Technology Management, Economics and Policy Papers

SEOUL NATIONAL UNIVERSITY
College of Engineering
Technology Management, Economics and Policy Program

The Impact of the Subgroup Structure on the Evolution of Networks

An Economic Model of Network Evolution

Kibae Kim, Jörn Altmann, and Junseok Hwang

TEMEP Discussion Paper No. 2010:56

서 울 대 학 교
공 과 대 학, 기술경영경제정책 대학원과정
151-742 서울시 관악구 관악로 599

Technology Management, Economics and Policy Program
College of Engineering, Seoul National University
599 Gwanak-Ro, Gwanak-Gu, Seoul 151-742, South-Korea
Phone: ++82-2-880-9140, Fax: ++82-2-880-8389

Technology Management, Economics and Policy Program (TEMEP) is a graduate program at Seoul National University. It includes both M.Sc. and Ph.D. programs, which are integrated into the graduate programs at the College of Engineering.

The *TEMEP Discussion Paper Series* is intended to serve as an outlet for publishing research about theoretical, methodological, and applied aspects of industrial economics, especially those related to the institute's areas of specialization, namely management of technology, information and communication technology, telecommunication, services, health industries, energy industries, as well as infrastructures for development. In particular, paper submissions are welcome, which analyze current technology and industry related issues, discuss their implications, and suggest possible alternative policies. The objective is to gain insights into important policy issues and to acquire a balanced viewpoint of policy making, technology management, and economics. It will enable us to identify the problems in industries accurately and to come up with optimal and effective guidelines. In addition, another important aim of this series is to facilitate communication with external research institutes, individual researchers, and policy makers worldwide.

Research results disseminated by TEMEP may include views on policies for information and communication technologies, technology management, telecommunication, energy, health, and development economics, but the institute itself takes no institutional policy position. If appropriate, the policy views of the institute are disseminated in separate policy briefs. Thus, any opinion expressed in the TEMEP Discussion Paper Series is those of the author(s) and not necessarily the opinion of the institute.

Finally, the current editor of this series would like to thank Prof. Dr. Almas Heshmati for establishing this working paper series in January 2009 and for guiding its development during the course of the first 55 issues. In particular, Prof. Heshmati provided invaluable contributions for setting up the local IT infrastructure, registering the series with RePEc, disseminating the working papers, and setting the quality requirements.

Jörn Altmann, Editor

Office: 37-305
Technology Management, Economics and Policy Program
College of Engineering
Seoul National University
599 Gwanak-Ro, Gwanak-Gu
Seoul 151-742
South-Korea

Phone: +82-70-76787-6676
Fax: +1-501-641-5384
E-mail: jorn.altmann@acm.org

The Impact of the Subgroup Structure on the Evolution of Networks

An Economic Model of Network Evolution

Kibae Kim, Jörn Altmann, Junseok Hwang

Department of Industrial Engineering
College of Engineering
Seoul National University
Seoul, South-Korea
email: kibaejjang@gmail.com, jorn.altmann@acm.org, junhwang@snu.ac.kr

March 2010

Abstract: One of the most important properties of self-organized networks is their scale-free property. Prior research proved empirically and theoretically that scale-free networks emerge under the preferential attachment rule. However, a few empirical studies also show that empirical networks diverge from the structure of scale-free networks. Empirical networks exhibit a lower exponent of the power law distribution than constructed scale-free networks. Our research aims at establishing a simple evolutionary network model that explains this difference. The results of our model suggest that there are two reasons for this discrepancy. First, as already known, additional links between existing nodes distort the scale-free feature. Second, boundaries between subgroups (groups of network nodes) distort the degree distribution. In general, we believe that our evolutionary model may be applicable not only to describe the structural evolution of networks but also to make network design recommendations in a variety of areas such as WWW-hyperlink networks, business collaboration networks, Peer-To-Peer Networks, and Web2.0 service networks.

Keywords: Social Network Analysis, Scale-Free Networks, Self-Organization, Evolutionary Model, Network Design, Network Science, Network Modeling.

JEL Classification Numbers: C02, C15, C65, D02, D85, L11, L14, L86, M15, M21.

1. Introduction

It can be assumed that the probability of gaining a link is proportional to the number of links that a network node has [1] [2] [3]. The rule, which considers this probability for creating links, is called preferential attachment rule. The application of the preferential attachment rule lets a network evolve into a scale-free network. A scale-free network is a network, in which the frequency $p(k)$ of the node degree k decreases under the power function of the node degree [1] [2] [3] [4] [5]. That can be expressed as $p(k) \sim k^{-\gamma}$, where γ is the exponent of the power function. In general, it describes a network, in which a few nodes (i.e. hubs) have high degrees and the majority of nodes has a low degree.

As empirical studies and numerical simulations of the network evolution under the preferential attachment rule have shown, the degree frequency distribution of a typical self-organized network is a decreasing straight line in log-log scale. The exponent γ of the node degree frequency distribution of these networks ranges between *2* and *3* [6] [7]. For example, the hyperlink network of the WWW evolves to a scale-free network with an exponent $\gamma=2.1$ [1] [6].

The shape of the degree frequency distribution and the exponent γ (if it obeys the power law) represents the characteristic of the network topology. For example, a power law decay term yields a stronger fat tail than the power law with an exponential term [8]. Additionally, if the exponent γ in the power law is small, a few hubs dominate the network. If it is large, nodes with a medium and small number of degrees dominate the network. Thus, the characterization of the exponent of a scale-free network provides information about how the node degree frequency is distributed within the network.

However, two empirical analyses found that the exponent of some networks is significantly smaller than *2*. The hyperlink network of Wikipedia shows a low exponent γ (about *1*) of the node degree distribution (though not explicitly mentioned in the work of Hendler et al.) [9]. The exponent γ of the Web2.0 service network was even lower than that, showing an exponent of about *0.5* [10].

This paper aims at establishing a simple evolutionary model, which can explain the observation about the hyperlink network of Wikipedia and the observation about the Web2.0 service network. The model considers two factors that cause this distortion: First, as already known, the additional links between existing nodes of a scale-free network; second, the characteristics of subgroups within a network. A *subgroup* is defined as a group of network nodes. Each node of a network has to belong to one but only one subgroup.

In particular, with respect to subgroups, we investigate how the limitation of link creations (Subgroup-Activity-Factor) between nodes of different subgroups impacts the network evolution. We also examine how the network evolution is impacted by the number of subgroups.

The results obtained from our simulations show that the evolution of networks is strongly impacted by external, none-technical factors, which can be formalized through the concept of subgroup structures. By understanding the specifics of those subgroups (i.e.

openness and the number of subgroups), counter actions can be applied. With respect to the Web2.0 service network, our results show that companies, which own popular Web2.0 services or own the first-mover Web2.0 services, can build subgroups that are highly beneficial to them. These subgroups can even be used strategically against competitors by closing the access to these Web2.0 services for competitors through, for example, the provisioning of proprietary interfaces.

The remainder of this paper is organized as follows: In Section II, we describe prior research on network evolution and establish two propositions related to the impact of subgroups. In Section III, we introduce our model and, based on that, a modified version of the preferential attachment rule is introduced. The propositions and the model are tested through simulations in Section IV. In particular, Section IV describes the results of the simulation and the analysis. Section V concludes with a discussion of the implications of our findings.

2. Scale-Free Networks, Their Construction, and Distortions

2.1. Network Evolution Models

One of the characteristic of self-organized networks is their scale-free network structure [1]. Scale-free networks appear in a variety of areas such as academic co-authorships, the World Wide Web, and blogs [1] [7] [11].

The degree distribution of nodes of these scale-free networks obeys a power-law [1]. The exponent γ of the power-law in real, scale-free networks generally ranges between 2 and 3 [6] [7]. The simulation of the preferential attachment rule showed that the network grows into a scale-free network whose exponent γ is 3 [1]. The authors, Barabási and Albert, assumed in their preferential attachment model that a new node, which is inserted every time unit, connects with an existing node i with a probability that is proportional to the degree of node i.

The degree distribution of a scale-free network is supposed to show a decreasing linear curve in the log-log coordinates plane. Measuring the scale-free property of an empirical network, however, causes problems. While the measurement points, which describe the curve, fit well to a straight line in the area of small degrees, they vary considerably in the area of large degrees. The fluctuations in the area of large degrees result from the small degree frequencies.

To resolve this problem of fluctuations at the high degree area, Newman suggest not to calculate the distribution function $p(k)$ but to use the cumulative distribution function $P(k)= \int p(k`)k`\, dk`$, where the integral ranges from k to ∞ [7] [12]. This technique can show the scale-free structure of a network, even if the degree frequencies are low. If $p(k) \sim k^{-\gamma}$, then $P(k) \sim k^{-(\gamma-1)}$. Thus, γ can be derived from the estimated exponent of the cumulative distribution function.

2.2. Links between Existing Nodes

A rule, which describes the link creation between existing nodes, was added to the original preferential attachment rule in prior research by Barabási et al. [2] [13]. It was designed to explain the evolution of networks, whose agents (nodes) are active over a long time period. In this model, besides the new agents entering the network, the existing agents can also generate new links.

These models are reasonable for empirical networks such as peer-to-peer networks, Web2.0 service networks, and academic co-authorship networks [2] [10] [14] [15] [16]. In an academic co-authorship network, scholars publish articles together with different peers throughout their career, from the moment the scholar enters the academic world until retirement [2] [14]. In peer-to-peer networks, links to other peers are constantly added or replaced depending on the content available at certain peers [15] [16].

In the Web2.0 service network, users compose new Web services with existing Web2.0 services, which offer open APIs (Application Programming Interfaces) [10]. These compositions of new Web2.0 services make the Web2.0 service network self-organized. Thus, we can expect that the Web2.0 service network has similar characteristics as other self-organizing networks (i.e. scale-free networks).

We assume that the selection of a Web2.0 service by another Web2.0 service (i.e. mashup) increases the visibility of the selected Web2.0 service. The fact that the Web2.0 service has been selected provides Web2.0 service developers of future services with insight about the selected Web2.0 service. Since a Web2.0 service that has been used by a mashup has proved its reliability, a developer also increases his own evaluation of the Web2.0 service selected. Therefore, we can state that the more visible a Web2.0 service is, the more it is preferred by Web2.0 service developers.

Since a link represents the existence of a mashup in the Web2.0 service network, the number of links represents the visibility of the Web2.0 services linked. In turn, the larger the degree of a node in the Web2.0 service network is, the larger the probability is that it gets linked to through a new mashup. Under this model, it is expected that the Web2.0 service network evolves similarly to the model of Barabási and Albert [1]. The Web2.0 service network evolves through new entrants (mashup), which link an existing Web2.0 service with a new Web2.0 service, and through new mashups that re-use only existing Web2.0 services.

Based on this definition of Web2.0 service network, we can also apply the second model of Barabási et al. [2]. In this model, the probability that node i is connected to node j at time t is proportional to their degree k_i and k_j, respectively:

$$\Pi_{ij} = N(t)a\frac{k_i k_j}{\sum_l \sum_{m \neq l} k_l k_m} \qquad (1)$$

where $N(t)$ is the network size at time t and a the average number of links generated between node i and node j at time t.

Additionally, the probability that node i is linked with a new entrant at time t is proportional to its degree k_i:

$$\Pi_i = \frac{b\beta k_i}{\sum_j k_j} \qquad (2)$$

where b is the average number of links generated by an entrant and β is the number of nodes that enters the network at each time unit.

By adding equation (1) and (2), the number of links that node i gains at time t in the continuous case can be calculated, as expressed with the following equation:

$$\frac{dk_i}{dt} = \frac{b\beta k_i}{\sum_j k_j} + N(t)a\sum_{j\neq i}\frac{k_i k_j}{\sum_l \sum_{m\neq l} k_l k_m} \qquad (3)$$

However, equation (1), (2), and (3) are not conform to the fundamental characteristic of probabilities of event occurrences. The probability should be in the range between *0* and *1*. If an event has already occurred, then the probability that it emerges is one. Therefore, it is not clear what the sum of probabilities over all the possible events in equation (1) means. For example, if *N(t) a* and *b* are *2*, then it would result in $\sum_{i,j\neq i}\Pi_{ij} = 2$ and $\sum_i \Pi_i = 2$. The meaning of a value of 2 has not been defined in the work of Barabási et al. It could either means that two links probably emerge in the network or that a link is certainly generated assuming a probability larger than *1* implies a certain event occurrence.

In general, because of the preference attachment between existing nodes, most of the links are generated between nodes with high degree. This increases the degree of nodes in the tail of the power law distribution. Therefore, the network, which obeys the model of Barabási et al. [2], has a lower exponent γ in the power law than the exponent of a traditional scale-free network [10].

Furthermore, we expect that any additional link between existing nodes under the preferential attachment rule increases the distortion. This has not been validated in prior research yet.

2.3. Subgroups

In the model of Barabási et al. [2], the linking of a node depends only on the degree of the node, ceteris-paribus. Another factor that might have an impact on the structure of a network is the existence of subgroups. A subgroup is defined as a group of network nodes. Each node is a member of one but only one subgroup. The classification of nodes into subgroups can be based on any criteria (e.g. company ownership, type of service). For example, all Web2.0 services that provide photo-related services are classified into the same subgroup.

Subgroups are generally believed to prevent nodes included in different subgroups from generating a link between them [17]. It was also analytically proved that the existence of subgroups affects the formation of networks [18]. Provided that a node gains uniform benefit from every node linked to it directly and indirectly, the sufficiently small cost to generate a link with a node in the same subgroup (i.e. the internal cost) yields a high-density network for each subgroup. Subgroups of the network are not connected, if the cost to connect two nodes involved in different subgroups (i.e. the external cost) is larger than the internal cost. Otherwise, the subgroups are strongly connected with each other if the external cost is substantially small.

To support the previous statements, two scenarios can be constructed in the area of innovation within the Web2.0 service network: First, if Web2.0 providers open the access to their proprietary resources in order to promote innovation through user participation [19], the network of Web2.0 services self-organizes under the preferential attachment rule in this open system. The network is expected to show the scale-free property [10].

Second, in an out-dated model of innovation, on the other hand, a network is preferred that consists of links between nodes of the same subgroup only, i.e. the subgroups are closed [20]. If further links are only created between nodes of the same subgroup, a network with several isolated sub-networks emerge. Figure 1 illustrates this example. The figure shows a network that has two subgroups *A* and *B*. Then, if a new link emerges between two nodes, it is within the same subgroup only (e.g. link *a* between node *1* and node *2* within subgroup *A*). The same is true if a link (link *b*) emerges between two existing nodes (node 4 and 5) of the same subgroup (subgroup *B*).

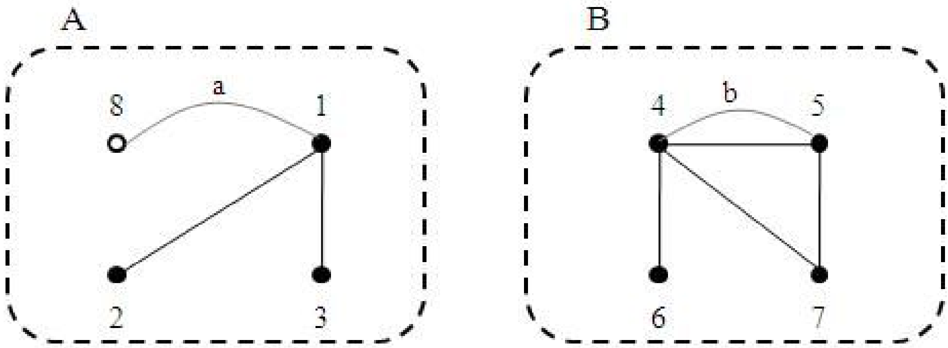

Figure 1. Example of two link creations (link *a* and *b*) between nodes of the same subgroups, while the subgroups are completely closed.

An example that explains the second scenario in more detail can also be found in the area of Web2.0 services. The company ownership of Web2.0 services (e.g. Google owns Google App Engine (GAE), and Amazon owns Amazon Web Services (AWS)) could be used to classify Web2.0 services into subgroups. Applying this, we get that subgroup Google includes Web2.0 services such as GAE and subgroup Amazon includes all AWS. Subgroup Google and subgroup Amazon interfere with the overall evolution of the Web2.0 service network, if Web2.0 services in one subgroup are incompatible with those in the other. A developer, who favorites the GAE, is likely to utilize Web2.0 services of Google only, since it would incur additional cost for him to learn the techniques to manipulate Web2.0 services of AWS.

Based on these two scenarios, we can state, if the boundaries between subgroups prevent agents of the network to link to each other (i.e. the preferences of the agents are affected), then the evolution of the entire network may be affected by this subgroup structure. Therefore, we postulate the following proposition:

- Proposition 1: Depending on the openness of subgroups, the subgroup structure of a network may distort the evolution of a network into a scale-free network.

Within another scenario, a network is considered that consists of subgroups with just a single node and a subgroup with many nodes, of which some are interconnected already. If a node belongs to a single-node subgroup and subgroup openness is assumed, then it is reasonable to assume that a new link of this single node would connect to a node, which has the highest degree. The single node would benefit from the link to the node with the high degree. Since even the subgroup with the high-degree node got larger, the probability of new links between its nodes increases as well. Thus, nodes belonging to this subgroup can be considered to have a first mover advantage. Consequently, this subgroup could dominate other subgroups (as if no subgroups exist), resulting in a scale-free network. This situation is depicted in Figure 2. The new link *b* connects node *5* of subgroup *B* with node *1* of subgroup *A*.

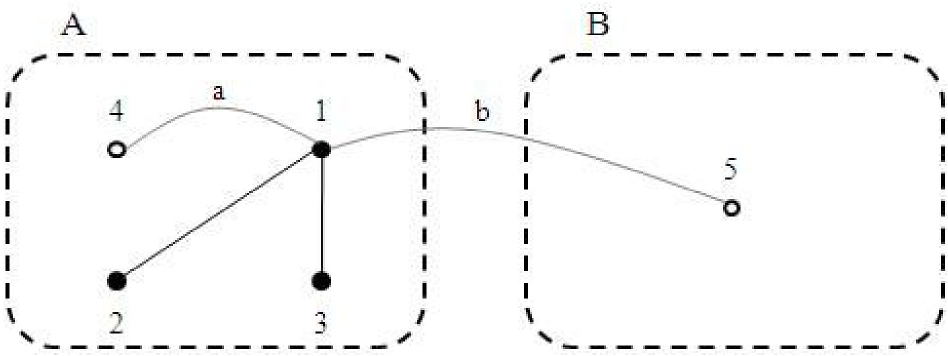

Figure 2. Example of a link creation (link b) across two subgroups between a node with a high node degree and a new node.

If, however, the subgroups are not totally open, the structure of the network is impacted by the subgroup number. To demonstrate this, we consider two networks. The first network consists of *N* nodes and two subgroups with $N/2$ nodes each. The second network consists of *N* nodes and *N* subgroups with one node each. The first network would evolve to a network with a scale-free property, since the effect of separation between the two subgroups is little. Within the second network, however, the total number of nodes is split across many subgroups. Therefore, there is only one node with a very high degree but many with a small-sized degree. Based on this, we postulate:

- Proposition 2: Depending on the openness of subgroups and the number of subgroups, the subgroup structure of a network distorts the evolution of a network into a scale-free network.

3. The Modified Preferential Attachment Rule

To model the effect of the subgroup structure, we define the *Subgroup-Activity-Factor*, which indicates the openness between two subgroups. The Subgroup-Activity-Factor Λ_s^t for two subgroup s and t is defined as:

$$\Lambda_s^t = \begin{cases} \Lambda & if\ s \neq t \\ 1 & if\ s = t \end{cases} \qquad (4)$$

where Λ is the subgroup openness. If $\Lambda = 1$, the effect of the subgroups on the network evolution is neutral (openness). If $\Lambda > 1$, then a link is likely to be generated Λ-times more often between nodes belonging to different subgroups than between nodes belonging to the same subgroup. If $\Lambda < 1$, then a link is likely to be generated $1/\Lambda$-times between nodes belonging to the same subgroup than between nodes of different subgroups (closeness).

In order to be able to express the effect of subgroups (i.e. Subgroup-Activity-Effect) on networks, we design the probabilities Π_{ij}^{pq} and Π_i^p and the preferential attachment rule for the discrete case, which follows the continuous case given by Barabási et al. (2001) [2].

In detail, we start by defining the probability that node i belonging to subgroup p is linked with node j belonging to subgroup q as:

$$\Pi_{ij}^{pq} = \frac{k_i^{\,p} k_j^{\,q} \Lambda_p^q}{\sum_l \sum_{m>l} k_l^{\,u} k_m^{\,v} \Lambda_u^v} \qquad (5)$$

where k_i^p and k_j^q are the degrees of node i, belonging to subgroup p, and of node j belonging to subgroup q, respectively. Λ_p^q is the Subgroup-Activity-Factor for subgroup p and subgroup q. u and v are the subgroups of node l and m, respectively.

Second, we define the probability that node i belonging to subgroup p is linked with a new node (entrant):

$$\Pi_i^p = \frac{k_i^{\,p} \Lambda_p^q}{\sum_j k_j^{\,u} \Lambda_p^u} \qquad (6)$$

where k_i^p is the degree of node i belonging to subgroup p, and the entrant belongs to subgroup q. u represents the subgroup of node j.

To resolve the shortcoming of the Barabási et al. model, the variables b, β, $N(t)$, and a, which appear in equation (1) and (2) of the Barabasi et al. model, were not considered in equation (5) and (6). Therefore, the probability is in the range between 0 and 1. However, besides defining equation (5) and (6), we also introduce two new variables δ_{ij}^{pq} and δ_i^p. δ_{ij}^{pq}

is defined as the increase in the number of new links created between node i and node j in one time period. δ_i^p represents the increase of the degree of node i by linking to entrants in one time period.

The calculation of δ_{ij}^{pq} depends on $N(t)\ a$, where $N(t)$ is the number of notes at time period t and a is rate of new nodes. The algorithm starts by calculating the probability of a new link between any two existing nodes as $(N(t)\ a - floor(N(t)\ a))\ \Pi_{ij}^{pq}$. If node i gains a link to node j, δ_{ij}^{pq} is increased by one. Afterwards, the following steps are executed *floor(N(t) a))* times: A link between two existing nodes is generated with the probability Π_{ij}^{pq}. If node i gains a link to node j, δ_{ij}^{pq} is increased by one. Eventually, the sum of δ_{ij}^{pq} for all nodes at any time period is greater than the largest integer which is smaller than $N(t)\ a$, and smaller than the minimal integer which is larger than $N(t)\ a$.

Likewise, we calculate δ_i^p, which depends on $\beta\ b$, where b is the average number of links generated by an entrant and β is the number of nodes that enters the network at each time unit. The algorithm for calculating of δ_i^p works equivalently to the algorithm for calculating δ_{ij}^{pq}, using the probability of equation (6) and $\beta\ b$ instead.

Based on the previous two definitions, the equation of the differential attachment rule is defined as:

$$\frac{dk_i^p}{dt} = \left[\delta_i^p + \sum_{j \neq i} \delta_{ij}^{pq} \right] \qquad (7)$$

4. Simulation Results and Analysis

Our simulation model comprises the equations (4) to (7) and the following two parameters: the number of subgroups G and the distribution of the subgroup size D. From the original model, we also include parameters about the initial network size N_0, the initial link density K, and the final network size F.

All simulations were initiated with an initial network consisting of 2 nodes ($N_0 = 2$) and 1 link ($K = 1$). It was also assumed that one node enters the network at one time unit ($b = 1$) and generates one link ($\beta = 1$). The distribution of the subgroup size (D) was assumed to be uniform. This avoided affecting the analysis of the simulation results with respect to the subgroup openness and the number of subgroups.

For our simulation runs, we controlled several variables. The independent variables in our simulations are the average number of links generated between existing nodes a, the number of subgroups G, and the strength of the subgroup openness Λ.

Dependent variables are the exponent γ of the power distribution of degrees and the shape of the cumulative degree distribution.

4.1. Validation Test

The basic preferential attachment model was simulated to check whether our simulation code is valid. The network evolved under the basic preferential attachment rule (i.e. not considering the second term of equation (7)). In this model, a new entrant enters the network to generate a link at each time unit under the preference, which is linearly proportional to the degree of the node, but no links are generated between existing nodes. The simulation run until the network size reached 30,000 nodes.

The result of our simulation matched the results of prior research, which proved empirically and theoretically that the exponent γ of the power law is between *2* and *3* [3] [5]. The degree distribution of our simulation result is shown in Figure 3. The regression of the distribution shows that its exponent γ is about *2.5*. The explanation power of the regression R^2 is 0.97, i.e. significantly high.

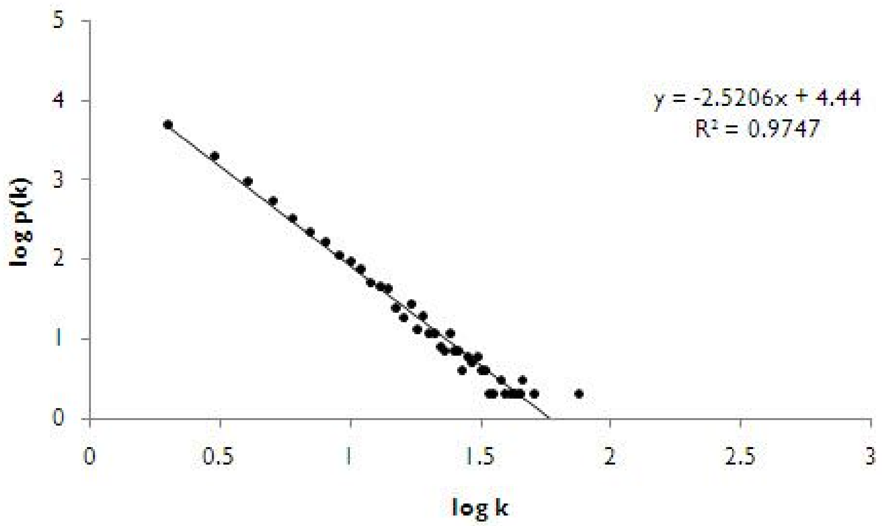

Figure 3. The distribution *p(k)* of the degree k of a scale-free network (*N_0=2, K=1, F=30000, a=0, b=1, β=1, G=1, Λ=1*).

In order to check whether the evolutionary pattern is stable near the point of time of observation, the history of the network evolution is traced until the iteration of the preferential attachment reaches *30000* time units. The exponent γ fluctuates between *1.5* and *2.5* until *10000* time units (Figure 4). After that, the exponent γ is quite stable at about *2.5*. The explanation power R^2 is stable at *0.97* after the network size reached *10000* time units (Figure 5). As a consequence, we set the simulation period of future simulations to *20,000* time units, still achieving the validity of results but increasing time efficiency.

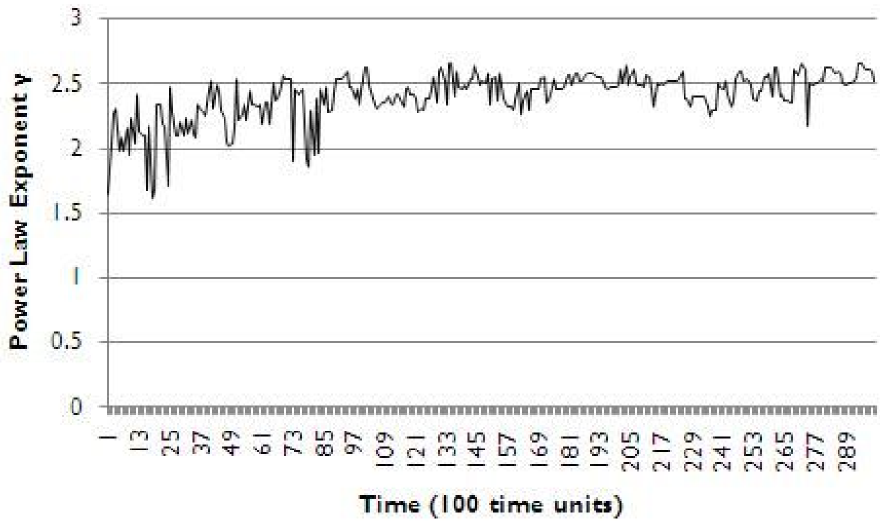

Figure 4. The trend of gamma of the scale-free network for each *100* time units until *30000* time units (*N_0=2, K=1, F=30000, a=0, b=1, β=1, G=1, Λ=1*).

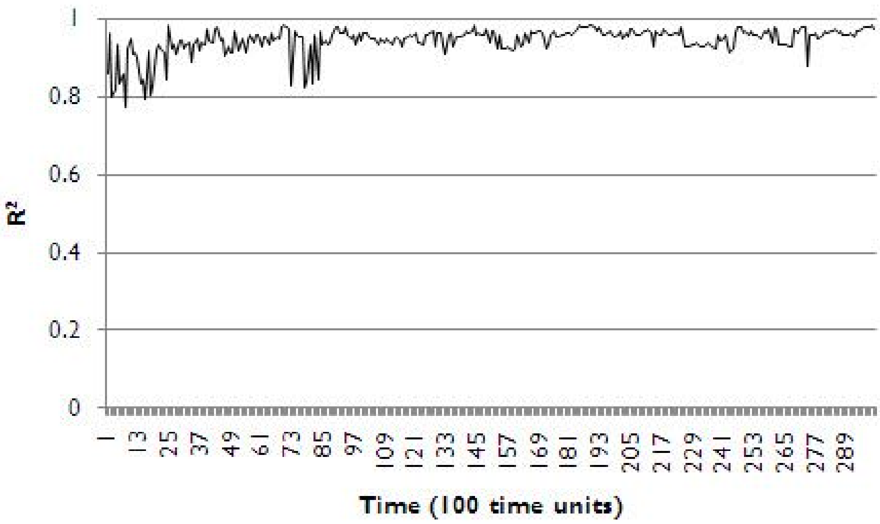

Figure 5. The trend of R^2 of the scale-free network for each *100* time units until *30000* time units (*N_0=2, K=1, F=30000, a=0, b=1, β=1, G=1, Λ=1*).

In order to check the validity of using the cumulative degree distribution, we calculate the regression of the cumulative degree distribution at a network size of *30000* nodes. Our result shows the usefulness of the cumulative degree distribution (Figure 6). The exponent γ of the degree distribution is *2.8* (i.e. *1.8 + 1*, where *1.8* is the negative slope of the cumulative degree distribution). The explanation power of the regression is *0.99*, slightly higher than R^2 of the degree distribution. Therefore, for remainder of this analysis, we continue to derive the exponent γ from the cumulative degree distribution.

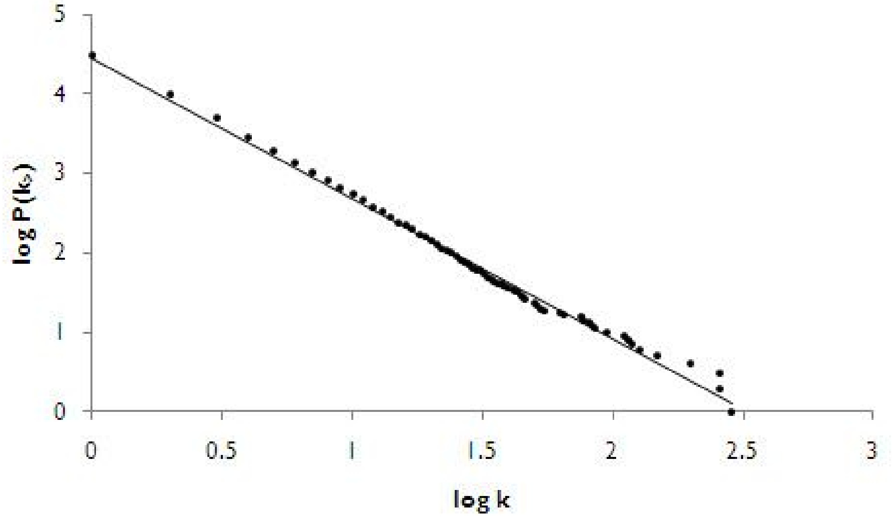

Figure 6. The cumulative degree distribution *P(k)* of the degree k of the scale-free network (*N_0=2, K=1, F=30000, a=0, b=1, β=1, G=1, Λ=1*).

4.2. Test of Impact of Links between Existing Nodes

For finding out the effect of adding links between existing nodes on the evolution of a network, we simulate a network that evolves from an initial network size of 2 nodes ($N_0=2$ and $K=1$) to a network of *20000* nodes ($F=20,000$). At each time step, an entrant establishes a link with an existing node. Concurrently, existing nodes create additional links with each other at a certain rate. We run simulations for a variety of different rates: $a=0$, 10^{-5}, 10^{-4}, 10^{-3}, and 10^{-2}.The results showed that the different networks (with respect to the rate a) slightly distort the shape of the power law of the degree distribution (Figure 7).

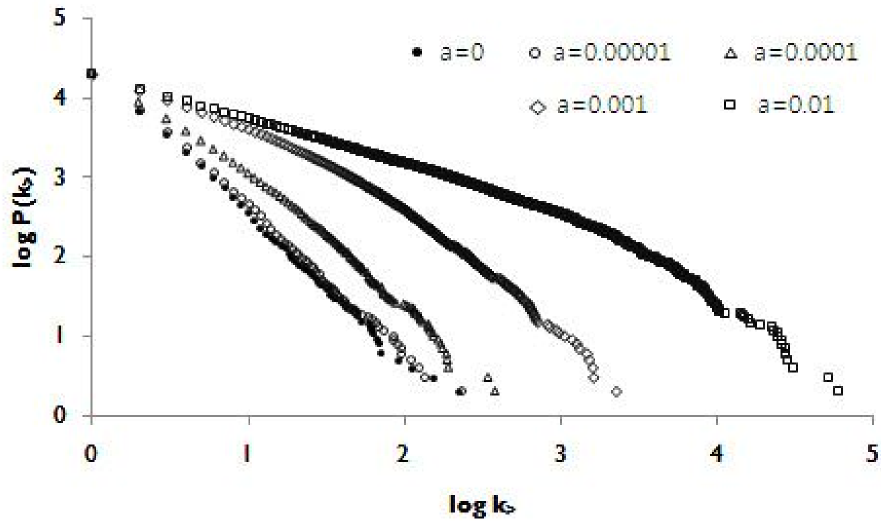

Figure 7. The cumulative degree distribution of the scale-free network for $a=0$, $a=10^{-5}$, $a=10^{-4}$, $a=10^{-3}$, $a=10^{-2}$ ($N_0=2$, $K=1$, $F=20000$, $b=1$, $\beta=1$, $G=1$, $\Lambda=1$).

As Figure 7 illustrates, the intercept point with the vertical axis with all distributions is invariable at *4.30*. This means that the links are hardly generated between existing nodes whose degree is *1*. If links had been generated between existing nodes whose degree is *1*, the intercept point would have increased as *a* increases. The lower ends of the different distributions move from *2.34* to *4.77* as *a* increases. It means that the generation of links between existing nodes reinforces the degree of nodes, especially of hubs. The disturbance gives the cumulative degree distribution a concave shape.

These results suggested that the effect of additional links between existing nodes lowers the exponent γ of the power law. The exponent γ decreases logarithmically as *a* increases. Figure 8 depicts the simulation results (diamonds) and the fitted line (bold line) estimated from the simulation results. A regression on this relationship showed that γ and the logarithm of *a* have a negative correlation (8). The slope is about *-0.05*. The fitting is significant in *1%* significance because the probability of the t-distribution is *0.003* and R^2 is *0.99*.

$$\gamma = -0.05 \log a + 2.1 \qquad (8)$$

These results about the impact of additional links between existing nodes are used as a reference for the next section, in which the effect of the subgroup structure on the overall shape of the distribution is determined.

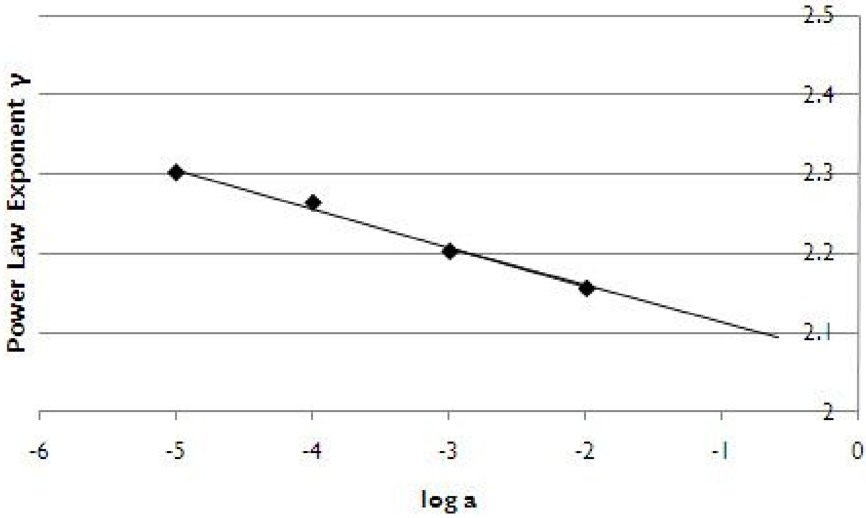

Figure 8. The correlation between the exponent γ and logarithm of the rate of link creation between existing nodes log a.

4.3. Test of Proposition 1

The initial network size and the final network size of the simulation for testing Proposition 1 are set identically to those of simulation in Section IV.B. At each time period, an entrant creates a link with an existing node and existing nodes create *a* links with each other. However, we set the number of subgroups to $G = 100$. We simulate the network evolution for different values of subgroup openness: $\Lambda=0,\ 10^{-4},\ 10^{-3},\ 10^{-2},\ 10^{-1}$, and 1. These values represent different levels of openness, ranging from completely open ($\Lambda = 1$) to completely closed ($\Lambda = 0$).

Our simulation results show that the subgroup structure affects the link creation between existing nodes, not the link creation between an entrant and an existing node. If $a = 0$, the power law of the cumulative degree distribution is almost invariant to the effect of subgroup structure. If $a > 0$, the subgroup structure affects the network evolution. Figure 9 illustrates the results for $a = 0.001$. However, in order to increase the readability of the graph, the cumulative degree distributions for three different values of Λ (i.e. $\Lambda = 0$, $\Lambda = 10^{-2}$, $\Lambda = 1$) are only shown. The cumulative distribution for $\Lambda = 1$ equals to the one for $\Lambda = 1$ and $G = 1$ (Figure 7). It means that, under complete openness, the number of subgroups has no impact.

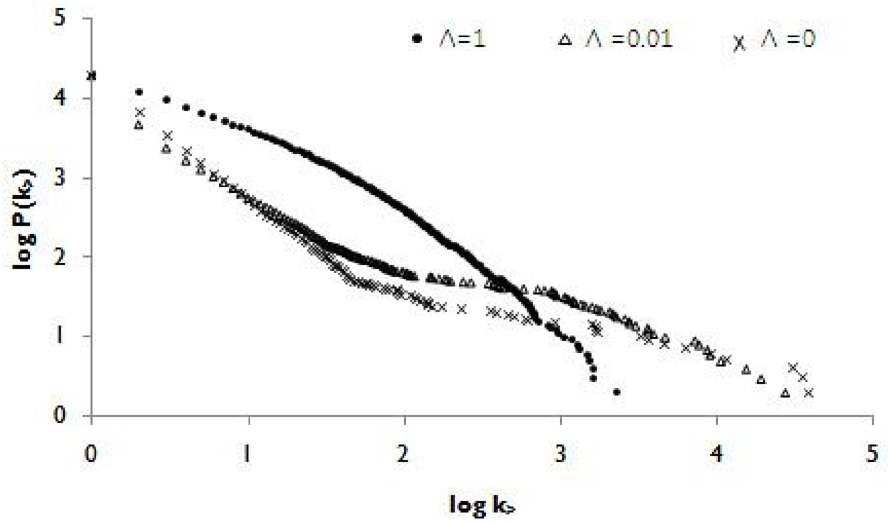

Figure 9. The cumulative degree distribution of the network for *Λ=1, Λ =10-2, Λ =0* (*N_0=2, K=1, F=20000, a=0.001, b=1, β=1, G=100*).

The distortion, which can be observed in Figure 9, results from the additional number of links between existing nodes. If $\Lambda = 10^{-2}$, or $\Lambda = 0$, the cumulative degree distribution changes its slope at a certain point (kinks). This is further highlighted in Figure 10. The two straight lines are trend lines of the simulation results for $\Lambda = 0$. As Figure 10 illustrates, the slope of the first part of the cumulative degree distribution (up to the kink) decreases faster than the second part of the distribution (right-hand side of the kink). The estimate of the slope of the first part is *-1.51* and the slope of the second part is *-0.40*. The larger Λ is, the more the cumulative distribution is distorted. Therefore, we can state that our simulation results support Proposition 1.

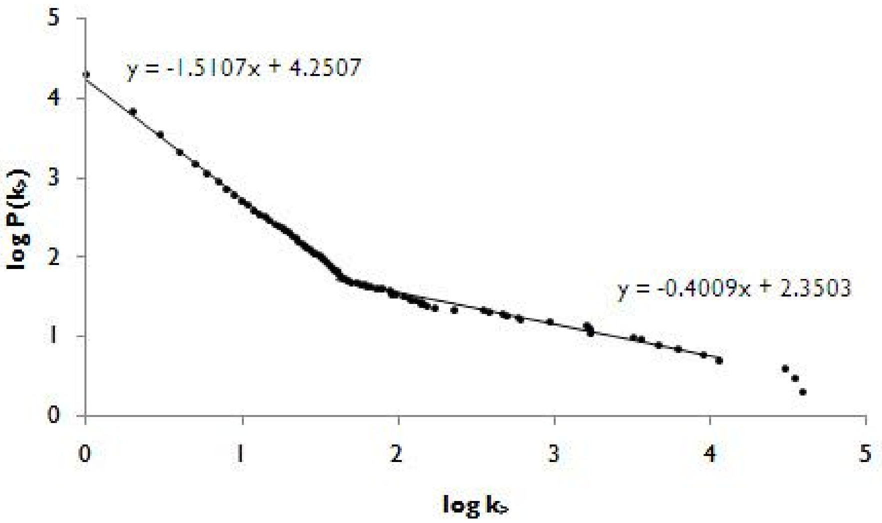

Figure 10. The cumulative degree distribution of the network simulated (*N_0=2, K=1, F=20000, a=0.001, b=1, β=1, G=100, Λ =0*).

It has also to be noticed that the distortion is different to the distortion obtained from additional number of links between existing nodes in the network. While the distortion shown here results in a convex cumulative degree distribution, the cumulative degree distribution for additional links between existing nodes is concave (Figure 7). Furthermore, the distortion from closeness outweighs the distortion from additional links between nodes in the network.

4.4. Test of Proposition 2

The initial and final network conditions of the simulation of Proposition 2 are equivalent to the simulation of Proposition 1. For simulating the effect of the number of subgroups on the network evolution, however, we varied the number of subgroups. We simulate the network evolution for five different subgroup numbers: $G=1$, 10, 10^2, 10^3, and 10^4. All nodes are allocated uniformly to one of those G subgroups. An entrant node is assigned a subgroup before its interaction with the network. Additionally, the network evolution also depends on the heterogeneity of subgroups, since the simulation starts with an initial network consisting of only two nodes. That means, some subgroups do not have any nodes initially. Furthermore, the nodes within the same subgroup are closed against those belonging to different subgroups ($\Lambda=0$).

Figure 11 shows clearly the impact of the number of subgroups on the evolution of the network. While the cumulative degree distribution for $G = 1$ (no subgroups) shows a concave curve, the distribution function at $G = 10^2$ *and* $G = 10^3$ exhibits a convex shape. The higher the number of subgroups G is, the more the cumulative degree distribution curve bends (convex shape), clearly distorting the shape of a scale-free network. Therefore, the results support Proposition 2 for completely closed subgroups ($\Lambda=0$).

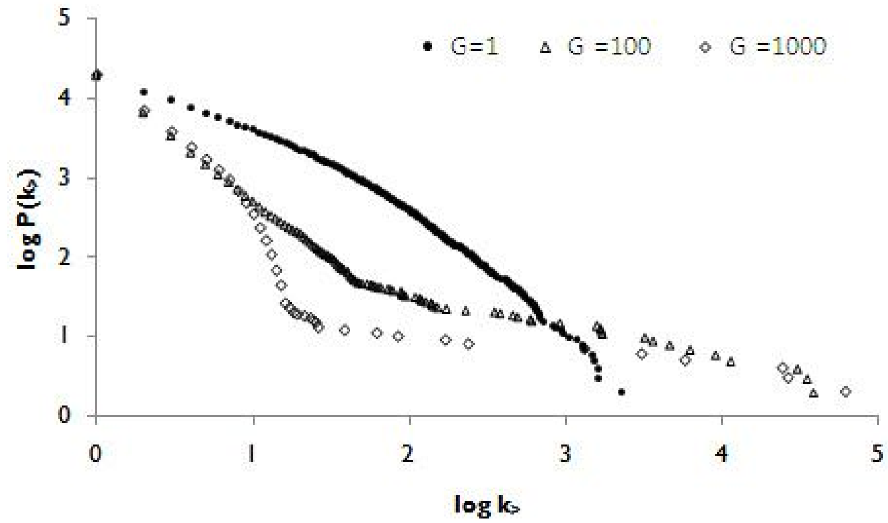

Figure 11. The cumulative degree distribution of the network for *G=1, G=10²,* *G=10³* (*N₀=2, K=1, F=20000, a=0.001, b=1, β=1, Λ=0*).

In order to explain this shape, we look at the degree distribution for $G = 100$. Figure 12 exhibits the degree distribution and the cumulative degree distribution. The bold vertical line at $\log k = 1.57$ indicates the difference in slopes on the right-hand side (slight slope) and on the left-hand side (steep slope).

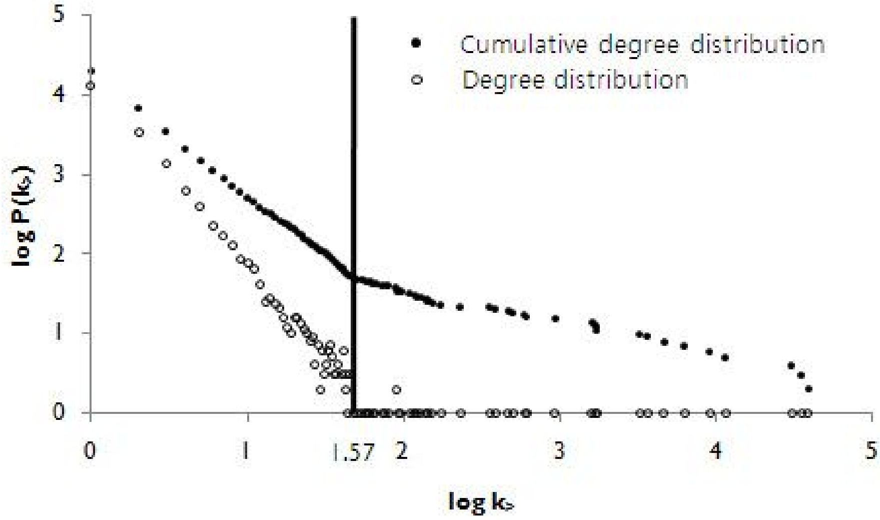

Figure 12. The degree distribution and the cumulative degree distribution of the network (*N₀=2, K=1, F=20000, a=0.001, b=1, β=1, G=100, Λ=0*).

The right-hand part of the cumulative degree distribution curve, which is to the right of the bold vertical line in Figure 12, emerges due to nodes with a degree frequency of one.

Therefore, the logarithm of the degree frequency is zero. This is an evidence of an imbalanced subgroup distribution. *41* out of *46* nodes, which represent the right-hand side of the degree distribution, belong to the same subgroup. This subgroup includes the node that has constituted the initial network. If the nodes of the initial network are not equally distributed across all subgroups, the nodes of the subgroup which contained initially the nodes with the highest degree can become dominant. That means this subgroup contains nodes with the highest degrees. The remaining nodes of other subgroups form regular scale-free network. In general, closeness between subgroups divides self-organized networks into a dominant subgroup and ordinary subgroups.

As a note, we did not analyze the trend of the exponent γ of the cumulative degree distribution functions for Proposition 1 and Proposition 2. Calculating the exponent γ would be meaningless, since the cumulative degree distribution is not a linear function.

5. Discussion and Conclusion

A scale-free network consists of a minority of nodes with a high node degree and a majority of nodes with a low and middle node degree. This inhomogeneous structure is self-organized through the application of a simple rule, the preferential attachment rule. The rule makes the degree-rich nodes get rich quicker than others.

For our analysis, a modified version of this rule, which we developed, was applied to links emerging between existing nodes as well as to links from new entrants. The results showed that links generated between existing nodes distort the scale-free property. This corresponds to results found in an empirical study of the Web2.0 service network [10]. In addition to this, we introduced the concept of subgroup structures. The subgroup structure concept assumes that all network nodes belong to only one subgroup and that the openness between subgroups affects the evolution of a network.

In detail, our theory, which was supported by simulation results, comprises two parts: First, depending on the openness between subgroups, the subgroup structure of a network distorts the evolution of a network into a scale-free network. Second, depending on the openness between subgroups and the number of subgroups, the subgroup structure of a network distorts the evolution of a network into a scale-free network.

Our theoretical and simulated results suggest that closeness between subgroups and the number of subgroups are a significant perturbation of a network that evolves under the preferential attachment rule. Closeness between subgroups increases the degree of nodes belonging to the same subgroup.

Bringing these results into the context of the Web2.0 services system, we can consider that Web2.0 services are provided to earn money based on usage. The more Web2.0 service developers include a specific Web2.0 service, the more revenue this specific Web2.0 service generates.

As a consequence of these research results, two recommendations can be given with respect to business and management.

First, it is much more beneficial for Web2.0 service providers to promote the re-use of their Web2.0 services than to introduce a new service into the network. A new service introduced in the network does not guarantee a large benefit in myopia, since it is placed in the periphery of the network (i.e. low node degree). The center of the network, where nodes with a high node degree are located, is most profitable.

Second, closeness between Web2.0 service providers is beneficial for the provider possessing the most popular Web2.0 service. Our simulations showed that a first mover service can become a dominant subgroup (i.e. it contains most of the nodes with the largest degree). The slope of the trend line of the nodes with high degrees ($k > 1.57$) increased. As Figure 10 shows, the slope of the trend line of the network ($a=0.001, G=100, \Lambda=0$) is -0.4. The slope of the remaining nodes is -1.51. This means that subgroups make the rich nodes richer and the poor nodes poorer.

However, our results do not suggest that a Web2.0 service provider, who owns the most popular service, should choose to close up its interfaces to other providers. Suppose that the owner of the most popular service closes its interfaces but all other Web2.0 service providers are open (i.e. allow interconnections), then the degree (i.e. profit) of the services of the closed subgroup containing the best services stagnate while the degree (i.e. profit) of the other nodes augment. This is similar to the failure of Sony's Betamax and Apple's Machintosh. Both superior technologies failed to get partners [21].

References

[1] Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks. Science, vol. 286, pp. 509-512.

[2] Barabási, A.-L., Jeong, H., Néda, Z., Ravasz, E. Z., Schubert, A., Vicsek, T., 2009. Evolution of the social network of scientific collaborations. arXiv:cond-mat/0104162v1, 2001 (accessed on 1. December 2009).

[3] Dorogovtsev, S. N., Mendes, J. F. F., Samukhin, A. N., 2000. Structure of growing networks with preferential linking. Physical Review Letters, vol. 85, no. 21, pp. 4633-4636.

[4] Fu, F., Liu, L., Wang, L., 2008. Empirical analysis of online social networks in the age of Web2.0. Physica A, vol. 387, no. 2-3, pp. 675-684.

[5] Štefančić, H., Zlatić, V., 2005. Winner takes it all: Strongest node rule for evolution of scale-free networks. Physical Review E. vol. 72, no. 3 pt. 2, 036105.

[6] Albert, R., Barabási, A.-L., 2002. Statistical mechanics of complex networks. Reviews of Modern Physics, vol. 74, no. 1, pp. 47-97.

[7] Newman, M. E. J., 2003. The structure and function of complex networks. SIAM Review, vol. 45, no. 2, pp. 167-256.

[8] Park, K., Lai, Y.-C., Ye, N., 2005. Self-organized scale-free networks. Physical Review E, vol. 72, no. 2, pp. 026131-1 – 026131-5.

[9] Hendler, J., Shadbolt, N., Hall, W., Berners-Lee, T., Weitzner, D., 2008. Web science: An interdisciplinary approach to understanding the Web. Communications of the ACM, vol. 51, no. 7, pp. 60-69.

[10] Hwang, J., Altmann, J., Kim, K., 2009. The structural evolution of the Web2.0 service network. Online Information Review, vol. 33, no. 6, pp. 1040-1057.

[11] Valverde, S. Solě, R. V., 2007. Self-organization versus hierarchy in open source social networks. Physical Review E, vol. 76, no. 4, pp. 046118.

[12] Newman, M. E. J., 2004. Power laws, Pareto distributions and Zipf's law. Contemporary Physics, vol. 46, no. 5, pp. 323-351.

[13] Albert, R. Barabási, A.-L., 2000. Topology of evolving networks: Local events and university. Physical Review letters, vol. 85, no. 24, pp. 5234-5237.

[14] Wagner, C. S., Leydesdorff, L., 2005. Network structure, self-organization, and the growth of international collaboration in science. Research Policy, vol. 34, no. 10, pp. 1608-1618.

[15] Adamic, L. A., Lukose, R. M., Huberman, B. A., 2003. Local search in unstructured networks. in Handbook of Graphs: From the Genome to the Internet, S. Bornholdt, and H. G. Schuster, Eds. Weinheim: Wiley-VCH, pp. 295-317.

[16] Wang, F., Sun, Y., 2008. Self-organizing peer-to-peer social networks," Computational Intelligence, vol. 24, no. 3, pp. 214-233.

[17] Krackhardt, D. Stern, R. N. 1988. Informal networks and organizational crises: An experimental simulation. Social Psychology Quarterly, vol. 51, no. 2, pp. 123-140.

[18] Galeotti, A. Goyal, S., Kamphorst, J., 2003. Network formation with heterogeneous players. Economics Discussion Papers of University of Essex.

[19] O'Reilly, T., 2007. What is Web2.0: Design patterns and business models for the next generation of software. Communications & Strategies, vol. 65, no. 1, pp. 17-37.

[20] Chesbrough, H. W., 2003. Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business School Press, Boston.

[21] Gawer, A., Cusumano, M. A., 2002. Platform Leadership: How Intel, Microsoft, and Cisco Drive Industry Innovation. Harvard Business School Press, Boston.

TEMEP Discussion Papers

2009-10: Donghyun Oh and Almas Heshmati, "A Sequential Malmquist-Luenberger Productivity Index".

2009-11: Arno Tausch and Almas Heshmati, "Re-Orient? MNC Penetration and Contemporary Shifts in the Global Political Economy"

2009-12: Donghyun Oh, Almas Heshmati and Hans Lööf, "Technical Change and Total Factor Productivity Growth for Swedish Manufacturing and Service Industries"

2009-13: Tai-Yoo Kim, Almas Heshmati and Jihyun Park, "The Faster Accelerating Knowledge-based Society"

2009-14: Junseok Hwang, Jörn Altmann and Kibae Kim, "The Structural Evolution of the Web2.0 Service Network"

2009-15: Alireza Abbasi, Jörn Altmann and Junseok Hwang, "Evaluating Scholars Based on Their Academic Collaboration Activities: The RC-Index and CC-Index for Quantifying Collaboration Activities of Researchers and Scientific Communities"

2009-16: Dong-hyun Oh, Hans Lööf and Almas Heshmati, "The Icelandic Economy: A Victim of the Financial Crisis or Simply Inefficient?"

2009-17: Sungki Lee, Donghyuk Choi and Yeonbae Kim, "Contextual Effects on the Complementarities between R&D Activities: An Empirical Analysis of the Korean Manufacturing Industry"

2009-18: Younghoon Kim, Yeonbae Kim and Jeong-Dong Lee, "Corporate Venture Capital and Its Contribution to Intermediate-Goods Firms in South Korea"

2009-19: Yunhee Kim, Jae Young Choi and Yeonbae Kim, "Complementarity and Contextuality in the Adoption of Information Systems in Korean Firms"

2009-20: Marcel Risch and Jörn Altmann, "Enabling Open Cloud Markets Through WS-Agreement Extensions"

2009-21: Almas Heshmati and Rachid El-Rhimaoui, "Effects of Ownership and Market Share on Performance of Mobile Operators in MENA Region"

2009-22: Donghyun Oh, Almas Heshmati and Hans Lööf, "Total Factor Productivity of Korean Manufacturing Industries: Comparison of Competing Models with Firm-Level Data"

2009-23: Almas Heshmati, Dan Johansson and Carl Magnus Bjuggren, "Effective Corporate Tax Rates and the Size Distribution of Firms"

2009-24: Seong-Sang Lee and Yeonbae Kim, "The Impact of Efficiency Parameters on Firms' Innovative Activities: Evidence from Korean Firm-Level Data"

2009-25: Junseok Hwang, Dongook Choi, Jongeun Oh and Yeonbae Kim, "Competition in the Korean Internet Portal Market: Network Effects, Profit, and Market Efficiency"

2009-26: Jae Young Choi, Yeonbae Kim, Yungman Jun and Yunhee Kim, "A Multivariate Probit Analysis of Korean Firms' Information System Adoption: An Empirical Analysis on the Determinants of the Adoption and Complementarity among the Information Systems"

2009-27: Flávio Lenz-Cesar and Almas Heshmati, "Determinants of Firms Cooperation in Innovation"

2009-28: Firas H. Al-Hammadany and Almas Heshmati, "Determinants of Internet Use in Iraq"

2009-29: Jeong-Dong Lee, "Industry Dynamics and Productivity Research"

2009-30: Nam TRAN-QUY and Almas Heshmati, "Measurements and Determinants of Pay Inequality and its Impacts on Firms Performance in Vietnam"

2009-31: Nils Karlson, Marcus Box and Almas Heshmati, "Generality, State Neutrality and Unemployment in the OECD"

2009-32: Chang Seob Kim, Yoonmo Koo, Ie-jung Choi, Junhee Hong and Jongsu Lee, "Consumer Preferences for Automobile Energy Efficiency Grades"

2009-33: Jongsu Lee, Jae Young Choi and Youngsang Cho, "A Forecast Simulation Analysis of the Next-Generation DVD Market Based on Consumer Preference Data"

2009-34: Jungwoo Shin, Chang Seob Kim and Jongsu Lee, "Model for Studying Commodity Bundling with a Focus on Consumer Preference"

2009-35: Yuri Park, Hyunnam Kim and Jongsu Lee, "Empirical Analysis of Consumer Preferences for Mobile Phones and Mobile Contents in the Presence of Indirect Network Effects"

2009-36: Jörn Altmann and Juthasit Rohitratana, "Software Resource Management Considering the Interrelation between Explicit Cost, Energy Consumption, and Implicit Cost: A Decision Support Model for IT Managers"

2009-37: Marcel Risch, Ivona Brandic and Jörn Altmann, "Using SLA Mapping to Increase Market Liquidity"

2009-38: Amit K. Bhandari and Almas Heshmati, "Willingness to Pay for Biodiversity Conservation"

2010-39: Marcel Risch, Jörn Altmann, Li Guo, Alan Fleming and Costas Courcoubetis, "The GridEcon Platform: A Business Scenario Testbed for Commercial Cloud Services"

2010-40: Junseok Hwang, Jörn Altmann and Ashraf Bany Mohammed, "Determinants of Participation in Global Volunteer Grids: A Cross-Country Analysis"

2010-41: Ruzana Davoyan, Jörn Altmann and Wolfgang Effelsberg, "Intercarrier Compensation in Unilateral and Bilateral Arrangements"

2010-42: Ruzana Davoyan, Jörn Altmann and Wolfgang Effelsberg, "Exploring the Effect of Traffic Differentiation on Interconnection Cost Sharing"

2010-43: Ruzana Davoyan and Jörn Altmann, "Investigating the Role of a Transmission Initiator in Private Peering Arrangements"

2010-44: Ruzana Davoyan, Jörn Altmann and Wolfgang Effelsberg, "A New Bilateral Arrangement between Interconnected Providers"

2010-45: Marcel Risch and Jörn Altmann, "Capacity Planning in Economic Grid Markets"

2010-46: Dang Minh Quan and Jörn Altmann, "Grid Business Models for Brokers Executing SLA-Based Workflows"

2010-47: Dang Minh Quan, Jörn Altmann and Laurence T. Yang, "Error Recovery for SLA-Based Workflows within the Business Grid"

2010-48: Jörn Altmann, Alireza Abbasi and Junseok Hwang, "Evaluating the Productivity of Researchers and their Communities: The RP-Index and the CP-Index"

2010-49: Jörn Altmann and Zelalem Berhanu Bedane, "A P2P File Sharing Network Topology Formation Algorithm Based on Social Network Information"

2010-50: Tai-Yoo Kim, Seunghyun Kim and Jongsu Lee, "The Gene of an Accelerating Industrial Society: Expansive Reproduction"

2010-51: Almas Heshmati and Sangchoon Lee, "The Relationship between Globalization, Economic Growth and Income Inequality"

2010-52: Flávio Lenz-Cesar and Almas Heshmati, "Agent-based Simulation of Cooperative Innovation"

2010-53: Erkhemchimeg Byambasuren and Almas Heshmati, "Economic Development in Mongolia"

2010-54: Almas Heshmati and Subal C. Kumbhakar, "Technical Change and Total Factor Productivity Growth: The Case of Chinese Provinces"

2010-55: Bory Seng and Almas Heshmati, "Digital Divide and Its Variations Amongst OECD, NIE and ASEAN Countries"

2010-56: Kibae Kim, Jörn Altmann, and Junseok Hwang, "The Impact of the Subgroup Structure on the Evolution of Networks: An Economic Model of Network Evolution"