This work is distributed as a Discussion Paper by the

**STANFORD INSTITUTE FOR ECONOMIC POLICY RESEARCH**

SIEPR Discussion Paper No. 11-004

# The Formation of Networks with Local Spillovers and Limited Observability

by
Michael D. Koenig

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

# The Formation of Networks with Local Spillovers and Limited Observability[☆]

Michael D. König[a]

[a] *SIEPR and Department of Economics, Stanford University, 579 Serra Mall, CA 94305-6072, United States.*

**Abstract**

In this paper I analyze 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 through 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 coinventors, firm alliances and trade relationships between countries, and find that the model can closely reproduce the observed patterns. The estimates show that with increasing levels of aggregation, the observation radius is increasing, indicating economies of scale in which larger organizations are able to process greater amounts of information.

*Key words:* diffusion, network formation, growing networks, limited observability
*JEL:* 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 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

[☆] 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 Yves Zenou, Ben Golub, Tomás R. Barraquer, and seminar participants at University of Vienna, University of Bielefeld, University of Zurich, ETH Zurich and Stanford University for their insightful comments. Financial support from Swiss National Science Foundation through research grant PBEZP1–131169 is gratefully acknowledged. A previous version of this paper was circulated under the title, "Centrality Based Network Formation of Boundedly Rational Agents with Limited Information".

*Email address:* `mdkoenig@stanford.edu` (Michael D. König)

[1] This literature has steadily grown in the last decade. The monographs of Jackson (2008), Goyal (2007), and Vega-Redondo (2007) are excellent surveys contrasting this literature with 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.

diverse actors, ranging from individuals to firms and countries. In this paper I introduce a simplistic 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 at different levels of aggregation.

On an individual level, a large body of literature has emphasized the detrimental effect of social networks of inventors on the productivity of innovative regions (see e.g. Allen, 1983, Almeida and Kogut, 1999, Marshall, 1919, Singh, 2005). A prominent example is the success story of Silicon Valley, which has been attributed to its informal networks of friendship and collaboration (Fleming et al., 2007, Saxenian, 1994). On the organizational level, R&D partnerships between firms have become a widespread phenomenon characterizing technology diffusion and dynamics (Fischer, 2006, Gulati, 2007, Hagedoorn, 2002, Nooteboom, 2004), especially in industries with rapid technological development such as the biotech and computer industries (see Ahuja, 2000, Powell et al., 2005, Riccaboni and Pammolli, 2002, Roijakkers and Hagedoorn, 2006). In R&D partnerships firms exchange information about new products or technologies and diffuse knowledge throughout the economy. On the aggregate level of countries, the spread and diffusion of technologies is a key factor for explaining economic growth (Bitzer and Geishecker, 2006, Franco et al., 2011, Grossman and Helpman, 1995). The basic idea is that economic growth in relatively backward economies takes the form of adoption and imitation of existing technologies (Kuznets, 1969). Imitation and innovation are affected by technology diffusion, trade and interdependencies, and these factors are crucial for the growth process.[2]

In this paper I identify a number of common empirical regularities shared by the networks of inventors, firms and countries, some of which have been documented already in the literature. First, the distributions of degree (the number of links of a node) in these networks exhibit fat tails, typically decaying as a power-law.[3] 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.[4] Moreover, the distribution of (small) connected components (in which there exists a path between every pair of nodes) follows a power-law decay. However, the average degree of the neighbors of a node varies among these networks. While the network of inventors exhibits an increasing average neighbors' degree with the degree of a node, this correlation is almost absent in the network of firms, and it is decreasing in the network of trade relationships between countries (cf. Serrano and Boguñá, 2003). The first is referred to as "assortativity" while the latter refers to "dissortativity" (Newman, 2002). In this paper I introduce a simple model that can reproduce all these empirical distributions and further gives an explanation for the variations observed in the neighbors connectivity.

[2] See e.g. Coe and Helpman (1995), Acemoglu (2009) and Aghion and Howitt (2009).

[3] A power-law degree distribution in patent citation networks has been documented in e.g. Brantle and Fallah (2007), Valverde et al. (2007), in the network of R&D collaborating firms in Gay and Dousset (2005), Powell et al. (2005) and the network of trade in Fagiolo et al. (2009), Serrano and Boguñá (2003).

[4] Goyal et al. (2006) make a similar observation in the network of scientific coauthorships among economists, and Serrano and Boguñá (2003) in the network of trade.

I consider 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. 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 derives from having access to the information of others. The noise term captures exogenous random perturbances, shortcomings in assessing the correct value of information possessed by other agents and exogenous matching effects.

Agents sequentially enter the network and obtain an opportunity to acquire information from the incumbent agents. 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).[5] 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 obtained for 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 which is difficult to observe called "snowball/star sampling" (Frank, 1977, Goodman, 1961, Kolaczyk, 2009).[6]

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.[7] 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".

I then estimate the model using three different empirical networks that can be regarded as a proxy for the underlying network of information transmission and knowledge diffusion at different levels of aggregation: a network of coinventors from patents in the drug development sector, firm alliances in the biotech sector and a network of trade relationships between countries. Notably, I find that the model can closely match all the observed distributions for the degree, clustering-degree, nearest neighbor average degree and the component size distribution. Furthermore, estimating the model's parameters for these networks shows that with increasing levels of aggregation the observation radius is increasing. This indicates the presence of *economies of scale*: larger organizational units are able to process greater amounts of information, as compared to the limited capacities indi-

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

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

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

viduals typically face for observation, communication and information processing (cf. Radner, 1992, Radner and Van Zandt, 1992, Wilson, 1975).

The paper in the economics literature most closely related to the one presented here is Jackson and Rogers (2007).[8] The authors introduce a model of a growing network which 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. Their model and method of analysis shares many features of a vast literature originating from statistical physics. As common in this literature, their process of network formation is rather mechanical, and a serious deficiency of this literature is the lack of a sound micro-foundation. One contribution of this work is that it starts directly from a discrete-choice approach, 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 processing capabilities of 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 approxkimation 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 or the component size distribution), and do not analyze the impact of different observation radii on these statistics, in particular, the transition from assortative to dissortative networks. 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. This indicates that their efficiency analysis is not robust under a degree dependent payoff function and the presence of noise.

Based on the 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 and Rogers (2009) introduce different types of agents and study the mechanisms underlying homophily, that is, the tendency of similar types of agents being connected. Moreover, Kovarik and van der Leij (2009) 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 (2011) 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 of why links are formed in the model by Jackson and Rogers (2007) in the context of knowledge diffusion in

[8] 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 and Redner (2001), Krapivsky et al. (2000), Oliveira and Spencer (2005), Vazquez (2003), Kumar et al. (2000), Wang et al. (2009) and Toivonen et al. (2006), to mention only a view. 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.

networks. Moreover, none of these works investigates 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 6 discusses several extensions of the model. Section 7 contains an empirical application of the model to different real world networks. Finally, in Section 8 I conclude. All proofs are relegated to Appendix A.

## 2. The Model

The network is modeled as a *directed graph* (unless otherwise stated), which is a pair $G \equiv \langle \mathcal{N}, \mathcal{E} \rangle$, where $\mathcal{N} \equiv \{1, \ldots, n\}$ 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 which 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 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 $\bar{d}_G \equiv \frac{1}{n} \sum_{i \in \mathcal{N}} d_G(i)$ and the degree variance is given by $\sigma_d^2(G) \equiv \frac{1}{n} \sum_{i \in \mathcal{N}} (d_G(i) - \bar{d}_G)^2$. Following Bala and Goyal (2000) I define the *closure* of a graph $G$, denoted by $\bar{G}$, by the condition $ij \in \mathcal{E}(\bar{G}) \Leftrightarrow ij \in \mathcal{E}(G) \vee ji \in \mathcal{E}(G)$. 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 $e(\bar{G})$ in the closure $\bar{G}$ is given by $e(\bar{G}) = \frac{1}{2} \sum_{i \in \mathcal{N}} d_G(i)$. 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 $\mathcal{E}$.

With these definitions at hand, we are now able to introduce the payoff agents derive from being connected in a network and their incentives to form links in the following sections.

### 2.1. Payoffs

For a given network $G = \langle \mathcal{N}, \mathcal{E} \rangle \in \mathcal{G}(n)$ we assign each each agent $i \in \mathcal{N}$ a payoff $\pi_i(\cdot, \delta) : \mathcal{G}(n) \to \mathbb{R}$ which depends on the network $G$ and a parameter $\delta \geq 0$ which measures the degree of interdependency between agents' payoffs in $G$ (we will encounter more specific examples below). 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

$$f_i(G, j) \equiv \pi_i(G \oplus ij, \delta) - \pi_i(G, \delta), \tag{2.1}$$

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) which satisfy the following conditions:

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

*(LM) Link monotonicity: $f_i(G, j) \geq 0$ for all $j \neq i \in \mathcal{N}$.*

*(LD) Linear differences: 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*

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

*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 non-negative. Intuitively it says that no link to be formed can harm an agent (cf. Dutta et al., 2005). Condition (LD), degree linearity, 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. Two potential links are judged as being equally attractive for the agent if the involved agents have the same degree in the current network.

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 $\Pi(G, \delta) = \sum_{i \in \mathcal{N}} \pi_i(G, \delta)$ and let $\sigma_d^2(G)$ be the degree variance of $G \in \mathcal{G}(n, e)$. Then we assume that*

*(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) implies that networks with a higher degree inequality, as measured by the degree variance, generate higher welfare.

In the following I give examples from the literature which satisfy the above assumptions.

#### 2.1.1. Information Diffusion in Networks

Following Fafchamps et al. (2010) I consider agents that exchange information in a network $G$, where information that travels longer paths is discounted by a factor $\delta \in [0, 1]$. It is assumed that information can travel both ways of a link and so I consider the (undirected) paths in the closure $\bar{G}$ of

$G$. The probability that an agent $j$ transmits information along a given path in $\bar{G}$ is independent of the probability that the same agent $j$ transmits the same information along another path. With this assumption, the probability that agent $i$ receives the information over distance $k$ when there are $c_{ij}^k(\bar{G})$ (undirected) paths of length $k$ connecting $i$ to $j$ becomes

$$P_{ij}^{\delta}(G) \equiv 1 - \prod_{k=1}^{\infty}(1-\delta^k)^{c_{ij}^k(\bar{G})}.$$

The payoff $\pi_i : \mathcal{G}(n) \times \mathbb{R}_+ \to \mathbb{R}$ of agent $i$ is defined as $\pi_i(G,\delta) \equiv V\sum_{j\in\mathcal{N}} P_{ij}^{\delta}(G) - cd_G^+(i)$ with $V > 0$ and a fixed cost $c \in [0, V\delta)$ for each link the agent has initiated. When the decay parameter $\delta$ is sufficiently small, we can write $(1-\delta^k)^c \approx 1 - c\delta^k$. With this approximation the payoff of agent $i$ becomes

$$\pi_i(G,\delta) \equiv V\sum_{j\in\mathcal{N}}\left(1 - \prod_{k=1}^{\infty}(1-\delta^k)^{c_{ij}^k(\bar{G})}\right) - cd_G^+(i) = V\left(\delta d_G(i) + \delta^2 \sum_{j\in\mathcal{N}_G(i)} d_G(j)\right) + O(\delta^3) - cd_G^+(i).$$

It then follows that the link incentive function is given by

$$f_i(G,j) = V\delta - c + V\delta^2 d_G(j) + O(\delta^3).$$

Link monotonicity (LM) holds if $c < V\delta$ and degree monotonicity (LD) holds for $g(x) = Vx$ and $\gamma = 2$, since $f_i(G,j) - f_i(G,k) = V\delta^2(d_G(j) - d_G(k)) + O(\delta^3)$. As our measure of welfare we consider aggregate payoff given by

$$\begin{aligned}
\Pi(G,\delta) &= V\delta\sum_{i\in\mathcal{N}} d_G(i) + V\delta^2\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}_G(i)} d_G(j) + O(\delta^3) - c\sum_{i\in\mathcal{N}} d_G^+(i) \\
&= (2V\delta - c)e(\bar{G}) + V\delta^2\sum_{i\in\mathcal{N}} d_G(i)^2 + O(\delta^3) \\
&= (2V\delta - c)e(\bar{G}) + \frac{4V\delta^2}{n}e(\bar{G})^2 + V\delta^2 n\sigma_d^2(G) + O(\delta^3)
\end{aligned}$$

where we have used the fact that $\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}_G(i)} d_G(j) = \sum_{i\in\mathcal{N}} d_G(i)^2$. The average degree is $\bar{d} = \frac{1}{n}\sum_{i=1}^n d_G(i) = \frac{2e(\bar{G})}{n}$. The degree variance is given by $\sigma_d^2(G) = \frac{1}{n}\sum_{i\in\mathcal{N}}(d_G(i) - \bar{d}_G) = \frac{1}{n}\sum_{i=1}^n d_G(i)^2 - \bar{d}^2 = \frac{1}{n}\sum_{i=1}^n d_G(i)^2 - \frac{4e(\bar{G})^2}{n^2}$. It follows that for small $\delta$, such that terms of the order $O(\delta^3)$ become negligible, maximizing aggregate payoff $\Pi(G,\delta)$ (given $n$ and $e$) becomes equivalent to maximizing the degree variance $\sigma_d^2(G)$, and condition (DC) holds.

### 2.1.2. Two-Way Flow Communication

The two-way flow model with decay has been introduced by Bala and Goyal (2000). In this model links are interpreted as lines of communication between two individuals. If $i$ wants to communicate

with $j$ then $i$ must first pay a fee of $c \geq 0$ to open the channel. By creating this link $i$ does not only get access to $j$ but also to all individuals that are approachable by $j$ via an (undirected) path in the closure $\bar{G}$. Formally, the payoff function $\pi_i : \mathcal{G}(n) \times \mathbb{R}_+ \to \mathbb{R}$ of agent $i \in \mathcal{N}$ is given by[9]

$$\pi_i(G, \delta) \equiv 1 + \sum_{i \neq j} \delta^{\ell(i,j,\bar{G})} - c d_G^+(i), \tag{2.2}$$

for some $\delta \in [0, 1]$, which is interpreted as the *degree of friction* in communication. The number $\ell(i, j, \bar{G})$ is the length of the shortest path connecting agent $i$ with $j$ in the graph $\bar{G}$. If $i$ and $j$ are not connected we adopt the convention that $\ell(i, j, \bar{G}) = \infty$. The difference to the payoff function in Fafchamps et al. (2010) of the previous section and the one in Equation (2.2) is that in the latter only the shortest paths matter.

In the following we assume that the network $\bar{G}$ does not contain any cycles, i.e. it is a tree (or a forest, if the network is unconnected). Denote by $\mathcal{T}(\mathcal{N})$ the class of (undirected) tree graphs with vertex set $\mathcal{N}$. Then a tree $\bar{G} \in \mathcal{T}(\mathcal{N})$ is defined by the conditions (i) that it is connected, and (ii) $|\mathcal{E}(\bar{G})| = |\mathcal{N}| - 1$ for all $\bar{G} \in \mathcal{T}(\mathcal{N})$. When $\bar{G} \in \mathcal{T}(\mathcal{N})$, the payoff of an agent $i \in \mathcal{N}$ can be written as

$$\pi_i(G, \delta) = 1 + \delta d_G(i) + \delta^2 \sum_{j \in \mathcal{N}_G(i)} (d_G(j) - 1) + O(\delta^3) - c d_G^+(i).$$

It follows that the linking incentive function of agent $i$ takes the form

$$f_i(G, j) = \delta(1 - \delta) - c + \delta^2 d_G(j) + O(\delta^3).$$

The link incentive function satisfies condition (LM) for $\delta(1-\delta) > c$ and condition (LD) with $g(x) = x$ and $\gamma = 2$, because $f_i(G, j) - f_i(G, k) = \delta^2(d_G(j) - d_G(k)) + O(\delta^3)$. Aggregate payoff $\Pi(G, \delta) = \sum_{i \in \mathcal{N}} \pi_i(G, \delta)$ is then given by

$$\begin{aligned}\Pi(G, \delta) &= n + \delta(1 - \delta) \sum_{i \in \mathcal{N}} d_G(i) + \delta^2 \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}_G(i)} d_G(j) + O(\delta^3) - c \sum_{i \in \mathcal{N}} d_G^+(i) \\ &= n + (2\delta(1 - \delta) - c)(n - 1) + \frac{4\delta^2}{n}(n - 1)^2 + n\delta^2 \sigma_d^2(G) + O(\delta^3),\end{aligned}$$

where $e(\bar{G})$ is the number of edges in $\bar{G}$, $n = |\mathcal{N}|$, and we have used the fact that for $\bar{G} \in \mathcal{T}(\mathcal{N})$ the number of edges is $e(\bar{G}) = n - 1$. It follows that for small $\delta$ such that terms of the order $O(\delta^3)$ become negligible, maximizing aggregate payoffs becomes equivalent to maximizing the degree variance . Hence, Condition (DC) holds for aggregate payoff when $\bar{G} \in \mathcal{T}[\mathcal{N}]$.[10]

[9] See also Jackson and Wolinsky (1996) for a similar payoff structure.

[10] We will see in the network growth model introduced in Section 2.2 that $\bar{G} \in \mathcal{T}[\mathcal{N}]$ is always guaranteed to hold if we allow an entering agent to form only a single link.

### 2.1.3. Public Goods Provision

The following network game is presented in Goyal and Joshi (2006) as an extension of Bloch (1997). An (undirected) link between two agents represents an agreement to share knowledge about the production of a public good. Each agent can decide how much to invest into the public good. Denote the level of contribution of agent $i \in \mathcal{N} = \{1, \ldots, n\}$ as $x_i \in \mathbb{R}_+$. The production technology of every agent is assumed to be $c_i(x_i, G) = \frac{1}{2}\left(\frac{x_i}{d_G(i)+1}\right)^2$. The payoff function $\pi_i : \mathbb{R}^n_+ \times \mathcal{G}(n) \to \mathbb{R}$ of agent $i$ is

$$\pi_i(\mathbf{x}, G) \equiv \sum_{j \in \mathcal{N}} x_j - \frac{1}{2}\left(\frac{x_i}{d_G(i)+1}\right)^2.$$

The Nash contribution of agent $i$ is $x_i^* = (d_G(i)+1)^2$. This optimal choice of an agent induces naturally preferences over networks by inserting the value of $x_i(G)$ into the payoff function $\pi_i$. This gives us

$$\pi_i(G) \equiv \pi_i(\mathbf{x}^*, G) = \frac{1}{2}\left(d_G(i)+1\right)^2 + \sum_{j \in \mathcal{N}\setminus\{i\}} (d_G(j)+1)^2.$$

With this payoff function, the linking incentive function for an agent $i$ is given by

$$f_i(G, j) = \frac{9}{2} + 2d_G(j).$$

This obviously satisfies conditions (LM) and (LD) with $g(x) = 2x$ and $\gamma = 0$. Aggregate payoff $\Pi(G) = \sum_{i \in \mathcal{N}} \pi_i(G)$ is then given by

$$\begin{aligned}\Pi(G) &= \frac{1}{2}\sum_{i \in \mathcal{N}} (d_G(i)+1)^2 + \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}\setminus\{i\}} (d_G(j)+1)^2 \\ &= \frac{n(2n-1)}{2} + 2(2n-1)\left(1 + \frac{\delta^2}{n} e(\bar{G})\right) e(\bar{G}) + \frac{n(2n-1)\delta^2}{2}\sigma_d^2(G).\end{aligned}$$

We see that aggregate payoffs are increasing in the degree variance and condition (DC) holds.

### 2.1.4. A Linear-Quadratic Complementarity Game

We consider a simplified form of the game introduced by Ballester et al. (2006) where each agent $i \in \mathcal{N}$ in the network $G$ selects an effort level $x_i \geq 0$, $\mathbf{x} \in \mathbb{R}^n_+$ (e.g. the R&D investment of a firm or the work hours of an inventor), and receives a payoff $\pi_i : \mathbb{R}^n_+ \times \mathcal{G}(n) \times \mathbb{R}_+ \to \mathbb{R}$ of the following form

$$\pi_i(\mathbf{x}, G, \delta) \equiv x_i - \frac{1}{2}x_i^2 + \delta \sum_{j=1}^{n} a_{ij} x_i x_j, \tag{2.3}$$

where $\delta \geq 0$ and $a_{ij} \in \{0, 1\}$, $i, j \in \mathcal{N} = \{1, \ldots, n\}$ are the elements of the symmetric $n \times n$ *adjacency matrix* $\mathbf{A}$ of $\bar{G}$. This payoff function is additively separable in the idiosyncratic effort component $(x_i - \frac{1}{2}x_i^2)$ and the peer effect contribution $(\delta \sum_{j=1}^{n} a_{ij} x_i x_j)$. Payoffs display strategic

complementarities in effort levels, i.e., $\frac{\partial^2 \pi_i(\mathbf{x},G,\delta)}{\partial x_i \partial x_j} = \delta a_{ij} \geq 0$. Ballester et al. (2006) have shown that if $\delta < 1/\lambda_{\mathrm{PF}}(G)$ then the unique interior Nash equilibrium solution of the simultaneous $n$–player move game with payoffs given by Equation (2.3) and strategy space $\mathbb{R}^n_+$ is given by the Bonacich centrality $x_i^* = b_i(G,\delta)$ for all $i \in \mathcal{N}$ (Bonacich, 1987).[11] Moreover, the payoff of agent $i$ in equilibrium is given by

$$\pi_i(G,\delta) \equiv \pi_i(\mathbf{x}^*, G, \delta) = \frac{1}{2}(x_i^*)^2 = \frac{1}{2} b_i^2(G,\delta). \tag{2.4}$$

In the case of small complementarity effects, corresponding to small values of $\delta$, the Bonacich centrality of an agent $i$ can be written as

$$b_i(G,\delta) = 1 + \delta d_G(i) + \delta^2 \sum_{j \in \mathcal{N}_G(i)} d_G(j) + O(\delta^3).$$

Note that equilibrium payoff can be written as

$$\pi_i(G,\delta) = \frac{1}{2} + \delta d_G(i) + \frac{\delta^2}{2} d_G(i)^2 + \delta^2 \sum_{j \in \mathcal{N}_G(i)} d_G(j) + O(\delta^3),$$

and the link incentive function is then given by

$$f_i(G,j) = \frac{\delta(2+\delta)}{2} + \frac{\delta^2}{2} d_G(i)(d_G(i)+1) + \delta^2 d_G(j) + O(\delta^3).$$

If we neglect terms of the order $O(\delta^3)$ then the linking incentive function also satisfies condition (LM). Further, $f_i(G,j) - f_i(G,k) = \delta^2(d_G(j) - d_G(k)) + O(\delta^3)$ so that condition (LD) holds with $g(x) = x$ and $\gamma = 2$. Aggregate payoff $\Pi(G,\delta) = \sum_{i \in \mathcal{N}} \pi_i(G,\delta)$ can be written as

$$\begin{aligned}\Pi(G,\delta) &= \frac{n}{2} + \delta \sum_{i=1}^n d_G(i) + \frac{\delta^2}{2} \sum_{i=1}^n d_G(i)^2 + \delta^2 \sum_{i=1}^n \sum_{j \in \mathcal{N}_G(i)} d_G(j) + O(\delta^3) \\ &= \frac{n}{2} + 2\delta \left(1 + \frac{3\delta}{n} e(\bar{G})\right) e(\bar{G}) + \frac{3n\delta^2}{2} \sigma_d^2(G) + O(\delta^3).\end{aligned}$$

Aggregate payoff is increasing in the degree variance, and hence, condition (DC) holds.

[11] Let $\lambda_{\mathrm{PF}}(G)$ be the largest real (Perron-Frobenius) eigenvalue of the adjacency matrix $\mathbf{A}$ of the undirected network $\bar{G}$. If $\mathbf{I}$ denotes the $n \times n$ identity matrix and $\mathbf{u} \equiv (1,\ldots,1)^\top$ the $n$-dimensional vector of ones then we can define the *Bonacich centrality* as follows: If and only if $\delta < 1/\lambda_{\mathrm{PF}}(G)$ then the matrix $\mathbf{B}(G,\delta) \equiv (\mathbf{I} - \delta\mathbf{A})^{-1} = \sum_{k=0}^\infty \delta^k \mathbf{A}^k$ exists, is non-negative (see e.g. Debreu and Herstein, 1953), and the vector of Bonacich centralities is defined as $\mathbf{b}(G,\delta) \equiv \mathbf{B}(G,\delta) \cdot \mathbf{u}$. We can write the vector of Bonacich centralities as $\mathbf{b}(G,\delta) = \sum_{k=0}^\infty \delta^k \mathbf{A}^k \cdot \mathbf{u} = (\mathbf{I} - \delta\mathbf{A})^{-1} \cdot \mathbf{u}$. For the components $b_i(G,\delta)$, $i = 1,\ldots,n$, we get $b_i(G,\delta) = \sum_{k=0}^\infty \delta^k (\mathbf{A}^k \cdot \mathbf{u})_i = \sum_{k=0}^\infty \delta^k \sum_{j=1}^n \left(\mathbf{A}^k\right)_{ij}$, where $\left(\mathbf{A}^k\right)_{ij}$ is the $ij$-th entry of $\mathbf{A}^k$. Because $\sum_{j=1}^n \left(\mathbf{A}^k\right)_{ij}$ is the number of all (undirected) walks of length $k$ in $\bar{G}$ starting from $i$, $b_i(G,\delta)$ is the number of all walks in $\bar{G}$ starting from $i$, where the walks of length $k$ are weighted by their geometrically decaying factor $\delta^k$.

## 2.2. The Network Formation Process

In this section I introduce the formation of the network. We consider a discrete time, non-stationary Markov chain $(G_t = \langle \mathcal{N}_t, \mathcal{E}_t \rangle)_{t \in \{1,2,\ldots,T\}}$ for some $T \in \mathbb{N} \cup \{\infty\}$, defining a nested sequence of graphs $G_1 \subset G_2 \subset \ldots G_T \in \mathcal{G}(T)$ 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 (cf. Jackson and Rogers, 2007).

Process time $t \in [T] \equiv \{1, 2, \ldots, T\}$ 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 if 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 $\mathcal{P}_{m+1} = \{1, 2, \ldots, m+1\}$, 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 $G_{t-1}$ be 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). An illustration is shown in Figure 1.[12]

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

[12] Further note that we assume that link formation follows a sampling procedure without replacement. Would we allow for sampling with replacement, multiple links could be created to the same agent.

to the model of Jackson and Rogers (2007),[13] 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 $\mathcal{S}_t$. We assume that this decision is made in a myopic way.[14] We assume that an entrant $t$ chooses to link to the an 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)

$$f_t(G_{t-1}, j) + \varepsilon_{ij}. \tag{2.5}$$

The term $\varepsilon_{ij}$ is an exogenous random variable, indicating the part of the agent's preference that is not represented by the systematic component $f_i(G, j)$. 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 $\varepsilon_{ij}$ are independent and identically distributed for all $i, j$. When these exogenous matching effects are weak and $\delta \to 0$, Equation (2.5) 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.

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 $G_1 = \ldots = G_{m+1} = K_{m+1}$, for all $t > m + 1$ the graph $G_t$ is obtained from $G_{t-1}$ by applying the following steps:*

**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 $S_t$ 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 $f_t(G_{t-1}, j) + \varepsilon_{tj}$.*

Let $\mathcal{R}_t \subseteq \mathcal{S}_t$, $|\mathcal{R}_t| = m$, be the set of agents that receive a link from the entrant at time $t$. The network at time $t$ is then given by $G_t = \langle \mathcal{P}_{t-1} \cup \{t\}, \mathcal{E}_{t-1} \cup \{tj : j \in \mathcal{R}_t\}\rangle$. We define the attachment

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

[14] With this we 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).

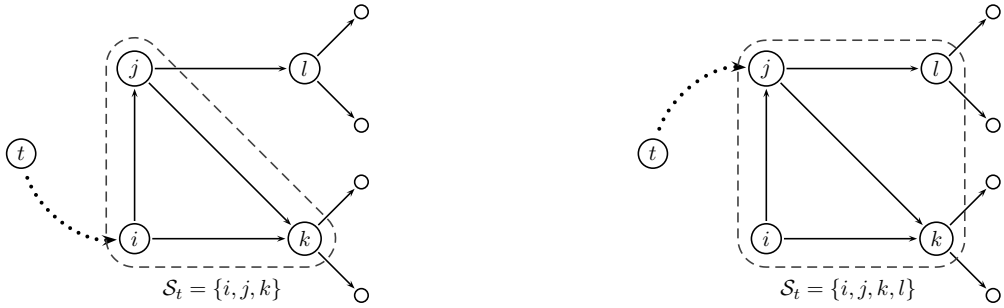



Figure 1: (Left panel) In the first draw, the entering agent $t$ observes agent $i$ and its out-neighbors $j, k$. The observed sample is $\mathcal{S}_t = \{i, j, k\}$. (Right panel) In the second draw, agent $t$ observes also agent $j$ and the out-neighborhood $\{k, l\}$ of $j$. The observed sample is then $\mathcal{S}_t = \{i, j, k, l\}$.

kernel as the probability that an agent $j \in \mathcal{P}_{t-1}$ receives a link from the entrant

$$\begin{aligned} K_t^{\beta}(j|G_{t-1}) \equiv \mathbb{E}_t[\mathbb{1}_{\mathcal{R}_t}(j)|G_{t-1}] &= \sum_{\mathcal{S}_t \subseteq \mathcal{P}_{t-1}} \sum_{\mathcal{R}_t \subseteq \mathcal{S}_t} \mathbb{1}_{\mathcal{R}_t}(j) \mathbb{P}_t(\mathcal{S}_t, \mathcal{R}_t|G_{t-1}) \\ &= \sum_{\mathcal{S}_t \subseteq \mathcal{P}_{t-1}} \underbrace{\sum_{\mathcal{R}_t \subseteq \mathcal{S}_t} \mathbb{1}_{\mathcal{R}_t}(j) \mathbb{P}_t(\mathcal{R}_t|\mathcal{S}_t, G_{t-1})}_{\equiv K_t^{\beta}(j|\mathcal{S}_t, G_{t-1})} \mathbb{P}_t(\mathcal{S}_t|G_{t-1}), \end{aligned}$$

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. Since the entrant forms links to the agents that maximize his link incentive function plus a random element, 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. The corresponding probability can be written as follows[15]

$$\begin{aligned} K_t^{\beta}(j|\mathcal{S}_t, G_{t-1}) = \sum_{l=1}^{m} \sum_{i_1, i_2, \ldots, i_{l-1}} \prod_{r=1}^{l-1} \mathbb{P}_t \left( f_t(G_{t-1}, i_r) + \varepsilon_{t,i_r} = \max_{k \in \mathcal{S}_t \backslash \{i_1, \ldots, i_r\}} f_t(G_{t-1}, k) + \varepsilon_{t,k} \right) \\ \times \mathbb{P}_t \left( f_t(G_{t-1}, j) + \varepsilon_{t,j} = \max_{k \in \mathcal{S}_t \backslash \{i_1, \ldots, i_{l-1}\}} f_t(G_{t-1}, k) + \varepsilon_{t,k} \right) \mathbb{1}_{\mathcal{S}_t}(j), \quad (2.6) \end{aligned}$$

with indices $i_1 \in \mathcal{S}_t \backslash \{j\}$, $i_2 \in \mathcal{S}_t \backslash \{j, i_1\}$, $i_3 \in \mathcal{S}_t \backslash \{j, i_1, i_2\}$, $\ldots$, $i_{l-1} \in \mathcal{S}_t \backslash \{j, i_1, i_2, \ldots, i_{l-2}\}$ and

[15] We assume that the entrant does not update the link incentive functions while forming links but evaluates it only once after he has observed the sample. The first sum in Equation (2.6) considers the case that agent $j$ receives a link in the $l$-th round while the second sum takes into account all possible sequences of agents $i_1, i_2, \ldots, i_{l-1}$ that receive a link in the $l-1$ previous rounds.

$1 \leq l \leq m$. In the following I assume that the exogenous random terms $\varepsilon_{tj}$ are identically and independently type I extreme value distributed (or Gumbel distributed) with parameter $\eta$.[16] This assumption is commonly made in random utility models in econometrics (see e.g. McFadden, 1981). Under this distributional assumption, 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)

$$\begin{aligned}
\mathbb{P}_t\left(f_t(G_{t-1}, j) + \varepsilon_{tj} = \max_{k \in \mathcal{S}_t} f_t(G_{t-1}, k) + \varepsilon_{tk}\right) &= \frac{e^{\eta f_t(G_{t-1}, j)}}{\sum_{k \in \mathcal{S}_t} e^{\eta f_t(G_{t-1}, k)}} \\
&= \frac{1}{\sum_{k \in \mathcal{S}_t} e^{-\eta(f_t(G_{t-1}, j) - f_t(G_{t-1}, k))}} \\
&= \frac{1}{\sum_{k \in \mathcal{S}_t} e^{-\eta\delta^b(d_{G_{t-1}}(j) - d_{G_{t-1}}(k)) + o(\delta^b)}} \\
&\approx \frac{e^{\beta d_{G_{t-1}}(j)}}{\sum_{k \in \mathcal{S}_t} e^{\beta d_{G_{t-1}}(k)}},
\end{aligned} \tag{2.7}$$

where we have applied condition (LD) for the link incentive function $f_t(G_{t-1}, \cdot)$, dropped terms of the order $o(\delta^b)$ and denoted by $\beta \equiv \eta\delta^b$. Knowledge of the selection probability in Equation (2.7) will allow us to analyze the network formation process introduced in Definition 1. As I will show in the following sections, this process gives rise to different network topologies, depending on the extent of the noise $\varepsilon_{tj}$, as measured by the scaling parameter $\beta$ 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 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. Moreover, as $\beta$ 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 $\beta$ tends to zero, then the noise term dominates and agents form links to the ones observed in $\mathcal{S}_t$ at random. These different parameter regions are indicated in Figure 2. In the following sections I give a more detailed account of the emerging networks depending on the level of noise scaled by $\beta$ 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 which 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

[16] The cumulative distribution function is given by $\mathbb{P}(\varepsilon \leq c) = \exp(-\exp(-\eta c - \gamma))$, where $\gamma \approx 0.577$ is Euler's constant. Mean and variance are given by $\mathbb{E}[\varepsilon] = 0$ and $\mathrm{Var}(\varepsilon) = \frac{\pi^2}{6\eta^2}$.

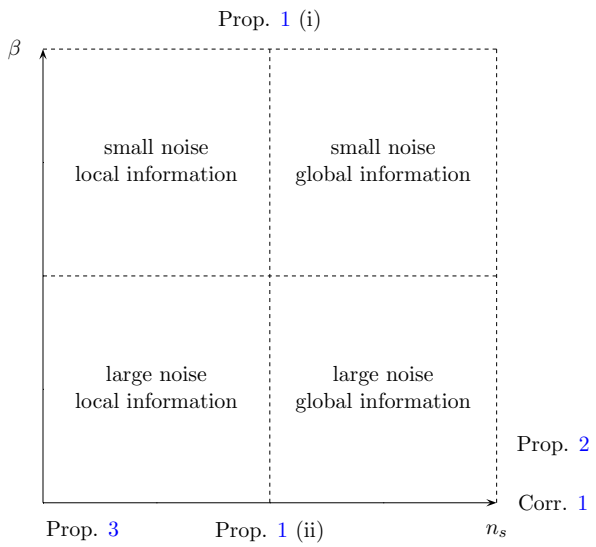



Figure 2: Illustration of the different parameter regions identified by the scaling parameter $\beta$ and the observation radius $n_s$. The figure also indicates the parameter regions to which the results discussed in Section 3 refer. Proposition 1 (i) deals with the case of $\beta = \infty$ and arbitrary 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 $\beta$, the latter assumes that $\beta = 0$. Proposition 3 deals with the case of $\beta = 0$ and small values of $n_s$.

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 degree of centralization has important efficiency implications and we will study these in Section 4.

### 3.1. The Emergence of Quasi-Stars

Our first result, which is central for the understanding of 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*. A quasi-star $S_n^m$, $n \geq m+1$, with node set $[n] \equiv \{1, \ldots, n\}$ 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 $K_{m+1} \subseteq S_n^m$.[17] An illustration of various quasi-stars can be seen in Figure 3. With this definition we are able to state the following proposition.

**Proposition 1.** *Let $(G_t^\beta)_{t \in [T]}$ be a sequence of networks generated with observation radius $n_s^{(1)}$, and $(H_t^\beta)_{t \in [T]}$ be a sequence of networks generated with observation radius $n_s^{(2)}$ such that $n_s^{(1)} > n_s^{(2)}$. Let*

[17] 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. (2008), König et al. (2009).

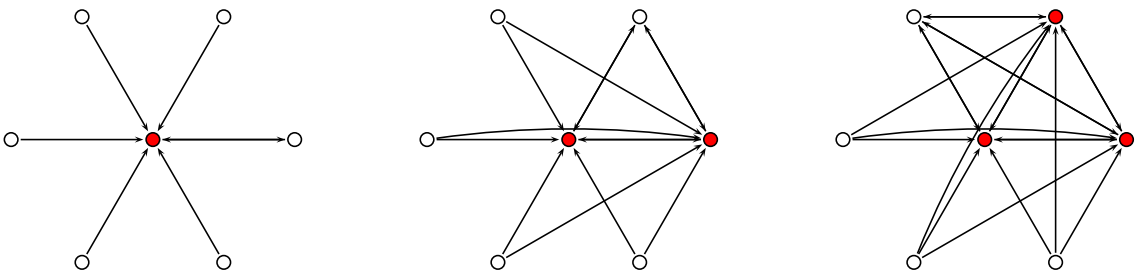


Figure 3: Illustration of the quasi-stars $S_7^1$, $S_7^2$ and $S_7^3$. Filled circles indicate the nodes with the highest degree.

$\Sigma_T^m \subset \mathcal{G}(T)$ *be the isomorphism class of quasi-stars of order* $T > m+1$. *Then,*

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

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

Proposition 1 shows that in the limit of vanishing noise ($\beta \to \infty$), the networks generated by our stochastic process 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 $n_s$. In the presence of noise, the set of networks generated by our model is much richer than the class of quasi-stars. In order 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 (for small values of $\beta$). 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 $P_t(k) \equiv \frac{1}{t} N_t(k)$ for all $t \geq 1$. The sequence $\{P_t(k)\}_{k\in\mathbb{Z}_+}$ is called the (empirical) *in-degree distribution*. Throughout the 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 $i \in \mathcal{P}_t$.

We first analyze the case of the observation radius $n_s$ being large enough, such that $\mathcal{S}_t = \mathcal{P}_{t-1}$.[18] 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 $\mathcal{P}_{t-1}$. Note that the probability that an agent

[18] Observe that the probability that an agent $i \in \mathcal{P}_{t-1}$ does not enter the sample $\mathcal{S}_t$ is given by

$$\mathbb{P}_t(i \notin \mathcal{S}_t|G_{t-1}) = \left(1 - \frac{1+d_{G_{t-1}}^-(i)}{t-1}\right)\left(1 - \frac{1+d_{G_{t-1}}^-(i)}{t-2}\right)\cdots\left(1 - \frac{1+d_{G_{t-1}}^-(i)}{t-1-(n_s-1)}\right) = \left(1 - \frac{1+d_{G_{t-1}}^-(i)}{t}\right)^{n_s} + o\left(\frac{1}{t}\right).$$

Applying Bonferroni's inequality and neglecting terms of the order $o\left(\frac{1}{t}\right)$, we then find that the probability that at least

$j$ with in-degree $d^-_{G_{t-1}}(j)$ receives a link in the $(k+1)$-st draw, given that the agents $l_1, \dots, l_k$ have received a link in the previous $k$ draws, $1 \le k \le m$, is[19]

$$\frac{e^{\beta d^-_{G_{t-1}}(j)}}{\sum_{i \in \mathcal{P}_{t-1} \setminus \{l_1,\dots,l_k\}} e^{\beta d^-_{G_{t-1}}(i)}} \approx \frac{1+\beta d^-_{G_{t-1}}(j)}{\sum_{i \in \mathcal{P}_{t-1} \setminus \{l_1,\dots,l_k\}} (1+\beta d^-_{G_{t-1}}(i))} = \frac{1+\beta d^-_{G_{t-1}}(j)}{(1+\beta m)t}\left(1+O_p\left(\frac{1}{t}\right)\right),$$

where we have used the approximation $e^{\beta x} \approx 1+\beta x$, and assumed that $d^-_{G_{t-1}}(i) = o_p(t)$ for all $i \in \mathcal{P}_{t-1}$. Moreover, we have used the fact that at every step $t$ every passive agent has out-degree equal to $m$. Since the average out-degree must be equal to the average in-degree, we see that also the average in-degree must be $m$, and so $\sum_{i \in \mathcal{P}_{t-1}} (1+\beta d_{G_{t-1}}(i)) = (1+\beta m)t$. 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

$$\begin{aligned} K^\beta_t(j|G_{t-1}) &\approx 1 - \left(1 - \frac{1+\beta d_{G_{t-1}}(j)}{(1+\beta m)t}\right)^m + o\left(\frac{1}{t}\right) = 1 - \left(1 - m\frac{1+\beta d_{G_{t-1}}(j)}{(1+\beta m)t}\right) + o\left(\frac{1}{t}\right) \\ &= \frac{m}{1+\beta m}\frac{1+\beta d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right). \end{aligned} \tag{3.1}$$

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

**Proposition 2.** *Fix $\epsilon > 0$ small and let $\beta \in (0, \epsilon)$, $m \ge 1$. 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 indefinite iteration of the network formation process $(G^\beta_t)_{t \in \mathbb{N}}$ 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*

$$P^\beta(k) = \frac{1+\beta m}{1+m(1+\beta)} \frac{\Gamma\left(\frac{1}{\beta}+k\right)\Gamma\left(2+\frac{1+\beta m}{\beta m}\right)}{\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(2+\frac{1+m}{1+\beta m}+k\right)}, \tag{3.2}$$

*for all $k \ge 0$.*

The expression for the degree distribution can be simplified when we focus on large degrees. Using

one of the agents in the set $\mathcal{P}_{t-1}$ is not observed by the entrant is bounded by

$$\mathbb{P}_t\Big(\bigcup_{i \in \mathcal{P}_{t-1}} \{i \notin \mathcal{S}_t\}|G_{t-1}\Big) \le \sum_{i=1}^{t-1} \mathbb{P}_t(i \notin \mathcal{S}_t|G_{t-1}) \approx \sum_{k=0}^{t-2}\left(1-\frac{1+k}{t}\right)^{n_s} P_t(k) \approx \sum_{k=0}^{t-2}\left(1-n_s\frac{1+k}{t}\right)P_t(k) = 1 - n_s\frac{1+m}{t},$$

where I have assumed that $k = o_p(t)$, and used the fact that the average in-degree $\sum_{k=0}^{t-2} kP_t(k)$ equals the out-degree $m$. Hence, if we require the probability of an agent not being sampled to be lower than $\epsilon > 0$, then we must have that $n_s > t\frac{1-\epsilon}{1+m}$.

[19] This probability is the same whether we use the in-degree $d^-_{G_{t-1}}(j)$ or the total degree $d_{G_{t-1}}(j)$, since they are related as $d_{G_{t-1}}(j) = d^+_{G_{t-1}}(j) + d^-_{G_{t-1}}(j) = m + d^-_{G_{t-1}}(j)$.

Stirling's formula, we get (for large $k$) the approximation (see Appendix A.2 for the details)

$$P^{\beta}(k) = (1+\beta k)^{-\left(2+\frac{1}{m\beta}\right)}\left(1+O\left(\frac{1}{k}\right)\right). \tag{3.3}$$

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 $2+\frac{1}{m\beta}$ for large degrees. Note, however, that this does not hold for small degrees. The degree distribution of Equation (3.2) 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 the exogenous noise (the smaller $\beta$) 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 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*

$$P^0(k) = \frac{1}{m+1}\left(\frac{m}{m+1}\right)^k, \tag{3.4}$$

*a geometric distribution with parameter $\frac{m}{m+1}$ 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 from Equation (2.7) (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 the order $o(t)$ and thus

$$\frac{1}{|\mathcal{S}_t|} = \frac{1}{n_s(m+1)} + o_p\left(\frac{1}{t}\right). \tag{3.5}$$

The unconditional probability that an agent $j \in \mathcal{P}_{t-1}$ receives a link by the entrant $t$ is then given by

$$K_t^0(j|G_{t-1}) = \frac{1}{n_s(m+1)}\mathbb{P}(j \in \mathcal{S}_t|G_{t-1}) + o\left(\frac{1}{t}\right).$$

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

$$\mathbb{P}(j \in \mathcal{S}_t | G_{t-1}) = n_s \frac{1 + d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right),$$

and we obtain

$$K_t^0(j|G_{t-1}) = \frac{n_s}{n_s(m+1)} \frac{1 + d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right) = \frac{1}{1+m} \frac{1 + d_{G_{t-1}}(j)}{t} + o\left(\frac{1}{t}\right). \tag{3.6}$$

We then can state the following result for the asymptotic degree distribution when the observation radius is small. The proof can be found in Appendix A.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*

$$P(k) = \frac{(1+m)\Gamma\left(3 + \frac{1}{m}\right)\Gamma(k+1)}{(1+2m)\Gamma\left(3 + \frac{1}{m} + k\right)}, \tag{3.7}$$

*for all $k \geq 0$.*

For large values of $k$ we can write Equation (3.7) as

$$P(k) = k^{-\left(2+\frac{1}{m}\right)}\left(1 + O\left(\frac{1}{k}\right)\right), \tag{3.8}$$

which is a power-law with exponent $2 + \frac{1}{m}$. A comparison with numerical simulations can be found in Figure 5. Compared to the power-law behavior in Equation (3.3) 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 hierarchical organization of the network.

Observe that the degree distribution in Equation (3.7) 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. These results are indicated in Figure 2. Hence, whether or not a limited observation radius impacts inequality in outcome networks depends crucially on the level of exogenous noise in agents's 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 $G_T^\beta$, respectively $H_T^\beta$, after $T$ iterations. Then, almost surely,*

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

Proposition 4 and Figure 4 show a major difference between the model considered here and the one by Jackson and Rogers (2007). In Jackson and Rogers (2007) a higher ratio of (local) neighborhood based linking to (global) random based linking is always increasing average payoff as long as payoff is a convex function of degree.[20] However, here we find that this does not hold in general when exogenous effects are taken into account, where this relationship might be reversed. 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 for efficiency results.

[20] See Corollary 1 and Footnote 51 in Jackson and Rogers (2007).

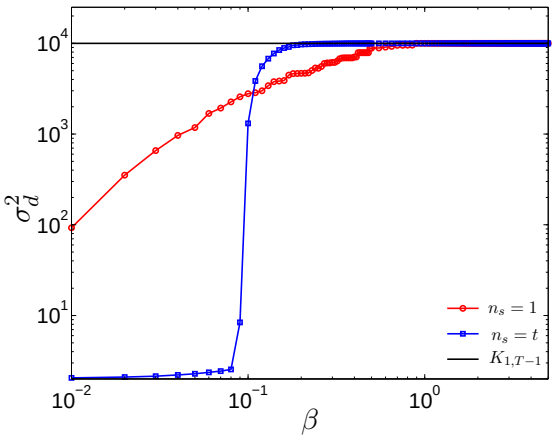



Figure 4: Degree variance $\sigma_d^2$ for local ($n_s = 1$) and global ($n_s = t$) search strategies for different values of $\beta$ with $m = 1$, $T = 10^4$ nodes (averaged over 10 simulation runs). The degree variance of the star $K_{1,T-1}$ is given by $\sigma_d^2(K_{1,T-1}) = (T-1)(T-2)^2/T^2$.

# 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.[21]

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).

Note that, in order 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 $\mathbb{R}_+$.[22] 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.

## 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 $P(k'|k' \to k)$ denote the probability that a node of in-degree $k$

[21] See Section 7 for an empirical application of the model to a network of inventors, a network of firm alliances and the network of trade relationships between countries.

[22] This is an approximation which has shown to be accurate in various growing network models as $T \to \infty$ (Dorogovtsev and Mendes, 2003, pp. 117). See Appendix A.4 for more discussion.

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 $k^-_{\rm nn}(k) = \int_0^\infty k' P(k'|k' \to k) dk'$ (Pastor-Satorras et al., 2001).[23] In the case that $k^-_{\rm nn}(k)$ is an increasing function of $k$ we speak of *assortative mixing*, while for $k^-_{\rm nn}(k)$ decreasing with $k$ we have *dissortative mixing* (Newman, 2002). Similarly, the average nearest out-neighbor connectivity $k^+_{\rm nn}(k)$ 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 $\beta$ (large noise) we obtain the following proposition:[24]

**Proposition 5.** *Consider the network formation process $(G^\beta_t)_{t\in\mathbb{R}_+}$ 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 $k$ is given by*

$$k^-_{nn}(k) = \frac{1}{\beta^2 k}\left(1 + (1+\beta k)(\ln(1+\beta k) - 1)\right), \tag{5.1}$$

*and the average nearest neighbor out-degree is given by*

$$k^+_{nn}(k) = \frac{1}{\beta^2 m}\left(\left(\beta m(1+p(\beta-1)) + \frac{a}{s}s^{2a}\zeta(s,2a)\right)\left(\frac{t}{s+1}\right)^a - m\beta\right), \tag{5.2}$$

*where $a = \frac{\beta m}{1+\beta m}$, $s = t(1+\beta k)^{-\frac{1}{a}}$ as $t \to \infty$.*

From Proposition 5 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 $t^{\frac{\beta m}{1+\beta m}}$. Figure 5 provides a comparison of numerical simulations with the theoretical predictions of Proposition 5.

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 $(G^\beta_t)_{t\in\mathbb{R}_+}$ 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 $k$ is given by*

$$k^-_{nn}(k) = \frac{1}{k}\left(1 + (k+1)\ln(k+1) - 1\right), \tag{5.3}$$

*and the nearest out-neighbor degree is given by*

$$k^+_{nn}(k) = \left(\frac{\Gamma(2+m)^2}{\Gamma\left(1+m+\frac{m}{m+1}\right)^2} + \frac{1}{m+1}\zeta\left(\frac{2m}{m+1}, 2+m\right)\right) t^{\frac{m-1}{m+1}}(1+k)^{\frac{1}{m}}, \tag{5.4}$$

[23] In the case of for uncorrelated networks we have that $P(k'|k' \to k) = k'P(k')$, where $P(k)$ is the probability to find a node with in-degree $k$ in the network $G$. Consequently, we get for uncorrelated networks that $k^-_{\rm nn}(k) = \int_0^\infty k' P(k'|k' \to k) dk' = \frac{\mathbb{E}[k^2]}{\mathbb{E}[k]}$, where $\mathbb{E}[k] = \int_0^\infty kP(k)dk = \bar{k}$ is the average in-degree in $G$ (see Boguñá and Pastor-Satorras, 2003).

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

*as* $t \to \infty$.

For large $k$ we find that $k^-_{\text{nn}}(k)$ grows logarithmically with $k$, is independent of the network size $t$, and $k^+_{\text{nn}}(k)$ grows as $O\left(t^{\frac{m-1}{m+1}} \cdot k^{\frac{1}{m}}\right)$. In Figure 5 a comparison of numerical simulations with the theoretical predictions of Proposition 6 is 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. We find, however, that even though the average nearest out-neighbor degree $k^+_{\text{nn}}(k)$ as well as the average nearest in-neighbor degree $k^-_{\text{nn}}(k)$ are increasing functions of the degree $k$, the total nearest neighbor connectivity $k_{\text{nn}}(k)$ (the sum of in- and out-neighbors' degrees divided by the total degree) is decreasing with degree (as e.g. in the network of international trade; see Section 7).[25]

As I will illustrate 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 $\bar{G}^\beta_t$ 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 $\bar{G}^\beta_t$ 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).[26] Note that in the case of $m = 1$ all networks will be trees, $\bar{G}^\beta_t \in \mathcal{T}([t])$, 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 $\beta$ (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*

[25] An increasing total nearest neighbor connectivity 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).

[26] Following Boguñá and Pastor-Satorras (2003), let $P(k', k''|k' \sim k, k'' \sim k)$ denote the joint probability that a vertex of degree $k$ has neighbors of degrees $k'$ and $k''$. Further, let $P(k' \sim k''|k' \sim k, k'' \sim k)$ denote the probability that vertices with degrees $k'$ and $k''$ are connected, given that they are neighbors of a vertex with degree $k$. Then we can write for the clustering coefficient as $C(k) = \int_0^\infty \int_0^\infty P(k', k''|k' \sim k, k'' \sim k) P(k' \sim k''|k' \sim k, k'' \sim k) dk'' dk'$. The average clustering coefficient is defined as $C = \int_0^\infty C(k)P(k)dk$. If degree correlations vanish, then we can obtain a simple expression for the clustering coefficient. Let $P(k'|k \sim k')$ be the conditional probability that a vertex of degree $k$ has a neighbor of degree $k'$. For an uncorrelated network $G \in \mathcal{G}(n)$ it follows that $P(k', k''|k' \sim k, k'' \sim k) = P(k'|k \sim k')P(k''|k \sim k'')$ and $P(k' \sim k''|k' \sim k, k'' \sim k) = \frac{(k'-1)(k''-1)}{\mathbb{E}[k]n}$, so that $C(k) = \frac{(\mathbb{E}[k^2]-\mathbb{E}[k])^2}{\mathbb{E}[k]^3 n}$, which is independent of $k$.

*degree $k$ is given by*

$$C(k) = \frac{2}{(k+pm)(k+pm-1)} \frac{a(m-1)}{mp\beta^3 b^2 s} \left( sb^2 \frac{mp\beta^3}{a(m-1)} M_s + \left( (1+\beta k)^b - 1 \right) \right.$$
$$\left. \times \left( b \left( \frac{s}{s+1} \right)^a \left( \beta^2 m + a s^{2a-1} \zeta(s, 2a) \right) - 1 \right) + b(1+\beta k)^b \ln(1+\beta k) \right), \tag{5.5}$$

*where $a = \frac{\beta m}{1+\beta m}$, $b = 2 - \frac{1}{a}$, 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 $s = t(1+\beta k)^{-\frac{1}{a}}$ as $t \to \infty$.*

The clustering coefficient in Equation (5.5) for $m = 4$ and $\beta = 0.1$ can be seen in Figure 5. 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

$$C(k) = O\left( t^{-\frac{2}{1+m\beta}} \cdot k^{2\left(\frac{1}{m\beta}-1\right)} \right). \tag{5.6}$$

This expression grows with $k$ as a power-law with exponent $2\left(\frac{1}{m\beta} - 1\right)$.[27] Moreover, we find that the clustering coefficient is decreasing with the network size as $t^{-\frac{2}{1+m\beta}}$. Hence, for large networks with a high clustering coefficient (such as the network of coinventors; 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 assume that $m > 1$. Let $a = \frac{m}{m+1}$ and $b = \frac{a(m-1)}{n_s(m+1)-1}$ with $a > b > 0$. If $\beta = 0$ then under the continuum approximation the clustering coefficient $C(k)$ of an agent with degree $k$ is bounded by $\underline{C}(k) \leq C(k) \leq \overline{C}(k)$, where*

$$\underline{C}(k) = \frac{2bk + 2(a(m-1) - bm)\left( (1+k)^{\frac{b}{a}} - 1 \right)}{(a-b)(k+m)(k+m-1)}, \tag{5.7}$$

*and*

$$\overline{C}(k) = \frac{2a(m-1) + 2b(k+m) + \left( a\left(m(m+1) - 2\right) - bm(1+m) \right)(1+k)^{\frac{b}{a}}}{(a-b)(k+m)(k+m-1)}, \tag{5.8}$$

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

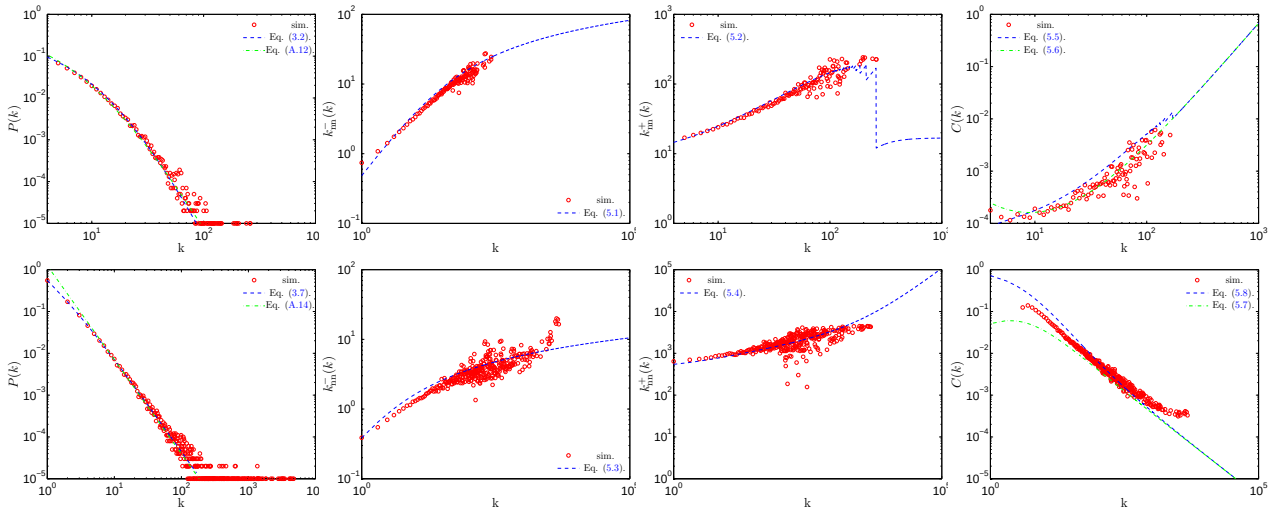



Figure 5: (Top row) Comparison of simulation results with the theoretical predictions for $T = 10^5$, $\mathcal{S}_t = \mathcal{P}_{t-1}$ and $m = 4$ with $\beta = 0.1$ under the linear approximation to the attachment kernel. (Bottom row) Comparison of simulation results for $T = 10^5$ and $n_s = m = 4$ ($\beta = 0$) with the theoretical predictions. Comparing the results of global and local information, we find that they differ mainly in the clustering degree distribution. First, smaller values of $n_s$ generate higher values of clustering. Second, for large values of $n_s$ the distribution is monotonic increasing, while for low values of $n_s$ it is monotonic decreasing.

*with the property that* $C(k) = O\left(\frac{1}{k}\right)$.

The bounds for the clustering coefficient in Equations (5.7) and (5.8) 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 minus one. 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 which is a typical feature of all the empirical networks we consider (see Section 7), indicating that a limited observation radius is an ubiquitous characteristic of real world networks in the present context.

# 6. Robustness Analysis and Extensions

In this section I briefly discuss two possible extensions of the model analyzed in the previous sections.

## 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 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 $\alpha = 3 + \frac{1}{m\beta}$ for a

large observation radius and $\alpha = 3 + \frac{1}{m}$ 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 not viable any more, because one cannot 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 Appendix C.

### 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 Appendix D. We find degree distributions that follow a power law decay $k^{-\alpha}$ with exponent $\alpha = 2 + \frac{1}{\beta m p}$ for a large observation radius and $\alpha = 1 + \frac{1+m}{pm}$ 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 $1 + \frac{1}{p}$. 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.

## 7. Empirical Implications

I consider three different real world networks in which knowledge diffusion and spillovers are an important source of knowledge generation and dissemination.

First, I analyze USPTO patent data in the year 2009 (see Lai et al., 2009, for a more detailed description of the data). For practical reasons 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.[28] The network of coinventors is constructed by creating a link between any pair of inventors

[28]The statistics computed for this subsample of the original data set are similar as in the full sample, or other subsamples for different sectors.

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.[29] This gives us a network with 27492 nodes, an average degree of $\bar{d} = 3.51$, a degree variance of $\sigma_d^2 = 30.03$ (with a coefficient of variation of $c_v \equiv \sigma_d/\bar{d} = 0.94$). The distribution of degree is highly skewed, following a power law for large degrees (see Figure 6). The network is highly clustered with an average clustering coefficient of $C = 0.64$ and a negative clustering-degree correlation (Figure 6). Moreover, the network is assortative, with an assortativity coefficient of $\kappa = 0.28$ (Newman, 2002).[30] The nearest neighbor average degree is monotonically increasing with degree (Figure 6). The largest component consists of 12060 nodes (which is 44% of all nodes).

Second, I consider a sample of a firm alliance network with alliances initiated before the year 2009. The data stems from the Thomson SDC alliance data base (cf. Gay and Dousset, 2005, Rosenkopf and Schilling, 2007, Schilling, 2009). I focus on the biotech sector (according to the Thomson SDC classification scheme), which is a sector with a high R&D collaboration intensity (Powell et al., 2005). The data base provides only information about the identity of the alliance partners (and not who initiated it) and so this network is undirected. The network of alliances is viewed as a proxy for the network of knowledge exchange and diffusion between firms. I obtain a network with 7374 nodes, an average degree of $\bar{d} = 1.79$ and a degree variance of $\sigma_d^2 = 8.33$ (the coefficient of variation is $c_v = 1.62$). The degree distribution follows a power-law (see Figure 6). Clustering is almost absent in the network of firms ($C = 0.0044$) and it is weakly assortative with $\kappa = 0.018$. The largest component consists of 3379 nodes (which is 46% of all nodes), which is similar to the network of coinventors.

Third, I consider the network of trade relationships between countries in the year 2000 (see Gleditsch, 2002, for a more detailed description of the data). Trade relationships in this data set are viewed as indicators of knowledge flows between countries (cf. Bitzer and Geishecker, 2006, Coe and Helpman, 1995). The trade network is defined as the network of import-export relationships between countries in a given year in millions of current-year U.S. dollars. I construct an undirected network in which a link is present between two countries if either one has exported to the other country. The trade network contains 196 nodes, has an average degree of $\bar{d} = 42.22$, a degree variance of $\sigma_d^2 = 1524.16$ and a coefficient of variation of $c_v = 0.92$. The network of trade is highly clustered with $C = 0.73$. The clustering degree correlation is negative (see Figure 6). Moreover, differently to the inventor and alliance network, it is dissortative, with a coefficient of $\kappa = -0.40$, and a monotonically

[29] As noted by Fafchamps et al. (2006), in the context of scientific coauthorship networks, the (unobserved) social network of personal acquaintances has more links than the coinventor network. However, the acquaintance network includes the coinventor 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.

[30] The assortativity coefficient $\kappa \in [-1, 1]$ is essentially the Pearson correlation coefficient of degree between nodes that are connected. Positive values of $\kappa$ 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.

Table 1: Descriptive statistics of the network of inventors and the network of firms before the year 2009, as well as the trade network in the year 2000.[a] Estimation of the model parameters $\mathbf{\Theta} = (m, \beta, n_s, p)$ for the network of inventors and the network of firms in the biotech sector. We have considered two model specifications: the case of entering agents observing only the out-neighbors of selected incumbents (Model A), as in Definition 1, and the case of entrants observing both, the out- and in-neighbors of the selected incumbents (Model B), as discussed in Section 6.1.

| | Inventor Network | | | | Firm Network | | | | Trade Network | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Model A | | Model B | | Model A | | Model B | | Model A | | Model B | |
| $T$ | 27495 | | | | 7374 | | | | 196 | | | |
| $n_s$ | 1.00 | (0.00) | 1.00 | (0.01) | 32.63 | (0.40) | 32.08 | (1.30) | 48.46 | (0.44) | 51.79 | (0.65) |
| $p$ | 0.60 | (0.00) | 0.58 | (0.00) | 0.69 | (0.05) | 0.82 | (0.02) | 0.34 | (0.01) | 0.58 | (0.02) |
| $m$ | 8.44 | (0.07) | 4.45 | (0.03) | 3.80 | (0.54) | 1.04 | (0.04) | 130.73 | (2.67) | 41.98 | (2.12) |
| $\beta$ | 0.75 | (0.09) | 1.46 | (0.23) | 0.01 | (0.00) | 0.01 | (0.00) | 0.57 | (0.09) | 1.69 | (0.27) |
| $n$ | 10000 | | 10000 | | 10000 | | 10000 | | 10000 | | 10000 | |

[a] Standard errors, reported in parenthesis, are calculated from batch means of length 10 (Chib, 2001).

decreasing average nearest neighbor degree (Figure 6). The network consists of a giant component with 181 nodes, encompassing 92% of all nodes in the network.

In order 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). The details of this algorithm are outlined in Appendix B.[31] I analyze both the basic model with directed links introduced in Definition 1 and the extension with undirected links, 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 one, as discussed in Section 6.2 (for both models, directed and undirected links). The initial condition (starting from a complete graph), which does not significantly impact the statistics in large networks, can affect the results in small networks such as the trade network. Hence, for the network of trade relationships between countries, I start from an empty network.[32]

The estimated parameter values are shown in Table 1. Moreover, Figure 6 shows various distributions for the inventor network, the firm alliance network and the network of trade relationships between countries, comparing fitted theoretical predictions of the model with empirical observations. The comparison of observed and the simulated distributions shown in Figure 6 indicate that the model can well reproduce the observed empirical networks.[33] The fit is in general better if entering agents are allowed to observe both, the out- and in-neighbors of the incumbents (see Section 6.1) and we allow for heterogeneity in the number of links being created (see Section 6.2).

Comparing the estimated observation radius $n_s$ for the inventor network to the one for the firm

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

[32] This restriction, however, is not crucial since its main effect is a slight reduction in the clustering coefficient for higher degree nodes.

[33] Estimating the model on an empirical network of coauthorships between physicists (Newman, 2001) shows a similarly good fit of the model as for the network of coinventors.

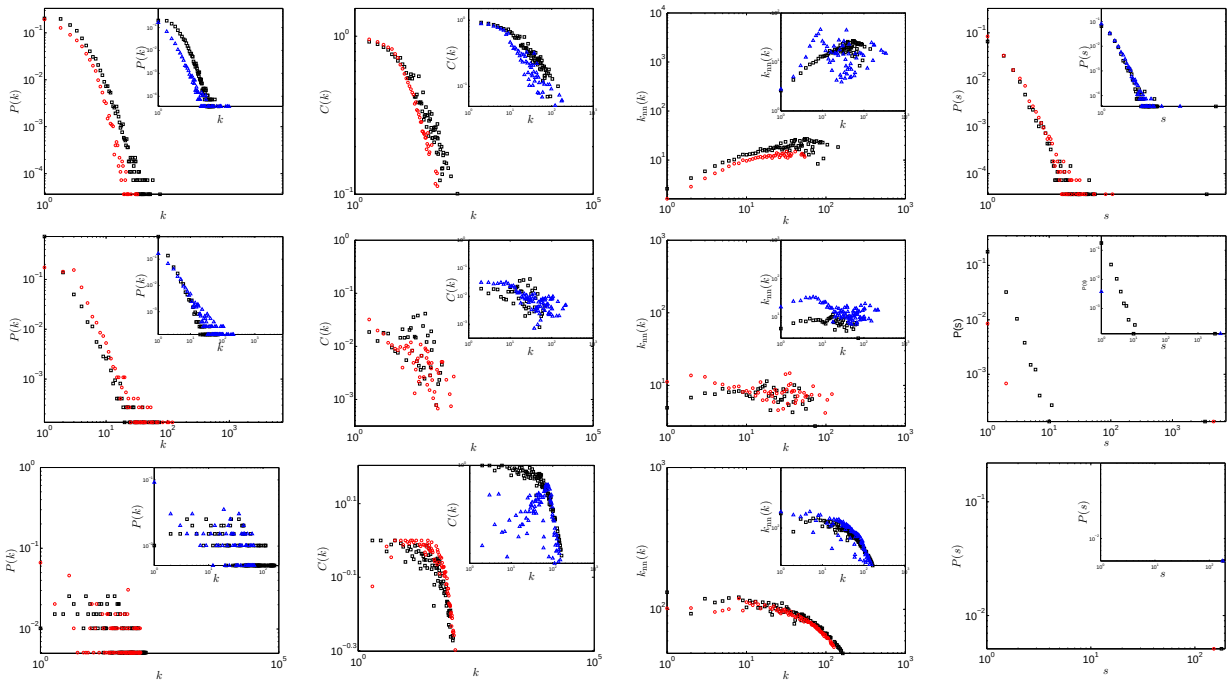


Figure 6: Empirical degree distribution $P(d)$ (first column), clustering-degree correlation $C(d)$ (second column), average nearest neighbor connectivity $k_{\mathrm{nn}}(d)$ (third column) and component size distribution $P(s)$ (fourth column) constructed from (first row) USPTO patents on drugs (patent classes 424 and 514), (second row) firm R&D alliances in the biotechnology sector and (third row) the world trade network in the year 2000 (data points indicated by □). The insets show the results obtained from the network formation process with directed links (△), corresponding to Model A in Table 1, while the larger figure shows the distributions obtained from the model with undirected links (○), corresponding to Model B in Table 1.

network in Table 1, we find that the number of observed agents by an entrant is much larger for firms than it is for inventors.[34] Hence, firms tend to use a significantly larger information set for their linking decisions than individual inventors. A similar observation can be made for the network of trade relationships between countries. This can be interpreted as an indicator for the presence of economies of scale in the information processing capabilities of larger organizations (such as firms compared to individual inventors). Moreover, the transition from assortative to dissortative networks for the network of coinventors, the network of firms and the trade network (see the change of $k_{\mathrm{nn}}(k)$ from an increasing to a decreasing function of $k$ in Figure 6, third column) can be explained from an increasing observation radius $n_s$ in the formation of these networks.

[34] Computing the $Z$-statistic for the differences in the sample means shows that the they are highly significantly different.

## 8. Conclusion

The current paper analyzes the growth of networks where agents' payoffs depend on communication or spillovers of valuable information form others through the links between them. 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 observes these agents and their neighbors. The set of observed agents constitute 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. 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 and a higher aggregate payoff. I then estimated the model using three different empirical networks: the network of coinventors, firm alliances and trade relationships between countries. I find that the model can reproduce the observed patterns for all these networks. The estimation shows that with increasing levels of aggregation (from individuals to firms or countries), the observation radius $n_s$ is increasing. This indicates the presence of economies of scale in which larger organizations are able to collect and process greater amounts of information.

The paper could be extended along several directions. First, I have assumed that the network is formed by incoming agents only, while neglecting the possibility of incumbent agents to form links. 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). Second, an extension of the analysis presented here could investigate further network measures and analyze additional network data sets beyond the ones studied in this paper (such as the coauthor network analyzed in Goyal et al. (2006)). 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 $\delta$) 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.

## References


Acemoglu, D. (2009). *Introduction to modern economic growth*. Princeton.

Aghion, P. and Howitt, P. (2009). *The economics of growth*. MIT Press.

Ahuja, G. (2000). Collaboration networks, structural holes, and innovation: A longitudinal study. *Administrative Science Quarterly*, 45:425–455.

Allen, R. C. (1983). Collective invention. *Journal of Economic Behavior & Organization*, 4(1):1–24.

Almeida, P. and Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management science*, 45(7):905–917.

Alós-Ferrer, C. and Weidenholzer, S. (2008). Contagion and efficiency. *Journal of Economic Theory*, 143(1):251–274.

Anderson, S. A., De Palma, A., and Thisse, J. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press.

Bala, V. and Goyal, S. (2000). A noncooperative model of network formation. *Econometrica*, 68(5):1181–1230.

Ballester, C., Calvó-Armengol, A., and Zenou, Y. (2006). Who's who in networks. wanted: The key player. *Econometrica*, 74(5):1403–1417.

Barrat, A. and Pastor-Satorras, R. (2005). Rate equation approach for correlations in growing network models. *Physical Review E*, 71(3):36127.

Bitzer, J. and Geishecker, I. (2006). What drives trade-related R&D spillovers? decomposing knowledge-diffusing trade flows. *Economics Letters*, 93(1):52–57.

Bloch, F. (1997). *The economic theory of the environment*, chapter Non-cooperative models of coalition formation with spillovers. Cambridge University Press.

Boguñá, M. and Pastor-Satorras, R. (2003). Class of correlated random networks with hidden variables. *Physical Review E*, 68(3):036112.

Bollobás, B., Riordan, O., Spencer, J., and Tusnády, G. (2001). The degree sequence of a scale-free random graph process. *Random Structures and Algorithms*, 18(3):279–290.

Bonacich, P. (1987). Power and centrality: A family of measures. *American Journal of Sociology*, 92(5):1170.

Bramoullé, Y. and Rogers, B. (2009). Diversity and Popularity in Social Networks. *Discussion Paper, Northwestern University*.

Brantle, T. and Fallah, M. (2007). Complex innovation networks, patent citations and power laws. In *Portland International Center for Management of Engineering and Technology*, pages 540–549.

Brooks, S. and Roberts, G. (1998). Assessing convergence of markov chain monte carlo algorithms. *Statistics and Computing*, 8(4):319–335.

Chaney, T. (2011). The network structure of international trade. Technical report, National Bureau of Economic Research.

Chernozhukov, V. and Hong, H. (2003). An MCMC approach to classical estimation. *Journal of Econometrics*, 115(2):293–346.

Chib, S. (2001). Markov chain monte carlo methods: computation and inference. *Handbook of econometrics*, 5:3569–3649.

Clauset, A., Shalizi, C., and Newman, M. (2009). Power-Law Distributions in Empirical Data. *SIAM Review*, 51(4):661–703.

Coe, D. and Helpman, E. (1995). International R&D spillovers. *European Economic Review*, 39(5):859–887.

Cooper, C. and Frieze, A. (2003). A general model of web graphs. *Random Structures and Algorithms*, 22(3).

Debreu, G. and Herstein, I. N. (1953). Nonnegative square matrices. *Econometrica*, 21(4):597–607.

Dorogovtsev, S. and Mendes, J. (2003). *Evolution of Networks - From Biological Nets to the Internet and WWW*. Oxford University Press.

Durrett, R. (2007). *Random Graph Dynamics*. Cambridge University Press.

Dutta, B., Ghosal, S., and Ray, D. (2005). Farsighted network formation. *Journal of Economic Theory*, 122(2):143–164.

Fafchamps, M., Goyal, S., and van der Leij, M. (2010). Matching and network effects. *Journal of the European Economic Association*, 8(1):203–231.

Fafchamps, M., Leij, M., and Goyal, S. (2006). Scientific Networks and Co-authorship. *Economics Series Working Papers*.

Fagiolo, G., Reyes, J., and Schiavo, S. (2009). World-trade web: Topological properties, dynamics, and evolution. *Phys. Rev. E*, 79:036115.

Fischer, M. (2006). *Innovation, networks, and knowledge spillovers: selected essays*. Springer Verlag.

Fleming, L., King, C., and Juda, Adam, I. (2007). Small worlds and regional innovation. *Organization Science*, 18(6):938–954.

Franco, C., Montresor, S., and Marzetti, G. V. (2011). On indirect trade-related r&d spillovers: The average propagation length of foreign R&D. *Structural Change and Economic Dynamics*, 22(3):227 – 237.

Frank, O. (1977). Survey sampling in graphs. *Journal of Statistical Planning and Inference*, 1(3):235–264.

Friedkin, N. (1983). Horizons of observability and limits of informal control in organizations. *Social Forces*, 62(1):54–77.

Galeotti, A., Goyal, S., Jackson, M., Vega-Redondo, F., and Yariv, L. (2010). Network games. *Review of Economic Studies*, 77(1):218–244.

Gay, B. and Dousset, B. (2005). Innovation and network structural dynamics: Study of the alliance network of a major sector of the biotechnology industry. *Research Policy*, pages 1457–1475.

Geweke, J. (1992). *Bayesian Statistics 4*, chapter Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, pages 169–193. Oxford University Press.

Ghiglino, C. (2011). Random walk to innovation: why productivity follows a power law. *Journal of Economic Theory*.

Gleditsch, K. (2002). Expanded trade and GDP data. *Journal of Conflict Resolution*, 46(5):712.

Goodman, L. (1961). Snowball sampling. *The Annals of Mathematical Statistics*, 32(1):148–170.

Goyal, S. (2007). *Connections: an introduction to the economics of networks*. Princeton University Press. forthcoming.
Goyal, S. and Joshi, S. (2006). Unequal connections. *International Journal of Game Theory*, 34(3):319–349.
Goyal, S. and Moraga-Gonzalez, J. L. (2001). R&D networks. *RAND Journal of Economics*, 32(4):686–707.
Goyal, S., van der Leij, M., and Moraga-Gonzalez, J. L. (2006). Economics: An emerging small world. *Journal of Political Economy*, 114(2):403–412.
Grimmett, G. and Stirzaker, D. (2001). *Probability and random processes*. Oxford University Press.
Grossman, G. and Helpman, E. (1995). Technology and trade. *Handbook of International Economics*, 3:1279–1337.
Gulati, R. (2007). *Managing network resources: Alliances, affiliations and other relational assets*. Oxford University Press.
Hagedoorn, J. (2002). Inter-firm R&D partnerships: an overview of major trends and patterns since 1960. *Research Policy*, 31(4):477–492.
Hall, B., Jaffe, A., and Trajtenberg, M. (2001). The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools. *NBER Working Paper*.
Jackson, M. (2008). *Social and Economic Networks*. Princeton University Press.
Jackson, M. O. and Rogers, B. W. (2007). Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915.
Jackson, M. O. and Wolinsky, A. (1996). A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74.
Jordan, J. (2006). The degree sequences and spectra of scale-free random graphs. *Random Structures & Algorithms*, 29(2):226–242.
Kolaczyk, E. (2009). *Statistical Analysis of Network Data: Methods and Models*. Springer.
König, M. D., Battiston, S., Napoletano, M., and Schweitzer, F. (2008). Efficiency and stability of dynamic innovation networks. CER-ETH Working Paper No. 08/95, ETH Zurich.
König, M. D., Tessone, C. J., and Zenou, Y. (2009). A dynamic model of network formation with strategic interactions. *CEPR Discussion Paper no. 7521, London, Centre for Economic Policy Research*.
Kovarik, J. and van der Leij, M. (2009). Risk aversion and networks: Microfoundations for network formation. University of Alicante Working Paper.
Krapivsky, P. and Redner, S. (2001). Organization of growing random networks. *Physical Review E*, 63(6):66123.
Krapivsky, P., Redner, S., and Leyvraz, F. (2000). Connectivity of growing random networks. *Physical Review Letters*, 85(21):4629–4632.
Kumar, R., Raghavan, P., Rajagopalan, S., Sivakumar, D., Tomkins, A., and Upfal, E. (2000). Stochastic models for the web graph. In *Proceedings of the 41st Annual Symposium on the Foundations of Computer Science, 2000.*, pages 57–65. IEEE.
Kuznets, S. (1969). *Modern Economic Growth: Rate, Structure, and Spread*. Yale University Press.
Lai, R., D'Amour, A., and Fleming, L. (2009). The careers and co-authorship networks of U.S. patent-holders, since 1975. Harvard Business School, Harvard Institute for Quantitative Social Science.
Marjoram, P., Molitor, J., Plagnol, V., and Tavaré, S. (2003). Markov chain Monte Carlo without likelihoods. *Proceedings of the National Academy of Sciences of the United States of America*, 100(26):15324.
Marshall, A. (1919). *Industry and trade*. McMillan.
Mas-Colell, A., Whinston, M. D., and R., G. J. (1995). *Microeconomic theory*. Oxford University Press New York.
McBride, M. (2006). Imperfect monitoring in communication networks. *Journal of Economic Theory*, 126(1):97–119.
McFadden, D. (1981). *Econometric models of probabilistic choice*, chapter Structural Analysis of Discrete Data with Econometric Applications. MIT Press, Cambridge, MA.
Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.
Newman, M. E. J. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences*, 98(2):404–409.
Newman, M. E. J. (2002). Assortative mixing in networks. *Physical Review Letters*, 89(20):208701.
Nooteboom, B. (2004). *Inter-firm collaboration, learning and networks: an integrated approach*. Psychology Press.
Oliveira, R. and Spencer, J. (2005). Connectivity transitions in networks with super-linear preferential attachment. *Internet Mathematics*, 2(2):121–163.
Pastor-Satorras, R., Vazquez, A., and Vespignani, A. (2001). Dynamical and correlation properties of the internet. *Physical Review Letters*, 87.
Powell, W. W., White, D. R., Koput, K. W., and Owen-Smith, J. (2005). Network dynamics and field evo-

lution: The growth of interorganizational collaboration in the life sciences. *American Journal of Sociology*, 110(4):1132–1205.
Radner, R. (1992). Hierarchy: The economics of managing. *Journal of economic literature*, 30(3):1382–1415.
Radner, R. and Van Zandt, T. (1992). Information processing in firms and returns to scale. *Annales d'Economie et de Statistique*, pages 265–298.
Ratmann, O., Andrieu, C., Wiuf, C., and Richardson, S. (2009). Model criticism based on likelihood-free inference, with an application to protein network evolution. *Proceedings of the National Academy of Sciences*, 106(26):10576.
Riccaboni, M. and Pammolli, F. (2002). On firm growth in networks. *Research Policy*, 31(8-9):1405–1416.
Robert, C. and Casella, G. (2004). *Monte Carlo statistical methods*. Springer Verlag.
Roijakkers, N. and Hagedoorn, J. (2006). Inter-firm R&D partnering in pharmaceutical biotechnology since 1975: Trends, patterns, and networks. *Research Policy*, 35(3):431–446.
Rosenkopf, L. and Schilling, M. (2007). Comparing alliance network structure across industries: Observations and explanations. *Strategic Entrepreneurship Journal*, 1:191–209.
Saxenian, A. (1994). *Regional Advantage: Culture and Competition in Silicon Valley and Route 128*. Harvard University Press, Cambridge.
Schilling, M. (2009). Understanding the alliance data. *Strategic Management Journal*, 30(3):233–260.
Serrano, M. A. and Boguñá, M. (2003). Topology of the world trade web. *Phys. Rev. E*, 68(1):015101.
Singh, J. (2005). Collaborative networks as determinants of knowledge diffusion patterns. *Management science*, 51(5):756–770.
Sisson, S. and Fan, Y. (2011). *Handbook of Markov Chain Monte Carlo*, chapter Likelihood-free markov chain monte carlo. Chapman & Hall/CRC Handbooks of Modern Statistical Methods.
Snijders, T. (2001). The Statistical Evaluation of Social Network Dynamics. *Sociological Methodology*, 31(1):361–395.
Snijders, T., Koskinen, J., and Schweinberger, M. (2010). Maximum likelihood estimation for social network dynamics. *The Annals of Applied Statistics*, 4(2):567–588.
Sokal, A. (1996). *Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms*.
Toivonen, R., Onnela, J., Saramäki, J., Hyvönen, J., and Kaski, K. (2006). A model for social networks. *Physica A*, 371(2):851–860.
Valverde, S., Sole, R. V., Bedau, M. A., and Packard, N. (2007). Topology and evolution of technology innovation networks. *Physical Review E*, 76:056118.
Vazquez, A. (2003). Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(5):056104–1–15.
Vega-Redondo, F. (2007). *Complex Social Networks*. Series: Econometric Society Monographs. Cambridge University Press.
Von Hippel, E., Thomke, S., and Sonnack, M. (1999). Creating breakthroughs at 3m. *Harvard Business Review*, 77:47–57.
Wang, L., Guo, J., Yang, H., and Zhou, T. (2009). Local preferential attachment model for hierarchical networks. *Physica A: Statistical Mechanics and its Applications*, 388(8):1713–1720.
Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393:440–442.
Westbrock, B. (2010). Natural concentration in industrial research collaboration. *The RAND Journal of Economics*, 41(2):351–371.
Wilson, R. (1975). Informational economies of scale. *The Bell Journal of Economics*, 6(1):pp. 184–195.

# Appendix

## A. Proofs

In this appendix I give the proofs of the propositions, corollaries and lemmas stated in the paper.

### A.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

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

Then, in the limit $\beta \to \infty$, we obtain from Equation (2.7) that

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

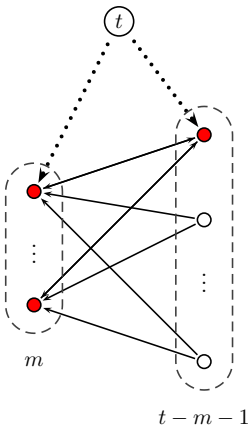
Hence, the entrant makes a uniform draw without replacement from the best response set $\mathcal{B}_t$ when deciding with whom to form a link with probability one, and the probability that an agent $j$ receives a link by the entrant is given by

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

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 one. 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 one.

In the following 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 one. The reason is the following: 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




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

Figure 7: (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) $X_0$ denotes the number of agents drawn from the set $[m+1]$ and $X_1$ 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 $X_1$.

from Equation (2.7) that

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

It follows that the entrant selects $m$ agents uniformly without replacement from the sample $\mathcal{S}_t$ with probability one 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 of this to happen is

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

Consequently, we then can write

$$\mathbb{P}_t(G_t \in \Sigma_t^m | G_{t-1} \in \Sigma_{t-1}^m) = \sum_{\mathcal{S}_t\in\mathcal{P}_{t-1}} \binom{|\mathcal{S}_t|}{m}^{-1} \mathbb{P}_t(\mathcal{S}_t | G_{t-1} \in \Sigma_{t-1}^m). \tag{A.2}$$

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 7. Let $X_0$ denote the number of agents drawn from the set $[m+1]$ and $X_1$ 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

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

The expected sample size is

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

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 $0 \leq l \leq n_s$, then first-order stochastic dominance is implied by

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

which is equivalent to

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

The last expression is non-negative 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 (e.g. Mas-Colell et al., 1995). Since Equation (A.1) is a decreasing function of the sample size $|\mathcal{S}_t|$, we can apply stochastic dominance and it follows that Equation (A.2) is decreasing with $n_s$. The network $G_{t\leq m+1}$ is the complete graph $K_{m+1}$ and therefore 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 $\lim_{\beta\to 0}\mathbb{P}(G_T^\beta \in \Sigma_T^m) < \lim_{\beta\to 0}\mathbb{P}(H_T^\beta \in \Sigma_T^m)$.

□

## A.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 $P_t^\beta(k) \equiv \frac{1}{t}N_t(k)$ for all $t \geq 1$. The sequence $\{P_t^\beta(k)\}_{k\in\mathbb{N}}$ is the (empirical) *degree distribution*.

We will now derive a recursive system which can be used to describe the time evolution of the expected degree distribution. Let $N_t \equiv \{N_t(k)\}_{k\geq 0}$. Denoting by $k = d^-_{G_{t-1}}(j)$ we write the attachment kernel as $K_t^\beta(j|G_{t-1}) = \frac{a(k)}{t\zeta(\beta,m)} + o\left(\frac{1}{t}\right)$. 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

$$\mathbb{E}[N_{t+1}(k)|N_t] = N_t(k)\left(1 - \frac{a(k)}{t\zeta(\beta,m)}\right) + N_t(k-1)\frac{a(k-1)}{t\zeta(\beta,m)} + \delta_{0,k} + o\left(\frac{1}{t}\right). \tag{A.3}$$

Taking expectations on both sides of Equation (A.3), dividing by $t+1$, and denoting by $P_t^\beta(k) = \mathbb{E}[N_t(k)]$, 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

$$P_{t+1}^\beta(k) - P_t^\beta(k) = b_t(k)\left[c_t(k) - P_t^\beta(k)\right] + o\left(\frac{1}{t}\right), \tag{A.4}$$

where

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

The following lemma gives us a simple way to determine the asymptotic solution (i.e. as $t \to \infty$) of the recursion in Equation (A.4).

**Lemma A.1.** *Let $(x_n), (y_n), (\eta_n), (r_n)$ 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 $\lim_{n\to\infty} x_n = x$.*

Proof of Lemma A.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 $c(0) = P^\beta(0), \ldots, c(k-1) = P^\beta(k-1), k > 0$. 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

$$P^{\beta}(k) = \frac{\zeta(\beta, m)}{\zeta(\beta, m) + a(0)} \prod_{j=1}^{k} \frac{a(j-1)}{\zeta(\beta, m) + a(j)}. \tag{A.5}$$

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 of Proposition 2. For $\beta \to 0$ the attachment kernel of Equation (3.1) is given by $K_t^{\beta}(j|G_{t-1}) = \frac{a(k)}{t\zeta(\beta,m)} + o\left(\frac{1}{t}\right)$, where $k = d_{G_{t-1}}(j)$, $a(k) = 1 + \beta k$ and $\zeta(\beta, m) = \frac{1+\beta m}{m}$. We then can apply Equation (A.5), noting that the product on the right-hand side admits a closed-form representation in terms of Gamma functions as

$$P^{\beta}(k) = \frac{1 + \beta m}{1 + m(1+\beta)} \frac{\Gamma\left(\frac{1}{\beta} + k\right) \Gamma\left(2 + \frac{1+\beta m}{\beta m}\right)}{\Gamma\left(\frac{1}{\beta}\right) \Gamma\left(2 + \frac{1+m}{1+\beta m} + k\right)}. \tag{A.6}$$

By Stirling's formula we can approximate the Gamma function for large $k$ as[35]

$$\frac{\Gamma(k)}{\Gamma(k+c)} = k^{-c}\left(1 + O\left(\frac{1}{k}\right)\right). \tag{A.7}$$

For the tails of the degree distribution in Equation (A.6) this implies that $P^{\beta}(k) \sim (1+\beta k)^{-(2+\frac{1}{\beta m})}\left(1 + O\left(\frac{1}{k}\right)\right)$ for large $k$. □

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

Proof of Corollary 1. The degree distribution in Equation (3.4) follows from the attachment kernel $K_t^0(j|G_{t-1}) = \frac{a(k)}{t\zeta(\beta,m)} + o\left(\frac{1}{t}\right) = \frac{m}{t} + o\left(\frac{1}{t}\right)$ and inserting $a(k) = 1$ and $\zeta(\beta, m) = \frac{1}{m}$ into Equation (A.5). □

[35] 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

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

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.

Proof of Proposition 3. With the attachment kernel from Equation (3.6) given by $K_t^0(j|G_{t-1}) = \frac{a(k)}{t\zeta(\beta,m)} + o\left(\frac{1}{t}\right) = \frac{m}{m+1}\frac{1+k}{t} + o\left(\frac{1}{t}\right)$, where $k = d_{G_{t-1}}(j)$, $a(k) = 1 + k$ and $\zeta(\beta, m) = \frac{m+1}{m}$, we can apply Equation (A.5) to obtain

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

Using Equation (A.7) we get $P(k) \sim k^{-\left(2+\frac{1}{m}\right)}$ for large $k$. □

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*

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

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

Proof of Proposition 9. 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 $\sigma$-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 $\sigma$-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 $\mathbb{E}_t[|Z_n|] = \mathbb{E}_t[Z_n] = \mathbb{E}_t[N_t(k)] \leq t < \infty$. Secondly, 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 $G_1$. $Z_t$ 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

$$Z_n - Z_{n-1} = \sum_{i\in\mathcal{N}_t} \left[\mathbb{P}_t(d_{G_{t-1}}(i) = k|\mathcal{F}_n) - \mathbb{P}_t(d_{G_{t-1}}(i) = k|\mathcal{F}_{n-1})\right]. \tag{A.9}$$

In $\mathcal{F}_{n-1}$ we know where the edges up to time $n-1$ have been attached to. In $\mathcal{F}_n$ we know in addition where the edges in the $n$-th step are attached to. These edges affect the total degree of $m+1$ vertices, namely the ones receiving a link and the one 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 the 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 $n$-th 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, Equation (A.9) 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 $\eta > 0$

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

and by choosing $\eta = \epsilon t$ Equation (A.8) 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 $k \geq 0$, $P_t(k) \xrightarrow{a.s.} \mathbb{E}_t\left[P_t(k)\right]$, as $t \to \infty$.*

Proof of Proposition 10. The proof follows from the Borel-Cantelli lemma (see e.g. Grimmett and Stirzaker 2001) and Proposition 9 by observing that for any $\epsilon > 0$

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

□

## A.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

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

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 $T$. □

## A.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, 2003, pp. 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

$$P_t(k) = \frac{1}{t}\int_0^t \delta(k - \bar{k}(s,t))ds = -\frac{1}{t}\left(\frac{\partial \bar{k}(s,t)}{\partial s}\right)^{-1}\Bigg|_{s=s(k,t)}. \tag{A.10}$$

In order to compare this approximation with our previous analysis, we will derive the degree distributions in the case of a large and 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 $\mathbb{E}_t[k_s(t+\Delta t) - k_s(t)|G_t] \approx \frac{m}{1+\beta m}\frac{1+\beta k_s(t)}{t}\Delta t$ for large $t$, where Equation (3.1) 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 following 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

$$k_s(t) = \frac{1}{\beta}\left(\left(\frac{t}{s}\right)^{\frac{m\beta}{1+m\beta}} - 1\right), \tag{A.11}$$

From Equation (A.10) we then get

$$P^\beta(k) = \frac{1+\beta m}{m}(1+\beta k)^{-(2+\frac{1}{\beta m})}, \tag{A.12}$$

with $\int_0^\infty P^\beta(k)dk = 1$. This is asymptotically equivalent to the degree distribution we have obtained in Equation (3.2).

Similarly, in the case of $n_s$ small enough (small observation radius), we have from Equation (3.6) that $\mathbb{E}_t[k_s(t+\Delta t) - k_s(t)|G_t] \approx \frac{m}{1+m}\frac{1+k_s(t)}{t}\Delta t$ 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

$$k_s(t) = \left(\frac{t}{s}\right)^{\frac{m}{m+1}} - 1, \tag{A.13}$$

From Equation (A.10) we then get

$$P(k) = \frac{m+1}{m}(1+k)^{-(2+\frac{1}{m})}, \tag{A.14}$$

with the property that $\int_0^\infty P(k)dk = 1$. Comparing this distribution with the one in Equation (3.7) 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 $O(\frac{1}{t})$ in our derivations in the following sections.

### A.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 Equation (3.1), we have up to leading orders in $O\left(\frac{1}{t}\right)$ 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 by $a = \frac{m\beta}{1+m\beta}$. Wit the initial condition $R_s^-(s) = 0$ we obtain

$$R_s^-(t) = \frac{1}{\beta^2}\left(1 + \left(a\ln\left(\frac{t}{s}\right) - 1\right)\left(\frac{t}{s}\right)^a\right), \tag{A.15}$$

and the average nearest neighbor in-degree is given by $k_{\mathrm{nn}}^-(k_s) = \frac{R_s^-(t)}{k_s}$. From Equation (A.11) we know that $\frac{t}{s} = (1+\beta k_s)^{\frac{1}{a}}$, and we obtain

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

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 $O\left(\frac{1}{t}\right)$ 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

$$R_s^+(t) = -\frac{m}{\beta} + C_s t^a, \tag{A.16}$$

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

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

where $\zeta(s,2a) \equiv \sum_{j=0}^{\infty} \frac{1}{(2a+j)^s}$ is the Hurwitz zeta function. Together with the solution Equation (A.16) we then get

$$R_s^+(t) = \frac{1}{\beta^2}\left(\left(\beta m(1+p(\beta-1)) + \frac{a}{s}s^{2a}H(s,2a)\right)\left(\frac{t}{s+1}\right)^a - m\beta\right). \tag{A.17}$$

The average nearest out-neighbor in-degree is then given by $k_{\text{nn}}^+(k) = \frac{R_s^+}{m}$. □

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 Equation (3.6), we have up to leading orders in $O\left(\frac{1}{t}\right)$ that[36]

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

where we have denoted by $a = \frac{m}{1+m}$. In the continuum approximation we have that $k_s(t) = \left(\frac{t}{s}\right)^a - 1$ (see Equation (A.13)), 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 this 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 $s(k,t) = \frac{t}{(k+1)^{\frac{1}{a}}}$ we obtain

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

It follows that

$$k_{\text{nn}}^- = \frac{R_s^-}{k} = \frac{1}{k}\left(1 + (k+1)\left(\ln(k+1) - 1\right)\right).$$

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 $R_s^+(t) = \sum_{j\in\mathcal{N}_{G_t}^+(s)} k_j(t)$. In order 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

[36] 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 $O\left(\frac{1}{t^2}\right)$.

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 $O\left(\frac{1}{t}\right)$. Second, we need to consider the cases where the out-neighbors of $s$ are found either directly or indirectly through other vertices than $s$. The probability of this is given by $\frac{m}{(m+1)t}k_j(t)$ for each $j$ in $\mathcal{N}^+_{G_t}(s)$ (discounting the link from $s$) Taking these cases together and denoting by $a = \frac{m}{m+1}$, 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

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

$C_s$ 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 $s(k,t) = \frac{t}{(k+1)^{\frac{1}{a}}}$ from Equation (A.13) gives

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

With $k^+_{\mathrm{nn}}(k) = \frac{R^+_s(k)}{m}$ we then get Equation (5.4). □

### A.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 $\bar{G}_t$. 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 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. $M_s(t)$ 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 8. The probability associated with case (i) up to leading orders in


Figure 8: (Left panel) Vertex $s$ and one of its out-neighbors $u \in \mathcal{N}^+_{G_t}(s)$ receive a link bei the entrant $t$. (Right panel) Vertex $s$ and one of its in-neighbors $u \in \mathcal{N}^-_{G_t}(s)$ receive a link.

$O\left(\frac{1}{t}\right)$ is given by

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

Similarly, the probability associated with case (ii) up to leading orders in $O\left(\frac{1}{t}\right)$ is given by

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

With $R^-_s$ and $R^+_s$ given by Equations (5.1) and (5.2), respectively, we obtain

$$\begin{aligned} \frac{dM_s(t)}{dt} &= \frac{m(m-1)(1+\beta k_s(t))}{(1+\beta m)t^2}(m + k_s(t) + \beta(R^+_s + R^-_s)) \\ &= \frac{a^2}{t^2}\frac{m-1}{m\beta^3}\left(\left(\beta^2 m + a s^{2a-1} H(s,2a)\right)\left(\frac{t}{s}\right)^a \left(\frac{t}{s+1}\right)^a + \left(\frac{t}{s}\right)^{2a} a \ln\left(\frac{t}{s}\right)^a\right). \end{aligned} \tag{A.18}$$

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

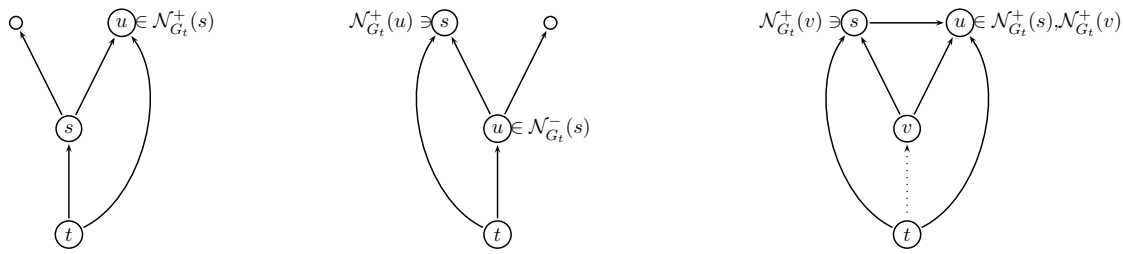



Figure 9: (Left panel) Vertex $s$ and one of its out-neighbors $u \in \mathcal{N}^+_{G_t}(s)$ receive a link. (Middle) Vertex $s$ and one of its in-neighbors $u \in \mathcal{N}^-_{G_t}(s)$ receive a link. (Right panel) The entrant $t$ observes a vertex $v$ and forms a link to both vertices $s$ and $u$ which are both out-neighbors of $v$.

indices $1 \leq i \leq m$ are initially connected), we can write the initial condition as follows[37]

$$\begin{aligned} M_{s+1} &= \frac{m(m-1)}{2} \sum_{j \neq i}^{s} \frac{1+\beta k_i(s)}{(1+\beta m)s} \frac{1+\beta k_j(s)}{(1+\beta m)s} \left(\Theta(m+1-i)\Theta(m+1-j)\right. \\ &\left. +\Theta(i-j)\Theta(j-m)m\frac{1+\beta k_j(i)}{(1+\beta m)(i-1)} + \Theta(j-i)\Theta(i-m)m\frac{1+\beta k_i(j)}{(1+\beta m)(j-1)}\right) \\ &= \frac{m(m-1)s^{2a-2}}{(1+\beta m)^2}\left(\sum_{i=1}^{m}\frac{1}{i^a}\sum_{j=i+1}^{m}\frac{1}{j^a} + \frac{2m}{1+\beta m}\sum_{i=m+1}^{s}\frac{1}{i^{2a}}\sum_{j=i+1}\frac{1}{j-1}\right), \end{aligned} \tag{A.19}$$

where we have denoted by $a = \frac{\beta m}{1+\beta m}$. Combining the initial condition in Equation (A.19) with Equation (A.18) yields Equation (5.5). □

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 9. In case (i) we consider that vertex $s$ is observed directly. The probability of this to happen is given by $\frac{n_s}{t}$. 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 follows

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

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

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 $\frac{n_s}{t}$, 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 $\frac{n_s}{t}$. The number of such triangles including node $s$ is given by $M_s(t)$ (in both $G_t$ and its closure $\bar{G}_t$). The expected number of triangles being formed in this way is then given as follows

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

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

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

where we have denoted by $a = \frac{m}{m+1}$ and used the fact that $k_s(t) = \left(\frac{t}{s}\right)^a - 1$ in the continuum approximation in Equation (A.13). Further denoting by $b = \frac{a(m-1)}{n_s(m+1)-1}$ we can write this as

$$\frac{dM_s(t)}{dt} = \frac{b}{t}\left(m-1+\left(\frac{t}{s}\right)^a+M_s(t)\right). \tag{A.20}$$

The general solution of Equation (A.20) is given by

$$M_s(t) = \frac{1}{a-b}\left((b-a)(m-1)+b\left(\frac{t}{s}\right)^a+(a(m-1)-bm+(a-b)M_s(s))\left(\frac{t}{s}\right)^b\right). \tag{A.21}$$

From Equation (A.21) 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

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

From Equation (A.13) we know that $s = t(1+k)^{-\frac{1}{a}}$. Inserting this into $\overline{M}_s(t)$ and $\underline{M}_s(t)$ and using the fact that $C(k) = \frac{2M_k}{(k+m)(k+m-1)}$ allows us to bound the clustering coefficient as $\underline{C}(k) \leq C(k) \leq \overline{C}(k)$, where

$$\underline{C}(k) = \frac{2bk + 2(a(m-1)-bm)\left((1+k)^{\frac{b}{a}} - 1\right)}{(a-b)(k+m)(k+m-1)},$$

and

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

For large $k$, these bounds decay as $O\left(\frac{1}{k}\right)$. Further, their difference is given by

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

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

□

# B. The LF-MCMC Algorithm

The purpose of the likelihood-free Markov chain Monte Carlo (LF-MCMC) algorithm is to estimate the parameter vector $\mathbf{\Theta} \equiv (\beta, p, n_s, m)_{1\times L}$, $L = 4$, of the model on the basis of the summary statistics $\mathbf{S} \equiv (\mathbf{S}_1, \ldots, \mathbf{S}_K)_{T\times K}$, $K = 4$, where $\mathbf{S}_1 \equiv (P(k))_{k=0}^{T-1}$, $\mathbf{S}_2 \equiv (C(k))_{k=0}^{T-1}$, $\mathbf{S}_3 \equiv (k_{\mathrm{nn}}(k))_{k=0}^{T-1}$ and $\mathbf{S}_4 \equiv (P(s))_{s=1}^{T}$. The algorithm generates a Markov chain which is a sequence of parameters $(\mathbf{\Theta}_s)_{s=1}^n$ with a stationary distribution that approximates the distribution of each parameter value $\theta \in \mathbf{\Theta}$ conditional on the observed statistic $\mathbf{S}^o$.

**Definition 2.** *Consider the statistics* $\mathbf{S}$ *and denote by* $\mathbf{S}^o$ *the observed statistics. Further, let* $\Delta(\mathbf{S}_i^o, \mathbf{S}_i)$ *be a measure of distance between the i-th realized statistic* $\mathbf{S}_i$ *of the network formation process* $(G_t)_{t=1}^T$ *with parameter vector* $\mathbf{\Theta}$ *and the i-th observed statistic* $\mathbf{S}_i^o$ *for* $i = 1, \ldots, K$. *Then we consider the Markov chain* $(\mathbf{\Theta}_s)_{s=1}^n$ *induced by the following algorithm:*

**(i)** *Given* $\mathbf{\Theta}$, *propose* $\mathbf{\Theta}'$ *according to the proposal density* $q_s(\mathbf{\Theta} \to \mathbf{\Theta}')$.

**(ii)** *Generate a network* $G_T(\mathbf{\Theta}')$ *according to* $\mathbf{\Theta}'$ *and calculate the summary statistics* $\mathbf{S}'$.

**(iii)** *Calculate*

$$h(\mathbf{\Theta}, \mathbf{\Theta}') = \min\left(1, \frac{q_s(\mathbf{\Theta}' \to \mathbf{\Theta})}{q_s(\mathbf{\Theta} \to \mathbf{\Theta}')} \prod_{i=1}^{K} \mathbb{1}_{\{\Delta(\mathbf{S}_i', \mathbf{S}_i^o) < \epsilon_{i,s}\}}\right), \tag{B.1}$$

*where* $\epsilon_{i,s} \geq 0$ *is a monotonic decreasing sequence of threshold values,* $\epsilon_{i,s} \downarrow \epsilon_i^{\min}$, *and* $\Delta : \mathbb{R}_+^T \times \mathbb{R}_+^T \to \mathbb{R}_+$ *is a distance metric in* $\mathbb{R}_+^T$.

**(iv)** *Accept $\mathbf{\Theta}'$ with probability $h(\mathbf{\Theta}, \mathbf{\Theta}')$, otherwise stay at $\mathbf{\Theta}$ and go to (i).*

Marjoram et al. (2003) have shown that the distribution generated by the above algorithm converges to the true conditional distribution of the parameter vector $\mathbf{\Theta}$, given the observations $\tau^o$ and the threshold values. Their result is stated more formally in the following proposition.

**Proposition 11.** *The stationary distribution $f : \mathbb{R}^K \to [0,1]^K$ of the Markov chain $(\mathbf{\Theta}_s)_{s=1}^n$ is given by*

$$f\left(\mathbf{\Theta} \middle| \prod_{i=1}^{K} \mathbb{1}_{\{\Delta(\mathbf{S}_i, \mathbf{S}_i^o) < \epsilon_i^{\min}\}}\right).$$

Proof of Proposition 11. Let us denote the transition probability of the Markov chain $(\mathbf{\Theta}_s)_{s=1}^n$ from state $\mathbf{\Theta}$ to state $\mathbf{\Theta}'$ by $p_s(\mathbf{\Theta} \to \mathbf{\Theta}')$. Assume w.l.o.g. that for $\mathbf{\Theta} \neq \mathbf{\Theta}'$ and $1 \leq s \leq n$ it holds that

$$\frac{q_s(\mathbf{\Theta}' \to \mathbf{\Theta})}{q_s(\mathbf{\Theta} \to \mathbf{\Theta}')} \leq 1. \tag{B.2}$$

Consider the distribution of the parameter vector $\mathbf{\Theta}$, conditional on the event $\{\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon\} \equiv \prod_{i=1}^{K} \mathbb{1}_{\{\Delta(\mathbf{S}_i, \mathbf{S}_i^o) < \epsilon_i^{\min}\}}$, that is

$$f(\mathbf{\Theta} | \Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon) = \frac{\mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon | \mathbf{\Theta})}{\mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon)}.$$

We have that

$$\begin{aligned}
f(\mathbf{\Theta} | \Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon) p_s(\mathbf{\Theta} \to \mathbf{\Theta}') &= \frac{\mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon | \mathbf{\Theta})}{\mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon)} \mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}') \leq \epsilon | \mathbf{\Theta}') q_s(\mathbf{\Theta} \to \mathbf{\Theta}') \frac{q_s(\mathbf{\Theta}' \to \mathbf{\Theta})}{q_s(\mathbf{\Theta} \to \mathbf{\Theta}')} \\
&= \frac{\mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}') \leq \epsilon | \mathbf{\Theta}')}{\mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon)} \mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon | \mathbf{\Theta}) q_s(\mathbf{\Theta}' \to \mathbf{\Theta}) \\
&= f(\mathbf{\Theta}' | \Delta(\mathbf{S}^o, \mathbf{S}') \leq \epsilon) q_s(\mathbf{\Theta}' \to \mathbf{\Theta}) \mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon | \mathbf{\Theta}) h(\mathbf{\Theta}', \mathbf{\Theta}) \\
&= f(\mathbf{\Theta}' | \Delta(\mathbf{S}^o, \mathbf{S}') \leq \epsilon) p_s(\mathbf{\Theta}' \to \mathbf{\Theta}),
\end{aligned}$$

where we have used the fact that $h(\mathbf{\Theta}', \mathbf{\Theta}) = 1$ if the inequality in (B.2) is satisfied. It follows that $f(\mathbf{\Theta} | \Delta(\mathbf{S}^o, \mathbf{S}) \leq \epsilon)$ satisfies a detailed balance condition and therefore is the stationary distribution of the Markov chain. □

The algorithm of Definition 2 is implemented as follows. First we need to choose the initial parameter values.[38] The network size $T$ is already given by the data. I set $\beta = 0$ for all empirical networks as a starting value. In this case, the empirical average degree is used as a restriction for the parameters $p$ and $m$ through $\bar{d} = mp$ when the network is directed (while $\bar{d} = 2mp$ when it is undirected). I compute the power-law exponent $\alpha$ of the tail of the empirical degree distribution for the network of coinventors (cf. Clauset et al., 2009). For the directed model with heterogeneous linking opportunities and $\beta = 0$ one can show that the distribution decays as $k^{-\alpha}$ with $\alpha = 1 + \frac{1+m}{mp}$ for large degrees $k$

[38] Alternatively, we could choose a uniform prior distribution similar to e.g. Ratmann et al. (2009). However, this would greatly amplify the number of iterations needed to reach the stationary distribution (which is independent of the initial conditions). For computational reasons I thus specify the initial parameters explicitly.

in the case of $\beta = 0$. Hence, I can compute $p$ and $m$ from these two conditions. For the network of coinventors I observe an empirical average degree of $\bar{d} = 4.79$ and $\alpha = 3.00$, so that I obtain $m = 8$ and $p = 0.56$. In a similar way, I observe for the alliance network a power-law decay with parameter $\alpha = 2.59$ and an average degree of $\bar{d} = 1.79$. From these values I can compute $m = 2$ and $p = 0.89$. Using the exponent of the power-law tail of the degree distribution together with the average degree for the trade network yields conditions on $p$ and $m$ which cannot be satisfied for the model with $n_s = 1$. Moreover, the monotonic decaying behavior of the empirical average nearest neighbor degree points at higher values of $n_s$ than one. I thus set the starting value of $n_s$ for the trade network to 50. I use the same initial values for both, the directed and the undirected network formation algorithms.

The proposal distribution $q_s(\mathbf{\Theta} \to \mathbf{\Theta}')$ is a truncated normal distribution $\mathbf{\Theta}' \sim \mathcal{N}(\mathbf{\Theta}, \mathbf{\Sigma}_s)\mathbb{1}_{[\mathbf{\Theta}^{\min}, \mathbf{\Theta}^{\max}]}(\mathbf{\Theta})$ for each parameter $\theta \in \mathbf{\Theta}$ with a diagonal variance-covariance matrix $\mathbf{\Sigma}_s = \text{diag}\{\sigma^2_{1,s}, \ldots, \sigma^2_{L,s}\}$. More precisely, for each continuous parameter $\theta_i \in \mathbb{R}_+$ (i.e. $p, \beta$) I choose a proposal distribution given by

$$q_s(\theta_i \to \theta_i') = \frac{\phi(\theta'|\theta, \sigma^2_{i,s})}{\Phi(\theta_i^{\max}|\theta_i, \sigma^2_{i,s}) - \Phi(\theta_i^{\min}|\theta_i, \sigma^2_{i,n})} \mathbb{1}_{[\theta_i^{\min}, \theta_i^{\max}]}(\theta_i'),$$

where $\phi(\theta|\mu, \sigma^2)$ and $\Phi(\theta|\mu, \sigma^2)$ are the pdf and cdf, respectively, of a normally distributed random variable with mean $\mu$ and variance $\sigma^2$. For the discrete parameters $\theta_i \in \mathbb{Z}_+$ (i.e. $n_s$, while $m$ is set through the condition $\bar{d} = mp$ when the network is directed while $\bar{d} = 2pm$ when it is undirected), I choose a proposal distribution given by

$$q_s(\theta_i \to \theta_i') = \frac{\Phi(\theta_i' + 1|\theta, \sigma^2_{i,s}) - \Phi(\theta_i'|\theta, \sigma^2_{i,s})}{\Phi(\theta_i^{\max}|\theta_i, \sigma^2_{i,s}) - \Phi(\theta_i^{\min}|\theta_i, \sigma^2_{i,s})} \mathbb{1}_{[\theta_i^{\min}, \theta_i^{\max}]}(\theta_i').$$

During the "burn-in" phase (Chib, 2001), I consider a monotonic decreasing sequence of thresholds given by $\epsilon_{i,s} \geq \epsilon_{i,s+1} \geq \ldots \geq \epsilon_i^{\min}$ with $\epsilon_{i,s+1} = \max\left\{(1-\gamma)\epsilon_{i,s}, \epsilon_i^{\min}\right\}$ and $\gamma = 0.05$. Similarly, I assume a decreasing sequence of variances $\sigma^2_{i,s} \geq \sigma^2_{i,s+1} \geq \ldots \geq (\sigma_i^{\min})^2$ with $\sigma^2_{i,s+1} = \max\left\{(1-\gamma)\sigma^2_{i,s}, (\sigma_i^{\min})^2\right\}$ for the proposal distribution $q_s(\theta_i \to \theta_i')$. As a measure of distance I choose the Euclidean distance $\Delta(\mathbf{S}_i, \mathbf{S}_i^o) = \sqrt{\sum_{j=1}^T \left(S_{i,j} - S^o_{i,j}\right)^2}$. The parameter ranges are $n_s \in \{1, \ldots, 100\}$, $p \in [0, 1]$ and $\beta \in [0, 50]$. The parameters $\epsilon_i^{\min}$ are choose sufficiently small after long experimentation with different starting values and burn-in periods.

The estimation results can be seen in Table 2. The table shows the average over the simulated parameter values, the standard error over these values, the corrected standard error computed over batches of length 10 (Chib, 2001), the integrated autocorrelation time $\iota_\theta$ (Sokal, 1996) and $p_n(\theta)$ is Geweke's spectral density diagnostic indicating the convergence of the chain (Brooks and Roberts, 1998, Geweke, 1992).

Table 2: Estimation of the model parameters $\theta \in \mathbf{\Theta} = (m, \beta, n_s, p)$ for the network of inventors, the network of firms and the trade network. We have considered two model specifications: the case of entering agents observing only the out-neighbors of selected incumbents (Model A), as in Definition 1, and the case of entrants observing both, the out- and in-neighbors of the selected incumbents (Model B), as discussed in Section 6.1.[a] The table shows simulated averages of the parameters and their standard deviations,[b] after the chain has converged.[c]

| | Model A | | | | | Model B | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu_\theta$ | $\bar{\sigma}_\theta$ | $\sigma_\theta$ | $\iota_\theta$ | $p_n(\theta)$ | $\mu_\theta$ | $\bar{\sigma}_\theta$ | $\sigma_\theta$ | $\iota_\theta$ | $p_n(\theta)$ |
| Inv. Netw. | | | | | | | | | | |
| $T = 27495$ | | | | | | | | | | |
| $n_s$ | 1 | 0 | 0 | 0 | 1 | 1.00 | 0.10 | 0.01 | 29.65 | 0.89 |
| $p$ | 0.60 | 0.05 | 0.00 | 33.66 | 0.99 | 0.58 | 0.08 | 0.00 | 8.06 | 0.95 |
| $m$ | 8.44 | 0.76 | 0.068 | 30.52 | 0.98 | 4.45 | 0.67 | 0.034 | 6.87 | 0.92 |
| $\beta$ | 0.75 | 0.52 | 0.09 | 199.59 | 0.45 | 1.46 | 1.28 | 0.23 | 137.33 | 0.56 |
| Firm Netw. | | | | | | | | | | |
| $T = 7374$ | | | | | | | | | | |
| $n_s$ | 32.63 | 1.94 | 0.40 | 649.15 | 0.99 | 32.08 | 5.84 | 1.30 | 647.16 | 0.98 |
| $p$ | 0.69 | 0.27 | 0.05 | 407.08 | 0.42 | 0.82 | 0.11 | 0.02 | 113.94 | 0.71 |
| $m$ | 3.80 | 3.09 | 0.54 | 190.96 | 0.68 | 1.04 | 0.20 | 0.04 | 171.44 | 0.58 |
| $\beta$ | 0.01 | 0.00 | 0.00 | 87.04 | 0.46 | 0.01 | 0.01 | 0.00 | 66.65 | 0.80 |
| Trade Netw. | | | | | | | | | | |
| $T = 196$ | | | | | | | | | | |
| $n_s$ | 48.46 | 2.21 | 0.44 | 323.43 | 0.95 | 51.79 | 3.39 | 0.65 | 874.22 | 0.95 |
| $p$ | 0.34 | 0.08 | 0.01 | 132.11 | 0.77 | 0.58 | 0.19 | 0.02 | 212.69 | 0.92 |
| $m$ | 130.73 | 25.56 | 2.67 | 124.94 | 0.80 | 41.98 | 18.07 | 2.12 | 232.91 | 0.92 |
| $\beta$ | 0.57 | 0.47 | 0.09 | 471.46 | 0.79 | 1.69 | 1.25 | 0.27 | 927.42 | 0.92 |

[a] The number of iterations of the MC for each model and each data set considered is $n = 10000$. $\iota_\theta$ is the integrated autocorrelation time, which should be much smaller than the number $n$ of iterations (Sokal, 1996).

[b] $\bar{\sigma}_\theta$ is the simulation standard deviation of the respective parameter, while $\sigma_\theta$ is the standard deviation calculated from batch means (of length 10) (Chib, 2001).

[c] $p_n(\theta)$ 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 (Brooks and Roberts, 1998, Geweke, 1992).

## C. Undirected Links

In the following network formation process we allow entering agents to observe not only the out-neighbors of incumbent agents but also their in-neighbors. The resulting network can then be viewed as an undirected graph. The precise definition of the network growth process is given below:

**Definition 3.** *For a fixed $T \in \mathbb{N} \cup \{\infty\}$ we define a network formation process $(G_t)_{t\in[T]}$ as follows. Given the initial graph $G_1 = \ldots = G_{m+1} = K_{m+1}$, for all $t > m+1$ the graph $G_t$ is obtained from $G_{t-1}$ by applying the following steps:*

**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 without replacement $n_s \geq 1$ agents $i \in \mathcal{P}_{t-1}$ uniformly at random and adding $i$ as well as the 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 $f_t(G_{t-1}, j) + \varepsilon_{tj}$.*

### C.1. Large Observation Radius

We first consider the case of $\mathcal{S}_t = \mathcal{P}_{t-1}$. Let $k_j(t)$ denote the degree of agent $j$ at time $t$. Considering only the leading terms in $O\left(\frac{1}{t}\right)$ we can write the probability that an agent $j \in \mathcal{P}_{t-1}$ to receive a link by the entrant $t$ as follows

$$K_t^{\beta}(j|G_{t-1}) \approx \frac{m}{1+2\beta m}\frac{1+\beta d_{G_{t-1}}(j)}{t}. \tag{C.1}$$

Using the recursive Equation (A.3) with the attachment kernel in Equation (C.1) yields the following proposition.

**Proposition 12.** *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^{\beta})_{t\in\mathbb{N}}$ introduced in Definition 3 with $n_s$ large enough such that $\mathcal{S}_t = \mathcal{P}_{t-1}$ for every $t > m+1$. Then, for all $k \geq 0$ we have in the limit $\beta \to 0$ that $P_t(k) \to P^{\beta}(k)$, where*

$$P^{\beta}(k) = \frac{(1+2m\beta)\Gamma\left(k+\frac{1}{\beta}\right)\Gamma\left(3+\frac{1}{\beta}+\frac{1}{m\beta}\right)}{(1+m+2m\beta)\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(k+3+\frac{1}{\beta}+\frac{1}{m\beta}\right)}. \tag{C.2}$$

Proof of Proposition 12. Equation (C.2) follows directly from the recursion in Equation (A.3) and the attachment kernel in Equation (C.1). □

From Equation (C.2) we find that the large $k$ behavior of the degree distribution follows a power-law as $P^{\beta}(k) \sim k^{-\left(3+\frac{1}{m\beta}\right)}$. In the continuum approximation we can write for the dynamics of $k_s(t)$ using Equation (C.1) as

$$\frac{dk_s(t)}{dt} = \frac{m}{1+2\beta m}\frac{1+\beta k_j(t)}{t},$$

with the initial condition $k_s(s) = m$. The solution is given by

$$k_s(t) = \frac{1}{\beta}\left((1+\beta m)\left(\frac{t}{s}\right)^{\frac{\beta m}{1+2\beta m}} - 1\right), \tag{C.3}$$

and we obtain for the degree distribution in the continuum approximation

$$P^\beta(k) = \frac{1+2\beta m}{m}(1+\beta m)^{2+\frac{1}{\beta m}}(1+\beta k)^{-\left(3+\frac{1}{m\beta}\right)}, \tag{C.4}$$

with $\int_0^\infty P^\beta(k)dk = 1$. This yields the same asymptotic behavior of the degree distribution as in Equation (C.2).

Next, we turn to the average nearest neighbor connectivity.

**Proposition 13.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 3 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that Equation (C.3) holds. Then in the limit $\beta \to 0$ the nearest-neighbor degree distribution is given by*

$$k_{nn}(k) = \frac{1}{\beta^2 k}\left(1 + \frac{1+\beta k}{1+\beta m}\left(\beta^2 R_s(s) - 1 + (1+\beta m)^2 \ln\left(\frac{1+\beta k}{1+\beta m}\right)\right)\right), \tag{C.5}$$

*where $a = \frac{m}{1+2\beta m}$, the initial condition*

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

*and $s = t\left(\frac{1+\beta m}{1+\beta k}\right)^{2+\frac{1}{m\beta}}$.*

Asymptotically, only the last term in Equation (C.5) is relevant and we obtain

$$k_{\mathrm{nn}}(k) \sim \frac{1+\beta m}{\beta}\ln\left(\frac{1+\beta k}{1+\beta m}\right), \tag{C.6}$$

as $k \to \infty$.

Proof of Proposition 13. Denote by $R_s(t) = \sum_{j\in\mathcal{N}_{G_t}(s)} k_j(t)$ the sum of the degrees of the neighbors of vertex $s$ at time $t$. We can write

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

where we have denoted by $a = \frac{m}{1+2\beta m}$ and using the fact that$1+\beta k_s(t) = (1+\beta m)\left(\frac{t}{s}\right)^{\beta a}$ from Equation (C.3) under the continuum approximation. The initial condition is given by

$$R_s(s) = \sum_{j=1}^{s}\frac{a}{s}(1+\beta k_j(s))(1+k_j(s)) = \frac{a(1-\beta)(1-2m\beta)}{\beta} + \frac{a}{s}\sum_{j=1}^{s}(1+\beta k_j(s))^2.$$

Using the fact that

$$1+\beta k_j(s) = (1+\beta m)\left(\frac{s}{j}\right)^{\beta a}, \tag{C.7}$$

we obtain

$$R_s(s) = \frac{a(1-\beta)(1-2m\beta)}{\beta} + \frac{a(1+\beta m)^2}{\beta} s^{2\beta a-1} H(s, 2\beta a).$$

We then get

$$R_s(t) = \frac{1}{\beta^2}\left(1 + \left(a\beta(1+\beta m)^2\left(\frac{1}{s}H(s,2a\beta) + (1+m\beta)\ln\left(\frac{t}{s}\right)\right) - 1 + \beta^2 b\right)\left(\frac{t}{s}\right)^{a\beta}\right). \tag{C.8}$$

Using once again Equation (C.7) and inserting into $k_{\text{nn}} = \frac{R_s}{k}$ delivers Equation (C.5). □

Moreover, we can compute the clustering degree distribution as provided in the next proposition.

**Proposition 14.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 3 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that Equation (C.3) holds. Then in the limit $\beta \to 0$ the clustering degree distribution is given by*

$$C(k) = \frac{2}{k(k-1)}\left(M_s + \frac{b}{s(1-2a\beta)}\left(d + a\beta s^{2a\beta-1}\left(1-\left(\frac{t}{s}\right)^{2a\beta-1}\right)H_s^{2\beta a} - \left(\frac{t}{s}\right)^{2a\beta-1}\left(d + \ln\left(\frac{t}{s}\right)^{a\beta}\right)\right)\right), \tag{C.9}$$

*where $s = t\left(\frac{1+m\beta}{1+k\beta}\right)^{2+\frac{1}{m\beta}}$, $a = \frac{a}{1+2\beta m}$, $b = \frac{m(m-1)(1+\beta m)^2}{\beta(1+2\beta m)}$, $c = \frac{\beta m + a\beta(1-\beta)(1-2m\beta)}{(1+\beta m)^2}$, $d = \frac{c+a\beta(1-2c)}{1-2a\beta}$, the Harmonic number is defined as $H_s^a \equiv \sum_{j=1}^s j^{-a}$ and the initial condition is given by*

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

The large $k$ behavior of the clustering coefficient is dominated by the second term in Equation (C.9), yielding

$$C(k) \sim \frac{2bd}{k(k-1)s(1-2a\beta)} = \frac{1}{t}\frac{2bd}{(1-2a\beta)(1+m\beta)^{2+\frac{1}{m\beta}}}\frac{(1+\beta k)^{2+\frac{1}{m\beta}}}{k(k-1)} = O\left(\frac{1}{t}k^{\frac{1}{m\beta}}\right), \quad k\to\infty. \tag{C.10}$$

Proof of Proposition 14. Let $M_s(t)$ denote the number of triangles containing $s$ at time $t$. We have that

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

With $R_s(t)$ from Equation (C.8) and Equation (C.7) we obtain

$$\frac{dM_s(t)}{dt} = \frac{b}{t^2}\left(\frac{t}{s}\right)^{2\beta a}\left(c + \ln\left(\frac{t}{s}\right)^{\beta a} + a\beta(s)^{2\beta a-1}H_s^{2\beta a}\right),$$

where $a = \frac{a}{1+2\beta m}$, $b = \frac{m(m-1)(1+\beta m)^2}{\beta(1+2\beta m)}$, $c = \frac{\beta m + a\beta(1-\beta)(1-2m\beta)}{(1+\beta m)^2}$ and the Harmonic number is defined as $H_s^a \equiv \sum_{j=1}^{s} j^{-a}$. The solution is given by

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

where $d = \frac{c+a\beta(1-2c)}{1-2a\beta}$. Similar to the derivation of Equation (A.19), the initial condition is given by

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

Using Equation (C.7) we then arrive at the expression in Equation (C.9). □

## C.2. Small Observation Radius

Next, we consider the case of a small observation radius $n_s$. The probability that agent $j$ receives a link from the entrant at time $t$, conditional on the sample $\mathcal{S}_t$ (and the current network $G_{t-1}$) when $\beta = 0$ is given by

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

In the following, we assume that $\mathcal{S}_t \approx n_s(\bar{d}+1)$, where the average degree is given by $\bar{d} = 2m$, so that $\mathcal{S}_t \approx n_s(2m+1)$. Note that this assumption is much stronger than the approximation we have made in Equation (3.5). The probability that an agent $j$ receives a link from $t$ is then given by

$$\begin{aligned} K_t^{\beta}(j|G_{t-1}) &= \frac{m}{|\mathcal{S}_t|}\frac{n_s(1+d_{G_{t-1}}(j))}{t} + O\left(\frac{1}{t^2}\right) \approx \frac{m}{n_s(2m+1)}\frac{n_s(1+d_{G_{t-1}}(j))}{t} + O\left(\frac{1}{t^2}\right) \\ &\approx \frac{m}{2m+1}\frac{1+d_{G_{t-1}}(j))}{t}. \end{aligned} \tag{C.11}$$

An analysis following the recursive Equation (A.3) with the attachment kernel in Equation (C.11) yields the following proposition.

**Proposition 15.** *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^{\beta})_{t\in\mathbb{N}}$ of Definition 3 with $\beta = 0$. If $n_s > 1$ or $m > 1$, further assume that Equation (C.11) holds. Then, for all, $k \geq 0$ we have $P_t(k) \to P(k)$, where*

$$P(k) = \frac{(1+2m)\Gamma\left(3+\frac{1}{m}\right)}{m\Gamma\left(3+k+\frac{1}{m}\right)}. \tag{C.12}$$

Proof of Proposition 15. Equation (C.12) follows directly from the recursion in Equation (A.3) and Equation (C.11). □

From Equation (C.12) we find that the degree distribution follows a power-law as $P(k) \sim k^{-\left(3+\frac{1}{m}\right)}$ for large $k$. For the dynamics of $k_s(t)$ in the continuum approximation we get with Equation (C.11) the following differential equation

$$\frac{dk_s(t)}{dt} = \frac{m}{2m+1}\frac{k_s(t)+1}{t}$$

with the solution

$$k_s(t) = (m+1)\left(\frac{t}{s}\right)^{\frac{m}{2m+1}} - 1 \tag{C.13}$$

The degree distribution in the continuum approximation is then given by[39]

$$P(k) = \frac{2m+1}{m}(m+1)^{2+\frac{1}{m}}(1+k)^{-\left(3+\frac{1}{m}\right)}, \tag{C.14}$$

satisfying the normalization condition $\int_0^\infty P(k)dk = 1$.

Next we consider the average nearest neighbor degree.

**Proposition 16.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 3 in the continuum approximation with $n_s$ small enough and assume that Equation (C.13) holds. If $\beta = 0$ then the nearest-neighbor degree distribution is given by*

$$k_{nn}(k) = \frac{1}{k}\left(\left(\frac{t}{s+1}\right)^a \left(a(m+1)^2 s^{2a-1} H_s^{2a} - 1\right) + (m+1)\left(\frac{t}{s}\right)^a \ln\left(\frac{t}{s+1}\right)^a\right), \tag{C.15}$$

*where $a = \frac{m}{2m+1}$, $s = t\left(\frac{k+1}{m+1}\right)^{-\frac{1}{a}}$ and the Harmonic number is defined as $H_s^{2a} \equiv \sum_{j=1}^{s}\frac{1}{j^{2a}}$.*

Proof of Proposition 16. Let $R_s(t) = \sum_{j\in\mathcal{N}_{G_t}(s)} k_j(t)$ be the sum of the degrees of the neighbors of vertex $s$ at time $t$. Denoting by $a = \frac{m}{1+2m}$, we have up to leading orders in $O\left(\frac{1}{t}\right)$ that[40]

$$\begin{aligned}\frac{dR_s(t)}{dt} &= \frac{n_s}{t}\sum_{j\in\mathcal{N}_{G_t}(s)}\frac{m}{|\mathcal{S}_t|}k_j(t) + \frac{n_s}{t}\sum_{j=1}^{m} j\frac{\binom{k_s(t)}{j}\binom{|\mathcal{S}_t|-k_s(t)}{m-j}}{\binom{|\mathcal{S}_t|}{m}} \\ &= \frac{a}{t}\left(k_s(t)+R_s(t)\right) = \frac{a}{t}\left((m+1)\left(\frac{t}{s}\right)^a - 1 + R_s(t)\right),\end{aligned}$$

where we have assumed that $|\mathcal{S}_t| \approx n_s(2m+1)$ and used the relation $s = t\left(\frac{k+1}{m+1}\right)^{-\frac{1}{a}}$. The solution is given by

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

[39] Note that the approximation for the degree distribution in Equation (C.14) has also been obtained in Wang et al. (2009).

[40] 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 $O\left(\frac{1}{t^2}\right)$.

and the initial condition is given by

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

Using this equation to solve for $C_s$ delivers Equation (C.15). □

Finally, we can compute the clustering coefficient as given in the following proposition.

**Proposition 17.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 3 in the continuum approximation with $n_s$ small enough and assume that Equation (C.13) holds. Let $a = \frac{m}{2m+1}$ and $b = \frac{2a(m-1)}{n_s(2m+1)-1}$ with $a > b > 0$. If $\beta = 0$ then the average clustering coefficient of an agent with degree $k$ is bounded by $\underline{C}(k) \le C(k) \le \overline{C}(k)$, where*

$$\underline{C}(k) = \frac{2}{(a-b)k(k-1)}\left(a - (a+mb)\left(\frac{1+k}{1+m}\right)^{\frac{b}{a}} + bk\right), \tag{C.16}$$

*and*

$$\overline{C}(k) = \frac{2}{(a-b)k(k-1)}\left(a + \left(\binom{m}{2}(a-b) - (a+mb)\right)\left(\frac{1+k}{1+m}\right)^{\frac{b}{a}} + bk\right), \tag{C.17}$$

*and the property that $C(k) = O\left(\frac{1}{k}\right)$.*

Proof of Proposition 17. We need to consider the cases we have encountered already in the proof of Proposition 8 for a vertex $s$ to form an additional triangle by an entrant $t$ (see Figure 9). The expected number of triangles associated with case (i) is given by

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

where we have assumed that $|\mathcal{S}_t| = n_s(2m+1)$. Similarly, for case (ii) we get

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

and for case (iii) we obtain

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

Denoting by $a = \frac{m}{2m+1}$ and $b = \frac{2a(m-1)}{n_s(2m+1)-1}$ we can add cases (i), (ii) and (iii) to get

$$\frac{dM_s(t)}{dt} = \frac{2a(m-1)}{t(n_s(2m+1)-1)}(k_s(t)+M_s(t)) = \frac{b}{t}\left(\left((m+1)\left(\frac{t}{s}\right)^a - 1 + M_s(t)\right)\right).$$

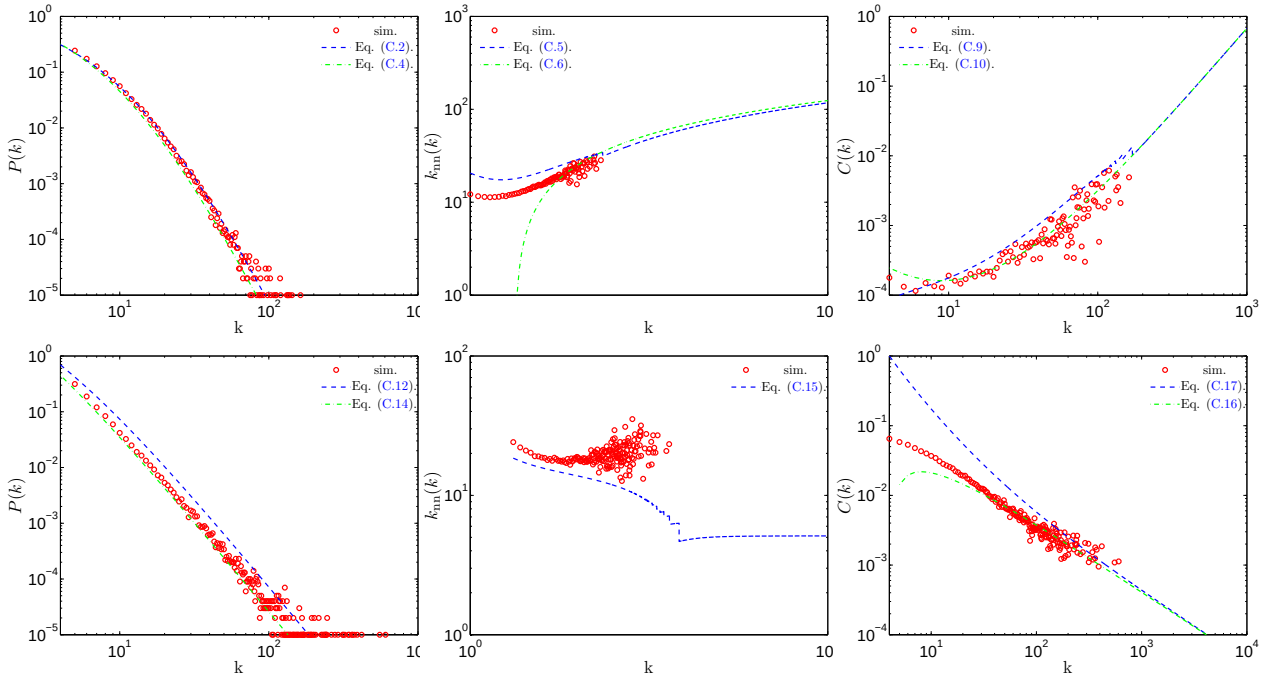



Figure 10: (Top row) Comparison of simulation results with the theoretical predictions for $T = 10^5$, $\mathcal{S}_t = \mathcal{P}_{t-1}$ and $m = 4$ with $\beta = 0.1$ under the linear approximation to the attachment kernel. (Bottom row) Comparison of simulation results for $T = 10^5$ and $n_s = m = 4$ ($\beta = 0$) with the theoretical predictions. Comparing the results of global and local information, we find that they differ mainly in the clustering degree distribution.

Using as a lower bound for the initial condition $M_s(s) \geq 0$ and an upper bound $M_s(s) \leq \binom{m}{2}$ as well as $s = \left(\frac{1+k}{1+m}\right)^{-1/a} t$, we obtain the corresponding bounds for the clustering coefficient in Equations (C.16) and (C.17). Both bounds decay as $\frac{2b}{a-b}\frac{1}{k}$ for large $k$ and their difference vanishes for large $k$, implying that also $C(k) = O\left(\frac{1}{k}\right)$. □

# D. Heterogeneous Linking Opportunities

In this section we assume that not all agents become active during the network formation process. More precisely, we assume that only a fraction $p \in (0,1)$ of the population of agents forms links, while the remaining agents stay passive throughout the whole evolution of the network. We assume that initially, agents in $[T] = \{1, 2 \ldots, T\}$ are randomly assigned to sets $\mathcal{P}_1$ with probability $1-p$ and to $\mathcal{A}_1$ with probability $p$, such that $|\mathcal{A}_1| = \lfloor pT \rfloor$ and $|\mathcal{P}_1| = \lceil (1-p)T \rceil$. The agents in $[m]$ are all connected to each other and form a complete graph $K_m$. At time $t \leq m+1$ these agents are all in the set $\mathcal{P}_t$. The network evolution process is then defined as follows.

**Definition 4.** *For a fixed $T \in \mathbb{N} \cup \{\infty\}$ we define a network formation process $(G_t)_{t\in[T]}$ as follows. Given the initial graph $G_1 = \ldots = G_{m+1} = K_{m+1}$, for all $t \in [T] \backslash \{1, \ldots, m+1\}$ the graph $G_t$ is obtained from $G_{t-1}$ by applying the following steps:*

**Growth:** *Given $\mathcal{P}_1$ and $\mathcal{A}_1$, for all $t > m$, if agent $t \in \mathcal{A}_{t-1}$ then 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. Otherwise, set $\mathcal{P}_t = \mathcal{P}_{t-1}$ and $\mathcal{A}_t = \mathcal{A}_{t-1}$.*

**Network sampling:** *If $t \in \mathcal{A}_{t-1}$ then $t$ 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:** *If $t \in \mathcal{A}_{t-1}$, given the sample $\mathcal{S}_t$, agent $t$ creates $X_m \geq 1$, $\mathbb{E}(X_m) = m$ links to agents in $\mathcal{S}_t$ without replacement. For each link, agent $t$ chooses the $j \in \mathcal{S}_t$ that maximizes $f_t(G_{t-1}, j) + \varepsilon_{tj}$.*

The number of links $X_m$ to be created by an entrant is a discrete random variable with expectation $\mathbb{E}(X_m) = m$. The results and approximations we obtain in this section do not depend on the specific distribution we choose for $X_m$. We illustrate this by comparing our theoretical approximations with simulations for a uniform distribution $X_m \sim \mathrm{U}\{1, \ldots, 2m-1\}$ and a Poisson distribution $X_m \sim \mathrm{Pois}(m)$.

## D.1. Large Observation Radius

We first consider the case of a large observation radius such that $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$. Similar to our discussion in Section 3.2, the probability that an agent $j \in \mathcal{P}_{t-1}$ with degree $d_{G_{t-1}}(j)$ receives a link by the entrant at time $t$ up to leading orders in $O\left(\frac{1}{t}\right)$ is given by

$$K_t^{\beta}(j|G_{t-1}) \approx \frac{pm}{1+\beta pm} \frac{1+\beta d_{G_{t-1}}(j)}{t}. \tag{D.1}$$

Following the recursive Equation (A.3) with the attachment kernel in Equation (D.1) yields the following proposition.

**Proposition 18.** *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^{\beta})_{t\in\mathbb{N}}$ introduced in Definition 4 with $n_s$ large enough such that $\mathcal{S}_t = \mathcal{P}_{t-1}$ for every $t > m+1$. Then, for all $k \geq m$ we have in the limit $\beta \to 0$ that $P_t^{\beta}(k) \to P^{\beta}(k)$ almost surely, where*

$$P^{\beta}(k) = \frac{1+\beta mp}{1+mp(1+\beta)} \frac{\Gamma\left(\frac{1}{\beta}+k\right)\Gamma\left(2+\frac{1+mp}{\beta mp}\right)}{\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(2+\frac{1+mp}{\beta mp}+k\right)}. \tag{D.2}$$

Proof of Proposition 18. Equation (D.2) follows directly from the recursion in Equation (A.3) and the attachment kernel in Equation (D.1). □

From the attachment kernel in Equation (D.1) we can write for the dynamics of the in-degree $k_s(t)$ of vertex $s$ at time $t$ in the continuum approximation

$$\frac{dk_s(t)}{dt} = \frac{pm}{1+\beta pm}\frac{1+\beta k_j(t)}{t},$$

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

$$k_s(t) = \frac{1}{\beta}\left(\left(\frac{t}{s}\right)^{\frac{\beta pm}{1+\beta pm}} - 1\right), \tag{D.3}$$

and we obtain for the degree distribution in the continuum approximation

$$P^\beta(k) = \frac{1+\beta mp}{mp}(1+\beta k)^{-\left(2+\frac{1}{\beta mp}\right)}, \tag{D.4}$$

with $\int_0^\infty P^\beta(k)dk = 1$. For $p = 1$ we recover the distribution in Equation (A.12). The degree distribution from Equations (D.2) and (D.4) can be seen in Figure 11.

Next we consider the average nearest neighbor degrees. We can state the following proposition.

**Proposition 19.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 4 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that Equation (D.3) holds. Then in the limit $\beta \to 0$ the nearest-neighbor degree distribution is given by*

$$k_{nn}^-(k) = \frac{1}{\beta^2 k}\left(1 + (1+\beta k)(\ln(1+\beta k) - 1)\right), \tag{D.5}$$

*and the average nearest neighbor out-degree is given by*

$$k_{nn}^+(k) = \frac{1}{\beta^2 m}\left(\left(\beta m(1+p(\beta-1)) + \frac{a}{s}s^{2a}\zeta(s,2a)\right)\left(\frac{t}{s+1}\right)^a - m\beta\right), \tag{D.6}$$

*where $a = \frac{\beta mp}{1+\beta mp}$, $s = t(1+\beta k)^{-\frac{1}{a}}$.*

Observe that Equation (D.5) is independent of $p$ and identical to Equation (5.1) from Proposition 5. From Proposition 19 we find that for large $k$,f the average nearest in-neighbor connectivity grows logarithmically with $k$ while the average nearest out-neighbor connectivity becomes independent of $k$ and grows with the network sizes as $t^{\frac{\beta mp}{1+\beta mp}}$.

Proof of Proposition 19. Let $R_s^-(t) = \sum_{j\in\mathcal{N}_{G_t}^-(s)} k_j(t)$. Up to leading orders in $O\left(\frac{1}{t}\right)$ we then have that

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

where we have denoted by $a = \frac{\beta mp}{1+\beta mp}$. The initial condition is given by $R_s^- = 0$. The solution is

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

Using the fact that $\frac{t}{s} = (1+\beta k)^{\frac{1}{a}}$ from Equation (D.3), we obtain

$$R_s^-(t) = \frac{1}{\beta^2}\left(1 + (1+\beta k)(-1 + \ln(1+\beta k))\right).$$

With $k_{\mathrm{nn}}(k) = \frac{R_s^-}{k}$, the expression in Equation (D.5) follows.

Next we turn to the average nearest out-neighbor degree. Consider a vertex $s$ which has received a linking opportunity upon entry. Let $R_s^+(t) = \sum_{j \in \mathcal{N}_{G_t}^+(s)} k_j(t)$. Then up to leading orders in $O\left(\frac{1}{t}\right)$ we obtain

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

where $a = \frac{\beta pm}{1+\beta pm}$. The solution is given by

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

The constant $C_s$ is determined by the initial condition

$$R_{s+1}^+ = \sum_{j=1}^{s} \frac{a}{s}\left(\frac{1}{\beta} + k_j(t)\right)(k_j(t)+1) = \frac{a}{\beta^2}\left(\beta - 1 + mp\beta(\beta-1) + s^{2a-1}H(s,2a)\right).$$

We then obtain

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

with $s = t(1+\beta k)^{-\frac{1}{a}}$ from Equation (D.3) and $k_{\mathrm{nn}}^+ = \frac{R_s^+(k)}{m}$. □

Moreover, we can derive the clustering degree distribution.

**Proposition 20.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 4 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that Equation (D.3) holds. Then in the limit $\beta \to 0$ the clustering degree distribution is given by*

$$C(k) = \frac{2}{(k+pm)(k+pm-1)}\frac{a(m-1)}{mp\beta^3 b^2 s}\left(sb^2\frac{mp\beta^3}{a(m-1)}M_s + \left((1+\beta k)^b - 1\right)\right.$$
$$\left.\times\left(b\left(\frac{s}{s+1}\right)^a\left(c + as^{2a-1}\zeta(s,2a)\right) - 1\right) + b(1+\beta k)^b \ln(1+\beta k)\right), \tag{D.7}$$

*where* $a = \frac{\beta mp}{1+\beta mp}$, $b = 2 - \frac{1}{a}$, $c = \beta m(1 + p(\beta - 1))$, *the initial condition is given by*

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

*and* $s = t(1+\beta k)^{-\frac{1}{a}}$.

For large $k$ (and small $s$, respectively) the first term in the initial condition $M_s$ dominates, and the behavior of the clustering coefficient is given by

$$C(k) \sim \frac{2t^{-2(1-a)}(1+k\beta)^{2\left(\frac{1}{a}-1\right)}}{(k+pm)(k+pm-1)}\frac{mp(m-1)}{(1+\beta pm)^2}\sum_{i=1}^{m} i^{-a}\sum_{j=i+1}^{m} j^{-a}. \tag{D.8}$$

We see that this expression grows with $k$ as a power-law with exponent $2\left(\frac{1}{a}-2\right) = -2 + \frac{2}{mp\beta}$.[41] Moreover, we find that the clustering coefficient is decreasing with the network size as $t^{-2(1-a)} = t^{-\frac{2}{1+mp\beta}}$.

Proof of Proposition 20. We need to consider the same cases as in the proof of Proposition 7. The probability associated with case (i) in Figure 8 is given by

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

Similarly, for the probability of case (ii) in Figure 8 we obtain

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

With $R^+_s$ and $R^-_s$ given by Equations (D.5) and (D.5), respectively, we obtain

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

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

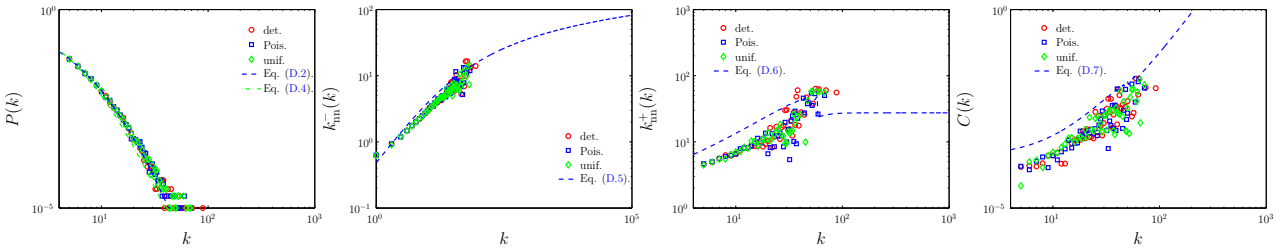



Figure 11: Comparison of simulation results with theoretical prediction of the link formation process in Definition 4 under global information with $p = 0.5$, $m = 4$, $\beta = 0.1$ and $T = 10^5$. We show simulations for a uniform distribution $X_m \sim \mathrm{U}\{1, 2m-1\}$ and a Poisson distribution $X_m \sim \mathrm{Pois}(m)$ with expectation $\mathbb{E}(X_m) = m$.

where we have denoted by $c = \beta m(1 + p(\beta - 1))$. The initial condition is given by

$$\begin{aligned} M_{s+1} &= p\frac{m(m-1)}{2}\sum_{j\neq i}^{s}\frac{1+\beta k_i(s)}{(1+\beta pm)s}\frac{1+\beta k_j(s)}{(1+\beta pm)s}\left(\Theta(m+1-i)\Theta(m+1-j)\right.\\ &\left.+\Theta(i-j)\Theta(j-m)pm\frac{1+\beta k_j(i)}{(1+\beta pm)(i-1)}+\Theta(j-i)\Theta(i-m)pm\frac{1+\beta k_i(j)}{(1+\beta pm)(j-1)}\right)\\ &= \frac{mp(m-1)s^{2a-2}}{(1+\beta pm)^2}\left(\sum_{i=1}^{m}\frac{1}{i^a}\sum_{j=i+1}^{m}\frac{1}{j^a}+\frac{2mp}{1+\beta pm}\sum_{i=m+1}^{s}\frac{1}{i^{2a}}\sum_{j=i+1}\frac{1}{j-1}\right), \end{aligned} \tag{D.9}$$

where we have denoted by $a = \frac{\beta pm}{1+\beta pm}$. The initial condition $M_{s+1}$ together with Equation (D.9) deliver

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

Together with the initial condition, this is the expression in Proposition 20. □

Next, we turn to the analysis of the connectivity of the networks generated by our model. We consider only the simple case where $m = 1$ and the limit of strong noise with $\beta \to 0$, where the network formation process follows a uniformly grown random graph.

**Proposition 21.** *Let $N_s(t)$ denote the number of components of size $s$ at time $t$. Consider the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ of Definition 4 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$. Assume that $m = 1$ and $\beta = 0$. If $p < 1$, then there exists no giant component and the asymptotic (finite) component size distribution $P(s) = \lim_{t\to\infty}\frac{N_s(t)}{t}$ is given by*

$$P(s) = \frac{(1-p)\Gamma\left(\frac{1}{p}\right)\Gamma\left(s\right)}{p^2\Gamma\left(1+\frac{1}{p}+s\right)}. \tag{D.10}$$

*When $p = 1$ then there exists a giant component encompassing all nodes.*

Proof of Proposition 21. Let $N_s(t)$ denote the number of components of size $s$ at time $t$. For

$m = 1$, the entrant $t$ forms only a single link and we need only consider the case of the component with size $s-1$ to receive a link in the contribution to the growth of $N_s(t)$. It then follows that

$$\begin{aligned}\mathbb{E}_t\left[N_1(t+1)|G_t\right] =& N_1(t) + (1-p) - p\frac{N_1(t)}{t},\\ \mathbb{E}_t\left[N_s(t+1)|G_t\right] =& N_s(t) + p\frac{(s-1)N_{s-1}(t)}{t} - p\frac{sN_s(t)}{t}, \quad s \geq 2.\end{aligned}$$

Denote by $n_s(t) = \frac{\mathbb{E}_t[N_s(t)]}{t}$. Taking expectations in the above equations delivers

$$\begin{aligned}n_1(t+1)(t+1) =& n_1(t)t + (1-p) - pn_1(t),\\ n_s(t+1)(t+1) =& n_s(t)t + p(s-1)n_{s-1}(t) - psn_s(t), \quad s \geq 2.\end{aligned}$$

For the stationary distribution $P(s) = \lim_{t\to\infty} n_s(t)$ we then get

$$\begin{aligned}P(1) =& \frac{1-p}{1+p},\\ P(s) =& \frac{p(s-1)}{1+ps}P(s-1), \quad s \geq 2.\end{aligned}$$

From this recursive equation we obtain

$$P(s) = P(1)p^{s-1}\prod_{k=2}^{s}\frac{k-1}{1+pk} = \frac{(1-p)\Gamma\left(\frac{1}{p}\right)\Gamma(s)}{p^2\Gamma\left(1+\frac{1}{p}+s\right)},$$

which is Equation (D.10).

We next consider the generating function of the component size distribution $g(x) = \sum_{s=1}^{\infty} sP(s)x^s$. Observe that $g(1) = \sum_{s=1}^{\infty} sP(s)$ the fraction of nodes in finite components. In the absence of a giant component (that grows with $t$), we must have that $g(1) = 1$. Inserting Equation (D.10) into $g(x)$ we find that $g(1) = 1$ as long as $p < 1$. Hence, the critical probability for the emergence of a giant component is $p = 1$. □

From Equation (D.10) we find that the component size decays as a power law with exponent $1 + \frac{1}{p}$, i.e.

$$P(s) = \frac{1-p}{p^2}\Gamma\left(\frac{1}{p}\right)s^{-\left(1+\frac{1}{p}\right)}\left(1 + O\left(\frac{1}{s}\right)\right).$$

We finally note that when $\beta \to 0$, the probability that a component $H \in G_{t-1}$ of size $s$ receives a link at time $t$, and thus grows by one, is given by

$$p\sum_{i\in H}\frac{1+\beta k_i(t)}{(1+\beta p)t} = \frac{p}{(1+\beta p)t}\sum_{i\in H}(s+\beta k_i(t)) \approx \frac{sp}{t},$$

where we have used the approximation $\sum_{i\in H} k_i(t) \approx sp$. This is the same probability for the growth of a component of size $s$ as in the case of $\beta = 0$ and hence we obtain the same component size distribution as in Equation (D.10).

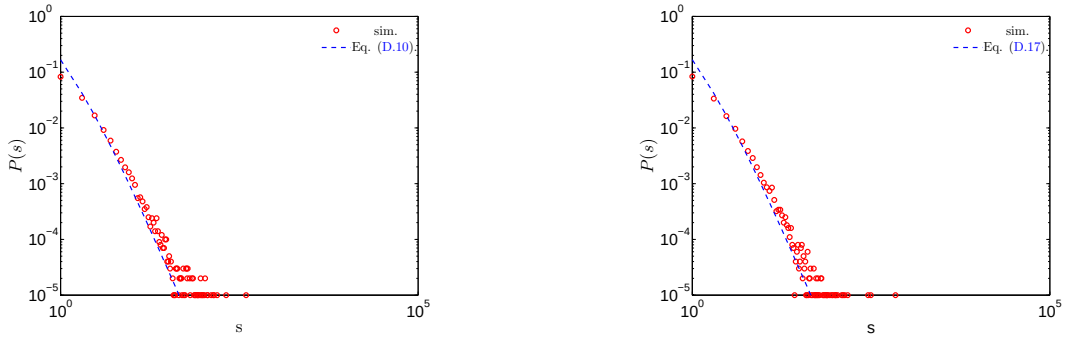



Figure 12: Comparison of simulation results with theoretical predictions for the component size distribution $P(s)$ of the link formation process in Definition 4 under global information with $p = 0.5$, $m = 1$, $\beta = 0$ and $T = 10^5$ (left panel); with $p = 0.5$, $n_s = 1$, $m = 4$, $\beta = 0$ and $T = 10^5$ (right panel).

## D.2. Small Observation Radius

Next, we consider the case of a small observation radius corresponding to small values of $n_s$. Similar to our discussion in Section 3.2, the probability that an agent $j \in \mathcal{P}_{t-1}$ with degree $d_{G_{t-1}}(j)$ receives a link by the entrant at time $t$ up to leading orders in $O\left(\frac{1}{t}\right)$ is given by

$$K_t^{\beta}(j|G_{t-1}) \approx \frac{pm}{1+m}\frac{d_{G_{t-1}}(j)+1}{t}. \tag{D.11}$$

Using the recursive solution of Equation (A.3) we can state the following proposition.

**Proposition 22.** *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^{\beta})_{t\in\mathbb{N}}$ of Definition 4 with $\beta = 0$. Further assume that Equation (D.11) holds. Then, for all, $k \geq 0$ we have $P_t(k) \to P(k)$, where*

$$P(k) = \frac{(1+m)k!\Gamma\left(2+\frac{m+1}{mp}\right)}{(1+m(1+p))\Gamma\left(2+\frac{m+1}{mp}+k\right)}. \tag{D.12}$$

Proof of Proposition 22. Equation (D.12) follows directly from the recursion in Equation (A.3) and Equation (D.11). □

With Equation (D.11) it follows for the dynamics of $k_s(t)$ in the continuum approximation

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

with the solution

$$k_s(t) = \left(\frac{t}{s}\right)^{\frac{pm}{1+m}} - 1. \tag{D.13}$$

The degree distribution in the continuum approximation is then given by

$$P(k) = \frac{1+m}{pm}(1+k)^{-\left(1+\frac{1+m}{pm}\right)}, \tag{D.14}$$

with $\int_0^\infty P(k)dk = 1$. For large $k$, Equations (D.12) and (D.14) are equivalent. Moreover, for $p = 1$ we recover the distribution in Equation (A.14). Next we turn to the analysis of the average nearest neighbor degree.

**Proposition 23.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 4 in the continuum approximation with $n_s$ small enough and assume that Equation (D.13) holds. If $\beta = 0$ then the average nearest in-neighbor degree distribution is given by*

$$k_{nn}^-(k) = \frac{1}{k}\left(1 + (k+1)(\ln(k+1) - 1)\right) \tag{D.15}$$

*and the average nearest out-neighbor degree distribution is given by*

$$k_{nn}^+(k) = \frac{mp+1}{m+1}k + \frac{p}{m+1}t^{2a-1}(k+1)^{-\frac{2a-1}{a}}\zeta(t(k+1)^{-\frac{1}{a}}, 2a) \tag{D.16}$$

*where $a = \frac{mp}{1+m}$.*

Proof of Proposition 23. In order to derive Equation (D.15), let us denote by $R_s^-(t)$ the sum of the in-neighbors' degrees of a vertex $s$ at time $t$. We then have that

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

where we have denoted by $a = \frac{mp}{1+m}$. The initial condition is $R_s^-(s) = 0$. The solution is given by

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

where we have used the fact that $s = t(k+1)^{-\frac{1}{a}}$ from Equation (D.13). Noting that $k_{\mathrm{nn}}^-(k) = \frac{R_s^-}{k}$ we readily obtain Equation (D.15).

Next, we consider the out-neighbors of $s$. Assume that vertex $s$ has out-degree $m$ and denote by $R_s^+$ the sum of the in-degrees of the out-neighbors of $s$ at time $t$. We then can write

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

The solutions is given by $R_s^+(t) = -\frac{m(1+mp)}{1+m} + C_s t^a$ and the initial condition is

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

so that we get

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

Inserting $s = t(k+1)^{-\frac{1}{a}}$ from Equation (D.13) and using the fact that $k_{\mathrm{nn}}(k) = \frac{R_s^+}{m}$ delivers Equation (D.16) . □

In a similar fashion as in Proposition 8 we can also compute the clustering degree distribution.

**Proposition 24.** *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition 4 in the continuum approximation with $n_s$ small enough and assume that Equation (D.13) holds. If $\beta = 0$ then the average clustering coefficient of an agent with degree $k$ is given by Proposition 8 setting $a = \frac{mp}{m+1}$.*

Proof of Proposition 24. We need to consider the same cases as in the proof of Proposition 8. We take $|\mathcal{S}_t| = n_s(m+1)$ ignoring terms of the order $O\left(\frac{1}{t^2}\right)$. For the probability of case (i) we obtain

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

For case (ii) we get

$$pk_s(t)\frac{n_s}{t}\frac{\binom{n_s(m+1)-2}{m-2}}{\binom{n_s(m+1)}{m}} = p\frac{k_s(t)}{t}\frac{m(m-1)}{n_s(m+1)(n_s(m+1)-1)}.$$

and similarly, for case (iii) we get

$$pM_s(t)\frac{n_s}{t}\frac{\binom{n_s(m+1)-2}{m-2}}{\binom{n_s(m+1)}{m}} = p\frac{M_s(t)}{t}\frac{m(m-1)}{(m+1)(n_s(m+1)-1)}.$$

The dynamics of $M_s(t)$ is then given by

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

with $a = \frac{mp}{m+1}$. This differential equation is identical to (A.20) and hence we obtain the same result as in Proposition 8. □

In the following we study the connectivity of the emerging networks in the network formation process introduced in Definition 4. We restrict our analysis to the case of $n_s = 1$. Observe that the probability that a component of size $s$ grows by one unit due to the attachment of an entrant $t$ is equivalent to the event that $t$ observes one of the nodes in the component when constructing the sample $\mathcal{S}_t$. The probability of this event is $\frac{ps}{t}$. Hence, we obtain the same component size distribution as in Proposition 21. We then can state the following proposition.

**Proposition 25.** *Let $N_s(t)$ denote the expected number of components of size $s$ at time $t$. Consider the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ of Definition 4 with $n_s = 1$. Then the asymptotic component size distribution $P(s) = \lim_{t\to\infty}\frac{N_s(t)}{t}$ is given by*

$$P(s) = \frac{(1-p)\Gamma\left(\frac{1}{p}\right)\Gamma(s)}{p^2\Gamma\left(1+\frac{1}{p}+s\right)}. \tag{D.17}$$

Proof of Proposition 25. The proof follows the one of Proposition 21. □

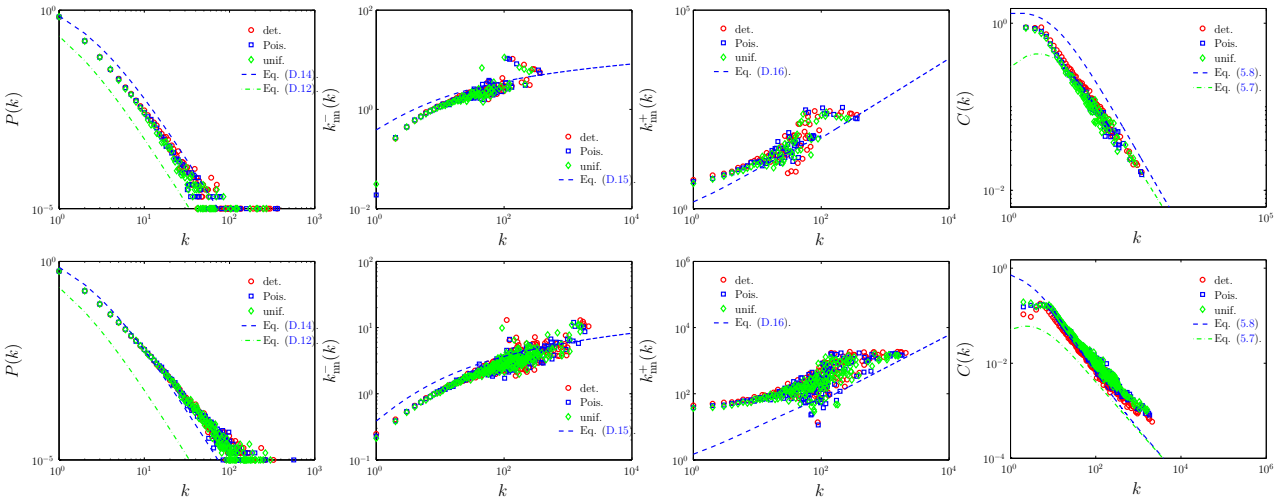



Figure 13: Comparison of simulation results with theoretical predictions of the link formation process in Definition 4 with $p = 0.5$, $n_s = 1$, $m = 4$, $\beta = 0$ and $T = 10^5$ (top row) and $T = 2 \times 10^5$ (bottom row). We show simulations for a uniform distribution $X_m \sim \mathrm{U}\{1, 2m-1\}$ and a Poisson distribution $X_m \sim \mathrm{Pois}(m)$ both with expectation $\mathbb{E}(x_m) = m$.