Theoretical Economics 11 (2016), 813–863 1555-7561/20160813

The formation of networks with local spillovers and limited observability

Michael D. König
Department of Economics, University of Zurich

This paper analyzes the formation of networks in which each agent is assumed to possess some information of value to the other agents in the network. Agents derive payoff from having access to the information of others through communication or spillovers via the links between them. Linking decisions are based on network-dependent marginal payoff and a network-independent noise capturing exogenous idiosyncratic effects. Moreover, agents have a limited observation radius when deciding to whom to form a link. I find that for small noise the observation radius does not matter and strongly centralized networks emerge. However, for large noise, a smaller observation radius generates networks with a larger degree variance. These networks can also be shown to have larger aggregate payoff. I then estimate the model using a network of co-inventors and scientific collaborations in physics and economics, and find that the model can closely reproduce a variety of observed patterns. I show that local search is important in all the empirical networks conside, but that economists tend to search more broadly for new collaboration opportunities.

Keywords. Diffusion, network formation, growing networks, limited observability.

JEL classification. C63, D83, D85, L22.

1. Introduction

Networks are important in explaining a large variety of social and economic phenomena. This insight has lead to an increasing interest in the study of networks in economics

Michael D. König: michael.koenig@econ.uzh.ch

I am grateful to Matt Jackson for his guidance and support. Moreover, I thank Mathias Staudigl for the excellent research assistance in the early stages of the paper. I would like to thank Christian Zimmermann, Sanjeev Goyal, Andrea Galeotti, Brian Rogers, Yves Zenou, Ben Golub, Tomás R. Barraquer, Lee Fleming, Fabrizio Zilibotti, Kjetil Storesletten, Anton Kolotilin, and seminar participants at University of Berkeley, University of Cambridge, University of Vienna, University of Bielefeld, University of Zurich, ETH Zurich, and Stanford University for their insightful comments. I would like to thank Thomas Krichel for providing access to the data. Financial support from Swiss National Science Foundation through research Grants PBEZP1-131169 and 100018_140266 is gratefully acknowledged. Finally, I would like to thank SIEPR and the Department of Economics at Stanford University for their hospitality during 2010–2012. A previous version of this paper was circulated under the title "Centrality Based Network Formation of Boundedly Rational Agents with Limited Information."

DOI: 10.3982/TE1524

and related sciences accompanied by a growing number of publications in the field.[1] Networks play a particularly important role in understanding the process of communication of information and knowledge diffusion among diverse actors, ranging from inventors to scientists. In this paper, I introduce a simple, parsimoniously parameterized and tractable model to study the emergence of networks of information and knowledge diffusion, which is able to match and explain the observed empirical patterns on an unprecedented scale.

A large body of literature has emphasized the crucial effect of social networks of inventors on the productivity of innovative regions (see, e.g., Marshall 1919, Allen 1983, Singh 2005, Almeida and Kogut 1999). A prominent example is the success story of Silicon Valley, which has been attributed to its informal networks of friendship and collaboration (Saxenian 1994, Fleming et al. 2007). Similarly, the formation of collaborations in academic research is a crucial component in the process of scientific discovery and knowledge production (Newman 2001a, 2004, Fafchamps et al. 2010, Goyal et al. 2006).

The networks of inventors and scientific collaborations share a number of empirical regularities. First, the distributions of degree (the number of links of a node) in these networks exhibit fat tails, typically decaying as a power law.[2] Similarly, the average clustering coefficient (Watts and Strogatz 1998), i.e., the fraction of connected neighbors of a node, tends to decrease with the degree and also exhibits a power-law decay (cf. Goyal et al. 2006). Moreover, the distribution of (small) connected components (in which there exists a path between every pair of nodes) decays as a power law. Finally, the networks of inventors and coauthors exhibit an increasing average neighbors' degree with the degree of a node, referred to *assortativity* (Newman 2002). In this paper, I introduce a simple model that can explain all these distributions. This is a novel contribution and extends previous studies that focused primarily on the distribution of degree or some aggregate statistics (cf. Jackson and Rogers 2007).

I consider a degree-based approximation (see below) to a general class of models (payoff functions) in which each agent is assumed to possess some information of value to the other agents in the network. Agents derive payoff from having access to the information of others through direct communication or spillovers along the links in the network.[3] Agents' incentives to form links can be partitioned into a network-dependent part as well as a network independent exogenous random term, referred to as *noise*. The network-dependent part of agents' payoffs (represented by the degree) derives from having access to the information of others. The noise term captures exogenous random

[1]This literature has steadily grown in the last decade. The monographs of Jackson (2008), Goyal (2007), and Vega-Redondo (2007) are excellent surveys for the economic theory of networks. See also Newman (2010) for a survey of the literature in physics, and Durrett (2007) for a concise review of the literature on networks in mathematics.

[2]A power-law degree distribution in patent citation networks has been documented in e.g. Valverde et al. (2007).

[3]For the purpose of tractability, in this paper I consider a simplified setup with myopically rational agents, and I ignore issues related to private signals, strategic communication, and inference problems with Bayesian updating (see, e.g., Hagenbach and Koessler 2010, Calvó-Armengol and De Martï 2007, for alternative setups).

perturbances, shortcomings in assessing the correct value of information possessed by other agents, and exogenous matching effects.[4]

Moreover, it is assumed that the information transmitted through the links in the network is exposed to decay, making information that travels longer distances less valuable (cf. Bala and Goyal 2000, Jackson and Wolinsky 1996). In this paper, I focus on the case of *strong decay*, or weak knowledge spillover effects, where the value for an agent of being connected in a network is determined by his immediate neighbors (cf. Galeotti et al. 2010). This assumption is consistent with the empirical evidence. For example, Singh (2005) and Breschi and Lissoni (2005) find in their studies of patent networks that the existence of a tie is associated with a greater probability of knowledge flow, while the probability is decreasing as the path length increases, and the probability is becoming small or nearly null for social distance greater than 2. In turn, this implies that the marginal return from connecting to an agent is determined by his degree.[5]

Agents sequentially enter the network and obtain an opportunity to acquire information from the incumbent agents through forming links. Upon entry, each agent can sample a given number of existing agents in the network and observes these agents and their neighbors (cf. Friedkin 1983).[6] I call the number of sampled agents the *observation radius*. He then forms links to the observed agents in the sample based on the marginal payoff (determined by the degree) obtained from each link. With this sampling procedure I follow a common approach in the statistics and sociology literature for how individuals collect information on an existing population that is difficult to observe called *snowball/star sampling* (Goodman 1961, Frank 1977, Kolaczyk 2009).[7]

I analyze the emerging networks for different observation radii and levels of noise. I find that for small noise the observation radius does not matter and strongly centralized networks emerge. However, for large noise, a smaller observation radius generates networks with a larger degree variance. One can show that the aggregate payoff maximizing networks in the class of models considered here increases with the degree variance.[8] Hence, I find that when the exogenous noise is large, then a smaller observation radius leads to networks that have larger aggregate payoff. This provides an example in the context of a network-based meeting process where "knowing less can be better."

Collaboration and the formation of teams involve opportunity, time, and friction costs, and information available in the circle of acquaintances will be more easily available. Hence, collaborations are more likely the closer individuals are in the network of

[4]The introduction of noise also allows me to compare the current model with other papers that consider a random network formation process (i.e., strong noise), such as the landmark model by Jackson and Rogers (2007).

[5]Newman (2001b) finds in his empirical study of co-authorship networks that the probability of a scientist acquiring a new collaborator increases with the number of his past collaborators, that is, his degree.

[6]In a similar way Jackson and Rogers (2007), Galeotti et al. (2010), McBride (2006), Alós-Ferrer and Weidenholzer (2008) assume that agents have only limited information of the network.

[7]See von Hippel et al. (1999) for a case study where a firm uses snowball sampling to collect information from customers and their contacts.

[8]Similarly, Westbrock (2010) shows that in the model by Goyal and Moraga-González (2001), where firms are competing on the product market while they can form research and development (R&D) collaborations to reduce their production costs, welfare positively correlates with the degree variance.

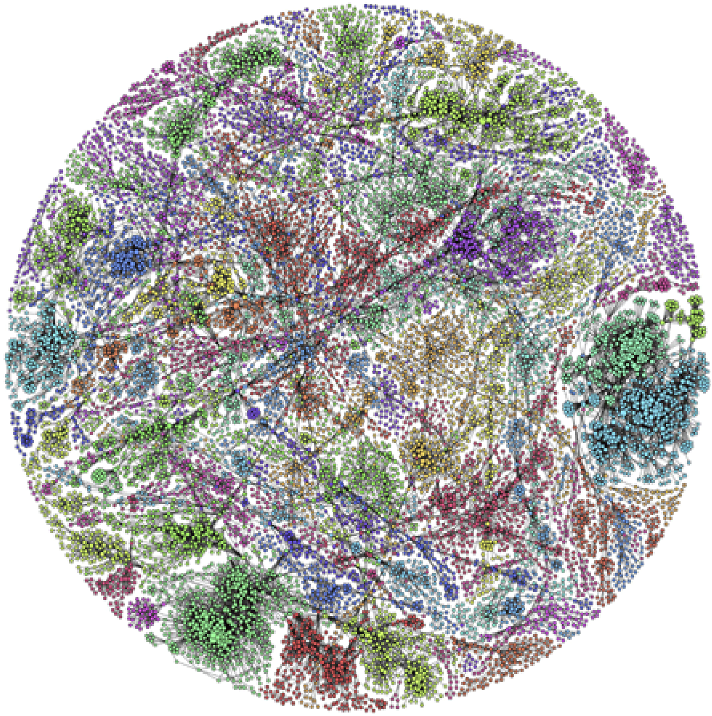


Figure 1. The network between co-inventors in the drugs development sector. Node colors indicate different clusters of densely connected nodes using the modularity algorithm proposed in Blondel et al. (2008); node sizes indicate their degree. The figure illustrates the existence of highly connected clusters of inventors. In this paper, I develop a model that can explain the formation of these densely connected groups through a local search process for new collaboration partners. See Section 7 for further details about the data.

social ties (Fafchamps et al. 2010, Goyal et al. 2006). In the empirical application of the model in this paper I compare the tendency of inventors (see Figure 1 for an example of a network between co-inventors in the drugs development sector), and scientists in condensed matter physics and economics to select their research collaborations in their local neighborhood as compared to searching for more distant partners. I show that local search is important in all the empirical networks considered, but that economists tend to search more broadly for collaboration opportunities as the estimated observation radius tends to be higher. This might reflect the diverse backgrounds and application areas of economists, which can also be witnessed in the large number of classification codes categorizing economic sciences,[9] and that economists are less constrained by their institutional boundaries.[10]

[9]Comparing, for example, the number of academic disciplines and sub-disciplines listed on Wikipedia we find 50 in economics and only 27 in physics. See also http://en.wikipedia.org/wiki/List_of_academic_disciplines_and_subdisciplines.

[10]Schilling and Green (2011) find that search scope, search depth, and atypical connections between different research domains significantly increase a paper's impact in the social sciences. Similarly,

The paper in the economics literature most closely related to the one presented here is the seminal work by Jackson and Rogers (2007).[11] In this influential article the authors introduce a model of a growing network that combines random search protocols for potential linking partners with local network-based search protocols. By means of theoretical and empirical analysis, they are able to show that their model is very flexible in fitting real-world data. However, while in Jackson and Rogers (2007) links are formed at random, here instead I make a first attempt to start directly from a (degree-based) discrete-choice approach,[12,13] with an explicit modeling of the reasons why links are formed. Further, albeit similar, the difference in the linking processes of their model and the present one allows me to measure empirically the information radius of the agents. Moreover, the results for the degree distribution and efficiency in Jackson and Rogers (2007) are based on a mean-field approximation, while such an approximation is not needed to obtain the corresponding results in the present paper. Further, Jackson and Rogers (2007) do not derive explicitly all the statistics that I do here (such as the average nearest neighbor connectivity, the clustering degree distribution,[14] or the component size distribution), and do not analyze the impact of different observation radii on these statistics. Moreover, with varying levels of the noise in the payoff function of

Jones et al. (2008) find that multi-university research teams produce the highest impact papers. Moreover, they observe that the growth in multi-institution collaboration was greater in the social sciences than science and engineering. For instance, as noted by Winkler et al. (2011), "... research in the biological and chemical sciences almost invariably requires a lab and thus has a strong local component. Research in economics is different: Except for experimental economics, labs are rarely part of economic research; neither is specialized equipment. But data and software can be readily shared and this encourages collaboration."

[11] Besides the economics literature there also exists a large literature in computer science, physics, and mathematics, where similar models are studied. I refer to Krapivsky et al. (2000), Krapivsky and Redner (2001), Oliveira and Spencer (2005), Vázquez (2003), Kumar et al. (2000), Wang et al. (2009), and Toivonen et al. (2006), to mention only a few. However, these authors typically do not make explicit behavioral assumptions about why links are formed, do not analyze welfare implications, and do not estimate their models for empirically observed networks.

[12] The payoff function I introduce focuses on the case of strong decay of the knowledge transmitted along the links between agents, such that the degree of an agent determines his propensity to acquire new links (see, e.g., Singh (2005), Breschi and Lissoni (2005), Newman (2001b) for an empirical motivation), so that the degree of an agent will be a sufficient statistic to assess the agent's marginal payoff from forming links (see Assumption 1 in Section 2.1). There is clearly much more work to be done in this area that incorporates more general payoff functions, but I believe that the model considered here provides a first step toward a better understanding of the structure of real-world networks and the incentives and information sets that are influencing the processes that generate these networks.

[13] The general class of models considered here (see also supplementary Appendix E) has the property that the payoff of an agent is increasing with the number of collaborations, i.e., his degree. This is a characteristic that has been found in empirical studies of co-authorship networks (see, e.g., Abbasi et al. 2011, Ductor 2015).

[14] Jackson and Rogers (2007, p. 900) conjecture that $C(k)$, the clustering coefficient for a node with in-degree k, is a strictly decreasing function of degree k. Here I complement their analysis by providing asymptotic expressions for $C(k)$ showing that $C(k)$ is not only decreasing with the degree, but actually decaying as a power law with a well defined exponent. The intuition behind the decreasing clustering coefficient with the degree is that in this growing network model older agents are connected to a larger number of younger agents, and these younger agents have not only a smaller number of links but are also less likely to be connected among each other. In contrast, the younger agents are primarily connected to older agents, who have a higher degree and are more likely to be connected among other old agents. Hence, we observe a

the agents, a transition from assortative to dissortative networks can be observed in the model. As Jackson and Rogers (2007) do not provide results for the average nearest neighbor connectivity, and they do not have a payoff function governing the decision with whom to form a link, this behavior cannot be studied in their setup. Besides, a feature of the model by Jackson and Rogers (2007) is that it generates dissortative undirected networks.[15] This is particularly problematic when we look at the networks in the empirical examples in Section 7 (networks of co-authors and co-inventors), which are all assortative. Also, when the marginal payoff of agents is increasing in the degree and there is no exogenous noise, then differently to the efficiency results obtained in Jackson and Rogers (2007), I show that the observation radius has no impact on aggregate payoffs and efficiency.

Based on the pioneering model by Jackson and Rogers (2007) a number of extensions and applications have been suggested. Ghiglino (2011) introduces an algorithm similar to Jackson and Rogers (2007) to study the creation and recombination of ideas from a pool of existing knowledge (more precisely, networks of citations between scientific publications). Bramoullé et al. (2012) and Vigier (2014) introduce different types of agents and study the mechanisms underlying homophily, that is, the tendency of similar types of agents being connected. Moreover, Kovářík and Van der Leij (2014) introduce risk aversion in the decisions of agents to form links locally or globally. They show that risk aversion can lead to increased clustering in the network. In contrast, in Chaney (2014) a spatial extension is suggested in which the network is embedded into geographical space and agents who are closer in space are more likely to form links. Differently to these authors, I introduce a behavioral foundation (albeit simplistic) of why links are formed in the model by Jackson and Rogers (2007) in the context of knowledge diffusion in networks. Moreover, none of these works investigates all the empirical networks that I do in the present paper and estimates the model for these data.

The paper is organized as follows. In Section 2, I introduce the general modeling framework. Section 2.1 defines the payoff agents derive from the network. Next, in Section 2.2, I describe the evolution of the network. In Section 3, I analyze the networks generated by the model, while Section 4 provides an efficiency analysis and shows how the level of noise and the observation radius affect aggregate payoffs. Section 5 analyzes correlations between an agent and his neighbors. Section 6 discusses several extensions of the model. Section 7 contains an application of the model to different real-world networks. Section 8 concludes. Appendix A contains a few basic definitions and notation. All proofs are relegated to Appendix B. Supplementary Appendices C and D,

---

negative clustering degree relationship. The numerator of the clustering coefficient then grows roughly linearly and the denominator grows roughly quadratically with degree, which results in a power-law behavior of the clustering degree distribution. See Section 5.2 for a more detailed discussion.

[15]Assortativity in the model by Jackson and Rogers (2007) is found for the average nearest in-neighbor connectivity that is increasing with the in-degree. The intuition is that older agents are more likely to form links to other old agents with high degrees, while younger agents are more likely to form links to other young agents with smaller degrees. This gives rise to an assortative trend. However, this does not hold in the undirected closure of the network. In the undirected network the average nearest neighbor degree of the younger agents is now much higher because it includes not only the in-neighbors, but also the out-neighbors, who are older agents with high degrees. This can reverse the positive trend of the average nearest neighbor connectivity and give rise to a dissortative network. See Section 5.1 for more details.

available in a supplementary file on the journal website, http://econtheory.org/supp/1524/supplement.pdf, present some technical details of the sampling scheme. Various examples in the literature that fall into the general class of games considered here are discussed in supplementary Appendix E. A detailed explanation of the empirical method and results are given in supplementary Appendix F. Finally, supplementary Appendices G and H provide a more detailed discussion of the model extensions introduced in Section 6.

2. The model

In the following sections we introduce the payoff agents derive from being connected in a network and their incentives to form links within a dynamic network formation process. The basic definitions and notation used throughout the paper can be found in Appendix A.

2.1 *Payoffs*

For a given network $G = \langle \mathcal{N}, \mathcal{E} \rangle \in \mathcal{G}(n)$ we assign each agent $i \in \mathcal{N}$ a payoff $\pi_i(\cdot, \delta) : \mathcal{G}(n) \to \mathbb{R}$ that depends on the network G and a (decay) parameter $\delta \geq 0$ that measures the degree of interdependency between agents' payoffs in G. We define the *link incentive function* $f_i : \mathcal{G}(n) \times \mathcal{N} \to \mathbb{R}$ for an agent $i \in \mathcal{N}$ as

$$f_i^{\delta}(G, j) \equiv \pi_i(G \oplus ij, \delta) - \pi_i(G, \delta),$$

which measures the marginal payoff to the agent i resulting from the potential link $ij \notin \mathcal{E}$. Here we focus on link incentive functions (and therefore on classes of games) that satisfy the following conditions.

Assumption 1. *For all $i \in \mathcal{N}$ the link incentive function $f_i^{\delta}(G, \cdot) : \mathcal{N} \to \mathbb{R}$ has the following properties:*

(LM) Link monotonicity. The incentives to link are nonnegative, i.e., $f_i^{\delta}(G, j) \geq 0$ for all $j \neq i \in \mathcal{N}$.

(LD) Linear differences. With strong decay, the incentives to link to an agent are increasing with his degree, i.e., for all $ij, ik \notin \mathcal{E}$, there exists a constant $\gamma \geq 0$ and a linear increasing function $g : \mathbb{R} \to \mathbb{R}$ such that

$$\frac{f_i^{\delta}(G, j) - f_i^{\delta}(G, k)}{\delta^{\gamma}} = g(d_G(j) - d_G(k)) + o(1)$$

holds in the limit of $\delta \to 0$.

Let us briefly discuss the implications of these two conditions in turn. Link monotonicity (LM) requires that the incentives to link are nonnegative. Intuitively it says that no link to be formed can harm an agent (cf. Dutta et al. 2005). Condition (LD), linear differences, allows us to order the linking incentives for the entering agent across all potential linking partners. It says that the agent i has the highest incentive to direct a link to the agent who has the current highest degree among all alternative linking partners.[16] Two potential links are judged as being equally attractive for the agent if the involved agents have the same degree in the current network.[17] Further, the assumption of strong decay should capture the fact that in this model an agent cares about knowledge spillovers from direct and indirect neighbors, that is, up to length-2 connections but not longer.[18]

For our efficiency analysis, we further make the following assumption.

Assumption 2. *Let $\Pi : \mathcal{G}(n) \times \mathbb{R}_+ \to \mathbb{R}$ denote aggregate payoff defined by $\Pi(G, \delta) \equiv \sum_{i\in\mathcal{N}} \pi_i(G, \delta)$ and let $\sigma_d^2(G)$ be the degree variance of $G \in \mathcal{G}(n, e)$. Then we assume that the following condition holds:*

(DC) Degree concentration. For $n \in \mathbb{N}$ and $0 \leq e \leq \binom{n}{2}$,

$$\arg\max_{G\in\mathcal{G}(n,e)} \Pi(G, \delta) = \arg\max_{G\in\mathcal{G}(n,e)} \sigma_d^2(G)$$

holds in the limit of $\delta \to 0$.

Assumption (DC) assumes that networks with a higher degree inequality, as measured by the degree variance, generate higher welfare.[19] For example, if welfare in an economy depends on the rapid diffusion of knowledge and new technologies, then a centralized structure can be optimal (cf., e.g., König et al. 2012). Assumption (DC) will be needed for our efficiency analysis in Section 4.

Supplementary Appendix E provides a number of examples from the economic literature that satisfy Assumptions 1 and 2. These examples illustrate how the assumptions made in this section arise naturally when knowledge diffuses in networks and the transmission of information along the links is exposed to strong decay or when there are weak knowledge spillover effects between neighboring agents (corresponding to small values of δ).

[16] This is consistent, for example, with the empirical evidence for co-authorship networks, where it is found that the probability of a particular scientist acquiring a new collaborator increases with the number of his past collaborators (Newman 2001b). Moreover, Ductor (2015) investigates the causal effect of co-authorship on individual productivity and provides evidence for the existence of peer effects. i.e., positive knowledge spillovers. He further shows that by simultaneously controlling for time invariant unobservable factors and for the potential endogeneity of co-authorship formation, co-authorship leads to a higher academic productivity. This result is robust and statistically significant.

[17] An agent i is indifferent between linking to j and k when j and k have the same degree. It does not matter with whom j and k are connected, even if j and k might share the same neighbor. In the network formation process I consider the probability of overlapping neighborhoods is small when the network becomes large. The model by Jackson and Rogers (2007) has the same feature, and it has been used for their mean field analysis. Hence, Assumption (LD) is not very restrictive.

[18] See also the examples discussed in supplementary Appendix E.

[19] Such a correlation between the degree variance and welfare has also been identified in R&D collaboration networks (cf., e.g., Westbrock 2010). Further examples from the literature with this feature can be found in supplementary Appendix E.

2.2 *The network formation process*

In this section the formation of the network is introduced. We consider a discrete time, nonstationary Markov chain $(G_t = \langle \mathcal{N}_t, \mathcal{E}_t \rangle)_{t\in\{1,2,\dots,T\}}$ for some $T \in \mathbb{N} \cup \{\infty\}$, defining a nested sequence of graphs $G_1 \subset G_2 \subset \cdots \subset G_T \in \mathcal{G}(T)$ in which each network G_t is obtained from the predecessor G_{t-1} by the addition of an agent and a specified number $m \geq 1$ of links emanating from that agent. Each network G_t is a random variable adapted to the filtration $\mathcal{F}_t = \sigma(\{G_s : 1 \leq s \leq t\})$. The probability measure $\mathbb{P}(\cdot|\mathcal{F}_{t-1}) : \mathcal{F}_t \to [0, 1]$ is denoted as $\mathbb{P}_t$. Expected values with respect to $\mathbb{P}_t$ are similarly denoted by $\mathbb{E}_t[\cdot|\mathcal{F}_{t-1}]$. Agents are labeled by their date of birth, so that t is the label of the agent entering the network at time t of the process.

We will need to agree on a given initial condition so that the network formation dynamics is well defined. I choose as the initial network the graph $G_1 \equiv K_{m+1}$, i.e., the complete graph on $m + 1$ agents in which all agents are bilaterally connected by m directed links (one in each direction) (cf. Jackson and Rogers 2007).[20]

Process time $t \in [T] \equiv \{1, 2, \dots, T\}$ divides the population of agents into a countable set in $\mathbb{N}$ of *active* and *passive* agents. These two sets are denoted, respectively, by $\mathcal{A}_t$ and $\mathcal{P}_t$. Passive agents have already entered the network and do not make any decisions in subsequent stages of the network formation process. At any date t the agent with label t, and only this agent, becomes active and considers forming a set of links. Once his decision has been made he joins the pool of passive agents. The initial composition of the population in active and passive agents is given by $\mathcal{P}_{m+1} = \{1, 2, \dots, m + 1\}$ and $\mathcal{A}_{m+1} = [T] \setminus \mathcal{P}_{m+1}$. Each graph G_t has exactly $|\mathcal{N}_t| = t$ (passive) vertices and $|\mathcal{E}_t| = e(G_t) = mt$ edges. It is formed from G_{t-1} by adding one agent with the label $t > m + 1$ and m edges from t to some passive agents $i \in \mathcal{P}_{t-1}$. Hence, every passive agent has constant out-degree equal to m, and thus we identify the in-degree simply by the degree of a passive agent via the identity $d_{G_t}(i) = d^-_{G_t}(i) + m$ for all agents $i \in \mathcal{P}_t$.

Before creating links, an entering agent t must make an observation of the prevailing network G_{t-1} and identify a set of agents to whom he can form links. We call this set the (observed) sample $\mathcal{S}_t \subseteq \mathcal{P}_{t-1}$. The sample $\mathcal{S}_t$ is obtained by selecting $n_s \geq 1$ passive agents in $\mathcal{P}_{t-1}$ uniformly at random (without replacement) and forming the union of these agents and their out-neighbors. We call n_s the *observation radius*. Note that an agent $j \in \mathcal{P}_{t-1}$ can enter the sample $\mathcal{S}_t$ either by being directly observed by the entrant t or by being observed indirectly as the neighbor of a directly observed agent $i \in \mathcal{P}_{t-1}$. This network sampling procedure is also known as *unlabeled star sampling* (Frank 1977, Kolaczyk 2009).[21] An illustration is shown in Figure 2. Further note that we assume that link formation follows a sampling procedure without replacement. Were we to allow for sampling with replacement, multiple links could be created to the same agent.

[20]The initial network does not matter as long as it is small, compared to the final network and the noise is large. However, when the noise term is small, and entering agents connect to the incumbent agent exclusively with the highest degree, then the agent who has the highest degree initially will attract most of the links most of the time.

[21]Instead of assuming that an agent can observe the degrees of the neighbors of the n_s sampled nodes, we could assume that he has to form beliefs about their degrees. Because this would complicate our analysis and introduce additional assumptions about how these beliefs are computed, we follow the statistical sampling literature and assume that the degrees are actually observable.

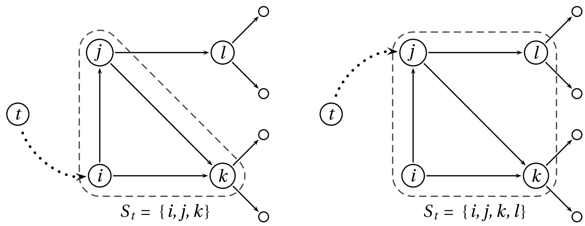


FIGURE 2. Illustration of the network sampling procedure. Left panel: In the first draw, the entering agent t observes agent i and its out-neighbors j, k. The observed sample is $\mathcal{S}_t = \{i, j, k\}$. Right panel: In the second draw, agent t observes also agent j and the out-neighborhood $\{k, l\}$ of j. The observed sample is then $\mathcal{S}_t = \{i, j, k, l\}$.

If the observed sample $\mathcal{S}_t$ constitutes only a small fraction of the passive agents P_{t-1} in the network G_{t-1}, we speak of link formation with *local information*. Local information is also a key ingredient to the model of Jackson and Rogers (2007),[22] and has been documented in various empirical studies (see, e.g., Friedkin 1983).

Given the observed sample $\mathcal{S}_t$, the entrant t must make a decision to whom he wants to create a link in $\mathcal{S}_t$. I assume that this decision is made in a myopic way.[23] I assume that an entrant t chooses to link to the incumbent agent $j \in \mathcal{S}_t$ that maximizes the value of his link incentive function plus a random element (cf. Snijders 2001, Snijders et al. 2010)

$$f_t^{\delta}(G_{t-1}, j) + \varepsilon_{tj}. \tag{1}$$

The term ε_{tj} is an exogenous random variable, indicating the part of the agent's preference that is not represented by the systematic component $f_t^{\delta}(G, j)$. This includes, for example, exogenous matching effects between characteristics of agents i and j that do not depend on the network structure G. We assume that the random variables ε_{tj} are independent and identically distributed for all t, j. When these exogenous matching effects are weak and $\delta \to 0$, (1) and Assumption (LD) introduce a preferential attachment mechanism to agents with a larger number of connections. In this case, agents who have a larger number of social ties are viewed as better sources for knowledge spillovers than agents with only a few neighbors (Galeotti et al. 2010).

More formally, we can give the following definition of the network formation process.

DEFINITION 1. For a fixed $T \in \mathbb{N} \cup \{\infty\}$ we define a network formation process $(G_t)_{t\in[T]}$, $[T] \equiv \{1, 2, \ldots, T\}$, as follows. Given the initial graph $G_1 = \cdots = G_{m+1} = K_{m+1}$, for all $t > m + 1$, the graph G_t is obtained from G_{t-1} by applying the following steps.

[22] See also McBride (2006) and Galeotti et al. (2010) for further examples.

[23] With this I mean that an agent t only considers the network G_{t-1} as source of information for his decision. He does not estimate the possible impact his linking decision at time t (which is an irreversible act) has on the future evolution of his personal utility level. For an alternative approach, see, e.g., Dutta et al. (2005).

Growth: Given $\mathcal{P}_1$ and $\mathcal{A}_1$, for all $t \geq 2$ the agent sets in period t are given by $\mathcal{P}_t = \mathcal{P}_{t-1} \cup \{t\}$ and $\mathcal{A}_t = \mathcal{A}_{t-1} \setminus \{t\}$, respectively.

Network sampling: Agent t observes a sample $\mathcal{S}_t \subseteq \mathcal{P}_{t-1}$. The sample $\mathcal{S}_t$ is constructed by selecting $n_s \geq 1$ agents $i \in \mathcal{P}_{t-1}$ uniformly at random without replacement and adding i as well as the out-neighbors $\mathcal{N}^+_{G_{t-1}}(i)$ of i to $\mathcal{S}_t$.

Link creation: Given the sample $\mathcal{S}_t$, agent t creates $m \geq 1$ links to agents in $\mathcal{S}_t$ without replacement. For each link, agent t chooses the $j \in \mathcal{S}_t$ that maximizes $f_t^\delta(G_{t-1}, j) + \varepsilon_{tj}$.

We next define the *attachment kernel* as the probability that an agent $j \in \mathcal{P}_{t-1}$ receives a link from the entrant. It can be written as $K_t^\beta(j|G_{t-1}) = \sum_{\mathcal{S}_t \subseteq \mathcal{P}_{t-1}} K_t^\beta(j|\mathcal{S}_t, G_{t-1})\mathbb{P}_t(\mathcal{S}_t|G_{t-1})$, where $K_t^\beta(j|\mathcal{S}_t, G_{t-1})$ is the probability, conditional on the sample $\mathcal{S}_t$ and the prevailing network G_{t-1}, that an agent j receives a link after the m draws (without replacement) by the entrant, and β is a parameter related to the distribution of the additive error term ε_{tj} from (1) (see below).[24] Since the entrant forms links to the incumbent agents that maximize his link incentive function plus a random element, so as to compute $K_t^\beta(j|\mathcal{S}_t, G_{t-1})$ we need to consider the cases where agent j has the highest value among all agents in the sample (or the second highest and so on). This means that we need to compute the probability that $f_t^\delta(G_{t-1}, j) + \varepsilon_{tj} \geq f_t^\delta(G_{t-1}, k) + \varepsilon_{tk}$ for all $k \in \mathcal{S}_t$. Assuming that the exogenous random terms ε_{tj} are identically and independently type I extreme value distributed (or Gumbel distributed) with parameter η,[25] the probability that an entering agent t chooses the passive agent $j \in \mathcal{S}_t$ for creating the link tj (in the first of the m draws of link creation) follows a multinomial logit distribution given by (cf. Anderson et al. 1992)

$$\mathbb{P}_t\Big(f_t^\delta(G_{t-1}, j) + \varepsilon_{tj} = \max_{k \in \mathcal{S}_t} f_t^\delta(G_{t-1}, k) + \varepsilon_{tk}\Big) \approx \frac{e^{\beta d_{G_{t-1}}(j)}}{\sum_{k \in \mathcal{S}_t} e^{\beta d_{G_{t-1}}(k)}}, \tag{2}$$

where we have applied condition (LD) for the link incentive function $f_t^\delta(G_{t-1}, \cdot)$, dropped terms of $o(\delta^b)$, and denoted $\beta \equiv \eta\delta^b$. Knowledge of the selection probability in Equation (2) will allow us to analyze the networks generated by the stochastic process introduced in Definition 1.

As I will show in the following sections, this stochastic process gives rise to different network topologies, depending on the extent of the noise ε_{tj}, as measured by the scaling parameter β and the observation radius, which depends on n_s. Small values of n_s (*local information*) refer to a local network formation process in which entering agents have

[24] A more detailed derivation can be found in supplementary Appendix C.

[25] This assumption is commonly made in random utility models in econometrics (see, e.g., McFadden 1981). Alternative distributional assumptions on the error term are possible. See also footnote 64 in supplementary Appendix C. However, relaxing the assumption of identically distributed error terms would complicate significantly the analysis, both from a theoretical and an empirical point of view. For alternative models with heterogeneity, which allow for agents with different types, I refer to Bramoullé et al. (2012) and Vigier (2014).

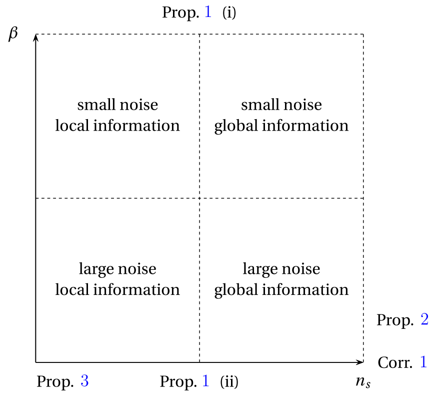


Figure 3. Illustration of the different parameter regions identified by the scaling parameter β and the observation radius n_s. The figure also indicates the parameter regions to which the results discussed in Section 3 refer. Proposition 1(i) deals with the case of $\beta = \infty$ and arbitrary values of n_s, while (ii) considers the case of $\beta = 0$. Both, Proposition 2 and Corollary 1 assume large values of n_s (such that $\mathcal{S}_t = \mathcal{P}_{t-1}$). While the first considers small but positive values of β, the latter assumes that $\beta = 0$. Proposition 3 deals with the case of $\beta = 0$ and small values of n_s.

only limited observability of the prevailing network, while large values of n_s (*global information*) constitute a network growth process in which entrants have full information of the network. It is important to note that we allow for n_s to grow with the network size, in particular, in the case of global information, where it is assumed that agents are able to observe the entire network.[26] Moreover, as β becomes large, the level of noise vanishes, and entrants choose to form links to the agents in the sample $\mathcal{S}_t$ that maximize their link incentive function. Conversely, when β tends to zero, then the noise term dominates and agents form links to the agents observed in $\mathcal{S}_t$ at random. These different parameter regions are indicated in Figure 3. In the following sections I give a more detailed account of the emerging networks depending on the level of noise scaled by β and the observation radius n_s.

3. Analysis of the network formation process

In this section I present a characterization of the different network architectures that may arise, in dependence of the noise in the attachment kernels and the observation radius. Section 3.1 analyzes the probability with which a class of strongly centralized networks emerges and shows that these networks are the unique outcome almost surely if the noise vanishes ($\beta \to \infty$), irrespective of the observation radius n_s. To gain further insight into the network topologies created by the model in the opposite case of large noise ($\beta \to 0$), Section 3.2 studies the degree distributions arising for both small and large observation radii. I show that networks tend to differ significantly for different observation radii when the exogenous noise term is large. Due to Assumption (DC) the varying degree of centralization across these different networks has important efficiency implications and we will study them in Section 4.

[26] See also Section 3.2 and supplementary Appendix D.1.

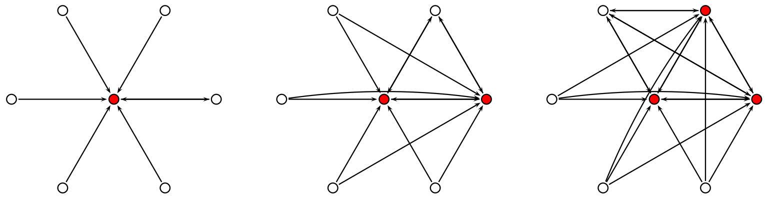


FIGURE 4. Illustration of the quasi-stars S_7^1, S_7^2, and S_7^3. Filled circles indicate the nodes with the highest degree.

3.1 *The emergence of quasi-stars*

Our first result, which is central for understanding the network formation process when the exogenous noise is small, is that it can produce a strongly centralized network topology, which we term a *quasi-star*.[27] A quasi-star S_n^m, $n \geq m+1$, with node set $[n] \equiv \{1, \dots, n\}$ is a directed graph in which all nodes in the set $[m+1]$ in S_n^m are bilaterally connected, while the nodes in the set $[n-1] \backslash [m+1]$ all maintain an outgoing link to the agents in the set $[m]$. Consequently, we have that $K_{m+1} \subseteq S_n^m$.[28] An illustration of various quasi-stars can be seen in Figure 4. With this definition we are able to state the following proposition.[29]

PROPOSITION 1. *Let $(G_t^\beta)_{t\in[T]}$ be a sequence of networks generated with observation radius $n_s^{(1)}$, and let $(H_t^\beta)_{t\in[T]}$ be a sequence of networks generated with observation radius $n_s^{(2)}$ such that $n_s^{(1)} > n_s^{(2)}$. Let $\Sigma_T^m \subset \mathcal{G}(T)$ be the isomorphism class[30] of quasi-stars of order $T > m+1$. Then the following statements hold:*

(i) *In the limit of vanishing noise, we have that* $\lim_{\beta\to\infty} \mathbb{P}(H_T^\beta \in \Sigma_T^m) = \lim_{\beta\to\infty} \mathbb{P}(G_T^\beta \in \Sigma_T^m) = 1$.

(ii) *In the limit of strong noise, we have that* $\lim_{\beta\to 0} \mathbb{P}(H_T^\beta \in \Sigma_T^m) > \lim_{\beta\to 0} \mathbb{P}(G_T^\beta \in \Sigma_T^m) > 0$.

Proposition 1 shows that in the limit of vanishing noise ($\beta \to \infty$), the networks generated are quasi-stars, irrespective of the observation radius n_s. However, as the level of noise becomes large ($\beta \to 0$), the probability of obtaining a quasi-star is higher, the smaller is n_s. The intuition for this result is that with increasing n_s the probability to

[27]Some authors also refer to this type of graphs as *core–periphery* networks (cf. Borgatti and Everett 2000, Galeotti and Goyal 2010).

[28]The complement $\bar{S}_n^m$ of a quasi-star S_n^m is the graph obtained from the complete graph K_d with d nodes and a subset of $n-d$ disconnected nodes by adding $n-d$ links connecting one node in K_d to each of the $n-d$ disconnected nodes. This graph falls into the class of *interlinked stars* introduced by Goyal and Joshi (2006) and the *nested split graphs* analyzed in König et al. (2012), König et al. (2009).

[29]Note that the essential assumption to obtain the result of Proposition 1 is the linear differences (LD), so that incumbent agents with a higher degree are the ones that will be selected to receive a link by the entrant. Hence, one could also allow for different distributional assumptions on the error term in (2) as long as (LD) is satisfied.

[30]See Appendix A.

sample the center(s) of the quasi-star decreases, and so does the probability to link to them and to further grow the quasi-star.

In the presence of noise, the set of networks generated by the model is much richer than the class of quasi-stars. So as to analyze these networks, we study in Section 3.2 the degree distribution in the case of large noise and in Section 5 we analyze higher order correlations.

3.2 *Large noise limit and the distributions of degree*

In this section we analyze the asymptotic degree distribution for large times t, when the level of noise is large (or small values of β, respectively). For this purpose, let us introduce some notation. For all $t \geq 1$ we denote by $N_t(k) \equiv \sum_{i=0}^{t} \mathbb{1}_k(d^-_{G_t}(i))$ the number of nodes in the graph G_t with in-degree k. The relative frequency of nodes with in-degree k is accordingly defined as $P_t(k) \equiv (1/t)N_t(k)$ for all $t \geq 1$. The sequence $\{P_t(k)\}_{k\in\mathbb{Z}_+}$ is called the (empirical) *in-degree distribution*. Throughout this section I assume that there are no hubs in the network, that is, I assume that $d^-_{G_t}(i) = o_p(t)$ for all $i \in \mathcal{P}_t$.[31]

We first analyze the case of the observation radius n_s being large enough, such that $\mathcal{S}_t = \mathcal{P}_{t-1}$, i.e., an entering agent can observe the entire population of incumbent agents (see supplementary Appendix D.1). When $\mathcal{S}_t = \mathcal{P}_{t-1}$, we have that $K_t^\beta(j|\mathcal{S}_t, G_{t-1}) = K_t^\beta(j|G_{t-1})$ for all $j \in \mathcal{P}_{t-1}$. The entrant t forms links by sampling m agents without replacement from $\mathcal{P}_{t-1}$. The probability that an agent j with in-degree $d^-_{G_{t-1}}(j)$ receives a link in the $(k+1)$st draw, given that k agents have received a link in the previous draws, $1 \leq k \leq m$, is given by $((1+\beta d^-_{G_{t-1}}(j))/((1+\beta m)t))(1+O(1/t))$ (see supplementary Appendix D.2). It then follows that the probability that an agent $j \in \mathcal{P}_{t-1}$ receives a link by the entrant t is given by[32]

$$K_t^\beta(j|G_{t-1}) \approx \frac{m}{1+\beta m}\frac{1+\beta d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right). \tag{3}$$

Having derived the attachment kernel, we are now able to obtain the asymptotic degree distribution in the following proposition. The proof of the proposition using the attachment kernel in (3) can be found in Appendix B.2.

Proposition 2. *Fix $\epsilon > 0$ small and let $\beta \in (0, \epsilon)$, $m \geq 1$. Assume that $d_{G_{t-1}}(j) = o_p(t)$ for all $j \in \mathcal{P}_{t-1}$. Consider the sequence of in-degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an*

[31] In the large noise limit we have a linear attachment kernel, that is, the probability that a node with in-degree k at time t receives a link is linear in k, and can be written as $(\beta + k)/t$ with $\beta \geq 0$. Móri (2005) has shown that the maximum degree in growing network models with a linear attachment kernel grows as $O(t^{1/(2+\beta)})$, which is sublinear, and hence $o(t)$.

[32] We have that

$$K_t^\beta(j|G_{t-1}) \approx 1 - \left(1 - \frac{1+\beta d_{G_{t-1}}(j)}{(1+\beta m)t}\right)^m + o\left(\frac{1}{t}\right) = 1 - \left(1 - m\frac{1+\beta d_{G_{t-1}}(j)}{(1+\beta m)t}\right) + o\left(\frac{1}{t}\right)$$
$$= \frac{m}{1+\beta m}\frac{1+\beta d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right).$$

indefinite iteration of the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ *assuming that* $\mathcal{S}_t = \mathcal{P}_{t-1}$ *for every* $t > m+1$. *Then* $P_t(k) \to P^\beta(k)$, *almost surely, where*

$$P^\beta(k) = (1+\beta k)^{-(2+1/(m\beta))}\left(1+O\left(\frac{1}{k}\right)\right) \tag{4}$$

for all $k \geq 0$ *as* $t \to \infty$.

Thus, Proposition 2 shows that in the limit of large noise and a large observation radius we obtain networks with a degree distribution that decays as a power law with exponent $2+1/(m\beta)$ for large degrees. This heavy tailed distribution indicates a highly uneven distribution of links, where old nodes typically have a larger degree as they have more time to accumulate links and thus become more likely to receive a link by the entrant. The tail flattens with increasing m, making high degree agents more likely as entering agents are forming more links. Note, however, that the power-law decay does not hold for small degrees. The degree distribution of (4) and a typical distribution obtained from a numerical simulation of the network formation process are shown in Figure 5.

The smaller is the number of links m created by an entrant and the stronger is the exogenous noise (the smaller β), the higher is the decay in the power-law tail of the distribution, making high degree agents less likely and reducing inequality. In the extreme case that we assume “strong noise,” corresponding to the situation with $\beta = 0$, we then obtain a process of uniform attachment (cf. Bollobás et al. 2001).

Corollary 1. *In the network formation process* $(G_t^\beta)_{t\in\mathbb{N}}$, *assuming that* $\mathcal{S}_t = \mathcal{P}_{t-1}$ *for every* $t > m+1$ *and* $\beta = 0$, *the agents perform a uniform attachment process whose degree distribution is given by*

$$P^0(k) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^k, \tag{5}$$

which is a geometric distribution with parameter $m/(m+1)$ *for all* $k \geq 0$.

When $\mathcal{S}_t$ does not encompass all agents in $\mathcal{P}_{t-1}$, then our analysis becomes more complicated. We therefore restrict our discussion to the case of strong noise when $\beta = 0$. In this case we have that the attachment kernel (which gives the probability that j receives a link from the entering agent given that j is in the sample $\mathcal{S}_t$) is

$$K_t^0(j|\mathcal{S}_t, G_{t-1}) = \frac{m}{|\mathcal{S}_t|}\mathbb{1}_{\mathcal{S}_t}(j).$$

The sample size is bounded by $|\mathcal{S}_t| \leq n_s(m+1)$. If no agent enters the sample more than once, then equality holds. The sample $\mathcal{S}_t$ is constructed by selecting n_s nodes from $\mathcal{P}_{t-1}$ without replacement, and forming the union of these nodes and their out-neighbors. Assuming that $n_s = o(t)$ and $d_{G_{t-1}}(j) = o_p(t)$, the probability that a node is entering $\mathcal{S}_t$ more than once is of $o(t)$ and thus

$$\frac{1}{|\mathcal{S}_t|} = \frac{1}{n_s(m+1)} + o_p\left(\frac{1}{t}\right). \tag{6}$$

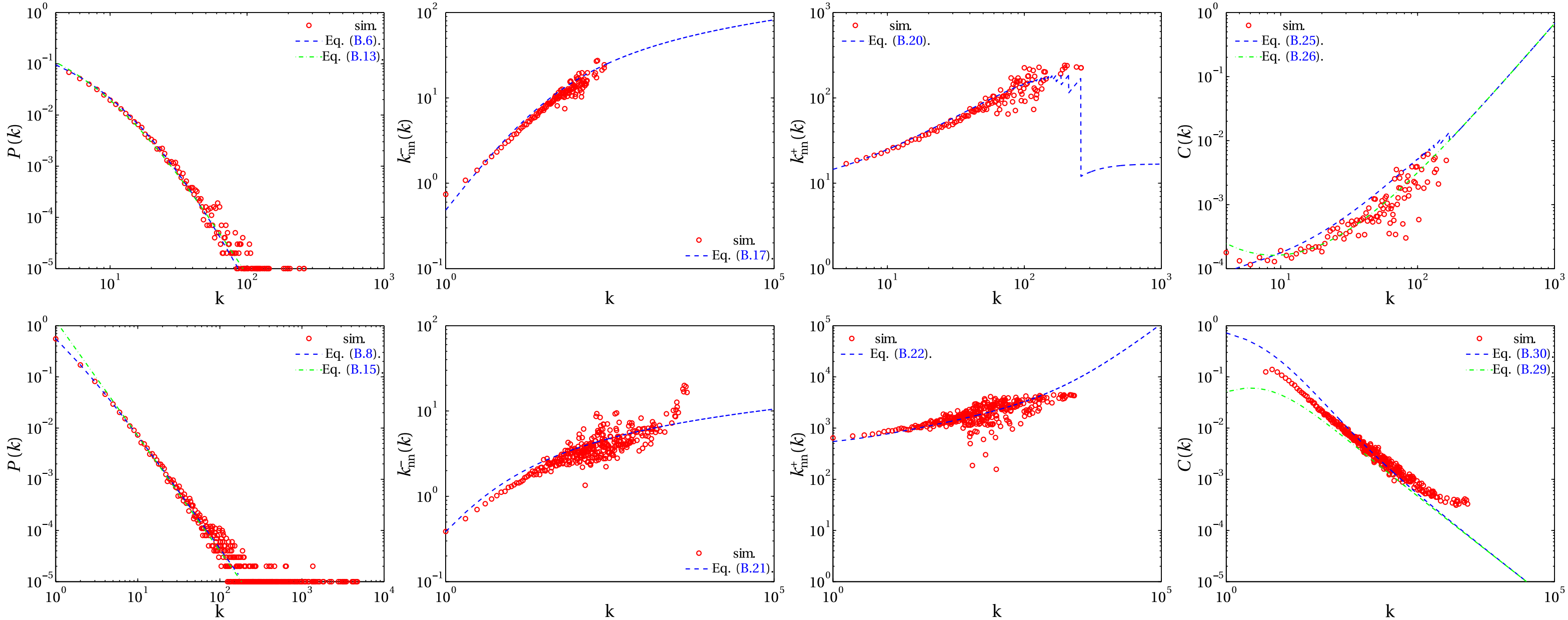


Figure 5. Top row: Comparison of the simulation results with the theoretical predictions for $T = 10^5$, $\mathcal{S}_t = \mathcal{P}_{t-1}$, and $m = 4$ with $\beta = 0.1$ under the linear approximation to the attachment kernel. Bottom row: Comparison of the simulation results for $T = 10^5$ and $n_s = m = 4$ ($\beta = 0$) with the theoretical predictions. The exact expressions for the different distributions can be found in the proofs in Appendix B.

The unconditional probability that an agent $j \in \mathcal{P}_{t-1}$ receives a link by the entrant t is then given by $K_t^0(j|G_{t-1}) = (1/(n_s(m+1)))\mathbb{P}(j \in \mathcal{S}_t|G_{t-1}) + o(1/t)$. If the degree of node j is small compared to the network size t, i.e., $d_{G_{t-1}}(j) = o_p(t)$, and the observation radius is small such that $n_s = o(t)$, then $\mathbb{P}(j \in \mathcal{S}_t|G_{t-1}) = n_s(1 + d_{G_{t-1}}(j))/t + o(1/t)$ and we obtain

$$K_t^0(j|G_{t-1}) = \frac{1}{1+m}\frac{1+d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right). \tag{7}$$

We then can state the following result for the asymptotic degree distribution when the observation radius is small. The proof, which is based on the attachment kernel in (7), can be found in Appendix B.2.

Proposition 3. *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ with a small observation radius $n_s = o(t)$. Assume that $\beta = 0$ and $d_{G_{t-1}}(j) = o_p(t)$ for all $j \in \mathcal{P}_{t-1}$. Then we have that $P_t(k) \to P(k)$, almost surely, where*

$$P(k) = k^{-(2+1/m)}\left(1 + O\left(\frac{1}{k}\right)\right) \tag{8}$$

for all $k \geq 0$ as $t \to \infty$.

Equation (8) is a power law decaying with an exponent $2 + 1/m$. A comparison with numerical simulations can be found in Figure 5. Compared to the power-law behavior in (4) obtained for a large observation radius, we find that the degree distribution in the case of a small observation radius has fatter tails, making high degree agents more likely and indicating a more unequal organization of the network. This is due to the fact that agents with a high degree can be found in a larger number of neighborhoods when entrants form the sample $\mathcal{S}_t$ and thus are more likely to receive a link. Also, when entrants form more links (by increasing m), the probability of agents with a high degree increases (which can be seen from a smaller exponent of the power-law decay).

Observe that the degree distribution in (8) does not depend on the number n_s of samples taken by the entering node. The reason is that two effects on the probability to receive a link of an incumbent cancel each other: On one hand, a larger value of n_s makes it more likely that an agent enters the sample $\mathcal{S}_t$ and, hence, increases the probability that he receives a link. On the other hand, a higher value of n_s also increases the sample size $|\mathcal{S}_t|$ and thus decreases the probability that he is selected by the entrant to receive a link.

The results obtained in this section show that when agents have global information, the presence of strong noise ($\beta \to 0$) induces networks with a smaller degree variance (following from the geometric distribution of Corollary 1) than when agents have only local information to form links (as implied by the power-law distribution of Proposition 3). However, as we have seen in part (i) of Proposition 1, in the absence of noise (as $\beta \to \infty$), the amount of information available to the agents when forming links does not matter, and the emerging network will be a quasi-star with a high degree variance.

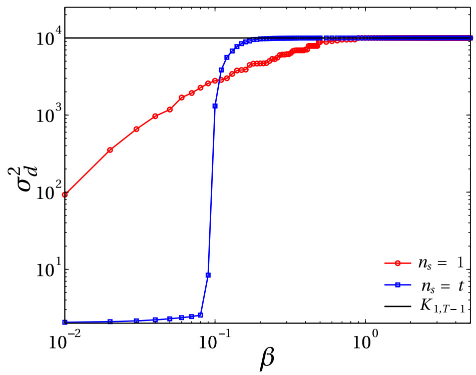


FIGURE 6. Degree variance σ_d^2 for local ($n_s = 1$) and global ($n_s = t$) search strategies for different values of β with $m = 1$, $T = 10^4$ nodes (averaged over 10 simulation runs). The degree variance of the star $K_{1,T-1}$ is given by $\sigma_d^2(K_{1,T-1}) = (T-1)(T-2)^2/T^2$.

These results are indicated in Figure 3. Hence, whether a limited observation radius impacts inequality in outcome networks depends crucially on the level of exogenous noise in agents' payoffs. The degree variance is also closely related to aggregate payoff and efficiency, and this will be discussed in more detail in the next section.

4. EFFICIENCY

Since we have computed the degree distribution in Section 3 for different values of the observation radius n_s, by virtue of Assumption (DC) we can readily state the following efficiency result.

PROPOSITION 4. *Consider the sequence of networks $(G_t^\beta)_{t\in[T]}$, generated with an observation radius $n_s^{(1)}$ large such that $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t \geq m+2$, and $(H_t^\beta)_{t\in[T]}$ with a small observation radius $n_s^{(2)} = o(t)$, and assume that $d_{H_t}(i) = o_p(t)$ for all $i \in \mathcal{P}_t$ as t becomes large. Let $\Pi(G_T^\beta, \delta)$ and $\Pi(H_T^\beta, \delta)$ be the aggregate payoff under G_T^β, respectively, H_T^β, after T iterations. Then, almost surely, the following statements hold:*

(i) For $\beta \to \infty$ we have $\Pi(H_T^\beta, \delta) = \Pi(G_T^\beta, \delta) = \Pi(\Sigma_T^m, \delta)$, where $\Sigma_T^m \subset \mathcal{G}(T)$ is the isomorphism class of quasi-stars of order T.

(ii) In the limit of large T, we have for $\beta \to 0$ that $\Pi(H_T^\beta, \delta) > \Pi(G_T^\beta, \delta)$.

A comparison of the degree variance σ_d^2 for different observation radii n_s (local vs. global) obtained by means of numerical simulations for $T = 10^4$ agents with different values of β can be seen in Figure 6. The figure shows that aggregate payoff is higher for G_T^β (global information) if β is high enough; however, the opposite holds for small values of β, where aggregate payoff is higher for H_T^β (local information).

Proposition 4 and Figure 6 show a major difference between the model considered here and the one by Jackson and Rogers (2007) (apart from differences in the sampling technology). In Jackson and Rogers (2007) a higher ratio of (local) neighborhood-based

n_s	$k^{\pm}_{\text{nn}}(k)$	$C(k)$
Global information	$k^{-}_{\text{nn}}(k) = O(\ln(k))$ $k^{+}_{\text{nn}}(k) = O(t^{(\beta m)/(1+\beta m)})$	$C(k) = O(t^{-2/(1+\beta m)} \cdot k^{2(1/(\beta m)-1)})$
Local information	$k^{-}_{\text{nn}}(k) = O(\ln(k))$ $k^{+}_{\text{nn}}(k) = O(t^{(m-1)/(m+1)} \cdot k^{1/m})$	$C(k) = O(\frac{1}{k})$

Table 1. Asymptotic behavior of the average nearest neighbor connectivity, $k^{\pm}_{\text{nn}}(k)$, and the clustering degree distribution, $C(k)$, in the large noise limit summarizing the results of Propositions 5, 6, 7, and 8 in Section 5.

linking to (global) random-based linking is always increasing average payoff as long as payoff is a convex function of the degree.[33] However, here we find that this does not hold in general when exogenous effects are taken into account, where this relationship might be reversed.[34] Also, when the marginal payoff of agents is increasing in the degree (and there is no exogenous noise), then differently to the welfare results obtained in Jackson and Rogers (2007), whether links are formed locally or globally has no impact on average payoffs and efficiency. Thus, the introduction of noise into decisionmaking in a network-based meeting process matters significantly for efficiency results.

5. Large noise limit and higher order statistics

In the following sections I analyze correlations between an agent and his neighbors. Such correlations are not only interesting as they help us to understand the behavior of our model for different parameter values, but also to compare it with correlations observed in real-world networks.

In Section 5.1 we first investigate the average in-degree of the in- and out-neighbors of a node with in-degree k, denoted by the average nearest in-neighbor connectivity $k^{-}_{\text{nn}}(k)$ and the average nearest out-neighbor connectivity $k^{+}_{\text{nn}}(k)$ (Pastor-Satorras et al. 2001). Next, in Section 5.2, we analyze the fraction of connected neighbors of a node with degree k (in the closure of the network), referred to the clustering coefficient $C(k)$ (Watts and Strogatz 1998). The results of these sections are anticipated in Table 1.

Note that, so as to derive the functional forms of these statistics, I consider a continuous representation of our discrete dynamical system, the so-called *continuum approximation,* in which both time t and degree k are treated as continuous variables in $\mathbb{R}_+$.[35] Using the continuum approximation, we can then apply the rate equation approach outlined in Barrat and Pastor-Satorras (2005) to compute higher order correlations in the network.

[33] See Corollary 1 and footnote 51 in Jackson and Rogers (2007).

[34] Observe, however, that a reversal of the assumption on the degree concentration (DC) would also reverse the inequality in part (ii) of Proposition 4 and this conclusion would no longer hold.

[35] This is an approximation that has been shown to be accurate in various growing network models as $T \to \infty$ (Dorogovtsev and Mendes 2013, p. 117). See Appendix B.4 for more discussion.

5.1 *Average nearest neighbor connectivity*

In this section we analyze two vertex degree correlations, i.e., correlations between the degree of an agent and his neighbors' degrees. Let $P(k'|k)$ denote the probability that a node of in-degree k has an in-neighbor with in-degree k'. The average in-degree of in-neighbors of nodes with in-degree k can then be written as $k^-_{\text{nn}}(k) = \int_0^\infty k'P(k'|k)\,dk'$ (cf. Pastor-Satorras et al. 2001). In the case that $k^-_{\text{nn}}(k)$ is an increasing function of k, we speak of *assortative mixing*, while for $k^-_{\text{nn}}(k)$ decreasing with k, we have *dissortative mixing* (Newman 2002). Similarly, the average nearest out-neighbor connectivity $k^+_{\text{nn}}(k)$ can be defined. We now derive these quantities for different observation radii.

In the case of global information (when the observation radius n_s is large) and small β (large noise) we obtain the following proposition.

Proposition 5. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ with $\mathcal{S}_t = \mathcal{P}_{t-1}$. Then under the continuum approximation in the limit $\beta \to 0$, the average nearest in-neighbor in-degree of an agent with in-degree k grows logarithmically with k and is independent of t, $k^-_{\text{nn}}(k) = O(\ln(k))$, and the average nearest neighbor out-degree becomes independent of k and grows with the network sizes as $k^+_{\text{nn}}(k) = O(t^{(\beta m)/(1+\beta m)})$ as $t \to \infty$.*

Similarly, we can compute the nearest neighbor connectivities under local information (when the observation radius n_s is small), assuming strong noise ($\beta = 0$).

Proposition 6. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ with n_s small. If $\beta = 0$, then under the continuum approximation the average nearest in-neighbor in-degree of an agent with in-degree k grows logarithmically with k, that is, $k^-_{\text{nn}}(k) = O(\ln(k))$, and the nearest out-neighbor degree grows as $k^+_{\text{nn}}(k) = O(t^{(m-1)/(m+1)} \cdot k^{1/m})$ as $t \to \infty$.*

In Figure 5 a comparison of numerical simulations with the theoretical predictions of Propositions 5 and 6 are shown.

In both cases—local as well as global information (corresponding to Propositions 5 and 6, respectively)—we find that networks are characterized by positive degree correlations, or assortative mixing.[36] The intuition for this result derives from the observation that older agents form links to other old agents with high degrees, while younger agents are more likely to form links to agents with smaller degrees. This gives rise to an assortative trend in the average nearest out-neighbor degree $k^+_{\text{nn}}(k)$. This intuition carries over to the average nearest in-neighbor degree $k^-_{\text{nn}}(k)$, but the average degree of the in-neighbors of older nodes is much smaller, because in this case the in-neighbors include also a large number of younger nodes. Consequently, we observe that the assortative trend is much weaker in the case of the average nearest in-neighbor degree $k^-_{\text{nn}}(k)$ (growing only logarithmically with the degree k).

If we compute the average nearest neighbor degree in the closure $\overline{G}$ of G, then the average nearest neighbor degree $k_{\text{nn}}(k)$ (the sum of in- and out-neighbors' total degrees

[36] Note that in contrast to the model considered here, both the Erdös–Rényi random graph $G(n, p)$ with link density $p > p_c$, where p_c is the critical link probability below which the graph becomes disconnected, and the Barabási–Albert power-law graph are zero assortative (Van Mieghem 2011).

divided by the total degree) of older nodes is similar to the case of the average nearest in-neighbor degree; however, the average nearest neighbor degree of younger nodes is now higher because the average nearest neighbor degree includes not only the in-neighbors, but also the out-neighbors which tend to have higher degrees. Therefore, we expect to see a dissortative trend in the average nearest neighbor connectivity $k_{\mathrm{nn}}(k)$ in the closure $\overline{G}$. This intuition is confirmed by combining the results we have obtained for $k^+_{\mathrm{nn}}(k)$ and $k^-_{\mathrm{nn}}(k)$.[37]

As we will see in the next section, the similarities between local and global observability do not carry over to the case of three vertex correlations, where networks generated under local and global information produce starkly different results.

5.2 *Clustering degree correlations*

In this section I study three vertex degree correlations in the undirected network obtained from the closure $\overline{G}^\beta_t$ of the directed network $(G^\beta_t)_{t\in\mathbb{R}_+}$. The clustering coefficient $C(k)$ is defined as the probability that a vertex of degree k in $\overline{G}^\beta_t$ is connected to vertices with degrees k' and k'', and that these vertices are themselves connected, averaged over all k' and k'' (Watts and Strogatz 1998).[38] Note that in the case of $m = 1$ all networks will be trees, $\overline{G}^\beta_t \in \mathcal{T}([t])$, which are characterized by a vanishing clustering coefficient. Hence, we will consider only the case of $m > 1$ in this section.

Similarly to the case of two vertex degree correlations in the previous section, we can derive the clustering coefficient using a rate equation approach (Barrat and Pastor-Satorras 2005). With global information ($\mathcal{S}_t = \mathcal{P}_{t-1}$) and small β (strong noise) we can state the following proposition.

Proposition 7. *Consider the network formation process $(G^\beta_t)_{t\in\mathbb{R}_+}$ with $\mathcal{S}_t = \mathcal{P}_{t-1}$ and $m > 1$. Then under the continuum approximation in the limit $\beta \to 0$ the clustering coefficient of an agent with degree k is given by $C(k) = O(t^{-2/(1+m\beta)} \cdot k^{2(1/(m\beta)-1)})$ as $t \to \infty$.*

The clustering coefficient for $m = 4$ and $\beta = 0.1$ can be seen in Figure 5. It grows with k as a power law with exponent $2(1/(m\beta) - 1)$.[39] Moreover, we find that the clustering coefficient is decreasing with the network size as $t^{-2/(1+m\beta)}$. Hence, for large networks

[37] An increasing total nearest neighbor connectivity $k_{\mathrm{nn}}(k)$ can be obtained in two possible extensions of the model, considering undirected links (see Section 6.1), or heterogeneous linking opportunities (see Section 6.2).

[38] So as to compute the clustering coefficient as a function of the degree, in this section we proceed by first computing the clustering coefficient of a vertex s, born at time s. The clustering coefficient of vertex s is defined as the number of links between the neighbors of s divided by the total number of links that can exist between them (in the undirected closure of the graph), which is $k_s(k_s - 1)/2$ when k_s is the degree of vertex s (Watts and Strogatz 1998). Under the continuum approximation (see also footnote 35), there exists a continuous mapping between the time of birth, s, and the degree, k_s, of a vertex s. Hence, under the continuum approximation, knowing the clustering coefficient of a vertex born at time s tells us the clustering coefficient of a vertex with degree k_s (cf., e.g., Barrat and Pastor-Satorras 2005, Boguná and Pastor-Satorras 2003). See Appendix B.4.2 for a detailed derivation.

[39] We need only consider values of k such that $C(k)$ does not exceed its upper bound given by 1.

with a high clustering coefficient (such as the network of co-inventors; see Section 7), the assumption of global information seems to be at odds with the empirical observation.

When agents have only local information and $\beta = 0$ (strong noise), we obtain clustering degree correlations as given in the next proposition.

Proposition 8. *Consider the network formation process* $(G_t^{\beta})_{t\in\mathbb{R}_+}$ *with* $n_s = o(t)$ *small and assume that* $m > 1$. *If* $\beta = 0$, *then under the continuum approximation the clustering coefficient* $C(k)$ *of an agent with degree* k *is given by* $C(k) = O(1/k)$ *as* $t \to \infty$.

The proof of Proposition 8 relies on deriving upper lower and upper bounds for the clustering coefficient that asymptotically decay as $O(1/k)$.[40] These bounds for the clustering coefficient for $m = n_s = 4$ can be seen in Figure 5. The figure confirms the asymptotic decay of the clustering coefficient as a power law with exponent -1. Note that, in contrast to the results obtained in Proposition 7, the clustering coefficient in Proposition 8 does not vanish as the network becomes large. Moreover, the clustering coefficient shows a power-law decay that is a typical feature of all the empirical networks we consider (see Section 7), indicating that a limited observation radius is a general constraint in the creation of various real-world networks.

Comparing the results for global and local information, we find that networks generated under global information produce relatively low clustering and a positive degree clustering correlation. This is what one would expect from a global link formation process in which the formation of cliques is very unlikely, and becomes even more unlikely the later an agent enters. Hence, we find an increasing clustering degree correlation since older agents tend to have higher degrees. However, networks formed with local information tend to produce higher clustering and a negative clustering degree correlation (see also Figure 5). Local link formation favors the creation of links between neighboring agents, making the network highly clustered. Moreover, the large number of younger agents that connect to the older ones with higher degrees are less clustered and thus gives rise to a negative clustering degree correlation (as conjectured by Jackson and Rogers 2007).

Finally, while the results for different observation radii are fairly robust as long as the observation radius does not become too large, this does not hold for the average clustering coefficient, which is decreasing sharply with the observation radius n_s. As the number of links are distributed across a larger number of agents when n_s increases, the formation of triangles becomes less likely and, hence, the clustering coefficient declines.

6. Robustness analysis and extensions

6.1 *Undirected links*

An extension to the network formation process we have introduced in Definition 1 is to allow entering agents to observe not only the out-neighbors of incumbent agents (the ones to which these agents have formed links), but also their in-neighbors (the ones

[40] See (29) and (30) in Appendix B.4.2.

from which they have received links). The resulting network can then be viewed as an undirected graph. One can show that the distributions of the network statistics we have considered follow a similar behavior as in the case of directed links. The degree distribution exhibits a power-law decay $k^{-\alpha}$ with exponent $\alpha = 3 + 1/(m\beta)$ for a large observation radius and $\alpha = 3 + 1/m$ for a small observation radius. Note, however, that by introducing undirected links, the rigorous approach to derive the degree distributions for a small observation radius in Section 3.2 is no longer viable, because there is no straightforward way to compute the sample size $|\mathcal{S}_t|$. Instead, one has to resort to an approximation as $|\mathcal{S}_t| \approx n_s(\bar{d} + 1)$. The results obtained using this approximation are given in supplementary Appendix G.

6.2 *Heterogeneous linking opportunities*

We can introduce heterogeneity in the linking opportunities of entering agents by assuming that a fixed fraction $1 - p$, with $p \in (0, 1)$, of the population of agents does not form any links and remains passive throughout the evolution of the network. Moreover, one can also allow for a varying number of links to be created by each entrant following a certain distribution function with given mean $m \geq 1$. This extension is studied in the accompanying supplementary Appendix H.[41] We find degree distributions that follow a power-law decay $k^{-\alpha}$ with exponent $\alpha = 2 + 1/(\beta mp)$ for a large observation radius and $\alpha = 1 + (1 + m)/(mp)$ for a small observation radius. The main difference with respect to the basic model in Definition 1 is that this extension gives rise to a nontrivial component structure of the network, where the component size distribution exhibits a power-law decay. In the special case of $\beta = 0$ and $n_s = m = 1$ one can show that the distribution $P(s)$ of components of size s is identical for both large and small observation radii and decays as a power law with exponent $1 + 1/p$. Moreover, we find an assortative trend for the nearest neighbor connectivity (in the closure of the graph) when the observation radius n_s and p are small enough in the large noise limit ($\beta \to 0$). Note, however, that differently to Proposition 1, a value of $p < 1$ can lead to the emergence of multiple quasi-stars in the limit of vanishing noise ($\beta \to \infty$) when the observation radius is small, and an analytic characterization as in Proposition 1 becomes harder to obtain.

6.3 *From assortative to dissortative networks*

Finally, when combining the model with undirected links and heterogeneity in the number of links that entrants can form, a transition from an assortative to a dissortative network can be observed when varying the parameter β, which is related to the noise in the payoff function of the agents. Figure 7 shows two examples for the average nearest neighbor degree distribution $k_{\mathrm{nn}}(k)$, either for $\beta = 0$ (left panel) or for $\beta = 5$ (right panel). Recall that β is a parameter related to the level of noise in the payoff function of the agents and their decisions with whom to form a link. For $\beta = 0$ (large noise) we

[41] A related, more general setup is introduced in Vigier (2014), where agents' propensities to form a link follow a certain distribution that allows the incorporation of homophily, making more similar agents more likely to connect.

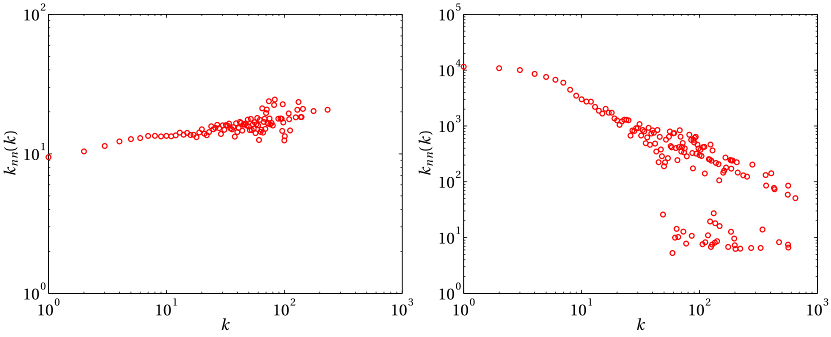


Figure 7. Left panel: The average nearest neighbor degree distribution $k_{\text{nn}}(k)$. The parameters used are $T = 50{,}000$, $n_s = 2$, $m = 4$, $p = 0.7$, and $\beta = 0$. Right panel: The average nearest neighbor degree distribution $k_{\text{nn}}(k)$ with the same parameters except for $\beta = 5$.

see that $k_{\text{nn}}(k)$ is increasing with the degree k, indicating an assortative network, while for $\beta = 5$ (weak noise) we find that $k_{\text{nn}}(k)$ is decreasing with the degree k, characterizing a dissortative network.[42] While dissortativity does not play any role in the empirical networks we consider in Section 7, dissortativity has been found, for example, in trade networks (see the working paper version König 2011).[43]

7. Empirical implications

To bring the model to data, I consider three different real-world collaboration networks in which knowledge diffusion and spillovers are an important source of knowledge generation and dissemination.

First, I analyze a network of inventors constructed from United States Patent and Trademark Office (USPTO) patent data in the year 2009.[44] I consider only patents in the drugs and medical sector with patent classification numbers 424 and 514 (see also the classification in Hall et al. (2001)). I focus on the drugs development sector due to the high collaboration intensity in this sector as well as for practical reasons, since for the size of the subsample corresponding to this sector our estimation process is feasible, while larger sample sizes would make the estimation of the model computationally difficult.[45] The network of co-inventors is constructed by creating a link between any pair of inventors that has appeared together on a patent. The resulting network is undirected. I use this network as a proxy for the social network of inventors, in which local knowledge spillovers take place.[46] This gives us a network with 27,492 nodes, an average degree of $\bar{d} = 3.51$, and a degree variance of $\sigma_d^2 = 30.03$ (with a coefficient of variation of

[42] Note that as Jackson and Rogers (2007) do not provide results for $k_{\text{nn}}(k)$ and they do not have a payoff function governing the decision with whom to form a link, this transition cannot be studied or observed in their setup.

[43] For a classification of assortative vs. dissortative networks, see Newman (2002).

[44] See Lai et al. (2009) for a more detailed description of the data.

[45] The statistics computed for this subsample of the original data set are similar to the full sample or other subsamples for different sectors.

[46] As noted by Fafchamps et al. (2006), in the context of scientific co-authorship networks, the (unobserved) social network of personal acquaintances has more links than the co-inventor network. However,

$c_v \equiv \sigma_d/\bar{d} = 0.94$). The distribution of degrees is highly skewed, following a power law for large degrees (see Figure 8). The network is highly clustered with an average clustering coefficient of $C = 0.64$ and a negative clustering-degree correlation (Figure 8, second column). Moreover, the network is assortative, with an assortativity coefficient of $\kappa = 0.28$ (Newman 2002).[47] The nearest neighbor average degree is monotonically increasing with degree (Figure 8, third column). The largest component consists of 12,060 nodes (which is 44% of all nodes).

Second, I consider the network of scientific co-authorships among physicists in the field of condensed matter physics in the year 2003 (Leskovec et al. 2007).[48] The co-author network is constructed from the e-print service arXiv and covers scientific collaborations between authors' papers submitted to the category of condensed matter physics.[49] If an author co-authored a paper with another, the network contains an undirected edge between them. The network constructed in this way contains 23,133 nodes. The average degree is $\bar{d} = 8.08$ and the degree variance is $\sigma_d^2 = 113.13$, and we obtain a coefficient of variation of $c_v = 1.32$. The degree distribution is heavy tailed (see Figure 8, first column). Similar to the network of inventors, the network of scientific co-authors in condensed matter physics is highly clustered with an average clustering coefficient of $C = 0.63$. Moreover, the clustering degree distribution exhibits a similar power-law decay for large degrees (see Figure 8, second column). Moreover, we obtain a positive assortativity coefficient of $\kappa = 0.13$. The size of the largest connected component is 21,363 encompassing 92% of all nodes. As in the inventor network we observe a power-law decay for the sizes of the small components (see Figure 8, fourth column).

Third, I consider the network of scientific co-authorships between economists as reported by the collaboration network service CollEc for authors registered in the RePEc author service in the year 2012.[50] When two authors claim the same paper in the RePEc digital library,[51] they are co-authors, and the relationship of co-authorship creates an undirected network between them. The network constructed in this way contains 24,721 nodes. The average degree is $\bar{d} = 5.83$, the degree variance is $\sigma_d^2 = 40.48$, and the coefficient of variation is $c_v \equiv \sigma_d/\bar{d} = 1.09$. The clustering coefficient is $C = 0.23$, which is lower by a factor of 3 than for the network of inventors or research collaborations among physicists. The assortativity coefficient is $\kappa = 0.19$, and the slope of the average nearest neighbor connectivity $k_{\mathrm{nn}}(k)$ is slightly smaller than in the other networks (see Figure 8, third column). We further find that the largest connected component encompasses all nodes in the network (see Figure 8, fourth column).

---

the acquaintance network includes the co-inventor network because it can reasonably be assumed that individuals who have appeared on a patent together know each other, and it can be used as a proxy for the network of acquaintances.

[47] The assortativity coefficient $\kappa \in [-1, 1]$ is essentially the Pearson correlation coefficient of degree between nodes that are connected. Positive values of κ indicate that nodes with similar degrees tend to be connected (and $k_{\mathrm{nn}}(k)$ is an increasing function of the degree k), while negative values indicate that nodes with different degrees tend to be connected (and $k_{\mathrm{nn}}(k)$ is a decreasing function of the degree k). See Newman (2002) and Pastor-Satorras et al. (2001) for further details.

[48] See http://snap.stanford.edu/data/ca-CondMat.html.

[49] See http://arxiv.org/.

[50] See http://collec.repec.org/.

[51] See http://repec.org/.

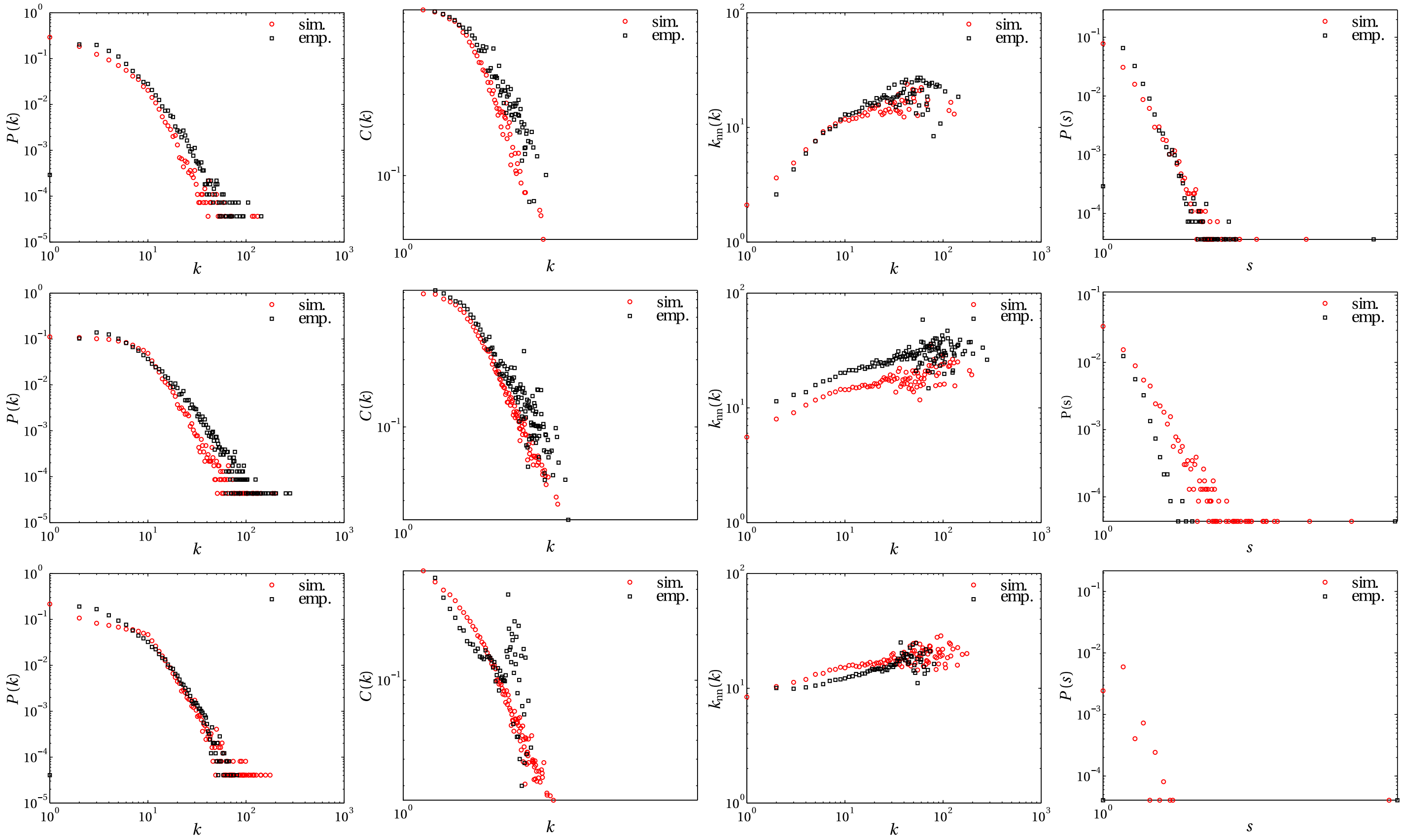


FIGURE 8. Empirical degree distribution $P(d)$ (first column), clustering-degree correlation $C(d)$ (second column), average nearest neighbor connectivity $k_{nn}(d)$ (third column), and component size distribution $P(s)$ (fourth column) constructed from (first row) USPTO patents on drugs (patent classes 424 and 514), (second row) co-authors in condensed matter physics from the arXiv data base, and (third row) the network of co-authors in economics from the CollEc data base (empirical data points indicated by □) and the corresponding distributions generated by the model (indicated by ◯).

To estimate the parameters of the model I follow the likelihood-free Markov chain Monte Carlo (LF-MCMC) algorithm suggested by Marjoram et al. (2003).[52,53] The details of this algorithm are outlined in supplementary Appendix F.[54] I analyze the model with undirected links, as it provides a better fit to the data, which has been discussed in Section 6.1. Moreover, I allow for heterogeneous linking probabilities, including the basic model when these probabilities are set to 1, as discussed in Section 6.2.

The estimated parameter values are shown in Table 2. Moreover, Figure 8 shows various distributions for the inventor network, and the networks of co-authors among physicists and economists. The comparison of observed and simulated distributions shown in Figure 8 indicates that the model can well reproduce the observed empirical networks for all the distributions that we consider.[55]

Comparing the estimated observation radius n_s for the inventor network and the network among physicists to that for the co-author network among economists in Table 2 shows that the number of observed agents by an entrant is larger for economists (with an observation radius n_s of 2 instead of 1).[56] Hence, this result indicates that economists tend to use a larger information set for their linking decisions than inventors or physicists. We also observe that the higher clustering coefficient of the inventor network as compared to the network of co-authors in economics is not due to a higher link density (the average degree in the first is 3.51, while in the latter it is 5.83), but can only be understood from a network formation process where agents use a different mix of local vs. global information as the basis for their decision with whom to form a link.

8. Conclusion

The current paper analyzes the growth of networks where an agent's linking incentives can be decomposed into a network-dependent part and an independent exogenous random term, referred to as noise. The network formation process sequentially adds agents to the network. Upon entry, each agent can sample a given number n_s (the observation radius) of existing agents in the network and observe these agents and their neighbors.

[52] Kolotilin (2013) discusses conditions for consistent estimation and identification using a generalized method of moments (GMM) approach of the related model by Cooper and Frieze (2003), and it would be desirable to use a similar approach here. However, because we do not have closed-form analytic solutions for the distributions of interest for all possible parameter configurations, we cannot proceed as in Kolotilin (2013), and instead need to resort to LF-MCMC methods. For this reason I do not address the issue of identification here, and from a more conservative point of view, one can also view this empirical section as a calibration exercise of the model's parameters.

[53] The same statistical method has been used in Ratmann et al. (2007). It is essentially a simulated method of moments (SMM) estimation procedure (cf. McFadden 1989, Pakes and Pollard 1989).

[54] See Sisson and Fan (2011) for an introduction to LF-MCMC, Robert and Casella (2004) for a general discussion of Markov chain Monte Carlo (MCMC) approaches, and Chib (2001) and Chernozhukov and Hong (2003) for applications of MCMC in econometrics.

[55] Note that, as can be seen from the top left panel in Figure 8, the estimated model for the network of inventors slightly underestimates the degree distribution, which leads to a slightly lower average degree than the one that can be observed in the data.

[56] The small variances of the estimates make clear that any two-sample Z-test for comparing two means would always reject the null that they are equal.

	USPTO Network					arXiv Network					CollEc Network				
	μ_θ	$\bar{\sigma}_\theta$	σ_θ	ι_θ	$p_\theta(n)$	μ_θ	$\bar{\sigma}_\theta$	σ_θ	ι_θ	$p_\theta(n)$	μ_θ	$\bar{\sigma}_\theta$	σ_θ	ι_θ	$p_\theta(n)$
n_s	1	0	0	3.3100	1	1	0	0	0.9994	1	2	0	0	0.9921	1
p	0.5703	0.0381	0.0069	758.1600	0.9636	0.7488	0.0454	0.0022	36.3550	0.9982	0.6089	0.0533	0.0117	961.2000	0.9325
m	4.6340	0.4838	0.0859	877.6300	0.9286	5.3148	0.4645	0.0244	40.0400	0.9923	4.7303	0.4439	0.0977	975.8500	0.9111
β	0.4758	0.3226	0.0592	548.8400	0.8367	0.0622	0.0780	0.0069	78.8640	0.7263	0.0023	0.0015	0.0003	996.2500	0.5212
T			27,495					23,133					24,721		
n			15,000					10,000					10,000		

[a]The notation μ_θ is the average and $\bar{\sigma}_\theta$ is the simulation standard deviation of the respective parameter, while σ_θ is the standard deviation calculated from batch means (of length 10) for each parameter $\theta \in \boldsymbol{\theta}$ (Chib 2001); ι_θ is the integrated autocorrelation time, which should be much smaller than the number n of iterations of the Markov chain (Sokal 1996).

[b]The notation $p_\theta(n)$ is the p-value of Geweke's spectral density diagnostic (converging in distribution to a standard normal random variable as $n \to \infty$) indicating the convergence of the chain (Geweke 1992, Brooks and Roberts 1998). The maximum number of iterations, n, has been chosen such that reasonably high values of $p_\theta(n)$ were obtained.

Table 2. Estimation of the model parameters $\theta \in \boldsymbol{\theta} = (n_s, p, m, \beta)$ for the network of inventors from USPTO patents, the network of co-authors in condensed matter physics from the arXiv data base, and the network of co-authors in economics from the CollEc data base. The table shows simulated averages of the parameters and their standard deviations,[a] after the chain has converged.[b]

The set of observed agents constitutes the sample $\mathcal{S}_t$. The entrant then forms links to the agents in $\mathcal{S}_t$ based on his linking incentives.

I analyze the emerging networks for different observation radii, n_s, and levels of noise. For small noise the observation radius does not matter and strongly centralized networks emerge. However, for large noise, a smaller observation radius generates networks with a larger degree variance and a higher aggregate payoff. Estimating the model for networks from co-inventors and co-authors in physics and economics, I find that the model can well reproduce the observed patterns. The estimation shows that local search for new collaboration partners is important in all the empirical networks considered, but that economists tend to search more broadly for new collaboration opportunities. This finding can guide future research on differences in communication culture, search costs and funding, and organization of research across different disciplines (cf. Stephan 2012, Schilling and Green 2011, Jones et al. 2008) by providing a simple model that can explain the empirical observations.

The paper could be extended along several directions. First, it would be interesting to extend the model by allowing both entering and incumbent agents to form links in a similar way (such as in Cooper and Frieze 2003). Moreover, the deletion of links is another important extension. Second, an extension of the analysis presented here could investigate further network measures and analyze additional network data sets. This could help to shed light on the generality of the patterns I have identified. Finally, the payoff functions considered in Section 2.1 typically assume that spillover effects (as measured by the parameter δ) are weak. An extension of the current paper could investigate the effect of stronger spillover effects on the emerging network structures and their impact on efficiency.

Appendix A: Basic definitions

The network is modeled as a *directed graph*, which is a pair $G \equiv \langle \mathcal{N}, \mathcal{E} \rangle$, where $\mathcal{N} \equiv \{1, \dots, n\}$ is a set of nodes (vertices) and $\mathcal{E} \subset \mathcal{N} \times \mathcal{N}$ is a set of edges (links). The set of all networks with n nodes is denoted by $\mathcal{G}(n)$. Similarly, the set of networks with n nodes and e edges (or links) is denoted by $\mathcal{G}(n, e)$. We identify every graph G with a *network*, and thus use these two terms interchangeably. We denote the *out-neighborhood* of a vertex i as the set of agents he can directly access, i.e., $\mathcal{N}_G^+(i) \equiv \{j \in \mathcal{N} | ij \in \mathcal{E}\}$. The *in-neighborhood* of i is conversely the set of agents who can access i directly, i.e., $\mathcal{N}_G^-(i) \equiv \{j \in \mathcal{N} | ji \in \mathcal{E}\}$. The *in-degree* of i is the cardinality of i's in-neighborhood set and is denoted as $d_G^-(i) \equiv |\mathcal{N}_G^-(i)|$. The *out-degree* of i is $d_G^+(i) \equiv |\mathcal{N}_G^+(i)|$. The (total) degree of i is $d_G(i) \equiv d_G^+(i) + d_G^-(i)$ and the total neighborhood is $\mathcal{N}_G(i) \equiv \mathcal{N}_G^+(i) \cup \mathcal{N}_G^-(i)$. The average degree of G is $\bar{d}_G \equiv (1/n) \sum_{i \in \mathcal{N}} d_G(i)$ and the degree variance is given by $\sigma_d^2(G) \equiv (1/n) \sum_{i \in \mathcal{N}} (d_G(i) - \bar{d}_G)^2$. Following Bala and Goyal (2000), the *closure* of a graph G, denoted by $\overline{G}$, is defined by the condition $ij \in \mathcal{E}(\overline{G}) \Leftrightarrow ij \in \mathcal{E}(G) \vee ji \in \mathcal{E}(G)$. The number of edges $e(G)$ in G satisfies $e(G) = \sum_{i \in \mathcal{N}} d_G^+(i) = \sum_{i \in \mathcal{N}} d_G^-(i)$, while the number of edges $e(\overline{G})$ in the closure $\overline{G}$ is given by $e(\overline{G}) = \frac{1}{2} \sum_{i \in \mathcal{N}} d_G(i)$. We denote by $G \oplus ij$ the network obtained by adding the link ij to $\mathcal{E}$. Similarly, $G \ominus ij$ is the network obtained from G by removing the link ij from $\mathcal{E}$. We call two graphs G and H isomorphic if an *isomorphism* exists between them. An isomorphism of graphs G and H is a bijection between the nodes of G and H such that $f: \mathcal{N}(G) \to \mathcal{N}(H)$, and any two nodes u and v of G are adjacent in G if and only if $f(u)$ and $f(v)$ are adjacent in H. Loosely speaking, G and H are isomorphic if we can rearrange a picture of G to match a picture of H, except for the nodes' labels. All graphs isomorphic to each other form an isomorphism class. For further graph theoretic terminology, see, e.g., West (2001) and Bollobás (1998).

Appendix B: Proofs

In this appendix the proofs of the propositions, corollaries, and lemmas stated in the paper are provided.

B.1 *Quasi-stars*

Proof of Proposition 1. We first give a proof for part (i) of the proposition. For each agent $j \in \mathcal{S}_t$, let the best response of the entrant t be the set-valued map $\mathcal{B}_t : \mathcal{N}_t \to \mathcal{N}_t$ given by

$$\mathcal{B}_t(\mathcal{S}_t) \equiv \arg\max_{k \in \mathcal{S}_t} f_t^\delta(G_{t-1}, k) = \arg\max_{k \in \mathcal{S}_t} d_{G_{t-1}}(k).$$

Then, in the limit $\beta \to \infty$, we obtain from (2) that

$$\lim_{\beta\to\infty} \mathbb{P}_t\Big(f_t^\delta(G_{t-1}, j) + \varepsilon_{tj} = \max_{k \in \mathcal{S}_t} f_t^\delta(G_{t-1}, k) + \varepsilon_{tk}\Big) = \frac{1}{|\mathcal{B}_t(\mathcal{S}_t)|} \mathbb{1}_{\mathcal{B}_t(\mathcal{S}_t)}(j).$$

Hence, the entrant makes a uniform draw without replacement from the best response set $\mathcal{B}_t$ when deciding with whom to form a link with probability 1, and the probability that an agent j receives a link by the entrant is given by

$$\begin{aligned}\lim_{\beta\to\infty} K_t^\beta(j|G_{t-1}, \mathcal{S}_t) &= \left(1 - \left(1 - \frac{1}{|\mathcal{B}_t(\mathcal{S}_t)|}\right) \cdots \left(1 - \frac{1}{|\mathcal{B}_t(\mathcal{S}_t)| - m + 1}\right)\right) \mathbb{1}_{\mathcal{B}_t(\mathcal{S}_t)}(j) \\ &= \left(1 - \frac{|\mathcal{B}_t(\mathcal{S}_t)| - m}{|\mathcal{B}_t(\mathcal{S}_t)|}\right) \mathbb{1}_{\mathcal{B}_t(\mathcal{S}_t)}(j) = \frac{m}{|\mathcal{B}_t(\mathcal{S}_t)|} \mathbb{1}_{\mathcal{B}_t(\mathcal{S}_t)}(j).\end{aligned}$$

We now give a proof by induction for $(G_t)_{t=m+2}^T$ and an arbitrary value of $n_s \geq 1$. The induction basis adds one agent at time $t = m+2$ to the complete graph K_{m+1}. By drawing a random sample $\mathcal{S}_t$ after selecting n_s agents from K_{m+1} uniformly at random, the entrant observes all agents in the set $[m+1] \equiv \{1, 2, \dots, m+1\}$. All of them have the same degree. Therefore, the entrant forms links to m of the agents in $[m+1]$ uniformly at random, and we obtain a quasi-star S_{m+2}^m with probability 1. Without loss of generality (w.l.o.g.), we can label the nodes that receive these links from 1 to m. Similarly, at time $t = m+3$, by sampling n_s agents in S_{m+2}^m, the entrant always observes the set of agents $[m]$. These agents have maximal degree in the prevailing network and hence obtain all the m links. It follows that we obtain the quasi-star S_{m+3}^m with probability 1.

In the following discussion, we consider the induction step. The induction hypothesis is that the network G_{t-1} is a quasi-star, with the highest degree agents in the set $[m]$.

After sampling n_s nodes uniformly at random, it must hold that $[m] \subseteq \mathcal{S}_t$ with probability 1. The reason is that either one of the agents in $[m]$ is observed directly. Since each of them has an outgoing link to all other agents in $[m]$, they all enter the sample $\mathcal{S}_t$. Otherwise, if one of the agents not in $[m]$ is observed directly, we know from the definition of the quasi-star that such an agent has outgoing links to all the agents in $[m]$ and, therefore, they all enter the sample $\mathcal{S}_t$. The agents in $[m]$ are the ones with the highest degree in G_{t-1} and so they receive all the m links. It follows that the network G_t must be a quasi-star. Hence, for all $n_s \geq 1$ and $T > m+1$, we must have that in the limit of $\beta \to \infty$, $G_T^\beta \in \Sigma_T^{m+1}$ almost surely.

Next, we consider part (ii) of the proposition. In the limit of strong shocks, as $\beta \to 0$, we obtain from (2) that

$$\lim_{\beta \to 0} \mathbb{P}_t\Big(f_t^\delta(G_{t-1}, j) + \varepsilon_{tj} = \max_{k \in \mathcal{S}_t} f_t^\delta(G_{t-1}, k) + \varepsilon_{tk}\Big) = \frac{1}{|\mathcal{S}_t|}.$$

It follows that the entrant selects m agents uniformly without replacement from the sample $\mathcal{S}_t$ with probability 1 as $\beta \to 0$. The probability that an agent j receives a link by the entrant is then given by

$$\lim_{\beta \to 0} K_t^\beta(j|G_{t-1}, \mathcal{S}_t) = \frac{m}{|\mathcal{S}_t|} \mathbb{1}_{\mathcal{S}_t}(j).$$

Let us consider the sequence $(G_t)_{t=m+2}^T$ with $n_s \geq 1$ and assume that $G_{t-1} \in \Sigma_{t-1}^m$. We are interested in the probability $\mathbb{P}_t(G_t \in \Sigma_t^m | G_{t-1} \in \Sigma_{t-1}^m)$. We have that $G_t \in \Sigma_t^m$ if only the m agents in the set $[m]$ receive a link by the entrant at time t. Given the sample $\mathcal{S}_t$, the probability that this happens is

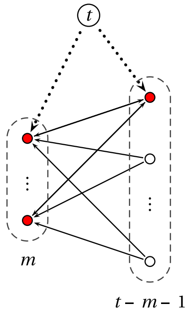
$$\frac{m}{|\mathcal{S}_t|}\left(\frac{m-1}{|\mathcal{S}_t|-1}\right) \cdots \left(\frac{1}{|\mathcal{S}_t|-m+1}\right) = \frac{m!(|\mathcal{S}_t|-m)!}{|\mathcal{S}_t|!} = \binom{|\mathcal{S}_t|}{m}^{-1}. \tag{9}$$

Consequently, we then can write

$$\mathbb{P}_t(G_t \in \Sigma_t^m | G_{t-1} \in \Sigma_{t-1}^m) = \sum_{\mathcal{S}_t \in \mathcal{P}_{t-1}} \binom{|\mathcal{S}_t|}{m}^{-1} \mathbb{P}_t(\mathcal{S}_t | G_{t-1} \in \Sigma_{t-1}^m). \tag{10}$$

Due to the properties of the quasi-star $G_{t-1} \in \Sigma_{t-1}^m$, the sample can only be of size $|\mathcal{S}_t| = m+1, m+2, \ldots, m+1+n_s$. The sample $\mathcal{S}_t$ has size $m+1$ if all the n_s draws are from the $m+1$ nodes in the set $[m+1]$ that are in the initial complete graph K_{m+1}. It is of size $m+2$ if $n_s - 1$ draws are from the set $[m+1]$, and one agent is drawn from the remaining agents, and so on. An illustration can be seen in Figure 9. Let X_0 denote the number of agents drawn from the set $[m+1]$ and let X_1 be the number of agents drawn from the remaining agents in the set $[t-1] \backslash [m+1]$. Then X_0 follows a hypergeometric distribution, and the sample size distribution is given by

$$\mathbb{P}_t(|S_t| = m+1+k|\cdot) = \mathbb{P}_t(X_0 = n_s - k, X_1 = k|\cdot) = \frac{\binom{m+1}{n_s-k}\binom{t-m-2}{k}}{\binom{t-1}{n_s}}.$$



$\lvert S_t\rvert$	X_0	X_1
$m+1$	n_s	0
$m+2$	n_s-1	1
$m+3$	n_s-2	2
$\vdots$	$\vdots$	$\vdots$
$m+1+n_s$	0	n_s

FIGURE 9. Left panel: Illustration of the selection of agents in a quasi-star by the entrant t. The filled circles indicate the nodes present in the initial complete graph K_{m+1}. Right panel: The variable X_0 denotes the number of agents drawn from the set $[m+1]$ and X_1 denotes the number of agents drawn from the remaining agents in the set $[t-1]\backslash[m+1]$. The table shows the possible values for $|S_t|$, X_0, and X_1.

The expected sample size is

$$\begin{aligned}\mathbb{E}_t[|S_t||\cdot] &= \sum_{k=0}^{n_s}(m+1+k)\mathbb{P}_t(|S_t|=m+1+k|\cdot) = (m+1+k)\frac{\binom{m+1}{n_s-k}\binom{t-m-2}{k}}{\binom{t-1}{n_s}} \\ &= n_s+m+1-\frac{n_s(m+1)}{t-1}.\end{aligned}$$

We thus find that the expected sample size is decreasing with n_s. Moreover, we have that the sample size distribution for n_s+1 first-order stochastically dominates the distribution for n_s. Let $0\le l\le n_s$. Then first-order stochastic dominance is implied by

$$\sum_{k=0}^{l}\frac{\binom{m+1}{n_s-k}\binom{t-m-2}{k}}{\binom{t-1}{n_s}} \ge \sum_{k=0}^{l}\frac{\binom{m+1}{n_s+1-k}\binom{t-m-2}{k}}{\binom{t-1}{n_s+1}},$$

which is equivalent to

$$\begin{aligned}0 &\le \sum_{k=0}^{l}\binom{t-2-m}{k}\left(\frac{\binom{m+1}{n_s-k}}{\binom{t-1}{n_s}}-\frac{\binom{m+1}{n_s+1-k}}{\binom{t-1}{n_s+1}}\right) \\ &= \frac{(l+1)(n_s-l-m-2)}{t(n_s-l)-m(n_s+1)-2(n_s+1)} \\ &\qquad\times\frac{\binom{t-m-2}{l+1}}{\binom{t-1}{n_s}\binom{t-1}{n_s+1}}\left(\binom{t-1}{n_s}\binom{m+1}{n_s-l}-\binom{t-1}{n_s+1}\binom{m+1}{n_s-l-1}\right) \\ &= \frac{(l+1)(n_s-l-m-2)}{t(n_s-l)-m(n_s+1)-2(n_s+1)}\binom{t-m-2}{l+1}\left(1+\frac{t-n_s-1}{n_s+1}\frac{n_s-l}{n_s-l-m-2}\right)\frac{\binom{m+1}{n_s-l}}{\binom{t-1}{n_s+1}} \\ &= \frac{l+1}{n_s+1}\frac{\binom{t-m-2}{l+1}\binom{m+1}{n_s-l}}{\binom{t-1}{n_s+1}}.\end{aligned}$$

The last expression is nonnegative for all admissible parameter values. If one distribution is first-order stochastically dominated by another, then the expected value of any decreasing function of a random variable governed by the first distribution is higher than the expectation under the latter (see, e.g., Mas-Colell et al. 1995). Since (9) is a decreasing function of the sample size $|\mathcal{S}_t|$, we can apply stochastic dominance and it follows that (10) is decreasing with n_s. The network $G_{t\leq m+1}$ is the complete graph K_{m+1} and, therefore, is a quasi star. The probability of observing a quasi-star in period T is given by $\mathbb{P}(G_T \in \Sigma_T^m) = \prod_{t=m+2}^T \mathbb{P}_t(G_t \in \Sigma_t^m | G_{t-1} \in \Sigma_{t-1}^m)$. As we have shown above, the probability $\mathbb{P}_t(G_t \in \Sigma_t^m | G_{t-1} \in \Sigma_{t-1}^m)$ is decreasing in n_s for any $t \geq m+2$. Thus, if $\beta \to 0$, it follows that for a sequence $(G_t^\beta)_{t=m+2}^T$ of networks generated under $n_s^{(1)}$ and a sequence $(H_t^\beta)_{t=m+2}^T$ of networks generated under $n_s^{(2)}$ with $n_s^{(1)} > n_s^{(2)}$, we must have that $\lim_{\beta\to 0}\mathbb{P}(G_T^\beta \in \Sigma_T^m) < \lim_{\beta\to 0}\mathbb{P}(H_T^\beta \in \Sigma_T^m)$. □

B.2 *The degree distributions*

Let us review some notation we have introduced in the main part of the paper. For all $t \geq 1$ we denote by $N_t(k) \equiv \sum_{i=0}^t \mathbb{1}_k(d_{G_t}(i))$ the number of nodes in the graph G_t with in-degree k. The relative frequency of nodes with in-degree k is accordingly defined as $P_t^\beta(k) \equiv (1/t)N_t(k)$ for all $t \geq 1$. The sequence $\{P_t^\beta(k)\}_{k\in\mathbb{N}}$ is the (empirical) *degree distribution*. In the following text, we will show almost sure convergence of the empirical degree distribution to its expected value (see Propositions 9 and 10), and explicitly characterize the limiting distribution (see the proofs of Propositions 2 and 3, and Corollary 1).

We will now derive a recursive system that can be used to describe the time evolution of the expected degree distribution. Let $N_t \equiv \{N_t(k)\}_{k\geq 0}$. Denoting $k = d^-_{G_{t-1}}(j)$, we write the attachment kernel as $K_t^\beta(j|G_{t-1}) = a(k)/(t\zeta(\beta,m)) + o(1/t)$. The expected number of nodes with in-degree k at time t can increase by the creation of a link to a node with in-degree $k-1$ or it decreases by the creation of a link to a node with in-degree k. It then follows that

$$\mathbb{E}[N_{t+1}(k)|N_t] = N_t(k)\left(1 - \frac{a(k)}{t\zeta(\beta,m)}\right) + N_t(k-1)\frac{a(k-1)}{t\zeta(\beta,m)} + \delta_{0,k} + o\left(\frac{1}{t}\right). \tag{11}$$

Taking expectations on both sides of (11), dividing by $t+1$, and denoting $P_t^\beta(k) = \mathbb{E}[N_t(k)]$ gives us

$$P_{t+1}^\beta(k) = \frac{t}{t+1}\left[P_t^\beta(k)\left(1 - \frac{a(k)}{t\zeta(\beta,m)}\right) + P_t^\beta(k-1)\frac{a(k-1)}{t\zeta(\beta,m)} + \frac{1}{t}\delta_{0,k}\right] + o\left(\frac{1}{t}\right).$$

Some algebraic manipulations allow us to write this as

$$P_{t+1}^\beta(k) - P_t^\beta(k) = b_t(k)[c_t(k) - P_t^\beta(k)] + o\left(\frac{1}{t}\right), \tag{12}$$

where

$$b_t(k) \equiv \frac{\zeta(\beta,m)+a(k)}{\zeta(\beta,m)}\frac{1}{t+1}$$

$$c_t(k) \equiv P_t^{\beta}(k-1)\frac{a(k-1)}{\zeta(\beta,m)+a(k)} + \frac{\zeta(\beta,m)}{\zeta(\beta,m)+a(k)}\delta_{0,k}.$$

The following lemma gives us a simple way to determine the asymptotic solution (i.e., as $t \to \infty$) of the recursion in (12).

Lemma 1. *Let* (x_n), (y_n), (η_n), *and* (r_n) *denote real sequences such that*

$$x_{n+1} - x_n = \eta_n(y_n - x_n) + r_n$$

and (i) $\lim_{n\to\infty} y_n = x$, *(ii)* $\eta_n > 0$, $\sum_{n=1}^{\infty} \eta_n = \infty$ *and there exists a* N_0 *such that for all* $n \geq N_0$, $\eta_n < 1$, *and (iii)* $r_n = o(\eta_n)$. *Then* $\lim_{n\to\infty} x_n = x$.

For the proof of Lemma 1, see Jordan (2006, p. 229).

For our purposes the lemma can be applied by identifying $x_t = P_t^{\beta}(k)$, $\eta_t = b_t(k)$ and $y_t = c_t(k)$. We have that $b_t(k) > 0$ and $\sum_{t\geq 0} b_t(k) = \infty$ since $\zeta(\beta,m) < \infty$. Under this condition it is evident that $c_t(k)$ has a well defined limit, which is determined in a recursive way. We give a proof by induction. The induction basis follows from the case of $k=0$, where

$$c(0) \equiv \lim_{t\to\infty} c_t(0) = \frac{\zeta(\beta,m)}{\zeta(\beta,m)+a(0)}.$$

To proceed with the induction proof, suppose we have already determined the lower tail of the distribution $c(0) = P^{\beta}(0), \dots, c(k-1) = P^{\beta}(k-1)$, $k > 0$. Then we see that

$$c(k) \equiv \lim_{t\to\infty} c_t(k) = P^{\beta}(k-1)\frac{a(k-1)}{\zeta(\beta,m)+a(k)},$$

and iterating this equation with respect to k gives us

$$c(k) = P^{\beta}(0)\prod_{j=1}^{k}\frac{a(j-1)}{\zeta(\beta,m)+a(j)}.$$

Hence, we get for the explicit expression for the asymptotic degree distribution

$$P^{\beta}(k) = \frac{\zeta(\beta,m)}{\zeta(\beta,m)+a(0)}\prod_{j=1}^{k}\frac{a(j-1)}{\zeta(\beta,m)+a(j)}. \tag{13}$$

This general scheme can be used to determine the degree distribution for the different parameters we consider, as we show now in the following proof.

Proof of Proposition 2. For $\beta \to 0$ the attachment kernel of (3) is given by $K_t^{\beta}(j|G_{t-1}) = a(k)/(t\zeta(\beta,m)) + o(1/t)$, where $k = d_{G_{t-1}}(j)$, $a(k) = 1 + \beta k$, and $\zeta(\beta,m) =$

$(1+\beta m)/m$. We then can apply (13), noting that the product on the right-hand side admits a closed-form representation in terms of Gamma functions as

$$P^{\beta}(k)=\frac{1+\beta m}{1+m(1+\beta)}\frac{\Gamma\big(\frac{1}{\beta}+k\big)\Gamma\big(2+\frac{1+\beta m}{\beta m}\big)}{\Gamma\big(\frac{1}{\beta}\big)\Gamma\big(2+\frac{1+m}{1+\beta m}+k\big)}. \tag{14}$$

By Stirling's formula we can approximate the Gamma function for large k as[57]

$$\frac{\Gamma(k)}{\Gamma(k+c)}=k^{-c}\left(1+O\left(\frac{1}{k}\right)\right). \tag{15}$$

For the tails of the degree distribution in (14) this implies that $P^{\beta}(k)\sim(1+\beta k)^{-(2+1/(\beta m))}(1+O(1/k))$ for large k. □

The case of $\beta=0$ can be treated analogously.

Proof of Corollary 1. The degree distribution in (5) follows from the attachment kernel $K_t^0(j|G_{t-1})=a(k)/(t\zeta(\beta,m))+o(1/t)=m/t+o(1/t)$ and inserting $a(k)=1$ and $\zeta(\beta,m)=1/m$ into (13). □

Similarly, we can derive the asymptotic degree distribution in Proposition 3 for $\beta=0$ when the observation radius n_s is small enough. The proof is given in the following text.

Proof of Proposition 3. With the attachment kernel from (7) given by $K_t^0(j|G_{t-1})=a(k)/(t\zeta(\beta,m))+o(1/t)=(m/(m+1))((1+k)/t)+o(1/t)$, where $k=d_{G_{t-1}}(j)$, $a(k)=1+k$, and $\zeta(\beta,m)=(m+1)/m$, we can apply (13) to obtain

$$P(k)=\frac{(1+m)\Gamma\big(3+\frac{1}{m}\big)\Gamma(k+1)}{(1+2m)\Gamma\big(3+\frac{1}{m}+k\big)},\quad k\geq 0.$$

Using (15) we get $P(k)\sim k^{-(2+1/m)}$ for large k. □

[57] By Stirling's formula we can approximate the Gamma function for large k as

$$\Gamma(k)=\sqrt{\frac{2\pi}{k}}\left(\frac{k}{e}\right)^{k}\left(1+O\left(\frac{1}{k}\right)\right).$$

Hence,

$$\frac{\Gamma(k)}{\Gamma(k+a)}=\left(1+O\left(\frac{1}{k}\right)\right)\sqrt{(1+a/k)}(1+a/k)^{-k}\left(\frac{k}{k+a}\right)^{k}\left(\frac{k+a}{e}\right)^{-a}.$$

Since $\sqrt{(1+a/k)}\to 1$ for $k\to\infty$, this term is asymptotically negligible. Additionally $(1+a/k)^{-k}\to e^{-a}$ for $k\to\infty$, and $(k+a)^{-a}\sim k^{-a}$ for $k\to\infty$. Hence, the leading order approximation of the ratio of Gamma functions is given by

$$\frac{\Gamma(k)}{\Gamma(k+a)}=k^{-a}\left(1+O\left(\frac{1}{k}\right)\right).$$

Finally, we can give an upper bound on the deviations for finite t and show that the empirical degree distribution is a consistent estimator of the expected degree distribution in the limit of large t.

Proposition 9. *Let the empirical in-degree distribution be given by $\{P_t(k)\}_{k\in\mathbb{N}}$. Then for any $\epsilon > 0$ we have that*

$$\mathbb{P}_t\big(\big|P_t(k) - \mathbb{E}_t[P_t(k)]\big| \geq \epsilon\big) \leq 2\exp\left(-\frac{\epsilon^2 t}{8(m+1)^2}\right), \tag{16}$$

and $P_t(k)$ converges in probability to $\mathbb{E}_t[P_t(k)]$ for large t.

Proof. Let the number of vertices with in-degree k in network $G_t = \langle \mathcal{N}_t, \mathcal{E}_t\rangle$ be denoted by $N_t(k) = \sum_{i\in\mathcal{N}_t} \mathbb{1}_{d^-_{G_{t-1}}(i)}(k) = |\mathcal{N}_t| P_t(k)$. Consider the filtration $\mathcal{F}_n = \sigma(G_1, G_2, \ldots, G_n)$, $1 \leq n \leq t$, which is the smallest σ-algebra generated by $G_1, G_2, \ldots, G_n$, with the property that $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, and let $\mathcal{F}_\infty$ be the σ-algebra generated by the infinite union of the $\mathcal{F}_n$'s. For $n = 1, \ldots, s$, we denote the conditional expectation of the number of vertices with in-degree k at time s, conditional on the filtration $\mathcal{F}_n$, by $Z_n = \mathbb{E}_t[N_t(k)|\mathcal{F}_n]$. First, from the fact that $N_t(k) \leq t$, it follows that $\mathbb{E}_t[|Z_n|] = \mathbb{E}_t[Z_n] = \mathbb{E}_t[N_t(k)] \leq t < \infty$. Second, since $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, we have that for all $n \leq t-1$, $\mathbb{E}_t[Z_{n+1}|\mathcal{F}_n] = \mathbb{E}_t[\mathbb{E}_t[N_t(k)|\mathcal{F}_{n+1}]|\mathcal{F}_n] = \mathbb{E}_t[N_t(k)|\mathcal{F}_n] = Z_n$. We thus find that $(Z_n)_{n=1}^t$ is a martingale with respect to $(\mathcal{F}_n)_{n=1}^t$.

Moreover, note that $Z_1 = \mathbb{E}_t[N_t(k)|\mathcal{F}_1] = \mathbb{E}_t[N_t(k)|G_1]$, since $\mathcal{F}_1$ contains no more information than the initial network G_1. The Z_t is given by $Z_t = \mathbb{E}_t[N_t(k)|\mathcal{F}_t] = N_t(k)$. Therefore, we have that $Z_t - Z_1 = N_t(k) - \mathbb{E}_t[N_t(k)|G_1]$. Next, we show that $|Z_n - Z_{n-1}| \leq 2(m+1)$. To see this note that $Z_n = \mathbb{E}_t[N_t(k)|\mathcal{F}_n] = \sum_{i\in\mathcal{N}_t} \mathbb{P}_t(d_{G_{t-1}}(i) = k|\mathcal{F}_n)$ and, similarly, $Z_{n-1} = \mathbb{E}_t[N_t(k)|\mathcal{F}_{n-1}] = \sum_{i\in N_s} \mathbb{P}_t(d_{G_{t-1}}(i) = k|\mathcal{F}_{n-1})$, so that we can write

$$Z_n - Z_{n-1} = \sum_{i\in\mathcal{N}_t} \big[\mathbb{P}_t(d_{G_{t-1}}(i) = k|\mathcal{F}_n) - \mathbb{P}_t(d_{G_{t-1}}(i) = k|\mathcal{F}_{n-1})\big]. \tag{17}$$

In $\mathcal{F}_{n-1}$ we know where the edges up to time $n-1$ have been attached. In $\mathcal{F}_n$ we know in addition where the edges in the nth step are attached. These edges affect the total degree of $m+1$ vertices, namely those receiving a link and those initiating the links.

For the conditional expectation given $\mathcal{F}_n$, we need to take the expectation over all possible ways of attaching the remaining edges in the periods $n+1, \ldots, s$. Only the distribution of the degrees of the vertices that have obtained or initiated an edge in period n are affected by the knowledge of $\mathcal{F}_n$, compared to the knowledge of $\mathcal{F}_{n-1}$. Neither the probability of the other vertices to receive a link nor the probability to initiate a link is affected by the creation of the edges in the nth step. Thus, also the law of their total degree is unaffected. There are at most $m+1$ vertices that receive or initiate a link in period n. Therefore, (17) shows that the distribution of at most $2(m+1)$ vertices in G_t is different by conditioning on $\mathcal{F}_n$ compared to conditioning on $\mathcal{F}_{n-1}$. This implies that $|Z_n - Z_{n-1}| \leq 2(m+1)$. We then can apply the Azuma–Hoeffding inequality (see, e.g., Grimmett and Stirzaker 2001) to obtain, for any $\eta > 0$,

$$\mathbb{P}_t\big(\big|N_t(k) - \mathbb{E}_t[N_t(k)|G_1]\big| \geq \eta\big) \leq 2\exp\left(-\frac{\eta^2}{8(m+1)^2 t}\right),$$

and by choosing $\eta = \epsilon t$, (16) follows. □

With Proposition 9 we are now able to show almost sure convergence of the empirical degree distribution to its expected value.

Proposition 10. *For a fixed $k \geq 0$, $P_t(k) \xrightarrow{a.s.} \mathbb{E}_t[P_t(k)]$ as $t \to \infty$.*

Proof. The proof follows from the Borel–Cantelli lemma (see, e.g., Grimmett and Stirzaker 2001) and Proposition 9 by observing that for any $\epsilon > 0$,

$$\sum_{t=1}^{\infty} \mathbb{P}_t\big(\big|P_t(k) - \mathbb{E}_t[P_t(k)]\big| \geq \epsilon\big) \leq 2\sum_{t=1}^{\infty} e^{-(\epsilon^2 t)/(8(m+1)^2)} = \frac{1}{e^{(\epsilon^2)/(8(m+1)^2)} - 1} < +\infty.$$ □

B.3 *Efficiency*

Proof of Proposition 4. Part (i) of the proposition is a direct consequence of part (ii) of Proposition 1.

Part (ii) of the proposition follows from the fact that networks generated under $(H_t)_{t=m+2}^T$ have a finite degree variance, while the degree variance of networks generated under $(G_t)_{t=m+2}^T$ diverge with T, since the first has a geometric degree distribution while the latter has a power-law degree distribution in the large T limit. More precisely, the degree variance under H_T is given by

$$\sigma_d^2 = \lim_{T\to\infty} \sum_{k=0}^{T} \frac{1}{1+m}\left(\frac{m}{m+1}\right)^k (k-m)^2 = m(m+1) < +\infty$$

while the variance under G_T is

$$\sigma_d^2 = \lim_{T\to\infty} \sum_{k=0}^{T} \frac{(m+1)\Gamma\big(3+\frac{1}{m}\big)\Gamma(k+1)}{(1+2m)\Gamma\big(3+\frac{1}{m}+k\big)}(k-m)^2 = \lim_{T\to\infty} O(T^{1-1/m}) = +\infty,$$

if $m > 1$, while for $m = 1$ we get

$$\sigma_d^2 = \lim_{T\to\infty}\left(4H_{T+1} - \frac{4(1+T)(5+3T)}{6+5T+T^2}\right) = +\infty,$$

where H_T is the harmonic number, diverging as $\ln T$ for large T. □

B.4 *Higher order statistics*

The results of this section are derived using a *continuum approximation* in which both time and degree are treated as continuous variables in $\mathbb{R}_+$ (see Dorogovtsev and Mendes 2013, p. 117). In this continuum approach, the probability that a vertex s has in-degree $d^-_{G_t}(s) = k$ at time t is given by $\delta(k - \bar{k}(s,t))$, where $\bar{k}(s,t) = \mathbb{E}_t[d^-_{G_t}(s)]$ denotes the expected degree of vertex s at time t. The degree distribution can then be obtained from

$$P_t(k) = \frac{1}{t}\int_0^t \delta(k - \bar{k}(s,t))\,ds = -\frac{1}{t}\left(\frac{\partial \bar{k}(s,t)}{\partial s}\right)^{-1}\Bigg|_{s=s(k,t)}. \tag{18}$$

So as to compare this approximation with our previous analysis, we will derive the degree distributions in the case of a large and a small observation radius. To ease the notation we will denote by $k_s(t)$ the in-degree $d^-_{G_t}(s)$ of a vertex s at time t for the remainder of this section, and we will focus only on the in-degree $k_s(t)$, since it uniquely determines the total degree $d_{G_t}(s) = k_s(t) + m$ and vice versa.

We first consider the expected change in the in-degree $k_s(t)$ of a vertex s receiving a link from an entrant t when $\mathcal{S}_t = \mathcal{P}_{t-1}$ (large observation radius). In the continuum approximation, the corresponding expectation in the time interval $[t, t+\Delta t)$ is given by $\mathbb{E}_t[k_s(t+\Delta t) - k_s(t)|G_t] \approx (m/(1+\beta m))((1+\beta k_s(t))/t)\Delta t$ for large t, where (3) describes a transition rate and $\Delta t = O(1/T)$. The evolution of the in-degree of vertex s at time t is governed by the differential equation

$$\frac{dk_s(t)}{dt} = \lim_{\Delta t \downarrow 0}\frac{\mathbb{E}_t[k_s(t+\Delta t) - k_s(t)|G_t]}{\Delta t} = \frac{m}{1+\beta m}\frac{1+\beta k_s(t)}{t},$$

with the initial condition $k_s(s) = 0$ for all $s \geq 0$. The solution is given by

$$k_s(t) = \frac{1}{\beta}\left(\left(\frac{t}{s}\right)^{(m\beta)/(1+m\beta)} - 1\right). \tag{19}$$

From (18) we then get

$$P^\beta(k) = \frac{1+\beta m}{m}(1+\beta k)^{-(2+1/(\beta m))}, \tag{20}$$

with $\int_0^\infty P^\beta(k)\,dk = 1$. This is asymptotically equivalent to the degree distribution we obtained in (4).

Similarly, in the case of n_s small enough (small observation radius), we have from (7) that $\mathbb{E}_t[k_s(t+\Delta t) - k_s(t)|G_t] \approx (m/(1+m))((1+k_s(t))/t)\Delta t$ for large t. The time evolution of the in-degree of a vertex s can then be written as

$$\frac{dk_s(t)}{dt} = \frac{m}{m+1}\frac{k_s(t)+1}{t},$$

with the initial condition $k_s(s) = 0$ for all $s \geq 0$. The solution is given by

$$k_s(t) = \left(\frac{t}{s}\right)^{m/(m+1)} - 1. \tag{21}$$

From (18) we then get

$$P(k) = \frac{m+1}{m}(1+k)^{-(2+1/m)}, \tag{22}$$

with the property that $\int_0^\infty P(k)\,dk = 1$. Comparing this distribution with the one in (8) shows that they are both asymptotically equivalent. Since the continuum approximation delivers only meaningful results in the large t limit, we will consider only the leading order terms in $O(1/t)$ in our derivations in the following sections.

B.4.1 *Average nearest neighbor degree distribution*

Proof of Proposition 5. Let $R_s^-(t)$ denote the sum of in-degrees of the in-neighbors of a vertex s at time t, that is, $R_s^-(t) = \sum_{j\in\mathcal{N}_{G_t}^-(s)} k_j(t)$. In the continuum approximation, with the attachment kernel from (3), we have up to leading orders in $O(1/t)$ that

$$\frac{dR_s^-(t)}{dt} = \sum_{j\in\mathcal{N}_{G_t}^-(s)} m\frac{1+\beta k_j(t)}{(1+\beta m)t} = \frac{a}{t}R_s^-(t) + \frac{a}{\beta t}k_j(t) = \frac{a}{t}R_s^-(t) + \frac{a}{\beta^2 t}\left(\left(\frac{t}{s}\right)^a - 1\right),$$

where we have denoted $a \equiv (m\beta)/(1+m\beta)$. Wit the initial condition $R_s^-(s) = 0$ we obtain

$$R_s^-(t) = \frac{1}{\beta^2}\left(1 + \left(a\ln\left(\frac{t}{s}\right) - 1\right)\left(\frac{t}{s}\right)^a\right),$$

and the average nearest neighbor in-degree is given by $k_{\mathrm{nn}}^-(k_s) = R_s^-(t)/k_s$. From (19) we know that $t/s = (1+\beta k_s)^{1/a}$, and we obtain

$$k_{\mathrm{nn}}^-(k) = \frac{1}{\beta^2 k}\big(1 + (\ln(1+\beta k) - 1)(1+\beta k)\big). \tag{23}$$

Next, we turn to the analysis of the average nearest out-neighbor in-degree. Let us denote by $R_s^+(t)$ the sum of the in-degrees of the out-neighbors of vertex s at time t, that is, $R_s^+(t) = \sum_{j\in\mathcal{N}_{G_t}^+(s)} k_j(t)$. Up to leading orders in $O(1/t)$ we can write

$$\frac{dR_s^+(t)}{dt} = \sum_{j\in\mathcal{N}_{G_t}^+(s)} \frac{a}{t}\left(\frac{1}{\beta} + k_j(t)\right) = \frac{a}{t}\left(\frac{m}{\beta} + R_s^+(t)\right).$$

The solution is given by

$$R_s^+(t) = -\frac{m}{\beta} + C_s t^a, \tag{24}$$

where the constant C_s is determined by the initial conditions. They are given by

$$R_{s+1}^+ = \sum_{j=1}^{s}\frac{a}{s}\left(\frac{1}{\beta} + k_j(s)\right)(k_j(s)+1) = \frac{a}{\beta^2}\big(\beta(1+m(\beta-1)) - 1 + s^{2a-1}\zeta(s,2a)\big),$$

where $\zeta(s, 2a)$ is the Hurwitz zeta function.[58] Together with the solution (24) we then get

$$R_s^+(t) = \frac{1}{\beta^2}\left(\left(\beta m(1+p(\beta-1)) + \frac{a}{s}s^{2a}\zeta(s, 2a)\right)\left(\frac{t}{s+1}\right)^a - m\beta\right).$$

The average nearest out-neighbor in-degree is then given by $k_{\mathrm{nn}}^+(k) = R_s^+/m$, that is,

$$k_{\mathrm{nn}}^+(k) = \frac{1}{\beta^2 m}\left(\left(\beta m(1+p(\beta-1)) + \frac{a}{s}s^{2a}\zeta(s, 2a)\right)\left(\frac{t}{s+1}\right)^a - m\beta\right).$$

Hence, we find that for large k, the average nearest in-neighbor connectivity grows logarithmically with k and is independent of t, while the average nearest out-neighbor connectivity becomes independent of k and grows with the network sizes as $t^{(\beta m)/(1+\beta m)}$. □

Proof of Proposition 6. Let $R_s^-(t)$ denote the sum of in-degrees of the in-neighbors of a vertex s at time t, that is, $R_s^-(t) = \sum_{j\in\mathcal{N}_{G_t}^-(s)} k_j(t)$. In the continuum approximation, with the attachment kernel from (7), we have up to leading orders in $O(1/t)$ that[59]

$$\frac{dR_s^-(t)}{dt} = \frac{a}{t}\sum_{j\in\mathcal{N}_{G_t}^-(s)}(1+k_j(t)) = \frac{a}{t}k_s(t) + \frac{a}{t}R_s^-(t),$$

where we have denoted $a \equiv m/(1+m)$. In the continuum approximation we have that $k_s(t) = (t/s)^a - 1$ (see (21)), so that we can write

$$\frac{dR_s^-(t)}{dt} = \frac{a}{t}\left(\left(\frac{t}{s}\right)^a - 1 + \frac{a}{t}R_s^-(t)\right).$$

The solution is given by

$$R_s^-(t) = C_s t^a + 1 + a\left(\frac{t}{s}\right)^a \ln t,$$

where the constant C_s is determined by the initial conditions given by $R_s^-(s) = 0$. With these initial conditions we get

$$R_s^-(t) = 1 - \left(\frac{t}{s}\right)^a + a\left(\frac{t}{s}\right)^a \ln\left(\frac{t}{s}\right).$$

Further, using the fact that $s(k, t) = t/((k+1)^{1/a})$ we obtain

$$R_s^-(t) = 1 + (k+1)(\ln(k+1) - 1).$$

It follows that

$$k_{\mathrm{nn}}^- = \frac{R_s^-}{k} = \frac{1}{k}\big(1 + (k+1)(\ln(k+1) - 1)\big).$$

[58] The Hurwitz zeta function is defined by $\zeta(s, a) \equiv \sum_{n=0}^{\infty} 1/(a+n)^s$.

[59] We ignore cases in which two or more neighbors of s are found as the neighbors of directly observed vertices (other than s), which happens with probability $O(1/t^2)$.

Next, we turn to the average nearest out-neighbor in-degree. Let us denote by $R_s^+(t)$ the sum of the in-degrees of the out-neighbors of vertex s at time t, that is, $R_s^+(t) = \sum_{j\in\mathcal{N}_{G_t}^+(s)} k_j(t)$. So as to compute the expected increase in the sum of the degrees of the out-neighbors of s we need to consider two different cases. First, s is observed directly and enters the sample $\mathcal{S}_t$ together with all the out-neighbors. The expected number of links created among the out-neighbors of s in this way is given by

$$\frac{n_s}{t}\sum_{k=1}^{m} k\frac{\binom{m}{k}\binom{|\mathcal{S}_t|-m}{m-k}}{\binom{|\mathcal{S}_t|}{m}} = \frac{m^2}{(m+1)t},$$

where we have used the fact that $|\mathcal{S}_t| = n_s(m+1)$ up to leading orders in $O(1/t)$. Second, we need to consider the cases where the out-neighbors of s are found either directly or indirectly through vertices other than s. The probability of this is given by $(m/((m+1)t))k_j(t)$ for each j in $\mathcal{N}_{G_t}^+(s)$ (discounting the link from s). Taking these cases together and denoting $a = m/(m+1)$, we can write

$$\frac{dR_s^+(t)}{dt} = \frac{ma}{t} + \sum_{j\in\mathcal{N}_{G_t}^+(s)} \frac{a}{t}k_j(t) = \frac{ma}{t} + \frac{a}{t}R_s^+(t),$$

with the solution

$$R_s^+(t) = -m + C_s t^a.$$

The term C_s is determined by the initial condition $R_s^+(s)$, which is given by

$$R_s^+(s) = \frac{a}{s}\sum_{j=1}^{s}(1+k_j(s))^2 = as^{2a-1}H(s,2a),$$

where $H(s,2a) \equiv \sum_{j=1}^{s} j^{-2a}$ is the generalized harmonic number. Inserting the initial condition delivers

$$R_s^+(t) = m\left(\left(\frac{t}{s}\right)^a - 1\right) + aH(s,2a)s^{a-1}t^a.$$

Further, using $s(k,t) = t/((k+1)^{1/a})$ from (21) gives

$$R_s^+(k) = \left(\frac{m\Gamma(2+m)^2}{\Gamma\left(1+m+\frac{m}{m+1}\right)^2} + \frac{m}{m+1}\zeta\left(\frac{2m}{m+1}, 2+m\right)\right)t^{(m-1)/(m+1)}(1+k)^{1/m}.$$

With $k_{\text{nn}}^+(k) = R_s^+(k)/m$ we then get

$$k_{\text{nn}}^+(k) = \left(\frac{\Gamma(2+m)^2}{\Gamma\left(1+m+\frac{m}{m+1}\right)^2} + \frac{1}{m+1}H\left(\frac{2m}{m+1}, 2+m\right)\right)t^{(m-1)/(m+1)}(1+k)^{1/m}.$$

For large k we find that $k_{\text{nn}}^-(k)$ grows logarithmically with k and is independent of the network size t, and $k_{\text{nn}}^+(k)$ grows as $O(t^{(m-1)/(m+1)}\cdot k^{1/m})$. □

Figure 10. Left panel: Vertex s and one of its out-neighbors $u \in \mathcal{N}^+_{G_t}(s)$ receive a link by the entrant t. Right panel: Vertex s and one of its in-neighbors $u \in \mathcal{N}^-_{G_t}(s)$ receive a link.

B.4.2 *Clustering degree distribution* We denote by $M_s(t)$ the number of links between neighbors of vertex s at time t in the closure $\overline{G}_t$. The clustering coefficient of vertex s can then be written as

$$C_s(t) = \frac{2M_s(t)}{(k_s(t)+m)(k_s(t)+m-1)}.$$

In the following text we derive the clustering coefficient for different observation radii. In the case of a large observation radius, we can give the following proof.

Proof of Proposition 7. The term $M_s(t)$ can increase at time t only through the addition of an edge to s and one of its neighbors. There are two possible cases to consider: (i) vertex s and one of its out-neighbors $u \in \mathcal{N}^+_{G_t}(s)$ receive a link, or (ii) s and one of its in-neighbors $u \in \mathcal{N}^-_{G_t}(s)$ receive a link. This is illustrated in Figure 10.

The probability associated with case (i) up to leading orders in $O(1/t)$ is given by

$$\frac{m(1+\beta k_s(t))}{(1+\beta m)t} \sum_{j \in \mathcal{N}^+_{G_t}(s)} \frac{(m-1)(1+\beta k_j(t))}{(1+\beta m)t} = \frac{m(m-1)(1+\beta k_s(t))}{(1+\beta m)^2 t^2}(m+\beta R^+_s(t)).$$

Similarly, the probability associated with case (ii) up to leading orders in $O(1/t)$ is given by

$$\frac{m(1+\beta k_s(t))}{(1+\beta m)t} \sum_{j \in \mathcal{N}^-_{G_t}(s)} \frac{(m-1)(1+\beta k_j(t))}{(1+\beta m)t} = \frac{m(m-1)(1+\beta k_s(t))}{(1+\beta m)^2 t^2}(k_s(t)+\beta R^-_s(t)).$$

With R^-_s and R^+_s given in the proof of Proposition 5, we obtain

$$\begin{aligned}\frac{dM_s(t)}{dt} &= \frac{m(m-1)(1+\beta k_s(t))}{(1+\beta m)t^2}(m+k_s(t)+\beta(R^+_s+R^-_s)) \\ &= \frac{a^2}{t^2}\frac{m-1}{m\beta^3}\left((\beta^2 m + a s^{2a-1} H(s,2a))\left(\frac{t}{s}\right)^a \left(\frac{t}{s+1}\right)^a + \left(\frac{t}{s}\right)^{2a} a \ln\left(\frac{t}{s}\right)^a\right).\end{aligned} \tag{25}$$

The initial condition M_s is determined by all connected pairs of vertices i, j, which both obtain a link from the entering vertex s at time s. Taking into account that all vertices with $i \leq m$ are connected while the vertices i, j introduced later in the network are connected only if either i has formed a link to j or j to i (depending on who has entered the network first and noting that all vertices with indices $1 \le i \le m$ are initially connected), we can write the initial condition as[60]

$$
\begin{aligned}
M_{s+1} &= \frac{m(m-1)}{2} \sum_{j \neq i}^{s} \frac{1+\beta k_i(s)}{(1+\beta m)s} \frac{1+\beta k_j(s)}{(1+\beta m)s} \Bigg(\Theta(m+1-i)\Theta(m+1-j) \\
&\qquad + \Theta(i-j)\Theta(j-m) m \frac{1+\beta k_j(i)}{(1+\beta m)(i-1)} \\
&\qquad + \Theta(j-i)\Theta(i-m) m \frac{1+\beta k_i(j)}{(1+\beta m)(j-1)} \Bigg) \\
&= \frac{m(m-1)s^{2a-2}}{(1+\beta m)^2} \left(\sum_{i=1}^{m} \frac{1}{i^a} \sum_{j=i+1}^{m} \frac{1}{j^a} + \frac{2m}{1+\beta m} \sum_{i=m+1}^{s} \frac{1}{i^{2a}} \sum_{j=i+1} \frac{1}{j-1} \right),
\end{aligned} \tag{26}
$$

where $a = (\beta m)/(1+\beta m)$. Combining the initial condition in (26) with (25) shows that the clustering coefficient of an agent with degree k is given by

$$
\begin{aligned}
C(k) = \frac{2}{(k+pm)(k+pm-1)} \frac{a(m-1)}{mp\beta^3 b^2 s} \Bigg(& sb^2 \frac{mp\beta^3}{a(m-1)} M_s + ((1+\beta k)^b - 1) \\
& \times \left(b\left(\frac{s}{s+1}\right)^a (\beta^2 m + a s^{2a-1} \zeta(s, 2a)) - 1 \right) + b(1+\beta k)^b \ln(1+\beta k) \Bigg),
\end{aligned}
$$

where $a = (\beta m)/(1+\beta m)$, $b = 2 - 1/a$, the initial condition is

$$
M_{s+1} = \frac{m(m-1)s^{2a-2}}{(1+\beta m)^2} \left(\sum_{i=1}^{m} \frac{1}{i^a} \sum_{j=i+1}^{m} \frac{1}{j^a} + \frac{2m}{1+\beta m} \sum_{i=m+1}^{s} \frac{1}{i^{2a}} \sum_{j=i}^{s-1} \frac{1}{j} \right),
$$

and $s = t(1+\beta k)^{-1/a}$ as $t \to \infty$. For large k (and small s, respectively) the first term in the initial condition M_{s+1} dominates and the asymptotic behavior of the clustering coefficient is given by

$$
C(k) = O(t^{-2/(1+m\beta)} \cdot k^{2(1/(m\beta)-1)}).
$$

This expression grows with k as a power law with exponent $2(1/(m\beta)-1)$.[61] Moreover, we find that the clustering coefficient is decreasing with the network size as $t^{-2/(1+m\beta)}$. □

Next, we turn to the derivation of the clustering coefficient when the observation radius is small.

Proof of Proposition 8. For the increase of $M_s(t)$ at time t we have to consider the following cases: (i) vertex s and one of its out-neighbors $u \in \mathcal{N}^+_{G_t}(s)$ receive a link or (ii) s and one of its in-neighbors $\in \mathcal{N}^-_{G_t}(s)$ receive a link, and (iii) the entrant observes a vertex v and forms a link to both vertices s and u which are both out-neighbors of v. This is illustrated in Figure 11. In case (i), we consider that vertex s is observed directly. The

[60] The Heaviside step function is defined as $\Theta(x) = 1$ if $x > 0$ and $\Theta(x) = 0$ if $x \le 0$.

[61] We need only consider values of k such that $C(k)$ does not exceed its upper bound given by 1.

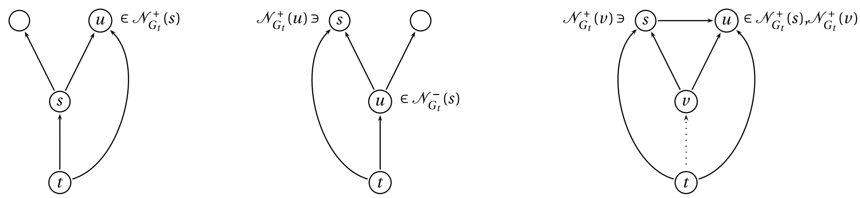


FIGURE 11. Left panel: Vertex s and one of its out-neighbors $u \in \mathcal{N}^+_{G_t}(s)$ receive a link. Middle panel: Vertex s and one of its in-neighbors $u \in \mathcal{N}^-_{G_t}(s)$ receive a link. Right panel: The entrant t observes a vertex v and forms a link to both vertices s and u, which are both out-neighbors of v.

probability of this happening is given by n_s/t. Assuming that s has been observed directly, s and all the out-neighbors $\mathcal{N}^+_{G_t}(s)$ of s are in the sample $\mathcal{S}_t$. We can then partition the sample $\mathcal{S}_t$ in three subsets: $\{s\}$, $\mathcal{N}^+_{G_t}(s)$, and $\mathcal{S}_t \backslash (\mathcal{N}^+_{G_t}(s) \cup \{s\})$, with corresponding cardinalities $|\{s\}| = 1$, $|\mathcal{N}^+_{G_t}(s)| = m$, and $|\mathcal{S}_t \backslash (\mathcal{N}^+_{G_t}(s) \cup \{s\})| = n_s(m+1) - (m+1)$. We need to take into account all cases where vertex s and at least one of the out-neighbors of s receive a link. The expected number of triangles formed in this way can then be computed with a trivariate hypergeometric distribution as

$$\frac{n_s}{t}\sum_{k=1}^{m-1} k \frac{\binom{1}{1}\binom{m}{k}\binom{|\mathcal{S}_t|-(m+1)}{m-(k+1)}}{\binom{|\mathcal{S}_t|}{m}} = \frac{n_s}{t}\sum_{k=1}^{m-1} k \frac{\binom{m}{k}\binom{(n_s-1)(m+1)}{m-(k+1)}}{\binom{n_s(m+1)}{m}} = \frac{m^2(m-1)}{(m+1)(n_s(m+1)-1)t}.$$

In case (ii), we consider that one of the in-neighbors $u \in \mathcal{N}^-_{G_t}(s)$ of s is observed directly by the entrant, which happens with probability n_s/t, and both u and s receive a link. The latter event follows a bivariate hypergeometric distribution where two nodes are drawn from the set $\{s, u\}$ and $m-2$ are drawn from the remaining nodes in the set $\mathcal{S}_t \backslash \{s, u\}$ with a total of m draws. Summing over all $k_s(t)$ in-neighbors of s delivers the total probability measure associated with case (ii) as given by

$$k_s(t)\frac{n_s}{t}\frac{\binom{2}{2}\binom{|\mathcal{S}_t|-2}{m-2}}{\binom{|\mathcal{S}_t|}{m}} = \frac{k_s(t)}{t}\frac{m(m-1)}{(m+1)(n_s(m+1)-1)}.$$

Next, in (iii) we need to consider all cases where a node v is observed directly by the entrant and the two out-neighbors s and u, which have a link between them, both receive a link. Similar to case (ii) we can then partition the set $\mathcal{S}_t$ in the subset $\{s, u\}$ and the set of remaining nodes $\mathcal{S}_t \backslash \{s, u\}$. The probability of both s and u receiving a link by the entrant follows a bivariate hypergeometric distribution as $\binom{2}{2}\binom{|\mathcal{S}_t|-2}{m-2}/\binom{|\mathcal{S}_t|}{m}$. The probability that node v is observed directly is n_s/t. The number of such triangles including node s is given by $M_s(t)$ (in both G_t and its closure $\overline{G}_t$). The expected number of triangles being formed in this way is then given as

$$M_s(t)\frac{n_s}{t}\frac{\binom{2}{2}\binom{|\mathcal{S}_t|-2}{m-2}}{\binom{|\mathcal{S}_t|}{m}} = \frac{M_s(t)}{t}\frac{m(m-1)}{(m+1)(n_s(m+1)-1)}.$$

Taking together the cases (i)–(iii), we can write in the continuum approximation for the dynamics of $M_s(t)$ that

$$\begin{aligned}\frac{dM_s(t)}{dt} &= \frac{a(m-1)}{t(n_s(m+1)-1)}(a(m+1)+k_s(t)+M_s(t)) \\ &= \frac{a(m-1)}{t(n_s(m+1)-1)}\left(a(m+1)-1+\left(\frac{t}{s}\right)^a+M_s(t)\right),\end{aligned}$$

where we have denoted $a = m/(m+1)$ and used the fact that $k_s(t) = (t/s)^a - 1$ in the continuum approximation in (21). Further denoting $b = (a(m-1))/(n_s(m+1)-1)$ we can write this as

$$\frac{dM_s(t)}{dt} = \frac{b}{t}\left(m-1+\left(\frac{t}{s}\right)^a+M_s(t)\right). \tag{27}$$

The general solution of (27) is given by

$$M_s(t) = \frac{1}{a-b}\left((b-a)(m-1)+b\left(\frac{t}{s}\right)^a+(a(m-1)-bm+(a-b)M_s(s))\left(\frac{t}{s}\right)^b\right). \tag{28}$$

From (28) we can obtain an upper and a lower bound for the number of triangles involving node s, i.e., $\underline{M}_s(t) \leq M_s(t) \leq \overline{M}_s(t)$, by noting that $0 \leq M_s(s) \leq \binom{m}{2}$. For the lower bound we set $M_s(s) = 0$ and obtain

$$\underline{M}_s(t) = \frac{a(m-1)\left(\left(\frac{t}{s}\right)^b-1\right)+b\left(m-1+\left(\frac{t}{s}\right)^a-m\left(\frac{t}{s}\right)^b\right)}{a-b}.$$

Similarly, for the upper bound we set $M_s(s) = \binom{m}{2}$. Then we get

$$\overline{M}_s(t) = \frac{2a(1-m)+(a(m(m+1)-2)-bm(m+1))\left(\frac{t}{s}\right)^b+2b\left(m-1+\left(\frac{t}{s}\right)^a\right)}{2(a-b)}.$$

From (21) we know that $s = t(1+k)^{-1/a}$. Inserting this into $\overline{M}_s(t)$ and $\underline{M}_s(t)$, and using the fact that $C(k) = (2M_k)/((k+m)(k+m-1))$ allows us to bound the clustering coefficient as $\underline{C}(k) \leq C(k) \leq \overline{C}(k)$, where

$$\underline{C}(k) = \frac{2bk+2(a(m-1)-bm)((1+k)^{b/a}-1)}{(a-b)(k+m)(k+m-1)} \tag{29}$$

and

$$\overline{C}(k) = \frac{2a(m-1)+2b(k+m)+(a(m(m+1)-2)-bm(1+m))(1+k)^{b/a}}{(a-b)(k+m)(k+m-1)}. \tag{30}$$

For large k, these bounds decay as $O(1/k)$. Furthermore, their difference is given by

$$\overline{C}(k)-\underline{C}(k) = \frac{2b(1+k)m-(1+k)^{b/a}m(b(m+1)-a(m-1))}{(a-b)(k+m-1)(k+m)},$$

with the property that $\lim_{k\to\infty} \overline{C}(k) - \underline{C}(k) = 0$, showing that also $C(k) = O(1/k)$. □

References

Abbasi, Alireza, Jörn Altmann, and Liaquat Hossain (2011), "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures." *Journal of Informetrics*, 5, 594–607. [817]

Allen, Robert C. (1983), "Collective invention." *Journal of Economic Behavior & Organization*, 4, 1–24. [814]

Almeida, Paul and Bruce Kogut (1999), "Localization of knowledge and the mobility of engineers in regional networks." *Management Science*, 45, 905–917. [814]

Alós-Ferrer, Carlos and Simon Weidenholzer (2008), "Contagion and efficiency." *Journal of Economic Theory*, 143, 251–274. [815]

Anderson, Simon P., Andre de Palma, and Jacques-Francois Thisse (1992), *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, Massachusetts. [823]

Bala, Venkatesh and Sanjeev Goyal (2000), "A noncooperative model of network formation." *Econometrica*, 68, 1181–1229. [815, 841]

Barrat, Alain and Romualdo Pastor-Satorras (2005), "Rate equation approach for correlations in growing network models." *Physical Review E*, 71, 036127. [831, 833]

Blondel, Vincent D., Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre (2008), "Fast unfolding of communities in large networks." *Journal of Statistical Mechanics: Theory and Experiment*, 2008, P10008. [816]

Boguná, Marián and Romualdo Pastor-Satorras (2003), "Class of correlated random networks with hidden variables." *Physical Review E*, 68, 036112. [833]

Bollobás, Béla (1998), *Modern Graph Theory*. Springer, New York. [842]

Bollobás, Béla, Oliver Riordan, Joel Spencer, and Gábor Tusnády (2001), "The degree sequence of a scale-free random graph process." *Random Structures & Algorithms*, 18, 279–290. [827]

Borgatti, Stephen P. and Martin G. Everett (2000), "Models of core/periphery structures." *Social networks*, 21, 375–395. [825]

Bramoullé, Yann, Sergio Currarini, Matthew O. Jackson, Paolo Pin, and Brian W. Rogers (2012), "Homophily and long-run integration in social networks." *Journal of Economic Theory*, 147, 1754–1786. [818, 823]

Breschi, Stefano and Francesco Lissoni (2005), "'Cross-firm' inventors and social networks: Localized knowledge spillovers revisited." *Annales d'Économie et de Statistique*, 79/80, 189–209. [815, 817]

Brooks, Stephen P. and Gareth O. Roberts (1998), "Assessing convergence of Markov chain Monte Carlo algorithms." *Statistics and Computing*, 8, 319–335. [840]

Calvó-Armengol, Antoni and Joan De Martí (2007), “Communication networks: Knowledge and decisions.” *American Economic Review*, 97, 86–91. [814]

Chaney, Thomas (2014), “The network structure of international trade.” *American Economic Review*, 104, 3600–3634. [818]

Chernozhukov, Victor and Han Hong (2003), “An MCMC approach to classical estimation.” *Journal of Econometrics*, 115, 293–346. [839]

Chib, Siddhartha (2001), “Markov chain Monte Carlo methods: Computation and inference.” In *Handbook of Econometrics* (James J. Heckman and Edward Leamer, eds.), 3569–3649, Elsevier Science, Amsterdam, The Netherlands. [839, 840]

Cooper, Colin and Alan Frieze (2003), “A general model of web graphs.” *Random Structures & Algorithms*, 22, 311–335. [839, 841]

Dorogovtsev, Sergei N. and José F. F. Mendes (2013), *Evolution of Networks: From Biological Nets to the Internet and WWW*. Oxford University Press, New York. [831, 850]

Ductor, Lorenzo (2015), “Does co-authorship lead to higher academic productivity?” *Oxford Bulletin of Economics and Statistics*, 77, 385–407. [817, 820]

Durrett, Rick (2007), *Random Graph Dynamics*. Cambridge University Press, New York. [814]

Dutta, Bhaskar, Sayantan Ghosal, and Debraj Ray (2005), “Farsighted network formation.” *Journal of Economic Theory*, 122, 143–164. [819, 822]

Fafchamps, Marcel, Marco J. Van der Leij, and Saneev Goyal (2006), “Scientific networks and co-authorship.” Department of Economics Dicussion paper series, University of Oxford. [836]

Fafchamps, Marcel, Marco J. Van der Leij, and Sanjeev Goyal (2010), “Matching and network effects.” *Journal of the European Economic Association*, 8, 203–231. [814, 816]

Fleming, Lee, Charles King III, and Adam I. Juda (2007), “Small worlds and regional innovation.” *Organization Science*, 18, 938–954. [814]

Frank, Ove (1977), “Survey sampling in graphs.” *Journal of Statistical Planning and Inference*, 1, 235–264. [815, 821]

Friedkin, Noah E. (1983), “Horizons of observability and limits of informal control in organizations.” *Social Forces*, 62, 54–77. [815, 822]

Galeotti, Andrea and Sanjeev Goyal (2010), “The law of the few.” *American Economic Review*, 100, 1468–1492. [825]

Galeotti, Andrea, Sanjeev Goyal, Matthew O. Jackson, Fernando Vega-Redondo, and Leeat Yariv (2010), “Network games.” *Review of Economic Studies*, 77, 218–244. [815, 822]

Geweke, John (1992), “Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments.” In *Bayesian Statistics 4* (José M. Bernardo and Morris H. DeGroot, eds.), 169–193, Oxford University Press, New York. [840]

Ghiglino, Christian (2011), "Random walk to innovation: Why productivity follows a power law." *Journal of Economic Theory*, 147, 713–737. [818]

Goodman, Leo A. (1961), "Snowball sampling." *Annals of Mathematical Statistics*, 32, 148–170. [815]

Goyal, Sanjeev (2007), *Connections: An Introduction to the Economics of Networks*. Princeton University Press, Princeton, New Jersey. [814]

Goyal, Sanjeev and Sumit Joshi (2006), "Unequal connections." *International Journal of Game Theory*, 34, 319–349. [825]

Goyal, Sanjeev and José L. Moraga-González (2001), "R&D networks." *RAND Journal of Economics*, 32, 686–707. [815]

Goyal, Sanjeev, Marco J. Van der Leij, and Jose Luis Moraga-González (2006), "Economics: An emerging small world." *Journal of Political Economy*, 114, 403–412. [814, 816]

Grimmett, Geoffrey R. and David R. Stirzaker (2001), *Probability and Random Processes*, third edition. Oxford University Press, Oxford. [849]

Hagenbach, Jeanne and Frédéric Koessler (2010), "Strategic communication networks." *Review of Economic Studies*, 77, 1072–1099. [814]

Hall, Bronwyn H., Adam B. Jaffe, and Manuel Trajtenberg (2001), "The NBER patent citation data file: Lessons, insights and methodological tools." NBER Working Paper 8498. [836]

Jackson, Matthew O. (2008), *Social and Economic Networks*. Princeton University Press, Princeton, New Jersey. [814]

Jackson, Matthew O. and Brian W. Rogers (2007), "Meeting strangers and friends of friends: How random are social networks?" *American Economic Review*, 97, 890–915. [814, 815, 817, 818, 820, 821, 822, 830, 831, 834, 836]

Jackson, Matthew O. and Asher Wolinsky (1996), "A strategic model of social and economic networks." *Journal of Economic Theory*, 71, 44–74. [815]

Jones, Benjamin F., Stefan Wuchty, and Brian Uzzi (2008), "Multi-university research teams: Shifting impact, geography, and stratification in science." *Science*, 322, 1259–1262. [817, 841]

Jordan, Jonathan (2006), "The degree sequences and spectra of scale-free random graphs." *Random Structures & Algorithms*, 29, 226–242. [846]

Kolaczyk, Eric D. (2009), *Statistical Analysis of Network Data: Methods and Models*. Springer-Verlag, New York. [815, 821]

Kolotilin, Anton (2013), "Estimation of a scale-free network formation model." Working Paper. [839]

König, Michael D. (2011), "The formation of networks with local spillovers and limited observability." SIEPR Discussion Paper 11-004. [836]

König, Michael D., Stefano Battiston, Mauro Napoletano, and Frank Schweitzer (2012), "The efficiency and stability of R&D networks." *Games and Economic Behaviors*, 75, 694–713. [820, 825]

König, Michael D., Claudio J. Tessone, and Yves Zenou (2009), "A dynamic model of network formation with strategic interactions." CEPR Discussion Paper DP7521. [825]

Kovářík, Jaromír and Marco J. Van der Leij (2014), "Risk aversion and networks." *Review of Network Economics*, 13, 121–155. [818]

Krapivsky, Paul L. and Sidney Redner (2001), "Organization of growing random networks." *Physical Review E*, 63, 066123. [817]

Krapivsky, Paul L., Sidney Redner, and Francois Leyvraz (2000), "Connectivity of growing random networks." *Physical Review Letters*, 85, 4629–4632. [817]

Kumar, Ravi, Prabhakar Raghavan, Sridhar Rajagopalan, D. Sivakumar, Andrew Tomkins, and Eli Upfal (2000), "Stochastic models for the Web graph." In *Proceedings of the 41st Annual Symposium on the Foundations of Computer Science, 2000*, 57–65, IEEE, Redondo Beach, California. [817]

Lai, Ronald, Alexander D'Amour, and Lee Fleming (2009), "The careers and coauthorship networks of U.S. patent-holders, since 1975." Harvard Business School, Harvard Institute for Quantitative Social Science. [836]

Leskovec, Jure, Jon Kleinberg, and Christos Faloutsos (2007), "Graph evolution: Densification and shrinking diameters." *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1, Article No. 2. [837]

Marjoram, Paul, John Molitor, Vincent Plagnol, and Simon Tavaré (2003), "Markov chain Monte Carlo without likelihoods." *Proceedings of the National Academy of Sciences*, 100, 15324–15328. [839]

Marshall, Alfred (1919), *Industry and Trade*. McMillan, London. [814]

Mas-Colell, Andreu, Michael Dennis Whinston, and Jerry R. Green (1995), *Microeconomic Theory*. Oxford University Press, New York. [845]

McBride, Michael (2006), "Imperfect monitoring in communication networks." *Journal of Economic Theory*, 126, 97–119. [815, 822]

McFadden, Daniel (1981), "Econometric models of probabilistic choice." In *Structural Analysis of Discrete Data With Econometric Applications* (Charles F. Manski and Daniel McFadden, eds.), 198–272, The MIT Press, Cambridge, Massachusetts. [823]

McFadden, Daniel (1989), "A method of simulated moments for estimation of discrete response models without numerical integration." *Econometrica*, 57, 995–1026. [839]

Móri, Tamás F. (2005), "The maximum degree of the Barabási–Albert random tree." *Combinatorics, Probability and Computing*, 14, 339–348. [826]

Newman, Mark (2001a), "The structure of scientific collaboration networks." *Proceedings of the National Academy of Sciences*, 98, 404–409. [814]

Newman, Mark (2001b), "Clustering and preferential attachment in growing networks." *Physical Review E*, 64, 025102. [815, 817, 820]

Newman, Mark (2002), "Assortative mixing in networks." *Physical Review Letters*, 89, 208701. [814, 832, 836, 837]

Newman, Mark (2004), "Coauthorship networks and patterns of scientific collaboration." *Proceedings of the National Academy of Sciences*, 101, 5200–5205. [814]

Newman, Mark (2010), *Networks: An Introduction*. Oxford University Press, New York. [814]

Oliveira, Roberto and Joel Spencer (2005), "Connectivity transitions in networks with super-linear preferential attachment." *Internet Mathematics*, 2, 121–163. [817]

Pakes, Ariel and David Pollard (1989), "Simulation and the asymptotics of optimization estimators." *Econometrica*, 57, 1027–1057. [839]

Pastor-Satorras, Romualdo, Alexei Vázquez, and Alessandro Vespignani (2001), "Dynamical and correlation properties of the Internet." *Physical Review Letters*, 87, 258701. [831, 832, 837]

Ratmann, Oliver, Ole Jørgensen, Trevor Hinkley, Michael Stumpf, Sylvia Richardson, and Carsten Wiuf (2007), "Using likelihood-free inference to compare evolutionary dynamics of the protein networks of h. pylori and p. falciparum." *PLoS Computational Biology*, 3, e230. [839]

Robert, Christian P. and George Casella (2004), *Monte Carlo Statistical Methods*. Springer-Verlag, New York. [839]

Saxenian, AnnaLee (1994), *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge, Massachusetts. [814]

Schilling, Melissa A. and Elad Green (2011), "Recombinant search and breakthrough idea generation: An analysis of high impact papers in the social sciences." *Research Policy*, 40, 1321–1331. [816, 841]

Singh, Jasjit (2005), "Collaborative networks as determinants of knowledge diffusion patterns." *Management Science*, 51, 756–770. [814, 815, 817]

Sisson, Scott A. and Yanan Fan (2011), "Likelihood-free MCMC." In *Handbook of Markov Chain Monte Carlo* (Steve Brooks, Andrew Gelman, Galin L. Jones, and Xiao-Li Meng, eds.), Chapman & Hall/CRC, New York. [839]

Snijders, Tom A. B. (2001), "The statistical evaluation of social network dynamics." *Sociological Methodology*, 31, 361–395. [822]

Snijders, Tom A. B., Johan Koskinen, and Michael Schweinberger (2010), "Maximum likelihood estimation for social network dynamics." *The Annals of Applied Statistics*, 4, 567–588. [822]

Sokal, Alan D. (1996), "Monte Carlo methods in statistical mechanics: Foundations and new algorithms." In *Functional Integration: Basics and Application* (Cecile DeWitt-Morette, Pierre Cartier, and Antoine Folacci, eds.), 131–192, Springer-Verlag, New York. [840]

Stephan, Paula E. (2012), *How Economics Shapes Science*. Harvard University Press, Cambridge, Massachusetts. [841]

Toivonen, Riitta, Jukka-Pekka Onnela, Jari Saramäki, Jörkki Hyvönen, and Kimmo Kaski (2006), "A model for social networks." *Physica A: Statistical Mechanics and its Applications*, 371, 851–860. [817]

Valverde, Sergi, Ricard V. Solé, Mark A. Bedau, and Norman Packard (2007), "Topology and evolution of technology innovation networks." *Physical Review E*, 76, 056118. [814]

Van Mieghem, Piet (2011), *Graph Spectra for Complex Networks*. Cambridge University Press, New York. [832]

Vázquez, Alexei (2003), "Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations." *Physical Review E*, 67, 056104. [817]

Vega-Redondo, Fernando (2007), *Complex Social Networks*. Cambridge University Press, New York. [814]

Vigier, Adrien (2014), "Meeting friends of friends and homophily: A complementarity." *Economic Theory Bulletin*, 2, 45–52. [818, 823, 835]

von Hippel, Eric, Stefan Thomke, and Mary Sonnack (1999), "Creating breakthroughs at 3M." *Harvard Business Review*, 77, 47–57. [815]

Wang, Li-Na, Jin-Li Guo, Han-Xin Yang, and Tao Zhou (2009), "Local preferential attachment model for hierarchical networks." *Physica A: Statistical Mechanics and its Applications*, 388, 1713–1720. [817]

Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks." *Nature*, 393, 440–442. [814, 831, 833]

West, Douglas B. (2001), *Introduction to Graph Theory*, second edition. Prentice-Hall, Upper Saddle River, New Jersey. [842]

Westbrock, Bastian (2010), "Natural concentration in industrial research collaboration." *RAND Journal of Economics*, 41, 351–371. [815, 820]

Winkler, Anne E., Wolfgang Glänzel, Sharon G. Levin, and Paula E. Stephan (2011), "The diffusion of information technology and the increased propensity of teams to transcend institutional and national borders." IZA Discussion Paper 5857. [817]

Co-editor Nicola Persico handled this manuscript.

Submitted 2013-5-1. Final version accepted 2015-9-2. Available online 2015-9-3.