Games and Economic Behavior 79 (2013) 106–131

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

Sources of variation in social networks

Enghin Atalay

Department of Economics, University of Chicago, 1126 E. 59th Street, Chicago, IL 60637, United States

ARTICLE INFO

Article history:
Received 31 December 2011
Available online 9 February 2013

JEL classification:
D85

Keywords:
Degree distribution
Heterogeneous fitness
Network formation
Social networks

ABSTRACT

What explains the large variation in the number of contacts (degree) that different participants of social networks have: age, randomness, or some unobservable fitness measure? To answer this question, I extend the model presented in Jackson and Rogers (2007) to allow individuals to vary in their ability to attract contacts. I estimate the parameters of the extended model, using a social network of citations among high-energy physics papers, and find that the extended Jackson–Rogers model can parsimoniously fit the degree distribution of each age cohort. Moreover, both the length of time spent in the network and the unobservable fitness measure are important in explaining the observed variation in participants' degrees.

1. Introduction

Why do some participants of social networks have so many contacts, while most others have so few? How important are age and randomness in explaining the variation in the number of contacts (i.e., the *degree*) that participants have? What is the underlying process that produces the degree distributions that are repeatedly observed in studies of social networks?

I answer these questions by extending the framework introduced in Jackson and Rogers (2007). In that paper, the authors construct a simple model that fits degree distributions observed in different social networks. In their model, nodes enter the network one at a time and form links with existing nodes in a two-step process. First, in *random* meetings, the entrant uniformly samples from the population of incumbents. Second, in *network-based* meetings, the entrant samples from the contacts of the incumbents that it met. In each of these meetings, the entrant forms a directed link to the sampled incumbent. When network-based meetings are more prevalent, incumbents with many contacts are relatively more likely to gain additional contacts. At one extreme, when all links are formed via networking, nodes' degrees are Pareto distributed. At the other extreme, when all links are formed randomly, the degree distribution is that of an exponential random variable.[1]

In Sections 3.1 and 3.2, I review the arguments that are used to solve for the exact joint distribution of participants' ages and degrees. According to the baseline model, participants are only distinguished by their age (when they entered the network) and chance (whether they happened to be contacted by entrants more or less often than expected). As I show in Section 3.2, the role of chance in the model is limited: almost all of the variation in the number of contacts is due to differences in participants' dates of entry into the social network. I illustrate in Section 2 and Appendix A that this fact conflicts with what is actually observed in social networks. There is wide variation in the number of contacts of individuals within any age group. Therefore, there must exist some other factor, independent of age or chance, that contributes to the observed variation in nodes' degrees.

E-mail address: atalay@uchicago.edu.

[1] The model introduced in Jackson and Rogers (2007) closely resembles those presented in earlier articles, including Dorogovtsev et al. (2000), Pennock et al. (2002), and Vázquez (2003).

0899-8256/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.geb.2013.01.009

I extend the baseline model by allowing nodes to differ in the rate at which they can expect to gain additional links. In Section 3.3, I define the *fitness* of a node as the probability that each of its meetings will generate a link. I cannot observe nodes' fitness measures. However, using the variation of degrees across nodes of a particular age, I can identify the variation of fitness. With more variability in fitness, there is more variability in the degree distribution of nodes of a particular age.

In Section 4, I use a dataset of citations among high-energy physics papers to show that heterogeneous node fitness is necessary to fit the observed within-cohort degree distributions. I estimate—via maximum likelihood—the network formation model, both with and without heterogeneous fitness. When nodes are assumed to have identical fitness, the model struggles to fit the relatively weak correlation between age and degree. By contrast, when nodes' fitness levels are drawn from a parametric distribution (that is the same for all age cohorts), I am able to parsimoniously fit the weak correlation between age and degree.[2] Maximum likelihood estimates indicate that a large fraction of the variation in degree is due to variation in fitness. For example, for a median-aged paper, the expected number of citations increases from 1.4 to 20.7 as fitness increases from the 25th to the 75th percentile. For a median-fitness paper, the expected number of citations increases from 2.7 to 13.6 as age increases from the 25th to the 75th percentile.

In Section 4.3, I show that, for most age–fitness combinations, a marginal increase in age has a smaller effect—compared to a marginal increase in fitness—on a node's expected number of contacts. For all but the youngest and lowest-fitness nodes, differences in the rate at which contacts are formed are due mainly to heterogeneity in fitness, rather than heterogeneity in the length of time spent in the network. Randomness in the link formation process plays a tertiary role in explaining the variation in nodes' degrees.

Apart from the shape of the degree distribution, the model introduced in Jackson and Rogers (2007) generates predictions on other features observed in social networks. First, their model predicts that the degrees of two linked nodes are positively correlated. Second, the probability that two nodes are linked is larger conditional on the two nodes having a common contact. In Section 5, I argue that introducing heterogeneity in nodes' fitness levels does not qualitatively alter either of these predictions of the Jackson and Rogers (2007) model. However, the theoretical prediction of the probability that two nodes form a link, conditional on the presence of a common contact, can be matched in the data only when the average fitness is substantially larger than what is estimated in Section 4.

Parsimony and tractability are two of the main strengths of the Jackson–Rogers model. With only two parameters, the model captures several features ubiquitous in social networks. The model that I present in this paper retains much of the tractability of the original Jackson–Rogers model.[3] In addition, it is able to capture not only the within-cohort degree distributions, but also the other characteristics of social networks that are studied in Jackson and Rogers (2007).

In many environments, the manner in which agents are linked to one another has important economic consequences. Jackson (2011, p. 512) catalogs a list of examples: Social networks play an important role by "…transmitting information about jobs, new products, technologies, and political opinions. They also serve as channels for informal insurance and risk sharing, and network structure influences patterns of decisions regarding education, career, hobbies, criminal activity, and even participation in micro-finance." The role that social networks play in economic activity makes it important to understand the mechanisms through which social networks form and the reasons why some agents have so many contacts while most others have so few.

1.1. Literature review

Before proceeding to Section 2, I discuss the theoretical literature from which the current paper borrows, and the empirical literature to which the current paper might lend some insights.

The two main theoretical ideas—first, that one can solve for the exact degree distribution using a mass-balance equation, and, second, that one can embed heterogeneous node fitness into a network formation model—are taken from earlier papers. The method of solving for the exact degree distribution is taken directly from Dorogovtsev et al. (2000). The recognition that one can embed heterogeneous node fitness into a network formation model is due to Bianconi and Barabási (2001) (in the context of a pure preferential attachment model) and Caldarelli et al. (2002) (in the context of a variant of the Erdös and Rényi, 1960 network formation model).

The primary contribution of the current paper is to apply the above-mentioned theoretical arguments to show that, for a social network of citations among high-energy physics papers, both age and fitness are important for explaining the observed variation of nodes' degrees. Related to this finding, allowing for heterogeneous node fitness results in a less prominent estimated role of network-based meetings versus random meetings. These observations have implications for the interpretation of Jackson and Rogers (2007) and its successors.

In an application of Jackson and Rogers (2007), Bramoullé et al. (2012) introduce a social network with different *types* of individuals and assume random meetings are more likely to occur between individuals of the same type. With this assumption, Bramoullé et al. generate clear, testable predictions on the relationship between an individual's degree and the

[2] These results are not specific to the network of citations among physics articles. In Appendix A, I show that the same patterns hold for two additional social networks, a network of citations among patents and a network of buyer–supplier relationships among publicly traded firms.

[3] Of course, allowing fitness to vary arbitrarily across nodes would allow one to explain 100% of the variation in nodes' degrees. The point of this paper is that, with only two extra parameters (which govern the first two moments of the distribution of nodes' fitness levels), the extended Jackson–Rogers model can better fit the degree distribution of each age cohort.

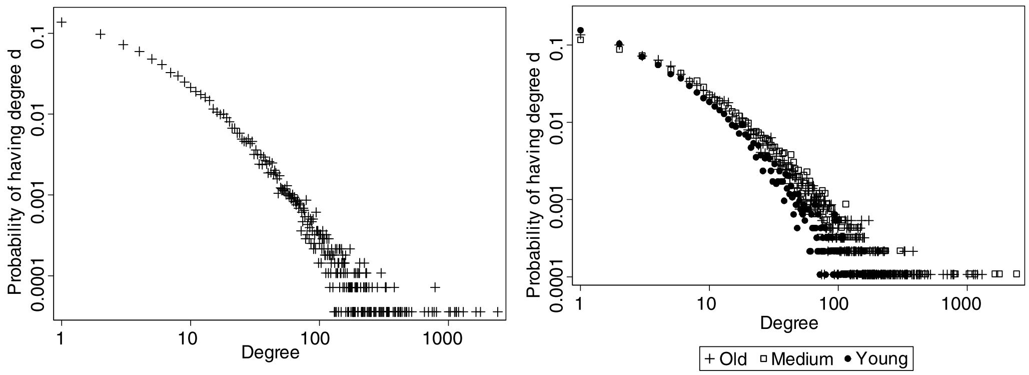

Fig. 1. Empirical degree distributions. In the left panel, I plot the distribution of the number of citations for all 27,770 papers in the dataset. In the right panel, I plot the degree distributions separately for old, medium-aged, and young papers. Papers uploaded onto the arXiv website on or before November 1996 are classified as old, while papers posted on or after May 2000 are classified as young.

fraction of its contacts that are of the same type. In a second application of Jackson and Rogers (2007), Chaney (2011) studies a network of inter-firm relationships and exogenously assigns each firm to a physical location. By assumption, random meetings are less likely to occur when two firms are far apart. The key empirical prediction generated by Chaney's model relates the number of counterparties a firm has to the average distance of the firm's contacts.

Allowing for heterogeneous fitness, as I do in the current paper, will moderate some of the conclusions of these extensions of Jackson and Rogers (2007). Take, as an example, Chaney (2011). In Chaney (2011), network-based meetings serve to increase the geographic dispersion of a firm's contacts.[4] As I argue in Section 4.2, accounting for heterogeneous fitness tends to decrease the estimate of the probability that any particular incumbent node is met in a given network-based meeting. This result, in turn, weakens the predicted relationship between the number of contacts a firm has and the average distance of the firm's contacts.

2. Data

In this paper, I use a dataset of citations among papers in the high-energy physics section of arXiv.[5] Between January 1992 and April 2003, 27,770 high-energy physics papers were uploaded onto arXiv. There were 352,807 citations among these papers, as of April 2003. In the terminology of social networks, each node represents an uploaded paper. A link from node i to node j represents a citation from paper i to paper j. The birth of node i corresponds to the date at which the paper was first uploaded onto the arXiv website.[6] Finally, the degree of a node is the citation count of the corresponding article.

In the left panel of Fig. 1, I plot the degree distribution. Like most other social networks, the degree distribution is skewed: Over half of the papers have four or fewer citations, while the most-cited paper has 2414 citations.

In the right panel of Fig. 1, I plot the degree distribution separately for old, medium-aged, and young papers. I define young papers to be the ones that were posted on or after May 2000; old papers were originally posted onto arXiv on or before November 1996. The remaining papers are classified as medium aged. While older papers tend to have more citations than younger papers, the relationship between age and citation count is weak. The correlation between age and degree is 0.08.[7] The weak relationship between age and citation count indicates that there is substantial within-cohort heterogeneity in the rate at which papers are prone to attract citations.

I have chosen to study the citation network not because it is the most interesting or important social network. Rather, the high-energy physics citation network is representative of a broad class of real-world social networks. Many of the empirical results of the paper are robust to the particular social network that is being studied. I provide support for this assertion in two ways. First, in Appendix A on page 122, I reproduce a few of the main empirical results, using data from two other social networks.

[4] As Chaney (2011, pp. 2–3) writes: "The theory therefore predicts that as firms acquire more foreign contacts, they expand into more remote countries, so that their exports become geographically more distant. The speed at which the geographic distance of exports increases depends on the relative importance of direct and remote search."

[5] arXiv is an Internet repository of working papers in physics, mathematics, and other mathematical sciences.
See Gehrke et al. (2003) for a description of the construction of the dataset.
The dataset can be downloaded at the Stanford Network Analysis Platform at http://snap.stanford.edu/data/index.html.

[6] Approximately 2% of the papers contain more than one date of submission. For the analysis, below, I choose the earliest date as the node's date of entry into the social network.

[7] The relationship between age and citation count may be nonlinear. Below, on page 112, I describe a correlation coefficient that accounts for this potential nonlinearity.

Second, I show that the high-energy physics social network is consistent with several stylized facts that characterize most social networks (see Jackson and Rogers, 2007). I list these stylized facts, below:

1. The degree distribution is heavy tailed.[8]
2. It is possible to construct a relatively short path between any two connected nodes, i and j.[9]
3. The probability that i is linked to k is, all else equal, higher when i is linked to j and j is linked to k.
4. If i and j are linked, the degree of i is correlated with the degree of j.

The citation network that I study is consistent with these stylized facts. The left panel of Fig. 1 displays that the first stylized fact holds. The second stylized fact holds as well: Gehrke et al. (2003) have computed that there exists a path of fewer than 15 links between any two connected nodes. In Section 5, I will argue that the third and fourth stylized facts also hold. Thus, given the archetypal nature of the citation network, it is likely that many of the empirical conclusions of the current paper will be shared by studies of other real-world social networks.

3. Theory

The goal of this section is to develop an estimable network formation model in which heterogeneity in age, randomness, and fitness have the potential to generate variation in nodes' degrees.

In Sections 3.1 and 3.2, I review the calculations that produce the stationary degree distribution—both for the overall population and for each cohort—of the Jackson and Rogers (2007) model. With the exact degree distribution in hand, I provide expressions for the relationship between nodes' ages and degrees, and argue that this relationship is stronger than what is observed in the data. In Section 3.3, I extend the network formation model, so that each node is endowed with an unobserved fitness measure that is drawn from a known distribution. Each node's fitness corresponds to the probability that any of its meetings with an entrant will result in a link. This extension breaks the strong relationship between age and degree.

3.1. A review of Jackson and Rogers (2007)

I begin this subsection with some notation. At each point in time, t, the network is defined by a $t \times t$ matrix, $\mathbf{D}^t$. Each node, i, is indexed by the period in which it entered the network. A directed link exists from node i to node j when $\mathbf{D}^t_{ij} = 1$. Otherwise, when no link exists from node i to node j, $\mathbf{D}^t_{ij} = 0$. For each link from node i to node j, i is called the predecessor node and j is called the successor node. The object of interest, for any node j, is the number of links for which j is the successor. Call this object d_j, the degree of j[10]:

$$d_j \equiv \sum_i \mathbf{D}^t_{ij}.$$

The Jackson–Rogers network formation model is defined as follows: In each period a single node enters the network. Nodes never leave the network and existing links are never broken. Time begins in period $t_0 > 0$. In the initial period there are t_0 nodes in the network. Thus, in period t, the number of nodes in the network equals t. Entrants have a degree of 0. Upon entry, the node meets $m \equiv m_r + m_n$ successors. The identity of each of the successors is the only source of randomness in Jackson–Rogers model.[11] In each of the m_r random meetings, the entering node forms directed links by uniformly sampling from the population of existing nodes.[12] Then, in the same period, it forms m_n directed links by uniformly sampling from the successors of the nodes that it just met. The probability that an existing node gains any one of the m_r links equals $\frac{1}{t}$. Similarly, the probability that an existing node of degree d receives any of the m_n links in a network-based meeting equals $\frac{d}{m_r+m_n}\frac{1}{t}$. On average, the number of additional predecessors gained by a degree-d node equals $\frac{d}{m_r+m_n}\frac{m_n}{t} + \frac{m_r}{t}$.

[8] A random variable, X, follows a heavy-tailed distribution if $\Pr\{X > x\} \propto x^{-\kappa}$ as $x \to \infty$.

[9] A *path* between nodes i and m is a sequence of nodes $\{i, j, k, \ldots, l, m\}$ such that i is linked to j, j is linked to k, etc., and l is linked to m. The distance between i and m, φ_{im}, is the number of links in the shortest path between i and m. For future reference, define the *diameter* of a network as $\max_{i,m:\ \varphi_{im}<\infty} \varphi_{im}$.

[10] In Jackson and Rogers (2007), d_j is called the in-degree. The "in" prefix distinguishes links for which j is the successor from links for which j is the predecessor. The out-degree of a node j is defined as $\tilde{d}_j \equiv \sum_i \mathbf{D}^t_{ji}$. Neither Jackson and Rogers (2007) nor the current paper analyzes the out-degree distribution. Since I will only refer to d_j, and not $\tilde{d}_j$, I drop the "in" prefix.

[11] Jackson and Rogers (2007) include a term, p, which represents the fraction of meetings that generate a link. The authors assume that this fraction is identical across all meetings. Jackson and Rogers note that p is not identified from the degree distribution. In other words, the Jackson–Rogers model predicts the same degree distribution for all m, p combinations for which $m \cdot p$ is constant. Thus, when describing the Jackson–Rogers model, I assume each meeting generates a link with probability 1.

[12] The entrants sample with replacement. It is true that there is a nonzero probability that an entrant will sample an existing node more than once. However, the probability that an entrant will sample any given node more than once is little-o of the probability that the entrant samples the node exactly once.

To solve for the stationary degree distribution, Jackson and Rogers employ a mean-field approximation.[13] Under this approximation, nodes deterministically gain $\frac{d}{m_r+m_n}\frac{m_n}{t}+\frac{m_r}{t}$ predecessors each period. Even though time and degree are discrete in the formulation of the model, the evolution of nodes' degrees are calculated as continuous functions of time and degree.

Given this set-up, Jackson and Rogers solve for (a) the relationship between a node's age and its degree, and (b) the stationary probability distribution function of nodes' degrees. These relationships are given in Eqs. (1) and (2), below. In these equations, $a \in [0,1]$ is the age quantile of the node and $r \equiv \frac{m_r}{m_n}$ is the ratio of random meetings to network-based meetings.

$$d \equiv \psi(a) = rm\left((1-a)^{-\frac{1}{1+r}} - 1\right), \tag{1}$$

$$f(d) = \frac{1+r}{d+mr}\left(\frac{mr}{d+mr}\right)^{1+r}. \tag{2}$$

As a result of the mean-field approximation, there is a one-to-one mapping between age and degree: a node, i, that has entered after i' will never have a degree, d_i, greater than $d_{i'}$. This result contrasts with the small difference between the degree distributions for young, medium-aged, and old nodes observed in the right panel of Fig. 1.

Perhaps though, the difference between theory and data is due solely to the mean-field approximation. I examine this hypothesis in the following subsection, by solving for the degree distribution without invoking a mean-field approximation.

3.2. The exact degree distribution

In this section, I restate the main results of Dorogovtsev et al. (2000). Without invoking a mean-field approximation, Dorogovtsev et al. are able to provide an expression for the degree distribution of the Jackson and Rogers (2007) model. With the exact solution to the degree distribution in hand, I will conclude this subsection by discussing the strength of the relationship between age and expected degree.

To solve for the exact stationary degree distribution, $h(d)$, it will be useful to describe the evolution of nodes' degrees from period to period. In Section 3.1, I argued that the expected number of predecessors that a node of degree d gains in a period is $\frac{d}{t(m_r+m_n)}m_n + \frac{1}{t}m_r = \frac{1}{t}[\frac{d}{1+r}+\frac{rm}{1+r}]$. As the size of the network becomes large, the probability of receiving a link approaches 0. Moreover, the probability of receiving more than 1 link approaches 0 at a much faster rate. In terms of little-o notation, the probability of receiving more than 1 link is little-o the probability of receiving exactly 1 link. Thus, as $t \to \infty$, a node with degree d will gain an additional predecessor with probability $\frac{1}{t}[\frac{d}{1+r}+\frac{rm}{1+r}]$ and will not gain any predecessors with probability $1-\frac{1}{t}[\frac{d}{1+r}+\frac{rm}{1+r}]$.

I begin by solving for the stationary degree distribution for nodes of a particular age quantile. Then, I will integrate across the age quantiles to solve for the entire population's degree distribution.[14]

Let $i \in [0,t]$ be the period in which the node was born. The mass-balance equation for period-i nodes is:

$$\tilde{h}(d;i,t+1) = \frac{1}{t}\frac{d-1+mr}{1+r}\tilde{h}(d-1;i,t)1_{d>0} + \left(1-\frac{1}{t}\frac{d+mr}{1+r}\right)\tilde{h}(d;i,t).$$

The left-hand side gives the number of nodes, in period $t+1$, that were born in period i and have degree d.[15] A node that is born in period i can have degree d at time $t+1$ either if (a) it had degree $d-1$ in period t and gained a predecessor, or (b) it had degree d in period t and did not gain a predecessor. Following Dorogovtsev et al., I replace $\frac{\tilde{h}(d;i,t+1)-\tilde{h}(d;i,t)}{1}$ with $\frac{\partial\tilde{h}(d;i,t)}{\partial t}$. Thus, I may write:

$$t\frac{\partial\tilde{h}(d;i,t)}{\partial t} = \frac{d-1+mr}{1+r}\tilde{h}(d-1;i,t)1_{d>0} - \frac{d+mr}{1+r}\tilde{h}(d;i,t). \tag{3}$$

In Appendix B on page 123, I show that one solution to Eq. (3) is:

$$\tilde{h}(d;i,t) = \left(\frac{mr+d-1}{d}\right)\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)\tilde{h}(d-1;i,t). \tag{4}$$

Let $a \equiv \frac{t-i}{t}$ be the age quantile of the node. With this change of variables, also define the function $h(d,a) \equiv \lim_{t\to\infty}\tilde{h}(d,i,t)$ to be the stationary degree distribution for age-quantile a nodes. Eq. (4) is equivalent to:

[13] Throughout this paper, the focus is on the stationary degree distribution. See Crespo and Cuenda (2010) for an analysis of the finite-population properties of the degree distribution in the Jackson and Rogers (2007) model.

[14] See page 4634 of Dorogovtsev et al. (2000) for an alternative method of solving for $h(d)$.

[15] Since one node is born per period, the number of nodes of a given age with degree d equals the fraction of nodes of a given age with degree d.

$$h(d;a)=\left(\frac{mr+d-1}{d}\right)\bigl(1-(1-a)^{1/(1+r)}\bigr)h(d-1;a). \tag{5}$$

For each age quantile, Eq. (5) provides a recurrence relationship between nodes with degree d and degree $d-1$. Iteratively applying Eq. (5) yields:

$$h(1;a)=mr\bigl(1-(1-a)^{1/(1+r)}\bigr)h(0;a),\quad \text{and}$$
$$h(2;a)=\frac{mr+1}{2}\bigl(1-(1-a)^{1/(1+r)}\bigr)h(1;a)=\frac{mr+1}{2}\frac{mr}{1}\bigl(1-(1-a)^{1/(1+r)}\bigr)^2h(0;a).$$

For a general d[16]:

$$h(d;a)=\frac{\Gamma[mr+d]}{d!\Gamma[mr]}\bigl(1-(1-a)^{1/(1+r)}\bigr)^d h(0;a).$$

The final step in solving for $h(d;a)$ is to compute the fraction of age-quantile a nodes that have a degree of 0. To compute $h(0;a)$, I use the fact that $h(d;a)$ is a probability distribution function. The sum of $h(d;a)$ must equal one.

$$\sum_{d=0}^{\infty}h(d;a)=\sum_{d=0}^{\infty}h(0;a)\bigl(1-(1-a)^{1/(1+r)}\bigr)^d\frac{\Gamma[mr+d]}{d!\Gamma[mr]}=h(0;a)(1-a)^{-mr/(1+r)}.$$

Since $\sum_{d=0}^{\infty}h(d;a)=1$, $h(0;a)=(1-a)^{mr/(1+r)}$. Thus:

$$h(d;a)=\frac{\Gamma[mr+d]}{d!\Gamma[mr]}(1-a)^{mr/(1+r)}\bigl(1-(1-a)^{1/(1+r)}\bigr)^d. \tag{6}$$

Eq. (6) is equivalent to Eq. (15) of Dorogovtsev et al. (2000).

I integrate the right-hand side of Eq. (6) across all age quantiles to arrive at the overall degree distribution:

$$\begin{aligned}h(d)&=\frac{\Gamma[mr+d]}{d!\Gamma[mr]}\int_0^1(1-a)^{mr/(1+r)}\bigl(1-(1-a)^{1/(1+r)}\bigr)^d\,da\\&=(1+r)\frac{\Gamma[(m+1)r+1]}{\Gamma[mr]}\frac{\Gamma[mr+d]}{\Gamma[(m+1)r+2+d]}.\end{aligned} \tag{7}$$

Eq. (7) is equivalent to Eq. (5) of Dorogovtsev et al. (2000).[17]

In Fig. 2, I plot $h(d;\frac{1}{6})$, $h(d;\frac{1}{2})$, $h(d;\frac{5}{6})$, and $h(d)$. The theoretical prediction of the stationary degree distribution matches the empirical degree distribution from the left panel of Fig. 1. At the same time, the model's predictions for the degree distributions of specific cohorts are not aligned with the data. In the citation network, the modal number of citations for young, medium-aged, and old papers is zero. The degree distribution for each age group has a heavy right tail. On the other hand, Fig. 2 indicates the modal degree increases monotonically with a. Furthermore, for any one age group, the degree distribution does not have a heavy right tail. In the citation network, there is more within-cohort degree variation—and less between-cohort degree variation—than the model would predict.[18]

To make this point more clearly, I invoke Eq. (6) to assess the strength of the relationship between a node's age quantile and its degree. In this exercise, one complicating factor is the nonlinearity in the relationship between age and degree, which has the potential to make the correlation between a and d significantly less than 1, even when these variables are strongly related to one another. Making this complicating factor more important, the extent of the nonlinearity depends on m and r. With these issues in mind, I apply ψ^{-1} to d. This transformation is aimed at removing the nonlinearity in the relationship between age and degree.

In Fig. 3, I plot the correlation between a and $\psi^{-1}(d)$. In the left panel, I vary r; in the right panel, I vary m. As the fraction of links formed in network-based meetings increases (r decreases), the relationship between age and degree weakens. The relationship between node age and degree is also weaker when the average degree in the network is smaller (m is smaller).

[16] $\Gamma[\cdot]$ is the gamma function: $\Gamma[x]=\int_0^\infty t^{x-1}e^{-t}\,dt$. When x is a positive integer, $\Gamma[x]=(x-1)!$.

[17] By comparing Eqs. (2) and (7), I can evaluate whether, as Jackson and Rogers argue, the mean-field approximation produces a degree distribution that is similar to the exact degree distribution. I compare Eqs. (2) and (7) in Appendix C, on page 124. I find that the mean-field approximation produces a degree distribution that closely matches the exact distribution, except when m is small or when d is close to 0 or ∞.

[18] To understand why under-estimation of within-cohort degree variation is associated with over-estimation of the strength of the relationship between age and degree, consider the following variance decomposition:

$$E\bigl(Var(d|a)\bigr)+Var\bigl(E(d|a)\bigr)=Var(d).$$

If, according to some model, $E(Var(d|a))$ is small, then, to correctly match the variance of d, the model must have a large estimate of $Var(E(d|a))$.

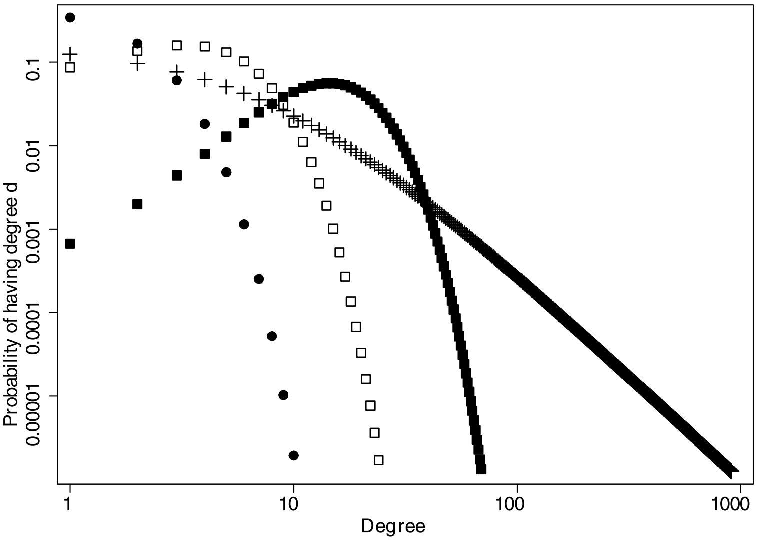

Fig. 2. Theoretical degree distributions: $h(d)$ ("+" signs), $h(d, \frac{5}{6})$ (solid squares), $h(d, \frac{1}{2})$ (hollow squares), and $h(d, \frac{1}{6})$ (solid circles). In this figure, $m = 15$, and $r = 0.5$.

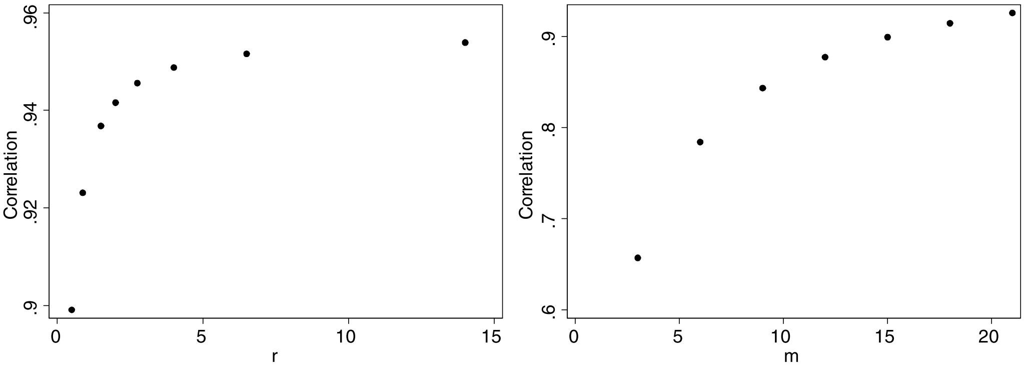

Fig. 3. Correlation between a and $\psi^{-1}(d)$, for different parameter values. In the left panel, $m = 15$; r is given on the x-axis. In the right panel, $r = 0.5$; m is given on the x-axis.

The correlations presented in Fig. 3 indicate that, in the Jackson–Rogers model, the relationship between age and degree is strong. By contrast, the correlation between age and degree is low in the citation network. For $(m, r) = (14.8, 0.5)$, the correlation between a and $\psi^{-1}(d)$ is 19%.[19]

3.3. Heterogeneity in node fitness

In this section, I introduce an additional source of heterogeneity to the network formation model. I assume that, upon entry into the network, each node is endowed with a fitness $p \in (0, 1]$ drawn from a distribution with probability distribution function, $b(p)$. The main restriction on $b(p)$ is that it is the same for all age cohorts. The model that I review in Section 3.1 corresponds to the case in which $p = 1$ for all nodes.

A node's fitness simply corresponds to the probability that it will receive a link when met by an entrant. In this paper, the term fitness does not have any deeper meaning. In the empirical application, for example, some articles will be called "high-fitness" while others will be called "low-fitness" articles. High-fitness articles are not necessarily superior to low-fitness articles. Instead, fitness simply reflects the propensity, for an article with a given number of citations, to gain additional citations.

To begin, consider the degree distribution for nodes of a particular fitness, p. Let $\bar{p}$ be the average probability that a meeting results in the creation of a link. Suppose that, upon entry, entering nodes meet m existing nodes, so that the average degree in the network equals $\bar{p}m$. As before, r denotes the ratio of random meetings to network-based meetings. The expected number of additional links that a node of degree d will gain in period t is:

[19] 14.8 and 0.5 are close to the maximum likelihood estimates of m and r. See Table 1, column 1.

$$p\left[\frac{d}{m\bar{p}}\cdot\frac{m_n}{t}+\frac{m_r}{t}\right]=\frac{1}{t}\left[\frac{d}{(1+r)\frac{\bar{p}}{p}}+\frac{mr\bar{p}}{(1+r)\frac{\bar{p}}{p}}\right].$$

So, a node with degree d and fitness p will gain an additional link in period t with probability $\frac{1}{t}[\frac{d}{(1+r)\frac{\bar{p}}{p}}+\frac{mr\bar{p}}{(1+r)\frac{\bar{p}}{p}}]$.

For nodes with a given fitness, p, all of the analysis from Sections 3.1 and 3.2 applies, with $(1+r)$ replaced by $(1+r)\frac{p}{\bar{p}}$ and mr replaced by $mr\bar{p}$. For example, the degree distribution, conditional on a and p, is:

$$f(d|a,p)=\frac{\Gamma[mr\bar{p}+d]}{d!\Gamma[mr\bar{p}]}(1-a)^{\frac{pmr}{1+r}}\left(1-(1-a)^{\frac{p}{\bar{p}(1+r)}}\right)^d. \tag{8}$$

Before proceeding to Section 4, I must specify how $\bar{p}$ is computed. To solve for $\bar{p}$, I integrate Eq. (8) over a and p and then use the formula for an expected value:

$$\sum_{d=0}^{\infty}\int_0^1 d(1+r)\frac{\bar{p}}{p}\frac{\Gamma[mr\bar{p}+(1+r)\frac{\bar{p}}{p}]}{\Gamma[mr\bar{p}]}\frac{\Gamma[mr\bar{p}+d]}{\Gamma[mr\bar{p}+(1+r)\frac{\bar{p}}{p}+d+1]}b(p)\,dp=m\bar{p}. \tag{9}$$

Eq. (9) implicitly defines $\bar{p}$ as a function of m, r, and the parameters describing the distribution of p. As $r\to\infty$, $\bar{p}$ approaches $\int_0^1 pb(p)\,dp$. In words, since links are assigned uniformly across all the nodes in the network, the probability that a meeting produces a link tends to the average fitness of the network. When $r<\infty$, $\bar{p}$ is greater than $\int_0^1 pb(p)\,dp$, since high degree nodes are both more likely to be sampled in network-based meetings and also have a higher fitness, p.

4. Estimation and results

This section contains the empirical content of the paper. I begin, in Section 4.1, by writing likelihood functions for three variants of the Jackson–Rogers model. The maximum likelihood estimates are then presented in Section 4.2. In Section 4.3, I discuss which of the three sources of variation—age, fitness, or luck—is most important for explaining degree heterogeneity. Finally, in Section 4.4, I compare each article's unobservable fitness measure with an observable proxy for fitness: the journal in which the article was published.

4.1. Likelihood functions

The parameters that I wish to estimate are m, the number meetings per entrant; r, the ratio of the number of random to network-based meetings; and the parameters of the fitness distribution, $b(p)$.

I estimate the model's parameters using maximum likelihood. Below, I describe the three different likelihood functions that I employ.

First, I estimate m and r without invoking any information on nodes' ages. From Eq. (7):

$$\Pr(d_i|m,r)=(1+r)\frac{\Gamma[(m+1)r+1]}{\Gamma[mr]}\frac{\Gamma[d_i+mr]}{\Gamma[d_i+2+(m+1)r]}.$$

The corresponding likelihood function is then:

$$\mathcal{L}_1(m,r)=n\log(1+r)+n\log\left(\frac{\Gamma[(m+1)r+1]}{\Gamma[mr]}\right)+\sum_{i=1}^{n}\log\left(\frac{\Gamma[d_i+mr]}{\Gamma[d_i+2+(m+1)r]}\right). \tag{10}$$

Next, I include information on the age quantile of each node in the likelihood function. From Eq. (6):

$$\Pr(d_i|a_i,m,r)=\frac{\Gamma[mr+d_i]}{\Gamma[mr](d_i)!}(1-a_i)^{mr/(1+r)}\left(1-(1-a_i)^{1/(1+r)}\right)^{d_i}.$$

The corresponding likelihood function is then:

$$\mathcal{L}_2(m,r)=\sum_{i=1}^{n}\log\left(\frac{\Gamma[mr+d_i]}{\Gamma[mr](d_i)!}(1-a_i)^{mr/(1+r)}\left(1-(1-a_i)^{1/(1+r)}\right)^{d_i}\right). \tag{11}$$

Finally, I estimate an unrestricted model, in which nodes are allowed to vary in the rate at which they are apt to gain additional links. I assume that p is drawn from the Beta(α,β) distribution. The probability distribution function for $p\in[0,1]$ is:

$$b(p)=p^{\alpha-1}(1-p)^{\beta-1}\frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]}\quad\text{for }\alpha,\beta>0. \tag{12}$$

As α increases relative to β, the expected value of p increases. As α and β both increase, the variance of p decreases. In the special case of $\alpha = \beta = 1$, the Beta distribution coincides with a uniform distribution. The Beta distribution is a convenient choice because it is sparingly parameterized, flexible, and has the unit interval as its support. However, I could have chosen a different parametric family for p without reducing the tractability of the estimation procedure.[20]

Given α, β, m, and r, the probability that a node with age quantile, a_i, has degree, d_i, is (see Eq. (8)):

$$\Pr(d_i|a_i, \alpha, \beta, m, r) = \frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]} \int_0^1 p^{\alpha-1}(1-p)^{\beta-1} \frac{\Gamma[\bar{p}mr + d_i]}{\Gamma[\bar{p}mr](d_i)!} (1-a_i)^{\frac{pmr}{1+r}} \left(1-(1-a_i)^{\frac{p}{\bar{p}(1+r)}}\right)^{d_i} dp. \tag{13}$$

The corresponding likelihood function is then:

$$\mathcal{L}_3(\alpha, \beta, m, r) = \sum_{i=1}^{n} \log\bigl(\Pr(d_i|a_i, \alpha, \beta, m, r)\bigr). \tag{14}$$

Using simulated data, I analyze the performance of the different maximum likelihood estimators. See Appendix D on page 125.

4.2. Results

Parameter estimates are collected in Table 1. In the first column, I give the maximum likelihood estimates for m and r that maximize Eq. (10). The estimate for the total number of meetings, 14.82, is somewhat higher than the average degree in the network, 12.70. The estimated ratio of random to network-based meetings indicates that roughly 34% ($= \frac{0.513}{1.513}$) of the links are formed in random meetings. The second column also gives maximum likelihood estimates of m and r. In this column, though, the likelihood function incorporates information on the dates at which nodes entered the network. Now the maximum likelihood estimate of m is extremely large, while the maximum likelihood estimate of r is close to 0. The model struggles to fit the degrees of the most-connected nodes. According to the model, the most connected nodes should be the oldest nodes. In the data, this is not the case: there are both young and old nodes at the right tail of the degree distribution. To fit the relatively young and well-connected nodes, the estimated fraction of meetings that are random is driven towards 0.[21,22]

The third column of Table 1 provides the maximum likelihood estimates for the full model. According to the model, an entrant meets over 2000 incumbents upon entry. Of these meetings, approximately 53% ($= \frac{1.121}{2.121}$) are random meetings. Given the parameter estimates, the average probability that a meeting produces a link equals $\bar{p} = 0.031$ (see Eq. (9)). According to the full model, an entrant forms $\bar{p}m = 67$ links upon entry, of which 20 were formed in random meetings and the remainder in network-based meetings.[23] In the full model, a link is formed with greater probability in a network-based meeting than in a random meeting.[24] Thus, even though the estimate of r increases, from 0.513 to 1.121, when allowing for heterogeneous node fitness, the fraction of links formed in network-based meetings increases from 66% to 71% ($= \frac{67.25-19.68}{67.25}$).

Allowing for heterogeneous node fitness reduces the importance of the “rich-get-richer” mechanism of the network formation model. The strength of the “rich-get-richer” mechanism is embodied in the probability that an incumbent will be met in a network-based meeting, conditional on the entrant meeting one of the incumbent's predecessors in a random meeting. The estimate of this term, $\frac{1}{\bar{p}rm}$, is smaller in the third column, compared to the first.[25] So, the most-connected nodes are prone to receive more contacts not only because they already have many contacts, but also because they tend to be of high fitness.

[20] In Appendix E, on page 126, I argue that the main results of the current section are robust to the assumption that fitness levels are drawn from the Beta distribution. To do so, I maximize Eq. (14), allowing for a much more flexible parameterization of $b(p)$.

[21] Concomitant to the small estimate of r, the estimate of m is large, of the same order of magnitude as $\frac{1}{r}$. Notice how the m term, whenever it appears in Eq. (11), is always multiplied by r.

[22] The near-singularity of the Fisher information matrix is another result indicating that the Jackson–Rogers model has difficulty incorporating nodes' ages into the estimation procedure.

The Fisher information matrix associated with $\mathcal{L}_2$ is $\left(\begin{smallmatrix} 1.7639\cdot 10^{-30} & 1.5962\cdot 10^{4} \\ 1.5962\cdot 10^{4} & 1.4445\cdot 10^{38} \end{smallmatrix}\right)$. The eigenvalues of this matrix are $1.4445 \cdot 10^{38}$ and $-1.5449 \cdot 10^{-35}$. Since the latter eigenvalue is so close to 0, the information matrix is near-singular. In this part of the parameter space, the likelihood function is nearly flat for m, r combinations for which $m \cdot r$ is constant.

[23] The number of links formed in random meetings equals $\frac{1.121}{2.121} \times \frac{0.491}{0.491+28.121} \times 2169.39 \approx 20$.

[24] On average, random meetings yield a link with probability $\frac{\alpha}{\alpha+\beta} \approx 1.7\%$, while network-based meetings produce a link with probability $\bar{p}(1+r) - r\frac{\alpha}{\alpha+\beta} \approx 4.7\%$.

[25] The entrant forms m_n network-based meetings. In each of these meetings, there are, on average, $m_r \cdot \bar{p} \cdot (m_r + m_n)$ potential successors of successors with whom the entrant will form a link. Thus, the probability that any one incumbent is met, conditional on the entrant meeting a predecessor of the incumbent, is $\frac{m_n}{m_r \cdot \bar{p} \cdot (m_r+m_n)} = \frac{1}{r\bar{p}m}$.

For the original Jackson–Rogers model, this term equals $\frac{1}{14.8 \cdot 0.5} = 13\%$. Allowing for heterogeneous fitness, the probability that an incumbent is met in a network-based meeting is $\frac{1}{67 \cdot 1.12} = 1.3\%$.

Table 1
Maximum likelihood estimates, with asymptotic standard errors in parentheses.

Parameter	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$
m	14.82	$1.86 \cdot 10^{17}$	2169.39
	(0.411)	(?)	(108.38)
r	0.513	$2.05 \cdot 10^{-17}$	1.121
	(0.025)	(?)	(0.076)
α			0.491
			(0.004)
β			28.121
			(0.947)
Log-likelihood	−91,894	−243,027	−96,324

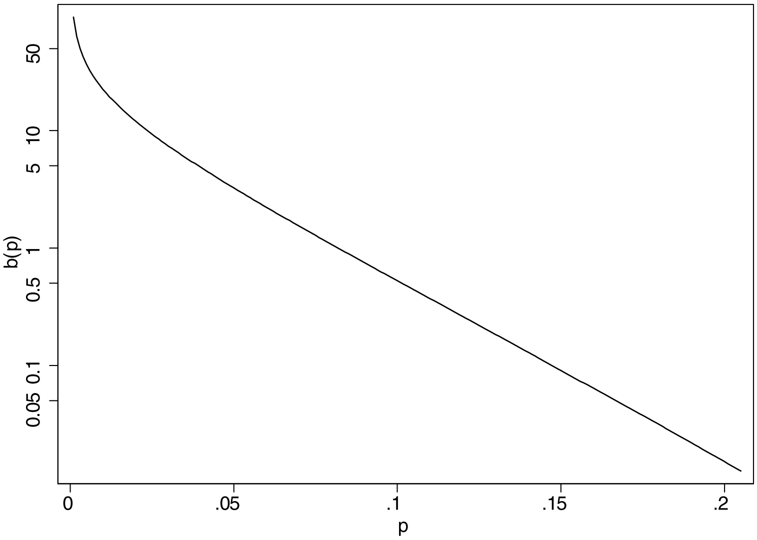

Fig. 4. Estimate of the fitness distribution, $b(p)$. To construct this plot, I use the estimates of α and β from the third column of Table 1.

Fig. 4 displays the estimated probability distribution function of node fitness. I construct this figure by plugging the maximum likelihood estimates of α and β into the probability distribution function for a Beta-distributed random variable (see Eq. (12)). According to estimates of α and β, fitness is skewed. The median and mean fitness measures are 0.0079 and 0.0172. Most likely, the skewness of the estimated fitness distribution is a result of the skewness of the within-cohort degree distributions.

The overall degree distributions predicted by each of the three models are plotted in Fig. 5. Despite the vastly different parameter estimates, the predicted degree distributions that are generated by the three likelihood functions are similar to one another. So, using only the overall degree distribution, one would have trouble distinguishing the original Jackson–Rogers model from a model with heterogeneous fitness nodes. Both models have similar predictions for the degree distribution.

However, the models have significantly different implications for the degree distributions of specific cohorts. To provide support for this claim, I form a likelihood-ratio test. Let the null hypothesis, for this test, be that articles have equal fitness: $H_0 : \alpha = 0$ or $\beta = 0$.[26] The alternative hypothesis states that both α and β are greater than 0. From the final row of Table 1, the value for the likelihood-ratio test statistic is 293,406 $(= 2 \cdot (243{,}027 - 96{,}324))$. The full sample of data on articles' ages and citation counts overwhelmingly rejects the homogeneous-fitness model. In addition, Fig. 6 plots the likelihood ratio test statistics, separately for each age cohort. (Here a cohort is a set of papers distributed in a given month–year.) Fig. 6 indicates that the null hypothesis is rejected by the data on the citation counts for any single cohort of articles.

The degree distributions for four different sets of papers are presented in Fig. 7. In the top-left panel, I plot the predicted degree distribution for $a = 0.2$. The solid line corresponds to the maximum likelihood estimates of the unrestricted model and the dashed line gives the predicted degree distribution from the homogeneous-fitness model. In the top-left panel, I also plot the empirical degree distribution for articles that have an age between the 15th and 25th percentiles. The other panels give the theoretical prediction of the degree distributions for $a = 0.4$, $a = 0.6$, and $a = 0.8$, as well as the empirical degree distributions for papers with ages in the 35–45th percentiles, 55–65th percentiles, and 75–85th percentiles.

[26] The variance of a Beta(α, β)-distributed random variable equals 0 when $\alpha = 0$ or $\beta = 0$.

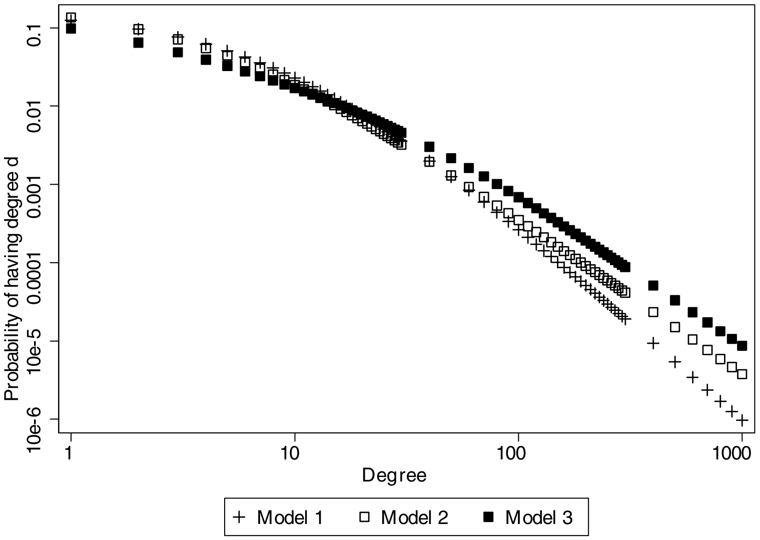

Fig. 5. Fitted degree distributions. To construct this plot, I plug the maximum likelihood estimates from Table 1 into Eq. (10), (11), or (14). To make the figure more readable, I only plot a subset of the points.

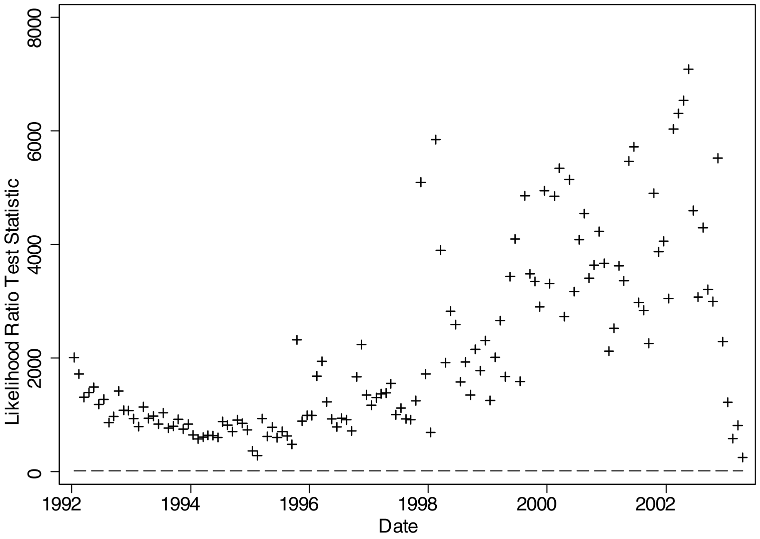

Fig. 6. Likelihood-ratio test statistics. Each point gives the likelihood-ratio test statistic using data on the citation count distribution of articles distributed in a given month. The dashed line gives the critical value, at a 0.1% significance level, of the likelihood ratio test.

Except for the bottom-right panel, the predicted degree distributions of the heterogeneous-fitness model fit the data reasonably well. The heterogeneous-fitness model, however, has trouble fitting the data for the older articles. Contrary to what the model would predict, the oldest articles do not, on average, have more citations than medium-aged papers. From youngest to oldest, the average number of citations for papers in each of the five age quintiles are 5.33, 11.37, 15.81, 18.14, and 12.77. Most likely, the sample of papers uploaded during arXiv's infancy is different, for idiosyncratic reasons, compared to the sets of papers uploaded once the website had become well-established.[27]

[27] Using a sample of papers published after April 1995 (the 20th percentile date), I have reproduced the maximum likelihood estimates given by Eqs. (10), (11), and (14). Removing the oldest papers produces a somewhat larger estimate of r and a somewhat smaller estimate of m; the estimates of α and β are unchanged. The parameter estimates, as well as any of the tables or figures given in Sections 4.2–4.3, are available upon request.

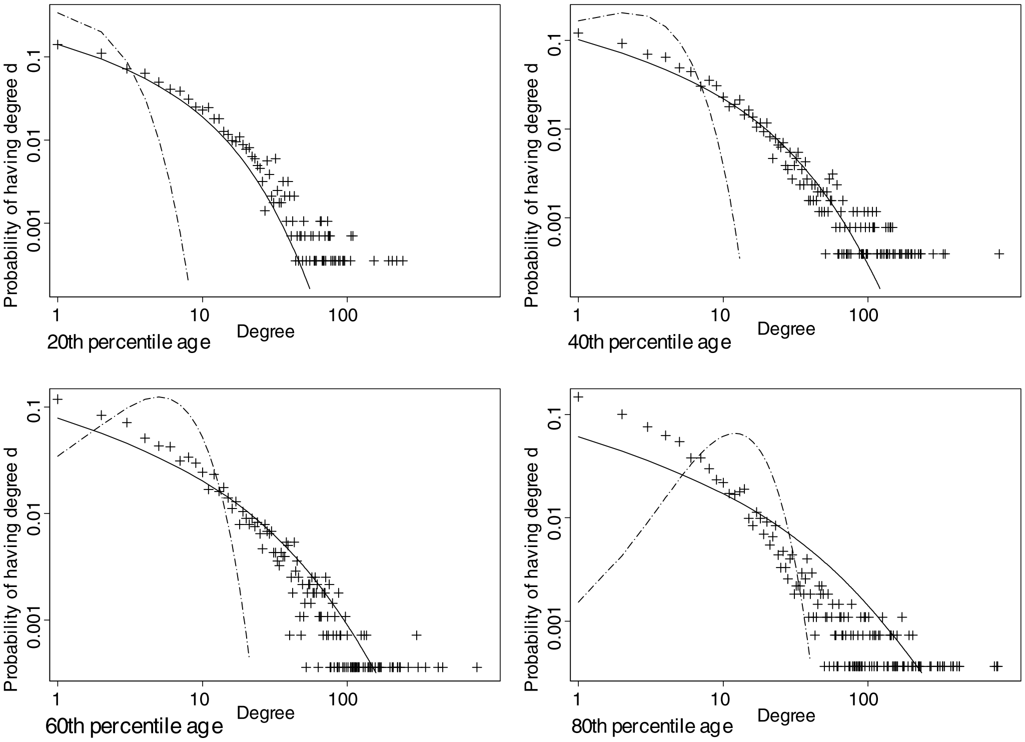

Fig. 7. Degree distributions for different cohorts. In the top-left panel, I plot the predicted degree distribution from the full model (solid line) and the original Jackson–Rogers model (dashed line) for $a = 0.2$. In the top-left panel, I also plot the empirical degree distribution ("+" signs) for articles that have an age between the 15th and 25th percentiles. The other three panels plot the theoretical degree distributions for $a = 0.4$, $a = 0.6$, and $a = 0.8$, and the empirical degree distributions for papers with ages in the 35–45th percentiles, 55–65th percentiles, and 75–85th percentiles.

4.3. Sources of degree heterogeneity

In this section, I analyze the extent to which variability in nodes' degrees is caused by variation in age, variation in fitness, or chance.

I begin by writing the conditional mean and variance for nodes of a given age–fitness combination. Using Eq. (8), one can check that

$$E[d|a,p] = m\bar{p}r\left((1-a)^{-\frac{p}{\bar{p}(1+r)}} - 1\right), \quad \text{and} \tag{15}$$

$$Var[d|a,p] = m\bar{p}r\left((1-a)^{-\frac{p}{\bar{p}(1+r)}} - 1\right)(1-a)^{-\frac{p}{\bar{p}(1+r)}}. \tag{16}$$

In Fig. 8, I use Eq. (15) to plot several isoquants of $E[d|a,p]$. Each plotted line gives a locus of a, p combinations for which the expected degree is constant. As expected, older nodes and nodes with higher fitness have a larger expected degree. For a median-age, median-fitness node the expected degree equals 6.51.

Since expected degree is a nonlinear function of age and fitness, there exists no simple statistic summarizing whether age or fitness plays a bigger role in explaining a node's expected degree. Thus, it will be necessary to consider the effect of increasing age or fitness for all a, p combinations. For $\theta \in [0,1]$, define $q(\theta)$ as the θ-quantile fitness. For example, since the median fitness paper had $p = 0.0079$, $q(0.5) = 0.0079$.

In Appendix F, I show how to compute[28]

$$\Psi(a,p) \equiv \frac{\frac{\partial E[d|a,p]}{\partial a}}{\frac{\partial E[d|a,p]}{\partial \theta}}. \tag{17}$$

The numerator on the right-hand side of Eq. (17) is the marginal effect on expected degree of marginally increasing the age quantile. The denominator gives the marginal effect of increasing the quantile of fitness. When $\Psi(a,p) > 1$, increasing

[28] Notice that, since the expected value of nodes' degrees exactly equals the expression for the degree of a node that would result from the mean-field approximation (compare Eqs. (1) and (15) with $p = \bar{p} = 1$), it would have been possible to compare the relative importance of fitness versus age without actually computing the exact degree distribution. Expressions for the exact degree distribution are, however, necessary to assess the relative importance of luck (see Fig. 10).

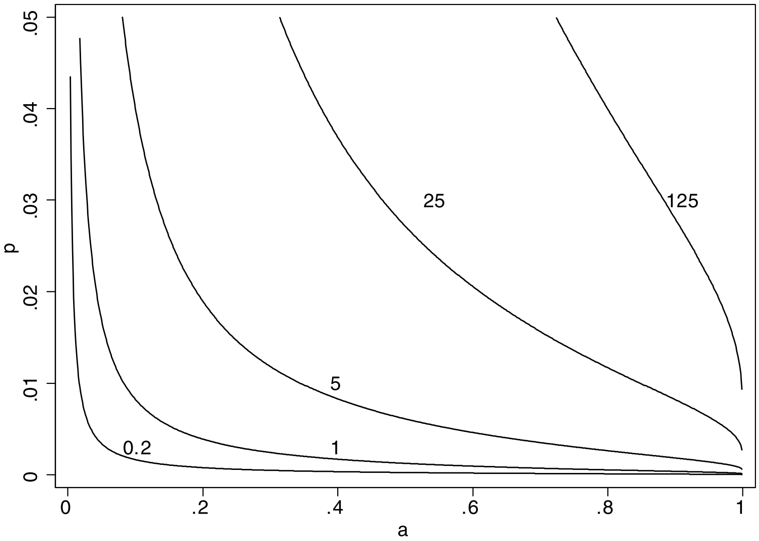

Fig. 8. Isoquants of $E[d|a, p]$, which are computed using Eq. (15).

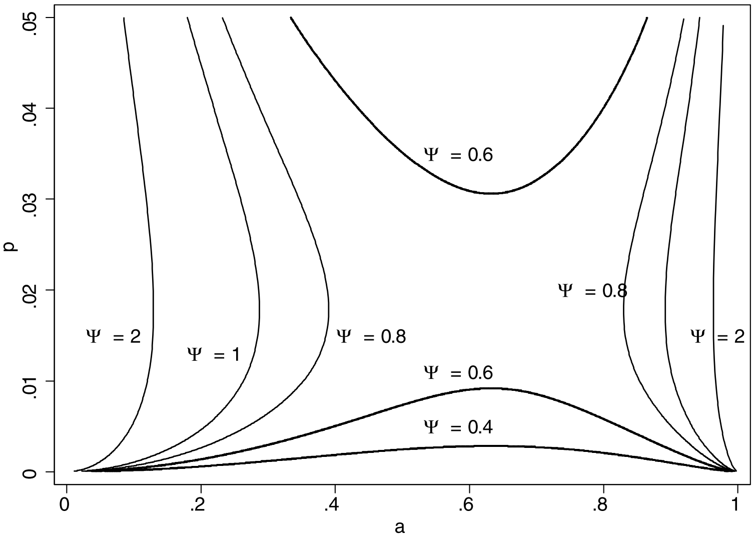

Fig. 9. Isoquants of $\Psi(a, p)$, which are defined using Eq. (17).

age marginally will have a greater effect on expected degree than marginally increasing fitness. The opposite is true when $\Psi(a, p) < 1$.

To give one example, consider Ψ evaluated at the median age and fitness: $\Psi(0.5, 0.0079)$. From before, the expected degree of this node equals 6.508. Increasing age by one percentile leads to an expected degree of $E[d|0.51, 0.0079] = 6.71$. On the other hand, increasing fitness by one percentile leads to an expected degree of $E[d|0.50, 0.0082] = 6.83$. Thus, $\Psi(0.5, 0.079) \approx \frac{6.71-6.51}{6.83-6.51} = 0.63$.

I plot the isoquants of $\Psi(a, p)$ in Fig. 9. Expected degree steeply increases with age when a is close to 0 or 1. For these values, $\Psi(a, p)$ is large. Between the 30th and 80th percentiles of age, $\Psi(a, p)$ is always less than 1. For medium-aged nodes, increasing fitness—instead of age—by one percentile will have a greater effect on expected degree.

So far, I have ignored the role that luck may play in generating variability in degree. In Fig. 9, I plot the isoquants of $\frac{SD[d|a,p]}{E[d|a,p]}$. For a, p combinations along a particular locus, the coefficient of variation of nodes' degrees is constant. As a or p increases, the coefficient of variation decreases. So, for old or high-fitness papers, randomness cannot explain much of the variability in citation counts. For example, for an article with age and fitness each at the 75th percentile, the standard deviation of degrees is only 19% of the expected degree. On the other hand, for young nodes or nodes with low fitness, randomness explains a large fraction of the variability that is observed in the data.

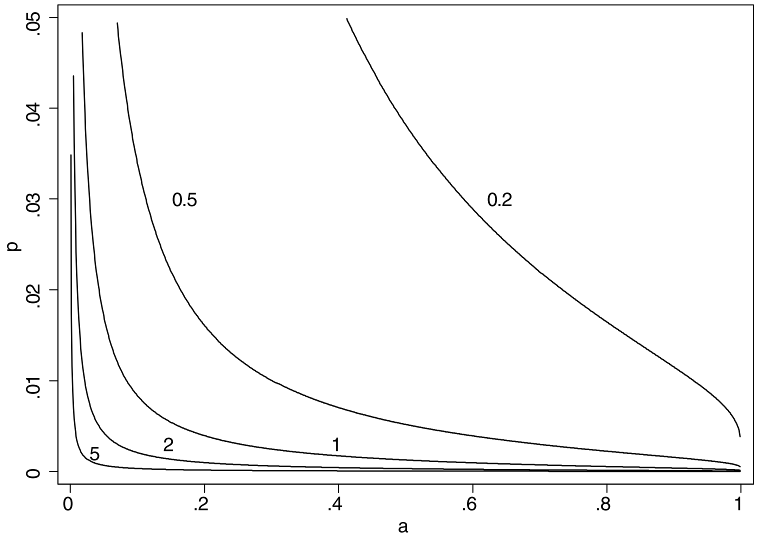

Fig. 10. Isoquants of $\frac{SD[d|a,p]}{E[d|a,p]}$, which are computed using Eqs. (15) and (16).

4.4. An observable proxy for fitness

Up to now, papers' fitness measures have been unobserved. In this subsection, I examine the relationship between an observable proxy for fitness and a model-based estimate of unobserved fitness. Papers that I estimate to have a high fitness measure are eventually published in more prestigious journals.

Of the 27,770 papers in the high-energy physics citation network, 20,228 were published in a journal. The three most popular journals—*Nuclear Physics B*, *Physics Letters B*, and *Physical Review D*—each published approximately 12% of the articles in the network. The average degree was highest for papers published in *Advances in Theoretical and Mathematical Physics* and *Physics Reports*. The average citation counts for papers from these journals were over five times the average for the entire citation network.

I begin the analysis of this subsection by constructing an estimate of each paper's unobserved fitness. A paper is likely to be of high fitness if it has a high degree relative to other papers in its cohort. The distribution of unobserved paper fitness, given its observed age and citation count, is given by the following formula[29]:

$$\eta(p|d,a) \propto (1-a)^{\frac{pmr}{1+r}}\left(1-(1-a)^{\frac{p}{\bar{p}(1+r)}}\right)^{d} p^{\alpha-1}(1-p)^{\beta-1}. \tag{18}$$

I then sum Eq. (18) over different sets of papers, yielding an estimated fitness distribution for different subsets of the high-energy physics literature. In Fig. 11, I plot the estimated distribution of fitness measures for four groups of papers: papers published in *Advances in Theoretical and Mathematical Physics*, papers published in *Physics Reports*, unpublished papers, and all other papers. As expected, the nodes' estimated fitness levels are lowest for the seven thousand unpublished papers. The median estimated fitness is 0.0069 for unpublished papers. By contrast, papers published in *Advances in Theoretical and Mathematical Physics* or *Physics Reports* have much higher estimated fitness measures. The median estimated fitness levels for these journals are 0.0256 and 0.0277, respectively.

The correlation between my estimates of nodes' unobservable fitness levels and observable node characteristics motivates the reduced-form regressions that have recently been employed in analyses of social networks. For instance, Atalay et al. (2011) study the buyer–supplier network of the United States. In this network, a node represents a firm and a directed link exists from one node to another provided one of the firms purchased intermediate inputs from the other firm. In this paper, Atalay et al. run a logit regression, in which the dependent variable equals 1 if firm i was a supplier of firm j. The authors find that firms with higher labor productivity and assets tend to have more suppliers. So, in the context of the buyer–supplier network, two proxies for node fitness are the labor productivity and asset value of the firms that the nodes are representing.

In a second example, Conley and Udry (2010) consider a social network among pineapple farmers in Ghana. A link exists between two individuals when one farmer asks the other about farming techniques. Conley and Udry employ a logit

[29] Via Bayes' theorem:

$$\eta(p|d,a) = \frac{f(d|a,p)b(p)}{f(d|a)}.$$

To arrive at the desired result, substitute $f(d|a,p)$, using Eq. (8); $b(p)$, using Eq. (12); and $f(d|a)$, using Eq. (13).

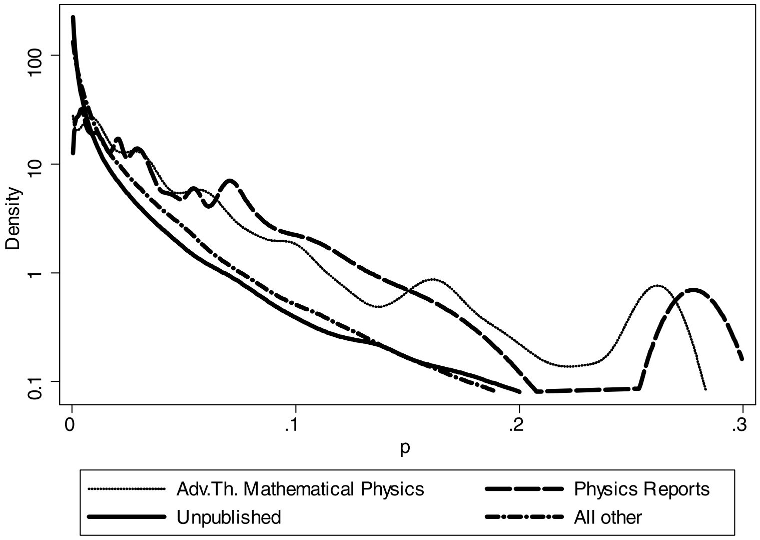

Fig. 11. Fitness distributions for four sets of papers. The fitness distribution, for each individual paper, is estimated using Eq. (18).

regression in which the dependent variable equals 1 when two farmers are linked. The authors find that the probability of a link is lower when either farmer holds a traditional office in the village. Thus, in the social network that Conley and Udry study, an official position can be used as a proxy for low node fitness.

5. Clustering and assortativity

Besides the degree distribution, the Jackson–Rogers model makes predictions on several other features observed in actual social networks. In this section, I describe two of these features, clustering and assortativity. I then argue that introducing heterogeneity in node fitness does not overturn either of the predictions of the Jackson–Rogers model.[30]

5.1. Clustering

Define the clustering coefficient of a network, at time t, as:

$$C_t = \frac{\sum_{i,j,k} \mathbf{D}_{ij}^t \mathbf{D}_{jk}^t \mathbf{D}_{ik}^t}{\sum_{i,j,k} \mathbf{D}_{ij}^t \mathbf{D}_{jk}^t}. \tag{19}$$

The clustering coefficient equals the fraction of pairs of $i \to j$ and $j \to k$ links for which a link exists from node i to node k. In a network where links are assigned randomly across pairs of nodes, the clustering coefficient approaches 0 as the number of nodes approaches infinity. In many social networks, the clustering coefficient is significantly greater than 0. In the high-energy physics citation network, for instance, the clustering coefficient is 9.0%.

In the Jackson–Rogers model, clustering is a direct result of the network-based meetings. For every instance in which the random meeting, $i \to j$, and the network-based meeting, $i \to k$, both produce links, a set of links $i \to j$, $j \to k$, and $i \to k$ is formed. This mechanism for generating triples of $i \to j, j \to k, i \to k$ links remains even when nodes are heterogeneous in their fitness.

I need to make additional assumptions to derive an approximate expression for the clustering coefficient, C_t. First, I re-introduce the mean-field approximation: the number of links that an existing node gains exactly equals the expected number of such links. Second, I assume that the number of links that an entrant forms, $m\bar{p}$, is an integer. Third, I assume that $r > \frac{\beta\bar{p}}{2\alpha+\beta-\beta\bar{p}}$.[31] Fourth, I assume that each entrant forms at most one link with the successors of any one particular successor. The first, second, and fourth assumptions are also made in Jackson and Rogers (2007). The assumption that r is greater than $\frac{\beta\bar{p}}{2\alpha+\beta-\beta\bar{p}}$ is made to simplify the proof, and holds for the set of parameters that are estimated in Section 4.2.[32]

[30] Jackson and Rogers (2007) discuss two other characteristics of social networks: (a) a small diameter and (b) a negative relationship between a node's degree and the probability that neighbors of the node are linked. The proofs that Jackson and Rogers provide, related to "a" and "b", are significantly more intricate than the proofs related to clustering and assortativity. Moreover, Jackson and Rogers are able to prove that the diameter of social networks is small only for $r \to 0$. Given these difficulties, I have not attempted to modify the proofs of Jackson and Rogers (2007), to allow heterogeneity in node fitness.

[31] This condition is equivalent to $m_r > \bar{p}_n m_n$, where $\bar{p}_n$ is the average fitness of the successor in a network-based meeting.

[32] For $(\alpha, \beta, \bar{p}) = (0.491, 28.121, 0.031)$, $\frac{\beta\bar{p}}{2\alpha+\beta-\beta\bar{p}} = \frac{28.121 \cdot 0.031}{0.892 \cdot + 28.121 - 28.121 \cdot 0.031} = 0.031$, which is considerably less than the estimate of r, 1.121.

Proposition 1. *Given the assumptions specified above, the clustering coefficient tends to*:

$$C = \frac{1}{m\bar{p}} \cdot \frac{\alpha}{\alpha+\beta}\left(1 - \frac{\alpha}{\alpha+\beta} \cdot \frac{1}{\bar{p}} \cdot \frac{r}{1+r}\right).$$

The proof is given in Appendix G. See page 128.

The addition of heterogeneity in p does not alter the Jackson–Rogers model's prediction of a positive clustering coefficient. In Jackson and Rogers (2007), $\bar{p}$ is a free parameter; the authors use this parameter to fit C exactly. I cannot do so here, as I have already estimated $\bar{p}$, α, and β using information on the relationship between age and expected degree. Using $(m, r, \alpha, \beta, \bar{p}) = (2169, 1.121, 0.491, 28.131, 0.031)$, the model's theoretical prediction for C is 0.02%. This value is much lower than the 9.0% value that is actually observed in the data. In this way, there is a trade-off: fitting the within-cohort degree distributions leads to a reduced ability to quantitatively match the substantial clustering that is observed in the high-energy physics social network.

In summation, the heterogeneous-fitness model maintains the theoretical prediction of a positive clustering coefficient. However, the theoretical prediction is an order of magnitude smaller than what is actually observed in the citation network.

5.2. Assortativity

Another feature observed in many social networks is the positive relationship between the degree of a node and the degree of its predecessors. A social network is *assortative* when such a positive relationship exists.[33] In the citation network, the correlation between the degree of a node and the degree of its predecessors is 4.1%. In the Jackson–Rogers model, assortativity is a direct result of the dynamic process under which nodes enter the network. In the model, the age of the predecessor node is restricted to be less than the age of the successor node. In a link from node i to node j, if j is a young node then node i must be young, as well. Indeed, in the citation network, the date of publication of the cited paper must be before that of the citing paper. The correlation between the publication dates of the cited and citing papers is 2.4%.

Using the mean-field approximation, Jackson and Rogers show that the degree of a randomly selected predecessor of node i first order stochastically dominates the degree of a randomly selected predecessor of node i', provided the degree of i is strictly greater than that of i'. I am able to show that the same result holds, even if nodes differ in their fitness.

Proposition 2. *Consider two nodes, i and i', with $d_i > d_{i'}$. Under the mean-field approximation, the degree distribution of the predecessors of i first order stochastically dominates the degree distribution of the predecessors of i'.*

The proof is given in Appendix G. See page 129.

In Jackson and Rogers (2007), the proof of this result is straightforward, since there is a one-to-one relationship between age and degree. If $d_i > d_{i'}$ it must be the case that i entered the network before i'. Thus the age distribution of a predecessor of i first order stochastically dominates the age distribution of a predecessor of i'. Again invoking the one-to-one relationship between age and degree, the degree distribution of a predecessor of i first order stochastically dominates the degree distribution of a predecessor of i'.

When nodes differ in fitness, things are not so simple. Even if $d_i > d_{i'}$, i might have entered the network after i'. This will be the case if the fitness of i is sufficiently larger than the fitness of i'. However, one can show, as I do in the appendix, that $d_i > d_{i'}$ implies that the age distribution of i first order stochastically dominates that of i'. Using this result, I then show that the age distribution of a randomly selected predecessor of node i first order stochastically dominates that of a randomly selected predecessor of node i'. Continuing with this logic gives the desired result.

6. Conclusion

Using a social network constructed from citations among high-energy physics papers, I estimate the extent to which articles vary in their *fitness* (ability to attract citations). Heterogeneity in fitness is identified using within-cohort variation in citation counts. If it had been the case that papers of a given age had similar citation counts, the model would have predicted a small variability in paper fitness. In contrast, the maximum likelihood estimates suggest that there is substantial heterogeneity in fitness. As a result, heterogeneity in fitness is a primary source of the variation in articles' citation counts.

The citation network that I study in this paper is, along several dimensions, a typical social network. First, the shape of the degree distribution for this network is similar to that observed in many other social networks. Second, the citation network exhibits the small-world property: it is possible to construct a relatively short path between any two papers, i and j. Third and fourth, the citation network is assortative and exhibits clustering. These four features of the citation network are also present in most other social networks (see Jackson and Rogers, 2007). Given the archetypal nature of the

[33] In contrast to social networks, biological and technological networks tend to be *disassortative*: the degrees of two linked nodes are negatively correlated. Newman (2003) computes, for different complex networks, the relationships between linked nodes' degrees.

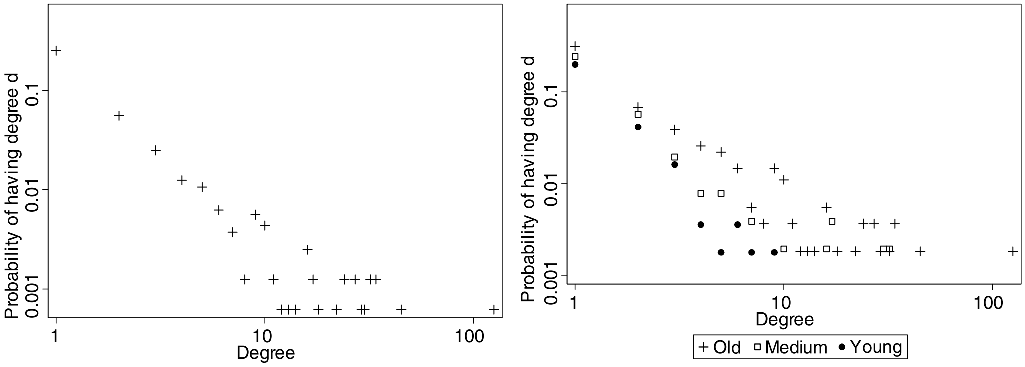

Fig. A.1. Empirical degree distributions, as of 2006. In the left panel, I plot the distribution of the number of customers for all 1607 firms in the Compustat dataset. In the right panel, I plot the degree distributions separately for old, medium-aged, and young firms. Firms which were first observed prior to 1986 are classified as old, while firms that were first observed after 1995 are classified as young.

citation network, it is likely that many of the empirical conclusions of the current paper will be shared by studies of other real-world social networks.

In this paper, fitness represents any characteristic of a node, independent of age and degree, that makes the node more or less likely to acquire additional links in the future. In future work, I hope to explain why some nodes have higher fitness than others. Fitness may be a choice variable of the participants of the social network. For example, in the context of a buyer–supplier network, firms may exert differing levels of effort in searching for suppliers of material inputs. The different levels of effort can be explained by a cost to searching combined with heterogeneous productivity or complexity of production. Whatever the source of the underlying heterogeneity, the different levels of search intensity will result in a distribution of fitness measures.

Incorporating agents' decisions over their fitness levels would result in a model that bridges the gap between two strands of the literature on network formation: (a) the purely mechanical models (e.g., Erdös and Rényi, 1960; Barabási and Albert, 1999; and Jackson and Rogers, 2007) and (b) the economic models (e.g., Jackson and Wolinsky, 1996 and Ostrovsky, 2008). The model would also retain the theoretical predictions on clustering, assortativity, and the degree distribution which are the hallmarks of real-world social networks.

Acknowledgments

I thank Aditya Bhave, Thomas Chaney, Kevin Czerniak, Thorsten Drautzburg, Ali Hortaçsu, Matthew Jackson, Sebastian Sotelo, and two anonymous referees for helpful comments.

Appendix A. Other examples of social networks

The purpose of this section is to argue that the main empirical results of Sections 2 and 4 are not special to the high-energy physics social network. In this section, I reproduce Fig. 1 and Table 1, using data from two other social networks. The two social networks consist of (a) buyer–supplier relationships among publicly traded U.S. firms and (b) citations among utility patents.

First, I consider a social network among large, publicly traded U.S. firms. In this social network, each node represents a firm. A directed link exists from node i to node j provided firm i sold goods or services to good j. I define the degree of a node, i, to be the number of suppliers that it has. See Atalay et al. (2011) for a description of the dataset.[34]

Using data from 2006, I plot the degree distributions for this social network in Fig. A.1. There were, in 2006, 1607 firms in the social network, of which 980 had no reported suppliers. The most connected firm, Wal-mart, had 125 reported suppliers. Other top firms were Cardinal Health (45 suppliers), General Motors (34 suppliers), and McKesson (34 suppliers).

In the right panel of Fig. A.1, I plot the degree distributions separately for old, medium-aged, and young firms. The correlation between a firm's age-quantile and its degree is 16%, which is higher than the correlation for the high-energy physics social network, but still lower than what would be predicted by Jackson and Rogers (2007).

Next, in Table A.1, I present the results of the maximum likelihood estimates of the three variants of the Jackson and Rogers (2007) model. Again, once information on nodes' ages is included in the likelihood function, the Jackson–Rogers generates nonsensical parameter estimates for m and r. The parameter estimates of the full model are more sensible: The average number of contacts per node is $118.2 \cdot 0.048 \approx 5.7$. Of these, $\frac{0.46}{1+0.46} \cdot \frac{0.85}{0.85+35.92} \cdot 118.2 \approx 0.9$ are formed as a result of random meetings. The remainder are formed following network-based meetings.

[34] One key difference between the buyer–supplier network and the physics citation network is that nodes and links are not permanent. Over time, firms exit and buyer–supplier relationships are severed. Atalay et al. (2011) amend the model of Jackson and Rogers (2007) and show that the predicted degree distributions are similar when node and link removal are possible.

Table A.1
Maximum likelihood estimates. For this table, I use data from the buyer–supplier social network.

Parameter	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$
m	1.01	$4.45 \cdot 10^{38}$	118.19
r	1.32	$1.30 \cdot 10^{-39}$	0.46
α			0.85
β			35.92
Log-likelihood	−2034.6	−2147.6	−2006.8

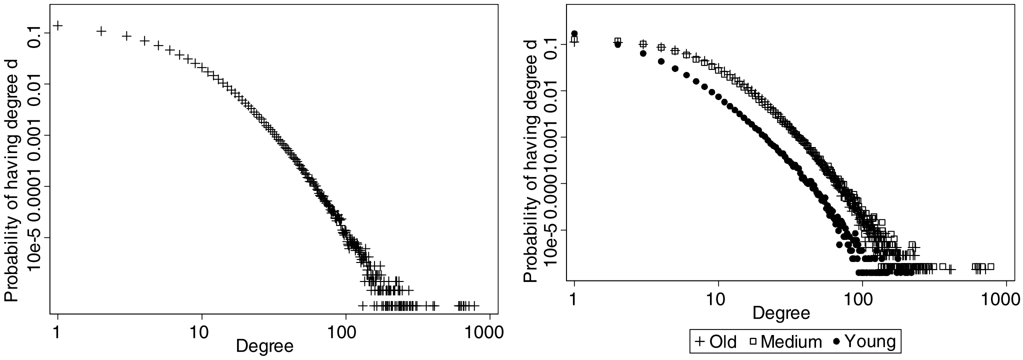

Fig. A.2. Empirical degree distributions, as of 1999. In the left panel, I plot the distribution of the number of citations for all 2,139,314 patents in the dataset. In the right panel, I plot the degree distributions separately for old, medium-aged, and young patents. Patents granted on or before 1985 are classified as old, while patents granted on or after 1993 are classified as young.

Table A.2
Maximum likelihood estimates. For this table, I use data from the NBER patent citation social network.

Parameter	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$
m	4.98	$3.15 \cdot 10^{26}$	364.42
r	2.44	$6.84 \cdot 10^{-27}$	1.37
α			0.86
β			31.14
Log-likelihood	$-5.66 \cdot 10^{6}$	$-7.20 \cdot 10^{6}$	$-5.45 \cdot 10^{6}$

As a second example, I consider a social network constructed from the NBER patent database. Here, a node represents a particular patent, while a directed link exists from node i to node j if patent i cites patent j. The data span all utility patents between 1975 and 1999. See Hall et al. (2001) for a thorough introduction to this dataset.

For the patent citation social network, I plot the degree distributions in Fig. A.2. Again, the degree distribution is skewed to the right. The median patent received 3 citations. The 90th percentile and 99th percentile patents received 12 and 35 citations, respectively. In the right panel, I plot the degree distributions separately for old, medium-aged, and young patents. Patents in the youngest tercile (those granted before 1986) received, on average, 2.1 citations. This is significantly lower than the average number of citations received by medium-aged (6.5 citations) or old (6.9 citations) patents. Compared to the buyer–supplier social network and the high-energy physics social network, the correlation between degree and age is higher in the patent citation social network. For this social network, the correlation between a node's degree and its age quantile is 29%.

Table A.2 contains the maximum likelihood estimates for the patent citation social network.

In both the buyer–supplier social network and the patent citation social network, I showed that the relationship between age and degree is too weak to be consistent with the model of Jackson and Rogers (2007). While, I have not ruled out the possibility that there exist some social networks in which the age–degree relationship conforms to the Jackson–Rogers model, the additional examples that I have given show that the main empirical results of the paper are not due to some special feature of the high-energy physics social network.

Appendix B. Verifying that Eq. (4) obeys Eq. (3)

The objective of this section is to show that

$$\tilde{h}(d;i,t) = \left(\frac{mr+d-1}{d}\right)\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)\tilde{h}(d-1;i,t) \tag{4}$$

obeys

$$t\frac{\partial\tilde{h}(d;i,t)}{\partial t}=\frac{d-1+mr}{1+r}\tilde{h}(d-1;i,t)1_{d>0}-\frac{d+mr}{1+r}\tilde{h}(d;i,t). \tag{3}$$

First, take the derivative of Eq. (4) with respect to t:

$$\frac{\partial\tilde{h}(d;i,t)}{\partial t}=\left(\frac{mr+d-1}{d}\right)\left[\frac{\partial\tilde{h}(d-1;i,t)}{\partial t}\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)+\left[\tilde{h}(d-1;i,t)\right]\left(\frac{i}{t}\right)^{1/(1+r)}\frac{1}{(1+r)t}\right]. \tag{B.1}$$

Plug (B.1) into Eq. (3):

$$\begin{aligned}\frac{(d-1+mr)\tilde{h}(d-1,i,t)}{(1+r)t}&=\left(\frac{mr+d-1}{d}\right)\frac{(d-2+mr)\tilde{h}(d-2,i,t)}{(1+r)t}\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)\\&\quad-\frac{(d-1+mr)\tilde{h}(d-1,i,t)}{(1+r)t}\left(\frac{mr+d-1}{d}\right)\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)\\&\quad+\left(\frac{mr+d-1}{d}\right)\left[\tilde{h}(d-1;i,t)\right]\left(\frac{i}{t}\right)^{1/(1+r)}\frac{1}{(1+r)t}+\frac{(d+mr)\tilde{h}(d,i,t)}{(1+r)t}.\end{aligned}$$

Use Eq. (4) to replace $\tilde{h}(d;i,t)$ with $\tilde{h}(d-1;i,t)$. Then cancel out the $\tilde{h}(d-1;i,t)$'s.

$$\begin{aligned}&\frac{d}{(1+r)t}-\frac{(d+mr)(1-(\frac{i}{t})^{1/(1+r)})}{(1+r)t}\\&\quad=\left[\frac{(d-2+mr)\frac{d-1}{mr+d-2}}{(1+r)t(1-(\frac{i}{t})^{1/(1+r)})}-\frac{(d-1+mr)}{(1+r)t}\right]\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)+\left(\frac{i}{t}\right)^{1/(1+r)}\frac{1}{(1+r)t}.\end{aligned}$$

Simplify:

$$\begin{aligned}&d-(d+mr)\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)=\left[\frac{d-1}{(1-(\frac{i}{t})^{1/(1+r)})}-(d-1+mr)\right]\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)+\left(\frac{i}{t}\right)^{1/(1+r)},\\&-(d+mr)\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)=-1-(d-1+mr)\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)+\left(\frac{i}{t}\right)^{1/(1+r)},\\&1=\left(1-\left(\frac{i}{t}\right)^{1/(1+r)}\right)+\left(\frac{i}{t}\right)^{1/(1+r)},\\&1=1.\end{aligned}$$

Since the two sides of the previous equation are equivalent, I conclude that Eq. (4) obeys Eq. (3).

Appendix C. Effect of the mean-field approximation on the degree distribution

In this section, I compare the exact degree distribution to the one computed in Jackson and Rogers (2007). Remember that Jackson and Rogers invoke a mean-field approximation to solve for the stationary degree distribution. Under the mean-field approximation, the number of contacts that a node gains each period is a deterministic function of its degree.

The stationary degree distribution that Jackson and Rogers compute is:

$$f(d)=\frac{1+r}{d+mr}\left(\frac{mr}{d+mr}\right)^{1+r}.$$

Since the rate at which nodes gain contacts is approximated by a continuous function of degree and time, the stationary degree distribution is a continuous function of d.

I argue in Section 3.3 that the exact degree distribution resulting from the Jackson–Rogers model is:

$$h(d)=(1+r)\frac{\Gamma[(m+1)r+1]}{\Gamma[mr]}\frac{\Gamma[mr+d]}{\Gamma[(m+1)r+2+d]}.$$

This function is defined over the set of nonnegative integers, $d\in 0,1,2,\ldots.$

I define the function $\epsilon(d;m,r)$ as the log ratio of the two degree distributions:

$$\epsilon(d;m,r)\equiv\log\left[\frac{f(d)}{h(d)}\right]=\log\left[\frac{(mr)^{1+r}\cdot\Gamma[mr]\cdot\Gamma[(m+1)r+2+d]}{(d+mr)^{2+r}\cdot\Gamma[mr+d]\cdot\Gamma[(m+1)r+1]}\right]. \tag{C.1}$$

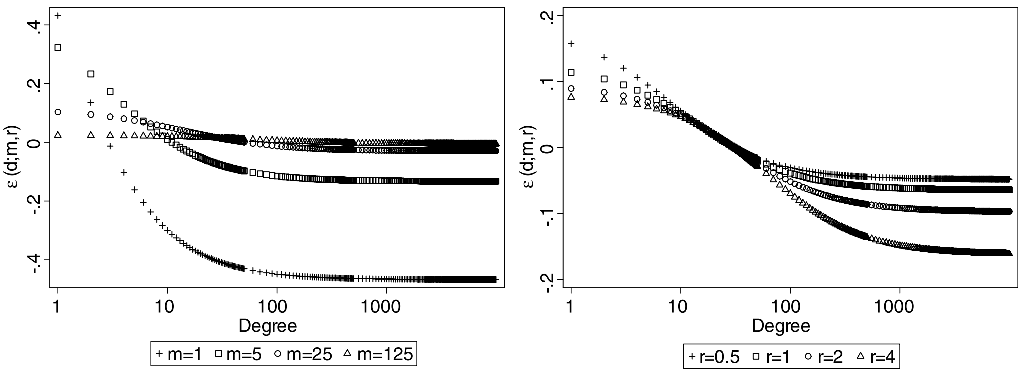

Fig. C.1. Values of $\epsilon(d;m,r)$ for different m,r combinations. In the left panel, I plot $\epsilon(d;m,r)$ for $r=0.5$ and different values of m. In the right panel, I plot $\epsilon(d;m,r)$ for $m=15$ and different values of r. To make the figure more readable, I only plot a subset of the points.

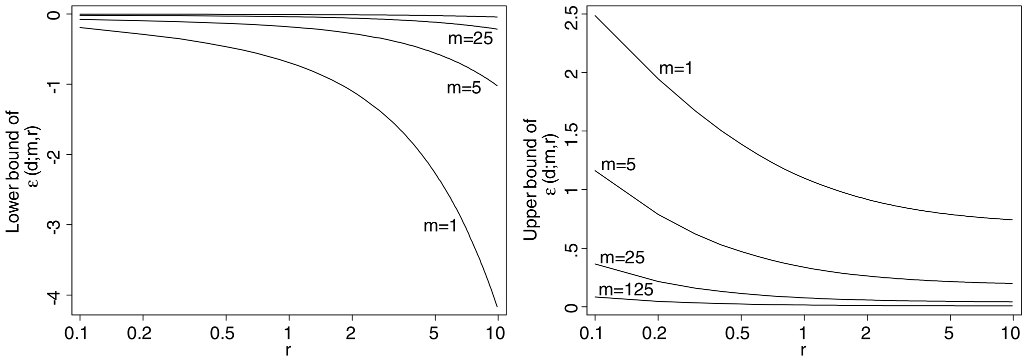

Fig. C.2. Upper and lower bounds of $\epsilon(d;m,r)$ for different m,r combinations. The lower bounds are given in the left panel and the upper bounds are given in the right panel. In the left panel, the top line gives the lower bound of $\epsilon(d,125,r)$.

In Fig. C.1, I plot $\epsilon(d;m,r)$ for different m, r combinations. For the eight combinations that are given in this figure, $\epsilon(d;m,r)$ is positive for small d and negative for large d. When m is small, the mean-field approximation produces a stationary degree distribution that is substantially different from the exact degree distribution. According to the right panel, $\epsilon(d;m,r)$ tends to increase in r.

I assert, without proof, that

$$\epsilon(0;m,r)=\log\left[\frac{\Gamma[2+r+mr]}{mr\cdot\Gamma[(m+1)r+1]}\right],$$

$$\lim_{d\to\infty}\epsilon(d;m,r)=\log\left[\frac{(mr)^{1+r}\Gamma[mr]}{\Gamma[(m+1)r+1]}\right],\quad\text{and}$$

$$\frac{\partial\epsilon(d;m,r)}{\partial d}=\frac{-2+r}{d+mr}-\frac{\Gamma'[d+mr]}{\Gamma[d+mr]}+\frac{\Gamma'[2+d+mr]}{\Gamma[2+d+mr]}<0.$$

Given these results, bounds for $\epsilon(d;m,r)$ are given by the following inequalities:

$$\log\left[\frac{\Gamma[2+r+mr]}{mr\cdot\Gamma[(m+1)r+1]}\right]\geqslant\epsilon(d;m,r)\geqslant\log\left[\frac{(mr)^{1+r}\cdot\Gamma[mr]}{\Gamma[(m+1)r+1]}\right].$$

For different m, r combinations, I plot the lower bound (left panel) and upper bound (right panel) of $\epsilon(d;m,r)$ in Fig. C.2. For extreme values of d, the mean-field approximation generates a degree distribution that is orders of magnitude off from the exact degree distribution. The difference between the two degree distributions is especially acute when m is small. Still, as I have shown in Fig. C.1, the mean-field approximation is innocuous for intermediate values of d.

Appendix D. Simulations

Using simulated data, I analyze the accuracy and precision of the maximum likelihood estimators that are presented in Section 4.1.

Table D.1
Maximum likelihood estimates, with asymptotic standard errors in parentheses. For the first two columns, the likelihood functions depend only on m and r. The estimated standard errors in the middle column cannot be computed because of the near-singularity of the Fisher information matrix.

Parameters m, r, α, β	Estimates		
	$\mathcal{L}_1$	$\mathcal{L}_2$	$\mathcal{L}_3$
20, 2, 1, 5	4.36, 0.47	$8.33 \cdot 10^{30}$, $1.40 \cdot 10^{-31}$	21.60, 2.13, 1.01, 5.66
	(0.14, 0.02)	(?, ?)	(2.57, 0.21, 0.02, 0.63)
20, 2, 1, 1	13.40, 0.51	$1.56 \cdot 10^{19}$, $2.22 \cdot 10^{-19}$	19.92, 1.87, 1.00, 0.96
	(0.38, 0.03)	(?, ?)	(0.29, 0.04, 0.02, 0.03)
20, 1, 1, 5	4.91, 0.28	$2.21 \cdot 10^{33}$, $4.42 \cdot 10^{-34}$	22.68, 0.98, 1.04, 5.90
	(0.25, 0.02)	(?, ?)	(3.83, 0.15, 0.03, 0.70)
20, 1, 1, 1	15.98, 0.29	$5.81 \cdot 10^{19}$, $5.02 \cdot 10^{-20}$	20.02, 0.96, 0.99, 0.96
	(0.80, 0.02)	(?, ?)	(0.37, 0.02, 0.02, 0.04)
40, 2, 1, 5	11.26, 0.29	$4.80 \cdot 10^{25}$, $4.48 \cdot 10^{-26}$	38.66, 1.99, 0.98, 4.63
	(0.55, 0.02)	(?, ?)	(3.00, 0.13, 0.02, 0.36)
40, 2, 1, 1	27.88, 0.46	$3.75 \cdot 10^{14}$, $1.78 \cdot 10^{-14}$	39.71, 1.92, 0.99, 0.96
	(0.87, 0.02)	(?, ?)	(0.39, 0.03, 0.01, 0.03)
40, 1, 1, 5	12.20, 0.19	$2.76 \cdot 10^{27}$, $6.28 \cdot 10^{-28}$	39.95, 1.12, 1.00, 5.34
	(0.89, 0.02)	(?, ?)	(5.34, 0.13, 0.02, 0.55)
40, 1, 1, 1	37.86, 0.20	$1.06 \cdot 10^{19}$, $5.25 \cdot 10^{-19}$	39.65, 1.00, 0.97, 0.96
	(2.63, 0.02)	(?, ?)	(0.50, 0.02, 0.01, 0.03)

In the synthetic datasets that I construct, each node i is endowed with an age quantile, a_i, fitness, p_i, and degree, d_i. Each node's fitness is independently drawn from a Beta(α, β) distribution. Then, given a_i and p_i, the degree distribution is randomly generated from the following probability distribution function,

$$f(d_i|a_i, p_i) = \frac{\Gamma[\bar{p}mr + d_i]}{d_i! \Gamma[mr\bar{p}]}(1-a_i)^{\frac{p_i mr}{1+r}}\left(1-(1-a_i)^{\frac{p_i}{\bar{p}(1+r)}}\right)^{d_i}.$$

For eight different m, r, α, β combinations, I generate a random sample of 25,000 nodes. I then perform maximum likelihood estimation using the $\mathcal{L}_1, \mathcal{L}_2$, and $\mathcal{L}_3$ likelihood functions. In Table D.1, I present the results. As one would hope, the estimates presented in the third column match up with the actual parameters. However, the estimated standard errors are large, even with a sample size of 25,000 nodes.

Appendix E. Nonparametric estimate of $b(p)$

In this section, I argue that the main results of Section 4 do not hinge on the assumption that fitness levels are drawn from the Beta distribution.

To provide support for this claim, I re-estimate the model corresponding to the $\mathcal{L}_3$ likelihood function, allowing for many more degrees of freedom for the fitness distribution, $b(p)$. In particular, I assume that:

$$b(p) = b_x \quad \text{if } p \in \left[\frac{x}{100}, \frac{x+1}{100}\right),$$
$$\text{for } x \in 0, 1, 2, \ldots, 99, \text{ where } b_x \geqslant 0 \text{ and } \frac{1}{100}\sum_{x=0}^{99} b_x = 1. \tag{E.1}$$

In words, $b(p)$ is constant within bins of size 0.01, but may vary arbitrarily across bins.

The MLE estimate of $b(p)$, for this new model, is of similar shape to the estimate of $b(p)$ given in Fig. 4 (see Fig. E.1). The MLE estimates of m and r are (729.31, 1.75). Compared to the estimates given in Table 1, the estimate of m is smaller, and the estimate of r is larger.

Finally, I re-produce Fig. 9, to compare the relative importance of age and fitness as a source of variation in the degree distribution (see Fig. E.2). Note that, because of the discontinuities in the new version of $b(p)$, $\Psi(a, p) \equiv \frac{\partial E[d|a,p]}{\partial a} \div \frac{\partial E[d|a,p]}{\partial \theta}$ will now be discontinuous as well. Again, for most of the relevant part of the parameter space, $\Psi(a, p)$ is less than 1.

Appendix F. Computing $\Psi(a, p)$

The task in this section is to compute:

$$\Psi(a, p) \equiv \frac{\frac{\partial E[d|a,p]}{\partial a}}{\frac{\partial E[d|a,p]}{\partial \theta}}.$$

Remember from Eq. (15) that $E[d|a, p] = m\bar{p}r((1-a)^{-\frac{p}{\bar{p}(1+r)}} - 1)$.

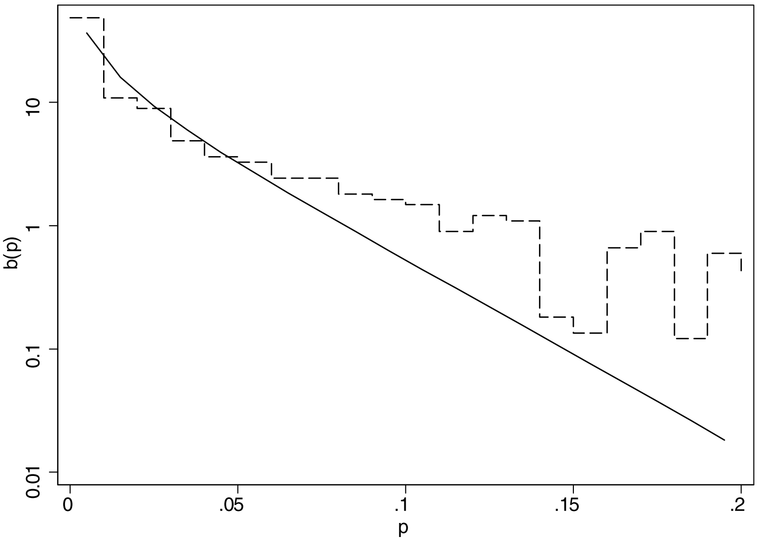

Fig. E.1. Estimates of $b(p)$. The solid line gives the estimate of $b(p)$ resulting from the assumption of Beta-distributed fitness. The dashed line gives the estimate corresponding to Eq. (E.1).

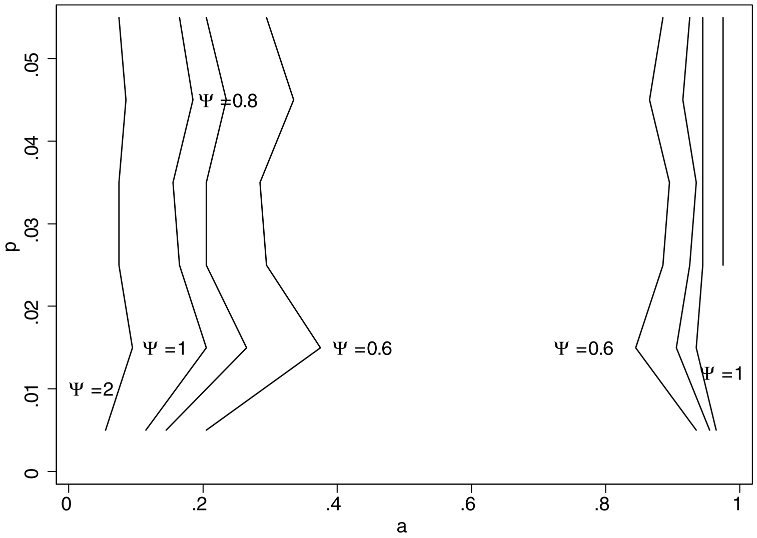

Fig. E.2. Isoquants of $\Psi(a, p)$, using the estimate of $b(p)$ parameterized by Eq. (E.1).

The marginal effect on expected degree of increasing age is:

$$\frac{\partial E[d|a,p]}{\partial a} = \frac{pmr}{1+r}(1-a)^{-\frac{p}{\bar{p}(1+r)}-1}. \tag{F.1}$$

For $\theta \in [0,1]$, define $q(\theta;\alpha,\beta)$ to be the θ-quantile fitness. The marginal effect, on expected degree, of increasing θ is:

$$\frac{\partial E[d|a,p]}{\partial \theta} = \frac{\partial E[d|a,p]}{\partial p}\frac{\partial p}{\partial \theta} = -\frac{mr}{1+r}(1-a)^{-\frac{p}{\bar{p}(1+r)}}\log(1-a)\frac{\partial p}{\partial \theta}. \tag{F.2}$$

The partial derivative, $\frac{\partial p}{\partial \theta}$, tells us how much the fitness increases from one percentile to the next. For example, the median-fitness node has $p = 0.00786$, while the 51st percentile node has a fitness of $p = 0.00823$. Thus, the derivative, evaluated at $\theta = 0.5$, is roughly $\frac{0.00823-0.00786}{0.01} = 0.0376$.

From the definition of q and the probability distribution function of a Beta distribution:

$$\frac{\partial p}{\partial \theta} = \frac{\Gamma[\alpha]\Gamma[\beta]}{\Gamma[\alpha+\beta]}\big(q(\theta;\alpha,\beta)\big)^{1-\alpha}\big(1-q(\theta;\alpha,\beta)\big)^{1-\beta}. \tag{F.3}$$

Plug $\theta = q^{-1}(p;\alpha,\beta)$ into Eq. (F.3) to get:

$$\frac{\partial p}{\partial \theta} = \frac{\Gamma[\alpha]\Gamma[\beta]}{\Gamma[\alpha+\beta]}\bigl(q\bigl(q^{-1}(p;\alpha,\beta);\alpha,\beta\bigr)\bigr)^{1-\alpha}\bigl(1-q\bigl(q^{-1}(p;\alpha,\beta);\alpha,\beta\bigr)\bigr)^{1-\beta}. \tag{F.4}$$

Combining Eqs. (F.1) and (F.2) leads us to

$$\Psi(a,p) = -\frac{p}{\frac{\partial p}{\partial \theta}}\frac{1}{(1-a)\log[1-a]},$$

where $\frac{\partial p}{\partial \theta}$ is given in Eq. (F.4).

Appendix G. Proofs of Propositions 1 and 2

In this section, I restate and prove Propositions 1 and 2.

Proposition 1 *(Reminded). Given the assumptions specified above, the clustering coefficient tends to*:

$$C = \frac{1}{m\bar{p}}\cdot\frac{\alpha}{\alpha+\beta}\left(1-\frac{\alpha}{\alpha+\beta}\cdot\frac{1}{\bar{p}}\cdot\frac{r}{1+r}\right).$$

Proof. To solve for C, I closely mimic the proof of Theorem 2 of Jackson and Rogers (2007).

First, I compute, $\bar{p}_n$, the probability that a network-based meeting ends with a link. To do so, I note that the average probability that a random meeting produces a link is simply

$$\bar{p}_r = \int_0^1 pb(p)\,dp = \frac{\alpha}{\alpha+\beta}. \tag{G.1}$$

Also, $\bar{p}$ is a weighted average of $\bar{p}_r$ and $\bar{p}_n$:

$$\bar{p} = \frac{m_r\bar{p}_r + m_n\bar{p}_n}{m_r+m_n} = \frac{r}{1+r}\cdot\frac{\alpha}{\alpha+\beta} + \frac{1}{1+r}\bar{p}_n.$$

Solve for $\bar{p}_n$:

$$\bar{p}_n = \bar{p}(1+r) - r\frac{\alpha}{\alpha+\beta}. \tag{G.2}$$

For future reference, note that $r > \frac{\beta\bar{p}}{2\alpha+\beta-\beta\bar{p}}$ corresponds to

$$r > \bar{p}(1+r) - r\left(\frac{\alpha}{\alpha+\beta}\right),$$

which is equivalent to:

$$r > \bar{p}_n \quad \text{or} \quad m_r > \bar{p}_n m_n.$$

Now, consider a node i. Upon entry, it forms $m\bar{p}$ links. From each node that i links to, there are $m\bar{p}$ directed links. Thus, there are $(m\bar{p})^2$ possible pairs of directed links $i \to j \to k$. To determine C, I need to calculate the fraction of these links that have $i \to k$ present. As in the proof of Theorem 2 of Jackson and Rogers (2007), I can alternatively count the number of times that either $j \to k$ or $k \to j$ is present, conditional on both $i \to j$ and $i \to k$ being present.

Since, by assumption, $r > \frac{\beta\bar{p}}{2\alpha+\beta-\beta\bar{p}}$ (which is equivalent to $m_r > \bar{p}_n m_n$), I only need to count the number of instances for which node i formed a link with j in a random meeting and then formed a link with a successor of j in a network-based meeting. (In the language of the proof of Theorem 2 of Jackson and Rogers (2007), I only need to consider case 2, and not case 3.)

Because of our assumption that the entrant forms at most one link with the successors of any one particular successor, the number of $j \to k$ links, conditional on both $i \to j$ and $i \to k$ being present, equals $\bar{p}_r \cdot m_n \cdot \bar{p}_n$. Why? I have assumed that $m_r > \bar{p}_n m_n$, so that there are strictly fewer links formed in network-based meetings than there are random meetings. Therefore, the number of network-based meetings is an upper bound for the number of $i \to k$ links for which $i \to j$ and $j \to k$ are also present. Thus, I simply need to multiply m_n by the probability that both the random meeting generates a link and the probability that the network-based meeting generates a link. This explains the $\bar{p}_r$ and $\bar{p}_n$ terms in our product.

Using Eqs. (G.1) and (G.2) $\bar{p}_r \cdot m_n \cdot \bar{p}_n$ equals:

$$\frac{\alpha}{\alpha+\beta} \cdot \frac{m}{1+r}\left(\bar{p}(1+r) - r\frac{\alpha}{\alpha+\beta}\right) = \frac{\alpha}{\alpha+\beta} \cdot m\bar{p}\left(1 - \frac{r}{1+r} \cdot \frac{1}{\bar{p}} \cdot \frac{\alpha}{\alpha+\beta}\right).$$

Divide by $(m\bar{p})^2$ to get the desired expression for C:

$$C = \frac{\alpha}{\alpha+\beta} \cdot \frac{1}{m\bar{p}}\left(1 - \frac{r}{1+r} \cdot \frac{1}{\bar{p}} \cdot \frac{\alpha}{\alpha+\beta}\right). \qquad \square$$

Proposition 2 *(Reminded). Consider two nodes, i and i′, with $d_i > d_{i'}$. Under the mean-field approximation, the degree distribution of the predecessors of i first order stochastically dominates the degree distribution of the predecessors of i′.*

Proof. The proof proceeds in three steps. In the first step, I argue that the age distribution of node i FOSD the age distribution of i'. In the second step, I use the finding of the first step to argue that the age of a randomly selected predecessor of node i FOSD the age of a randomly selected predecessor of i'. Then, in third step, I integrate over the age distributions of the predecessors of i and i' to arrive at the desired result.

Under the mean-field approximation, d is deterministically given by any a, p combination. In particular, $d = \bar{p}rm((1-a)^{-\frac{p}{\bar{p}}\frac{1}{1+r}} - 1)$. I invert this relationship to solve for p in terms of a and d:

$$\frac{\bar{p}rm}{d+\bar{p}rm} = (1-a)^{\frac{p}{\bar{p}}\frac{1}{1+r}},$$

$$p = \bar{p}(1+r)\frac{\log(\frac{\bar{p}rm}{d+\bar{p}rm})}{\log(1-a)} \equiv \pi(d,a). \tag{G.3}$$

Let $k(a|d)$ be the probability distribution function of a node's age quantile, conditional on the degree of the node equaling d. Let the analogous cumulative distribution function be $K(a|d)$. Using Eq. (G.3), the probability that a node's age is less than a equals the probability that the fitness of the node is greater than $\pi(d,a)$.

$$K(a|d) = \int_{\pi(d,a)}^{1} b(p)\,dp.$$

Differentiate with respect to d:

$$\frac{\partial K(a|d)}{\partial d} = -b\big(\pi(d,a)\big)\frac{\partial \pi(d|a)}{\partial d} < 0.$$

This concludes the first step of the proof, which states that $d_i > d_{i'}$ implies $K(a|d_i) < K(a|d_{i'})$.

Before proceeding to the second step of the proof, in Fig. G.1, I graphically depict the idea behind the first step. This is purely for pedagogical purposes. Consider the expression $K(0.005, d_i)$. To compute this function, I integrate $b(p)$ from $p = \pi(d_i, 0.005)$ to $p = 1$. This is the thick part of the solid curve in Fig. G.1. The difference between $K(0.005, d_{i'})$ and $K(0.005, d_i)$ is simply the integral of $b(p)$ from $p = \pi(d_{i'}, a)$ to $p = \pi(d_i, a)$ (which in Fig. G.1 is the solid vertical line that connects $p = \pi(d_{i'}, a)$ to $p = \pi(d_i, a)$ at $a = 0.005$). The important thing to notice is that this integral must be positive, so $K(0.005, d_{i'}) > K(0.005, d_i)$.

To begin the second step, define $J(a|\xi_i)$ as the cumulative distribution function of the age of a randomly selected predecessor of node i, given that the age of i is ξ_i. Note that $J(a|\xi_i)$ is a weakly decreasing function of ξ_i: knowing that node i is younger (ξ_i is smaller) means that it is more likely that a is small ($J(a|\xi_i)$ is bigger). Also, note that $J(a|\xi_i) = 1$ for $\xi \in [0, a]$, since the predecessor of i cannot be older than i. Furthermore, define $G(a|d_i)$ as the cumulative distribution function of the age of a predecessor of i, conditional on the degree of i being d_i. Using these definitions:

$$G(a|d_i) - G(a|d_{i'}) = \int_0^1 J(a|\xi)\big[k(\xi|d_i) - k(\xi|d_{i'})\big]\,d\xi. \tag{G.4}$$

Because (a) K is increasing in d, in the sense of first-order stochastic dominance, and (b) $J(a|\xi)$ is nonincreasing in d, the right-hand side of Eq. (G.4) is negative. This completes the second part of the proof.

For the final step, define $L(d|d_i)$ as the cumulative distribution function of the degree of a randomly chosen predecessor of i, conditional on the degree of i being d_i. I need to show that $d_i > d_{i'}$ implies $L(d|d_i) < L(d|d_{i'})$.

Conditional on its fitness level, a node has degree less than d if its age is less than $1 - (\frac{\bar{p}rm}{d+\bar{p}rm})^{\frac{\bar{p}}{p}(1+r)}$ (see Eq. (G.3)). Thus, conditional on the degree of node i, the cumulative distribution function of a randomly selected predecessor of node i is:

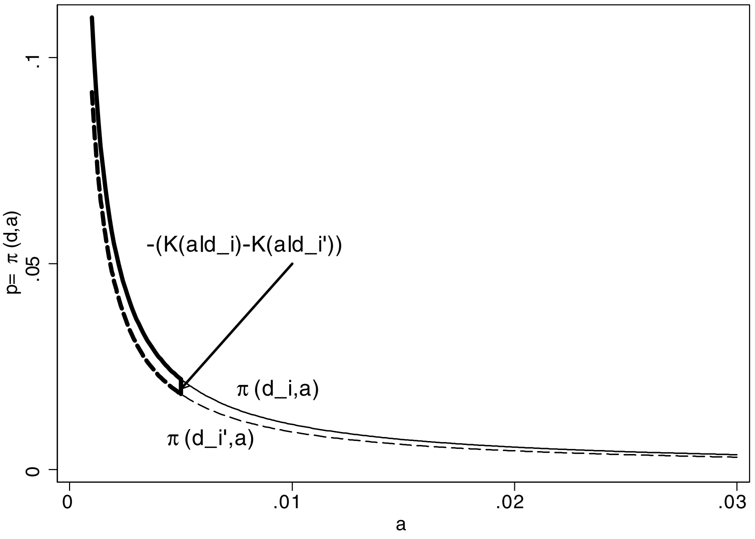

Fig. G.1. Graphical representation of the first part of the proof. $p = \pi(d|a)$ is drawn for two different values of d. As d gets bigger, the interval of p such that age is less than a gets smaller. Because of this, as d increases, $K(a|d)$ decreases.

$$L(d|d_i) = \frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]} \int_0^1 G\left(1 - \left(\frac{\bar{p}rm}{d+\bar{p}rm}\right)^{\frac{\bar{p}}{p}(1+r)}; d_i\right) p^{\alpha-1}(1-p)^{\beta-1}\,dp.$$

Also:

$$L(d|d_{i'}) = \frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]} \int_0^1 G\left(1 - \left(\frac{\bar{p}rm}{d+\bar{p}rm}\right)^{\frac{\bar{p}}{p}(1+r)}; d_{i'}\right) p^{\alpha-1}(1-p)^{\beta-1}\,dp.$$

Thus:

$$\begin{aligned} L(d|d_i) - L(d|d_{i'}) = {} & \frac{\Gamma[\alpha+\beta]}{\Gamma[\alpha]\Gamma[\beta]} \int_0^1 p^{\alpha-1}(1-p)^{\beta-1} \\ & \times \left\{ G\left(1 - \left(\frac{\bar{p}rm}{d+\bar{p}rm}\right)^{\frac{\bar{p}}{p}(1+r)}; d_i\right) - G\left(1 - \left(\frac{\bar{p}rm}{d+\bar{p}rm}\right)^{\frac{\bar{p}}{p}(1+r)}; d_{i'}\right)\right\} dp < 0. \end{aligned}$$

This implies that $L(d|d_i) < L(d|d_{i'})$, which was the desired result. □

References

Atalay, Enghin, Hortaçsu, Ali, Roberts, James W., Syverson, Chad, 2011. Network structure of production. Proc. Natl. Acad. Sci. 108 (13), 5199–5202.

Barabási, Albert-László, Albert, Réka, 1999. Emergence of scaling in random networks. Science 286 (5439), 509–512.

Bianconi, Ginestra, Barabási, Albert-László, 2001. Competition and multiscaling in evolving networks. Europhys. Lett. 54 (4), 436–442.

Bramoullé, Yann, Currarini, Sergio, Jackson, Matthew O., Pin, Paolo, Rogers, Brian W., 2012. Homophily and long-run integration in social networks. J. Econ. Theory 147 (5), 1754–1786.

Caldarelli, Guido, Capocci, Andrea, De Los Rios, Paolo, Muñoz, Miguel A., 2002. Scale-free networks from varying vertex intrinsic fitness. Phys. Rev. Lett. 89 (25), 258702.

Chaney, Thomas, 2011. The network structure of international trade. Working paper.

Conley, Timothy G., Udry, Christopher R., 2010. Learning about a new technology: Pineapple in Ghana. Amer. Econ. Rev. 100 (1), 35–69.

Crespo, Juan A., Cuenda, Sara, 2010. A comment on Meeting strangers and friends of friends: How random are social networks? Working paper.

Dorogovtsev, Sergey N., Mendes, Jose F., Samukhin, Alnis N., 2000. Structure of growing networks with preferential linking. Phys. Rev. Lett. 85 (21), 4633–4636.

Erdös, Paul, Rényi, Alfréd, 1960. On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci. 5, 17–61.

Gehrke, Johannes, Ginsparg, Paul, Kleinberg, Jon, 2003. Overview of the 2003 KDD cup. SIGKDD Explor. Newsl. 5 (2), 149–151.

Hall, Bronwyn H., Jaffe, Adam B., Trajtenberg, Manuel, 2001. The NBER patent citation data file: Lessons, insights and methodological tools. Working paper.

Jackson, Matthew O., 2011. An overview of social networks and economic applications. In: Benhabib, J., Bisin, A., Jackson, M.O. (Eds.), The Handbook of Social Economics. North-Holland, pp. 511–585.

Jackson, Matthew O., Rogers, Brian W., 2007. Meeting strangers and friends of friends: How random are social networks? Amer. Econ. Rev. 97 (3), 890–915.

Jackson, Matthew O., Wolinsky, Asher, 1996. A strategic model of social and economic networks. J. Econ. Theory 71 (1), 44–74.

Newman, Mark E.J., 2003. Mixing patterns in networks. Phys. Rev. E 67, 026126.

Ostrovsky, Michael, 2008. Stability in supply chain networks. Amer. Econ. Rev. 98 (3), 897–923.

Pennock, David M., Flake, Gary W., Lawrence, Steve, Glover, Eric J., Giles, C. Lee, 2002. Winners don't take all: Characterizing the competition for links on the web. Proc. Natl. Acad. Sci. 99 (8), 5207–5211.

Vázquez, Alexei, 2003. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev. E 67, 056104.