CONNECTIONS 27(3): 53-59
http://www.insna.org/Connections-Web/Volume27-3/Ko.pdf

Rethinking Preferential Attachment Scheme: Degree centrality versus closeness centrality[1]

Kilkon Ko[2]
University of Pittsburgh, USA

Kyoung Jun Lee
Kyung Hee University, Korea

Chisung Park
University of Pittsburgh, USA

Construction of realistic dynamic complex network has become increasingly important. One of widely known approaches, Barabasi and Albert's "scale-free" network (BA network), has been generated under the assumption that new actors make ties with high degree actors. Unfortunately, degree, as a preferential attachment scheme, is limited to a local property of network structure, which social network theory has pointed out for a long time. In order to complement this shortcoming of degree preferential attachment, this paper not only introduces closeness preferential attachment, but also compares the relationships between the degree and closeness centrality in three different types of networks: random network, degree preferential attachment network, and closeness preferential attachment network. We show that a high degree is not a necessary condition for an actor to have high closeness. Degree preferential attachment network and sparse random network have relatively small correlation between degree and closeness centrality. Also, the simulation of closeness preferential attachment network suggests that individuals' efforts to increase their own closeness will lead to inefficiency in the whole network.

INTRODUCTION

During the last decade, a considerable number of empirical studies have suggested that "scale-free" is one of the most conspicuous structural property in large complex networks (Barabasi, 2002; Buchanan, 2002; Newman, 2003; Strogatz, 2004; Watts, 2003). A scale-free network is a network whose degree distribution follows a power law, i.e. that the probability of having a node with degree k satisfies $P(k)=k^{-g}$. Common characteristics of the scale-free network are; 1) there are centrally located and interconnected high degree "hubs", 2) an average distance among nodes is small, and 3) a clustering coefficient is high (Barabasi, 2002; Watts, 1999).

While many mechanisms can be used for simulating the scale-free structure (Keller, 2005), the degree preferential attachment assumption has been widely used in both mathematical (S.N. Dorogovtsev & Mendes, 2002) and simulation approaches (Albert et al., 2000; Barabasi & Albert, 1999; Barabasi, 2002; S.N. Dorogovtsev & Mendes, 2002). The attachment rule assumes that actors try to make a tie with other actors who maintain high degree centrality.

But for the popularity of preferential attachment, yet few have questioned whether degree-based preferential attachment is

[1] We are grateful to anonymous reviewers who help us to clarify concepts and assumptions. This research is supported by the Ubiquitous Autonomic Computing and Network Project, the Ministry of Information and Communication (MIC) 21st Century Frontier R&D Program in Korea.

[2] Direct correspondent to Kilkon Ko, 3601 Posvar Hall, Graduate School of Public and International Affairs, University of Pittsburgh, PA 15260, USA, Email:kilkon@gmail.com

appropriate. Social network theory, however, has recognized that indirect relations can be more important than the direct ones. As the degree centrality focuses on the local property of the network structure and overemphasizes the direct relations (Freeman, 1979), it underestimates the importance of indirect relations or the global property of the network structure. For this reason, social network theory has developed additional measures such as betweenness, closeness, information or power centrality. Literatures in scale-free network, however, have not dealt with the appropriateness of degree preferential attachment. The assumption might not be a problem if degree centrality is highly correlated to other centrality measures. To our knowledge, however, there is no study that suggests positive or negative relationship between degree centrality and other centrality measures.

This study aims to bridge this knowledge gaps. When degree centrality was introduced by Shaw (1954), it was so intuitively appealing that network researchers admitted this concept as a fundamental feature to explain the network structure. If an actor occupies a structural location to connect with other actors with many adjacent ties, she can be seen as a major channel of information in communication network. For example, in the friendship network, if a person receives many choices from others, she can be considered as a focal point of friendship network.

However, as Granovetter (Granovetter, 1973) raises the importance of the indirect ties (i.e., the strength of weak ties), social network theory becomes interested in the fact that a high degree is not a necessary condition to be a powerful actor. Unlike Winship's hierarchy model for group classification (Winship 1977), Granovetter's weak tie model looses the condition of intransitivity (Freeman, 1992). It argues that if A and B have strong ties and B and C have strong ties, then it will be enough to assume that A and C are weakly linked. The structural location of weak tie can play an important role in this intransitivity. Burt (Burt, 2000) argues that if a social network is composed of multiple groups which are internally cohesive, many actors can have a high degree and short distance within their cohesive groups. However, an actor who bridges two internally cohesive groups – but externally weakly linked between the two cohesive groups – would have more opportunities in getting information as well as mobilizing embedded resources in a timely manner.

Such insights of social network theory make us rethink the attachment rule. If actors of low degree can reach each other with small number of paths, the comparative advantage of high degree actors in its accessibility to others will not be large. Actors may prefer to choose an influential actor(s) in order to minimize their costs to access information or embedded resources. Thus, the "influential actors" can be actors who have high degree and/or high closeness centrality. When degree and closeness centrality are not highly correlated, we can not say the one is always better than the other (Freeman, 1979). But most studies on current scale free network (Barabasi, 2002; Newman, 2001; Watts, 2003) have used a degree centrality as an attachment scheme without providing detailed grounds. In other words, the existing studies do not pay much attention to the relationship between the degree and closeness centrality in modeling dynamic complex network. The main goal of this study is not to discuss the superiority between degree and closeness as a centrality measure. Instead, this paper attempts to answer the following questions:

> Are degree and closeness centrality highly correlated in scale free network?
>
> If they are (not) correlated, under what conditions are they (not) correlated?
>
> If we use closeness centrality as an attachment scheme in modelling scale-free network, what will be results distinguished from the scale-free network using degree centrality?

In order to answer the above questions, this study simulates[3] the two different types of networks commonly used to define the network topology: random network and degree preferential attachment network. We compare the two networks by looking at the relationship between degree and closeness. In addition to these two networks, we also introduce a closeness preferential attachment network, which shows the structural difference to the other two networks.

To preserve the comparability among different networks, we simulate those three different types of networks having the same number of actors (N=500) and ties (≈1000). According to random graph theory and empirical studies on large networks, the low density networks are more likely to be observed. In order to achieve reliable results, we repeated the simulation 50 times for each type of networks. Since the goal of this paper is to explore new knowledge that is understudied in the literature, we do not attempt to perform a full scale simulation.

In the following sections, we examine three different ways of generating dynamic network: random network, degree preferential attachment network (hereafter DPN), and closeness preferential attachment network (hereafter CPN). After providing simulation results on the relationship between degree and distances, and the structural difference among them, we will present the implications and conclusion.

DISTANCE AND THREE TYPES OF NETWORK

Within a network, several paths may exist between a pair of nodes. In that case, the shortest path between two nodes is called geodesic distance (Wasserman and Faust, 1994). This study uses the geodesic distances as a distance measure for the closeness centrality (Valente & Foreman, 1998). A geodesic distance is infinite if two nodes are not connected to each other. Although one simple way is to consider pairs of

[3] We used SAS 8.2 for simulation and data analyses

nodes reachable only, the weakness of this approach is to underestimate the role of isolated nodes. A better alternative to define the distance between unreachable nodes is to use a size of network, which is used in UCINET and also in this paper. If the network size is *N*, then the maximum path to reach other node will be *N*-1.

The closeness of an actor is measured with the actor's closeness centrality. Let's denote a geodesic distance between two nodes n_i, n_j as $d(n_i, n_j)$. The actor closeness centrality is the inverse of the sum of geodesic distances from actor *i* to the *N*-1 other actors.

$$C_c(n_i) = [\sum_{j=1}^{N} d(n_i, n_j)]^{-1}, \text{ where } i \neq j$$

Also the normalized closeness centrality controlling the network size effect is defined as:

$$C^N{}_c(n_i) = (N-1) * [\sum_{j=1}^{N} d(n_i, n_j)]^{-1}, \text{ where } i \neq j$$

The average of the normalized closeness centrality of all nodes is the reciprocal of the average distance of the network.

Random Network

Random network is the simplest but most widely discussed network form (Chung & Lu, 2001; Erdos & Renyi, 1960). Just as many other real world processes have been effectively modeled by appropriate random models, random network provides useful insights to understand complex networks (Aiello et al., 2000). We can generate random network in two different ways. One way is to start with couple of nodes at an initial stage and have new nodes connected to existing nodes randomly. The other way is to create a network connecting two arbitrary nodes with equal probability p after fixing the total number of nodes (N). While the former approach is more appropriate for describing dynamic growth of network, the latter classical random network is easier to mathematically operationalize. Thus, we use the classical random network in the following.

Theoretically, the random network will have a total number of links

$$L = p * [N(N-1)/2]$$

The degree distribution, a probability that a certain node has k degree, follows binomial distribution defined:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

And the average degree is:

$$\bar{k} = p(N-1) \approx pN$$

If the N is large enough, the degree distribution will follow the Poisson distribution:

$$P(k) = e^{-\bar{k}} \bar{k}^k / k!$$

which approximates a normal distribution when N is large. The density of random network is the same to p because density is defined as

$$\Delta = \frac{L}{N(N-1)/2} = \frac{2L}{N(N-1)}.$$

The average distance of a random network will be given by:

$$\bar{\ell} \sim \ln N / \ln[pN]$$

(S.N. Dorogovtsev & Mendes, 2002,15). Compared to degree and density, the distance of random network is not affected so much by the size of network (N). To examine the impact of the size of network and the probability of attachment, we simulate the random network 50 times by changing N, and p. After controlling the p, we analyze the impact of network size on closeness centrality measured as the reciprocal of distances. Figure 1 shows that the impact of the size of network on the closeness decreases significantly as the size of network grows. In contrast, average distance is a decreasing function of p. As the first order derivative of average distance with respect to p is

$$\frac{Log(N)}{pLog(NP)^2},$$

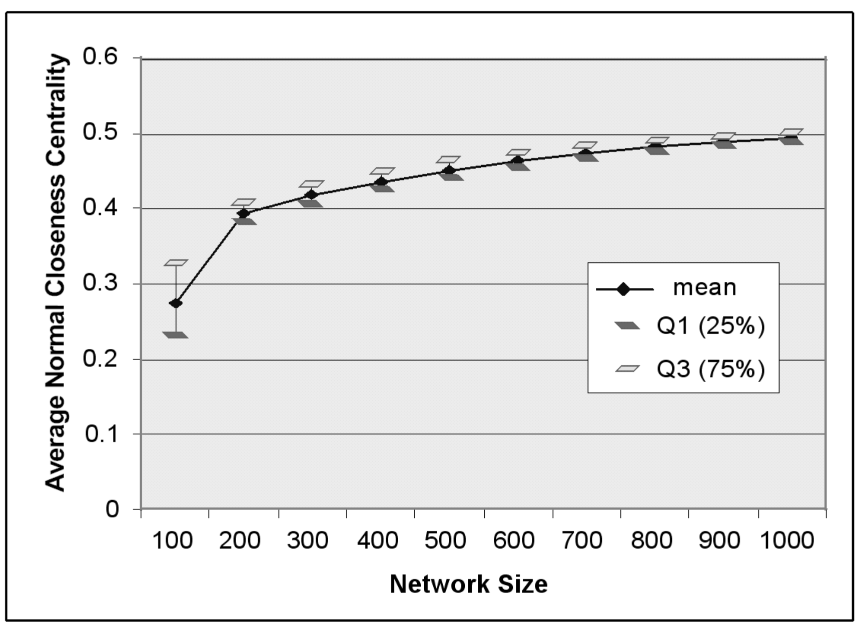

Figure 1. The positive relationship between size of network and the normalized closeness

average distance is a monotonic decreasing function of p. After controlling the size of network, we simulate random network to observe the impact of the p on closeness centrality.

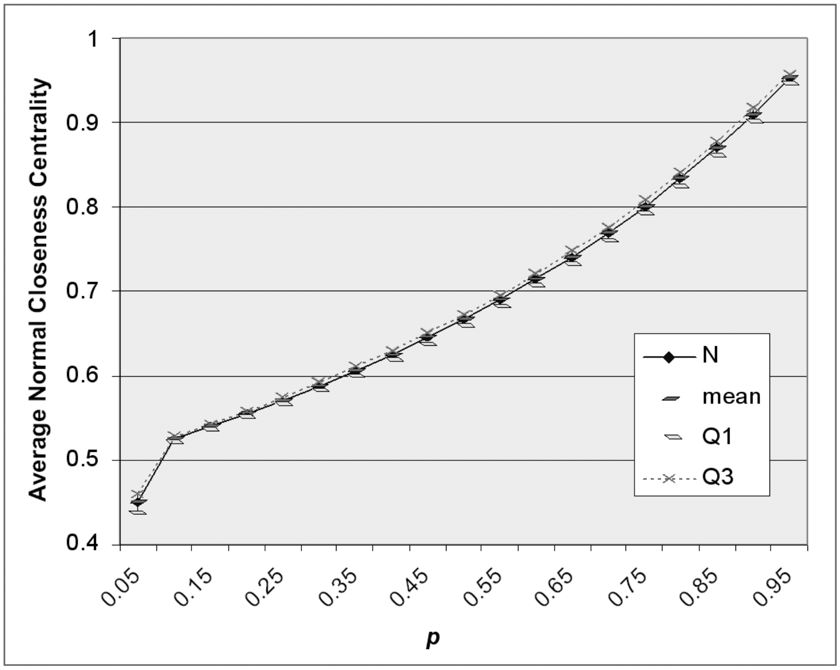

Figure 2. The impact of probability of attachment on normalized closeness

As the probability of a link increases, the average distances between nodes decrease as shown in Figure 2.

Finally, we analyze the relationship between degree and closeness. The relationship between degree and closeness can be measured by the Pearson correlation coefficient. Figure 3 shown below illustrates the correlation between closeness centrality and degree centrality of actor by changing p from 0.005 to 0.101. The correlation is high when p exceeds a critical value (greater than 0.025 in our simulation) regardless of size of network. In contrast, when p is small, the correlation coefficient is small regardless of the size of network.

An interesting question at this point is whether a network which has small correlation coefficient between degree and closeness centrality is rare in the real network or not. The question is related to the size of probability of making ties between actors, i.e. p. According to empirical studies (Newman, 2003), the average distance of social network such as film actors (N=449, 913), CEOs, academic coauthorships (N=52,905~ 253,339), and e-mail message networks (N=59,912) ranges from 3.48 to 16.01. We can infer the probability of attachments p using the size of network and the average distance. As the average distance of random

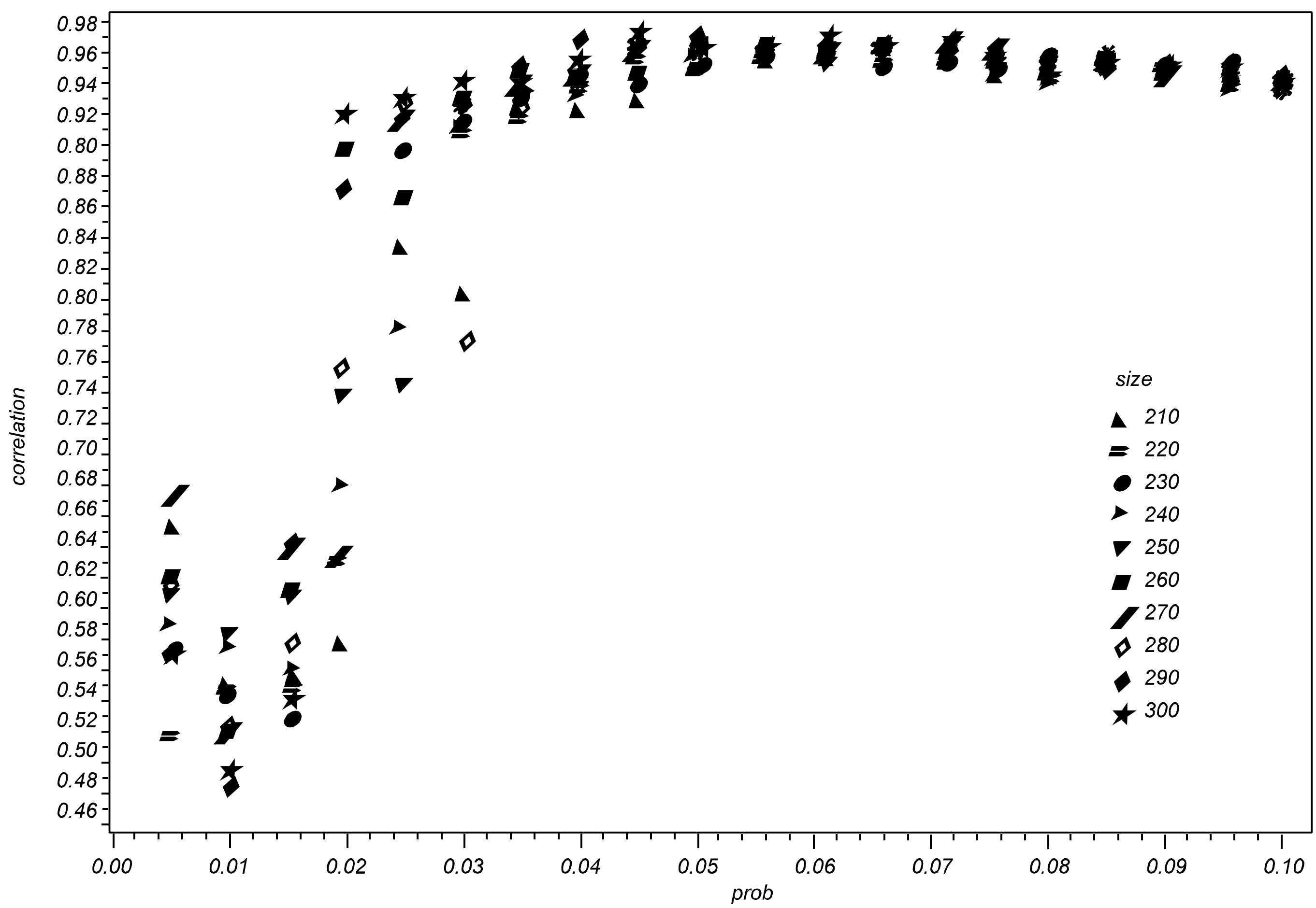

Figure 3. Correlation between closeness and degree centrality in random networks

network is $\overline{\ell} \sim \ln N / \ln[pN]$, the large p (e.g. p=0.101) will lead less than 2 degree average distance of network when N is greater than 100. Even if we have the small average distance such as 3.48 from the network size from the network size N=100,000, *p* is 0.0003 which is very small enough to have small correlation coefficients.[4] Thus, networks whose degree and closeness are weakly correlated can be frequently found even though small world effect exists. In sum, degree and closeness centrality are not correlated in low density networks.

Degree Preferential Attachment Network (DPN)

During the last decade, a preferential attachment has been regarded as a basic rule governing the formation and evolution of real networks (S.N. Dorogovtsev & Mendes, 2002; Newman, 2003; Watts, 2003). The widely used algorithms for generating the preferential attachment network are (Barabasi, 2002,86):

> Growth: For each given period of time, adding a new node to the network
>
> Degree preferential attachment: Each new node connects to the existing nodes with two ties. The probability that it will choose a given node is proportional to a degree the chosen node has.

This paper starts from a highly cohesive seed network with size 10 having the probability of making ties 0.9 and following the same rule of degree preferential attachment[5]. The most distinguishing feature as compared to the random network is that the degree preferential attachment network has a heavier tail than random network. It has been known that the probability that DPN has a node with *k* degree follows a power law with a form of $p(k) \sim k^{-\alpha}$. Although the fluctuations in the tail of degree distribution are not small (SN Dorogovtsev & Samukhin, 2002), our simulation produces pretty stable maximum degree distribution. DPN and the random network have a similar average degree, 2 because, for a random network, the expected average degree is Np=500*0.004 =2. Also our simulation of DPN has an average degree 2.1. However there exists an actor having maximum 45 in-degrees in DPN, whereas there is an actor having less than 5 in-degrees in the random network.

To compare the efficiency of network, we calculate the average distance. As the long average distance implies the less efficient in accessibility, we use the term 'efficiency' to refer to the closeness in this paper. The average distance of the random network is 8.97 but DPN has 4.6. If the size of network and the number of ties are equal, we can confirm DPN is more efficient than the random network.

DPN also does not have a strong correlation between degree and closeness centrality. If high degree does not guarantee high closeness, it brings out an important question: what if actors want to make a tie to others who have higher closeness rather than a high degree? We will call this preferential attachment scheme as a closeness preferential attachment.

Closeness Preferential Attachment Network (CPN)

Unlike DPN, CPN uses the closeness as a weight of attachment instead of degree. It follows, then, a new actor makes ties with high closeness actors with high probability. Other conditions are the same to those used in DPN.

Under the assumption of high correlation between degree and closeness centrality, the network structure of DPN and CPN should not be different. DPN and CPN, however, do show different structures. First, the correlation between degree and closeness in CPN is larger than that in DPN. As seen in Table 1, the correlation coefficient between in-closeness and in-degree is 0.47 in DPN but 0.67 in CPN. This result may be because the attachment rule of CPN is based on the closeness centrality.

Table 1. Correlation coefficients between closeness and degree

		In Degree	Out Degree
In Closeness	DPN	0.47325	0.43704
	CPN	0.67421	0.56243
Out Closeness	DPN	0.42947	0.46331
	CPN	0.55853	0.66811

Note: All coefficient are statistically significant under alpha= 0.05

Second, the average distance of CPN is larger than that of DPN as shown in Table 2 below. As DPN has the same number of actors and ties, the longer distance of CPN implies that CPN is less efficient than DPN.

Table 2. Average distance of Networks

	In distance	Out distance
Random Network	8.97	
DPN	4.46 (0.86)	4.46 (0.97)
CPN	6.43 (1.03)	6.43 (1.22)

Note: Standard deviation is in parenthesis, the size of network is 500.

The result is paradoxical in that we assume that individual actors try to make ties based on closeness centrality to reduce distance to others. While individuals make their ties to high closeness actor to reduce his or her distance to others, these

[4] Of course, the empirical networks are not random networks. So, the probability of making ties based on random network assumption might not be applicable. However, when we measured the probability of making ties in empirical networks using the size of network and the number of edges, we also get very small probability. For instance, film actor network whose average distance is 3.48, has 449,913 nodes and 25,516,482 edges. As the probability of making ties are the same to density, p is 0.000126.

[5] This paper repeats the simulation by 50 times to control the random effects.

efforts of each individual, however, are not transformed into increasing efficiency of the overall system.

Such result comes from the fact that the closeness centrality is sensitive to other actors' structural position. The closeness centrality is interdependent to other actors and ties. It does not simply depend on one actor's own choice. When the dynamic network is constructed, network is not evenly distributed. Compared to random network, DPN and CPN have higher clustering coefficient which implies the emergence of subgroups within a network. When internally cohesive subgroups emerge, an actor with small degree but high closeness centrality such as a cutpoint becomes important in the network. Although DPN puts little weight on the cutpoint in the attachment process, CPN puts much weight on it. The structural advantage of cutpoint in CPN, however, will decrease as new ties are added between two subgroups. As a result, CPN will have more equally distributed degree distribution than DPN. The maximum degree of DPN and CPN in Table 3 shows that CPN can not have an extremely high degree node.

Table 3. Maximum In and Out Degree of Networks

	In Degree	Out Degree
Random Network	5	
DPN	45	52
CPN	18	20

In sum, compared to DPN, CPN has higher correlation coefficient of degree and closeness centrality. Individual actors make their own best choices by connecting themselves to high closeness centrality actors and reduce distance to others. However, the individual's best choice does not guarantee the whole network's efficiency. As our results show, CPN has lower efficiency than DPN. The low efficiency of CPN mainly comes from the fact that the high degree actors in CPN have relatively small number of their neighbors.

CONCLUSION AND DISCUSSION

While BA's "scale-free" network has used degree centrality as a basic preferential attachment rule, it is only one of possible measures of important actors (Bonacich, 1987; Scott, 2000; Stephenson & Zelen, 1989; Wasserman & Faust, 1994). Our study shows that high degree nodes are not necessarily high closeness centrality nodes in a sparse random network or a degree preferential attachment network. In particular, the correlation between degree and closeness centrality in degree preferential attachment network is not so large.

We also simulated and compared the structural features of three different types of networks: random network, DPN (degree preferential attachment network), and CPN (closeness preferential attachment network). The results revealed that CPN has distinctive characteristics different from both DPN and random network. CPN showed relatively higher correlation between degree and closeness centrality than DPN and random network. A noticeable finding is that CPN had longer average distances than DPN, although individual actors in CPN tried to minimize distance to others. Within a closeness-oriented scheme, highly centered actors are less likely to emerge than in a degree-oriented network, as shown in Table 4.

Table 4. Summary of three models by average distance and maximum degree

<table>
<tr><td></td><td></td><td colspan="3">Average distance</td></tr>
<tr><td></td><td></td><td>Low</td><td>Medium</td><td>High</td></tr>
<tr><td rowspan="3">Maximum degree</td><td>Low</td><td>–</td><td>–</td><td>Random Network</td></tr>
<tr><td>Medium</td><td>–</td><td>CPN</td><td>–</td></tr>
<tr><td>High</td><td>DPN</td><td>–</td><td>–</td></tr>
</table>

Although scale-free networks are a useful framework to understand the property of complex networks such as hyperlink network, power grid, telecommunication, internet, or biological network, it is not clear how such networks are formed and evolve. The ambiguity of evolutionary mechanism is mainly due to the lack of knowledge about the possible impetus of networking. As our study proposes, one may prefer to be connected to high degree nodes but others may prefer to be connected to high closeness nodes. Thus, in future studies, we should pay more attention to the relationship among various measures of preferential attachment. At the same time, we have to analyze how different attachment schemes generate different network structures.

REFERENCES

Aiello, W, Chung, F, & Lu, L. 2000. A random graph model for massive graphs. Paper presented at the The 32d Annual ACM Symposium on Theory of Computing.

Albert, R, Jeong, H, & Barabasi, A-L. 2000. Error and attack tolerance in complex networks. *Nature*, 406: 378-381.

Barabasi, A-L. 2002. *Linked:The new science of networks*. Cambridge:MA: Perseus Publishing.

Barabasi, A-L, & Albert, R. 1999. Emergence of scaling in random networks. *Science*, 286: 509-512.

Bonacich, P. 1987. Power and centrality: A family of measures. *American Journal of Sociology*, 92: 1170-1182.

Buchanan, M. 2002. *Nexus: Small worlds and the groundbreaking science of networks* (1st ed.). New York: W.W. Norton.

Burt, RS. 2000. The network structure of social capital. In R. I. Sutton & B. M. Staw (Eds.), *Research in organizational behavior*, volume 22. Greenwich, CT: JAI Press.

Chung, F, & Lu, L. 2001. The diameter of random sparse graphs. *Advances in Applied Math.*, 26: 257-279.

Dorogovtsev, SN, & Mendes, JFF. 2002. Evolution of networks. *Advances In Physics*, 51: 1079-1146.

Dorogovtsev, SN, & Samukhin, AN. 2002. Mesoscopics and fluctuations in networks. cond-mat/0211518.

Erdos, P, & Renyi, A. 1960. On the evolution of random graphs. *Publ. Math. Inst. Internat. Acad. Sci.*, 5: 17-61.

Freeman, LC. 1979. Centrality in social networks conceptual clarification. *Social Networks*, 1(3): 215-239.

Freeman, LC. 1992. The sociological concept of "group": An empirical test of two models. *American Journal of Sociology*, 98(1): 152-166.

Granovetter, MS. 1973. The strength of weak ties. *American Journal of Sociology*, 78(6): 1360-1380.

Keller, EF. 2005. Revisiting scale-free networks. *BioEssays*, 27(10): 1060 - 1068.

Newman, MEJ. 2001. Clustering and preferential attachment in growing networks. *Physical Review E*, 64(025102).

Newman, MEJ. 2003. The structure and function of complex networks. *SIAM Review*, 45: 167-256.

Scott, J. 2000. *Social network analysis: A handbook* (2nd ed.). London; Thousands Oaks, Calif.: SAGE Publications.

Stephenson, K & Zelen, M. 1989. Rethinking centrality: Methods and examples. *Social Networks*, 11(1): 1-37.

Strogatz, SH. 2004. *Sync: The emerging science of spontaneous order*. Hyperion.

Valente, TW, & Foreman, RK. 1998. Integration and radiality: Measuring the extent of an individual's connectedness and reachability in a network. *Social Networks*, 20: 89-105.

Wasserman, S, & Faust, K. 1994. *Social network analysis: Methods and applications*. Cambridge: Cambridge University Press.

Watts, DJ. 1999. *Small worlds: The dynamics of networks between order and randomness*. Princeton, N.J.: Princeton University Press.

Watts, DJ. 2003. *Six degrees: The science of a connected age* (1st ed.). New York: W.W. Norton.