jin, M. Girvan, and M. E .J. Newman, Phys. Rev. E **64**, 046132 (2001).

[20] A. Grönlund and P. Holme, Phys. Rev. E **70**, 036108 (2004).

[21] D.-H. Kim, G. J. Rodgers, B. Kahng, and D. Kim, cond-mat/0310233.

[22] The dynamics of opening a new group or joining an existing group is analogous to the copy and growth dynamics of the Simon model (see e.g., Ref. [23]).

[23] S. Bornholdt and H. Ebel, Phys. Rev. E **64**, 035104 (2001).

[24] The Yahoo groups can be divided into 16 categories and 18 165 subcategories and the Daum cafes into 22 categories and 825 subcategories. We also find that the group size distribution in each category also has a similar fat tail.