Abstract
We propose a dynamic model of network formation with strategic interactions in growing population environments. We combine the network formation model by Jackson and Rogers [2007] with the linear-quadratic complementarity game by Ballester et al. [2006]. When complementarity effects are weak, this game has a unique Nash equilibrium that can be interpreted as a statistic of centrality in the underlying interaction network. We assume that the agents are boundedly rational so that they have only a noisy perception of their payoffs. Moreover, we consider two different scenarios in which agents have either global or local information of the network. We show that in the limit of vanishing noise both global and local information yield identical outcomes and agents are able to coordinate on the efficient network. However, in the limit of strong noise, local information leads to higher aggregate payoffs as networks being formed with global information. We thus find that restricting the information of agents when forming links can lead to better outcome networks.

Key words: network formation, preferential attachment, Bonacich centrality, bounded rationality, limited observability

JEL: C63, D83, D85, L22

1. Introduction
Understanding network structures is 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.\footnote{The recent books by Jackson [2008], Goyal [2007] and Vega-Redondo [2007] as well as De Martí and Zenou [2009] provide a very good introduction into this lively research area.} Despite the strong influence of networks in shaping many economic and social phenomena we are still missing a satisfactory understanding of the underlying mechanisms that form these networks. Existing approaches can be either classified as too “mechanistic”, in the sense that they propose some randomized algorithm governing the formation of links in a network without providing a behavioral foundation of why links are formed in that way. On the other hand, typical models of network formation in the economics and game theory literature can be regarded as too “simplistic” as they lack the ability to generate networks complex enough to match with characteristics found in real world networks [De Martí and Zenou, 2009].

The network formation model by Jackson and Rogers [2007] is a first attempt to bridge the above mentioned approaches and it is quite successful in reproducing the main stylized facts observed in real world networks. In this model the network is formed by new agents entering the existing network and linking to other agents in the population through a mixture of global and local (to the neighbors of already identified agents) linking protocols. However, the process governing the formation of links in Jackson and Rogers [2007] is largely mechanical as it essentially assumes that links are formed at random.

Based on their model a number of extensions and applications have been suggested. Ghiglino [2007] introduces an algorithm similar to Jackson and Rogers [2007] to study the creation and recombination of ideas from a pool of existing knowledge (cf. 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 Vigier [2009] a spatial extension is suggested in which the social network is embedded into geographical space and agents who are closer in space are more likely to form links.

Differently to these authors, we introduce a behavioral foundation of why links are formed in the model by Jackson and Rogers [2007] by introducing a utility function corresponding to the local complementarity game of Ballester et al. [2006]. We assume that agents are forming links on the basis of the marginal utility they derive from these links. More precisely, similar to König et al. [2009] we consider a repeated two-stage game, in which, in
the first stage, agents choose their optimal effort levels given the prevailing network while in the second stage, links are formed.\textsuperscript{3} We further depart from two common assumptions in the literature on strategic interaction in networks, namely (i) that agents have global information of the whole interaction network and (ii) that agents are perfectly rational. We relax the assumption of perfect rationality by introducing noise of varying intensity in the perception of the marginal payoffs agents obtain from forming links. Moreover, we relax the assumption of global information by studying two models of network formation varying in the amount of information agents possess at the date of their entry in the prevailing network. In the first, called the \textit{noisy growing network model with global information (NGNG)} entering agents possess global information of the prevailing network and their potential linking partners. In contrast, in the second model, which we call the \textit{noisy growing network model with local information (NGNL)}, entering agents observe only a randomly selected incumbent agent and his neighbors. The NGNL model is akin to the local search mechanism suggested by Jackson and Rogers \cite{Jackson:2007}. It takes into account the fact that individuals in social networks only imperfectly monitor other individual’s relationships \cite{Galeotti:2010,McBride:2006}, usually encompassing their direct neighbors and their neighbors’ neighbors \cite{Friedkin:1983}.\textsuperscript{4} The NGNG model is amenable to fairly detailed analysis since we can rely on rigorous work by Bollobás et al. \cite{Bollobas:2001} and Oliveira and Spencer \cite{Oliveira:2005}. We provide a complete picture of the behavior of this process for every level of noise. In the high noise scenario, a good approximation of typical realizations of the process is given by a linear preferential attachment process as studied by Krapivsky and Redner \cite{Krapivsky:2003}. Here we find that the degree distribution of typical networks has asymptotically a power-law tail with scaling exponent larger than three. In the extreme case of zero noise, the information that entering agents receive is so noisy that their link creation behavior will appear as a uniform draw. Consequently, in this scenario the degree distribution is a geometric distribution. For low noise the link creation pattern of agents will be geared towards linking with high degree nodes, i.e. the tendency towards preferential attachment \cite{Barabasi:1999} will become stronger. As shown by Oliveira and Spencer \cite{Oliveira:2005} typical

\textsuperscript{3}König et al. \cite{Koenig:2009} consider a network with a fixed size while we model a growing network. Besides, as we will discuss later, in König et al. \cite{Koenig:2009} agents have full information of the network and always choose the link that maximizes their payoffs, while we restrict the amount of information available to the agents and allow for mistakes in their linking decisions.

\textsuperscript{4}In a recent study by Ballester et al. \cite{Ballester:2009} it has been confirmed experimentally that individuals playing the network game by Ballester et al. \cite{Ballester:2006} take only into account their neighborhood up to a distance of two, that is, they consider the actions of their neighbors and their neighbors’ neighbors.
outcomes of the network formation process will have a star architecture.

The second model we propose, the NGNL model, differs from the NGNG algorithm by making the information entering agents possess local. As in Jackson and Rogers [2007] the working of the process can be disentangled into two parts. First, an entering agent who has no ex ante information of the prevailing network observes an incumbent agent uniformly at random. The information neighborhood of the entering agent is determined by the randomly sampled agent and his neighbors. Having identified the information neighborhood of the agent a process of local link formation takes place. Similar to the NGNG model, the probability with which a link is formed depends on the marginal utility the agent derives from it and the level of noise with which the agent perceives his payoff. For the NGNL model we can characterize the emerging networks in the small and large noise limits. We further compare the outcome networks in the NGNG and the NGNL models. We find that in the vanishing noise limit both setups lead to the star network. However, in the presence of significant noise, NGNL generates networks with more degree heterogeneity than NGNG. These observations have implications for aggregate payoff and efficiency. For vanishing noise, both NGNG and NGNL produce efficient networks. For low levels of noise networks generated by NGNG have a higher aggregate payoff than those by NGNL. In contrast, for high levels of noise NGNL produces higher aggregate payoffs than NGNL. Remarkably, we therefore find that when agents have a noisy perception of their payoffs then restricting the amount of information that is available to them when deciding with whom to form a link leads to better outcome networks.

The paper is organized as follows. In Section 2 we introduce the payoff agents derive from the network and define efficiency. Next, in Section 3 we describe the evolution of the network and introduce two link formation processes differing in the information agents have when forming links. The appendix contains the detailed mathematical description of these models. In Section 4 we compare both models of network formation and identify the condition under which they yield identical outcome networks. We further show for which level of noise one has higher aggregate payoffs than the other. Finally, in Section 5 we conclude.

2. The Linear-Quadratic Complementarity Game

Ballester et al. [2006] consider the following interaction game. There is a finite population of players, contained in the set \( V := \{v_1, \ldots, v_n\} \), who interact on an undirected graph \( G = (V, E) \) to play the normal form game \( \Gamma(\alpha, G) := (V, (\pi_i(\cdot; \alpha, G), i \in V)) \), with payoff function for player \( i \in V \)

\[
\pi_i(x; \alpha, G) := x_i - \frac{1}{2}x_i^2 + \alpha \sum_{j \in N_i(G)} x_i x_j. \tag{1}
\]
The action profile \( x := (x_i)_{i \in V} \in \mathbb{R}^{|V|} \) collects the effort levels of the players. \( N_i(G) \) is the set of neighboring agents of player \( i \) in the graph \( G \), and the parameter \( \alpha \) measures the influence of the actions of neighboring players on the payoff of player \( i \). We consider only the case where effort levels are strategic complements, i.e. where \( \alpha > 0 \).\(^5\) Introducing the adjacency matrix \( A(G) \equiv A \) of the network \( G \), whose elements are defined as

\[
a_{ij}(G) := \begin{cases} 1 & \text{if } (i, j) \in E(G), \\ 0 & \text{otherwise}, \end{cases}
\]

allows one to write the payoff function as

\[
\pi_i(x; \alpha, G) = x_i - \frac{1}{2} x_i^2 + \alpha x_i (A(G)x)_i.
\]

(2)

The following measure of network centrality is of high importance for the structure of Nash equilibria in this game.\(^6\)

**Definition 1.** Given an undirected graph \( G \) of size \( n \) with adjacency matrix \( A \). Let \( \lambda_{\text{PF}}(G) \) denote the largest modulus eigenvalue of the matrix \( A \). For \( \alpha < 1/\lambda_{\text{PF}}(G) \) the matrix

\[
B(G, \alpha) := \sum_{k=0}^{\infty} \alpha^k A^k = (I - \alpha A)^{-1}
\]

exists and is non-negative. Let \( u \) denote the \( n \times 1 \) vector where each entry equals to 1. The vector

\[
b(G, \alpha) := B(G, \alpha) \cdot u
\]

is called the Bonacich centrality vector [Bonacich, 1987].

We can write the Bonacich centrality vector as

\[
b(G, \alpha) = \sum_{k=0}^{\infty} \alpha^k A^k \cdot u = (I - \alpha A)^{-1} \cdot u.
\]

(3)

The interpretation of the components \( b_i(G, \alpha), i \in V, n = |V| \), can be seen by writing

\[
b_i(G, \alpha) = \sum_{k=0}^{\infty} \alpha^k (A^k \cdot u)_i = \sum_{k=0}^{\infty} \alpha^k \sum_{j=1}^{n} a_{ij}^{[k]},
\]

(4)

\(^5\)The case where effort levels are strategic substitutes would make no sense in the network formation algorithms proposed below.

\(^6\)The definition uses basic properties of non-negative matrices. See for instance Horn and Johnson [1985].
where \( a_{ij}^{[k]} \) is the \( ij \)-th entry of \( A^k \). Because \( \sum_{j=1}^n a_{ij}^{[k]} \) is the number of all walks of length \( k \) in \( G \) starting from \( i \), \( b_i(G, \alpha) \) is the number of all walks in \( G \) starting from \( i \), where walks of length \( k \) are geometrically weighted by a factor \( \alpha^k \). Ballester et al. [2006] show that Bonacich centrality is closely related to Nash equilibrium in the game \( \Gamma \) by proving the following result.

**Theorem 1** (Ballester et al. [2006]). Given a network \( G \) of size \( n \) with adjacency matrix \( A \). For \( \alpha < 1/\lambda_{PF}(G) \), the unique interior Nash equilibrium of the normal form game \( \Gamma(\alpha, G) \) is given by

\[
(\forall i \in V) : x_i^*(G, \alpha) := b_i(G, \alpha). \tag{5}
\]

A proof of this theorem, stated for completeness only, can be given by recognizing \( \Gamma(\alpha, G) \) as an exact potential game [Monderer and Shapley, 1996] with potential function

\[
P(x; \alpha, G) := x^\top u - \frac{1}{2} x^\top (I - \alpha A)x \tag{6}
\]

for all \( x \in \mathbb{R}^{|V|} \). Fix a pair \( (\alpha, G) \) such that \( \alpha < 1/\lambda_{PF}(G) \). It is well known [see e.g. Monderer and Shapley, 1996] that the set of solutions of the program

\[
\max_{x \in \mathbb{R}^{|V|}} P(x; \alpha, G)
\]

forms a subset of the set of Nash equilibria of the game \( \Gamma(\alpha, G) \). This program has a unique interior solution if the potential function is strictly concave on the relevant domain. The Hessian matrix of \( P \) is easily computed to be \(- (I - \alpha A)\). The matrix \( I - \alpha A \) is positive definite if for all non-zero \( x \) we have

\[
x^\top (I - \alpha A)x > 0 \iff \alpha < \left( \frac{x^\top Ax}{x^\top x} \right)^{-1}.
\]

By the Rayleigh-Ritz theorem [see e.g. Horn and Johnson, 1985, p. 177], we have \( \lambda_{PF}(G) = \sup_{x \neq 0} \frac{x^\top Ax}{x^\top x} \). Thus a necessary and sufficient condition for having a strict concave potential is that \( \alpha < 1/\lambda_{PF}(G) \), as stated in the theorem.

If there exists an interior solution all agents choose effort levels given by their Bonacich centrality in the network \( G \). Equilibrium payoffs are

\[
\pi_i(x^*(G, \alpha); \alpha, G) = \frac{1}{2} b_i(G, \alpha)^2. \tag{7}
\]

From this expression it is clear that aggregate payoffs in the Nash equilibrium (which we take as a welfare measure) depend crucially on the underlying network structure. In equilibrium the total sum of utilities of the players
can be written as

$$
\Pi(G, \alpha) = \sum_{i=1}^{n} \pi_i^*(G, \alpha) = \frac{1}{2} \mathbf{u}^\top \mathbf{B}(G, \alpha)^\top \mathbf{B}(G, \alpha) \mathbf{u}.
$$

(8)

A graph that maximizes aggregate equilibrium payoffs is called an efficient graph. For this paper the class of tree graphs is of particular importance.\footnote{A connected graph on \( n \) vertices is a tree if and only if it has \( n - 1 \) edges.}

For a fixed number of players \( n \) let \( T_n \) denote the class of tree networks.

In the case of low complementarity effects, corresponding to small values of \( \alpha \), we see that Bonacich centrality of a player can be written as

$$
b_i(G, \alpha) = 1 + \alpha d_i + \alpha^2 \sum_{j \in N_i} d_j + O(\alpha^3).
$$

(9)

Hence, aggregate payoffs from Equation (8) can be written as

$$
\Pi(G, \alpha) = \frac{1}{2} \sum_{i=1}^{n} b_i(G, \alpha)^2 = \frac{n}{2} + \alpha \sum_{i=1}^{n} d_i + \frac{\alpha^2}{2} \sum_{i=1}^{n} d_i^2 + \alpha^2 \sum_{j \in N_i} \sum_{i=1}^{n} d_j + O(\alpha^3)
$$

$$
= \frac{n}{2} + \alpha \sum_{i=1}^{n} d_i + \frac{3\alpha^2}{2} \sum_{i=1}^{n} d_i^2 + O(\alpha^3).
$$

where we have used the fact that \( \sum_{i=1}^{n} \sum_{j \in N_i} d_j = \sum_{i=1}^{n} d_i^2 \). If \( G \in T_n \) then the number of edges is \( e(G) = n - 1 \), and by means of \( \sum_{i=1}^{n} d_i = 2e(G) = 2(n - 1) \) we obtain

$$
\Pi(G, \alpha) = \frac{n}{2} + 2\alpha(n - 1) + \frac{3\alpha^2}{2} \sum_{i=1}^{n} d_i^2 + O(\alpha^3).
$$

(10)

If complementarity effects are weak, we can state the following proposition.

**Proposition 1.** Consider the linear-quadratic game \( \Gamma \) with interactions restricted to tree networks \( G \in T_n \) with \( n > 7 \). Assume that \( \alpha \) is small such that terms \( O(\alpha^3) \) can be neglected in Equation (10). Then aggregate equilibrium payoffs \( \Pi(G, \alpha) \) are maximized for the star \( K_{1,n-1} \).

**Proof of Proposition 1.** From Equation (10) we see that maximizing aggregate payoffs in the class of tree graphs \( G \in T_n \) for small values of \( \alpha \) (neglecting terms of order \( O(\alpha^3) \)) is equivalent to maximizing the sum of squares of degrees. The graph which maximizes \( \sum_{i=1}^{n} d_i^2 \) is the quasistar \( S_n^\alpha \) if the number of edges satisfies \( e < \frac{1}{2}\binom{n}{2} - \frac{\alpha^2}{2} \) \cite{bell1992}. For \( G \in T_n \) we have that \( e(G) = n - 1 \), the quasistar \( S_n^\alpha \) is equivalent to the star network \( K_{1,n-1} \), and the above inequality is satisfied for \( n > 7 \). \( \square \)
Proposition 1 shows that highly asymmetric networks are efficient in terms of maximizing aggregate equilibrium payoffs.\footnote{When complementarity effects are weak, such that we can neglect terms of the order $O(\alpha^3)$, and we do not require $G$ to be a tree then it can be shown that the efficient network is a \textit{nested split graph} \cite[for a definition and application of this type of networks]{König et al., 2009}.} Maximal aggregate payoffs can be computed as follows; Note that, by symmetry of the matrix $A$

$$B(G, \alpha)^\top B(G, \alpha) = \left(\sum_{k=0}^{\infty} \alpha^k A^k \right)^2.$$ 

We diagonalize $A$ as $A = SDS^\top$, where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$ are the eigenvalues of $A$ and the matrix $S$ has columns given by the orthonormal eigenvectors $v_i$ corresponding to eigenvalues $\lambda_i$. If $G$ is connected then $A$ is irreducible and $\lambda_{PF}$, the Perron-Frobenius eigenvalue, is unique with a strictly positive eigenvector. We can then write

$$B(G, \alpha)^\top B(G, \alpha) = \left(\sum_{k=0}^{\infty} \alpha^k SD^k S^\top \right)^2.$$ 

This yields

$$\Pi(G, \alpha) = \frac{1}{2} u^\top B(G, \alpha)^\top B(G, \alpha) u = \frac{1}{2} \left(\sum_{k=0}^{\infty} \alpha^k \lambda_1^k \left(1 + \sum_{j=2}^{n} \left(\frac{\lambda_j}{\lambda_1}\right)^k\right)\right)^2.$$ 

For the star $K_{1,n-1} \in T_n$ we have that $\lambda_1 = \sqrt{n-1}, \lambda_2 = \lambda_3 = \cdots = \lambda_{n-1} = 0$ and $\lambda_n = -\sqrt{n-1}$. Hence, we get

$$\Pi(K_{1,n-1}, \alpha) = \frac{1}{2} \left(\sum_{k=0}^{\infty} \left(\alpha \sqrt{n-1}\right)^k\right)^2 = \frac{1}{2} \left(\frac{1}{1 - \alpha \sqrt{n-1}}\right)^2. \quad (11)$$

From the above equation we see that $\lim_{\alpha \to 1/\sqrt{n-1}} \Pi(K_{1,n-1}, \alpha) = \infty$. Finally, observe that for both cases, $\alpha$ small and large, we have that $\Pi(G, \alpha) \leq \Pi(K_{1,n-1}, \alpha),\footnote{The case of higher values of $\alpha$ has been analyzed in Corbo et al. [2006].}$ where the upper bound is given in Equation (11), for all
The case of $\alpha$ small is summarized in the following corollary.

**Corollary 1.** Consider the linear-quadratic game $\Gamma$ with interactions restricted to tree networks $G \in T_n$ with $n > 7$. Let $\alpha$ be small enough such that terms $O(\alpha^3)$ can be neglected in Equation (10). Then aggregate equilibrium payoffs are bounded from above by

$$\Pi(G, \alpha) \leq \frac{1}{2} \left( \frac{1}{1 - \alpha \sqrt{n} - 1} \right)^2.$$ 

**Proof of Corollary 1.** The corollary is a direct consequence of Equation (11) and Proposition 1. \qed

### 3. The Evolutionary Process

We now embed the static model of Ballester et al. [2006] in a model of network evolution. We envision a countable infinite population of players $I = \mathbb{N}$, which consists of the set of *active* players at time $t \in \mathbb{N}$, denoted as $V(t)$, and *passive* players $I \setminus V(t)$. Interaction takes place only among active players, and the pattern of interaction is represented as a graph $G(t) := (V(t), E(t))$. The general structure of the network formation algorithms we consider are non-stationary Markov chains in discrete time. Each $G(t)$ is a random variable adapted to the filtration $\mathcal{F}_t = \sigma(G(t_0), G(t_0 + 1), \ldots, G(t))$.\footnote{Our algorithms allow for creating multiple edges but no self-loops. Hence, technically $G(t)$ is a random multigraph and we count edges with their multiplicities. The results obtained in this paper are unaffected by this assumption, but they make the mathematical statements much simpler.} The probability measure $\mathbb{P}(.|\mathcal{F}_t)$ is denoted as $\mathbb{P}_t$. Expected values with respect to $\mathbb{P}_t$ are similarly denoted by $\mathbb{E}[.|\mathcal{F}_t]$. Agents are labeled by their date of birth, so that $v_t$ is the label of the agent born at time $t$ of the process. Hence, at each point of time there are $n(t) := |V(t)| = t_0 + t$ agents in the population, and the number of edges in the graph $G(t)$ is $|E(t)| = e(t) = 2m(t - t_0)$. The degree of an agent $i \in V(t)$ is denoted as $d_i(t)$, and by an abuse of notation we will use this notation for the random variable as well as its realization. By our choice of labeling agents we have $d_i(t) = 0$ for $i \leq t$. We are interested in a model of growing networks and we want that at each step of the network formation algorithm there is a unique Nash equilibrium in the interaction game. This requires that $\alpha < 1/\lambda_{PF}(G(t))$ in all periods $t$ in which the network is growing. The network formation process is then defined as follows.

**Initialization:** Start from the network $G(t_0) = (V(t_0), \emptyset)$ and fix positive
integers $T$ and $m$.\textsuperscript{11} The time window of the process is set equal to the time interval $[t_0, T]$, where $t_0 = m$ and the initial set of active players is $V(t_0) = \{v_1, \ldots, v_{t_0}\}$.

For $t \in [t_0, T - 1]$ the dynamic process consists of two parts:

**Equilibration:** All players get to know their neighborhood under the graph $G(t)$. Given that graph the actions of the players (active and passive) are in equilibrium, i.e.

$$\left(\forall i \in \mathbb{N}\right): x_i^*(G(t), \alpha) = \begin{cases} 1 & \text{if } i \in I \setminus V(t), \\ b_i(G(t), \alpha) & \text{if } i \in V(t). \end{cases} \quad (12)$$

**Network formation:** The network $G(t + 1)$ is obtained from $G(t)$, given the equilibrium actions $x_i^*(G(t), \alpha)$, as follows: The passive player with label $v_{t+1}$ becomes active and connects to maximally $m$ agents in his information neighborhood $I(t + 1) \subseteq V(t)$, according to a probability distribution $K_t$ depending on the marginal payoff agent $v_{t+1}$ receives from linking to an active player $i \in I(t + 1)$. The information neighborhood is a subset of active agents which are observed by the entrant. Assuming that the strategies of all players $i \in I$ are fixed at the time of entry, the marginal payoff is exactly given by\textsuperscript{12,13}

$$\Delta \pi_{v_{t+1}}(i, \alpha) \equiv \pi_{v_{t+1}}(G(t) + v_{t+1}, \alpha) - \pi_{v_{t+1}}(G(t), \alpha) = \alpha b_i(G(t), \alpha). \quad (13)$$

Once agent $v_{t+1}$ has executed the network formation steps, the equilibration phase sets in.

At $t = T$ the process stops.

In the equilibration step we implicitly assume that agents behavior settles down to the (unique) Nash equilibrium on a temporarily fixed network. This requires that players instantaneously react on the change in the network resulting from the network formation step. The intuition for this is that behavior adapts at much faster rate than the network grows.\textsuperscript{14}

\textsuperscript{11}Alternatively we could choose as initial condition a regular graph. These two choices eliminate dependency of evolution on the initial condition of the network.

\textsuperscript{12}For alternative interpretations see Appendices B and C.

\textsuperscript{13}It is shown in König et al.[2009] that we could introduce a fixed linking cost $c > 0$ and the creation of an edge would always be beneficial as long as $c < \alpha/(1 - \alpha)$. Hence, even in the presence of fixed linking costs, our results would not change as long as the cost is below this bound.

\textsuperscript{14}Conceptually, our evolutionary process has the same structure as models studying the evolution of preferences. These papers also assume that the evolution of actions and preferences proceeds at two distinct time scales, where actions equilibrate at an arbitrary
Note that from Equation (9) it follows that the Bonacich centrality of player $i$ in the marginal payoff of Equation (13) can be written as $b_i(G(t), \alpha) = 1 + \alpha d_i(t) + O(\alpha^2)$. Hence, when complementarity effects are sufficiently small we can make the following assumption.\footnote{Similarly, Bloch and Quéro \cite{2008} have analyzed small complementarity effects in the static network model of Ballester et al. \cite{2006}.}

**Assumption 1.** Given a predefined stopping time $T \geq t_0$, we assume that the degree of strategic complementarity $\alpha$ is small enough such that for all $t \in [t_0, T)$ and $i \in V(t)$ in the marginal utility of a link between passive agent $v_{t+1}$ and active agent $i \in V(t)$,

$$
\Delta \pi_{v_{t+1}}(i, \alpha) = \alpha + \alpha^2 d_i(t) + O(\alpha^3).
$$

(14)

Hence, terms of order $O(\alpha^3)$ can be neglected.

This assumption is a structural assumption on the preferences of the players, implying essentially that interactions with close neighbors weigh much more than interactions with agents at larger (geodesic) distance in the graph.\footnote{For small values of $\alpha$ Bonacich centrality becomes proportional to degree centrality. This situation also arises in the model by Jackson and Wolinsky \cite{1996} with a small decay parameter $\delta > 0$, where the utility of an agent $i$ in the network $G$ is given by his decay centrality $\sum_{j \neq i} s^{(i,j)}$ and $l(i,j)$ being the number of edges on the shortest path between $i$ and $j$. For small values of $\delta$, decay centrality becomes proportional to degree centrality and in this case we could alternatively use decay centrality as our utility function yielding identical results.} Neglecting terms of the order $O(\alpha^3)$ yields a marginal payoff of $\Delta \pi_{v_{t+1}}(i, \alpha) \approx \alpha + \alpha^2 d_i(t)$. Applying an affine transformation by subtracting $\alpha$ and dividing by $\alpha^2$ yields a marginal payoff of $\Delta \tilde{\pi}_{v_{t+1}}(i, \alpha) = d_i(t)$, which is entirely determined by the degree of the target agent $i$ agent $v_{t+1}$ observes. We will thus assume that the marginal payoff for agent $v_{t+1}$ from forming a link to agent $i$ is given by $\Delta \tilde{\pi}_{v_{t+1}}(i, \alpha)$ and therefore determined by the degree $d_i(t)$ of agent $i$.

### 3.1. Noisy Growing Network Formation with Global Information

In the global information scenario, the information neighborhood of an entering agent is given by the full set of incumbent agents, that is $I(t+1) = V(t)$. Under Assumption 1 the marginal payoff of a link formed by an entering agent is determined by the degree of the incumbent agents. However, the agents perceive the marginal payoff with noise. We thus assume that the
information of the entering agent is the perturbed statistic
\[ \theta(t) := (\theta_i(t); i \in V(t)) = (d_1(t) + \epsilon_1, \ldots, d_n(t) + \epsilon_n(t)), \]
where \( \epsilon_i \) is an i.i.d. random error term. Agents prefer to link with other agents who have a large degree. Hence, agent \( v_{t+1} \) selects a player with a probability which is increasing in \( \theta_i(t) \).

**Definition 2 (The NGNG algorithm).** Given the initial graph \( G(t_0) = (V(t_0), \emptyset) \). For all \( t \geq t_0 \) the graph \( G(t + 1) \) is obtained from \( G(t) \) by the following steps:

**Growth:** \( V(t + 1) = V(t) \cup \{v_{t+1}\} \) deterministically.

**Link creation:** Agent \( v_{t+1} \) creates \( m \) edges simultaneously. The marginal probability that agent \( i \in V(t) \) is sampled is given by the attachment kernel
\[ K_t(i) := P_t \left( d_i(t) + \epsilon_i = \max_{j \in V(t)} (d_j(t) + \epsilon_j) \right). \]

The random vector \( X^{(m)}(t + 1) \) of length \( m \) is the outcome of this sampling procedure.

The NGNG algorithm generates a Markov chain with transition probabilities
\[ P_t(G_m(t + 1) = G) = \begin{cases} K_t(i) & \text{if } G = G(t) + iv_{t+1}, i \in V(t), \\ 0 & \text{otherwise}. \end{cases} \]

For analytical results it is convenient to make a distributional assumption on the nature of perturbations. A particularly nice form for the choice probabilities \( K_t(\cdot) \) is obtained when the error terms follow a double exponential distribution with mean 0 and variance \( \beta^{-2} \pi^2 / 6 \), where \( 1/\beta \in (0, \infty) \) is the scale parameter of the distribution. Then the function \( K_t(\cdot) \) has the form
\[ K_t^\beta(i) = \frac{\exp(\beta d_i(t))}{\sum_{j \in V(t)} \exp(\beta d_j(t))} = \frac{f(\beta, d_i(t))}{Z(\beta, t)} \] (15)
for all \( i \in V(t) \), where \( f(\beta, x) := \exp(\beta x) \) and \( Z(\beta, t) = \sum_{j \in V(t)} f(\beta, d_j(t)) \).

This choice function has the great advantage that there exists a single parameter \( \beta \) with which one can model many types of choice behavior. For

---

17 This is the log-linear model in the theory of discrete choice. An alternative standard formulation in the discrete choice literature is a power choice function, where a choice is made with probability proportional to \( \theta_i(t)^\beta, \beta > 0 \). Our results for the NGNG algorithm hold also when we assume this choice function (or a general class of parameterized choice functions which satisfies certain limit conditions).
\( \beta \approx 0 \) the choice rule can be well approximated by the linear choice function

\[
K_i^\beta(t) \propto 1 + \beta d_i(t)
\]

which is the linear attachment rule of Krapivsky and Redner [2003]. For moderate \( \beta \), Equation (15) cannot be simplified and we would have a model akin to the super-linear preferential attachment model of Krapivsky and Redner [2003], which has been rigorously studied by Oliveira and Spencer [2005]. As \( \beta \to \infty \) choices of the players become more concentrated around the maximal degree of the prevailing degree sequence. Figure 1 provides an illustration of possible networks generated under the NGNG algorithm for various values of \( \beta \). Our main interest lies in the structure of resulting networks formed by the NGNG-algorithm. Much, though not all, information
about the characteristic of a network is contained in its degree distribution. Here we will derive this statistic for small noise $\beta$. This will give us a good description of the networks generated by NGNG over finite time periods when the noise parameter is sufficiently small. The general case is studied separately in Section 4. Denote by $N_k(t)$ the number of agents at time $t$ with degree equal to $k$. $P_k(t) = \frac{1}{n(t)} N_k(t)$ is the associated empirical frequency of agents with degree $k$ at time $t$. The sequence $P(t) = \{P_k(t)\}_{k \geq 1}$ is called the empirical degree distribution generated by NGNG at time $t$. We denote by $\bar{P}(t) := \mathbb{E}[P(t)]$ the expected degree sequence. For $\beta$ sufficiently small we may approximate the preference function linearly as $f(\beta, k) = 1 + \beta k + O(\beta^2)$. Under this first-order approximation it follows that

$$Z(\beta, t) = \sum_k N_k(t) f(\beta, k)$$
$$= n(t) + \beta \sum_k k N_k(t) + O(\beta^2)$$
$$= n(t) + \beta e(t) + O(\beta^2)$$
$$= n(t) + 2\beta m(t - t_0) + O(\beta^2)$$

Let $\zeta(\beta, t) := \frac{Z(\beta, t)}{t}$. It follows that $\lim_{t \to \infty} \zeta(\beta, t) = 1 + 2\beta m \equiv \zeta^*(\beta)$. Theorem 2. Fix $m \geq 1, \epsilon > 0$ small and $\beta \in [0, \epsilon)$. Consider the degree distribution $p^\beta := \{p^\beta_k\}_{k \geq 1}$ defined by

$$p^\beta_k = \frac{(1 + 2m\beta)\Gamma \left[ k + \frac{1}{\beta} \right] \Gamma \left[ 2 + m + \frac{1}{\beta} + \frac{1}{m\beta} \right]}{m\beta \Gamma \left[ m + \frac{1}{\beta} \right] \Gamma \left[ 3 + k + \frac{1}{\beta} + \frac{1}{m\beta} \right]}$$

for all $k \geq m$. Then, there exists a constant $C = C(m, \beta)$ such that, as $t \to \infty$,

$$\mathbb{P} \left( \max_k \left| P_k(t) - p^\beta_k \right| > C \frac{\sqrt{t \log t}}{n(t)} \right) = o(1).$$

This theorem shows that in the large noise limit NGNG generates networks with a power-law degree distribution with coefficient $\gamma = 3 + \frac{1}{m\beta}$. The proof of the Theorem proceeds in essentially two steps, which we formulate as the following two propositions, whose proof can be found in Appendix D.1. Proposition 2 below identifies the shape of the expected degree sequence in the limit of large $t$ and shows that the sequence $P(t)$ converges as $t \to \infty$ to the limiting degree distribution $p^\beta := \{p^\beta_k\}_{k \geq 1}$.

**Proposition 2.** Fix $m \geq 1, \epsilon > 0$ small and $\beta \in [0, \epsilon)$. Then, for all $k \geq m$ we have

$$\lim_{t \to \infty} P_k(t) = p^\beta_k.$$
Proposition 3 shows that the empirical degree distribution concentrates around its mean for $t$ sufficiently large. This completes the proof of Theorem 2.

**Proposition 3.** Fix $\epsilon > 0$ small and $\beta \in [0, \epsilon), m \geq 1$. Then, for any $C > \sqrt{8m^2}$, with probability $1 - o(1)$ we have

$$|P_k(t) - \bar{P}_k(t)| \leq C \frac{\sqrt{t \log t}}{n(t)}$$

for $t$ sufficiently large.

**Remark 1.** We would like to emphasize that the linear preference function $f(\cdot, \cdot)$ arises only for small $\beta$. This is however only a consequence of our assumption that players sample optimally given their noisy data $\theta(t)$ and the noise is double exponentially distributed. This interpretation is, however, mathematically unimportant. Instead we could study the NGNG algorithm by simply assuming that entering nodes link with the incumbent with probability proportional to the affine preference function $f = 1 + \beta x$. This would allow us to treat $\beta$ as a flexible parameter. If we would adopt this alternative formulation Proposition 2 tells us that the NGNG algorithm is able to generate scale-free networks with power-law coefficient $3 < \gamma < \infty$.

The expression for the degree distribution can be simplified when we focus on large degrees. Using Stirling’s formula, we we get (for large $k$) the approximation (see Appendix D.1 for the details)

$$p^\beta_k = (1 + \beta k)^{-\left(3 + \frac{1}{m}\right)} \left(1 + O\left(\frac{1}{k}\right)\right). \quad (17)$$

This power-law behavior is also visible in numerical simulations as illustrated in Figure 2. If we would start the NGNG process with “infinite noise”, corresponding to the situation $\beta = 0$, we obtain a process of uniform attachment. Hence, we obtain the rigorous results of Bollobás et al. [2001] as the following corollary.

**Corollary 2.** In the NGNG algorithm with $\beta = 0$ the agents perform a uniform attachment process whose degree distribution is given by

$$p^0_k = \frac{1}{1 + m} \left(\frac{m}{1 + m}\right)^{k-m}, \quad (18)$$

a geometric distribution with parameter $\frac{1}{1+m}$.

A numerical example for $n = 10000$ of this geometric degree distribution is provided in Figure 2.
3.2. Noisy Growing Network Formation with Local Information

We now study a variation of the NGNG model where entering agents have only local information in the following sense.\footnote{As in the model of \textcite{Jackson2007} the NGNL model combines a process of “random meetings” and “network-based meetings”. The rational for the NGNL model is the fact that there exist numerous empirical evidences that individuals in real world networks have only restricted information about the network, usually encompassing their direct contacts and their contacts’ contacts. See e.g. \textcite{Friedkin1983} and \textcite{Ballester2009}.}

**Definition 3** (The NGNL algorithm). \textit{Given the initial graph }$G(t_0) = (V(t_0), \emptyset)$. \textit{For all }$t \geq t_0$ \textit{the graph }$G(t+1)$ \textit{is obtained from }$G(t)$ \textit{by the following steps:}

\textbf{Growth:} $V(t+1) = V(t) \cup \{v_{t+1}\}$ \textit{deterministically.}

\textbf{Information neighborhood:} Agent $v_{t+1}$ \textit{samples an agent }$i \in V(t)$ \textit{uniformly at random and observes the local degree sequence }$\tilde{\theta}(t) = (d_j(t) + \epsilon_j; j \in N_i(t) \cup \{i\})$. \textit{The set }$I(t+1) = N_i(t+1) \cup \{i\}$ \textit{forms the information neighborhood of }$v_{t+1}$.

\textbf{Link creation:} Agent $v_{t+1}$ \textit{creates }$m_{t+1}^* := \min\{m, d_i(t) + 1\}$ \textit{edges simultaneously}. \textit{The probability that agent }$j \in V(t)$ \textit{is sampled is given by}

\[ K_t(j|I(t+1)) := \mathbb{P}_t \left( d_i(t) + \epsilon_i = \max_{j \in I(t+1)} (d_j(t) + \epsilon_j) \right) \mathbb{1}_{I(t+1)}(j). \]

The random vector $X^{(m_{t+1}^*)}(t+1)$ \textit{of length }$m_{t+1}^*$ \textit{is the outcome of this sampling procedure.}
In order to compare the performance of NGNG with NGNL we make the same distributional assumption on the nature of the random perturbations. However, under the NGNL algorithm, the attachment kernel, i.e. the total probability that an incumbent vertex $i \in V(t)$ is found by the entering agent, is a slightly more complicated object. It is given by

$$\tilde{K}_t^\beta(i) = \frac{1}{n(t)} \sum_{j \in \bar{N}_i(t)} \frac{f(\beta, d_i(t))}{Z_j(\beta, t)}$$

(19)

where $f(\beta, x) = \exp(\beta x)$, $\bar{N}_i(t) := N_i(t) \cup \{i\}$, and

$$Z_j(\beta, t) := \sum_{v \in \bar{N}_j(t)} f(\beta, d_v(t))$$

(20)

The sum in the attachment probability $\tilde{K}_t^\beta(i)$ can be decomposed into two parts. The first part for $j = i$ corresponds to the case that node $i$ is observed directly by the entering node $v_{t+1}$. This happens with probability $1/n(t)$. Given that $i$ has been observed directly by $t$, the probability that the edge $(v_{t+1}, i)$ is created is given by $f(\beta, d_i(t))/Z_i(\beta, t)$, where the partition sum $Z_i(\beta, t)$ in Equation (20) depends on the degrees of all nodes in the neighborhood of $i$ (see Figure 3, left). The second part in Equation (19) for $j \in N_i(t)$ corresponds to the case that node $i$ is observed indirectly by $v_{t+1}$, as a neighbor of an already observed node $j$. There are $|N_i(t)|$ such nodes whose neighborhood contains $i$ and for each the probability that the link $(v_{t+1}, i) \notin E(t)$ is created is given by $f(\beta, d_i(t))/Z_j(\beta, t)$, where the par-

---

$^{19}$As for the NGNG algorithm we believe that our results are not dependent on the specific functional form we assume. This remains, however, a conjecture.
tition sum $Z_j(\beta, t)$ depends on the degrees of all nodes in the neighborhood of $j$ (see Figure 3, right). Finally, the factor $1/n(t)$ stems from the fact that the probability that a node is observed among the $n(t) = |V(0)| + t$ nodes present at time $t$ is $1/n(t)$.

NGNL defines a Markov chain with transition probabilities

$$
P_t(G_m(t+1) = G) = \begin{cases} \tilde{K}_t^\beta(i) & \text{if } G = G(t) + iv_{t+1}, i \in V(t) \\ 0 & \text{otherwise.} \end{cases}$$

Figure 4 provides a snapshot of possible networks generated under the NGNL algorithm. In the following we will analyze the degree distribution arising under the NGNL model in the large noise case (when $\beta$ becomes small).

Since the transition probabilities describing the random process are much more complicated as in the NGNG model we rely on methods from stochastic approximation theory.

20 Assume that $t_i \geq 0$ is the arrival time of node $i$ and let $d_i(t), t \geq t_i$ be the degree of a node $i \in V(t)$, born at time $t_i \geq 1$. Taking into account that an entering node can form up to $m$ links, the expected increase of $d_i(t)$ is given by

$$
E(d_{i}(t+1) - d_{i}(t)|F_t) = m \tilde{K}_t^\beta(i).
$$

Note that $m \tilde{K}_t^\beta(i)$ is bounded and locally Lipschitz continuous for all $i \in V(t)$ (the expected change in the degree of any node is at most $m$ in one time period). This allows us to introduce the continuous time counterpart of Equation (21), the so called mean-field equation [Benaim and Weibull, 2003a,b], governing the evolution of the expected degree $d_i(t), t \in \mathbb{R}_+, t \geq t_i$, in the limit of large $t$ (respectively $n(t)$)

$$
\frac{dd_i(t)}{dt} = m \tilde{K}_t^\beta(i) = \frac{m f(\beta, d_{i}(t))}{t Z_i(\beta, t)} + \frac{m}{t} \sum_{j \in N_i} \frac{f(\beta, d_{j}(t))}{Z_j(\beta, t)}.
$$

In order to solve Equation (22) we assume that for $\beta$ small enough there are only weak degree-degree correlations in $G(t)$, so that we can write for the partition sum

$$
Z_i(\beta, t) = \sum_{j \in (N_i \cup \{i\})} e^{\beta d_{j}(t)}
= 1 + \beta d_{i}(t) + \sum_{j \in N_i} (1 + \beta d_{j}(t)) + O(\beta^2)
\approx 1 + \beta d_{i}(t) + d_{i}(t) (1 + \beta \bar{d}(t)) + O(\beta^2),
$$

20In fact the analysis of Jackson and Rogers [2007] is based on such methods.
K have higher centrality. The labels of the nodes indicate the date of entry of the agents.

Figure 4: Examples of networks in the NGNL model for $T = 100$, $m = 1$ and values of $\beta = 0, 0.01, 0.25, 0.5, 0.75, 1$, from top left to bottom right. Nodes with brighter shapes have higher centrality. The labels of the nodes indicate the date of entry of the agents. With increasing values of $\beta$, networks become more and more centralized. For $\beta = 1$ we obtain a star $K_{1,n-1}$. Observe that already for $\beta = 0.5$ a star is generated in the NGNG model (see Figure 1) while this does not hold for the NGNL model.
where the average degree is given by \( \bar{d}(t) = 2e(t)/n(t) \sim 2m \) for large \( t \). Similarly, we have for \( j \in N_i \)

\[
Z_j(\beta, t) = \sum_{s \in (N_j \cup \{j\})} e^{\beta d_s(t)} \\
= 1 + \beta d_i(t) + \sum_{s \in (N_j \setminus \{i\})} (1 + \beta d_s(t)) + O(\beta^2) \\
\approx 1 + \beta d_i(t) + \bar{d}(1 + \beta \bar{d}) + O(\beta^2).
\]

(24)

With the expressions in Equations (23) and (24) we can solve Equation (22) and establish the following result.

**Theorem 3.** Assume that for small \( \beta \) we have only weak degree correlations so that Equations (23) and (24) hold in good approximation. Then the asymptotic cumulative degree distribution \( \lim_{t \to \infty} F_t(d) = F(d) \) for the NGNL model when \( \beta \) is small enough is given by

\[
F(d) = 1 - C(m, \beta, d) \left( \frac{1 + \beta d}{1 + \beta m} \right) \frac{2}{1 + 2m} \\
\times \left( \frac{(1 + \beta + 2m\beta)d^2 + (1 + \beta)d + 1 + 2m + 4m^2 \beta}{(1 + \beta + 2m\beta)m^2 + (1 + \beta)m + 1 + 2m + 4m^2 \beta} \right)^{-\frac{1 + 2m \beta + (1 + 4m) \beta}{m(1 + 2m \beta)}}
\]

(25)

with \( C(m, \beta, d) \in O(1) \) for large \( d \).

**Proof of Theorem 3.** Following Equation (23) we assume that

\[
Z_i(\beta, t) = \sum_{j \in (N_i \cup \{i\})} e^{\beta d_j(t)} \approx 1 + (1 + \beta(1 + 2m)) d_i(t) + O(\beta^2),
\]

for small \( \beta \). Similarly, from Equation (23) we obtain

\[
\sum_{j \in N_i} \frac{f(\beta, d_i(t))}{Z_i(\beta, t)} = \sum_{j \in N_i} \sum_{s \in (N_j \cup \{j\})} e^{\beta d_s(t)} \\
\approx \sum_{j \in N_i} \frac{1 + \beta d_i(t) + 1 + \beta d_j(t) + (d_j(t) - 1)(1 + 2m \beta)}{1 + 2m + 4m^2 \beta + \beta d_i(t)} + O(\beta^2) \\
\approx \frac{(1 + \beta d_i(t)) d_i(t)}{1 + 2m + 4m^2 \beta + \beta d_i(t)} + O(\beta^2).
\]

Inserting this into Equation (22) and neglecting terms of the order \( O(\beta^2) \),
further gives
\[
\frac{dd_i(t)}{dt} = \frac{m}{t} \left( \frac{1 + \beta d_i(t)}{1 + (1 + \beta(1 + 2m)) d_i(t)} + \frac{(1 + \beta d_i(t)) d_i(t)}{1 + 2m + 4m^2\beta + \beta d_i(t)} \right).
\]  
(26)

Equation (26) is a separable ODE and we can write
\[
\int_{t_i}^t \frac{m}{s} ds = \int_m^{d_i} \frac{1}{1 + \beta k} \left( \frac{1}{1 + k + 2\beta mk + \beta k} + \frac{k}{1 + 2m + 4m^2\beta + \beta k} \right)^{-1} dk.
\]  
(27)

Denoting by \( R(d_i) \) the integral on the right-hand side of the above equation, we obtain
\[
\ln \left( \frac{t}{t_i} \right)^m = R(d_i),
\]  
(28)

where we have denoted the time of birth of node \( i \) by \( t_i = i \). We then get \( i(d_i) = te^{-R(d_i)/m} \). Let \( F_i(d) \) denote the cumulative degree distribution function (cdf). Note that the proportion of nodes with a degree \( d > d_i \) is given by \( 1 - F_i(d) \). Equivalently, it is also given by \( i(d_i)/t \), since these are exactly the nodes that are younger than \( i \). Using the fact that \( F_i(d) = 1 - i(d_i)/t \), we arrive at
\[
F(d) = 1 - \exp \left( -\frac{1}{2m(1+2m\beta)} \left( 4m \ln \left( \frac{1 + \beta m}{1 + \beta d} \right) + B(m, \beta) \times \left( \arctan \left( \frac{2(1 + \beta + 2m\beta)d + 1 + \beta}{A(m, \beta)} \right) - \arctan \left( \frac{1 + \beta + 4m^2\beta + 2m(1 + \beta)}{A(m, \beta)} \right) \right) \right) + (1 + 2m(1 + \beta + 4m\beta)) \ln \left( \frac{(1 + \beta + 2m\beta)d^2 + (1 + \beta)d + 1 + 2m + 4m^2\beta}{1 + 2m^3\beta + m(3 + \beta) + m^2(1 + 5\beta)} \right) \right),
\]  
(29)

where
\[
A(m, \beta) = \sqrt{3 + 2\beta + 8m(1 + (2 + 4m)\beta)},
\]
\[
B(m, \beta) = \frac{2(1 - \beta + 16m^2\beta + 2m(1 + 2\beta))}{A(m, \beta)}.
\]

We define
\[
C(m, \beta, d) = \exp \left( -\frac{B(m, \beta)}{2m(1+2m\beta)} \left( \arctan \left( \frac{2(1 + \beta + 2m\beta)d + 1 + \beta}{A(m, \beta)} \right) - \arctan \left( \frac{1 + \beta + 4m^2\beta + 2m(1 + \beta)}{A(m, \beta)} \right) \right) \right)
\]

Note that \( \arctan(d) \to \frac{\pi}{2} \) for large \( d \) so that \( C(m, \beta, d) \in O(1) \). Inserting \( C(m, \beta, d) \) into Equation (29) gives Equation (25). In the weak selection
limit, when \( \beta = 0 \), we get from Equation (29)

\[
F(d) = 1 - \exp \left( -\frac{1 + 2m}{m} \left( \frac{1}{\sqrt{3 + 8m}} \left( \arctan \left( \frac{1 + 2d}{\sqrt{3 + 8m}} \right) - \arctan \left( \frac{1 + 2m}{\sqrt{3 + 8m}} \right) \right) \right) \right.
\]

\[
+ \frac{1}{2} \ln \left( \frac{1 + d + d^2 + 2m}{1 + m(3 + m)} \right) \right). \quad (30)
\]

For large \( d \) we then obtain from Equation (30) that \( 1 - F(d) \sim d^{-\frac{1 + 2m}{m}} \).

From Equation (25) it follows in the large noise limit, when \( \beta = 0 \), the cumulative distribution function (cdf) is given by

\[
1 - F(d) \sim d^{\frac{-1 + 2m}{m}}. \quad \text{(31)}
\]

The probability density function (pdf) corresponding to Equation (31) is given by\(^{21}\)

\[
p(d) \sim d^{\frac{-1 + 3m}{m}}. \quad \text{(32)}
\]

The comparison of the pdfs derived from the theoretical predictions of the cdfs from Equations (25) and (31) (the complete Equations (29) and (30) can be found in the proof of Theorem 3) is shown in Figure 5. From the figure we see that with increasing \( \beta \) the tail of the distribution becomes flatter as nodes with higher degrees become more probable.

4. Small Noise Equivalence and Efficiency

We see that as \( \beta \) decreases, NGNG and NGNL tend to produce different networks. In particular, in the large noise limit, networks formed under global information are geometric random graphs, with a geometric degree distribution given by Equation (18), while networks under local information are characterized by a power-law degree distribution as in Equation (32). However, as will be shown in this section, in the small noise limit, i.e. as \( \beta \) tends to infinity, global and local observability yield almost surely identical outcome networks. Note that for \( m = 1 \) both algorithms generate tree networks \( G(t) \in T_{n(t)} \) for all \( 1 \leq t \leq T \). In order to simplify our analysis we hence focus on the case of \( m = 1 \). However, the main results obtained in this section can also be generalized to \( m > 1 \) (see also Remark 3 at the end of this section). As a first result we derive some bounds for the probability that a star network, and therefore an efficient graph by Proposition 1, is formed under the two algorithms.

\(^{21}\)Note that the power-law exponent in Equation (32) has also been obtained by Wang et al. [2009] for a similar model.
Figure 5: Empirical degree distribution (averaged over 20 simulation runs) plotted with circles and the theoretical predictions of Equations (29) and (30) for $n = 10000$, $m = 5$ and $\beta = 0.01$, plotted with a dashed line.

**Proposition 4.** Let $\{G_1(t)\}_{t \geq 0}$ be a sequence of networks generated under the NGNG algorithm, and $\{H_1(t)\}_{t \geq 0}$ a sequence of networks generated under the NGNL algorithm for $m = 1$. Fix $\beta > 0$ and an integer $T \geq 1$. Let $v_1$ be the label of the unique vertex initially present in the respective initial graphs (i.e. the root vertex). Let $S$ be a star network with $T$ vertices having $v_1$ as its root. Then, there is a positive probability that

(a) NGNG and NGNL reach $S$, and

(b) this probability is at least $1/2$ as $\beta \to \infty$ for NGNL and NGNG.

**Proof of Proposition 4.** For $m = 1$ we know from the arguments presented in the Appendix that in NGNG almost surely and in finite time there exists a single dominant vertex that attracts almost all edges. We present here an alternative, more direct argument, which will suffice for the proof. We examine first the probability that NGNL reaches the star $S$. Observe
that

\[ P(\{S \subset H_1(t); \text{for some } t > 0\}|\text{NGNL}) \geq P(\{\text{link } T \text{ times in a row to the root } v_1\}|\text{NGNL}) \]

\[ = \prod_{j=1}^{T} \left( 1 - \frac{1}{j} \frac{e^{\beta(j-1)}}{e^\beta + e^{\beta(j-1)}} + \frac{1}{j} \frac{e^{\beta(j-1)}}{(j-1)e^\beta + e^{\beta(j-1)}} \right) \]

\[ = \prod_{j=1}^{T} \left( 1 - \frac{1}{j} p_1(\beta, j) + \frac{1}{j} p_2(\beta, j) \right) =: \mathcal{P}(\beta, T). \]

The first probability \( p_1(\beta, j) := \frac{e^{\beta(j-1)}}{e^\beta + e^{\beta(j-1)}} \) corresponds to the conditional probability that the new entering vertex at stage \( j + 1 \) samples one of the “children” of the root, and links to the root. Given that all the previous \( j \) nodes have linked to the root, there is a chance of \( 1 - 1/j \) to identify one of the children in the information neighborhood step of the algorithm. The second term \( p_2(\beta, j) := \frac{e^{\beta(j-1)}}{(j-1)e^\beta + e^{\beta(j-1)}} \) is the conditional probability that the entering node identifies the root in the information neighborhood step (what occurs with probability \( 1/j \)) and creates a link with the root. The product \( \prod_{j=1}^{T} p_2(\beta, j) \) is the probability that the root vertex receives \( T \) edges in a row under the NGNG algorithm. The sequences \( \{p_1(\beta, j)\}_{j \geq 2}, \{p_2(\beta, j)\}_{j \geq 2} \) are monotonically increasing with common limit 1. Moreover, \( p_2(\beta, 1) = 1, \)

\( p_1(\beta, 2) = p_2(\beta, 2) = \frac{1}{2}, \)

and \( p_1(\beta, j) > p_2(\beta, j) \) for all \( j > 2 \). Examining the product, one sees that for all finite \( T > 1 \) and all \( \beta \geq 0 \)

\[ \mathcal{P}(\beta, T) \leq p_2(\beta, 2). \]

From this the lower bound

\[ \mathcal{P}(\beta, T) \geq \prod_{j=2}^{T} p_2(\beta, j) \]

follows for all \( \beta \geq 0 \) and \( T \geq 1 \). Now

\[ p_2(\beta, j) = 1 - \frac{(j - 1)e^\beta}{(j-1)e^\beta + e^{\beta(j-1)}} = 1 - a_j, \]

where the sequence \( \{a_j\}_{j \geq 2} \) converges to 0 for all \( \beta > 0 \), and the ratio test shows that \( \sum_{j \geq 2} a_j < \infty \). It is then easy to deduce that \( \lim_{T \to \infty} \mathcal{P}(\beta, T) > 0 \)
for all $\beta > 0$.\footnote{Since $\sum_j a_j$ converges there exists an $N$ such that for all $T \geq N$, $a_N + \ldots + a_T < 1/2$. Hence,}

Hence, for all $\beta > 0$ and finite $T$ we have

$$0 < \mathcal{P}(\beta, T) \leq \frac{1}{2}.$$  \hfill (33)

This verifies part (a) of the proposition. For part (b) we simply remark that as $\beta \to \infty$ the upper bound in Equation (33) is indeed attained, since $p_2(\beta, 2) = 1/2$ and $p_2(\beta, j) \to 1$ for all $j > 2$ as $\beta \to \infty$. This shows that the lower bound approaches the upper bound of $\frac{1}{2}$. \hfill $\Box$

**Remark 2.** For $\beta = 0$ there is also a positive probability that the two algorithms generate the star $S$ for finite $T$. However, the larger $T$ the less likely it is that this will happen. For NGNG the probability that the root vertex receives $T$ links in a row is $\prod_{j=1}^{T} p_2(0, j) = \prod_{j=1}^{T} \frac{1}{2}$, which is positive for finite $T$. For large $T$ this product is very well approximated by

$$\prod_{j=1}^{T} e^{\gamma - \sum_{k=1}^{j} 1/k}$$

where $\gamma \approx 0.55721$ the Euler-Mascheroni constant and decays therefore very fast. For NGNL, by Jensen’s inequality, we can write

$$\log \mathcal{P}(0, T) = \sum_{j=1}^{T} \log \left[ (1 - 1/j)p_1(0, j) + \frac{1}{j}p_2(0, j) \right]$$

$$\geq \sum_{j=1}^{T} (1 - \frac{1}{j}) \log p_1(0, j) + \sum_{j=1}^{T} \frac{1}{j} \log p_2(0, j)$$

$$= - \sum_{j=1}^{T} (1 - \frac{1}{j}) \log 2 - \sum_{j=1}^{T} \frac{1}{j} \log j \to -\infty$$

as $T \to \infty$, but at a much slower rate. Hence, the situation is much better under NGNL than it is under NGNG.
This proposition shows that under both algorithms it is quite easy to obtain star networks. In the case of significant $\beta$ almost all generated networks will be of a star architecture. Proposition 4 shows that the fastest path that generates a star network has already a probability of nearly $1/2$ when $\beta$ is sufficiently large. The factor $1/2$ comes entirely from the choice behavior of the second agent entering the network, who creates a link either with the root or with the first entering agent with equal probability $1/2$. Hence, either the root or the first entering vertex must be the hub node in the network for $\beta$ sufficiently large.

For the linear quadratic game under small complementarity effects we know from our discussion in Section 2 that star networks have very good efficiency properties among the class of tree networks. Combining Proposition 1 with Proposition 4, we get the following nice result.

**Proposition 5.** Fix $m = 1$, $\epsilon > 0$, and a star network $S$ of finite order $T \geq 1$ where either the vertex with label $v_1$ or $v_2$ are the centers. Let $G$ be a graph generated either by NGNG or NGNL after $T$ iterations. Then there exists a $\beta(\epsilon) > 0$ such that for all $\beta > \beta(\epsilon)$ we have

$$P(G = S | \cdot) \geq 1 - \epsilon,$$

where the relevant conditioning is on either NGNG or NGNL.

**Proof of Proposition 5.** The event $\{S = G\}$ is equivalent to the event that $S$ is either the star with the initial vertex as hub, or the star with the vertex $v_2$ as hub. Both star networks can be reached with the same probability $\prod_{j=1}^{T} p_2(\beta, j)$ under NGNG, and $P(\beta, T)$ under NGNL. Hence,

$$P(G = S | \text{NGNG}) = 2 \prod_{j=1}^{T} p_2(\beta, j) \leq 2P(\beta, T) \leq 1.$$  

From Proposition 4 it follows that for all $\epsilon > 0$ we can find a $\beta(\epsilon) > 0$ such that for all $\beta > \beta(\epsilon)$

$$|2 \prod_{j=1}^{T} p_2(\beta, j) - 1| \leq \epsilon$$

giving us the desired result for NGNG and henceforth for NGNL.

The results we have obtained so far can be summarized in the following theorem.

**Theorem 4.** Fix $\alpha, \beta \geq 0$, $m = 1$ and $T > 0$. Denote by $\Pi^{\text{NGNG}}(\alpha, \beta, T)$ and $\Pi^{\text{NGNL}}(\alpha, \beta, T)$ the expected aggregate payoffs under the NGNG, respectively NGNL, algorithm after $T$ iterations. Let $\alpha$ be small enough such that Assumption 1 holds. Assume that for small $\beta$ we have only weak degree
correlations so that Equations (23) and (24) hold in good approximation. Then

(a) for $\beta = 0$ we have $\Pi^{NGNL}(\alpha, \beta, T) > \Pi^{NGNG}(\alpha, \beta, T)$, while
(b) for $\beta \to \infty$ we have $\Pi^{NGNL}(\alpha, \beta, T) = \Pi^{NGNG}(\alpha, \beta, T) = \Pi(K_{1,T-1}, \alpha)$, almost surely.

PROOF OF THEOREM 4.

(a) For NGNG, we know that the degree distribution is given by Equation (18). From this geometric distribution we can compute the expectation $E(d^2) = m(1 + 5m) = 6$. For NGNL, the degree distribution is given by Equation (25). In the case of $m = 1$ and $\beta = 0$ we get

$$E(d^2) = \int_0^\infty \frac{15\sqrt{5}d^2(1 + d)}{(3 + d + d^2)^2} \times \exp \left( \frac{3}{\sqrt{11}} \left( \arctan \left( \frac{3}{\sqrt{11}} \right) - \arctan \left( \frac{1 + 2d}{\sqrt{11}} \right) \right) \right) \, dd$$

This integral can be solved numerically to obtain $E(d^2) = 7.36$, which is larger than what we obtain for the NGNG model. Dropping terms of the order $O(\alpha^3)$ in Equation (10) we thus find that expected aggregate payoffs are larger for the NGNL than for the NGNG algorithm when $m = 1$ and $\beta = 0$.

(b) This follows from Propositions 4 and 5.
Theorem 4 shows that in the case where $\beta \to \infty$, i.e. where the new entering agents with high probability sample incumbent agents having high Bonacich centrality, both global and local information yield identical outcomes and agents are able to coordinate on the efficient network (the star $K_{1,T-1}$, see Proposition 1) in terms of maximizing aggregate payoffs. By means of numerical simulations we also observe that for lower levels of noise networks generated by NGNG have a higher aggregate payoff than those by NGNL. However, in the presence of strong noise for vanishing $\beta$, local information leads to higher aggregate payoffs as networks being formed with global information. As support for the results of the displayed by numerical simulations, we present the following lemma, which might serve as the basis for a more rigorous proof.

**Lemma 1.** Given a nonempty graph $G = (V,E)$ of size $|V| = n$. Let $K_G^\beta$ denote the attachment kernel of the NGNG algorithm and $\tilde{K}_G^\beta$ the attachment kernel of the NGNL algorithm. Let $i = \arg \max \{d_j(G) | j \in V\}$ denote the agent with the largest degree in the graph $G$. Consider the function

$$\psi_G^\beta(i) := K_G^\beta(i) - \tilde{K}_G^\beta(i).$$

There exists a $\beta^*$, depending only on the graph $G$, such that $\psi_G^\beta(i) > 0$ for $\beta > \beta^*$ and $\psi_G^\beta(i) \leq 0$ for $0 \leq \beta < \beta^*$.

**Proof of Lemma 1.** A simple algebraic manipulation gives us

$$\psi_G^\beta(i) = K_G^\beta(i) \sum_{k \in \bar{N}_i(G)} \left(\frac{1}{d_k(G)} + 1 - \frac{1}{2n} \frac{Z(\beta,G)}{Z_{k}(\beta,t)}\right). \quad (35)$$

At $\beta = 0$ this can be evaluated as

$$\psi_G^0(i) = \frac{1}{n} \sum_{j \in \bar{N}_i(G)} \frac{d_k(G) - d_i(G)}{(d_i(G) + 1)d_k(G) + 1} \leq 0$$

as $d_i(G)$ is the largest degree in the graph.

For $\beta > 0$ but sufficiently small we know that

$$Z(\beta,G) = \sum_{j \in V} (1 + \beta d_j(G)) \leq n(1 + \beta d_i(G)).$$

Hence $-\frac{Z(\beta,G)}{n} \geq -(1 + d_i(G))$. Further, since $d_i(G)$ is the largest degree in
the graph we get the bound

\[ Z_k(\beta, G) = \sum_{j \in \tilde{N}_k(G)} (1 + \beta d_j(G)) + o(\beta) \]

\[ = d_k(G) + 1 + \beta \sum_{j \in \tilde{N}_k(G)} d_j(G) + o(\beta) \]

\[ \leq d_i(G) + 1 + \beta (d_i(G) + 1)d_i(G) + o(\beta) \]

\[ \leq (d_i(G) + 1)^2 \]

Hence for \( \beta \) sufficiently small we have for all \( k \in \tilde{N}_i(G) \)

\[ \frac{1}{d_i(G) + 1} - \frac{1}{n} Z_k(\beta, G) \geq \frac{1}{d_i(G) + 1} + \frac{-(1 + d_i(G))}{(d_i(G) + 1)^2} = 0 \]

Thus there exists a \( \beta^* \) as claimed in the lemma. \( \square \)

Evaluated at the largest degree vertex, the attachment kernels of both models are increasing functions of the noise parameter \( \beta \). Lemma 1 shows that for \( \beta < \beta^* \) the probability that the largest degree agent receives an additional neighbor is higher in the NGNL process than in the NGNG process, but for \( \beta > \beta^* \) the situation is reversed and large degree nodes are preferred by the NGNG algorithm. An alternative illustration of this phenomenon can be given by considering the exponential rate of decay of the attachment kernels in the large \( \beta \) case. For the NGNG attachment kernel we see that

\[ \lim_{\beta \to \infty} \frac{1}{\beta} \log K^\beta_t(i) = d_i(t) - \max_{j \in \tilde{V}(t)} d_j(t), \]

so that for \( \beta \) sufficiently large we have

\[ K^\beta_t(i) = \exp \left( \beta (d_i(t) - \max_{j \in \tilde{V}(t)} d_j(t) + o(1)) \right). \]

In contrast to this, the exponential rate of decay of the probability that agent \( i \in V(t) \) receives an additional edge in the NGNL process is

\[ \lim_{\beta \to \infty} \frac{1}{\beta} \log \tilde{K}^\beta_t(i) = d_i(t) - \max_{j \in \tilde{N}_i(t)} \max_{k \in \tilde{N}_j(t)} d_k(t). \]

Hence, the rate of decay of the attachment kernel under NGNG is driven by the difference of the degree of an agent compared to the largest degree in the whole network. Thus, as \( \beta \) grows large \( K^\beta_t \) puts almost all mass on the largest degree agents in the network. In NGNL, however, it is not the largest degree of the network, but the largest degree in the relevant neighborhood of the agent. Hence, as in NGNG, nodes with the largest degree will be selected for attachment with very high probability for large \( \beta \), but also nodes that
have a high local degree will have a chance to receive a link under NGNL.

**Remark 3.** We can allow for general values of $m > 1$ and consider the case of $\beta = 0$. For NGNG, we know that the degree distribution is given by Equation (18). From this geometric distribution we can compute the expectation $E(d^2) = m(1 + 5m)$. For NGNL, the degree distribution is given by Equation (25). There does not exist a closed form solution of $E(d^2)$. However, we can compute this expectation numerically. The result, in comparison with NGNG is shown in Figure 6 (right). As the figure reveals, $E(d^2)$ is always higher for NGNL than for NGNG. Hence, for $\alpha$ small enough such that Assumption 1 applies, we find that NGNL generates higher aggregate payoffs than NGNG.

5. Conclusion

In this paper we have introduced a network formation model in which the decision of agents to form links is endogenous and depends on the marginal payoff associated with a link, the degree of rationality of agents governing the level of noise with which agents perceive their marginal payoff and the amount of information available to agents when deciding with whom to connect. In this framework, we show that fully rational agents form the same networks irrespective of the information they have of the network. This means that the same networks are obtained under both, local and global information. However, this is only true in the limit of vanishing noise. As the level of noise becomes dominant, local information typically generates networks with a higher aggregate payoff than networks formed with global information.

The paper could be extended along several directions. First, we have assumed that the network is formed by incoming agents only, while neglecting the possibility of incumbent agents forming links. It would be interesting to extend our model by allowing both, newcomer and incumbent agents to form links in a similar way. Second, we have analyzed the network formation process for small levels of the complementary parameter in Equation (2). This assumption is needed in order to guarantee that a unique Nash equilibrium exists and that we can approximate the Nash equilibrium strategies by the degree of an agent. A further analysis could study the case of stronger complementarity effects. Finally, an extension of the analysis presented in Section 3.2 could try to fully capture the effects of non-vanishing degree correlations on the emerging network statistics.

References

König, Michael, D., Tessone, Claudio, J., Zenou, Y., 2009. A Dynamic Model of Network Formation with Strategic Interactions, CEPR Discussion Pa-
Appendix

A. Dynamic Adjustment of Strategies to the Nash Equilibrium

We define by $\text{BR} := (\text{BR}_i)_{i \in \mathbb{N}}$ the countable infinite column vector of best response functions with elements

$$\text{BR}_i(x_{-i}; \alpha, G) := \arg \max_{y \geq 0} \pi_i(y, x_{-i}; \alpha, G) = 1 + \alpha \sum_{j \in N_i} x_j.$$

(36)

The best-response dynamics is defined by the vector field

$$\dot{x} = F(x) := \text{BR}(x; \alpha, G(t)) - x.$$

(37)

Under these dynamics, passive players are immediately in a steady state, while active players $i \in V(t)$ adjust their behavior according to

$$\dot{x}_i = \text{BR}_i(x_{-i}, G(t)) - x_i,$$

(38)

where

$$\text{BR}_i(x_{-i}, G(t)) = 1 + \alpha \sum_{j \in V(t)} a_{ij}(t)x_j.$$

(39)

It is important to note that the best response function of active players depends only on a finite number of elements of the action vector $x$. More precisely, it depends only on the actions of the neighbors. This implies that an agent needs only local information of the network, i.e. knowledge of the strategies of his neighbors, in order to determine his best reply strategy. With the help of the potential function, identified in Equation (6), we can apply many well-known theorems on the asymptotic behavior of the dynamical system (37). In particular, it is easy to see that the vector field for the active players can be written as

$$\dot{x}_i = \frac{\partial P(x; \alpha, G(t))}{\partial x_i},$$

(40)

showing that fixed points of the dynamic coincide with the stationary states of the potential function. The potential function $P(x; \alpha, G)$ is strictly concave and so the unique stationary state must be a global maximizer of the potential function, which is exactly achieved when each player chooses his Bonacich centrality in the graph $G(t)$ as action. That is, the fixed point of the dynamics in Equation (38) give the Nash equilibrium strategies in Equation (5) of Theorem 1.

B. Dynamic Adjustment of Effort Levels and Links

The probability with which links are formed depends on the marginal payoff the entering agent receives from a link. Let $v_{t+1}$ be the agent entering
the network \( G(t) = (V(t), E(t)) \) at time \( t + 1 \). Let \( I(t + 1) \subseteq V(t) \) be the information neighborhood of the entering agent \( v_{t+1} \). \( I(t + 1) \) contains all agents in \( G(t) \), agent \( v_{t+1} \) observes and hence can potentially link to. Assume that \( v_{t+1} \) creates a link to agent \( i \in I(t + 1) \). For any \( i \in I(t + 1) \) denote by \( G(t) + vi = (V(t) \cup \{v_{t+1}\}, E(t) \cup \{(v_{t+1}, i)\}) \) the network obtained by the addition of the link \((v_{t+1}, i)\) to \( G(t) \). Given the strategies \( x = (x_j)_{j \in V(t)} \) of the agents \( j \in V(t) \) the payoff the agent \( v_{t+1} \) receives from forming the link \((v_{t+1}, i)\) and choosing strategy \( y \in \mathbb{R}_+ \) follows from Equation (2) as

\[
\pi_{v_{t+1}}((x, y); \alpha, G(t) + vi) = y - \frac{1}{2}y^2 + \alpha y x_i. \tag{41}
\]

We then assume that all players in the set \( V(t + 1) = V(t) \cup \{v_{t+1}\} \) follow the adjustment rule of Equation (38) in the action evolution process, which converges to the Nash equilibrium identified in Theorem 1 at a reasonable fast rate. As a consequence of Equation (5), in this Nash equilibrium agent \( v_{t+1} \)’s strategy is given by his Bonacich centrality \( y^* = b_{v_{t+1}}(G(t) + vi, \alpha) \) and his equilibrium payoff is given by

\[
\pi^*_{v_{t+1}}(G(t) + vi, \alpha) = \frac{1}{2} b_{v_{t+1}}(G(t) + vi, \alpha)^2. \tag{42}
\]

The payoff agent \( v_{t+1} \) receives from not being connected to any other agent is one (the autarky payoff). Hence, the marginal payoff agent \( v_{t+1} \) receives from creating the link \((v_{t+1}, i)\) is \( \pi^*_{v_{t+1}}(G(t) + vi, \alpha) - 1 \). We then assume that the above procedure is repeated separately, for each agent \( i \in I(t + 1) \). This means that for each \( i \in I(t + 1) \) the link \((v_{t+1}, i)\) is formed and the payoff \( \pi^*_{v_{t+1}}(G(t) + vi, \alpha) \) is obtained as a fixed point of the adjustment dynamics in Equation (38). In this way, agent \( v_{t+1} \) gets to know all marginal payoffs for all possible links he can create to agents \( i \in I(t + 1) \). For doing this, agent \( v_{t+1} \) needs only to observe the strategies of the agents in the set \( I(t + 1) \).

One may think of the evolutionary process as running on two distinct time scales, a fast one for action adjustments and a much slower one for network evolution. We assume that the time in which agents are forming new links evolves much slower than the rate at which the adjustment process is repeated (see Vega-Redondo [2006] and König et al. [2009], for a similar approach).23,24

Following from Equation (4), the Bonacich centrality of agent \( i \) can be writ-
ten as $b_i(G(t) + vi, \alpha) = 1 + \alpha(d_i(t) + 1) + O(\alpha^2)$. Accordingly, the payoff of the entering agent $v_{t+1}$ from Equation (42) in the network $G(t) + vi$ can then be written as

$$\pi_{v_{t+1}}(G(t) + vi, \alpha) = \frac{1}{2} b_{v_{t+1}}(G(t) + vi, \alpha)^2$$

$$= \frac{1}{2} (1 + \alpha b_i(G(t) + vi, \alpha))^2$$

$$= \frac{1}{2} + \alpha + \alpha^2 + \alpha^2 d_i(t) + O(\alpha^3).$$

In the following we neglect terms of the order $O(\alpha^3)$. This assumption is a structural assumption on the preferences of the players, implying essentially that interactions with close neighbors weigh much more than interactions with agents at larger (geodesic) distance in the graph. Neglecting terms of the order $O(\alpha^3)$ yields a marginal payoff of $\Delta \pi_{v_{t+1}}(G(t) + vi, \alpha) := \pi_{v_{t+1}}(G(t) + vi, \alpha) - 1 = \alpha - \frac{1}{2} + \alpha^2 + \alpha^2 d_i(t)$, which is entirely determined by the degree of the target agent $i$ agent $v_{t+1}$ connects to. We will thus assume that the marginal payoff for agent $v_{t+1}$ from forming a link to agent $i$ is given by $\Delta \tilde{\pi}_{v_{t+1}}(G(t) + vi, \alpha) = d_i(t)$, which is entirely determined by the degree $d_i(t)$ of agent $i$.

C. Dynamic Adjustment of Strategies of Entrants to Frozen Strategies of Incumbents

We assume that at every point of time $t + 1$ at which the network $G(t)$ changes, the actions of the players are frozen to the Nash strategy of the previous period $t$. The entering agent $v_{t+1}$ at time $t+1$ observes the strategy of an incumbent agent $i$ in his information set $I(t+1)$ at the previous period $t$, which is given by her Bonacich centrality $b_i(G(t), \alpha)$. If agent $v_{t+1}$ forms a link to agent $i \in I(t+1)$ and chooses strategy $y \in \mathbb{R}_+$ then his payoff is given by

$$\pi_{v}(x, y; G(t) + vi, \alpha) = y - \frac{1}{2} y^2 + \alpha y b_i(G(t), \alpha).$$  \hfill (43)

The best strategy agent $v_{t+1}$ can choose is $y^* = 1 + \alpha b_i(G(t), \alpha)$. Agent $v_{t+1}$’s payoff is then given by

$$\pi_{v_{t+1}}^*(G(t) + vi, \alpha) = \frac{1}{2} + \alpha b_i(G(t), \alpha) - \frac{1}{2} \alpha^2 b_i(G(t), \alpha)^2.$$  \hfill (44)

From the above payoff function we get

$$\frac{d\pi_{v_{t+1}}^*}{db_i} \geq 0 \iff b_i \leq \frac{1}{\alpha}.$$  \hfill (45)
Thus we find that for $\alpha$ small enough the payoff $\pi^*_v t + 1$ of the entering agent $v t + 1$ is increasing in the Bonacich centrality of the agent $i$, $v t + 1$ connects to. The Bonacich centrality can be written as $b_i(G(t), \alpha) = 1 + \alpha d_i(G(t), \alpha) + O(\alpha^2)$. Neglecting terms of the order $O(\alpha^2)$ for small values of $\alpha$, the inequality $b_i < 1/\alpha$ in Equation (45) implies that

$$d_i(t) < \frac{1}{\alpha} \left( \frac{1}{\alpha} - 1 \right) \sim \frac{1}{\alpha^2}.$$  

for all $i \in I(t + 1) \subseteq V(t)$. This is equivalent to

$$\alpha < \frac{1}{\sqrt{d_{\max}(t)}},$$  

(46)

where $d_{\max}(t) = \max_{i \in V(t)} d_i(t)$. For the special case of a star $K_{1, n-1}$ we have that $\lambda_{PF} = \sqrt{d_{\max}} = \sqrt{n - 1}$ and therefore $\alpha < \sqrt{n - 1}$. For generic networks we can use the approximation [see e.g Restrepo et al., 2007]

$$\lambda_{PF}(G(t)) \approx \max \left\{ \frac{\mathbb{E}(d^2(t))}{\mathbb{E}(d(t))}, \sqrt{\mathbb{E}(d(t))} \right\}.$$  

(47)

Then the condition in Equation (46) is implied by the inequality $\alpha < 1/\lambda_{PF}(G(t))$. This is exactly the upper bound on $\alpha$ required for the existence of the Nash equilibrium in Theorem 1. Thus, if the condition for the existence of a Nash equilibrium in Theorem 1 is satisfied, it follows that the entering agent $v t + 1$ obtains a higher payoff the higher is the degree of the agent $i$ agent $v_{t+1}$ connects to. Note that an agent needs only to observe the strategies of the agents in his information neighborhood to determine his best response.

### D. Analysis of the NGNG Algorithm

The NGNG model is tractable enough to derive some of its properties rigorously. In this section we present a general derivation of the degree distribution of the NGNG. The Ansatz is valid for any preference function $f(\beta, \cdot)$ and general $m \geq 1$. However, only the asymptotic results will go through for preference functions which are of a "non-explosive type", as made precise below.

#### D.1. The Degree Distribution

Here we present the derivation of the expected degree distribution for general $m$. Without loss of generality we shift time so that $t_0 = 0$. Let $\{G_m(t)\}_{t \in \mathbb{N}}$ denote the generated sequence of graphs by the NGNG algorithm. As initial graph $G_m(0)$ we assume that there are $n_0 \geq m$ agents in the population, who share no connections. Our goal will be to derive an recursive definition
of the expected number of nodes of degree \( k \) in period \( t + 1 \) in terms of the degree sequence \( N(t) = \{N_k(t)\}_{k \geq 0} \). This equation will be of the following form
\[
\mathbb{E}[N_k(t + 1)|\mathcal{F}_t] = N_k(t) + \omega^+(k, t) - \omega^-(k, t)
\] (48)
where \( \omega^+(k, t) \) is the expected inflow into the set \( N_k(t) \) and \( \omega^-(k, t) \) is the expected outflow from the set \( N_k(t) \). Let us derive these quantities step-by-step.

- The outflow from the set \( N_k(t) \) is exactly the number of nodes with degree \( k \) at time \( t \), \( N_k(t) \), times the probability that a node from this set is chosen at least once by the newcomer in the global search algorithm. This probability is
\[
1 - \left(1 - \frac{f(\beta, k)}{Z(\beta, t)}\right)^m.
\]
Hence,
\[
\omega^-(k, t) := N_k(t) \left[1 - \left(1 - \frac{f(\beta, k)}{Z(\beta, t)}\right)^m\right].
\] (49)

- To determine the expected inflow, simply observe that in order for a node with current degree \( k - s \) to have degree \( k \) in \( t + 1 \) it must be sampled exactly \( s \) times in the global search algorithm. The expected flow from \( N_{k-s}(t) \) to \( N_k(t + 1) \) is thus
\[
N_{k-s}(t) \binom{m}{s} \left(\frac{f(\beta, k - s)}{Z(\beta, t)}\right)^s \left(1 - \frac{f(\beta, k - s)}{Z(\beta, t)}\right)^{m-s}
\]
for \( s \in \{1, \ldots, m\} \). In addition to this, we have to take into account that at every step one new vertex is added to the network and this vertex creates \( m \) links for sure.\(^{25}\) In total, we count the expected inflow into the set \( N_k(t) \) as
\[
\omega^+(k, t) = \sum_{s=1}^{m} N_{k-s}(t) \binom{m}{s} \left(\frac{f(\beta, k - s)}{Z(\beta, t)}\right)^s \left(1 - \frac{f(\beta, k - s)}{Z(\beta, t)}\right)^{m-s} + \delta_{k,m},
\] (50)
where \( \delta_{x,y} \) is the Kronecker delta, equal to 1 if \( x = y \) and 0 otherwise.

\(^{25}\)Recall that our graphs are multigraphs in which we count every edge with its multiplicity.
Plugging this into equation (48) we get

\[
E[N_k(t+1) | \mathcal{F}_t] = N_k(t) \left( 1 - \frac{f(\beta, k)}{Z(\beta, t)} \right)^m + \sum_{s=1}^{m} N_{k-s}(t) \left( \frac{m}{s} \right) \left( \frac{f(\beta, k - s)}{Z(\beta, t)} \right)^s \left( 1 - \frac{f(\beta, k - s)}{Z(\beta, t)} \right)^{m-s} + \delta_{k,m}.
\]

Define \( \zeta(\beta, t) := \frac{1}{t}Z(\beta, t) \) and rewrite the equation above in terms of the proportion of agents with degree \( k \) to get

\[
E[P_k(t+1) | \mathcal{F}_t] = \frac{n(t)}{n(t+1)} \left( \frac{1}{t \zeta(\beta, t)} \right)^m \left\{ P_k(t) [t \zeta(\beta, t) - f(\beta, k)]^m + \frac{(t \zeta(\beta, t))^m}{n(t)} \delta_{k,m} \middle| X \right\} + \sum_{s=1}^{m} P_{k-s}(t) \left( \frac{m}{s} \right) f(\beta, k - s)^s [t \zeta(\beta, t) - f(\beta, k - s)]^{m-s}
\]

Subtract \( P_k(t) \) from the left and the right side of this equation and then take expectations to arrive at

\[
\tilde{P}_k(t+1) - \tilde{P}_k(t) = \frac{1}{n(t+1)} \left( \frac{1}{t \zeta(\beta, t)} \right)^m \left\{ \tilde{P}_k(t) n(t) [(t \zeta(\beta, t) - f(\beta, k))^m - n(t+1)t^m \zeta(\beta, t)^m] \right\} + \sum_{s=1}^{m} n(t) \tilde{P}_{k-s}(t) \left( \frac{m}{s} \right) f(\beta, k - s)^s (t \zeta(\beta, t) - f(\beta, k - s))^{m-s} + t^m \zeta(\beta, t)^m \delta_{k,m}
\]

Now collect terms according to their order in \( t \). Recall that \( n(t) = n_0 + t \) and \( n(t+1) = n(t) + 1 \). By the Binomial theorem, the multiplicative term in \( n(t) \tilde{P}_k(t) \) can be simplified to

\[
n(t) [(t \zeta(\beta, t) - f(\beta, k))^m - v(t + 1)t^m \zeta(\beta, t)^m] = n(t) \left[ t^m \zeta(\beta, t)^m - m \zeta(\beta, t)^{m-1} \zeta(\beta, t) + O(t^{m-2}) \right] - (n(t) + 1)t^m \zeta(\beta, t)^m = -t^m \zeta(\beta, t)^m - (n_0 + t)m \zeta(\beta, t)^{m-1} + O(t^{m-1}) = -m \zeta(\beta, t)^{m-1} \left( \frac{\zeta(\beta, t)}{m} + \zeta(\beta, t) \right) + O(t^{m-1})
\]

Similarly, the multiplicative term for \( n(t) \tilde{P}_{k-1}(t) \) can be simplified to

\[
n(t)m f(\beta, k - 1) \left[ t^{m-1} \zeta(\beta, t)^{m-1} + O(t^{m-2}) \right] = t^m m f(\beta, k - 1) \zeta(\beta, t)^{m-1} + O(t^{m-1})
\]

38
The last terms for \( s \geq 2 \) are of order \( O(t^{m-2}) \) and so shall be combined under this symbol. In summary, we arrive at the recursive formula

\[
P_k(t+1) - \bar{P}_k(t) = \frac{1}{v(t+1)} \left( \frac{1}{t \zeta(\beta, t)} \right)^m \left[ \bar{P}_{k-1}(t) t^m f(\beta, k-1) \zeta(\beta, t)^{m-1} \right.
\]

\[
- \bar{P}_k(t) m t^m \zeta(\beta, t)^{m-1} \left( \frac{\zeta(\beta, t)}{m} + f(\beta, k) \right) + t^m \zeta(\beta, t)^m \delta_{k,m} + O(t^{m-1})
\]

\[
= \frac{1}{n(t+1)} \frac{m}{\zeta(\beta, t)} \left[ \bar{P}_{k-1}(t) f(\beta, k-1) - \bar{P}_k(t) \left( \frac{\zeta(\beta, t)}{m} + f(\beta, k) \right) \right.
\]

\[
+ \frac{\zeta(\beta, t)}{m} \delta_{k,0} + O(t^{m-1})
\]

\[
= \frac{1}{n(t+1)} \frac{m}{\zeta(\beta, t)} \left( f(\beta, k) + \frac{\zeta(\beta, t)}{m} \right) \left[ \frac{f(\beta, k-1)}{f(\beta, k) + \zeta(\beta, t)/m} \bar{P}_{k-1}(t) \right.
\]

\[
- \bar{P}_k(t) + \frac{\zeta(\beta, t)/m}{f(\beta, k) + \zeta(\beta, t)/m} \delta_{k,m} + O(1/t).
\]

Calling

\[
b_k(t+1) := \frac{1}{n(t+1)} \frac{m}{\zeta(\beta, t)} \left( f(\beta, k) + \frac{\zeta(\beta, t)}{m} \right),
\]

\[
c_k(t) := \frac{f(\beta, k-1)}{f(\beta, k) + \zeta(\beta, t)/m } \bar{P}_{k-1}(t) + \frac{\zeta(\beta, t)/m}{f(\beta, k) + \zeta(\beta, t)/m} \delta_{k,m}.
\]

we arrive at the recursion

\[
P_k(t+1) - \bar{P}_k(t) = b_k(t+1)(c_k(t) - \bar{P}_k(t)) + O(1/t^2) \quad (t \geq 1).
\]

Note that we have not specified the form of the attachment rates \( f(\beta, k) \) yet. In principle any function can be used to arrive at the recursive definition of the empirical distribution function. However, not all preference functions \( f(\beta, \cdot) \) will admit a solution. The following lemma gives us a simple way to determine the asymptotic solution (i.e. \( t \to \infty \)) of the recursion in Equation (53).

**Lemma 2.** Let \((x_n), (y_n), (\eta_n), (r_n)\) denote real sequences such that

\[
x_{n+1} - x_n = \eta_{n+1} (y_n - x_n) + r_{n+1}
\]

and

(a) \( \lim_{n \to \infty} y_n = x \)

(b) \( \eta_n > 0, \sum_{n=1}^{\infty} \eta_n = \infty \) and there exists a \( N_0 \) such that for all \( n \geq N_0 \)

\[
\eta_n < 1.
\]

(c) \( r_n = o(\eta_n) \)

39
Then $\lim_{n \to \infty} x_n = x$.

**Proof of Lemma 2.** See Jordan [2006], p. 229.

For our purposes the lemma can be applied by identifying $\eta_t = b_k(t)$ and $\gamma_t = c_k(t)$. Since $n(t) = n_0 + t$, we have $b_k(t) > 0$ and $\sum_{t \geq 0} b_k(t) = \infty$ if $\lim_{t \to \infty} \zeta(\beta, t) = \zeta^*(\beta) < \infty$. Under this condition it is evident that $c_k(t)$ has a well-defined limit, which is determined in a recursive way. Hence, existence of a proper solution of the system (53) is coupled with the existence of the limit $\lim_{t \to \infty} \frac{2(\beta, t)}{m}$. With these insights we are now ready to prove our main propositions needed for Theorem 2.

**Proof of Proposition 2.** We will determine the asymptotic expected degree distribution under the linearized preference function $f(\beta, \cdot)$ via induction on $k$. Call $\lim_{t \to \infty} \bar{P}_k(t) = p^\beta_k$. In view of Lemma 2, we only have to determine the limit of

$$c_k(t) = \frac{f(\beta, k - 1)}{f(\beta, k) + \zeta(\beta, t)/m} \bar{P}_{k-1}(t) + \frac{\zeta(\beta, t)/m}{f(\beta, k) + \zeta(\beta, t)/m} \delta_{k,m}$$

as $t \to \infty$. Since $v(t) = n_0 + t$, we have $b_k(t) := \frac{1}{m(t+1)} \frac{m}{f(\beta, k) + \zeta(\beta, t)/m} > 0$ and $\sum_{t \geq 0} b_k(t) = \infty$ if $\lim_{t \to \infty} \zeta(\beta, t) = \zeta^*(\beta) < \infty$. Under this condition it is evident that $c_k(t)$ has a well-defined limit, which is determined in a recursive way. All newborn nodes enter the population by adding $m$ edges to the network. The only way that nodes have no degree is that they are in the initial population $V(0)$ and do not receive any links in course of the algorithm. Hence, for sufficiently large $t$ the fraction of agents with degree $k < m$ is negligible.

To start the induction begin with $k = m$. Taking Equation (53) as starting point we see that

$$c_m(t) = \frac{\zeta(\beta, t)}{mf(\beta, m) + \zeta(\beta, t)} \to \frac{\zeta^*(\beta)}{mf(\beta, m) + \zeta^*(\beta)}$$

as $t \to \infty$. Call this $p^\beta_m$.

To proceed with the induction proof, suppose we have already determined the upper tail of the distribution $p^\beta_m, \ldots, p^\beta_{k-1}, k > m$. Then we see that

$$c_k(t) = \frac{f(\beta, k - 1)}{f(\beta, k) + \zeta(\beta, t)/m} \bar{P}_{k-1}(t) \to \frac{f(\beta, k - 1)}{f(\beta, k) + \zeta^*(\beta)/m} p^\beta_{k-1}$$

for all $k \geq 1$. Call $p^\beta_k = \lim_{t \to \infty} c_k(t)$, and iterating this equation with respect to $k$, gives us the expression

$$p^\beta_k = \frac{\zeta^*(\beta)/m}{f(\beta, k)} \prod_{j=m}^{k} \frac{f(\beta, j)}{f(\beta, j) + \zeta^*(\beta)/m}.$$
If we substitute \( f(\beta, x) = 1 + \beta x \), this expression simplifies to

\[
p_{\beta}^{k} = \frac{(1 + 2m\beta)\Gamma \left[ k + \frac{1}{\beta} \right] \Gamma \left[ 2 + m + \frac{1}{\beta} + \frac{1}{m\beta} \right]}{m\beta \Gamma \left[ m + \frac{1}{\beta} \right] \Gamma \left[ 3 + k + \frac{1}{\beta} + \frac{1}{m\beta} \right]}.
\]

(55)

From Equation (55) we can derive the large \( k \) approximation of the degree distribution presented in equation (17). Set \( t := k + 1/\beta \) and \( a := 3 + 1/m\beta \). The leading term in \( k \) in equation (55) is then of the form \( \frac{\Gamma(t)}{\Gamma(t+a)} \). By Stirling’s formula we can approximate the Gamma function for large \( t \) as

\[
\Gamma(t) = \sqrt{\frac{2\pi}{t}} \left( \frac{t}{e} \right)^{t} \left( 1 + O \left( \frac{1}{t} \right) \right).
\]

Hence,

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

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

\[
\frac{\Gamma(t)}{\Gamma(t+a)} = t^{-a} \left( 1 + O \left( \frac{1}{t} \right) \right).
\]

For the tails of the degree distribution (55) this implies that

\[
p_{\beta}^{k} \sim (1 + \beta k)^{-\left(3 + \frac{1}{\beta} + \frac{1}{m\beta} \right)} \left( 1 + O \left( \frac{1}{k} \right) \right).
\]

**Proof of Corollary 2.** Since \( \zeta^*(0) = 1 \), we get

\[
p_{m}^{0} = \frac{1}{1 + m}
\]

and setting \( \beta = 0 \) in the first line of the induction step (54) gives the desired result.

**Proof of Proposition 3.** Here we follow the work of Bollobás et al. [2001], who were the first to employ martingale techniques to provide such a prob-
abilistic bound. Fix $t > t_0$ and a degree $k$. Define
\[ M(s) := \mathbb{E}[N_k(t)|\mathcal{F}_s] \]
We will show that $\{M(s)\}_{s=t_0}^t$ is a Doob martingale [see e.g. Grimmet and Stirzaker, 2001, Section 7]). First of all, by taking expectations, we see that
\[ \mathbb{E}[M(s)] = \mathbb{E}[N_k(t)] \leq n(t). \]
Hence, the first moment of $\{M(s)\}_{s=t_0}^t$ is bounded. Similarly one can show that $\mathbb{E}[M(s)^2] < \infty$. Further, by the Tower property,
\[
\mathbb{E}[M(s)|\mathcal{F}_{s-1}] = \mathbb{E}\{\mathbb{E}[N_k(t)|\mathcal{F}_s]|\mathcal{F}_{s-1}\}
\]
\[
= \mathbb{E}[N_k(t)|\mathcal{F}_{s-1}]
\]
\[
= M(s-1).
\]
This proves the claim that $\{M(s)\}_{s=t_0}^t$ is indeed a martingale with respect to the filtration $\{\mathcal{F}_s\}_{s=t_0}^t$. By definition of the random variables we have
\[ M(t) = \mathbb{E}[N_k(t)|\mathcal{F}_t] = N_k(t) \]
and
\[ M(t_0) = \mathbb{E}[N_k(t)] \]
since $\mathcal{F}_{t_0}$ gives no more information than the initial graph $G(t_0)$. We want to apply Hoeffding’s inequality. Therefore, we need to find a universal bound on the first-order martingale differences. We claim that
\[ |M(s) - M(s-1)| \leq 2m \]
To see why this bound holds, note that
\[ |M(s) - M(s-1)| = |\mathbb{E}[N_k(t)|\mathcal{F}_s] - \mathbb{E}[N_k(t)|\mathcal{F}_{s-1}]|. \]
Since $\mathcal{F}_{s-1} \subset \mathcal{F}_s$ and the additional information we obtain by conditioning on $\mathcal{F}_s$ is that vertex $v_s$ joins the network and creates $m$ edges to nodes in the set $V(s-1)$. By conditioning on $\mathcal{F}_s$ only the distribution of the degrees of vertices in $V(t)$ to which $v_s$ connects changes compared to conditioning on

---

26However, this technique seems to be limited to the case where the function $f$ is almost linear.

27Hoeffding’s inequality says that if $X_0, X_1, \ldots, X_n$ is a martingale (with respect to some filtration) with $|X_i - X_{i-1}| \leq c$. Then
\[ \mathbb{P}(|X_n - X_0| \geq x) < 2 \exp\left(-\frac{x^2}{2nc^2}\right). \]
These are \( m \) edges, and thus the distribution of at most \( 2m \) vertices is affected. This gives the claimed upper bound.

Now we are able to apply Hoeffding’s inequality. Choose \( x = C\sqrt{\log t} \) and \( C > \sqrt{8m^2} \) to get

\[
\mathbb{P}\left( |N_k(t) - \mathbb{E}[N_k(t)]| > C\sqrt{\log t} \right) < 2t^{-C^2/8m^2} = o(1)
\]

This shows that with probability \( 1 - o(1) \) we have

\[
|P_k(t) - \bar{P}_k(t)| \leq C\sqrt{\log t/n(t)}.
\]

D.2. \( m = 1 \) is Enough

Bollobás et al. [2001] have argued that a model where an entering node creates \( m \) edges can be equivalently defined via a model where each entering node creates a single edge. Since in our algorithms agents create \( m \) edges without intermediate updating of degrees, we can also make a similar identification of the two network formation processes. The NGNG model was indexed by times \( t \in \mathbb{N} \). Now let us introduce artificial time \( N^*_m := \frac{1}{m} \mathbb{N} = \{1/m, 2/m, \ldots\} \), and call \( \tau := 1/m \). An agent \( v_{t+r\tau}, 1 \leq r \leq m \), is called a principal of agent \( v_{t+1} \). The objective of the principal is to create edges. He does this by sampling agents (not principals) relative to his information set \( I(t+1) \subseteq V(t) \). Let \( (X(t+\tau), X(t+2\tau), \ldots, X(t+m\tau)) \) denote the choices made by the principals of agent \( v_{t+1} \). The random variables \( \{X(t+r\tau)\}_{r=1}^m \) are an i.i.d. sequence with common law dictated by the attachment kernels of the respective network formation algorithms. For instance, in the NGNG model the \( r \)-th principal of agent \( v_{t+1} \) samples agent \( j \in V(t) \) with probability

\[
\mathbb{P}(X^{(1)}(t+r\tau) = j|G(t)) = \frac{f(\beta, d_j(t))}{Z(\beta, t)} = K^\beta(t)(j) \quad (1 \leq r \leq m).
\]

At \( t + r\tau = t + 1 \) the degree sequence of the network is updated and we identify the vertex \( v_{t+1} \) with the vertices \( v_{t+r}, v_{t+2r}, \ldots, v_{t+mr} \). Further, identifying \( X^{(m)}(t+1) \) with the outcome generated by the random vector \( (X(t+\tau), X(t+2\tau), \ldots, X(t+m\tau)) \) shows that the sequentially made choices by the principles are equivalent in law to the choice of the agent if he picks simultaneously \( m \) nodes in the network without replacement. In this way we can always obtain NGNG with general \( m \) by running NGNG with \( m = 1 \) by a suitable identification of agents with its principals and updating of the degree sequences only at the integers of the artificial time \( N^*_m \).
D.3. The Existence of a Hub Vertex

In this section we argue that as soon as $\beta > 0$ there is a single hub vertex $v$ in the network to which all other agents are linked to with probability 1. This is a long-run statement where we give the network formation algorithm enough time to work, independent of the stopping time $T_\alpha$. Therefore the statement is not a contradiction to the analytical results we have obtained in the small noise limit. There we have used a linear approximation for small $\beta$, so that the preference for linking is effectively linear. As the network grows, the degrees of the nodes are growing, and so higher-order terms, neglected by the linear approximation, become decisive and predict a different network architecture. On shorter time-scales the linear approximation gives however a quite good understanding of the process. We present our results only for $m = 1$ since the general case is obtained by the collapsing of vertices technique explained in Appendix D.2.

We borrow heavily from the techniques used by Oliveira and Spencer [2005], and provide only a heuristic explanation of their methods. For more details we refer the reader to this nice paper. The purpose of the construction is to write NGNG as a continuous-time birth process on the set of labeled rooted trees, which can be identified with the tree networks generated by NGNG as described below. We will show that this birth process almost surely explodes in finite time, which will imply that there exists a vertex in the graph that has infinitely many children in the tree. This will be the hub vertex in the created star network.

For $m = 1$ NGNG as well as NGNL create randomly growing trees. The first vertex in the tree is called the root node. If vertices $v$ and $w$ are linked then, depending on the date of entry of the nodes, either $v$ is a child of $w$ or $v$ is a parent of $w$. The use of a suitable labeling procedure is a fruitful way to keep track of the evolution of the tree. A labeling is a countable sequence of integers $(a_i)_{i \geq 1}$. The set of all labels is

$$\mathcal{N} := \bigcup_{n \geq 0} \mathbb{Z}_n^+, \quad \mathbb{Z}_0^+ := \emptyset.$$ 

We attach labels to the vertices in the following way:

- The root vertex $v_0$ receives the label $\emptyset$.
- If the vertex $a = (a_1, \ldots, a_n)$ has $k$ children then they receive the labels $ad := (a_1, \ldots, a_n, d), 1 \leq d \leq k$.

In this way we can identify every graph $G$ we may encounter in course of the

---

28Technically, the birth process will have as its jump chain exactly the NGNG-algorithm.
algorithm by a set of labels $A(G) = A$.\footnote{This representation is not necessarily unique, so a selection has to be made.} For a set $A \subset \mathcal{N}$ to be an admissible representation of a tree $G$ it must satisfy the conditions that $\emptyset \in A$ and if $(a_1, \ldots, a_n) \in A$ then also $(a_1, \ldots, a_{n-1}) \in A$.\footnote{This simply means that the root must be identified and every vertex in the tree has a parent.} Let us call the set of finite subsets $A \subset \mathcal{N}$ which satisfy these two conditions as $A^{\text{fin}}$. If the graph is the set of labels $A \in A^{\text{fin}}$ then the degree of vertex $a \in A$ is denoted as $d_A(a)$. To each $a \in \mathcal{N}$ we associate a countable infinite sequence of independent exponentially distributed random variables \( \{X(a,j) \mid a \in \mathcal{N}, j \in \mathbb{Z}_+\} \) with \( X(a,0) \) the time at which vertex $a$ gives birth to its first child. \( X(a,j) \) is the random waiting time until vertex $a$ gives birth to its \((j+1)\)-st child. In this terminology the NGNG process can be written as a Markov chain on $A^{\text{fin}}$ with transition probabilities

\[
P(A, A \cup \{a(d_A(a) + 1)\}) = \frac{f(\beta, d_A(a))}{\sum_{b \in A} f(\beta, d_A(b))}.
\]

This is the probability that vertex $a \in \mathcal{N}$ receives its \((d_A(a) + 1)\)-st child. Now define the birth times of vertices as the random variable $B(a), a \in \mathcal{N}$ such that

- $B(\emptyset) = 0$,
- If $a = (a_1, \ldots, a_n)$ then
  \[
  B(a) = B(a_1, \ldots, a_{n-1}) + \sum_{j=0}^{a_n-1} X((a_1, \ldots, a_{n-1}), j).
  \]

Consider the continuous-time process

\[
W(t) := \{a \in \mathcal{N} \mid B(a) \leq t\}, \ t \in \mathbb{R}_+,
\]

which records the labels of vertices born before some time $t$. Every time a new vertex is born the process $W(\cdot)$ records the label of this vertex and adds it to the set of already active vertices. Hence, birth times have the role of so-called jump times in the continuous time Markov process. Conditional on $\{W(t) = A\}$ for some $A \in A^{\text{fin}}$, the transition probabilities of this process are as follows; $|A|$ nodes give independently from each other birth to a child at time $X(a, d_A(a)), a \in A$. The vertex with the smallest waiting time will be the first one that is observed to give birth. Hence, by the competition theorem of the independent exponential distributions we have that the probability that the first observed transition from $A$ to
$A \cup \{a(d_A(a) + 1)\}$ is exactly Equation (56). The explosion time of vertex $a \in \mathcal{N}$ is the time when it has infinitely many children, i.e.

$$E(a) := \sup_{k \in \mathbb{Z}_+} B(ak) - B(a) = \sum_{j=0}^{\infty} X(a, j).$$

By a standard result on infinite sums of independent exponential random variables [see e.g. Grimmet and Stirzaker, 2001], we know that the expected explosion time is almost surely finite if $\sum_{j \geq 0} \frac{1}{f(\beta, j)} < \infty$. As $f(\beta, j) = e^{\beta j}$ this condition is fulfilled for all $\beta > 0$. Hence, the tree explosion time

$$\tau_\infty := \inf_{a \in \mathcal{N}} (B(a) + E(a))$$

is almost surely finite. Hence, in finite time there exists a vertex that has infinitely many children. Call the progeny of vertex $a \in \mathcal{N}$ the subtree with $a$ as its root. Lemma 5.1 of Oliveira and Spencer [2005] implies that for all $k > 0$ there exists a $\beta(k) > 0$ such that for all $\beta > \beta(k)$ the size of the progeny of any vertex is finite. This is the main step for their proof that almost surely there can be only a single vertex with infinite degree in $W(\tau_\infty)$. A sketch of their argument is as follows. We know from the just mentioned lemma that there are only finitely many vertices that have degree larger than $k$ if we choose $\beta$ sufficiently large. Suppose there does not exist a vertex with infinite degree. König’s infinity lemma states that an infinite rooted tree in which every vertex has finite degree contains an infinite path starting from the root. $W(\tau_\infty)$ is infinite. Any vertex on this path must have an infinite progeny, and there are infinitely many such vertices. A contradiction to the fact that the number of vertices with progeny larger than $k$ is finite. This shows that, independent of the initial condition of the algorithm, almost surely there is at least one vertex with infinite degree, i.e. a hub in the network. They then go on and use properties of the exponential distribution to show that almost surely there is a single vertex with infinite degree. To summarize

**Theorem 5** (Oliveira and Spencer [2005]). For all $\beta > 0$ and $t$ sufficiently large but finite the NGNG algorithm generates almost surely a network consisting of finitely many stars. Asymptotically there is a single star in the network.