Who Will Be Your Next Friend: The Bonding Role of Linkage Influence in Social Networks

Cheng Zhang
School of Management
Fudan University
Shanghai, 200433, China
zhangche@fudan.edu.cn

Xu Liu
School of Management
Fudan University
Shanghai, 200433, China
10210690034@fudan.edu.cn

Yunjie Xu
School of Management
Fudan University
Shanghai, 200433, China
yunjiexu@fudan.edu.cn

Heng Xu
School of Information Sciences and Technology
Pennsylvania State University
307C IST Building, University Park, PA 16802
hxu@ist.psu.edu

ABSTRACT
To develop a general mathematical model for social networks is one of the fundamental tasks currently on demand within social network research. Ignoring the strength of the relationships, existing social network models simply use a Boolean value to describe the existence of relationships between peers. This shortage can be overcome by importing repeated social interactions into the model and building the model on a path-based link analysis. In doing this, the authors developed a new semi-random graph model, which offers a general description of the evolution of social networks, with substantial power, to the well accepted hypothesis of preferential attachment in social networks. In addition to these theoretical results, the authors created a quantitative description of the bonding role of social relationship in networks, a parameter within the model denoted as ν. Empirical results indicate that the presented model has a degree of distribution in line with those of real-world networks, which is superior to those of major existing models, and the parameter ν, which essentially represents the cohesiveness of social networks, makes an ideal indicator for the cohesion in social networks.

Categories and Subject Descriptors
J.4 SOCIAL AND BEHAVIORAL SCIENCES

General Terms
Measurement, Theory

Keywords
social network, growth, linkage, path influence, bonding index.

1. INTRODUCTION
Our society is now in the era of digitally connected village when people are increasingly sharing their personal interests, opinions, contacts, job information and other information online.

ICEC '12, August 07 - 08 2012, Singapore, Singapore

People are influenced in turn by the relationships developed from their sharing. The emerging information technologies under the rubrics of Web 2.0 testify the power of online social networking. According to Forrester's recent report, "The Future of the Social Web," the first objective in the development of an online social network is to exploit the social relationships among users, followed by developing social functionality to help users interact with their peers (Owyang et.al, 2009). Because social network structures are important to online commerce, it is important to understand network formation process (Jackson and Rogers, 2007). In online social networks, users are more concerned about online privacy. For example, Facebook users often hide their updates (Armano, 2009). This phenomenon does not actually mean that users are becoming less social, but it is likely that some users behave unfriendly toward strangers online. Will the friendliness between peers affect network growth?

By modeling social network participants as nodes and their relationships as ties with a network graph, this study attempts to answer two questions: (1) How does a social network grow from repeated interactions between peers and (2) how does a member's relationship with his/her peers in the network affect the growth of a network? Supported by large-scale datasets, the authors propose a general network-growth model which offers a good indicator for the cohesion in social networks.

2. THEORETICAL BACKGROUND
A social network can be regarded as a set of actors and the links among them. From the perspective of graph theory, a social network can be viewed as a graph, with its actors and links being nodes and edges of the graph, respectively. Both the nodes and edges can be assigned a specific meaning in various situations. For example, in an academic collaboration network, researchers are nodes and collaborations among them are edges, or journal papers could be viewed as nodes and the citation relationship between two papers are edges. In the World Wide Web, nodes are billions of web pages and edges are the hyperlinks between two web pages.

Developing a quantitative models to explain the formation of network is a fundamental task in social network research. One of the earliest investigations was conducted by Rapport et al. (Solomonoff and Rapport, 1951; Rapport, 1957) who proposed a random network model. Based on that, more rigorous studies were carried out by Erdos and Renyi several years later (1959; 1960;

1961). Since Erdos and Renyi named their random network model "random graph model," the name was continually referred to by succeeding researches (Karonski, 1982; Molloy and Reed, 1995; Janson et.al., 1999; Bollobas, 2001).

One advantage of random graph models is that they reproduce the small-world effect well (Watts and Strogatz, 1998), i.e. the distance between two nodes through the network is $\ell \sim O(\log n)$. However, random graph models have their weakness. For example, other network properties derived from random graph models do not match those of real-world networks, such as the scale-free distribution of degrees (Barabasi and Albert, 1999; Strogatz, 2001; Dorogovtsev and Mendes, 2002). In addition, they assume entirely random mixing of patterns, and there is no correlation between degrees of adjacent nodes, no community structure, and navigation is impossible in a random graph using local algorithms (Newman, 2003).

While random graph models make a strong assumption of social networks, i.e., the randomness in the establishment of links among nodes, the differences between theoretical random networks and real-world networks suggest that they missed something important in modeling social networks. Many alternative network growth models have been proposed (Price, 1965; Barabasi and Albert, 1999; Krapivsky and Redner, 2001; Pennock et. al., 2002; Jackson and Rogers, 2007). Network growth models are different from random graph models. They assume that the growth of a network is by gradual additions of nodes and edges in some manner. Growth models argue that it is the growth process that leads to the structural features of real networks.

In the literature, network growth models share three basic assumptions: (1) the total number of nodes in a network increases as new nodes are added, (2) the formation of new links is a uniform distribution, and (3) preferential attachment is the mechanisms of new link formation. Among them, assumption 3 is of most importance. Preferential attachment means that the probability a node acquires a new link is proportional to the links the node already has (Barabasi and Albert, 1999; Price, 1965) After Barabasi and Albert's study (the BA model hereafter), many researchers have attempted to generalize their model (Krapivsky and Redner, 2001; Pennock et. al., 2002). Preferential attachment predicts network properties that are supported by many real-world networks.

Taking into consideration both randomness and preferential attachment in link formation, Jackson and Rogers (2007) (the JR model hereafter) offered a new model which is consistent with the small-world effect and the scale-free degree distribution in large socially generated networks.

Both BA models and the JR model rely on the assumption of preferential attachment. However, these models do not offer an explanation of why links are formed in this way. Because the underlying sociological reason of link formation is important to many researches, the lack of an in-depth explanation of why preferential attachment occurs is a major drawback of the existing network growth models. Furthermore, when extant network growth models use an adjacency matrix as a mathematical description of social networks, they treat links as binary. Such practice ignores another important nature of social relations, that social relationships are usually built on and maintained through repeated interactions rather than one-shot action. Therefore, it is of great theoretical importance to investigate beyond link formation and look into the link enhancement.

The main social network growth models, their assumptions, and model characteristics can be summarized in Table 1, and key terminologies appeared in the literature is listed in Table 2. As discussed above, extant major models suffer the weakness of, dichotomizing link strength in network, ignoring the theoretical explanation of link formation with a focus on the degree distribution, and relying on an imperfect assumption of preferential attachment. Although random graphs do not exhibit the latter two shortcomings, they are criticized for being 'too random to be true.' These shortcomings have motivated researchers to develop new semi-random graph models.

Table 1. Important social network models

Social Network Models	Sources	Modeling Link Strength	Model Level
Random graph	Erdos & Renyi, 1959	No	Links
Price's Model	Price, 1965	No	Degree distribution
BA model	Barabasi & Albert, 1999	No	Degree distribution
Generalized BA models	Dorogovtsev et. al., 1999	No	Degree distribution
	Krapivsky & Redner, 2001	No	Degree distribution
	Pennock et. al., 2002	No	Degree distribution
JR model	Jackson & Rogers, 2007	No	Degree distribution

Table 2. Key terminologies in the literature

Terminology	Definition
Degree	The degree of a node in a network (sometimes referred to incorrectly as the connectivity) is the number of connections or edges the node has to other nodes. If a network is directed, meaning that edges point in one direction from one node to another node, then nodes have two different degrees, the in-degree, which is the number of incoming edges, and the out-degree, which is the number of outgoing edges.

Degree distribution	The degree distribution P(k) of a network is then defined to be the fraction of nodes in the network with degree k. Thus if there are n nodes in total in a network and n_k of them have degree k, we have $P(k) = n_k/n$.
Accumulated degree distribution	The degree distribution P(k) of a network is then defined to be the fraction of nodes in the network with degree no bigger than k. Thus if there are n nodes in total in a network and n_k of them have degree k, we have $P(k) = \sum_{i \le k} n_i/n$.
Random graph model	A random graph is obtained by starting with a set of n vertices and adding edges between them at random. Different random graph models produce different probability distributions on graphs.
Preferential attachment	A preferential attachment process is any of a class of processes in which some quantity, typically some form of wealth or credit, is distributed among a number of individuals or objects according to how much they already have, so that those who are already wealthy receive more than those who are not. "Preferential attachment" is only the most recent of many names that have been given to such processes. They are also referred to under the names "Yule process", "cumulative advantage", "the rich get richer", and, less correctly, the "Matthew effect".

We propose a model that offers a micro level explanation of link formation, taking into consideration both link formation and link enhancement. We also compare our model with existing models on various performance measures.

3. MODEL DEVELOPMENT

The core logic of our model follows two fundamental observations of social networks: friendship between peers increases as more social interactions occur, and stronger friendship leads to a stronger peer influence. In modeling language, these principles can be stated as (1) tie strength between any two adjacent nodes increases as news links forms between the two nodes, and (2) the influence of one node on another increases as the tie strength increases. The rest of the section describes how the model is developed based on the two fundamental principles.

3.1 Description of Networks

In the proposed model, a social network is viewed as a directed graph, where nodes are actors in the social network and edges are links between actors. We assume that all nodes are added to the graph one by one and each node is assigned an index $i \in \{1,2,\cdots,N\}$ representing its time of entry. N is the total number of nodes. In other words, at time t, there are t nodes in the graph. We will then define a few important concepts, including linkage coefficient, linkage influence and path influence. All variables are included in Table 3.

Table 3. Variables in the model

Notation	Name	Discription
N	Size of the network	The total number of nodes in the network
t	Time step	At time step t, there are t nodes in the network, one of which is new.
i, j, k	Node index	Node i enters the network at time t.
$\mathbf{a_{ij}}$	Linkage coefficient	The total times linkage occurs between i and j
$\mathbf{a_0}$	Initial linkage coefficient	a_0 is introduced to represent the possibility of linkage between two nodes even when the nodes have not established a link yet.
φ	Linkage influence	$\varphi(i,j)$ is the influence node i has on node j as a result of the link from j to i.
θ	Linkage influence scalar	$\theta > 0$
ν	Bonding index	The linkage influence $\varphi(i,j)$ when $a_{ij} = 1$
P	Path	A path $P(i,j)$ from node i to j is defined as a sequence of distinct nodes $(i,k_1,\cdots,k_n,j)$ beginning with i and ending with j.
T	Set of paths	$T(i,j)$ is the set of all paths from node i to node j.
R	Path influence	$R(P)$ is the product of all the linkage influences along P.

V	Network influence	$V(i,j)$ is the sum of all the path influence from node i to node j.
m	Average links	The average number of links a node in the network has

(a) Linkage Coefficient

A graph with N nodes can be expressed by a $N \times N$ matrix $A = (a_{ij})_{N\times N}$, where a_{ij} is the linkage coefficient between nodes i and j. Linkage coefficient a_{ij} refers to how many times link occurs between nodes i and j. Linkage coefficient marks the first difference from other models which considered the edge between two nodes to be binary. In the current model, a larger a_{ij} means that nodes i and j have a stronger linkage.

An initial linkage coefficient a_0 is introduced to represent the possibility of linkage between two nodes even when the nodes have not established a link yet. The introduction of a_0 ensures that a node has a probability to establish a link with another nodes when it is added to a graph. It is reasonable to assume that the initial linkage coefficient is smaller than any existing link coefficient; therefore, $0 \le a_0 \le 1$.

(b) Linkage Influence and Path Influence

For any given nodes i and j, the linkage influence from j to i is defined as a function of its linkage coefficient a_{ij}, i.e.,

$$\varphi(i,j) = \theta(a_{ij} + a_0) \tag{1}$$

where a_0 is the initial linkage coefficient defined above and θ is a scalar. When there is exactly one link from node i to node j, i.e. $a_{ij} = 1$, then the corresponding linkage influence is $\varphi(i,j) = \theta(1 + a_0)$, which is independent of i and j. Define $\nu = \theta(1 + a_0)$, then ν refers to the linkage influence from one node to another when there is just one link. In networks with a large ν, existing links can effectively help nodes gain a new link or enhance the existing linkage coefficient. In this sense, we call ν as the bonding index of the network. Note that, for any models that use binary linkage coefficients, the definition of ν holds valid and it represents the linkage influence with one link between two nodes.

A path $P(i,j)$ from node i to j is defined as a sequence of distinct nodes $(i, k_1, \cdots, k_n, j)$ beginning with i and ending with j, and the path influence $R(P)$ of path $P(i,j) = (i, k_1, \cdots, k_n, j)$ is defined as the product of all the linkage influences along this path, i.e.,

$$R(P) = \prod_{l=0}^{n} \theta(a_{k_l} a_{k_{l+1}} + a_0) \tag{2}$$

Where $a_{k_0} = i$ and $a_{k_{n+1}} = j$. The network influence V, or influence for short, from node j to node i is defined as the sum of the path influence of all the paths from node i to node j, which is,

$$V(i,j) = \sum_{P \in T(i,j)} R(P) \tag{3}$$

Where $T(i,j)$ is the set of all the paths from node i to node j.

3.2 Establishment of Links

A social network evolutes when new nodes join the network or when new links are established. The establishment of links is of much more importance. In this model, we hold that a link is established through a probabilistic process. We also hold that node t is added to the network at time t.

We define p_{ij} as the probability that node i establishes a new link to node j at time $t+1$. Between two nodes i to j, a link might be established under the influence from i to j. For simplicity, suppose that $V(i,j)$ is proportional to p_{ij}, i.e.,

$$p_{ij} \propto V(i,j) \tag{4}$$

Meanwhile, node i could possibly establish a link with any one of the other t nodes. The probability p_{ij} that node i will establish a link with some node j is inversely proportion to the total number of nodes t, i.e.,

$$p_{ij} \propto \frac{1}{t} \tag{5}$$

From (4) (5) , we can have:

$$p_{ij} = \frac{\lambda V(i,j)}{t} \tag{6}$$

where $\lambda > 0$ is a scalar.

3.3 Explanation of Preferential Attachment

Most existing models assume the preferential attachment in the formation of new links (Barabasi and Albert, 1999; Krapivsky and Redner, 2001; Pennock et. al., 2002; Jackson and Rogers, 2007). Preferential attachment means that the probability p of a node i establishing a link with a newly added node is proportional to d_i, the degree of node i. Our model explains the preferential attachment as follows.

Without losing generality, we suppose that each new node forms m links to old nodes when it joins the network, where m is the average number of links a node in the network has.

At time $t+1$, the new node $t+1$ forms m links to existing nodes. According to (6), the probability that node $t+1$ forms a link to node i is

$$p_{ij} = \frac{\lambda V(t+1,j)}{t} \tag{7}$$

Where $V(t+1,j)$ is defined by (3) and (4). For simplicity, we consider only the paths from $t+1$ to i no longer than 2, and we have

$$V(t+1,i) = \theta(a_{t+1,i} + a_0) + \sum_{a_{ki}>0} \left(\theta^2 (a_{t+1,k} + a_0)(a_{ki} + a_0)\right) \tag{8}$$

Note that $a_{t+1,k} = 0$ holds for $\forall k \in \{1, \cdots, t\}$, since node t+1 has no links to any existing node when it is first enters the network, and the in-degree of node i is $d_i = \sum_{k=1}^{t} a_{ki}$ at time t+1. By reorganization of (8) we have

$$V(t+1,i) = \theta a_0 (1 + \theta(1 + a_0) d_i) \tag{9}$$

Based on (7) and (9), the probability of node $i+1$ establishing a new link from $t+1$ is

$$\frac{\lambda\theta a_0}{t} + \frac{\lambda\theta^2 a_0 (1 + a_0) d_i}{t} \tag{10}$$

The first component of equation (10) is uniform to all nodes. It can be regarded as the uniformity in the formation of new links. The second component, $\frac{\lambda\theta^2 a_0(1+a_0)d_i}{t}$, is proportional to d_i, the in-degree of node i, which is exactly what preferential attachment implies. In summary, (10) is a more general expression of link formation by considering both the uniformity and preferential attachment process.

It can be seen that assumption (2) and (3) of existing models listed in Table 1 are subsumed in our model. Based on this micro analysis of links, we can explain preferential attachment. Next we examine the degree distribution implied by our model.

4. DEGREE DISTRIBUTION ANALYSIS

Analysis of network evolution, especially when the structure of the network plays an important role in the evolution, is complicated. Here the authors use "mean-field" approximation to the evolution of social networks. On average, the earlier a node joins a network, the larger is its expected in-degree.

If node $t+1$ establishes a linkage to node $i(i \leq t)$, the in-degree of i, namely d_i, increases by1; otherwise d_i remains the same. That is to say, the expectation of the increment of d_i is $1 \times p_{t+1,i} + 0 \times (1 - p_{t+1,i}) = p_{t+1,i}$. Since d_i is related to time t, we use $d_i(t)$ to represent the in-degree of node i at time t, and thus at $t+1$,

$$d_i(t+1) = d_i(t) + \frac{\lambda\theta a_0}{t} + \frac{\lambda\theta^2 a_0(1+a_0)}{t} d_i(t) \quad (11)$$

Denote $\alpha = \lambda\theta a_0$ and$\beta = \lambda\theta^2 a_0(1+a_0)$, then (11) can be rewritten as

$$d_i(t+1) = d_i(t) + \frac{1}{t}\left(\alpha + \beta d_i(t)\right) \quad (12)$$

By solving $d_i(t)$, we get the degree distribution function as follows:

$$\text{-}F(d) = 1 - \left(\frac{1}{1+\nu d}\right)^{1+1/(\nu m)} \quad (13)$$

Where F(d) is the proportion of nodes with an in-degree no larger than d. Please refer to Appendix A for detailed deduction process.

Of the m links node $t+1$ forms, α links are independent of the existing links, while βm links are directly related to the in-degree of the old nodes, which means these βm links are formed owing to some existing paths. In this sense, we use α to represent the "random connectivity" of the network, and βm the "transitive connectivity". So "the randomness of the network" can be defined as $r = \frac{\alpha}{\beta m}$, which means the ratio of links formed by random connectivity and transitive connectivity. We can rewrite (13) as follows:

$$F(d) = 1 - \left(\frac{rm}{rm+d}\right)^{1+r} \quad (14)$$

From (14), it is clear that if the randomness of the network is defined as proportional to random connectivity and inversely proportional to transitive connectivity, the Degree distribution showed by (13) is to the same as the result of Jackson & Rogers(2007). The difference is that the new model is based on links between nodes, while the JR model (Jackson and Rogers, 2007) is based on the assumption of preferential attachment. Given the equivalence of two models, it is reasonable to conclude that the new model is a more general mode being free from the assumption of preferential attachment and instead offers an in-depth explanation to preferential attachment.

5. EMPRICAL DATA ANALYSIS

To empirically test the applicability of our model, we fit this model to real-world networks. The datasets used are author networks in five academic fields including business, management, management science & operation research, information systems, and information science & library science. The data were collected from Web of Knowledge (WOK), the world's largest organization which stores journal and paper information in various disciplines. We collected the co-authorship of papers published in these fields over the past ten years from 1999 to 2008. We then obtained the degree distribution and average degree m of each network, and fit the degree distribution function of expression 11, in which d_0 is set to 0 for the five networks respectively. The results are shown in Table 3.

Table 4. Results of fitting degree distribution to real-world networks

Networks	Nodes	Avg. in-degree m	Bonding Index ν	Model fit R^2
Business	27724	3.311	0.333	0.985
Management	36394	3.230	0.262	0.988
Management Science and Operation Research	44662	4.085	0.213	0.992
Information Systems	72470	5.251	0.114	0.995
Information Science and Library Science	23161	3.666	0.254	0.976

Table 4 shows that for all five networks, the fit value measured by R^2 is excellent, ranging from 0.976 to 0.995, indicating the degree distribution function in expression 13 fits these real networks well.

We next compared the degree distribution predicted by our model with that of five major models summarize in Table 1. Please refer to Appendix B for degree distribution functions of these models.

Table 5. Comparison of degree distributions of five major models

Networks	Price Model	BA Model	Generalized BA Model	JR Model	Our Model
Business	0.741	0.846	0.974	0.985	0.985
Management	0.789	0.857	0.969	0.988	0.988
Management Science and Operation Research	0.910	0.821	0.962	0.992	0.992
Information Systems	0.926	0.711	0.909	0.995	0.995
Information Science and Library	0.787	0.773	0.938	0.976	0.976

Science					

Note: Random graph model cannot estimate degree distribution

Among them, the random graph model failed to produce any convergent result, hence not reported here . From these results, it is clear that our model excels the Price model and BA models. It is equivalent to the JR models. However, our model is based on the assumption of preferential attachment, but based on a more general assumption of influence.

We also investigated the impact of bonding index on social networks. After calculating the bonding index, v, for author networks of 78 journals in the management category and 68 journals in the business category, we examined the correlations between bonding index and scientific influence indices including 2-year impact factor, 5-year impact factor, Eigenfactor and article influence. 2/5-year journal impact factor is the average number of times an average article in a journal has been cited two/five years since it was published. Eigenfactor measures the number of times articles from the journal published in the past five years have been cited in WOK's journal citation report year. Article influence is the journal's Eigenfactor score divided by the fraction of articles published by the journal. The four indices measure the influence of a journal. They represent the external influence of a journal.

Results show that the bonding index has a significant correlated relationship with most of the influence indices. The negative correlation coefficients mean that more bonded by friends in social networks, the less network influence recognized by others. It has been noted that social relationship may be not always invested towards positive ends. "Bonding" groups can become isolated and disenfranchised from the rest of society without an appropriate balance with "bridging" relationships (Bolin et.al., 2004). As social relationship bonds and stronger homogeneous groups form, the likelihood of bridging social relationships is attenuated. Therefore, the strengthening of insular ties can lead to a cohesive community but a social isolation as well. The empirical evidence from author networks reveals the phenomenon and prove itself to be a good index to measure cohesiveness of social networks.

Table 6: correlation between v and 2-year journal impact factor, 5-year journal impact factor, Eigenfactor and article influence

Correlations Disciplines	2-Year Impact Factor	5-Year Impact Factor	Eigenfactor	Article Influence
Management	-0.286**	-0.291**	-0.162	-0.179
Business	-0.292**	-0.382***	-0.161	-0.304**
Management + Business	-0.292***	-0.324***	-0.155*	-0.228**

*: p<0.1; **: p<0.05; ***: p<0.01

6. CONCLUSIONS

In response to the call for deeper development into the micro details of link formation (Jackson and Rogers, 2007), the current investigation presented a model of network formation from the linkage level that exhibits features matching observed socially generated networks. The model is independent of the hypothesis that preferential attachment is the underlying mechanism in the process of new link formation. Nevertheless, the process of new link formation exhibits all the characteristics under the preferential attachment mechanism. Based on the linkage level analysis, we can explain preferential attachment as a result of network influence, while in prior studies the preferential attachment mechanism is a natural start-point.

Fitting the model to data indicated a wide range in terms of the relative ratio of uniformly random versus network-based meetings in the process of link formation. In addition, the current model shows close fits of the model in terms of a variety of network properties. The authors also demonstrated that the degree distributions and clustering coefficients, coming from the model, fit actual situations well. This can be useful in further applications, since, if there is a well-structured relationship between degree, clustering, and payoffs, then we can relate the network formation process to total societal utility.

The current model and analysis suggest a pressing and interesting answer regarding what accounts for the differences in the network formation process across applications and how preference bias forms.

7. LIMITATIONS

In this study our empirical results are based on the analysis of a historical co-authorship network. The network we analyzed to fit our theoretical model may be not relevant exactly. Fitting the model to friendship networks will be more supportive.

Another limitation of this study is that we did not address the difference of power or strength among the nodes. Since our model is based on linkage influence, assuming all the nodes are equal in power may not capture full information of the network.

We are to address these concerns in our future study.

8. REFERENCES

[1] Armano, D., Six Social Media Trends for 2010, November 2009, Business Week, accessible at The Conversation in Harvard Business Online at Hhttp://blogs.hbr.org/cs/2009/11/six_social_media_trends.htmlH

[2] Barabasi, A.L. and Albert,R. (1999). Emergence of scaling in random networks, Science 286,509-512

[3] Bollobas, B. (2001). Random Graphs, Academic Press, New York, 2nd ed.

[4] Bolin, B., Hackett, E.J., Harlan, S.L., Kirby, A., Larsen, L., Nelson, A., Rex, T.R., Wolf., S. (2004). Bonding and Bridging: Understanding the Relationship between Social Capital and Civic Action. Journal of Planning Education and Research 24:64-77

[5] Cortizo, J.C. CFP: Special Issue of International Journal of Electronic Commerce on Mining Social Media, International Journal of Electronic Commerce, 2009

[6] Dorogovtsev, S.N. and Mendes,J.F.F. (2002). Evolution of networks, Advances in Physics 51,1079-1187

[7] Erdos, P. and Renyi,A. (1959). On random graphs, Publicationes Mathematicae 6,290-297

[8] Erdos, P. and Renyi,A. (1960). On the evolution of random graphs, Publications of the Mathematical Institute of the Hungrarian Academy of Sciences 5,17-61

[9] Erdos, P. and Renyi,A. (1961). On the strength of connectedness of a random graph, Acta Mathematica Scientia Hungary 12,261-267

[10] Jackson, M.O. and Rogers,B.W.(2007) Meeting strangers and friends of friends: How random are social networks? The American Economic Review, Vol.97,No. 3, 890-915

[11] Janson, S., Luczak, T., and Rucinski, A. (1999). Random Graphs, John Willey, New York

[12] Karonski, M. (1982). A review of random graphs, Journal of Graph Theory 6,349-389

[13] Krapivsky, P. L. and Redner, S. (2001). Organization of growing random networks, Phys. Rev. E 63, 066123.

[14] Newman, M.E.J. (2003). The structure and function of complex networks, SIAM Review 45, 167–256.

[15] Molloy, M. and Reed,B. (1995). A critical point for random graphs with a given degree sequence, Random Structures and Algorithms 6,161-179

[16] Owyang, J.K., Bernoff, J., Pflaum, C.N, and Bowen, E., The Future of The Social Web, Forrester Research Report, April, 2009.

[17] Pennock, D.M., Gary,W.F., Steve Lawrence, Eric J. Glover, and C. Lee Giles(2002). Winners don't take all: Characterizing the competition for links on the web, PNAS, vol. 99, no. 8, 5207-5211

[18] Price, D.J. de S. (1965). Networks of scientific papers, Science 149,510-515

[19] Rapoport, A. (1957). Contribution to the theory of random and biased nets, Bulletin of Mathematical Biophysics 19,257-277

[20] Solomonoff, R. and Rapoport,A. (1951). Connectivity of random nets, Bulletin of Mathematical Biophysics 13,107-117

[21] Strogatz, S.H. (2001). Exploring complex networks, Nature 410,248-276

[22] Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks, Nature 393,440-442

9. APPENDIX

Appendix A: Degree distribution function $F(d) = 1 - \left(\frac{1}{1+\nu d}\right)^{1+1/(\nu m)}$ (13).

To solve $d_i(t)$ in Eq(12): $d_i(t+1) = d_i(t) + \frac{1}{t}\left(\alpha + \beta d_i(t)\right)$, a continue function $y(t)$ can be applied as a approximation to $d_i(t)$. Assuming that $y(t)$ increases steadily between time t and $t+1$, we can get the approximate form of Eq(12) as follows:

$$\frac{dy}{dt} = \frac{\alpha+\beta y}{t} \quad \text{(L1)}$$

As t increases from i to N, $d_i(t)$ increases from d_0 to d_i, and accordingly $y(t)$ increases from y_0 to y. When N is very large, we can safely assume that $N-1 \approx N$. By (L1) we get

$$\int_{y_0}^{y} \frac{dy}{\alpha+\beta y} = \int_{i}^{N} \frac{dt}{t} \quad \text{(L2)}$$

Solving (L2) we have:

$$\frac{\alpha+\beta y}{\alpha+\beta y_0} = \left(\frac{N}{i}\right)^{\beta} \quad \text{(L3)}$$

Replace y with d_i, we can get the approximate form of (L3):

$$\frac{\alpha+\beta d_i}{\alpha+\beta d_0} = \left(\frac{N}{i}\right)^{\beta} \quad \text{(L4)}$$

By (L4) we can rewrite the inequality $d_i \le d$ as $\frac{i}{N} \ge \left(\frac{\alpha+\beta d}{\alpha+\beta d_0}\right)^{1/\beta}$, and thus the proportion of nodes with an in-degree no bigger than dis:

$$F(d) = 1 - \left(\frac{\alpha+\beta d_0}{\alpha+\beta d}\right)^{1/\beta} \quad \text{(L5)}$$

Note that $\nu = \frac{\beta}{\alpha}$, (L5) can be rewritten as

$$F(d) = 1 - \left(\frac{1+\nu d_0}{1+\nu d}\right)^{1/\beta} \quad \text{(L6)}$$

Now consider the relation between β and ν. The expectation of the number of the links from node $t+1$ is $\sum_{i=1}^{t} \frac{dd_i}{dt} = \sum_{i=1}^{t} (\alpha + \beta d_i)/t = \alpha + \beta\left(\sum_{i=1}^{t} d_i\right)/t$, where $\sum_{i=1}^{t} d_i = m$ because each new node forms m links and hence the average in-degree of the network is m. Meanwhile, node $t+1$ forms m links as well, and hence we have $\alpha + \beta m = m$, where $\nu = \frac{\beta}{\alpha}$, therefore we get:

$$\beta = \frac{\nu m}{1+\nu m} \quad \text{(L7)}$$

Now, by (L6) and (L7), the degree distribution function can be expressed as $F(d) = 1 - \left(\frac{1+\nu d_0}{1+\nu d}\right)^{1+1/(\nu m)}$. Let $d_0 = 0$ and we get:$F(d) = 1 - \left(\frac{1+\nu d_0}{1+\nu d}\right)^{1+1/(\nu m)}$ (13)

Appendix B: degree distribution functions of social network models

Models	Degree distribution function
Random Graph	$F(d) = e^{-m}\sum_{k=0}^{d}\left(m^{-k}/k!\right)$
Price's Model	$F(d) = 1 - d^{-1-1/m}$
BA Model	$F(d) = 1 - \frac{m(m+1)}{d(d+1)}, d \ge m$
Generalized BA Model	$F(d) = 1 - \left(\frac{2m(1-\alpha)}{\alpha d + 2m(1-\alpha)}\right)^{1/\alpha}$
JR Model	$F(d) = 1 - \left(\frac{d_0 + rm}{d + rm}\right)^{1+r}$
Our Model	$F(d) = 1 - \left(\frac{1 + \nu d_0}{1 + \nu d}\right)^{1+1/(\nu m)}$