A Signalling Explanation for Preferential Attachment in the Evolution of Social Networks

Jeong-Yoo Kim[1]
Department of Economics
Kyung Hee University

Hang-Hyun Jo
School of Physics
Korea Institute for Advanced Study (KIAS)

August 9, 2007

[1]Corresponding author: Department of Economics, Kyung Hee University, 1 Hoegidong, Dongdaemunku, Seoul 130-701, Korea (Tel & Fax) 82-2-961-0986, (Email) jyookim@khu.ac.kr

Abstract

The network theory developed by physicists has several critical drawbacks in characterizing the structure of social networks. First, they largely neglect considering the link cost and the link benefit that agents usually take into account in forming their links. Second, they view a social network as a consequence of unilateral decisions of agents, not of bilateral decisions of linking parties, although a link of an agent can be formed only after he obtains the consent of the other side. Third, there is no logical justification for the assumption of preferential attachment upon which their analysis relies heavily. In this paper, we provide several models that overcome the three drawbacks. By analyzing the models, we can explain preferential attachment as rational equilibrium behavior. The main idea is that people are not certain of the value that they can obtain from forming a link with someone. Based on this assumption, we will argue that a person has an incentive to form a link with another who has many links because the number of his links can convey some information about his value.

JEL Classification Code: C72

Key Words: Preferential Attachment, Hub, Power Law, Network Evolution, Signal

1 Introduction

Many economic interactions rely on social networks. Just as physical networks like telecommunications network or Internet do, social networks based on solid interpersonal relationships enable or facilitate exchange of valuable information or transactions among agents on a network. Thus, understanding the structure of networks has been recognized as important to explaining many natural or economic/social phenomena.

In the economics literature, there has been a significant development in the theory of network formation since the seminal article by Jackson and Wolinsky (1996).[1] On the other hand, in the physics literature, network evolution models have been developed to offer an explanation for some specific features of networks, for example, the emergence of hubs and the power law.[2] Indeed, link distributions in many networks including even social networks are found to follow the power law.[3]

Barabási and Albert (1999) model the scale-free network[4] based on the evolution and preferential attachment to explain the emergence of a hub and the power law.[5] According to them, a network can have hubs if nodes enter the population sequentially so that earlier comers have higher chances to be linked with others. If new nodes are more likely to attach a link to a node connected to more links, a hub will emerge more rapidly and the network will follow the power law. Their argument appears convincing, but it has several critical drawbacks at least in explaining social network structures. First, the argument neglects the reality that people form links by comparing their benefit and cost from the links. Second, it fails to capture the feature that a social network is a consequence of bilateral decisions of a pair of people, not of their unilateral decisions, in other words, a person's link cannot be attained solely by his wish to form the link, but can be formed only after he obtains

[1]To name a few, see Dutta & Mutuswami (1997), Bala & Goyal (2000), Jackson & van den Nouweland (2005) and Jackson & Watts (2002). For an extensive survey, see Jackson (2005).

[2]The power law says that the distribution of nodes with a certain number of links follows a power function.

[3]For example, Liljeros *et al.* (2001) showed the distribution of the number of sexual partners decays as a scale-free power law. Ahn (2005) recently studied the network structure of the closest friends (that can be linked with each other after mutual agreement) in the CyWorld which is one of the most popular online communities, and found that the degree distribution in the network follows a pattern close to the power law.

[4]A scale-free network is a network in which the distribution of connectivity is extremely uneven in the sense that some nodes act as very connected hubs using a power-law distribution.

[5]Evidences for preferential attachment have been documented in literature. See, for example, Jeong *et al.* (2003) identified the evidence of preferential attachment in the science citation network, the actor collaboration network, the science coauthorship network.

the consent of the other. Third, a logical justification for the assumption of preferential attachment is lacking. No explanation is provided for why people prefer to attach their links to others who have more links. The assumption of preferential attachment may be reasonable if links are formed unilaterally, but it seems not convincing in situations in which forming a link requires mutual consent, because a proposal of a newcomer (who has no link) to form a link with an incumbent who has many links will be unlikely to be accepted by the incumbent.

In this paper, we provide several models that overcome the three drawbacks. By analyzing the models, we can explain preferential attachment as rational equilibrium behavior.

The fundamental assumption driving our result is that people are not certain of the value that they can obtain to form a link with someone. Based on this assumption, we will argue that a person has an incentive to form a link with another who has many links because the number of links of a person can convey some information about his value, in an economic jargon, the number of links can be a signal of the value of the person.[6] Then, why a person with many links would be willing to accept the new comer with no link? This is possible because no link of a new comer is not a signal that the value of the new comer is low. His linklessness is a consequence of no chance of being connected, but not a consequence of his low value. Therefore, if the *ex ante* probability that a person is valuable is reasonably high, it is expected to be beneficial for an incumbent to accept the proposal of a new comer to form a link with him. To the best of our knowledge, our work seems the first paper in literature on network formation that views a link with a neighbor as an experience good,[7] in the sense that each agent does not realize the true value of others before he pays a price (linking cost) to form a link, thereby interacting with another.

The paper is organized as follows. In Section 2, we briefly review the physical literature related to network formation/evolution. After introducing some relevant definitions from the graph theory in Section 3, we provide the basic model in Section 4. In Section 5, we analyze the problem of network formation in the case that new entrants are informed of the order of entry of the existing population members. In Section 6, we analyze the alternative but more realistic case that new entrants are not informed. In Section 7, we consider far-sighted

[6]Economists use the word "signal" to mean an observable variable containing some information about an unobservable characteristic. This concept has been widely used in economics since Spence (1973).

[7]Economists classify goods roughly into two categories, search goods and experience goods. The former refers to goods whose quality can be ascertained by consumers before purchase, and the latter refers to goods whose quality is learned only after purchase. This taxonomy follows from Nelson (1970).

agents who make a decision to sever their link strategically not in a way to maximize their one-period payoff. In Section 8, we discuss the implication of bounded rationality that is implicitly assumed in the physics literature and derive some simulation results based on the assumption of bounded rationality. Section 9 contains concluding remarks and some caveats.

2 Related Literature

In a pioneering work, Price (1965) demonstrated that the distributions of in-degrees and out-degrees in citation networks of science articles both follow the power law. In his subsequent paper (1976), he provided an analytic explanation for this phenomena and called the underlying force cumulative advantage which is essentially the same as preferential attachment named by Barabási and Albert (1999). His exposition is based on the assumption that when a new node (scientific article) is added into a network, the probability that the node will cite another existing node is proportional to the in-degree of the cited node, k. If the distribution of in-degrees is denoted by P_k, $P_k \sim k^{-(2+1/m)}$ where m is the average of out-degrees.

Although Price (1965) is relevant to explaining citation networks, it is not so satisfactory as to explain diverse networks (including www) that exist in the real world. Barabási and Albert (1999) provide a model of undirected networks by assuming that the probability that a new node j and an existing node i are linked is given by $k_i/\sum_{l\in N_{j-1}} k_l$, where N_{j-1} is the set of nodes that has existed prior to node j.

Bianconi and Barabási (2001) introduce the concept of fitness into the model. The fitness η_i of a node i is determined randomly from a distribution $\rho(\eta)$. Then, the probability that a new node j is linked with a node i is given by $\eta_i k_i/\sum_{l\in N_{j-1}} \eta_l k_l$. Under this assumption, even a late comer can be a new hub if its fitness is high enough. This can explain the observation of low correlation between the number of links of a node and its age. The concept of fitness is analogous to the value in our model. But, our model has an important distinction from their model in that the value itself is not observable to a new entrant, so that the decision of an entrant cannot be based on the value itself.

In the economics literature, Jackson and Wolinsky (1996) is seminal in the sense that they first introduced a formal model of the social network formation with the important feature that forming a link is costly.[8] Their model is innovative as well in the sense that

[8] Aumann and Myerson (1988) who considered the network formation process as a non-cooperative game

they depart from the traditional non-cooperative game theoretic approach by assuming that a link between two nodes (agents) is formed by the bilateral decision of the agents. This view has been inherited to many subsequent authors. Our paper also borrowed their view of bilaterality.

The original paper by Jackson and Wolinsky (1996) involves the static model and the static solution concept, what they call, pairwise stability, although their model and their solution concept can be easily interpreted as a dynamic one. Watts (2001) first provided a dynamic network formation model and showed the possible disparity between the static stability and the dynamic stability in a deterministic dynamic model. Jackson and Watts (2002) considered a stochastic dynamic model by introducing randomness into the model and showed that the dynamic process of network formation will not get stuck at an unstable network, if the process is stochastic so that any network can be reached with positive probability. While, in most dynamic models including Watts (2001) and Jackson and Watts (2002), agents are assumed to be myopic so as to neglect the reaction of others when they decide to add or sever a link, there are a few of articles that consider farsighted agents who make linking decisions by taking account of its subsequent effect on others' linking decision in the future. Watts (2002) and Page, Wooders and Kamat (2005) are notable ones among them.

3 Preliminaries

Let N be the set of players (or agents). We define a graph (or network) g on N by a set of links (a pair of linked players), i.e., $g \equiv \{(i,j) \mid i.j \in N, i \neq j\}$. We denote by $g + (i,j)$ the graph obtained by adding the link (i,j) to g, and by $g - (i,j)$ the graph obtained by deleting the link (i,j) from g. If $g' \subset g$, g' is called a subgraph of g. We denote the number of links in g by $|g|$. A player j is called adjacent to player i or a neighbor to player i in g if $(i,j) \in g$. The number of neighbors to player i in g is called the degree of player i and denoted by $d(i;g)$. If $d(i;g) = 0$, we call player i isolated in g.

We say that i and j are (indirectly) connected if there is a path between them, i.e., a sequence $\{i, i_1, i_2, \cdots, i_k, j\}$ such that $(i,i_1), (i_1,i_2), \cdots, (i_k,j) \in g$. If there is a path between i and j in g for all distinct pair of players $i,j \in N$, we call g connected. If g is not connected, it is partitioned into more than one connected subgraph called component, $C(g)$. A component

preceded Jackson and Wolinsky (1996), but they did not consider the linking cost.

$C(g)$ is said to be empty if $|C(g)| = 0$. If g has only one nonempty component, we say that g satisfies [UNC].

4 Model

We consider a dynamic model in which one agent enters the population sequentially in each period, $t = 1, 2, \cdots, \infty$. Let $N_t = \{1, 2, \cdots, t\}$ be the set of agents in period t. Agent i is called the youngest (or foremost junior) of N_t. Agents in N_{t-1} are seniors (or predecessors) to agent t. In particular, agent $t-1$ is called the immediate senior to agent t. The value of agent t is v_t. If agent $i \neq t$ is linked with the agent t, he can share the value of v_t, for example, by accessing the information that agent t has. In other words, from the link (i, t), agent i and agent t both enjoy the total value of $v_i + v_t$. We assume that there is no benefit from indirect connections. For example, if $(i, j), (j, k) \in g$, the total value that agent i can get from this network is simply $v_i + v_j$, not $v_i + v_j + v_k$.[9]

The value of v_t is private information of agent t, i.e., no agent $i \neq t$ is *a priori* informed of v_t. In the game theoretic jargon, we call v_t the type of agent t. For simplicity, we assume that v_t is either L or H where $L < H$, and that v_t is i.i.d. with $\text{Prob}(v_t = H) = \alpha \in (0, 1)$.

At each period t, a new entrant into the population, who maximizes his net benefit, chooses a single agent from N_{t-1} to propose to link with him. If the proposed accepts the proposal made by agent t, the link is formed. Forming a link incurs $c(> 0)$ to both parties. We assume that $L < c < H$ and that

$$M \equiv \alpha H + (1 - \alpha)L > c. \qquad [A1]$$

Under this assumption, if a new comer has no information about another agent's type, he prefers linking with him to not linking. Thus, this assumption excludes the possibility that an entrant is not willing to link with any other. Also, due to this assumption, the proposed always accepts the proposal.

An agent can realize the true value of his neighbor in one period of interaction after forming a link. Upon realizing that the neighbor's value is L, an agent may sever his link. Thus, each period consists of two stages. In the first stage, the entrant forms a link with an

[9]If agents can get some benefit from indirect connections as well, they may have an incentive to behave strategically, since their utility would also depend on the decision of other agents. This would complicate the analysis significantly.

existing agent, and in the second stage, an existing agent can sever his link. More specifically, (i, j) is severed in the period of $\max\{i, j\} + 1$, if either $v_i = L$ or $v_j = L$.

We assume that all agents are myopic in the sense that they do not consider the possibility that their current decision will affect future entrants' decisions.[10] Although there are many agents interacting with each other in this model, the model will be analyzed as if it were a single-person decision problem in each period, as far as we assume myopic agents.

We will perform our analysis, based on two alternative assumptions; one corresponding to a case that each entrant can tell the seniority, that is, the entry order of existing agents perfectly (Section 5), and the other corresponding to a more realistic case that he cannot tell the seniority at all (Section 6).

5 When the Seniority of the Population is known to the New Entrant

Let the graph formed at period t be $g_t \equiv (g_{t,1}, g_{t,2})$. The former is a graph obtained by agent t's linking decision and the latter is a graph obtained after agent $t-1$ and his neighbor exercise the option of deleting their link.

Let $\alpha_t(i)$ be the agent t's posterior belief or probability that $v_i = H$ (evaluated in the beginning of period t). We will use notation of $i \succ_t j$ if player t strictly prefers linking with i to linking with j, or equivalently, if $\alpha_t(i) > \alpha_t(j)$, and $i \sim_t j$ if player t is indifferent between linking with i and j, of equivalently, if $\alpha_t(i) = \alpha_t(j)$.

We will start our analysis from $t = 2$. It is clear that $(1, 2) \in g_{2,1} = g_{2,2}$. At $t = 3$, agent 3 links to either agent 1 or agent 2, since $\alpha_3(1) = \alpha_3(2) = \alpha$ so that $1 \sim_3 2$. Without loss of generality, assume that $(2, 3) \in g_{3,1}$. In the second stage of $t = 3$, agent 1 or agent 2 can sever his link if either $v_1 = L$ or $v_2 = L$. So, we have $g_{3,2} = \{(2, 3)\}$ with probability $(1-\alpha^2)$ and $g_{3,2} = \{(1, 2), (2, 3)\}$ with probability α^2.

Consider the decision of agent 4. If $g_{3,2} = \{(2, 3)\}$, it implies that $(1, 2)$ has been deleted, in turn implying that either $v_1 = L$ or $v_2 = L$. Thus, the posterior probability that $v_1 = H$ (or $v_2 = H$) is $\alpha/(1 + \alpha)$. Since the probability that $v_3 = H$ is α, agent 4 forms a link with agent 3. In this case, no link between agent 1 and agent 2 is a bad signal for both v_1 and v_2. If $g_{3,2} = \{(1, 2), (2, 3)\}$, agent 4 links either to agent 1 or to agent 2, because they are

[10] If agents are not myopic, their decisions to sever a link can be affected. We will consider this possibility in Section 7.

sure to be H. In this case, a link between agent 1 and agent 2 is a good signal for both of their values. In the second stage, link $(2,3)$ may be deleted in both cases if either agent 2 or agent 3 turns out to be L. Then, agent 5 faces similar situations, and so on.

We can generalize our reasoning as follows.

Lemma 1 *If $(i,j) \in g_{t-1,2}$ for $i,j < t-1$, then $\alpha_t(i) = \alpha_t(j) = 1$.*

Proof. If $v_i = L$, $(i,j) \notin g_{t-1,2}$, since the link (i,j) should have been severed by $j \neq t-1$. The same argument holds for the case that $v_j = L$. $\|$

This lemma says that a player linked with other than the youngest must be a H type. If $(i,t-1) \in g_{t-1,2}$, we cannot be sure that $v_i = H$, because neither i nor $t-1$ is given the opportunity to sever the link until the second stage of period t. We provide a sequence of lemmas characterizing the optimal decision of agents and the resultant network structure.

Lemma 2 *For $i \neq t-1$, (i) $t-1 \succ_t i$ if $d(i; g_{t-1,2}) = 0$, (ii) $t-1 \prec_t i$ if $(i,j) \in g_{t-1,2}$ for some $j < t-1$, and (iii) $t-1 \succ_t i$ if $(i,t-1) \in g_{t-1,2}$ and $d(i; g_{t-1,2}) = 1$.*

Proof. See the appendix.$\|$

Lemma 3 *If $d(i; g_{t,\tau}) = 0$, then $d(i; g_{t',\tau'}) = 0$ for all $t' \geq t$, $\tau, \tau' = 1, 2$.*

Proof. If $d(i; g_{t,\tau}) = 0$, $t \succ_{t+1} i$ by Lemma 2(i), implying that $d(i; g_{t+1,\tau}) = 0$. Induction leads to $d(i; g_{t',\tau}) = 0$ for all $t' > t$. $\|$

This lemma implies that a once isolated agent will remain isolated forever.

Lemma 4 *There is the unique nonempty component in $g_{t,\tau}$ for any t and τ.*

Proof. See the appendix.

This lemma says that the network evolves with maintaining connectivity except for isolated players.

From Lemma 2 together with Lemma 3 and 4, we can infer that an entrant links to one of his remote (or not immediate) seniors if there is a link between them, while otherwise he links just to his immediate senior and that the latter case occurs only when all the previous links except the most recent link has been deleted; otherwise, [UNC] implies that the agent to whom the youngest senior links cannot have degree one. The intuition for this decision rule is that the preservation of such an old (not recent) link signals the high value of the linking agents.

Now, we will define the notion of "preferential attachment." By the property of "preferential attachment," we will mean that all agents rationally link to one of the agents with the largest number of links. Then, we have the following proposition.

Proposition 1 *When each entrant can tell the seniority of his predecessors, there is an equilibrium in which the property of preferential attachment is satisfied.*

Proof. See the appendix.

Note that preferential attachment is not the unique rational behavior outcome. For example, if agent t links randomly with some remote senior linked with another remote senior and links with his immediate senior if there is no such remote senior, it is also a rational behavior but does not satisfy preferential attachment, because agent $t-1$ may link with a remote senior with degree one. In other words, there are also other equilibria in which preferential attachment is not satisfied. In this sense, we can say that preferential attachment holds but only weakly in this case.

6 When the Seniority of the Population is unknown to the New Entrant

In this section, we consider a more realistic case that a new entrant does not tell the order of the entries of the existing agents.

To illustrate, let us start by considering g_2. Clearly, $g_{2,1} = g_{2,2} = \{(1,2)\}$. At period 3, player 3 links with either player 1 or player 2. Assume that $(2,3) \in g_{3,1}$. Then, $(1,2)$ can be deleted if $v_1 = L$ or $v_2 = L$. If $g_{3,2} = \{(2,3)\}$, player 4 will link to either player 2 or player 3, because $\alpha_4(1) = \alpha/(1+\alpha)$, $\alpha_4(2) = \alpha_4(3) = \alpha/2 + \alpha/2(1+\alpha)$. If $g_{3,2} = \{(1,2),(2,3)\}$, player 4 will link with player 2 since $\alpha_4(2) = 1$.

It is easy to see that Lemma 1 and Lemma 2 cannot be carried over to this case of uninformed players. However, we can provide a counterpart for them in terms of the degree which is observable.

Lemma 5 *At any t, if $d(i; g_{t-1,2}) \geq 2$, $\alpha_t(i) = 1$.*

Proof. See the appendix.

This lemma says that the degree of a low type must be at most one. However, this lemma does not imply that a player whose degree is one must be of L type because he may be the immediate senior to the new entrant (whose type can be H) or a remote senior who severed his links except one.

Lemma 6 *For any t, (i) $i \sim_t j$ for any i, j such that $d(i; g_{t-1,2}) = d(j; g_{t-1,2}) \geq 2$, (ii) $i \succ_t j$ for any i, j such that $d(i; g_{t-1,2}) \geq 2 > d(j; g_{t-1,2})$ and (iii) $i \succ_t j$ for any i, j such that $d(i; g_{t-1,2}) = 1 > d(j; g_{t-1,2}) = 0$.*

Proof. See the appendix.

Due to Lemma 6(iii), the counterpart for Lemma 3 holds.

Lemma 7 *Lemma 3 holds when the seniority is unknown.*

Proof. It is clear from Lemma 6(iii) since $d(t-1; g_{t-1,2}) = 1$.

Lemma 6 and Lemma 7 suggest that an entrant links to one of his seniors whose degree is more than one if any, while otherwise he links to one of seniors with degree one. The latter case is possible only when there is only one link involved with the last entrant. In fact, in the equilibrium, one player becomes a hub (a center) and each entrant links to a hub, thereby forming a star except isolated players. In this network, only one player can have the degree of more than one.

The result of rational preferential attachment is strengthened in this case of uninformed players, as summarized in the following proposition, and accordingly we can say that preferential attachment holds strongly in this case. Again, this is the consequence of the signalling effect of a hub.

Proposition 2 *When each entrant cannot tell the seniority of his predecessors, the property of preferential attachment is satisfied in equilibrium. Moreover, preferential attachment is the only equilibrium outcome.*

Proof. This is clear from the above argument.

7 Far-Sighted Agents

In this section, we assume that agents are not myopic but far-sighted. So, when they make a decision, they take into account the possibility that their decision will affect the decision of future entrants.

We assume that each agent maximizes discounted future net benefits. Let the discount factor be $\delta(>0)$. To accommodate the possibility of strategic behavior, we will assume that agent t can sever his link at any period $t' \geq t+1$.

Since it is always beneficial in *ex ante* terms for an existing agent to form a link with the new entrant under [A1], an L type agent has an incentive to pretend to be a H type, while a H type has no incentive to pretend to be an L type. If agents are so far-sighted that they behave strategically to affect the linking decision of future entrants, equilibria characterized so far will not be valid.

To illustrate, consider the case of informed agents. Suppose that $g_{3,1} = \{(1,2),(1,3)\}$. If agents were myopic, the link $(1,2)$ would be severed when $v_1 = L$ or $v_2 = L$. If they are far-sighted, however, they may want to maintain this link, because a deletion conveys a bad signal to the entrant and thus, he will propose a new link to neither of them. Of course, such a feint will be profitable only when the benefit (of increasing the probability of obtaining a new link) exceeds the cost of maintaining a link with a low type neighbor. Specifically, they have an incentive to postpone severing the link one period if

$$c - L \leq \frac{\delta}{2}(M-c), \qquad [A2]$$

or equivalently, $\delta \geq 2\frac{c-L}{M-c}$ or $\alpha \geq (\frac{2}{\delta}+1)\frac{c-L}{H-L}$. Thus, as δ or α is higher, the benefit from feigning gets higher, so players are more likely to postpone deleting their links.

Now, consider the case of uninformed agents. Suppose again that $g_{3,1} = \{(1,2),(1,3)\}$. When $v_2 = L$, agent 1 has an incentive to maintain his link with agent 2 if

$$c - L \leq \delta(M-c).^{11} \qquad [A3]$$

Note that it is more difficult for the myopic equilibrium to be sustainable in the case of uninformed case, because the gain that agent 1 can get from one-period postponing is larger as far as he can get an additional link with higher probability (probability one).

[11] If $v_1 = L$, agent 2 gains nothing by postponing severing his link (regardless of $v_2 = H$ or L), because subsequent entrants will keep forming links with agent 1 anyhow.

Proposition 3 *Assume that agents are far-sighted. (i) When agents are informed, the myopic equilibrium characterized by Lemma 2 is sustainable if and only if [A2] does not hold. (ii) When agents are uninformed, the myopic equilibrium characterized by Lemma 6 is sustainable if and only if [A3] does not hold.*

Proof. See the appendix.

Henceforth, let us focus on the interesting case that [A2] (or [A3] respectively) holds so that the myopic equilibrium is not sustainable. It is too complicated to fully characterize the equilibrium in either case (of informed agents or uninformed agents), since the characterization will depend on each realization of v_t. But, we can provide the intuitive description for the equilibrium configuration. Postponing severing a link is beneficial to early entrants since it significantly increases the chance of obtaining another link with a new entrant. So, all agents in earlier periods maintain their links for some periods regardless of their neighbor's type, which is called a pooling equilibrium in the game theoretic jargon.[12] However, if t is very large, the benefit from postponing deleting a link is very small, because the chance that he can obtain another link is very low as long as a large number of existing agents maintain their links. Hence, all agents after some large t will delete their link as soon as they realize that their neighbor's type is L, which corresponds to a separating equilibrium.

Proposition 4 *With probability one, there exists t^* such that for all $t \geq t^* + 1$, agent t behaves just as if he were myopic and $(i, j) \notin g_t$ for $i, j \leq t^*$ if $v_i = L$ or $v_j = L$.*

Proof. See the appendix.

The upshot is that players begin from the pooling phase and then eventually enter the separating phase. For period $t \geq t^* + 1$, any agent i with $d(i; g_t) \geq 2$ is clearly of type H. This implies that there is an equilibrium in which preferential attachment is not satisfied when $t \geq t^* + 1$. This suggests that we can hardly expect the power law.

[12]In a pooling equilibrium, players choose the same action (postponing severing their links) regardless of their private information (neighbor's types), while a different type of player chooses a different action in a separating equilibrium.

8 Discussions and Simulations

Although we have succeeded in explaining strong preferential attachment in Section 6,[13] the resulting network, a star with only one hub, is far from realistic. In this section, we add a more realistic ingredient.

Suppose that agents cannot observe the number of links of all their predecessors but only k predecessors due to significant search costs. To be realistic, we consider myopic agents who are not informed of the seniority of their predecessors, as assumed in Section 6. In any period t, if entrant t observes his predecessor i with $d(i) \geq 2$, he can be sure that $i = H$, i.e., $\alpha_t(i) = 1$ if $d(i) \geq 2$. If he observes a predecessor i with $d(i) = 1$, he may be either the immediate predecessor or otherwise type H, so that $\alpha_t(i) > \alpha$ if $d(i) = 1$, since $\alpha_t(i)$ is a weighted average of α and 1. Also, we know that $0 < \alpha_t(i) < \alpha$ if $d(i) = 0$. Therefore, preferential attachment holds even in the case of limited observability due to search costs.

According to this reasoning, we will take the following simulation procedure. First, in every period t, the entrant picks k nodes randomly. If $t < k$, he chooses all existing nodes. Second, the entrant links to the node whose degree is maximal among the randomly selected k nodes. If there are multiple nodes with the maximal degree, link any one of them randomly. If all selected nodes have degree zero, link any node randomly.

Figure 1 illustrates resulting network structures when $\alpha = .5$ and $N = 500$ when $k = 2, 10, 100, 400$ respectively. As we can correctly conjecture, resulting networks are almost random networks when k is small and then hubs begin to emerge as k becomes larger, and finally the star network emerge when $k = 400$. Figure 2 shows the degree distribution. In the larger figure, only the vertical axis is log-scale, while both axes are log-scale in the smaller inset. Interestingly, we can confirm that the degree distribution is smoothly transformed from exponential functions into following the power law (except the tail part) as k gets larger. Appearance of a small hump in the tail part when $k = 100, 1000$ is due to the fact that the hubs generated in the initial phase (when $t < k$) grow faster than other nodes as time goes.

An alternative approach to generate realistic network structures is to introduce bounded rationality, i.e., the possibility that agents can make errors in calculations or choices. Hereafter, we will briefly discuss a rationale for "probabilistic preferential attachment" which can be defined as "a node with more links is more likely to be connected with the new node."

[13] In addition, we explained weak preferential attachment in Section 5 and long-run weak preferential attachment in Section 7.

For the purpose, we will utilize the concept of "Properness" coined by Myerson (1978) in game theory.[14] The properness requires agents to choose an alternative yielding a lower net benefit with lower probability. It can be justified by the argument that an agent will be less likely to make a more serious mistake since he will be more cautious. We will incorporate "bounded rationality" defined based on properness into our model. For simplicity, we will assume that agents can make non-optimal linking decision, but that they do not make a mistake in a severing decision.[15]

Let r_t be a behavioral strategy of entrant t,[16] i.e., $r_t = (r_t(i))_{i \leq t-1}$ where $r_t(i)$ is the probability that an entrant t chooses agent i as his linking partner and

$$\sum_{i \in t-1} r_t(i) = 1. \tag{1}$$

At $t = 2$, it is clear that $r_2(1) = 1$ by equality (1). At $t = 3$, $r_3(1) = r_3(2) = 1/2$ by Properness. Without loss of generality, assume that $(2, 3) \in g_{3,1}$. If $(1, 2)$ is deleted in the second stage of $t = 3$, $r_4(1) < r_4(2) = r_4(3)$. If $(1, 2)$ is preserved in the second stage of $t = 3$, $r_4(2) > r_4(1) = r_4(3)$ since $\alpha_4(2) > \alpha_4(1) = \alpha_4(3)$. More generally, suppose that $d(i) = 2$ and $d(j) = 1$ for $i, j < t$. Then, $r_t(i) > r_t(j)$ by Properness, since $d(j) = 1$ implies that agent j (or his partner) was L type or he is the immediate senior. Now, suppose that $d(i) = 3$ and $d(j) = 2$. If $d(i) = d(j) = 2$ in the immediately preceding network, it is clearly off the equilibrium, since only one node can have more than one link in equilibrium, (because the equilibrium network is a star). But, we cannot tell which is more likely to be H. If the earlier phase was that $d(i) = 3$ and $d(j) = 1$ in the immediately preceding network, it is highly likely to be an equilibrium. Thus, node i is more likely to be H. In turn, Properness implies that $r_t(i) > r_t(j)$. This argument can be extended to any i, j such that $d(i) > d(j)$. The point here is that an entrant links with higher probability to a node more likely to be H which turns out to be equivalent to a node with more links. Although our approach closely follows the spirit of Properness, it is distinguished from the proper equilibrium of Myerson in the sense that we allow mistakes in equilibrium while Myerson allows mistakes only off the equilibrium.

[14]This notion follows the spirit of "trembling hand perfectness" by Selten (1975) allowing some mistakes of players.

[15]All that is needed in making a decision of severing a link with an agent is to know whether he is L or H, while a linking decision requires much more information (probabilities that each senior is of H type).

[16]A behavioral strategy, which involves randomization over pure actions at an information set, is distinguished from a mixed strategy involving randomization over pure strategies. For a relation between the behavioral strategy and the mixed strategy, see Kuhn (1953).

Under the assumption of bounded rationality, we conduct simulations for our model. Figure 3 is a simulation result that is repeated 1,000 times when $\alpha = .5$ and $N = 100,000$, showing that the resulting degree distribution follows the power law $P(d) \sim d^{-\gamma}$ with the exponent $\gamma \approx 2.9$. Figure 4 illustrates the resulting network structure when $\alpha = .5$ and $N = 1,000$. A few hubs of H type emerge, leading to a hierarchical structure,while isolated nodes are ones found to be L types. From simulations, it is also found that the oldest nodes do not necessarily become hubs due to several reasons. The oldest ones may be L types or they may not be chosen to be linked with new entrants. Figure 5 shows the index distribution of hubs over 1,000,000 realizations of the model with $N = 1,000$. In an extreme case, even the 214-th entrant becomes a (local) hub as shown in Figure 5.

9 Concluding Remarks and Caveats

In this paper, we provided an economic explanation for preferential attachment, which is usually assumed in the physics literature, in the context of social networks requiring mutual consent of the linking partners. We argued that a newly born agent may prefer to form a link with an agent with a high degree on the ground that the high degree may signal his high value. Meanwhile, we also argued that if agents have perfect rationality in the sense that they do not make mistakes, preferential attachment can be supported only weakly unless uninformed agents are myopic.

If uninformed agents are myopic, preferential attachment is supported strongly in equilibrium, but it does not imply the power law. Since preferential attachment is so strongly supported then, only a star network is viable. We may consider generalizing our model of binary value of v_t to a continuum value. However, this will not lead to the power law. Since existing agents i make their severance decision, depending only on whether $v_i \geq c$ or $v_i < c$, the number of links of an agent only conveys information about whether his value exceeds c or not, but no more than than. Consequently, for any $d_i > d_j > 2$, there is no convincing evidence for $v_i > v_j$. We may be able to infer that $v_i > v_j$ if the linking cost of agent i and j, c_i and c_j, are different, because if v_i is large, agent $j \neq i$ may want to link with him even if c_j is high, implying that a higher degree can be a signal of a higher value. Although this may be an interesting extension, the resulting network will be still close to a star network, because once a particular node obtains a relatively large number of links, new entrants will want to link only with the node with probability one. This suggests that introducing bounded rationality similar to Properness seems indispensable to generating scale-free networks whose link distributions follow the power law.

We look forward to the emergence of more realistic and sophisticated models explaining social networks with bilateral links in the near future.

Appendix

Proof of Lemma 2: (i) Since player i has been once linked with some j, either $v_i = L$ or $v_j = L$ if $d(i; g_{t-1,2}) = 0$. Thus, $\alpha_t(i) = \frac{\alpha}{1+\alpha} < \alpha_t(t-1) = \alpha$. (ii) If $(i,j) \in g_{t-1,2}$ for some $j < t-1$, $\alpha_t(i) = 1 > \alpha_t(t-1) = \alpha$ by Lemma 1. (iii) Let $(i,j) \in g_{i,1}$. Clearly, $j < i$. If $(i, t-1) \in g_{t-1,2}$ and $d(i; g_{t-1,2}) = 1$, this implies that (i,j) has been deleted. This implies that either $v_i = L$ or $v_j = L$. Thus, $\alpha_t(i) = \frac{\alpha}{1+\alpha} < \alpha_t(t-1) = \alpha$. ‖

Proof of Lemma 4: It is clear that $g_{2,1} = g_{2,2}$ satisfies [UNC], since it is connected. Suppose $g_{t-1,2}$ satisfies [UNC]. From Lemma 2(i), it is clear that $g_{t,1}$ also satisfies [UNC]. It only remains to show that $g_{t,2}$ satisfies [UNC]. Two possible cases are that either (i) $(i,t) \in g_{t,1}$ for $i \neq t-1$ or (ii) $(t-1,t) \in g_{t,1}$. In the case (i), suppose $(t-1,k) \in g_{t,1}$ for some $k \in N_{t-2}$ and deleting $(t-1,k)$ decomposes $g_{t,1} - (t-1,k)$ into more than one nonempty component. This means that $d(t-1; g_{t,1}) \geq 2$, which is not possible. Consider the other case (ii). Since $g_{t-1,2}$ satisfies [UNC], $d(t-1; g_{t-1,2}) = 1$. Let the neighbor of player $t-1$ be $k \in N_{t-1}$, i.e., $(t-1,k) \in g_{t-1,2}$. If $d(k; g_{t-1,2}) = 1$, $g_{t,2} - (t-1,k)$ still satisfies [UNC]. If $d(k; g_{t-1,2}) \geq 2$, it is a contradiction to $(t, t-1) \in g_{t,1}$ due to Lemma 2(ii). ‖

Proof Proposition 1: For any t, define

$$\begin{aligned} N^0_{t-1} &= \{i \in N_{t-2} \mid d(i; g_{t-1,2}) = 0\}, \\ N^1_{t-1} &= \{i \in N_{t-2} \mid d(i; g_{t-1,2}) = 1\}, \\ N^2_{t-1} &= \{i \in N_{t-2} \mid d(i; g_{t-1,2}) \geq 2\}. \end{aligned}$$

Then, $N_{t-1} = \{t-1\} \cup N^0_{t-1} \cup N^1_{t-1} \cup N^2_{t-1}$. Also, define

$$\begin{aligned} N^{1,0}_{t-1} &= \{i \in N^1_{t-1} \mid (i,j) \in g_{t-1,2} \text{ for } j = t-1\}, \\ N^{1,2}_{t-1} &= \{i \in N^1_{t-1} \mid (i,j) \in g_{t-1,2} \text{ for } j \neq t-1\}. \end{aligned}$$

From Lemma 2, it is clear that (i) $t-1 \succ_t i$ for $i \in N^0_{t-1} \cup N^{1,0}_{t-1}$ and (ii) $i \succ_t t-1$ for $i \in N^2_{t-1} \cup N^{1,2}_{t-1} \equiv \bar{N}_{t-1}$. If $\bar{N}_{t-1} \neq \emptyset$, it is optimal for player t to choose to link with player

$i \in \bar{N}_{t-1}$. If $\bar{N}_{t-1} = \emptyset$, it is optimal for him to choose to link with player $t-1$. Thus, the decision of agent t to choose to link with a player $i \in N^2_{t-1}$ if $\bar{N}_{t-1} \neq \emptyset$ and to link with a player $t-1$ otherwise satisfies preferential attachment.$\|$

Proof of Lemma 5: Suppose $v_i = L$. If $i = t-1$, $d(i; g_{t-1,2}) = 1 (< 2)$, so it must be that $i < t-1$. Then, the links of player i except the link with $t-1$ would have been deleted at latest by period $t-1$. Thus, $d(i; g_{t-1,2})$ could be at most one. Contradiction. $\|$

Proof of Lemma 6: (i) It is clear, because $\alpha_t(i) = \alpha_t(j) = 1$ by Lemma 5. (ii) This is also clear because $\alpha_t(i) = 1 > \alpha_t(j)$. (iii) We have $\alpha_t(j) = \frac{\alpha}{1+\alpha}$ for j such that $d(j; g_{t-1,2}) = 0$. On the other hand, consider player i such that $d(i; g_{t-1,2}) = 0$. Either $i = t-1$ or $i \neq t-1$. If $i = t-1$, $\text{Prob}(v_i = H) = \alpha$. If $i \neq t-1$, $\text{Prob}(v_i = H \mid i \neq t-1) = \frac{\alpha}{1+\alpha}$. Thus, $\alpha_t(i) = \theta\alpha + (1-\theta)\frac{\alpha}{1+\alpha}$ for some $\theta \in (0,1)$. Since $\alpha > \frac{\alpha}{1+\alpha}$, $\alpha_t(i) > \frac{\alpha}{1+\alpha} = \alpha_t(j)$. $\|$

Proof of Proposition 3: (i) ($\Longrightarrow$) Under [A2], the myopic equilibrium is not viable, as argued in the text, so this is clear. ($\Longleftarrow$) Note that $\frac{\delta}{2}(M-c)$ is the maximum gain attainable by postponing severing a link for one or more than one period, while $c - L$ is the minimum loss from postponing severing a link. (Postponing severance for more than one period incurs $c - L + \delta(c-L) + \cdots$.) Thus, if $c - L > \frac{\delta}{2}(M-c)$, no other deviation will be profitable, implying that the myopic equilibrium is sustainable. (ii) The proof is similar. $\|$

Proof Proposition 4: Let $x(t)$ be the number of H types in N_t, that is, $x(t) = \sum_{i=1}^{t} x_i$ where $x_i = 1$ if $v_i = H$ and $x_i = 0$ if $v_i = L$. Since $E(x_i) = \alpha$ and $\text{Var}(x_i) = \alpha(1-\alpha)$, we have $\mu \equiv E(x(t)) = \alpha t$ and $\sigma^2 \equiv \text{Var}(x(t)) = \alpha(1-\alpha)t$.

Suppose that for some t^* $x(t^*) = n$ for some large n such that $L - c + \delta\frac{M-c}{n} < 0$, or equivalently, $n > \delta\frac{M-c}{c-L}$. Then, for $i, j \leq t^* - 1$, link (i,j) with $v_i = L$ or $v_j = L$ will not survive the second stage of period t^*. Knowing this, it is optimal for agent i ($i \geq t^* + 1$) to choose the equilibrium strategy of a myopic agent.

It remains to show that with probability one, there exists t^* such that $x(t^*) = n$, in other words, that $x(t) < n$ for all t with probability zero. By Chebyshev's inequality, we have

$$P(|x(t) - \mu| > k\sigma) \leq 1/k^2, \text{ for all } k > 0. \tag{2}$$

Let $1/k^2 = \epsilon$, i.e., $k = 1/\sqrt{\epsilon}$. Then, inequality (2) can be written as

$$P(|x(t) - \mu| > \sigma/\sqrt{\epsilon}) \leq \epsilon, \ \forall \epsilon.$$

Note that inequality (3) holds for any t. Thus, we have $P(x(t) < \mu - \sigma/\sqrt{\epsilon}) < \epsilon$. Since $\mu = \alpha t$ and $\sigma^2 = \alpha(1-\alpha)t$, we have

$$P(x(t) < \bar{x}(t)) < \epsilon,$$

where $\bar{x}(t) = \alpha t - \sqrt{\frac{\alpha(1-\alpha)t}{\epsilon}}$. Note that $d\bar{x}(t)/dt > 0$ if $t \geq t_1$ for some large t_1 and that $\lim_{t\to\infty} \bar{x}(t) = \infty$. Since $\lim_{t\to\infty} \bar{x}(t) = \infty$, we can take t_2 such that $\bar{x}(t_2) \equiv \alpha t_2 - \sqrt{\frac{\alpha(1-\alpha)t_2}{\epsilon}} > n$. Take $\bar{t} = \max\{t_1, t_2\}$. Then, $P(x(t) < n) < P(x(t) < \alpha t - \sqrt{\frac{\alpha(1-\alpha)t}{\epsilon}}) < \epsilon$ for all $t \geq \bar{t}$. This implies that there exists t^* such that $x(t^*) = n$ with probability one. $\|$

References

[1] Ahn, Y.-Y.: CyWorld (in Korean), in progress, available at http://janice.kaist.ac.kr/∼gomeisa/wiki/wiki.php/CyWorld/Summary?action=print (2005)

[2] Aumann, R., Myerson, R.: Endogenous Formation of Links between Players and Coalition: An Application of the Shapley Value. In Roth, A. (Eds.), The Shapley Value. Cambridge University Press, Cambridge, U.K. (1988)

[3] Bala, V. and Goyal, S.: A Noncooperative Model of Network Formation, Econometrica **68**, 1181-1229 (2000)

[4] Barabási, A-L. and Albert, R.: Emergence of Scaling in Random Networks, Science **286**, 509-512 (1999)

[5] Dutta, B. and Mutuswami, S.: Stable Networks, Journal of Economic Theory **76**, 322-344 (1997)

[6] Bianconi, G. and Barabási, A-L. : Bose-Einstein Condensation in Complex Networks, Physical Review Letters **86**, 5632-5635 (2001)

[7] Jackson, M.: The Economics of Social Networks, based on a lecture in the 9th World Congress of the Econometric Society, available at http://www.hss.caltech.edu/ jacksonm/netect.pdf (2005)

[8] Jackson, M. and van den Nouweland A.: Strongly Stable Networks, Games and Economic Behavior **51**, 420-444 (2005)

[9] Jackson, M. and Watts A.: The Evolution of Social and Economic Networks, Journal of Economic Theory **106**, 265-295 (2002)

[10] Jackson, M. and Wolinsky, A.: A Strategic Model of Social and Economic Networks, Journal of Economic Theory **71**, 44-74 (1996)

[11] Jeong, H., Neda, Z. and Barabási, A-L.: Measuring Preferential Attachment in Evolving Networks, Europhysics Letters **61**, 567-572 (2003)

[12] Kuhn, H.W.: Extensive Games and the Problem of Information, Annals of Mathematics Studies **28**, 193-216 (1953)

[13] Liljeros, F., Edling, C., Amaral, L.A., Stanley H.E. and Åberg Y.: The Web of Human Sexual Contacts, Nature **411**, 907-908 (2001)

[14] Myerson, R.: Refinements of the Nash Equilibrium Concepts, International Journal of Game Theory **7**, 73-80 (1978)

[15] Nelson, P.: Information and Consumer Behavior, Journal of Political Economy **78**, 311-329 (1970)

[16] Page, F., Wooders, M. and Kamat, S.: Networks and Farsighted Stability, Journal of Economic Theory **120**, 257-269 (2005)

[17] Price, D.: Networks of Scientific Papers, Science **149**, 510-515 (1965)

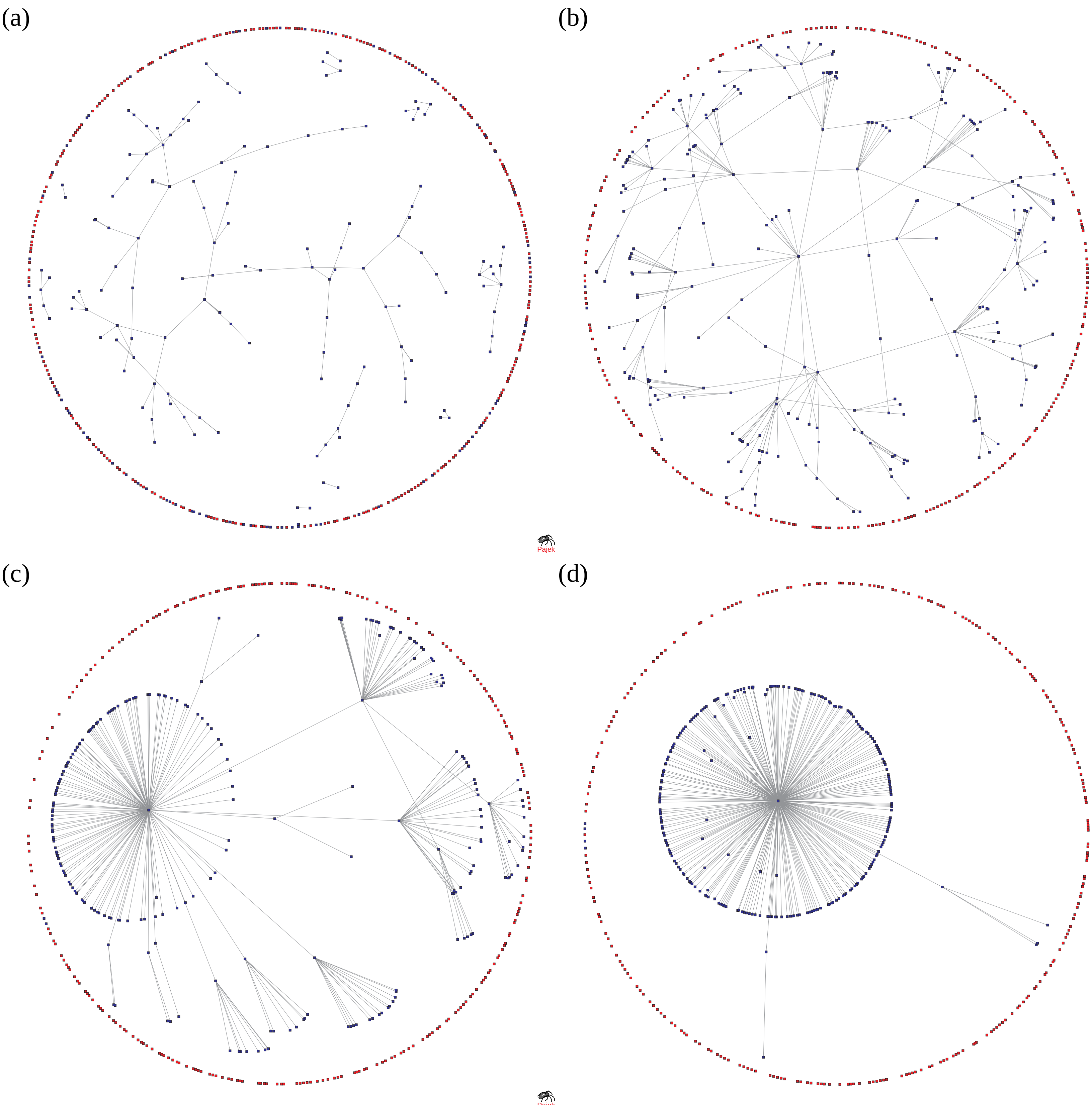
[18] Price, D.: A General Bibliometric and Other Cumulative Advantage Processes, Journal of the American Society for Information Science **27**, 292-306 (1976)

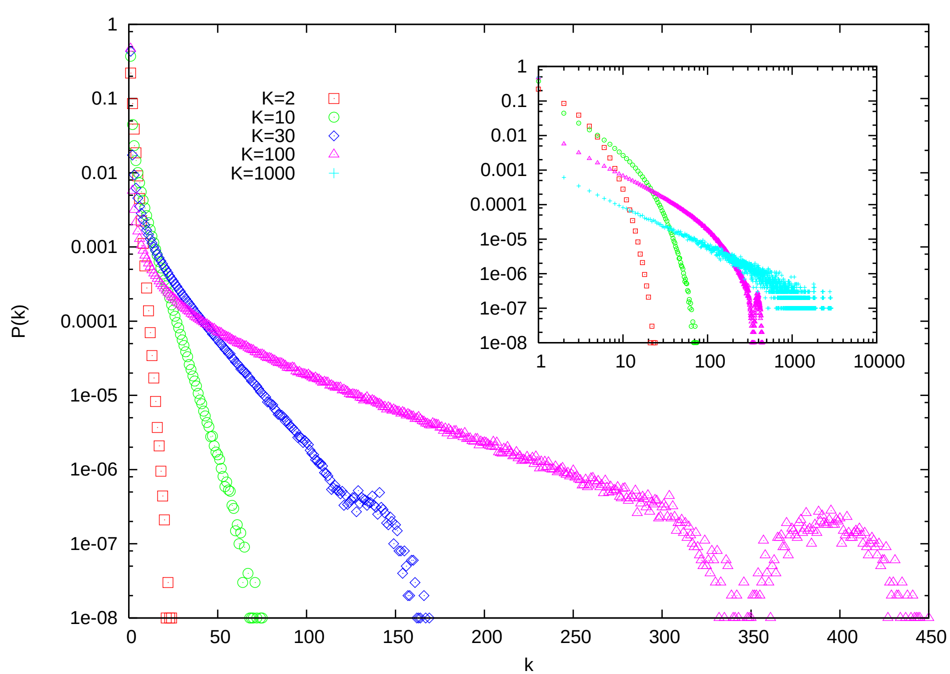
[19] Selten, R.: Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games, International Journal of Game Theory **4**, 25-55 (1975)

[20] Spence, M.: Job Market Signaling, Qarterly Journal of Economics **87**, 355-374 (1973)

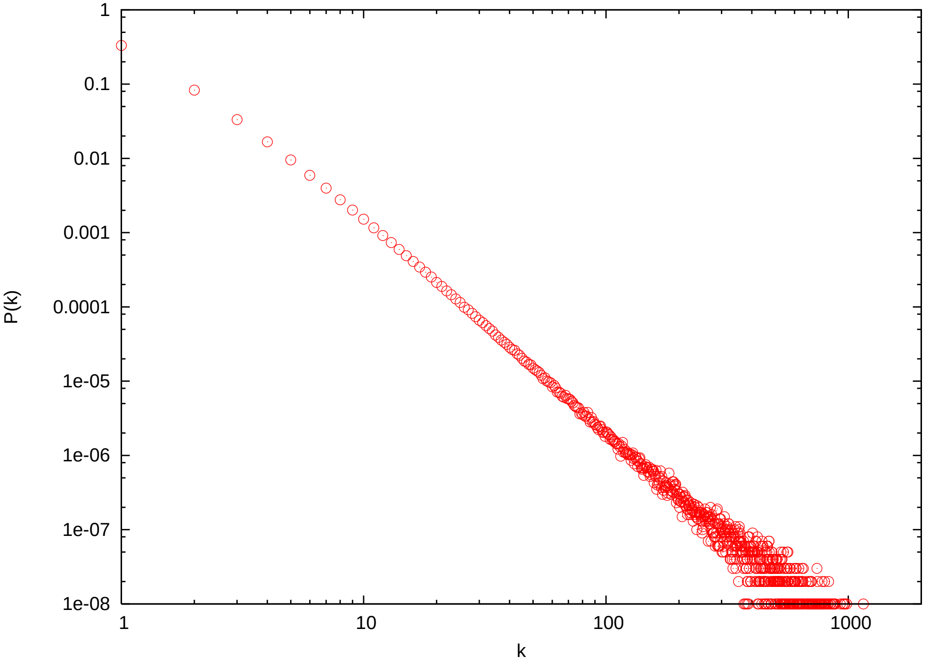
[21] Watts, A.: A Dynamic Model of Network Formation, Games and Economic Behavior **34**, 331-341 (2001)

[22] Watts, A.: Non-Myopic Formation of Circle Networks, Economics Letters **74**, 277-282 (2002)

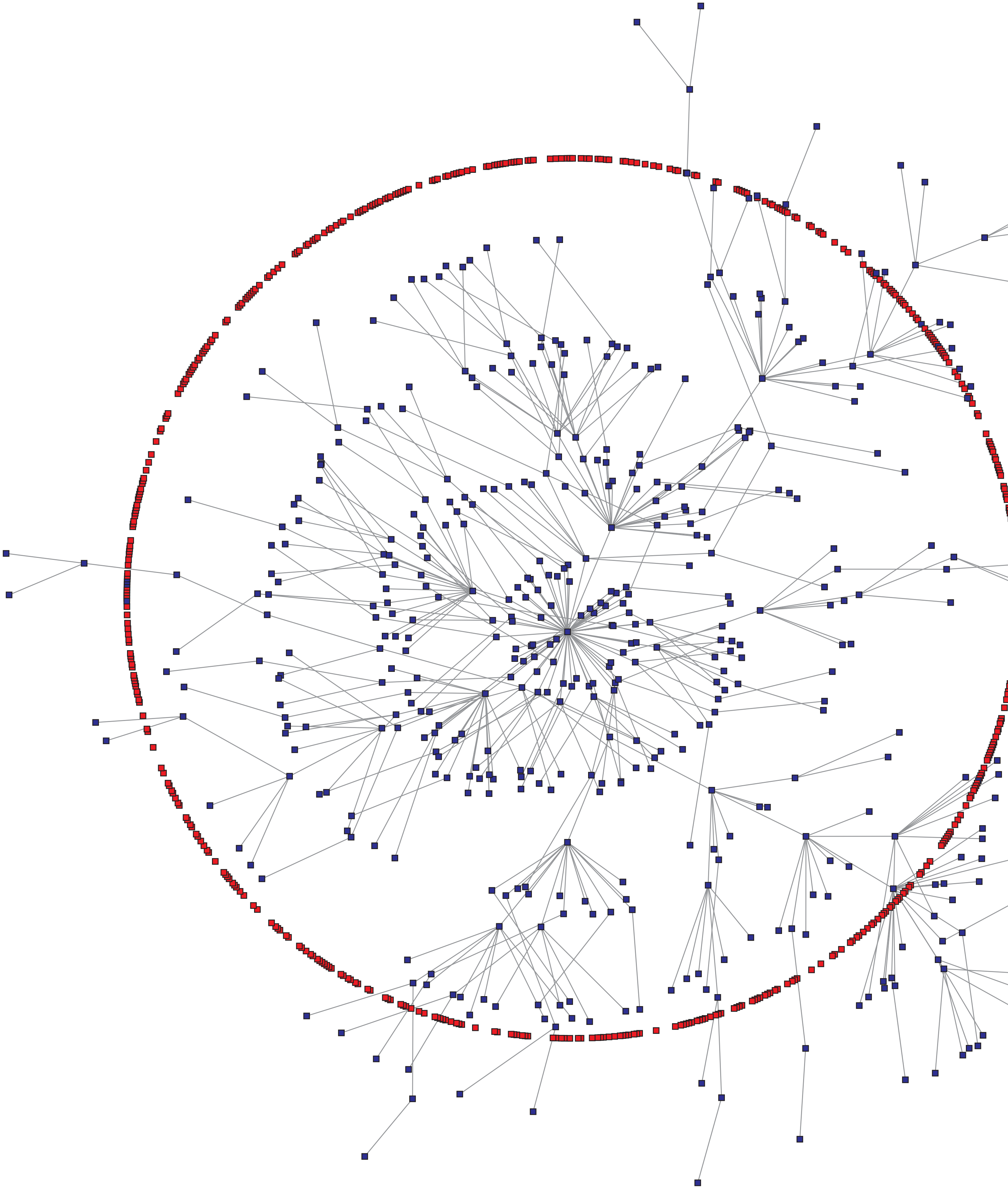

(a)
(b)
Pajek
(c)
(d)
Pajek

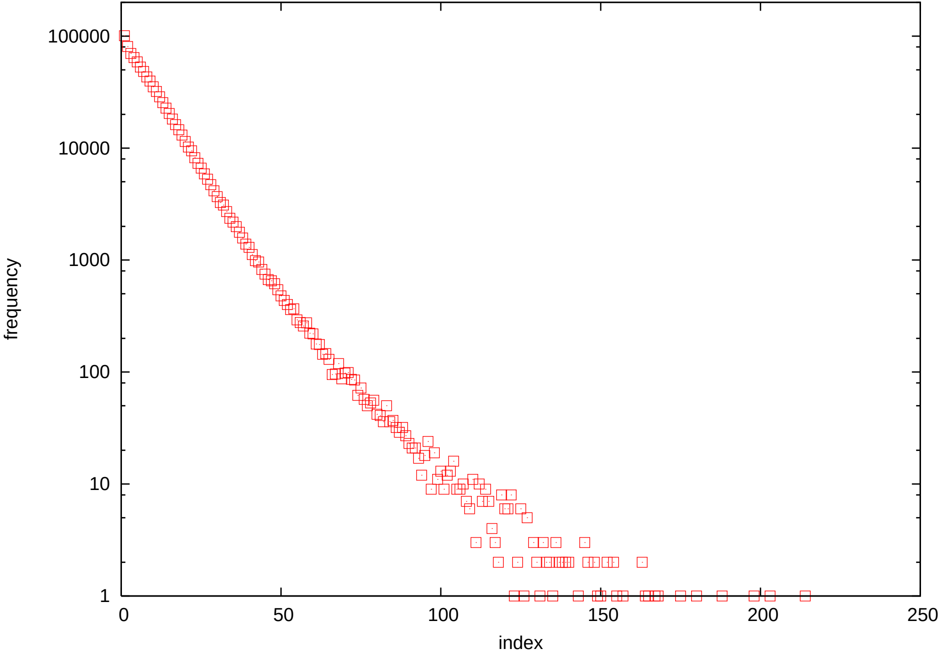


K=2
K=10
K=30
K=100
K=1000
P(k)
k



P(k)
1
0.1
0.01
0.001
0.0001
1e-05
1e-06
1e-07
1e-08
1
10
100
1000
k





100000
10000
1000
100
10
1
0
50
100
150
200
250
frequency
index