

FBM: A Flexible Random Walk Based Generative Model for Social Network

Hongke Xia^{1,*} and Xiang Hu^{2,3}

¹Computer School, Beijing Info Science and Tech University, Beijing, 100101, P.R. China; ²State Key Laboratory of Network and Switching, Beijing University of Posts and Telecommunications, Beijing, 100876, P.R. China; ³School of Control and Computer Engineering, North China Electric Power University, Beijing, 102206, P.R. China

Abstract: This paper studied the static and dynamic characteristics of the real social networks as well as their proposed generative models, among which the Butterfly Model [1] is useful while not being flexible enough to generate the social networks with the expected power-law exponent. Therefore, a novel Flexible Butterfly Model (FBM) is proposed based on the Butterfly Model and combined with the Monte Carlo method and a Bayesian Graph Model for the training of the FBM Model is built in order to learn parameters from real social networks. Experiments have shown that the FBM model can adjust the law power exponent of the generated social network effectively by the introduced parameters. Meanwhile, the FBM model also maintains the vast majority of important characteristics that the Butterfly model has.

Keywords: Generative model, Graph model, Random walk, Social network.

1. INTRODUCTION

In recent years, virtual social networks, such as Facebook, Twitter, *etc.*, have emerged in large numbers, and have impacted human society in various aspects profoundly along with the penetrating applications of the internet. As the extension of complex network in human society, the social networks generate the relationships among individuals or groups and the activities taking place in a human society, for example: friendly relationships of individuals, business relationships of companies, or cooperation relationships of authors.

With social entity represented by node and social relationship represented by edge, the social network can be described as a graph, and it possesses some prominent static and dynamic characteristics. Here the question arises as how to establish a reasonable generative model to simulate the formative procedure of the social network. The answers to this question are significant to explain the formative mechanism of the social network, forecast the trend and detect the abnormal behaviors in the social networks.

Finding the characteristics and patterns of social networks is the foundation of constructing generative model, and a method informally named “observations and imitations” is widely adopted by the researchers of the social network; firstly, some characteristics and metrics are established *via* observations on the data of real social networks, then their patterns are sought, further, more possible explanations for the discovered patterns are provided. After that, setting up of reasonable generative model is required to mimic the formative procedure of the social network according to the explanations and sequentially generate the synthetic

social networks complying with those patterns. Currently, studies on generative model of social networks mainly concentrate on the two areas: one is to find new characteristics, metrics and patterns, and the other, to build new generative models. Researchers have found many interesting characteristics and metrics and most of them are based on two original properties of the graph: degrees and distances of nodes. With the accumulation of data and the renovation of methods, researchers are constantly discovering new interesting characteristics and metrics as well as static or dynamic patterns.

Analyzing characteristics and constructing generative model of social networks lay the groundwork for numerous applied researches and have penetrated into many application fields: these contribute to identifying the reputation of the webpage, thereby promoting the performance of search engine [2] and are helpful to determine the importance of customers in the virtual market and priority of recommender in social recommender system [3]; these can help to forecast the trend of social networks and can benefit social network operators to plan the scale and time of purchasing new hardware and storage devices; these can be engaged to detect abnormalities in the social network so as to find and prevent the attack and abuse of social network such as spam links [4]; these can be used to generate data for scientific studies with similar characteristics of the real social network, which cannot only protect privacies in the real data, but can also cheaply produce the experimental data with the different characteristics and scales by adjusting parameters of the generative model, and then, reduce the costs for obtaining the data; it is also possible to generate the small-scale samples of the large-scale social networks to accelerate the experiments, or extrapolate the small-scale social networks to predict the trend further. Therefore, studies on the characteristics of the social network and generative model are hot areas in the social network research.

*Address correspondence to this author at the Computer School, Beijing Info. Science and Tech. University, 100101, P.R. China; Tel: 13717504422; E-mail: kk325@126.com

The main contributions of this paper are: (1) it is concluded that the faults of power law exponent generated by Butterfly Model hold fixed value, and by introducing parameters, a new FBM model can be proposed with adjustable power law exponent. (2) a new random number generator function is derived to adjust the power law exponent of the FBM model taking advantage of the Monte Carlo method. (3) a Bayesian graph model can be built easily learnt by Gibbs Sampling method with the purpose of training the FBM mode.

The contents of this paper are organized as follows: section 2 introduced some found static and dynamic characteristics and their generative models; section 3 analyzed the defects of Butterfly Model; section 4 deduced a novel FBM model on the basis of Butterfly Model and proposed a Bayesian graph model for its training; section 5 designed several experiments to verify the FBM, and the last section drew a conclusion for this paper.

2. RELATED WORKS

In this section, some discovered static and dynamic characteristics, metrics and patterns are first introduced and the main generative models are proposed.

2.1. Researches Related to Characteristics of Social Network

The studies on the characteristics and metrics mainly focus on three aspects: the power-law characteristics that the nodes and edges satisfy, network diameters and community structure. By means of different metrics, researchers have discovered various significant power-law characteristics that edges and nodes meet, variation laws of network diameter and community structures in the social network.

The characteristics of power-law related to nodes and edges include: Degree of nodes follows a power-law distribution [5], the discrete probability distributions of degree $f(d)$ follow $f(d) \propto d^{-\gamma}, \gamma < 0$; Density obeys a power-law distribution [6], that is, at any time t , the number of total edges $E(t)$ and the number of total nodes $N(t)$ in the social network meet $E(t) \propto N(t)^\alpha, \alpha > 0$; The weight of the network follows the power-law distribution [1], namely at any time t , the sum of weights on all edges $W(t)$ and the number of all edges $E(t)$ comply with $W(t) \propto E(t)^\beta, \beta > 1$; Weight of any individual node N follows the power-law distribution [1], the sum of weights on the edges connected to the node W_n and degree of the node d_n submit to $W_n \propto d_n^\theta, \theta > 1$; The number of triangles follows the power-law distribution [7], the number of triangles existing in the social network Δ and the number of nodes involved in the triangles $f(\Delta)$ are obedient to $f(\Delta) \propto \Delta^\sigma, \sigma < 0$; The eigenvalues of the adjacency matrix representing the social network follow the power-laws perfectly [8] and the maximum of eigenvalues follows the power-law distribution too [9]; Moreover at any time t ,

the maximum of eigenvalues $\lambda_1(t)$ and the number of all edges $E(t)$ observe $\lambda_1(t) \propto E(t)^\delta, \delta < 0.5$.

Apart from this, researches have also found some meaningful characteristics in the growth of the social network: with the scale increasing, social networks exhibit small-world phenomenon [10, 11], gradual shrink of the network diameter [12] and clear community structures [13, 14]; the number of nodes in the second and the third connected components maintain relative invariableness [1] and similar to the network traffic, newly arrived edges or weights which exhibit the property of self-similarity in different timescales [15], have burstiness rather than uniform peak, etc.

2.2. Researches Related to Generative Models

The role of the generative model is to simulate and generate the social network with some special characteristics. In order to establish reasonable generative models, substantial studies are conducted to find out the formative mechanism of the social networks. The procedure of exploring new characteristics and the procedure of establishing new generative models have moved forward alternately, and the first try of the generative model named as ER model [16] has been made by Erdos and Renyi, while the ER model is ideal, explanative and not practical; In 1999, Albert and Barabasi put forward the BA model [17] which explained the formative mechanism of power-law as a Yule process. The BA model can partly simulate the power-law characteristics in the social network, as well as the small-world phenomenon, but there are strict restrictions; nevertheless, the BA model laid the foundation for later researches. To generate community structure, based on the BA model, Kleinberg, et al. proposed the Copy Model [18] which mimics a newcomer's behavior when he joins an unfamiliar society; when a new node arrives, it randomly selects a node in the social network to connect, and then, the chosen node continues to copy his friends (relationships of neighbors) to the new node with certain probability. Copy Model cannot only generate the community structure, but can also simulate more power-law characteristics in social networks. Subsequently, Jure, et al. found density power-law and shrinkage of the network diameter and considered that the Self-Organized Criticality [19] existing in the system accounts for its findings, thus they put forward the Forestfire model [6]; Because the power laws could also derive from the self-similarity, the researchers studied the formative procedures of various self-similar phenomena naturally [20, 21], and on these bases, proposed R-MAT model [22], Kronecker graph model [23] and RTG model [24].

Different from others, McGlohon and Kang, et al. proposed the Butterfly model and Community Connect model [25] based on random walk. Besides the different perspectives, two models are the same and they can simulate the majority of the found characteristics and patterns in the social network but there are some obvious shortcomings that the models can only generate social networks with fixed power-law exponents and lack effective methods to fit the models according to real data. Even so, they are still the

novel ones compared with other models. This paper analyzed the lacking of the two models, combined with Monte Carlo method, and proposed a novel Flexible Butterfly Model and its Bayesian probabilistic graphical model for the training.

3. DEFECT ANALYSIS OF THE BUTTERFLY MODEL

3.1. The Procedure Description of the Butterfly Model

The Butterfly Model contains three parameters (P_{link} , P_{step} , P_{host}), and can be simply described as follows:

P1. When a node *node_newer* wants to join the social network, it is assigned with a probability P_{step} following a specific distribution, and then it is transferred to the process **P2**.

P2. A new walk process is restarted with probability P_{host} , and if the restart is successful, a node *node_host* in the social network as the starting point for the new walk process is uniformly selected, transferring it to the process **P3**. If the restart is unsuccessful, a failure occurs in the restart and the formative processes are terminated.

P3. The *node_host* is taken as the current node *node_current*, and moves one step forward with the probability P_{step} repeatedly until the end of this walk process, and then is transferred to the process **P2**. Each step in the walk process can be stated as: first uniformly select a node as the *node_current* in this step from all neighbors of *node_current* in the previous step, and then, establish connections between the *node_newer* with the *node_current* in this step with probability P_{link} .

Sketch of Butterfly Model is shown in Fig. (1).

3.2. The Probability Analysis of Butterfly Model

In terms of the above process, some conclusions can be drawn: the length of a walk process L follows a geometric

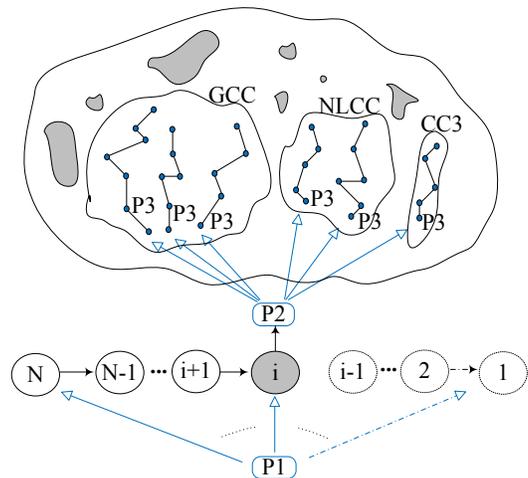


Fig. (1). Butterfly model in conceptual perspective.

distribution $P(L = l) = Geom(l, 1 - P_{step})$, the expectation of L is $E(L) = 1 / (1 - P_{step})$, the time of all walk processes H follows a geometric distribution $P(H = h) = Geom(h, 1 - P_{host})$, the expectation of H is $E(H) = 1 / (1 - P_{host})$, total length of all walk processes LA follows $P(LA = la) = NBin(la, h, 1 - P_{step}) \cdot Geom(h + 1, 1 - P_{host})$, $NBin$ denotes a negative binomial distribution and degree of the new node *node_newer* follows $P(D = d) = NBin(d, h, 1 - P_{step}) \cdot Geom(h + 1, 1 - P_{host}) \cdot P_{link}$, and approximately follows $P(D = d) \approx h \cdot E(L) = h / (1 - P_{step}) = Geom(h, 1 - P_{host}) / (1 - P_{step})$.

Evidently, let P_{link} be constant, given P_{host} and P_{step} following the same distribution, D and L will comply with the same power-law, and if P_{host} is assigned by $E(P_{step})$, and D and L obey the near power-laws. The Butterfly Model in detailed procedure is shown in Fig. (2).

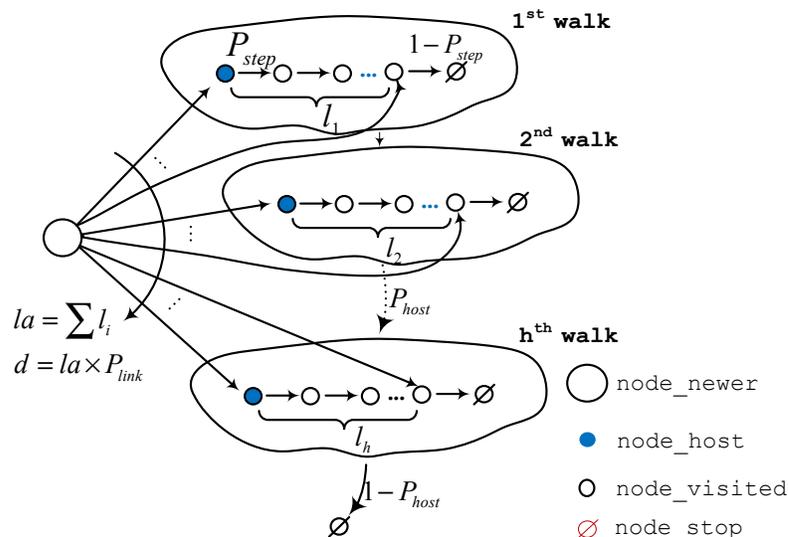


Fig. (2). Butterfly model in detailed procedures.

3.3. Defects and Improvements of the Butterfly Model

Here, defects of the Butterfly model are analyzed, starting from deducing the quantitative relation between the distribution of degree D and length of a walk process L : for the convenience of deducing, P_{step} is relabeled by X , the length of a walk process L is relabeled by Y , and in the Butterfly Model, P_{step} follows the uniform distribution $f_{P_{step}} \sim uniform(0,1)$, and $Y = E(L) - 1 = 1 / (1 - P_{step}) - 1 = x / (1 - x)$.

Now let $y = g(x) = \frac{x}{1-x}$, hence $x = g^{-1}(y) = h(y) = \frac{y}{y+1}$,

and according to the formula of probability density on function of random variables, we have formula (1):

$$f_y(y) = f_x(h(y)) \cdot \left| \frac{dh(y)}{dy} \right| = (y+1)^{-2} \quad (1)$$

Therefore, the Butterfly model can only generate social network with power-law exponent near -2. Multiple edges and the number of restarts occupy a very small part, therefore, the distribution of LA can be approximately given by the distribution of L and the above conclusion is also applicable to the repeatedly walk processes in the Butterfly Model.

Obviously, the power-law exponent generated by the Butterfly Model is not flexible enough to meet the needs of the social networks with mutative power-law exponents. In order to generate social network with flexible power-law exponents, authors have introduced two new parameters to the Butterfly model and then have used the Monte Carlo inverse function method to adjust the probability distribution of X and Y , thus adjusting the power-law exponents of the model. The improved Butterfly Model, named as Flexible Butterfly Model (FBM), can learn parameters from the real data through building a Bayesian graph model, hence appearing to be more practical.

4. FLEXIBLE BUTTERY MODEL

4.1. Description of the Flexible Butterfly Model

The main difference between FBM model and Butterfly model exists in **P1**, in which the new model is no longer subjected to uniform distribution but a deduced special distribution. FBM model may change the generated power-law exponent by changing the parameters of the special distribution, though besides that, **P2** and **P3** are the same.

Assuming that Y wants to follow target distribution: $f_y(y) \sim b \cdot (y+1)^{-a}$ $1 < y < \infty$, a and b are the introduced parameters, a is the power-law exponent, b is magnification factor related to scale of social network (number of nodes in the social network). In the light of Monte Carlo inverse function method, let $x = h(y) = \frac{y}{y+1}$ as primitive function, let $y = h^{-1}(x) = g(x) = \frac{x}{1-x}$ as inverse function, so that X needs to be as follows:

$$f_x(x) = f_y(g(x)) \cdot |g'(x)| = b \cdot (1-x)^{a-2} \quad (2)$$

To generate the above distribution of X , another random variable R can be taken following uniform distribution $f_R \sim uniform(0,1)$. Similarly in line with the Monte Carlo inverse function method, taking $x = \phi(r)$ as primitive function and $r = \phi^{-1}(x)$ as inverse function, we have $f_x(x) = f_r(\phi^{-1}(x)) \cdot \left| \frac{d}{dx} \phi^{-1}(x) \right|$ and $b \cdot (1-x)^{a-2} = 1 \cdot \frac{d}{dx} \phi^{-1}(x)$, then it can be deduced that:

$$\phi(r) = 1 - \left(\frac{a-1}{b} \cdot r \right)^{\frac{1}{a-1}} \quad a > 1, b \geq a-1 \quad (3)$$

As mentioned above that L is relabeled as Y and L and D obey the near power-laws, the graph model of the D 's target distribution can be shown in Fig. (3), and it means that new model can change the distribution of P_{step} by changing parameter a , hence changing the power-law exponents of D . Conversely, the introduced parameters could be trained from degrees distribution through sampling methods, such as MCMC with Gibbs Sampling [26], on graph model. From the above derivation, it is not difficult that: let $a = 2.0$, $b = 1.0$, then P_{step} obeys the uniform distribution between (0, 1), and the FBM model is equivalent to the Butterfly Model. According to the above derivations and explanations, the algorithm description of the FBM model is proposed as shown in Table 1.

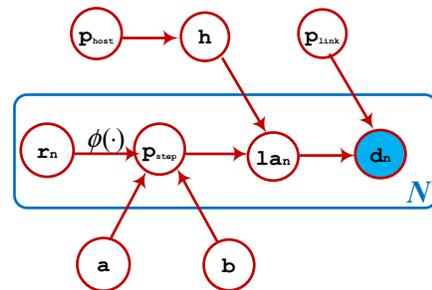


Fig. (3). Graph model of the degree's target distribution.

4.2. Training of the FBM Model

The FBM model takes the modification of the distribution P_{step} as the entry point, by introducing parameter a and b , and has achieved the regulation of the power-law exponent. Previously, in order to simplify the discussion, P_{link} and P_{host} were set in the new model as empirical values in the original model, respectively as 0.3 and 0.5. Now for training the FBM model, P_{host} , P_{link} together with the introduced parameters a , b , four parameters must be learned, so as to establish a complete training model. Based on Fig. (3), the Bayesian graph model is established to train the FBM, and the Bayesian graph model is described via a pseudo-code similar to OpenBUGS [27] syntax, as shown in Table 2.

Table 1. Algorithm description of Flexible Butterfly Model.

P1: getPStepByMonteCarlo(node_newer, a, b)
input: <i>node_newer</i> with uninitialized p_{step} , and a, b is parameters to adapt the exponents of the power lawer output: <i>node_newer</i> with initialized p_{step}
$r \leftarrow Uniform(0,1)$ if $a > 1.0$ and $b \geq a - 1$ then $node_newer.p_{step} \leftarrow 1 - \left(\frac{a-1.0}{b} \cdot r\right)^{\frac{1}{a-1.0}}$ else $node_newer.p_{step} \leftarrow -1$ End
P2: restartWalks(phost, N, a, b)
input: Restart walks with p_{host} , N is total of nodes output: Directed Graph G
$G \leftarrow Graph()$ $N \leftarrow$ the total of nodes for $i \leftarrow 1$ to N do $node_newer \leftarrow Node(i)$ getPStepByMonteCarlo(<i>node_newer</i> , a , b) //call P1 $G.add_node(node_newer)$ while $rand() < p_{host}$ do $node_host \leftarrow G.random_node()$ $travel(node_newer, node_host)$ //call P3 end end return G
P3: travel(node_newer, node_host)
input: <i>node_newer</i> is the newer, the <i>node_host</i> is the starter of a walk process output: link the <i>node_newer</i> to the visited nodes
$node_current \leftarrow node_host$ repeat if $rand() < p_{link}$ then $G.add_directed_edge(node_newer, node_current)$ End if <i>node_current</i> has neighbors then $node_current \leftarrow chooseRandom(node_current.neighbors())$ continue else break end until $rand() > node_newer.p_{step}$

Table 2. Bayesian graph model of the FBM.

Init: $\{N = N_0\}$
Data: $\{d[i] = \text{degrees of real social networks}\}$
Model { $a: Normal(1.0, 10^{-6})$ $b: Normal(2.0, 10^{-6})$ $p_{link}: Normal(0.3, 10^{-6})$ $p_{host}: Normal(0.5, 10^{-6})$ $h: Geom(1 - p_{host})$ for (i in $1 : N$) { $r[i]: Uniform(0, 1.0)$ $p_{step}[i] \leftarrow 1 - \left(\frac{a-1}{b} \cdot r[i]\right)^{\frac{1}{a-1}}$ $la[i]: NBin(1 - p_{step}[i], h - 1)$ $d[i] \leftarrow la[i] \cdot p_{link}$ }

The Bayesian graphical model consists of three parts: part **Init** expresses initialization of model, part **Data** lists the observed data in the model, which is mainly the random sequence of degrees in the social network and part **Model** presents the distributions of random variables and dependencies among the random variables in the model. The Bayesian model takes the degree of nodes D as observed variable, takes other variables as the latent random variable, takes P_{host} , P_{link} , a and b as the parameters to be learnt, assumes that the parameters are subjected to the normal distribution with very small variance, sets the mean of P_{host} , P_{link} as the empirical values of original model, and sets the mean of a , b as 1.0 and 2.0 respectively. h is the random variable representing the number of restarts and follows the geometric distribution with $1 - p_{host}$, N is number of sampling, r obeys the uniform distribution between 0 and 1, p_{step} follows the special distribution defined by $\phi(\cdot)$, la follows the negative binomial distribution with $1 - p_{step}$, and $h - 1$. The training of the Bayesian graph model can use MCMC with Gibbs sampling, calculate the posterior distribution of the parameters and treat posterior mean value of the normal distribution as the learned value of the parameters. The learned parameters can be employed for FBM to generate the social network with the power-law exponents similar to the real social network. Due to limited space, MCMC with Gibbs sampling on the Bayesian graph model and related tools are not discussed here.

5. DESIGN OF EXPERIMENTS

The main purposes of the experiment are to verify the effectiveness of adjusting the power-law exponent in the FBM model and verify whether the FBM model can keep all the main characteristics of Butterfly Mode. Therefore, the experimental data is generated by the FBM model.

Table 3. Estimation of e with the growth of a .

GID		1	2	3	4	5
a		1.6	1.7	1.8	1.9	2.0
e	1-5%	1.42-1.59	1.58-1.69	1.65-1.80	1.81-1.95	1.87-2.01
	mean	1.51	1.64	1.73	1.88	1.94

Table 4. Estimation of e with the growth of b .

GID		1	2	3	4	5
B		1.0	1.3	1.6	1.9	2.2
e	1-5%	1.87-2.01	1.90-1.20	1.91-1.97	1.89-2.01	1.91-1.97
	mean	1.94	1.95	1.94	1.95	1.94

5.1. Validation of Adjustment on Power-Law Exponent

The experiments verify the effectiveness of the adjusting power-law exponent which is estimated by 1-5% normal interval estimates in statistics. Firstly, the experiments assign respectively five different values in (1.6-2.0) to a , and treat b as the constant and maintain the value of 1.0, so that each pair of a and b is used to generate 20 social networks with 10000 nodes repeatedly. Experiments calculate the degree frequency of nodes, and take the degree as x axis and degree frequency as the y axis so as to draw scatter plot under the loglog coordinates. In order to measure the power-law exponent, experiments fit straight lines on the scatter plot by the least squares method and treat the fitted slope e as the estimated power-law exponent. Experiments suppose that e follows normal distribution. With the estimated values of power-law exponent denoted by $e[1]-e[20]$ for each pair of a, b , the 1-5% confidence interval and the mean value of the random variable e are calculated. Experiments take the fifth group as reference, where the FBM model is equivalent to the Butterfly model by having parameters $a=2.0, b=1.0$. The results of experiment are shown in Table 3 and Fig. (4).

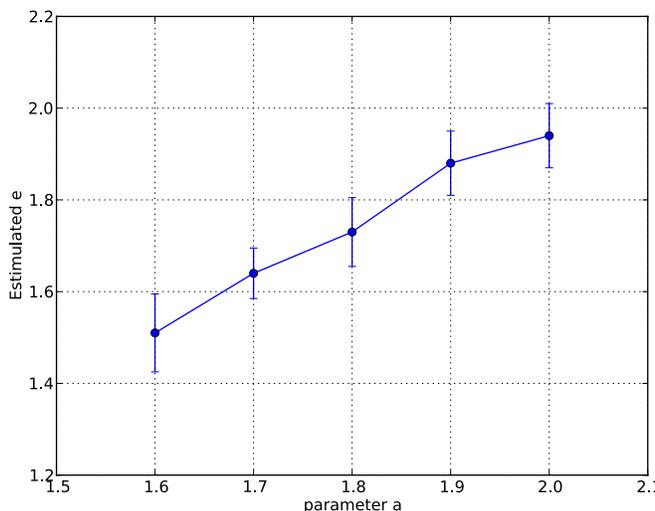


Fig. (4). Estimation of e with the growth of a .

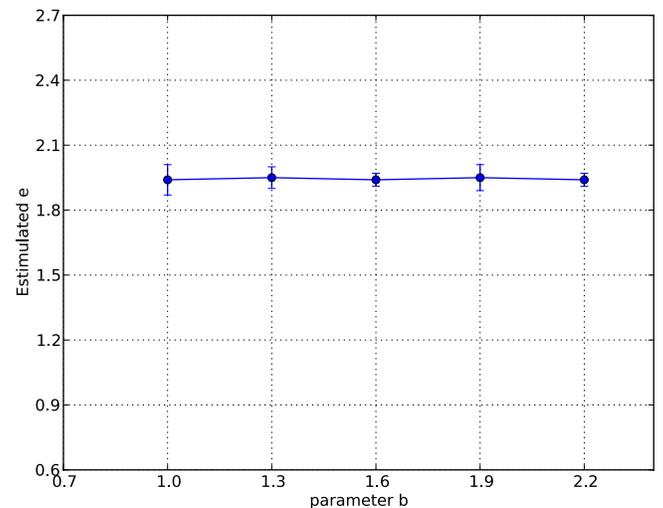


Fig. (5). Estimation of e with the growth of b .

Then, the experiments further test whether parameter b has impact on the power-law exponent. Experiments assign respectively five different values in (1.0, 2.2) to b , hold a to 2.0, perform the processes similar to the previous ones and take the first group as reference. The result is shown in Table 4 and Fig. (5).

Fig. (4) shows that the estimated mean of power-law exponent e increases while parameter a increases and b remains unchanged and Fig. (5) demonstrates that the estimated mean of power-law exponent e essentially remains unchanged while b increases and a remains the same. Therefore, experiments have testified that the FBM model can effectively adjust the power-law exponent through the parameter a . Although the experiments show that parameter b has no obvious effect on the adjustment of power-law exponent, as related to the size of social networks, the parameter b plays important role while fitting the FBM model to various real social networks with different power-law exponents and scales.

5.2. Social Network Characteristics of the FBM Model

Similar to the Butterfly Model in [1], this paper does not verify all the social network characteristics of the FBM, but

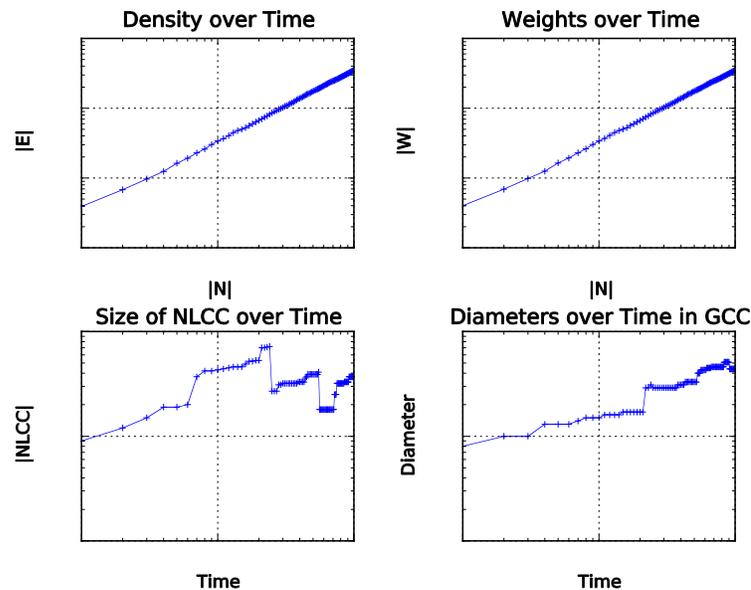


Fig. (6). Dynamic characteristics generated by FBM model.

verifies the four major dynamic characteristics: power law of density, power law of weight, shrinking diameter phenomenon in giant connected component (GCC) and constant size of next large connected component (NLCC). Fig. (6) shows the temporal variations of dynamic characteristics in a FBM-generated social network assigning $a=2.0$, $b=1.0$.

Fig. (6) shows that the density and weight of social networks have approximately linear growth with the nodes under loglog coordinates; and NLCC integrates into the GCC after growing to a certain extent, and its nodes are almost constant; network diameter of GCC increases with time and begins to shrink after the peak. Therefore, the FBM model is not only able to adjust the power-law exponent, but also maintains the major dynamic social networking characteristics that the Butterfly Model has.

CONCLUSION

This paper analyzed the defect of Butterfly model and proposed a novel FBM model based on Butterfly Model by introducing new adjustable parameters. The experiments show that the FBM model has overcome the shortcomings of Butterfly Model, power-law exponent of which can only be near to the fixed value and the introduced parameters may adjust the power-law exponent of generated social networks, making FBM able to simulate social networks with different power-law exponents. The experiments also show that the FBM model maintained some important social network characteristics, such as the power-law of density and weight, network diameter shrinks in the largest connected component (GCC) and that the nodes of next large connected component (NLCC) remain relatively constant.

In order to learn required parameters from the sequence data of the real social network, this article has also established a Bayesian graph model for the FBM model, which makes the FBM model more practical because of its capability of fitting real social networks. Evaluating and validating the performance of Bayesian graph model will be the focus of further work based on this paper.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This work is supported by Beijing Higher Education Young Elite Teacher Project (No. 71A1411117).

REFERENCES

- [1] M.M. Glohon, L. Akoglu, and C. Faloutsos, "Weighted graphs and disconnected components: patterns and a generator," In: *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, New York, NY, USA, ACM, 2008, pp. 524-532. [Online] Available: <http://portal.acm.org/citation.cfm?id=1401955>
- [2] D. Horowitz, and S.D. Kamvar, "The anatomy of a large-scale social search engine," In: *Proceedings of the 19th International Conference on World Wide Web, ser. WWW '10*, New York, NY, USA, ACM, 2010, pp. 431-440. [Online] Available: <http://doi.acm.org/10.1145/1772690.1772735>
- [3] I. Guy, and D. Carmel, "Social recommender systems," In: *Proceedings of the 20th International Conference Companion on World Wide Web, ser. WWW '11*, New York, NY, USA, ACM, 2011, pp. 283-284. [Online] Available: <http://doi.acm.org/10.1145/1963192.1963312>
- [4] Z. Gyöngyi, P. Berkhin, H.G. Molina, and J. Pedersen, "Link spam detection based on mass estimation," Stanford University, California, Tech. Rep., 2005.
- [5] M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," In: *SIGCOMM*, 1999, pp. 251-262.
- [6] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," In: *KDD '05: Proceeding of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, New York, NY, USA, ACM Press, 2005, pp. 177-187. [Online] Available: <http://dx.doi.org/10.1145/1081870.1081893>
- [7] C.E. Tsourakakis, "Fast counting of triangles in large real networks without counting: Algorithms and laws," In: *ICDM. IEEE Computer Society*, 2008, pp. 608-617.
- [8] G. Siganos, M. Faloutsos, P. Faloutsos, and C. Faloutsos, "Power laws and the as-level internet topology," *IEEE/ACM Trans. Netw.*, vol. 11, no. 4, pp. 514-524, 2003. [Online] Available: <http://dblp.uni-trier.de/db/journals/ton/ton11.html#SiganosFFF03>

- [9] L. Akoglu, M. McGlohon, and C. Faloutsos, "Rtm: Laws and a recursive generator for weighted time-evolving graphs." In: *ICDM. IEEE Computer Society*, 2008, pp. 701-706. [Online] Available: <http://dblp.uni-trier.de/db/conf/icdm/icdm2008.html#AkogluMF08>
- [10] S. Milgram, "The small-world problem," *Psychology Today*, vol. 1, no. 1, pp. 61-67, 1967.
- [11] D.J. Watts, and S.H. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, no. 6684, pp. 440-442, June 1998. [Online] Available: http://tam.cornell.edu/tam/cms/manage/upload/SS_nature_smallworld.pdf
- [12] R. Albert, H. Jeong, and A.L. Barabási, "Diameter of the world wide web," *Nature*, vol. 401, pp. 130-131, 1999.
- [13] A. Clauset, M. Newman, and C. Moore, "Finding community structure in very large networks," *Physical Review E*, vol. 70, no. 6, p. 066111, 2004.
- [14] G. Flake, S. Lawrence, C. Giles, and F. Coetzee, "Self-organization and identification of web communities," *Computer*, vol. 35, no. 3, pp. 66-70, 2002.
- [15] M.E. Crovella, and A. Bestavros, "Self-similarity in world wide web traffic: evidence and possible causes," In: *SIGMETRICS '96: Proceedings of the 1996 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Systems*, vol. 24, no. 1. New York, NY, USA, ACM Press, May 1996, pp. 160-169. [Online] Available: <http://dx.doi.org/10.1145/233013.233038>
- [16] P. Erdős, and A. Rényi, "On the evolution of random graphs" *Publication of the Mathematical Institute of the Hungarian Academy of Sciences*, vol. 5, p. 17, 1960.
- [17] A.-L. Barabási, and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, p. 509, 1999. [Online] Available: <http://www.sciencemag.org/content/286/5439/509.abstract>
- [18] J.M. Kleinberg, R. Kumar, P. Raghavan, S. Rajagopalan, and A. Tomkins, "The web as a graph: measurements, models, and methods." In: *COCOON*, 1999, pp. 1-17. [Online] Available: <http://dblp.uni-trier.de/db/conf/cocoon/cocoon99.html#KleinbergKRRT99>
- [19] S.N. Dorogovtsev, A.V. Goltsev, and J.F.F. Mendes, "Critical phenomena in complex networks," *Reviews of Modern Physics*, vol. 80, no. 4, pp. 1275-1335, 2008.
- [20] M. Mitzenmacher, "A brief history of generative models for power law and lognormal distributions." *Internet Mathematics*, vol. 1, no. 2, pp. 226-251, 2004. [Online] Available: <http://www.eecs.harvard.edu/~michaelm/NEWWORK/postscripts/history-revised.pdf>
- [21] M.E.J. Newman, "Power laws, pareto distributions and zipf's law," *Contemporary Physics*, vol. 46, p. 323, 2005.
- [22] D. Chakrabarti, Y. Zhan, and C. Faloutsos, "R-mat: a recursive model for graph mining," In: *SIAM International Conference on Data Mining*, 2004. [Online] Available: <http://www.cs.cmu.edu/~christos/PUBLICATIONS/siam04.pdf>
- [23] J. Leskovec, D. Chakrabarti, J.M. Kleinberg, C. Faloutsos, and Z. Ghahramani, "Kronecker graphs: an approach to modeling networks," *Journal of Machine Learning Research*, vol. 11, pp. 985-1042, 2010. [Online] Available: <http://mlg.eng.cam.ac.uk/pub/pdf/LesChakKleetal10.pdf>
- [24] L. Akoglu and C. Faloutsos, RTG: *A Recursive Realistic Graph Generator Using Random Typing*." In: W.L. Buntine, M. Grobelnik, D. Mladenic, and J. Shawe-Taylor, Eds., *ECML/PKDD (1)*, Springer, 2009, pp. 13-28. [Online] Available: <http://dblp.uni-trier.de/db/conf/pkdd/pkdd2009-1.html#AkogluF09>
- [25] U. Kang, M. McGlohon, L. Akoglu, and C. Faloutsos, "Patterns on the connected components of terabyte-scale graphs," In: *Data Mining (ICDM), 2010 IEEE 10th International Conference*, 2010, pp. 875 -880.
- [26] C. Andrieu, N. de Freitas, A. Doucet, and M.I. Jordan, "An introduction to mcmc for machine learning." *Machine Learning*, vol. 50, no. 1-2, pp. 5-43, 2003. [Online] Available: <http://dblp.uni-trier.de/db/journals/ml/ml50.html#AndrieuFDJ03>
- [27] [Online]. Available: <http://www.openbugs.info>

Received: September 16, 2014

Revised: December 23, 2014

Accepted: December 31, 2014

© Xia and Hu; Licensee Bentham Open.

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.