

Modified Penna Bit-string Network Evolution Model for Scale-free Networks with Assortative Mixing

Yup KIM, Woosik CHOI and Soon-Hyung YOON*

Department of Physics and Research Institute for Basic Sciences, Kyung Hee University, Seoul 130-701, Korea

(Received 13 October 2011, in final form 24 November 2011)

Motivated by biological aging dynamics, we introduce a network evolution model for social interaction networks. In order to study the effect of social interactions originating from biological and sociological reasons on the topological properties of networks, we introduce the activity-dependent rewiring process. From the numerical simulations, we show that the degree distribution of the obtained networks follows a power-law distribution with an exponentially decaying tail, $P(k) \sim (k+c)^{-\gamma} \exp(-k/k_0)$. The obtained value of γ is in the range $2 < \gamma < 3$, which is consistent with the values for real social networks. Moreover, we also show that the degree-degree correlation of the network is positive, which is a characteristic of social interaction networks. The possible applications of our model to real systems are also discussed.

PACS numbers: 89.75.Hc, 87.23.Ge, 89.20.Hh, 89.65.-s

Keywords: Complex networks, Social networks, Network evolution

DOI: 10.3938/jkps.60.621

I. INTRODUCTION

During the last decade, there has been an explosion in studies on the structural and the dynamical properties of complex networks due to a wide range of discoveries in real systems [1]. The examples range from physical and technological systems [2,3] to biological [4] and social [5] systems. Numerous studies have revealed that those networks share several common topological properties, such as small-world phenomena [6]. An interesting finding was that complex weblike structures are divided into two different classes based on their degree distribution. The degree distribution, $P(k)$, is defined as the probability to find a node of degree k and is the simplest measure for the topological property of networks. Networks whose degree distribution satisfies a power-law $P(k) \sim k^{-\gamma}$ are called scale-free networks (SFNs). Examples of SFNs include the world wide web (WWW) [7], protein-interaction networks (PIN) [4], and social interaction networks [8]. Since the discovery of SFNs in real networks, many studies have tried to uncover the underlying mechanism that causes the power-law degree distribution [9]. One important mechanism that is believed to produce a SFN is the preferential attachment [10]. On the other hand, when the evolution of the network is completely random, the degree distribution is known to follow a Poisson distribution [11]. Thus, a network whose degree distribution is Poissonian is called a

random network (RN). Examples of RNs are power-grid networks and highway networks [1]. The origin of such a difference in $P(k)$ is known to be closely related to the physical properties of the networks, such as tolerance against attack [1,9].

Based on degree-degree correlations, networks can be divided into two groups [12]. When the degree-degree correlation is positive, the network is assortative while it is disassortative if the correlation is negative. In general, disassortative networks include technological networks such as the WWW, the Internet, and biological networks, such as PINs [1]. On the other hand, social interaction networks are well known examples of assortative networks [12]. The degree-degree correlation also affects the topological properties [13] and various dynamical properties on networks [14]. Thus, an understanding of the underlying mechanisms that affect the degree-degree correlation and the degree distribution during the evolution of the network is important to enlarge the horizon of knowledge on complex networks.

Recently, Li and Maini investigated the topological properties of complex network generated by using a modified Penna bit-string model [15]. The Penna bit-string model was originally introduced for biological aging based on Darwinian evolution with mutations [16,17]. Due to its simplicity and predictability for well-known phenomena such as Gompertz law [18,19], the Penna model has been widely used in the population dynamics in biological systems with aging. In Ref. 15, Li and Maini modified the Vurhulst factor to control the birth of individuals in the population and introduced activity-based

*E-mail: syook@khu.ac.kr

preferential attachment of offspring to incorporate the social interaction during the evolution of networks. The results show that such biologically originating interaction can produce a network with assortative mixing. However, a scale-free degree distribution was not obtained, even though the degree distribution became fatter than the Poisson distribution [15]. Therefore, in this paper, we introduce a social interaction based on sociological and biological reasons and show that the resultant network has a scale-free degree distribution with assortative mixing. We will also discuss some possible applications of the studied models to real systems.

This paper is organized as follows: In Sec. II, we define two models for a systematic study of social interactions. The simulation results are given in Sec. III. In Sec. IV, a summary and discussion are given.

II. MODELS

The asexual Penna model [17] uses a bit-string to represent a chronological genome for each individual (each bit can be 0 or 1). Each position in the bit-string corresponds to a unit time interval in the life of the individual, for example, a year. Thus, the maximum age to which an individual can live is thirty-two. A bit set to 1 represents a disease gene, and one set to 0 corresponds to a healthy one. In the standard Penna model, the rules for the individual to be alive are (i) the age of the individual is less than or equals to 32, (ii) the number of accumulated disease gene, m , is less than a threshold T , and (iii) at each time step, the individual will stay alive with the probability $V(t) = 1 - N(t)/N_{max}$, where $N(t)$ is the number of individuals at time t and N_{max} is the parameter characterizing the maximum population allowed by the environment. The probability $V(t)$ is called the Verhulst factor. Li and Maini argued that the Verhulst factor in the standard Penna model is too severe [15] because the simulation results show that the population size never reaches $0.3N_{max}$ [20]. Based on this argument, we discarded the Verhulst factor to allow the network to grow and adopted the activity function in Ref. 15 for the interaction. Moreover, by introducing the rewiring process based on the sociological reasons, we will show that the Penna bit-string network evolution model has a scale-free degree distribution with assortative mixing.

The definition of model is as follows: *Initialization*—We start from N_0 individuals ($1 \ll N_0 \ll N_{max}$). Each individual has a 32-bit string with value 0. In order to locate the disease genes in the bit-string of individual i , we randomly choose an integer T_i in the interval $[1, T]$. T_i locations in the bit-string of i are selected at random, and the value is changed to 1. Each individual is connected to another one with probability $\phi = 0.1$. At each time step, all individuals in the population increase their age by unity, which means that the corresponding gene in the bit-string is expressed. If the expressed gene is a

disease gene, then m_i is increased by 1. *Death*—At each time step, all individuals whose ages reach 33 die. In addition, the individuals with $m \geq T$ die. The dead individuals and links connected to them are removed from the network. *Birth*—After the completion of the death process, all individuals whose ages are larger than the reproduction age R make B offspring, each of which inherits the genome and links of its parent. Each inherited healthy gene mutates into a disease gene with probability P . *Interaction*—In order to introduce the sociologically and biologically originating interaction to the network evolution, we use the activity of each individual defined in Ref. 15. The activity of each individual strongly depends on its age and health. The age-dependent activity for individual i is defined as

$$A_{i1} = \begin{cases} 0 & \text{if } 0 < a_i < R \\ 1 - \exp[(a_i - 32)/2] & \text{if } R \leq a_i \leq 32, \end{cases} \quad (1)$$

where a_i is the age of individual i , and the health-dependent activity is defined as

$$A_{i2} = \exp\left(-\frac{m_i}{2}\right). \quad (2)$$

Here, m_i is the accumulated number of expressed disease genes for individual i . Using Eqs. (1) and (2), we define the activity A_i of individual i as

$$A_i = A_{i1}A_{i2}. \quad (3)$$

The new-born offspring inherits the links of its parent because in social systems, the new-born baby starts its social interaction with the people who already have acquaintance with its parents. However, some of the inherited links can be removed for various reasons. One such reason might come from social activity; *i.e.*, links to individuals with small activity can be easily removed. Thus, we first consider the effect of the removal of inherited links due to the activity (**model A**). The removal probability of each inherited link of individual i is defined as

$$R_i = \frac{A_i^{-1}}{\sum_{j=1}^{k_i} A_j^{-1}}, \quad (4)$$

where k_i is the degree of individual i . The main differences between the model in Ref. 15 and our model are in the Verhulst factor and the removal probability of inherited links.

In **model B**, we add new connections between unconnected individuals. In social systems, making a new connection is regarded as making a new friend. It is natural to assume that if the difference in the age between the two individuals is small, then they have more chance to become friends. The activity of each individual also affect the probability of making a new friendship between unconnected individuals. Thus, we assume that

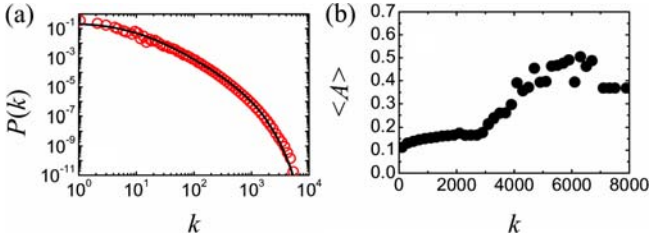


Fig. 1. (Color online) (a) Plot of $P(k)$ for $N = 10^6$. The solid line represents the relation $P(k) \sim (k+c)^{-\gamma} \exp(-k/k_0)$ with $\gamma = 2.27$, $c = 9$, and $k_0 = 550$. (b) Plot of the average activity, $\langle A(k) \rangle$, against k .

the probability of making a new link between unconnected individuals i and j is

$$\Pi(i, j) = A_i A_j \exp(-|a_i - a_j|). \quad (5)$$

In the following simulations, we use $N_0 = 1000$, $T = 3$, $R = 8$, $B = 1$, and $P = 0.1$.

III. SIMULATION RESULTS

1. Model A

In Fig. 1(a), we show the $P(k)$ distribution when $N(t) = 10^6$ for model A. The obtained $P(k)$ can be best fitted by a power-law distribution with an exponential cutoff [21]:

$$P(k) \sim (k+c)^{-\gamma} \exp(-k/k_0), \quad (6)$$

with $\gamma \approx 2.27$. This result can be easily understood from the gene duplication-divergence model for a PIN [22]. In the duplication-divergence model, the newborn gene inherits its connections from its parent; then, due to some mutation, its inherited links are removed with non-zero probability. This process resembles the link-removal process in our model during the social and the biological interaction. In the duplication-divergence model, the $P(k)$ distribution is known to follow Eq. (6) [23].

The origin of the power-law in the degree distribution is known to be closely related to the preferential attachment of new links [9]. In model A, the probability to inherit the parent's links is proportional to the activity of the parent's nearest neighbors. Thus, to investigate the relationship between A and k , we measure the average activity, $\langle A(k) \rangle$, for each k in Fig. 1(b). The data clearly show that $\langle A \rangle$ increases as k increases. This implies that the more active individual has more links. As a result, the evolutionary dynamics of the network has a preferential attachment. This behavior of $\langle A \rangle$ can be qualitatively understood from the definition of the model. Since the probability to keep the connection in inherited links is proportional to A , the individual with large A has more chance to be connected to the newborn offspring, which causes an effective preferential attachment.

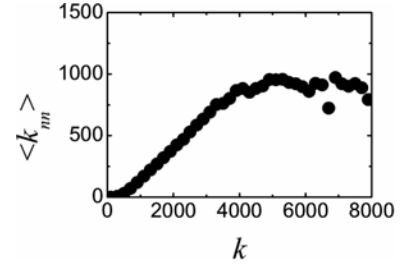


Fig. 2. Plot of $\langle k_{nn} \rangle$ against k .

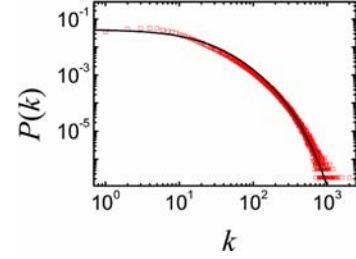


Fig. 3. (Color online) Plot of activity $P(k)$ with social interaction. The solid line corresponds to $P(k) \sim (k+c)^{-\gamma} \exp(-k/k_0)$ with $\gamma = 2.1$, $c = 37$, and $k_0 = 170$.

In Fig. 2, we measure the average degree of the nearest neighbors of nodes of degree k , $\langle k_{nn}(k) \rangle$, to measure the degree-degree correlation [24]. $\langle k_{nn} \rangle$ is defined as $\langle k_{nn} \rangle = \sum_{k'} k' P(k'|k)$, where $P(k'|k)$ is the conditional probability that a node of degree k is connected to a node of degree k' . Thus, if there is no degree-degree correlation, then $\langle k_{nn} \rangle$ is independent of k . On the other hand, when the network has assortative mixing (or positive degree correlation), $\langle k_{nn} \rangle$ increases as k increases. The data in Fig. 2 clearly shows that $\langle k_{nn} \rangle$ increases as k increases when $k < 4000$. For $k > 4000$, $\langle k_{nn} \rangle$ does not seem to depend on k . This indicates that the degree-degree correlation becomes positive. We also measure the Pearson coefficient defined as

$$r = \frac{M^{-1} \sum_i j_i k_i - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}{M^{-1} \sum_i \frac{1}{2} (j_i^2 + k_i^2) - \left[M^{-1} \sum_i \frac{1}{2} (j_i + k_i) \right]^2}, \quad (7)$$

where M is the total number of links and j_i and k_i are the degrees of nodes at the ends of the i th edge, with $i = 1, \dots, M$ [12]. When $r > 0$ the network is assortative, and when $r < 0$, it is disassortative. For model A, we obtain $r \approx 0.44$, which clearly shows that the obtained network is assortative.

2. Model B

A behavior similar to that of model A can be found in model B. In Fig. 3, we show the $P(k)$ distribution

obtained from model B. From the best fit of the data to Eq. (6), we obtain $\gamma \approx 2.1$. The value of γ in model B is slightly smaller than that for model A. The smaller value of γ in model B comes from the effective rewiring process in model B. Even in the preferential attachment model, γ decreases as the rewiring probability increases [9]. In model B, the attachment of new links enhances the effective rewiring probability. As a result, the value of γ for model B becomes smaller than that for model A. The Pearson coefficient, r , for model B is also positive ($r \simeq 0.31$). This shows that the addition of new links simply enhances the heterogeneity in the degree distribution.

IV. SUMMARY AND DISCUSSION

In summary, inspired by the aging dynamics in biological systems, we introduce social interactions to the evolutionary dynamics of networks based on the Penna bit-string model. From numerical simulations, we find that the obtained networks satisfy a power-law degree distribution of $\gamma \simeq 2.1 \sim 2.3$ with an exponential cutoff. The emergence of a scale-free behavior in $P(k)$ is understood from the well-known dynamics of the duplication-divergence model for PIN. By measuring $\langle k_{nn} \rangle$ and the Pearson coefficient, we also find that the networks are assortative, which is a well-known characteristics of social networks. Because most social networks are assortative and SFN with $\gamma \simeq 2 \sim 3$ [1], we expect that our model can be applied to study various properties in social networks. For example, google plus requires an invitation from an already registered user to enroll in the service as a new user [25]. Thus, the evolutionary dynamics of the social network in google plus is directly mapped into our model in the following way: The registered user and the new user correspond to the parent and the offspring in our model, respectively. Aging can be interpreted as a gradual losing of their interests on SNS. In this mapping, it is natural to assume that the friends of registered user have more chance to be a friend of new user because invitations would more frequently occur between people who are already acquaintances. An age-dependent activity has been observed in Gowalla [26]. Gowalla is a location-based SNS created in 2009. The empirically measured user activity distribution of Gowalla decays exponentially as in Eq. (1), and the $P(k)$ distribution is known to have a power-law tail [26]. Therefore, we expect our model to provide a theoretical framework to investigate various social phenomena in SNS.

ACKNOWLEDGMENTS

We thank Mr. H. Kim and Mr. B. Han for the computational work. This work was supported by National Research Foundation of Korea (NRF) Grant funded

by the Korean Government (MEST) (No. 2009-0073939 and No. 2010-0021586) and by the Korea Research Foundation Grant funded by the Korean Government (MOEHRD, Basic Research Promotion Fund) (KRF-2008-321-B00031). One of the authors (Yup Kim) acknowledges the support from Kyung Hee University through the requested research year 2011.

REFERENCES

- [1] M. E. J. Newman, *Networks: An Introduction* (Oxford University Press, Oxford, 2010).
- [2] S-H. Yook, H. Jeong and A-L. Barabási, *Proc. Nat. Acad. Sci.* **99**, 13382 (2002).
- [3] Y. Kim, J-H. Kim and S-H. Yook, *Phys. Rev. E* **83**, 056115 (2011).
- [4] S-H. Yook, Z. Oltvai and A-L. Barabási, *Proteomics* **4**, 928 (2003).
- [5] H. Ebel, L-I. Mielsch and S. Bornholdt, *Phys. Rev. E* **66**, 035103(R) (2002).
- [6] D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998).
- [7] R. Albert, H. Jeong and A-L. Barabási, *Nature* **401**, 130 (1999).
- [8] M. E. J. Newman, *Phys. Rev. E* **64**, 016131 (2001).
- [9] R. Albert and A-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [10] A-L. Barabási, R. Albert and H. Jeong, *Physica A* **272**, 173 (1999).
- [11] P. Erdős and A. Rényi, *Publ. Math-Debrecen* **6**, 290 (1959).
- [12] M. E. J. Newman, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [13] S-H. Yook and H. Meyer-Ortmanns, *Phys. Rev. E* **72**, 045105(R) (2005).
- [14] L. K. Gallos, C. Song, S. Havlin and H. A. Makse, *Proc. Nat. Acad. Sci.* **104**, 7746 (2007).
- [15] C. Li and P. K. Maini, *Phys. Rev. E* **72**, 045102(R) (2005).
- [16] D. Stauffer, S. M. de Oliveira, P. M. C. de Oliveira and J. S. Sá Martins, *Biology, Sociology, Geology by Computational Physicists* (Elsevier, Amsterdam, 2006).
- [17] T. J. P. Penna, *J. Stat. Phys.* **78**, 1629 (1995).
- [18] D. Stauffer, *Biological Evolution and Statistical Physics*, edited by M. Lässig and A. Valleriani (Springer, Berlin-Heidelberg, 2002), p. 258.
- [19] B. Gompertz, *Philos. T. R. Soc. Lond.* **115**, 513 (1825).
- [20] M. Magdoń-Maksymowicz, A. Z. Maksymowicz and K. Kułakowski, *Theor. Biosci.* **119**, 139 (2000).
- [21] H. Jeong, S. P. Mason, A.-L. Barabási and Z. N. Oltvai, *Nature* **411**, 41 (2001).
- [22] A. Vázquez, A. Flammini, A. Maritan and A. Vespignani, *ComplexUs* **1**, 38 (2003).
- [23] V. Colizza, A. Flammini, A. Maritan and A. Vespignani, *Physica A* **352**, 1 (2005).
- [24] R. Pastor-Satorras, A. Vázquez and A. Vespignani, *Phys. Rev. Lett.* **87**, 258701 (2001).
- [25] <http://plus.google.com>
- [26] S. Scellato and C. Mascolo, in *2011 IEEE Conference on Computer Communication Workshops (INFOCOM WK-SHPS, 2011)*, p. 918.