Theoretical Economics 9 (2014), 695–752 1555-7561/20140695

Nestedness in networks: A theoretical model and some applications

Michael D. König
Department of Economics, University of Zurich

Claudio J. Tessone
Department of Management, Technology, and Economics, ETH Zurich

Yves Zenou
Department of Economics, Stockholm University and IFN

We develop a dynamic network formation model that can explain the observed nestedness in real-world networks. Links are formed on the basis of agents' centrality and have an exponentially distributed lifetime. We use stochastic stability to identify the networks to which the network formation process converges and find that they are nested split graphs. We completely determine the topological properties of the stochastically stable networks and show that they match features exhibited by real-world networks. Using four different network data sets, we empirically test our model and show that it fits well the observed networks.

Keywords. Nestedness, Bonacich centrality, network formation, nested split graphs.

JEL classification. A14, C63, D85.

1. Introduction

Nestedness is an important aspect of real-world networks.[1] For example, the organization of the New York garment industry (Uzzi 1996) and of the Fedwire bank network

Michael D. König: michael.koenig@econ.uzh.ch
Claudio J. Tessone: tessonec@ethz.ch
Yves Zenou: yves.zenou@ne.su.se

We would like to thank Nicola Persico, two anonymous referees, Phillip Bonacich, Yann Bramoullé, Ulrik Brandes, Guido Caldarelli, Sanjeev Goyal, Patrick Groeber, Matt Jackson, Matteo Marsili, Fernando Vega-Redondo, Douglas White, Eric Gilson, and George Papanicolaou as well as the participants of the 2009 Dynamics of Institutions and Markets in Europe (DIME) conference in Paris, the 2009 Trento Summer School, the 2009 Human Sciences and Complexity (HSC) conference at the University of California–San Diego, the 2009 Workshop on the Economics of Social Networks at Laval University, the Microeconomic Theory Seminar at University of California–Berkeley in 2009, the Theory Workshop at the Kellog School of Management (MEDS) at Northwestern University in 2010, and the 2011 Society for Advanced Economic Theory (SAET) conference in Faro for their helpful comments. We would like to thank the Austrian National Bank for providing access to the data. Michael D. König acknowledges financial support from Swiss National Science Foundation through research Grant PBEZP1–131169, and thanks SIEPR and the Department of Economics at Stanford University for their hospitality during 2010–2012.

[1]A network exhibits *nestedness* if the neighborhood of a node is contained in the neighborhoods of the nodes with higher degrees.

DOI: 10.3982/TE1348

(Soramaki et al. 2007) is nested in the sense that their organization is strongly hierarchical. If we consider, for example, the Fedwire network, it is characterized by a relatively small number of strong flows (many transfers) between banks, with the vast majority of linkages being weak to nonexisting (few to no interbank payment flows). Furthermore, the topology of this network is highly *dissortative* since large banks are disproportionately connected to small banks and vice versa; the average bank was connected to 15 others. In other words, most banks have only a few connections, while a small number of "hubs" have thousands. Åkerman and Larsson (forthcoming), who study the evolution of the global arms trade network using a unique data set on all international transfers of major conventional weapons over the period 1950–2007, also find that these networks are nested and dissortative in the sense that big countries mainly trade arms with small countries, but small countries do not trade with each other. Using aggregate bilateral imports from 1950 to 2000, De Benedictis and Tajoli (2011) analyze the structure of the world trade network over time, detecting and interpreting patterns of trade ties among countries. Figure 3 in their paper shows a clear *core–periphery structure*, indicating nestedness of their networks. Interestingly, in all these networks, dissortativity arises naturally since "big" agents tend to interact with "small" agents and vice versa. For example, banks seek relationships with each other that are mutually beneficial. As a result, small banks interact with large banks for security, lower liquidity risk, and lower servicing costs, and large banks may interact with small banks in part because they can extract a higher premium for services and can accommodate more risk.

Surprisingly, nestedness has not been studied from a theoretical point of view, even though other salient features of networks such as "small world" properties with high clustering and short average path lengths (Watts and Strogatz 1998) as well as "scale-free" or power-law degree distributions (Barabási and Albert 1999) have received a lot of attention.[2]

The first aim of this paper is to propose a dynamic network formation model that exhibits not only the standard features of real-world networks (small worlds, high clustering, short path lengths, and a power-law degree distributions), but also *nestedness* and *dissortativity*. The second aim is to provide a microfoundation for the network formation process where linking decisions are based on the utility maximization of each agent rather than on a random process, which is often assumed in most dynamic models of network formation. The last aim of this paper is to provide some evidence that our model matches well some real-world network features (interbank loans, trade in conventional goods, and arms trade between countries), especially their nestedness.

To be more precise, we develop a dynamic model where, at each period of time, agents play a two-stage game: in the first stage, as in Ballester et al. (2006), agents play their equilibrium contributions proportional to their *Bonacich centrality*,[3] while in the

[2] See Jackson and Rogers (2007), who propose a model that has all these features, but not nestedness.

[3] Centrality is a fundamental measure of the importance of actors in social networks. See Wasserman and Faust (1994) for an introduction and survey. The Bonacich centrality, introduced by Bonacich (1987), of a particular node counts the total number of paths that start from this node in the graph, weighted by a decay factor based on path length.

second stage, a randomly chosen agent can update her linking strategy by creating a new link as a best response to the current network. Links do not last forever, but have an exponentially distributed lifetime. The most valuable links (i.e., the ones with the highest Bonacich centrality) decay at a lower rate than those that are less valuable. As a result, the formation of social networks can be regarded as a tension between the search for new linking opportunities and the volatility that leads to the decay of existing links.

We introduce noise into the decision process to form links (see, e.g., Sandholm 2010), and analyze the limit of the invariant distribution, the *stochastically stable* networks, as the noise vanishes to zero.[4] We first show that in this limit, starting from arbitrary initial conditions, at each period of time, the network generated by this dynamic process is a *nested split graph*. These graphs, which are relatively well known in the applied mathematics literature (Cvetković and Rowlinson 1990, Mahadev and Peled 1995), have a very nice and simple structure that make them very tractable. To the best of our knowledge, this is the first time that a complex dynamic network formation model can be characterized by such a simple structure in terms of networks it generates. By doing so, we are able to bridge the economics literature and the applied mathematics/physics literatures in a simple way. Because of their simple features, we then show that degree, closeness, eigenvector, and Bonacich centrality induce the same ordering of nodes in a nested split graph (this is also true for betweenness centrality if the ordering is not strict). This implies, in particular, that if we had a game where agents formed links according to measures of centrality (such as degree, closeness, or betweenness) other than the Bonacich centrality, then all our results would be unchanged. We then show that the stochastically stable network is a nested split graph. Instead of relying on a mean-field approximation of the degree distribution as most dynamic network formation models do, because of the nature of nested split graphs, we are able to derive explicit solutions for all network statistics of the stochastically stable networks (by computing the adjacency matrix).[5] We also find that by altering the rate at which linking opportunities arrive and links decay, a sharp transition takes place in the network density. This transition entails a crossover from highly centralized networks when the linking opportunities are rare and the link decay is high to highly decentralized networks when many linking opportunities arrive and only few links are removed.

The intuition of these results is as follows. Agents want to link to other agents who are more central since this leads to higher efforts (as efforts are proportional to centrality) and higher efforts raise payoffs. Similarly, links to agents with lower centrality last

[4]In the literature on coordination games (see, e.g., Kandori et al. 1993), the noise is introduced as an equilibrium selection device, when, in the absence of noise, multiple equilibria can emerge. This is not the case here since we have a unique steady-state equilibrium even when the noise tends to zero. Introducing some noise allows us to better calibrate our model to the data since, when the noise goes to zero, the diameter of the steady-state network is equal to 2, a feature that is not always observed in the data. In other words, the noise allows us to have some flexibility with the model so that it can be calibrated to empirically observed networks. This is what is done in Figure 11, where we show that our model matches well various features of four real-world networks. Also, in Table 1, our estimates of the model's parameters indicate that the level of noise does not vanish.

[5]In a nested split graph, the degree distribution uniquely defines the adjacency matrix (up to a permutation of the node labels).

shorter. Notice, moreover, that once someone loses a link with an agent, she becomes less central and this makes it more likely that the next link she has will also disappear. Thus link gains and losses are self-reinforcing. This intuition suggests that if α, the probability of adding links, is large, then the process should approximate a complete network, while if it is small, then the process should approximate the star network. The key insight of our model is that *for intermediate values of α, the stochastically stable network is a nested split graph*.

We then proceed by showing that our model reproduces some empirical observations of real-world networks. We show that the stochastically stable networks emerging in our link formation process are characterized by *short path length* with *high clustering*, *exponential degree distributions* with *power-law tails*, *negative degree-clustering correlation*, and *nestedness*. These networks also show a clear core–periphery structure. Moreover, we show that stochastically stable networks are *dissortative*.

Using four different data sources, we empirically test our model. We analyze the network of Austrian banks, the global banking network, the trade network (import–export relationships between countries), and the network of arms trade between countries. Despite the fact that these networks are very different, they all exhibit strong nestedness and dissortativity, and we find a reasonable goodness of fit of our model with these networks (even though it is only parsimoniously parameterized).

Our paper is organized as follows. Section 2 discussed the relation of our model to the literature. In Section 3, we introduce the model and discuss the basic properties of the network formation process. Next, Section 4 shows that stochastically stable networks exist, can be computed analytically, and are nested split graphs. After deriving the stochastically stable networks in Section 5, we analyze their properties in terms of topology and centralization. Using four different network data sets, we empirically test our model in Section 6. All proofs can be found in Appendix A. Appendix B gives all the necessary definitions and characterizations of networks used throughout the paper. In Appendix C, we provide some general results for nested split graphs in terms of their topology properties and centralization measures. For the purpose of motivating the empirical test of our model with the four data sets mentioned above, we provide an interpretation of our theoretical model in terms of networks of banks and trade networks in Appendix D. Moreover, we extend our analysis in Appendix E by including linking costs.

2. Relation to the literature

The literature on network formation is basically divided into two strands that are not communicating very much with each other. In the random network approach (mainly developed by mathematicians and physicists), which is mainly dynamic, the reason why a link is formed is pure chance. In this approach, researchers study how emerging networks match real-world networks (see, e.g., Vega-Redondo 2007). While sharing some common features with this literature, our model is quite different, since agents do not create links randomly but in a strategic way, i.e., they maximize their utility function.

In the other approach (developed by economists; see, in particular, Jackson and Wolinsky 1996), the reason for the formation of a link is strategic interactions. Individuals carefully decide with whom to interact and this decision entails some consent by both parties in a given relationship. There are some dynamic network formation models with strategic interactions. Bala and Goyal (2000), Watts (2001), Jackson and Watts (2002a), and Dutta et al. (2005) are prominent papers of this literature. Our model is different than the ones developed in these papers in the sense that we consider both dynamic models of network formation and optimal actions from agents. This allows us to give a microfoundation of the network formation process as equilibrium actions transform into equilibrium utility functions. Another crucial difference is that we are able to match most features of real-world networks while these models do not.[6]

There is also another strand of the literature (called *games on networks*) that takes the network as given and studies how the network structure impacts outcomes and individual decisions.[7] A prominent paper of this literature is Ballester et al. (2006).[8] They mainly show that if agents' payoffs are linear-quadratic, then the unique interior Nash equilibrium of an n-player game in which agents are embedded in a network is such that each individual effort and outcome is proportional to her Bonacich centrality measure. In the present paper, we introduce *strategic interactions* in a nonrandom dynamic network formation game where agents also choose how much effort they put into their activities.

There are some papers that, as in our framework, combine both network formation and endogenous efforts. These papers include Bramoullé et al. (2004), Cabrales et al. (2011), Calvó-Armengol and Zenou (2004), Galeotti and Goyal (2010), Goyal and Vega-Redondo (2005), Goyal and Joshi (2003), and Jackson and Watts (2002a). Most of these models are, however, static and the network formation process is different.

Our paper is also related to Jackson and Rogers (2007), who also motivate their modeling approach by means of statistics of empirical networks.

Finally, the paper by König and Tessone (2011) shows that our model can be applied not only to an economic context, but also to a variety of models studied in the physics literature, ranging from the analysis of ecological systems to physical synchronization processes being coupled to network dynamics. They extend our model by introducing heterogeneous selection probabilities of the nodes depending on the number of links they already have, derive the dynamics of the degree distribution in the continuous limit, and analyze its properties. They show that the stationary degree distribution is given by a double power law with a flexible exponent. It has to be clear, however, that

[6]Mele (2010) and Liu et al. (2012) provide interesting dynamic network formation models where individuals decide with whom to form links by maximizing a utility function. However, contrary to our model, these papers do not characterize analytically the degree distribution and the resulting network statistics.

[7]See Jackson and Zenou (2014), for a recent overview of this literature.

[8]Bramoullé and Kranton (2007), Bramoullé et al. (2014), and Galeotti et al. (2010) are also important papers in this literature. The first paper focuses on strategic substitutabilities, while the second one provides a general framework for solving any game on networks with perfect information and linear best-reply functions. The last paper investigates the case when agents do not have perfect information about the network. Because of its tractability, in the present paper, we use the model of Ballester et al. (2006), who analyze a network game of local complementarities under perfect information.

the paper by König and Tessone (2011) is just an extension of our framework that analyzes the nature of the phase transition from sparse to dense networks in the continuous limit. It was written after our paper,[9] and our main result that characterizes the steady-state networks as nested split graphs is only proved in the current paper and then used by König and Tessone (2011).

To summarize, our main contribution to the literature is that we are able to explain the emergence of nestedness in networks by analyzing a dynamic network formation model with endogenous actions. We are also able to analytically characterize the stochastically stable networks, which can be shown to be nested split graphs, and to provide a microfoundation for the link formation process. Even if nested split graphs have a much more regular structure than the complex networks we observe in many real-world applications, they are easy to study, they are the result of endogenous rational actions, and they have most of the properties of real-world networks. Finally, we empirically test our model with four different data sets and show that our model fits these observed networks well.

3. The model

In this section, we introduce the network formation process, which can be viewed as a two-stage game on two separate time scales. On the fast time scale, all agents simultaneously choose their effort level in a fixed network structure. It is a game following Ballester et al. (2006) with local complementarities where players have linear-quadratic payoff functions. On the slow time scale, agents receive linking opportunities at a given rate and decide with whom they want to form a link, while the links they have created decay after having reached their finite lifetime. This introduces two different time scales, one in which agents are choosing their efforts in a simultaneous move game, and the second in which an agent forms a link and anticipates the equilibrium outcome in the following simultaneous move game.

3.1 *Nash equilibrium and Bonacich centrality*

Consider a static network G in which the nodes represent a set $\mathcal{N} = \{1, 2, \dots, n\}$ of agents/players. Following Ballester et al. (2006), each agent $i \in \mathcal{N}$ in the network G selects an effort level $x_i \geq 0$, $\mathbf{x} \in \mathbb{R}_+^n$. Denote by Ω the countable state space of all networks with n nodes. Then each agent i receives a payoff $\pi_i : \mathbb{R}_+^n \times \Omega \times \mathbb{R}_+ \to \mathbb{R}$ of the form

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

where $\lambda \geq 0$ and $a_{ij} \in \{0, 1\}$, $i, j = 1, \dots, n$, are the elements of the symmetric $n \times n$ *adjacency matrix* $\mathbf{A}$ of G. This utility function is additively separable in the idiosyncratic effort component $(x_i - \frac{1}{2}x_i^2)$ and the peer effect contribution $(\lambda \sum_{j=1}^{n} a_{ij} x_i x_j)$. Payoffs display strategic complementarities in effort levels, i.e., $\partial^2 \pi_i(\mathbf{x}, G, \lambda)/\partial x_i \partial x_j = \lambda a_{ij} \geq 0$.

[9] See our working paper (König et al. 2009).

The general payoff structure in (1) has a variety of applications. For example, (1) can be interpreted as the profit function of a bank competing in quantities of lending à la Cournot with other banks in a loan market where different types of loans cannot be substituted. Equation (1) can also be interpreted as the payoff function of a firm in country i acting as a local monopolist supplying a nonsubstitutable good. In both cases, interdependencies induce a reduction in marginal costs of production due to technology spillovers and learning by doing effects. Appendix D provides a more detailed explanation of these two examples.

So as to find the Nash equilibrium solution associated with the above payoff function (1), we define a network centrality measure introduced by Bonacich (1987). Let $\lambda_{\text{PF}}(G)$ be the largest real eigenvalue of the adjacency matrix $\mathbf{A}$ of network G. The adjacency matrix is a matrix that lists the direct connections in the network. If $\mathbf{I}$ denotes the $(n \times n)$ identity matrix and $\mathbf{u} \equiv (1, \dots, 1)^\top$ denotes the n-dimensional vector of 1s, then we can define *Bonacich centrality* as follows.

Definition 1. If and only if $\lambda < 1/\lambda_{\text{PF}}(G)$, then the matrix $\mathbf{B}(G, \lambda) \equiv (\mathbf{I} - \lambda\mathbf{A})^{-1} = \sum_{k=0}^{\infty} \lambda^k \mathbf{A}^k$ exists, is nonnegative, and the vector of Bonacich centralities is defined as

$$\mathbf{b}(G, \lambda) \equiv \mathbf{B}(G, \lambda) \cdot \mathbf{u}.$$

We can write the vector of Bonacich centralities as $\mathbf{b}(G, \lambda) = \sum_{k=0}^{\infty} \lambda^k \mathbf{A}^k \cdot \mathbf{u} = (\mathbf{I} - \lambda\mathbf{A})^{-1} \cdot \mathbf{u}$. For the components $b_i(G, \lambda)$, $i = 1, \dots, n$, we get

$$b_i(G, \lambda) = \sum_{k=0}^{\infty} \lambda^k (\mathbf{A}^k \cdot \mathbf{u})_i = \sum_{k=0}^{\infty} \lambda^k \sum_{j=1}^{n} (\mathbf{A}^k)_{ij},$$

where $(\mathbf{A}^k)_{ij}$ is the ijth entry of $\mathbf{A}^k$.

Now we can turn to the equilibrium analysis of the game.

Theorem 1 (Ballester et al. 2006). *Consider the n-player simultaneous move game with payoffs given by (1) and strategy space $\mathbb{R}^n_+$. If $\lambda < 1/\lambda_{\text{PF}}(G)$, there exists a unique interior Nash equilibrium, which, for each agent $i = 1, \dots, n$, is given by*

$$x_i^* = b_i(G, \lambda).$$

Moreover, the equilibrium payoff of each agent i is given by

$$\pi_i^*(G, \lambda) = \pi_i(\mathbf{x}^*, G, \lambda) = \tfrac{1}{2}(x_i^*)^2 = \tfrac{1}{2} b_i^2(G, \lambda).$$

Observe that the condition $\lambda < 1/\lambda_{\text{PF}}(G)$ is an endogenous object. Below, we will consider a dynamic network formation model where this condition has to hold at each period of time.[10]

[10] For this condition not to depend on an endogenous variable (i.e., $\lambda_{\text{PF}}(G)$ varies with the evolution of the network), we can use the sufficient condition $\lambda < 1/\sqrt{2m(n-1)/n}$, where m is an upper bound on the number of links in G. See Cvetković and Rowlinson (1990) for various other bounds on the largest eigenvalue $\lambda_{\text{PF}}(G)$.

Furthermore, Ballester et al. (2006) have shown that the equilibrium outcome and the payoff for each player increases with the number of links in G (because the number of network walks increases in this way). This implies that if an agent is given the opportunity to change her links, she will add as many links as possible. On the other hand, if she is only allowed to form one link at a time, she will form the link to the agent that increases her payoff the most. In both cases, eventually, the network will then become complete, i.e., each agent is connected to every other agent. However, to avoid this latter unrealistic situation, we assume that the agents are living in a volatile environment that causes links to decay such that the complete network can never be reached. Instead the architecture of the network adapts to the volatile environment. We will treat these issues more formally in the next section.

3.2 *The network formation process*

We now introduce a network formation process that incorporates the idea that agents with high Bonacich centrality (their equilibrium effort levels) are more likely to connect to each other, while the links they have established between each other have a longer lifetime if they are viewed as more valuable to them.

We consider a continuous time Markov chain $(G(t))_{t\in\mathbb{R}_+}$ with $G(t) = (\mathcal{N}, \mathcal{E}(t))$ comprising the set of agents $\mathcal{N} = \{1, \ldots, n\}$ together with the set of edges/links $\mathcal{E}(t) \subset \mathcal{N} \times \mathcal{N}$ at time t between them. $(G(t))_{t\in\mathbb{R}_+}$ is a collection of random variables $G(t)$, indexed by time $t \in \mathbb{R}_+$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, where Ω is the countable state space of all networks with n nodes, $\mathcal{F}$ is the σ-algebra $\boldsymbol{\sigma}(\{G(t) : t \in \mathbb{R}_+\})$ generated by the collection of $G(t)$, and $\mathbb{P} : \mathcal{F} \to [0, 1]$ is a countably additive, nonnegative measure on $(\Omega, \mathcal{F})$ with total mass $\sum_{G\in\Omega} \mathbb{P}(G) = 1$. At every time $t \geq 0$, links can be created or decay with specified rates that depend on the current network $G(t) \in \Omega$.

Definition 2. Consider a continuous time Markov chain $(G(t))_{t\in\mathbb{R}_+}$ on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\boldsymbol{\pi}^*(G(t), \lambda) \equiv (\pi_1^*(G(t)), \ldots, \pi_n^*(G(t)))$ denote the vector of Nash equilibrium payoffs of the agents in $G(t)$ derived from the payoff function (1) with parameter $0 \leq \lambda < 1/\lambda_{\mathrm{PF}}(G(t))$.

(i) At rate $\alpha_i \in (0, 1)$, link creation opportunities arrive to each agent $i \in \mathcal{N}$. If such an opportunity arrives, then agent i computes the marginal payoff $\pi_i^*(G(t) \oplus (i, j), \lambda)$ for each agent $j \notin \mathcal{N} \setminus (\mathcal{N}_i \cup \{i\})$ she is not already connected to, where this computation includes an additive, exogenous stochastic term ε_{ij}, incorporating possible mistakes in the computation of the agent. We assume that the exogenous random terms ε_{ij} are identically and independently type I extreme value distributed (or Gumbel distributed) with scaling parameter ζ.[11] Given that agent $i \in \mathcal{N}$

[11]For the distribution of the error term, it holds that $\mathbb{P}(\varepsilon_{ij} \leq c) = e^{-e^{c/\zeta - \gamma}}$, where $\gamma \approx 0.58$ is Euler's constant. The expectation is $\mathbb{E}(\varepsilon_{ij}) = 0$ and the variance is given by $\mathrm{Var}(\varepsilon_{ij}) = \pi^2\zeta^2/6$.

receives a link creation opportunity, she then links to agent $j \in \mathcal{N} \setminus (\mathcal{N}_i \cup \{i\})$ with probability

$$b_i^{\zeta}(j|G(t)) \equiv \mathbb{P}\Big(\pi_i^*(G(t) \oplus (i,j), \lambda) + \varepsilon_{ij} = \max_{k \in \mathcal{N}\setminus(\mathcal{N}_i \cup \{i\})} \pi_i^*(G(t) \oplus (i,k), \lambda) + \varepsilon_{ik}\Big)$$
$$= \frac{e^{\pi_i^*(G(t)\oplus(i,j),\lambda)/\zeta}}{\sum_{k\in\mathcal{N}\setminus(\mathcal{N}_i\cup\{i\})} e^{\pi_i^*(G(t)\oplus(i,k),\lambda)/\zeta}}.$$

It follows that the probability that, during a small time interval $[t, t+\Delta t)$, a transition takes place from $G(t)$ to $G(t) \oplus (i,j)$ is given by $\mathbb{P}(G(t+\Delta t) = G \oplus (i,j)| G(t) = G) = \alpha_i b_i^{\zeta}(j|G(t))\Delta t + o(\Delta t)$.[12]

(ii) We assume that a link (i,j), once established, has an exponentially distributed lifetime $\tau_{ij} \in \mathbb{R}_+$ with parameter $\nu_{ij}^{\zeta}(G(t)) \equiv 1/\mathbb{E}(\tau_{ij}|G(t)) = \beta_i f_{ij}(G(t))$, including an agent-specific component $\beta_i \in (0,1)$ and a link-specific component

$$f_{ij}^{\zeta}(G(t)) \equiv \frac{e^{\pi_i^*(G(t)\ominus(i,j),\lambda)/\zeta}}{\sum_{k\in\mathcal{N}_i} e^{\pi_i^*(G(t)\ominus(i,k),\lambda)/\zeta}}$$

for any $i \in \mathcal{N}$ and $j \in \mathcal{N}_i$. The probability that, during a small time interval $[t, t+\Delta t)$, a transition takes place from $G(t)$ to $G(t) \ominus (i,j)$ is given by $\mathbb{P}(G(t+\Delta t) = G \ominus (i,j)|G(t) = G) = \beta_i f_{ij}^{\zeta}(G(t))\Delta t + o(\Delta t)$.

Transitions to networks that differ by more than one link have probability $o(\Delta t)$.

In words, if agent i is chosen to form a link (at rate α_i), she will choose the agent that increases her utility the most. There is, however, a possibility of error, captured by the stochastic term in the profit function. Furthermore, it is assumed that links do not last forever, but have an exponentially distributed lifetime with an expectation that depends on the relative payoff loss from removing that link. The specific functional form of the pairwise component $f_{ij}^{\zeta}(\cdot)$ in the expected lifetime of a link incorporates the fact that links that are more valuable to an agent (i.e., the ones with the highest Bonacich centrality) live longer than the ones that are viewed as less valuable to her. The value of a link is measured by the perceived loss in payoff incurred by the agent from removing the link.[13,14]

[12] $f(t) = o(g(t))$ as $t \to \infty$ if $\lim_{t\to\infty} f(t)/g(t) = 0$.

[13] In a similar way, Staudigl (2011) assumes that the linking activity levels of agents depend on their relative marginal payoffs. Snijders (2001) and Snijders et al. (2010) introduced exponential link update rates, which "depend on actor-specific covariates or on network statistics expressing the degree to which the actor is satisfied with the present network structure." See also (3.4) in Staudigl (2011) and Section 7.1 in Snijders (2001).

[14] The fact that links do not last forever is a quite natural feature of real-world networks. For example, in the context of interfirm alliances, Hagedoorn (2002) for research partnerships, Kogut et al. (2007) for joint ventures, Harrigan (1988) for alliances, and Park and Russo (1996) for (equity-based) joint ventures provide empirical evidence on this phenomenon. For example, Harrigan (1988) studies 895 alliances from 1924 to 1985 and concludes that the average lifespan of the alliance is relatively short, 3.5 years, with a standard deviation of 5.8 years and that 85% of these alliances last less than 10 years. Park and Russo (1996) focus

It should be clear that when a new link may be added to the network, then that link proposal will always be accepted by the receiver. This is because it always increases the utility of the receiver due to local complementarities in the utility function. In fact, we will show below that it will also be the best reply for the receiver (i.e., the best alternative in terms of link formation).

Observe that when agents decide to create a link, they do it in a *myopic* way, that is, they only look at the agents that give them the *current* highest payoff. There is literature on farsighted networks where agents calculate their lifetime expected utility when they want to create a link. We adopt a myopic approach here because of its tractability and because our model also incorporates effort decision.[15]

We now discuss the networks generated by our model for large times t and how they depend on the error term parameterized by ζ. For this purpose, observe that the Markov chain $(G(t))_{t\in\mathbb{R}_+}$ can be described infinitesimally in time by the *generator matrix* $\mathbf{Q}^\zeta$ with elements given by the *transition rates* $q^\zeta:\Omega\times\Omega\to\mathbb{R}$ defined by $\mathbb{P}(G(t+\Delta t)=G'|G(t)=G)=q^\zeta(G,G')\Delta t+o(\Delta t)$ for $G\neq G'$ and $\mathbb{P}(G(t+\Delta t)=G|G(t)=G)=1+q^\zeta(G,G)\Delta t+o(\Delta t)$ in the limit of $\Delta t\downarrow 0$. Consequently, $q^\zeta(G,G\oplus(i,j))=\alpha_i b_i^\zeta(j|G)$ and $q^\zeta(G,G\ominus(i,j))=\beta_i f_{ij}^\zeta(G)$. The transition rates have the property that $q^\zeta(G,G')=q^\zeta(G,G\pm(i,j))\geq 0$ if G' differs from G by the link (i,j) and that $q^\zeta(G,G')=0$ if G' differs from G by more than one link. Moreover, it must hold that $\sum_{G'\in\Omega}q^\zeta(G,G')=0$, and one can show that $\mathbb{P}(G(t)=G'|G(0)=G)=e^{\mathbf{Q}^\zeta t}$. If a nonnegative solution to $\mu^\zeta\mathbf{Q}^\zeta=0$ with $\sum_{G\in\Omega}\mu^\zeta(G)=1$ exists, then μ^ζ is the stationary distribution of the Markov chain satisfying $\mu^\zeta(G')=\lim_{t\to\infty}\mathbb{P}(G(t)=G'|G(0)=G)$ (see, e.g., Liggett 2010).

The simplest case is the one where ζ diverges, the error term ε_{ij} becomes dominant, and the link formation and decay rates are payoff independent. The link creation and decay rates for any $i\in\mathcal{N}$ are then given by

$$\lambda_i\equiv\lim_{\zeta\to\infty}q^\zeta(G,G\oplus(i,j))=\alpha_i\frac{1}{|\mathcal{N}\setminus(\mathcal{N}_i\cup\{i\})|},\quad j\in\mathcal{N}\setminus(\mathcal{N}_i\cup\{i\})$$
$$\mu_i\equiv\lim_{\zeta\to\infty}q^\zeta(G,G\ominus(i,j))=\beta_i\frac{1}{|\mathcal{N}_i|},\quad j\in\mathcal{N}_i.$$

These transition rates correspond to a birth–death Markov chain with birth rates λ_i and death rates μ_i (see, e.g., Liggett 2010, Chapter 2.7.1), and the stationary degree distribution is that of the corresponding birth–death chain. In the special case of $\alpha_i=\beta_i=\frac{1}{2}$ for all $i\in\mathcal{N}$, we obtain a Poisson degree distribution corresponding to a random graph $G(n,p)$ with an independent link probability $p=\frac{1}{2}$.

A more interesting case, from a behavioral and topological point of view, is the one where ζ converges to zero and the error term ε_{ij} vanishes. For each agent $i\in\mathcal{N}$, let the best response be the set-valued map $\mathcal{B}_i:\Omega\to\mathcal{N}$ defined as

on 204 joint ventures among firms in the electronics industry for the period 1979–1988. They show that less than half of these firms remain active beyond a period of 5 years and for those that last less than 10 years ($\frac{2}{3}$ of the total), the average lifetime turns out to be 3.9 years.

[15] Jackson and Watts (2002b) argue that this form of myopic behavior makes sense if players heavily discount the future.

$\mathcal{B}_i(G) \equiv \arg\max_{k\in\mathcal{N}\setminus(\mathcal{N}_i\cup\{i\})} \pi_i^*(G \oplus (i,k), \lambda)$; similarly, we define the map $\mathcal{M}_i : \Omega \to \mathcal{N}$ as $\mathcal{M}_i(G) \equiv \arg\max_{k\in\mathcal{N}_i} \pi_i^*(G \ominus (i,k), \lambda)$. In the limit $\zeta \to 0$, we then have that the link creation and decay rates for any $i \in \mathcal{N}$ are given by

$$
\begin{aligned}
q(G, G\oplus(i,j)) &\equiv \lim_{\zeta\to 0} q^{\zeta}(G, G\oplus(i,j)) = \alpha_i \frac{1}{|\mathcal{B}_i(G)|}, \quad j \in \mathcal{B}_i(G) \\
q(G, G\ominus(i,j)) &\equiv \lim_{\zeta\to 0} q^{\zeta}(G, G\ominus(i,j)) = \beta_i \frac{1}{|\mathcal{M}_i(G)|}, \quad j \in \mathcal{M}_i(G).
\end{aligned}
\tag{2}
$$

We call a network $G \in \Omega$ *stochastically stable* if $\mu(G) > 0$, where $\mu \equiv \mu^0$ is the stationary distribution of the Markov chain with transition rates given in (2).[16] The set of stochastically stable networks is denoted by $\hat{\Omega} \equiv \{G \in \Omega : \mu(G) > 0\}$. We will analyze these states in Section 4, while we will study the sample paths generated by the chain when ζ is zero in the next section. We refer to this case ($\zeta = 0$) as the *unperturbed dynamics*, while the case of noise ($\zeta > 0$) is referred to as *perturbed dynamics*.

3.3 *Network formation and nested split graphs*

In this section, we will focus on the unperturbed dynamics of the Markov chain introduced in Definition 2. An essential property of the chain is that it produces networks in a well defined class of graphs denoted *nested split graphs* (Cvetković and Rowlinson 1990).[17] We will give a formal definition of these graphs and discuss an example in this section. Nested split graphs include many common networks such as the star network. Moreover, as their name already indicates, they have a *nested neighborhood structure*. This means that the set of neighbors of each agent is contained in the set of neighbors of each higher degree agent. Nested split graphs have particular topological properties and an associated adjacency matrix with a well defined structure.

So as to characterize nested split graphs, it will be necessary to consider the degree partition of a graph, which is defined as follows.

Definition 3 (Mahadev and Peled 1995). Let $G = (\mathcal{N}, \mathcal{E})$ be a graph whose distinct positive degrees are $d_{(1)} < d_{(2)} < \cdots < d_{(k)}$ and let $d_0 = 0$ (even if no agent with degree 0 exists in G). Further, define $\mathcal{D}_i = \{v \in \mathcal{N} : d_v = d_{(i)}\}$ for $i = 0, \ldots, k$. Then the set-valued vector $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ is called the degree partition of G.

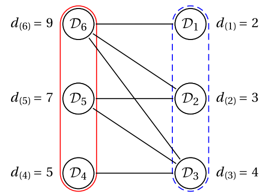
With the definition of a degree partition, we can now give a more formal definition of a nested split graph.[18,19]

[16] See also Young (2001, Chapter 3) and Sandholm (2010, Chapter 12).

[17] Nested split graphs are also called threshold networks (Mahadev and Peled 1995, Hagberg et al. 2006).

[18] Let x be a real-valued number $x \in \mathbb{R}$. Then $\lceil x \rceil$ denotes the smallest integer larger than or equal to x (the ceiling of x). Similarly, $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x (the floor of x).

[19] In general, split graphs are graphs whose nodes can be partitioned into a set of nodes that are all connected among each other and sets of nodes that are disconnected. A nested split graph is a special case of a split graph.



$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Figure 1. Representation of a connected nested split graph (left) and the associated adjacency matrix (right) with $n = 10$ agents and $k = 6$ distinct positive degrees. A line between $\mathcal{D}_i$ and $\mathcal{D}_j$ indicates that every node in $\mathcal{D}_i$ is linked to every node in $\mathcal{D}_j$. The solid frame indicates the dominating set, and the nodes in the independent sets are included in the dashed frame. Next to the set $\mathcal{D}_i$, the degree of the nodes in the set is indicated. The neighborhoods are nested such that the degrees are given by $d_{(i+1)} = d_{(i)} + |\mathcal{D}_{k-i+1}|$ for $i \neq \lfloor k/2 \rfloor$ and $d_{(i+1)} = d_{(i)} + |\mathcal{D}_{k-i+1}| - 1$ for $i = \lfloor k/2 \rfloor$. In the corresponding adjacency matrix $\mathbf{A}$ (on the right), the 0 entries are separated from the 1 entries by a step function.

Definition 4 (Mahadev and Peled 1995). Consider a nested split graph $G = (\mathcal{N}, \mathcal{E})$ and let $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ be its degree partition. Then the nodes $\mathcal{N}$ can be partitioned in independent sets $\mathcal{D}_i$, $i = 1, \ldots, \lfloor k/2 \rfloor$, and a dominating set $\bigcup_{i=\lfloor k/2 \rfloor+1}^{k} \mathcal{D}_i$ in the graph $G' = (\mathcal{N} \setminus \mathcal{D}_0, \mathcal{E})$. Moreover, the neighborhoods of the nodes are nested. In particular, for each node $v \in \mathcal{D}_i$, $i = 1, \ldots, k$,

$$\mathcal{N}_v = \begin{cases} \displaystyle\bigcup_{j=1}^{i} \mathcal{D}_{k+1-j} & \text{if } i = 1, \ldots, \left\lfloor \dfrac{k}{2} \right\rfloor \\ \displaystyle\bigcup_{j=1}^{i} \mathcal{D}_{k+1-j} \Big\backslash \{v\} & \text{if } i = \left\lfloor \dfrac{k}{2} \right\rfloor + 1, \ldots, k. \end{cases}$$

Figure 1 (left) illustrates the degree partition $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_6)$ and the nested neighborhood structure of a nested split graph. A line between $\mathcal{D}_i$ and $\mathcal{D}_j$ indicates that every node in $\mathcal{D}_i$ is linked to every node in $\mathcal{D}_j$ for any $i, j = 1, \ldots, 6$. The nodes in the dominating set included in the solid frame induce a clique while the nodes in the independent sets that are included in the dashed frame induce an empty subgraph. In the following discussion, we will call the sets $\mathcal{D}_i$, $i = \lfloor k/2 \rfloor + 1, \ldots, k$, *dominating subsets*, since the set $\mathcal{D}_i$ induces a dominating set in the graph obtained by removing the nodes in the set $\bigcup_{j=0}^{k-i} \mathcal{D}_j$ from G.

A nested split graph has an associated adjacency matrix that is called a *stepwise matrix* and it is defined as follows.

Definition 5 (Brualdi and Hoffman 1985). A stepwise matrix $\mathbf{A}$ is a symmetric, binary $(n \times n)$ matrix with elements a_{ij} satisfying the following condition: if $i < j$ and $a_{ij} = 1$, then $a_{hk} = 1$ whenever $h < k \leq j$ and $h \leq i$.

Figure 1 (right) shows the stepwise adjacency matrix $\mathbf{A}$ corresponding to the nested split graph shown on the left hand side. If we let the nodes be indexed by the order of the rows in the adjacency matrix $\mathbf{A}$, then it is easily seen that, for example, $\mathcal{D}_6 = \{1, 2 \in \mathcal{N} : d_1 = d_2 = d_{(6)} = 9\}$ and $\mathcal{D}_1 = \{9, 10 \in \mathcal{N} : d_9 = d_{10} = d_{(1)} = 2\}$.

If a nested split graph is connected, we call it a connected nested split graph. The representation and the adjacency matrix depicted in Figure 1 actually show a connected nested split graph. From the stepwise property of the adjacency matrix, it follows that a connected nested split graph contains at least one spanning star, that is, there is at least one agent that is connected to all other agents. In Appendix C, we also derive the clustering coefficient, the neighbor connectivity, and the characteristic path length of a nested split graph. In particular, we show that connected nested split graphs have small characteristic path length, which is at most 2. We also analyze different measures of centrality (see Wasserman and Faust 1994, Chapter 5.2) in a nested split graph. One important result is that degree, closeness, and Bonacich centrality induce the same ordering of nodes in a nested split graph. If the ordering is not strict, then this holds also for betweenness centrality (see Appendix C.2.5).

In the next proposition, we identify the relationship between the Bonacich centrality of an agent and her degree in a nested split graph. Denote by $G \oplus (i, j)$ the network G for which a link between i and j has been added, and denote by $G \ominus (i, j)$ the network G for which the link between i and j has been deleted.

Proposition 1. *Consider a pair of agents $i, j \in \mathcal{N}$ of a nested split graph $G = (\mathcal{N}, \mathcal{E})$.*

(i) If and only if agent i has a higher degree than agent j, then i has a higher Bonacich centrality than j, i.e., $d_i > d_j \Leftrightarrow b_i(G, \lambda) > b_j(G, \lambda)$.

(ii) Assume that neither the links (i, k) nor (i, j) are in G, $(i, j) \notin \mathcal{E}$ and $(i, k) \notin \mathcal{E}$. Further assume that agent k has a higher degree than agent j, $d_k > d_j$. Then adding the link (i, k) to G increases the Bonacich centrality of agent i more than adding the link (i, j) to G, i.e., $d_k > d_j \Leftrightarrow b_i(G \oplus (i, k), \lambda) > b_i(G \oplus (i, j), \lambda)$.

(iii) Consider two agents $j, k \in \mathcal{N}_i$ and assume that agent k has a higher degree than agent j, $d_k > d_j$. Then removing the link (i, k) from G decreases the Bonacich centrality of agent i more than removing the link (i, j) from G, i.e., $d_k > d_j \Leftrightarrow b_i(G \ominus (i, k), \lambda) < b_i(G \ominus (i, j), \lambda)$.

From part (ii) of Proposition 1, we find that when agent i has to decide to create a link to either agent k or j, with $d_k > d_j$, in the link formation process $(G(t))_{t\in\mathbb{R}_+}$, then i will always connect to agent k because this link gives i a higher Bonacich centrality than the other link to agent j. A similar argument holds for the removal of a link in part (iii). We can make use of this property to show that the networks emerging from the link formation process defined in the previous section actually are nested split graphs. This result is stated in the next proposition.

Proposition 2. *Consider the unperturbed dynamics of the network formation process $(G(t))_{t\in\mathbb{R}_+}$ introduced in Definition 2. Assume that at $t = 0$, we start with the empty network $G(0) = \bar{K}_n$. Then, at any time $t \geq 0$, the network $G(t)$ is a nested split graph almost surely and the set $\Psi \in \Omega$, consisting of all possible unlabeled nested split graphs on n nodes with $|\Psi| = 2^{n-1}$, has measure $\mathbb{P}(\Psi) = 1$.*

This result is due to the fact that agents, when they have the possibility of creating a new link, always connect to the agent who has the highest Bonacich centrality (and by Proposition 1 the highest degree). This creates a nested neighborhood structure that can always be represented by a stepwise adjacency matrix after a possible relabeling of the agents.[20] The same applies for link decay.

Let us give some more intuition of this crucial result. Agents want to link to others who are more central since this leads to higher actions (as actions are proportional to centrality) and higher actions raise payoffs more. Similarly, links decay to those with lower centrality as these agents have lower actions and hence lower payoff effects. Notice, moreover, that once a link decays to an agent, she becomes less central and this makes it more likely that another link decays. Thus link gains and losses are self-reinforcing. This intuition suggests that if α, the probability of adding links, is large, then the process should approximate the complete network, while if it is small, then the process should approximate the star network. The key insight of our model is that for intermediate values of α, the network is a nested split graph.

Observe that it is assumed that there is no cost of forming links. If links represent a social tie, then there typically is a cost to maintaining a link since agents must spend time with the person they are linked to. Because of the assumption of the absence of any linking cost, each agent wants to connect to every other agent, which leads to the formation of nested split graphs. In Appendix E, we extend the model to see what would happen to our results if links were costly to maintain and only the links that increase the payoff of an agent were formed. We show that as long as the cost is not too high, marginal payoffs are positive and the networks always converge to nested split graphs so that all our results hold.

Due to the nested neighborhood structure of nested split graphs, any pair of agents in (the connected component of) a nested split graph is at most two links separated from each other. From Proposition 1 it then follows that in a nested split graph $G(t)$, the best response of an agent i is the agents with the highest degrees in i's second-order neighborhood $\mathcal{N}_i^{(2)}$.[21] Moreover, if $G(t)$ is a nested split graph, then $i \in \mathcal{B}_j(G(t))$ if and only if $j \in \mathcal{B}_i(G(t))$. Hence, we could require in addition that links are only formed under mutual consent.

From the fact that $G(t)$ is a nested split graph with an associated stepwise adjacency matrix, it further follows that at any time t in the network evolution, $G(t)$ consists of a single connected component and possibly isolated nodes.

Corollary 1. *Consider the unperturbed dynamics of the network formation process* $(G(t))_{t\in\mathbb{R}_+}$ *introduced in Definition 2. Assume that at* $t=0$*, we start with the empty network* $G(0)=\bar{K}_n$*. Then, at any time* $t \geq 0$*, the network* $G(t)$ *has at most one nonsingleton component almost surely.*

[20] Further, we will show in Proposition 3 that as ζ converges to zero, $(G(t))_{t\in\mathbb{R}_+}$ induces a finite state Markov chain where the recurrent states $\hat{\Omega}$ consist of nested split graphs.

[21] Let $\mathcal{N}_i = \{k \in \mathcal{N} : (i,k) \in \mathcal{E}(t)\}$ be the set of neighbors of agent $i \in \mathcal{N}$ and let $\mathcal{N}_i^{(2)} = \bigcup_{j\in\mathcal{N}_i} \mathcal{N}_j \setminus (\mathcal{N}_i \cup \{i\})$ denote the second-order neighbors of agent i in the current network $G(t)$. Note that the connectivity relation is symmetric such that j is a second-order neighbor of i if i is a second-order neighbor of j, i.e., $i \in \mathcal{N}_j^{(2)}$ if and only if $j \in \mathcal{N}_i^{(2)}$ for all $i, j \in \mathcal{N}$.

Nested split graphs are not only prominent in the literature on spectral graph theory, but they have also appeared in the recent literature on economic networks. Nested split graphs are called interlinked stars in Goyal and Joshi (2003).[22] Subsequently, Goyal et al. (2006) identified interlinked stars in the network of scientific collaborations among economists. It is important to note that nested split graphs are characterized by a distinctive core–periphery structure (see the Introduction (Section 1) and Section 6).

Finally, note that the network formation process $(G(t))_{t\in\mathbb{R}_+}$ introduced in Definition 2 is independent of initial conditions $G(0)$.[23] This means that even when we start from an initial network $G(0)$ that is not a nested split graph, then after some finite time the Markov chain will reach a nested split graph. This is because there exists a positive probability that all links in the current graph are removed. The resulting graph is then empty. This graph is a special case of a nested split graph. Due to Proposition 2 for $\zeta = 0$, from then on all consecutive networks visited by the chain are nested split graphs and the class of nested split graphs will never be left by the chain, that is, it forms an absorbing set. Moreover, since the chain stays forever in the class of nested split graphs and it takes only a finite number of transitions to reach this class from any other graph, all other graphs form a transient set.

4. Stochastically stable networks: Characterization

In this section, we show that the network formation process $(G(t))_{t\in\mathbb{R}_+}$ of Definition 2 induces an ergodic Markov chain with a unique invariant distribution. We then proceed by analyzing the stochastically stables states in $\hat{\Omega}$ (in the limit of $\zeta \to 0$) of this process as the number n of agents becomes large.

Proposition 3. *The network formation process $(G(t))_{t\in\mathbb{R}_+}$ introduced in Definition 2 induces an ergodic Markov chain on the finite state space Ω with a unique stationary distribution μ^ζ such that $\mu^\zeta(G') = \lim_{t\to\infty} \mathbb{P}(G(t) = G'|G(0) = G)$ for any $G, G' \in \Omega$. Moreover, the stochastically stable states $\hat{\Omega}$ are given by the set of nested split graphs Ψ such that $\mu(\Psi) = 1$.*

In the following discussion, we will assume for simplicity that $\alpha_i = 1 - \beta_i = \alpha$ for all $i \in \mathcal{N}$ in Definition 2, expressing the relative weights of link creation versus link decay.[24] In this case, the symmetry of the network formation process with respect to the link arrival rate α and the link decay parameter $1 - \alpha$ allows us to state the following proposition.

Proposition 4. *Consider the unperturbed dynamics of the Markov chain $(G(t))_{t\in\mathbb{R}_+}$ in Definition 2 with $\alpha \equiv \alpha_i = 1 - \beta_i$ for all $i \in \mathcal{N}$. Let $G(t)$ be a sample path generated with*

[22]Nested split graphs are interlinked stars, but an interlinked star is not necessarily a nested split graph. Nested split graphs have a nested neighborhood structure for all degrees, while in an interlinked star, this holds only for the nodes with the lowest and highest degrees.

[23]See Proposition 3 in Section 4 and its proof in Appendix A.

[24]Note that taking into account the possibility of an agent remaining quiescent only modifies the time scale of the process discussed, thus yielding identical results to the model proposed. This implies that, without any loss of generality, it is possible to assume $\alpha_i + \beta_i = 1$ for all $i \in \mathcal{N}$. For simplicity, we also assume that these probabilities are the same across agents.

the homogeneous link arrival rate α and let $G'(t)$ be a sample path with arrival rate $1-\alpha$. Let μ be the stationary distribution of $G(t)$ and let μ' be the stationary distribution of $G'(t)$. Then for each network G in the support of μ, the complement $\bar{G}$ of G has the same probability in μ', i.e., $\mu'(\bar{G}) = \mu(G)$.

Proposition 4 allows us to derive the stationary distribution μ for any value of $\frac{1}{2} < \alpha < 1$ if we know the corresponding distribution for $1-\alpha$. This follows from the fact that the complement $\bar{G}$ of a nested split graph G is a nested split graph as well (Mahadev and Peled 1995). In particular, the networks $\bar{G}$ are nested split graphs in which the number of nodes in the dominating subsets corresponds to the number of nodes in the independent sets in G and, conversely, the number of nodes in the independent sets in $\bar{G}$ corresponds to the number of nodes in the dominating subsets in G.

With this symmetry in mind, we restrict our analysis in the following discussion to the case of $0 < \alpha \leq \frac{1}{2}$. Let $\{\mathbf{N}(t)\}_{t\in\mathbb{R}_+}$ be the degree distribution with the dth element $N_d(t)$, giving the number of nodes with degree d in $G(t)$, in the tth sequence $\mathbf{N}(t) \equiv \{N_d(t)\}_{d=0}^{n-1}$. Further, let $P_t(d) \equiv N_d(t)/n$ denote the proportion of nodes with degree d ($\mathbf{P}(t) \equiv \mathbf{N}(t)/n$) and let $P(d) \equiv \lim_{t\to\infty} P_t(d)$ be its asymptotic value. In the following proposition, we determine the asymptotic degree distribution of the nodes in the independent sets for n large enough.[25]

Proposition 5. *Consider the unperturbed dynamics of the Markov chain $(G(t))_{t\in\mathbb{R}_+}$ in Definition 2 with $\alpha \equiv \alpha_i = 1-\beta_i$ for all $i \in \mathcal{N}$ and let $0 < \alpha \leq \frac{1}{2}$. Let $P_t(d)$ denote the proportion of nodes with degree d in $G(t)$. Then the asymptotic expected proportion of nodes in the independent sets with degrees $d = 0, 1, \ldots, d^*$ in the stochastically stable networks $G \in \hat{\Omega}$ for large n is given by*

$$\lim_{t\to\infty} \mathbb{E}(P_t(d)) = \frac{1-2\alpha}{1-\alpha}\left(\frac{\alpha}{1-\alpha}\right)^d, \tag{3}$$

where[26]

$$d^*(n,\alpha) = \frac{\ln\left(\frac{(1-2\alpha)n}{2(1-\alpha)}\right)}{\ln\left(\frac{1-\alpha}{\alpha}\right)} \tag{4}$$

and $P_t(d) \to \mathbb{E}_t(P_t(d))$ almost surely as $n \to \infty$.

The proof of the proposition follows from a series of intermediate steps, where we can take advantage of the intuitively simple stepwise structure of the adjacency matrix associated with a nested split graph (see Figure 1). First we use the fact that we can approximate the continuous time Markov chain with a sampled time Markov chain whose stationary distributions are the same (see Lemma 1 in Appendix A). We then proceed

[25] As Proposition 5 speaks of the asymptotic degree distribution in the limit of large n, this is to be understood as letting $n \to \infty$ after considering the limit of $t \to \infty$.

[26] Note that $d^*(n,\alpha)$ from (4) might, in general, not be an integer. In this case, we take the closest integer value to (4), that is, we take $[d^*(n,\alpha)] = \lfloor d^*(n,\alpha) + \frac{1}{2} \rfloor$. The error we make in this approximation is negligible for large n.

by induction to show that (3) holds for all degrees smaller than an upper bound d^* in the support of the stationary distribution μ of the sampled time Markov chain. The induction basis is concerned with the number of isolated nodes (separately derived in Lemma 3 in Appendix A) and the number of nodes with degree 1. The induction step assumes that (3) holds for $d-1$ and d to show that it then must also hold for $d+1$. To draw this conclusion, we compute the fixed point of the expected change in the number of nodes with degree d in an incremental time step using the fact that the underlying network is a nested split graph. This is possible because of the particular structure inherent in the adjacency matrix of a nested split graph and our payoff maximizing link formation protocol, which allows us to consider only a few cases for the formation or removal of a link to compute that change. Finally, by requiring that the degree distribution is a proper probability measure with mass 1, we can derive d^* in (4). The details of the proof can be found in Appendix A.

The structure of nested split graphs implies that if there exist nodes for all degrees between 0 and d^* (in the independent sets), then the dominating subsets with degrees larger than d^* contain only a single node. Further, using Proposition 4, we know that for $\alpha > \frac{1}{2}$, the expected number of nodes in the dominating subsets is given by the expected number of nodes in the independent sets in (3) for $1-\alpha$, while each of the independent sets contains a single node. This determines the asymptotic degree distribution for the independent or dominating subsets, respectively, for all values of α in the limit of large n.

From (4), we can directly derive the following corollary.

Corollary 2. *Consider the unperturbed dynamics of the Markov chain $(G(t))_{t\in\mathbb{R}_+}$ in Definition 2 with $\alpha_i = 1-\beta_i = \alpha$ for all $i \in \mathcal{N}$. Then there exists a phase transition in the asymptotic average number of independent sets, $d^*(n,\alpha)$, for $G \in \hat{\Omega}$ as n becomes large such that*

$$\lim_{n\to\infty} \frac{d^*(n,\alpha)}{n} = \begin{cases} 0 & \text{if } \alpha < \frac{1}{2} \\ \frac{1}{2} & \text{if } \alpha = \frac{1}{2} \\ 1 & \text{if } \alpha > \frac{1}{2}. \end{cases}$$

Corollary 2 implies that as n grows without bound, the networks in the stationary distribution μ are either sparse or dense, depending on the value of the link creation probability α. Moreover, from the functional form of $d(n,\alpha)$ in (4), we find that there exists a sharp transition from sparse to dense networks as α crosses $\frac{1}{2}$ and the transition becomes sharper the larger is n.

Observe that because a nested split graph is uniquely defined by its degree distribution,[27] Proposition 5 delivers us a complete description of a typical network generated by our model in the limit of large t and n. We call this network the *stationary network*. We can compute the degree distribution and the corresponding adjacency matrix of the stationary network for different values of α.[28] The latter is shown in Figure 2. From the

[27] The degree distribution uniquely determines the corresponding nested split graph up to a permutation of the indices of nodes.

[28] Noninteger values for the partition sizes can be approximated with the closest integer while preserving the nested structure of the degree partitions.

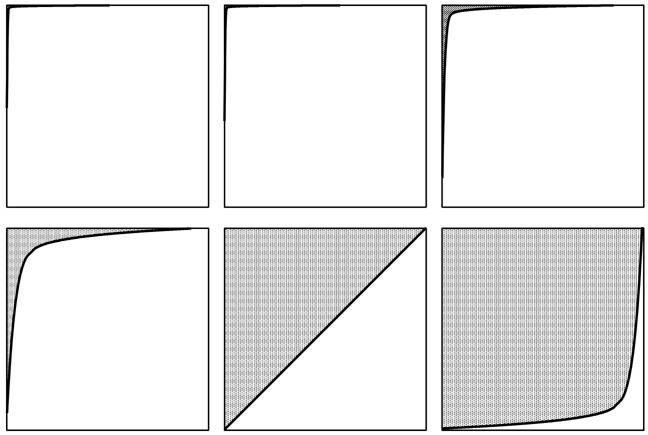


FIGURE 2. Representation of the adjacency matrices of stationary networks with $n = 1{,}000$ agents for different values of parameter α: $\alpha = 0.2$ (top-left plot), $\alpha = 0.4$ (top-center plot), $\alpha = 0.48$ (top-right plot), $\alpha = 0.495$ (bottom-left plot), $\alpha = 0.5$ (bottom-center plot), and $\alpha = 0.52$ (bottom-right plot). The solid line illustrates the step function separating the 0 from the 1 entries in the matrix. The top-left matrix for $\alpha = 0.4$ corresponds to a starlike network while the bottom-right matrix for $\alpha = 0.52$ corresponds to an almost complete network. Thus, there exists a sharp transition from sparse to densely connected stationary networks around $\alpha = 0.5$. Networks of smaller size for the same values of α can be seen in Figure 3.

structure of these matrices, we observe the transition from sparse networks containing a hub and many agents with small degree to a quite homogeneous network with many agents having similar high degrees. Moreover, this transition is sharp around $\alpha = \frac{1}{2}$. In Figure 3, we show particular networks arising from the network formation process for the same values of α. Again, we can identify the sharp transition from hub-like networks to homogeneous, almost complete networks.

Figure 4 (left) displays the number $\bar{m}$ of links m relative to the total number of possible links $n(n-1)/2$, i.e., $\bar{m} = 2m/(n(n-1))$, and the number of distinct degrees k as a function of α. We see that there exists a sharp transition from sparse to dense networks around $\alpha = \frac{1}{2}$, while k reaches a maximum at $\alpha = \frac{1}{2}$. This follows from the fact that $k = 2d^*$ with d^* given in (4) is monotonic increasing in α for $\alpha < \frac{1}{2}$ and monotonic decreasing in α for $\alpha > \frac{1}{2}$.

Note that Proposition 5 makes a statement in the limit of $n \to \infty$. In the following section (see in particular Figures 4–8), we compare various networks statistics computed from the analytical solution in Proposition 5 with the results obtained from numerical simulations for finite values of n. These figures illustrate that for relatively small values of n there is almost no deviation from the theoretical prediction of Proposition 5, providing evidence that our limit results also make reasonably good predictions in the case of a finite number n of agents.[29]

[29] This also weakens the eigenvalue condition imposed on the spillover parameter λ introduced in Section 3.1. See also footnote 10.

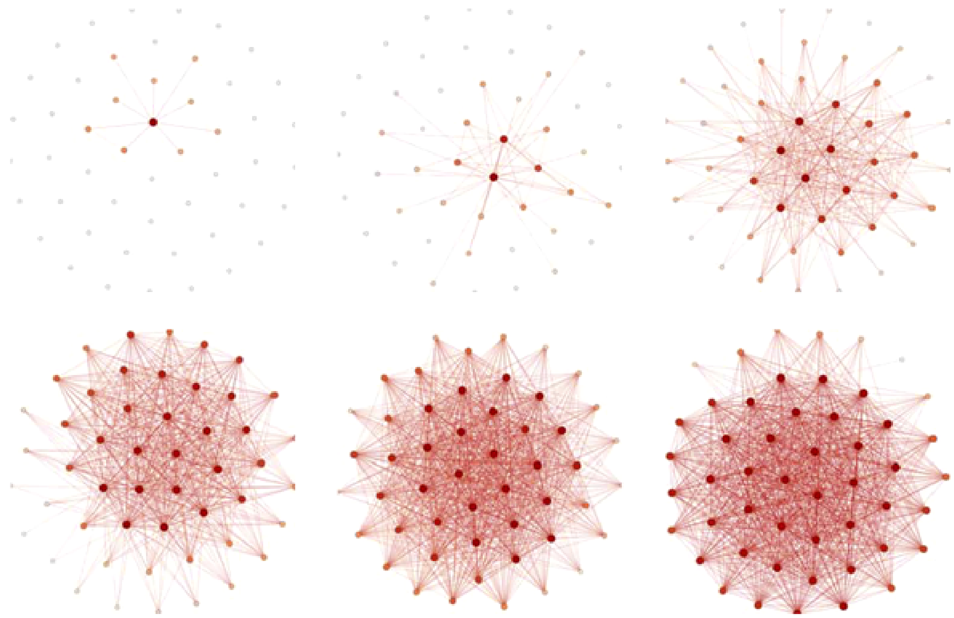


FIGURE 3. Sample networks with $n = 50$ agents for different values of parameter α: $\alpha = 0.2$ (top-left plot), $\alpha = 0.4$ (top-center plot), $\alpha = 0.48$ (top-right plot), $\alpha = 0.495$ (bottom-left plot), $\alpha = 0.5$ (bottom-center plot), and $\alpha = 0.52$ (bottom-right plot). The shade and size of the nodes indicate their eigenvector centrality. The networks for small values of α are characterized by the presence of a hub and a growing cluster attached to the hub. With increasing values of α, the density of the network increases until the network becomes almost complete.

5. Stochastically stable networks: Statistics

In the following sections, we analyze some of the topological properties of the stochastically stable networks in our model that are in the support of the stationary distribution μ. We simply refer to these networks as stationary networks. With the asymptotic expected degree distribution derived in Proposition 5, we can calculate the expected clustering coefficient, the clustering-degree correlation, the neighbor connectivity, the assortativity, and the characteristic path length by using the expressions derived for these quantities in Appendix C, where we show that these statistics are all functions of the degree distribution.[30]

Note that since the stationary distribution μ is unique, we can recover the expected value of any statistic by averaging over a large enough sample of empirical networks generated by numerical simulations. We then superimpose the analytical predictions of the statistic derived from Proposition 5 with the sample averages so as to compare the validity of our theoretical results, also for small network sizes n. As we will show, there is a good agreement of the theory with the empirical results for all n.

[30]Any network statistic $f:\Omega \to \mathbb{R}$ we consider can be expressed as a function of the (empirical) degree distribution $\mathbf{P}_t:\Omega \to [0,1]^n$. Hence, we can compute the expectation as $\mathbb{E}_t(f) = \sum_{\mathbf{k}\in(0,\dots,n)^n} f(\mathbf{k}/n)\mathbb{P}_t(\mathbf{P}_t = \mathbf{k}/n)$. In Proposition 5, we show that the degree distribution converges to its expected value with probability 1. Therefore, we have that $\mathbb{E}_t(f) = \sum_{\mathbf{k}\in(0,\dots,n)^n} f(\mathbf{k}/n)\mathbb{1}_{\mathbb{E}_t(\mathbf{P}_t)}(\mathbf{k}/n) = f(\mathbb{E}_t(\mathbf{P}_t))$ as $n \to \infty$.

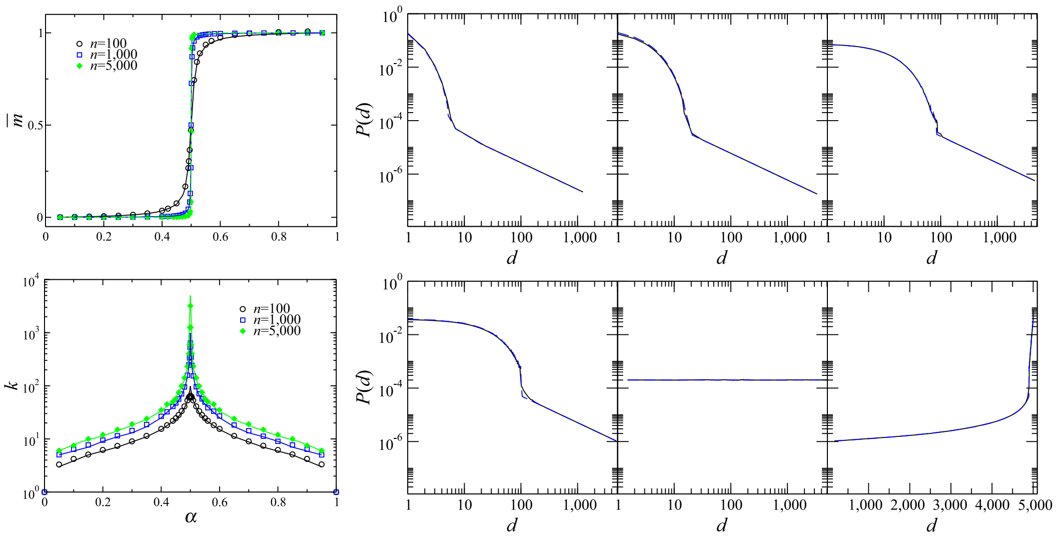


FIGURE 4. (Left) In the top panel we show the number $\bar{m}$ of links m relative to the total number of possible links $n(n-1)/2$ of the stationary network. The number of distinct degrees $k = 2d^*$, with d^* from (4), found in the stationary network for different values of α are shown in the bottom panel. The figures display both the results obtained by recourse of numerical simulations (symbols) and by respecting theoretical predictions (lines) of the model. (Right) Degree distribution $P(d)$ for different values of parameter α and a network size $n = 10{,}000$: $\alpha = 0.2$ (top-left plot), $\alpha = 0.4$ (top-center plot), $\alpha = 0.48$ (top-right plot), $\alpha = 0.49$ (bottom-left plot), $\alpha = 0.5$ (bottom–center plot), and $\alpha = 0.52$ (bottom-right plot). The solid line corresponds to the average of simulations, while the dashed line indicates the theoretical degree distribution from Proposition 5. The degrees have been binned to smooth the degree distribution.

We would like now to investigate the properties of our networks and see how they match real-world networks.

Degree distribution

From Proposition 5, we find that the degree distribution follows an exponential decay with a power-law tail.[31] The power-law tail has an exponent of -1.[32] Figure 4 displays the relative degree in the network (left panel) and the degree distribution (right panel).

[31]For $0 < \alpha \leq \frac{1}{2}$ and n large enough, the asymptotic expected degree distribution for the degrees d smaller than or equal to d^* is given by an exponential function $P(d) = ((1-2\alpha)/(1-\alpha))\exp(-\ln((1-\alpha)/\alpha)d)$. On the other hand, if we assume (i) that the degree of a node in a dominating subset is symmetrically distributed around its expected value, (ii) we compute the integral over the probability density function by a rectangle approximation, and (iii) we further assume that the degree distribution obtained in this way has the same functional form for all degrees d larger than d^*, then one can show that for $0 < \alpha \leq \frac{1}{2}$ and n large enough, the asymptotic expected degree distribution $P(d)$ is given by $P(d) = (\alpha/((1-2\alpha)n))d^{-1}$. The power-law tail of the degree distribution can be confirmed by the empirical distribution from a logarithmic binning of numerical simulations, as can be seen in Figure 4 (right).

[32]We can extend our model to obtain a degree distribution with an arbitrary power-law tail by making the probability of selecting an agent depend on the number of links she already has, while preserving the nested

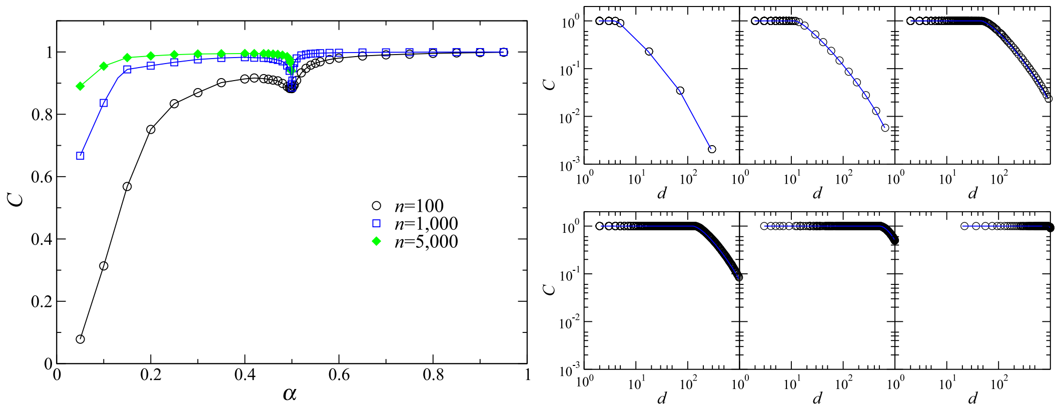


FIGURE 5. The left panel shows the clustering coefficient C and the right panel the clustering-degree correlation of stationary networks. The symbols correspond to the results obtained by recourse of numerical simulations. The solid lines correspond to the analytical results. We show that the clustering-degree correlation is negative for different values of α and a network size of $n = 1{,}000$. The different plots show different values of α: $\alpha = 0.2$ (top-left plot), $\alpha = 0.4$ (top-center plot), $\alpha = 0.48$ (top-right plot), $\alpha = 0.49$ (bottom-left plot), $\alpha = 0.5$ (bottom-center plot), and $\alpha = 0.52$ (bottom-right plot).

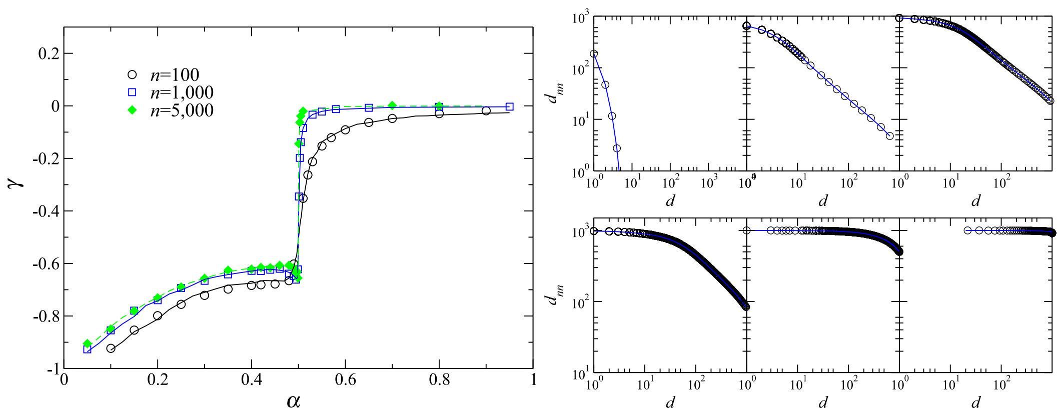


FIGURE 6. In the left panel, we show the assortativity γ of stationary networks. In the right panel, we show the average nearest neighbor connectivity d_{nn} for $\alpha = 0.2$ (top-left plot), $\alpha = 0.4$ (top-center plot), $\alpha = 0.48$ (top-right plot), $\alpha = 0.49$ (bottom-left plot), $\alpha = 0.5$ (bottom-center plot), and $\alpha = 0.52$ (bottom-right plot). The symbols correspond to the results obtained by recourse of numerical simulations. The solid lines correspond to the analytical results.

Clustering

The clustering coefficient is shown in Figure 5 (left). We find that for practically all values of α, the clustering in the stationary networks is high. This finding is in agreement with

structure of the network she is embedded in. This extension is further discussed in König and Tessone (2011).

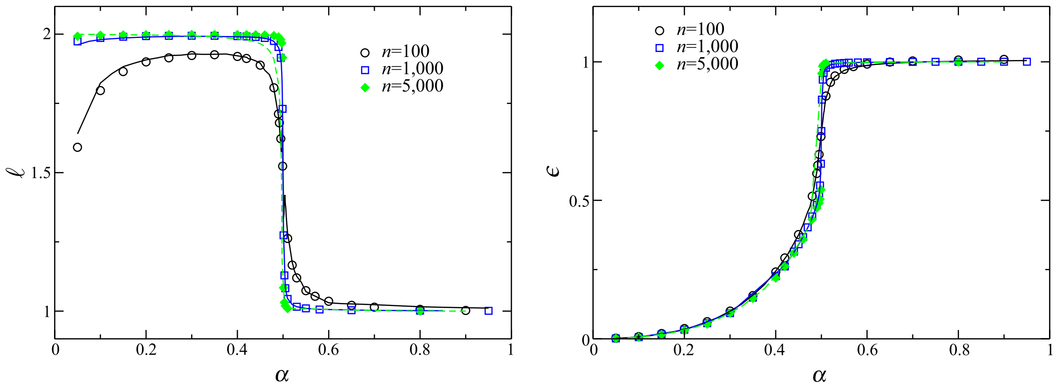


FIGURE 7. (Left panel) The characteristic path length ℓ of stationary networks and (right panel) the results for the network efficiency ϵ, obtained by recourse of numerical simulations (symbols) and respecting theoretical predictions (lines) of the model.

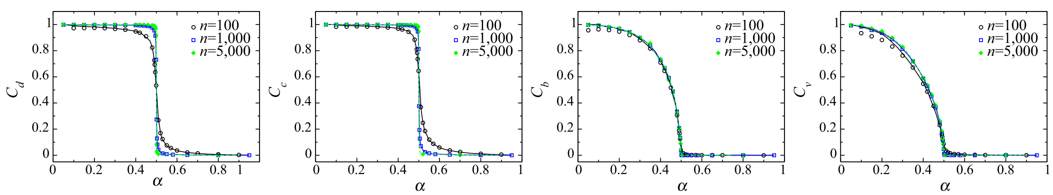


FIGURE 8. (From left to right) Degree, closeness, betweenness, and eigenvector centralization in the stationary networks for different values of α. For all centralization measures, we obtain a sharp transition between strongly centralized networks for lower values of α and decentralized networks for higher values of α. Note that we have only considered the connected component for the computation of the different centralization measures.

the vast literature on social networks that has reported high clustering to be a distinctive feature of social networks. Moreover, Goyal et al. (2006) have shown that there exists a negative correlation between the clustering coefficient of an agent and her degree. We find this property in the stationary networks as well, as it is shown in Figure 5 (right).

Assortativity and nearest neighbor connectivity

We now turn to the study of correlations between the degrees of the agents and their neighbors. This property is usually measured by the network assortativity γ (Newman 2002)[33] and nearest neighbor connectivity $d_{nn}(d)$ (Pastor-Satorras et al. 2001). Dissortative networks are characterized by negative degree correlations between a node and

[33]The assortativity coefficient $\gamma \in [-1, 1]$ is essentially the Pearson correlation coefficient of degree between nodes that are connected. Positive values of γ indicate that nodes with similar degrees tend to be connected (and $d_{nn}(d)$ is an increasing function of the degree d), while negative values indicate that nodes with different degrees tend to be connected (and $d_{nn}(d)$ is a decreasing function of the degree d). See Newman (2002) for further details.

its neighbors, and assortative networks show positive degree correlations. In dissortative networks, γ is negative and $d_{nn}(d)$ is monotonic decreasing, while in assortative networks γ is positive and $d_{nn}(d)$ is monotonic increasing. In our model, we observe dissortative networks.[34]

Assortativity and average nearest neighbor connectivity for different values of the link creation probability α are shown in Figure 6. Clearly, stationary networks are dissortative, while the degree of dissortativity decreases with increasing α. The dissortativity of stationary networks simply reflects the fact that stationary networks are strongly centralized for values of α below $\frac{1}{2}$.

Characteristic path length

Figure 7 shows the characteristic path length ℓ and the network efficiency ϵ (defined in Appendix C.1.4). From these figures, one can see that the characteristic path length ℓ never exceeds a distance of 2. Together with the high clustering shown in this section, the stationary networks can be seen as "small worlds" (Watts and Strogatz 1998). Stationary networks are efficient for values of α larger than $\frac{1}{2}$, in terms of short average distance between agents, while for values of α smaller than $\frac{1}{2}$ they are not. However, this short average distance is attained at the expense of a large number of links.

Centralization of stochastically stable networks

In the following section, we analyze the degree of centralization in stationary networks.

For our analysis, we use the centralization index introduced by Freeman (1979). The centralization $C:\Omega \to [0, 1]$ of a network $G = (\mathcal{N}, \mathcal{E}) \in \Omega$ is given by

$$C \equiv \frac{\sum_{u \in G}(C(u^*) - C(u))}{\max_{G'} \sum_{v \in G'}(C(v^*) - C(v))},$$

where u^* and v^* are the agents with the highest values of centrality in the current network and the maximum in the denominator is computed over all networks $G' = (\mathcal{N}, \mathcal{E}') \in \Omega$ with the same number of agents.

From Figure 8 (right), showing degree, closeness, betweenness, and eigenvector centralization, we clearly see that there exists a phase transition at $\alpha = \frac{1}{2}$ from highly centralized to highly decentralized networks. This means that for low arrival rates of linking opportunities α (and a strong link decay), the stationary network is strongly polarized, composed mainly of a star (or an interlinked star as in Goyal and Joshi 2003), while for high arrival rates of linking opportunities (and a weak link decay), stationary networks are largely homogeneous. We can also see that the transition between these states is sharp. It is interesting to note that the same pattern emerges for all centrality measures considered, irrespective of whether the measures take into account only the local neighborhood of an agent, such as in the case of degree centrality, or the entire network structure, as for the other centrality measures.

[34] In König et al. (2010), we show that by introducing capacity constraints in the number of links an agent can maintain, we are able to produce both assortative as well as dissortative networks.

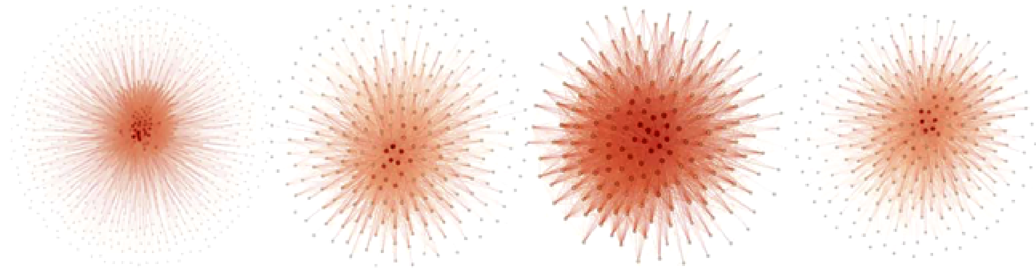


Figure 9. The Austrian banking network, the global network of banks obtained from the Bank of International Settlements (BIS) locational statistics, the gross domestic product (GDP) trade network, and the arms trade network (from left to right). The shade and size of the nodes indicate their eigenvector centrality. The GDP trade network is much more dense than the network of banks and the network of arms trade. All four networks show a core of densely connected nodes.

6. Empirical implications

6.1 *Data*

In this section, we would like to provide real-world evidence for our model and estimate the model's parameters for four different empirical data sets, all of which are characterized by a strongly nested network architecture. We essentially consider two types of networks: bank and trade networks.[35] In the following discussion, we describe in detail the different data sets that we use.

The first network we analyze is a network of Austrian banks in the year 2008 (see Boss et al. 2004). Links in the network represent exposures between Austrian-domiciled banks on a nonconsolidated basis (i.e., no exposures to foreign subsidiaries are included). We obtain a sample of $n = 770$ banks with $m = 2{,}454$ links between them and an average degree of $\bar{d} = 20.54$. The degree variance is $\sigma_d^2 = 1{,}273.22$. The largest connected component comprises 768 banks, which is 99.7% of the total of banks, and it is illustrated in Figure 9.

Second, we consider the global banking network in the year 2011 obtained from the Bank of International Settlements (BIS) locational statistics on exchange-rate-adjusted changes in cross-border bank claims (see Minoiu and Reyes 2011). BIS locational statistics are compiled on the basis of residence of BIS reporting banks and cover the cross-border positions of all banks domiciled in the reporting area, including positions with respect to foreign affiliates, loans, deposits, debt securities, and other assets provided by banks. We obtain a network with $n = 239$ nodes and $m = 2{,}454$ links between them. An illustration can be seen in Figure 9. The average degree of the network is $\bar{d} = 20.54$ and we observe a high degree variance of $\sigma_d^2 = 1{,}273.22$.

The third empirical network we consider is the network of trade relationships between countries in the year 2000. The trade network is defined as the network of import–export relationships between countries in a given year in millions of current-year U.S.

[35]In Appendix D, we discuss an application of our model to networks of banks (see Appendix D.1), where links are loans between banks, and in terms of trade networks (see Appendix D.2), where links between countries represent trade relationships (in imports or exports).

dollars. We construct an undirected network in which a link is present between two countries if either one has exported to the other country. The trade network contains $n = 196$ nodes, $m = 4{,}138$ links, has an average degree of $\bar{d} = 42.22$, and a degree variance of $\sigma_d^2 = 1{,}524.16$.

Fourth, we consider the network of arms trade between countries (see Åkerman and Larsson forthcoming). We use data obtained from the Stockholm International Peace Research Institute (SIPRI) Arms Transfers Database that holds information on all international transfers between countries of seven categories of major conventional weapons accumulated from 1950 to 2010. A link in the network represents a recipient or supply relationship of arms between two countries during this period. We obtain a network with $n = 246$ nodes and $m = 2{,}245$ links. The average degree is $\bar{d} = 18.25$ and the degree variance is $\sigma_d^2 = 589.97$. An illustration can be seen in Figure 9.

All these four real-world networks are of similar size, show short average path lengths of around 2, are dissortative, and have a monotonic decreasing average nearest neighbor connectivity. They also show a relatively high clustering and the clustering degree distribution is decreasing with the degree (see Figure 11). An important feature of these networks is that they all show a high degree of *nestedness*. This can be witnessed from the adjacency matrices depicted in Figure 10, which resemble the nested matrices we derive from our theoretical model (see Figure 2). Similarly, when we compare the networks simulated from our model (see Figure 3) and the ones described in real-world networks (Figure 9), they are relatively similar (in terms of a clear core–periphery structure, indicating nestedness).[36,37]

6.2 *Estimating the model's parameters*

We then estimate the main parameters $\boldsymbol{\theta} \equiv (\alpha, \zeta)$ of our model by using the likelihood-free Markov chain Monte Carlo (LF-MCMC) algorithm suggested by Marjoram et al. (2003). The purpose of this algorithm is to estimate the parameter vector $\boldsymbol{\theta}$ of our model on the basis of the summary statistics $\mathbf{S} \equiv (\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3)_{n\times 3}$, where $\mathbf{S}_1 \equiv (P(d))_{d=0}^{n-1}$, $\mathbf{S}_2 \equiv (C(d))_{d=0}^{n-1}$, and $\mathbf{S}_3 \equiv (d_{nn}(d))_{d=0}^{n-1}$ are the degree distribution, the clustering degree

[36]We have performed a k-core decomposition of the empirical networks. A k-core is a maximal subnetwork in which all nodes have a degree of at least k with the other nodes in the subnetwork. Examining the k-cores with increasing values of k does not split the network into separate components. This is another indicator for the nested structure observed in these networks.

[37]We can complement these observations of nested patterns in real-world networks by a rigorous statistical analysis. The key question is how to measure, precisely, the degree of nestedness of a network. For that, we have computed the degree of nestedness by calculating the *matrix temperature* T_n using the BINMATNEST algorithm proposed by Rodríguez-Gironés and Santamaría (2006). Typically, the lower the temperature T_n, the higher the degree of nestedness. More precisely, T_n is normalized in such a way that it ranges between 0 for a perfectly nested matrix and 100 for a maximally "unnested" matrix. For the network of Austrian banks, we obtain $T_n = 0.05$; for the network of banks from BIS statistics, we get $T_n = 0.75$; for the trade network, we obtain $T_n = 7.26$; and for the arms trade network, we obtain $T_n = 1.72$. This indicates that the networks of banks have the highest degree of nestedness. Moreover, we can also compute the probability of a certain degree of nestedness being generated at random. For all the networks considered, we obtain a p-value not distinguishable from 0 (using 500 null matrices), showing that all empirical networks are significantly nested. See Rodríguez-Gironés and Santamaría (2006) for further details of the BINMATNEST algorithm.

distribution, and the average nearest neighbor degree distribution, respectively. Moment conditions are obtained from the Euclidean distances $\Delta(\mathbf{S}_i, \mathbf{S}_i^o) \equiv \sqrt{\sum_{j=1}^{n}(S_{i,j} - S_{i,j}^o)^2}$ for each statistic $\mathbf{S}_i$ (generated by the algorithm) and its observed value $\mathbf{S}_i^o$. The algorithm generates a Markov chain that is a sequence of parameters $\{\boldsymbol{\theta}_s\}_{s=1}^{S}$ with a stationary distribution that approximates the distribution of the parameter values $\boldsymbol{\theta}$ conditional on the observed statistic $\mathcal{S}^o$.[38] Since this estimation algorithm would require the computation of the Bonacich centrality an extensive number of times, we assume that the complementarity parameter λ is small such that we can approximate the Bonacich centrality by the degree centrality when simulating the network formation process.[39,40]

The estimated parameter values are shown in Table 1. We observe that the estimates for ζ are higher for the network of GDP trading countries and the network of arms trade than the corresponding estimates for the networks of banks. This confirms our intuition that with increasing values of ζ, stationary networks become less nested (and we obtain a random graph as $\zeta \to \infty$), and the values for the matrix temperature T_n become lower for these networks (see also the adjacency matrices in Figure 10). Hence, our estimates support our earlier observation that the networks of banks have a higher degree of nestedness than the networks of trade relationships between countries.

Moreover, Figure 11 shows the empirical distributions (squares) and typical simulated distributions (circles) for the bank network, the network of GDP trade, and the arms trade network. The comparison of observed and simulated distributions shown in Figure 11 indicates that the model can relatively well reproduce the observed empirical networks, even though the model is parsimoniously parameterized in relying only on two exogenous variables α and ζ. The fit seems to be best for the networks of banks, which also show the most distinct nestedness pattern (see Figure 10).

Appendix A: Proofs of propositions, corollaries, and lemmas

In this section we give the proofs of the propositions, corollaries, and lemmas stated earlier in the paper.

Proof of Proposition 1. (i) A graph having a stepwise adjacency matrix is a nested split graph G. A nested split graph has a nested neighborhood structure. The neighborhood $\mathcal{N}_j$ of an agent j is contained in the neighborhood $\mathcal{N}_i$ of the next higher degree agent i with $|\mathcal{N}_i| = d_i > |\mathcal{N}_j| = d_j$ with $\mathcal{N}_j \subset \mathcal{N}_i$. For the adjacency matrix $\mathbf{A}$, the vector of

[38] For the implementation of the algorithm, we have chosen an initial uniform (prior) parameter distribution. The proposal distribution is a normal distribution. During the "burn-in" phase (Chib 2001), we consider a monotonic decreasing sequence of thresholds with appropriately chosen values from careful numerical experimentation. For the Austrian banking network, we have chosen a burn-in period of 1,000 steps, while for the network of GDP trade, we have used a period of 3,000.

[39] The Bonacich centrality is defined by $b_i(G,\lambda) = \sum_{k=0}^{\infty}\lambda^k(\mathbf{A}^k \cdot \mathbf{u})_i = 1 + \lambda d_i + \lambda^2 \sum_{j\in\mathcal{N}_i} d_j + \lambda^3 \sum_{j\in\mathcal{N}_i}\sum_{k\in\mathcal{N}_j} d_k + \cdots = 1 + \lambda d_i + \lambda^2 \sum_{j\in\mathcal{N}_i} d_j + O(\lambda^3)$. Marginal payoff from forming a link (i,j) for agent i can then be written as $\pi_i^*(G \oplus (i,j),\lambda) - \pi_i^*(G,\lambda) = \lambda(2+\lambda)/2 + (\lambda^2/2)d_i(d_i+1) + \lambda^2 d_j + O(\lambda^3)$. When computing marginal payoffs from forming a link (and the decay rates), we ignore terms of $O(\lambda^3)$.

[40] Note also that the reported estimates of ζ hold only up to a scaling factor, which depends on the choice of λ. Hence, only the relative values of ζ between different samples is meaningful, but not its absolute value.

	Austrian bank network				Global bank network				GDP trade network				Arms trade network			
θ	μ_θ	σ_θ	τ_θ	$p_\theta(S)$	μ_θ	σ_θ	τ_θ	$p_\theta(S)$	μ_θ	σ_θ	τ_θ	$p_\theta(S)$	μ_θ	σ_θ	τ_θ	$p_\theta(S)$
α	0.45	0.00	29.63	0.99	0.45	0.01	10.65	0.91	0.44	0.00	6.49	0.96	0.44	0.00	50.67	0.99
ζ	1.13	0.06	68.98	0.99	1.70	0.18	91.64	0.96	20.51	0.32	143.30	0.89	24.87	0.16	657.82	0.93
n	770				239				196				246			
S	3,000				700				2,000				5,000			

[a] τ_θ is the integrated autocorrelation time, which should be much smaller than the number S of iterations of the Markov chain algorithm used to compute the parameter estimates (Sokal 1996).

[b] μ_θ is the mean and σ_θ is the simulation standard deviation calculated from batch means (of length 10) for each parameter $\theta \in \boldsymbol{\theta}$ (Chib 2001).

[c] $p_\theta(S)$ is the p-value associated with Geweke's spectral density diagnostic indicating the convergence of the chain (Brooks and Roberts 1998). The number S of iterations of the chain have been chosen for each data set individually such that reasonably high values of $p_\theta(S)$ are obtained.

Table 1. Estimation of the model parameters $\theta \in \boldsymbol{\theta} = (\alpha, \zeta)$ for the Austrian network of banks, the global network of banks obtained from the Bank of International Settlements (BIS) locational statistics, the network of GDP trading countries, and the arms trade network.[a] The table shows simulated averages of the parameters and their standard deviations,[b] after the chain has converged.[c]

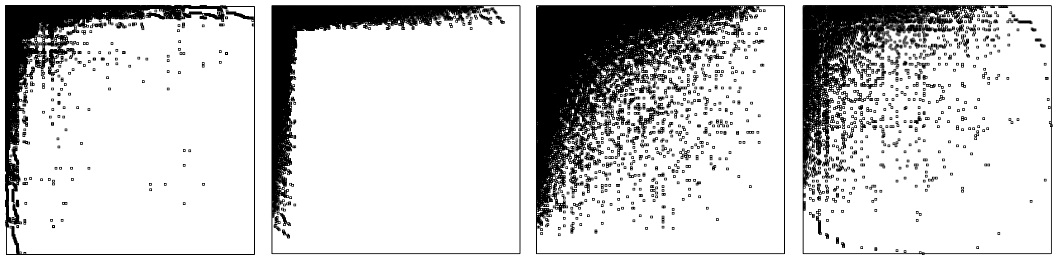


FIGURE 10. Adjacency matrices (sorted by the eigenvector centralities of the nodes) for the Austrian banking network, the global network of banks obtained from the Bank of International Settlements (BIS) locational statistics, the GDP trade network, and the arms trade network (from left to right). All adjacency matrices are significantly nested.

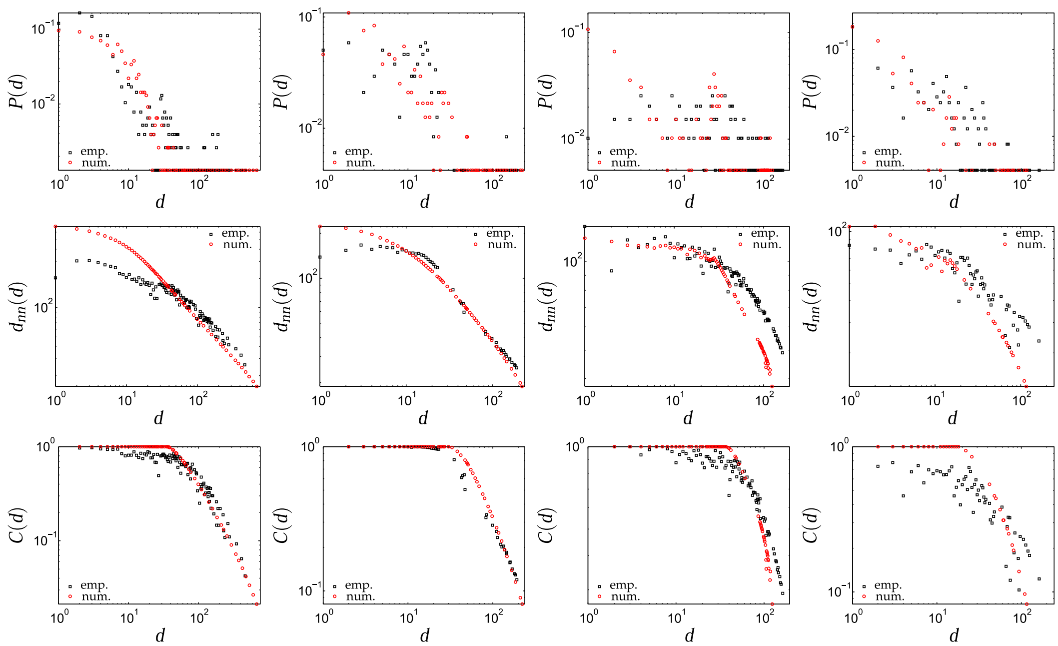


FIGURE 11. The empirical (□) and an exemplary simulated (○) degree distribution $P(d)$, average nearest neighbor degree $d_{nn}(d)$, and clustering degree distribution $C(d)$ for the Austrian banking network (first column), the network of banks obtained from the Bank of International Settlements (BIS) (second column), the GDP trade network (third column), and the arms trade network (fourth column).

Bonacich centralities is given by $\mathbf{b}(G,\lambda)=\lambda\mathbf{A}\mathbf{b}+\mathbf{u}$, with $\mathbf{u}=(1,\ldots,1)^{\top}$. For agent i, we then get

$$b_i(G,\lambda)=1+\lambda\sum_{k=1}^{n}a_{ik}b_k(G,\lambda)=1+\lambda\sum_{k\in\mathcal{N}_i}b_k(G,\lambda)$$

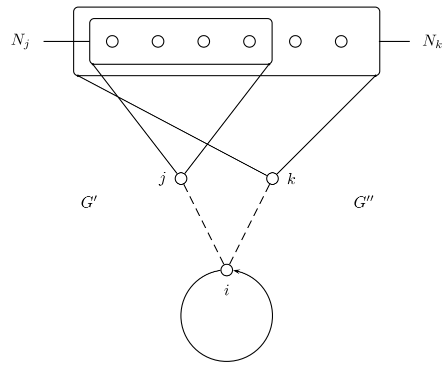


FIGURE 12. An illustration of the two networks G' and G'', which differ in the links (i, j) and (i, k). The neighborhood $\mathcal{N}_j$ of agent j and the neighborhood $\mathcal{N}_k$ of agent k are indicated by corresponding boxes. Note that the neighborhood of agent j is contained in the neighborhood of agent k. The loop at agent i indicates a walk starting at i and coming back to i before proceeding to either agent j or k.

and similarly for agent j,

$$b_j(G, \lambda) = 1 + \lambda \sum_{k \in \mathcal{N}_j} b_k(G, \lambda).$$

Since $\mathcal{N}_j \subset \mathcal{N}_i$ and $d_j = |\mathcal{N}_j| < |\mathcal{N}_i| = d_i$, we get

$$\frac{b_i(G, \lambda)}{b_j(G, \lambda)} = \frac{1 + \lambda \sum_{k \in \mathcal{N}_i} b_k(G, \lambda)}{1 + \lambda \sum_{k \in \mathcal{N}_j} b_k(G, \lambda)} > 1.$$

The inequality follows from the fact that the Bonacich centrality is nonnegative and the numerator contains the sum over the same positive numbers as the denominator plus some additional values.

Conversely, in a nested split graph, we must either have $\mathcal{N}_i \subset \mathcal{N}_j$ or $\mathcal{N}_j \subset \mathcal{N}_i$. Assuming that $b_i(G, \lambda) > b_j(G, \lambda)$, we can conclude from the above equation that $\mathcal{N}_j \subset \mathcal{N}_i$ and, therefore, $|\mathcal{N}_i| = d_i > |\mathcal{N}_j| = d_j$. If there are l distinct degrees in G, then the ordering of degrees $d_1 > d_2 > \cdots > d_l$ is equivalent to the ordering of the Bonacich centralities $b_1(G, \lambda) > b_2(G, \lambda) > \cdots > b_l(G, \lambda)$.

(ii) Consider the agents i, j, and k in the nested split graph $G(t)$, such that $d_j \leq d_k$. Let G' be the graph obtained from $G(t)$ by adding the link (i, j) and let G'' be the graph obtained from $G(t)$ by adding the link (i, k). We want to show that the Bonacich centrality of agent i in G'' is higher than in G', that is, $b_i(G', \lambda) < b_i(G'', \lambda)$. For this purpose, we count the number of walks emanating at agent i when connecting to either agent j or agent k. Since G is a nested split graph, we have that $\mathcal{N}_j \subset \mathcal{N}_k$. An illustration is given in Figure 12. We consider a walk W_l of length $l \geq 2$ starting at agent i in G'. We want to know how many such walks there are in G' and G'', respectively. For this purpose, we distinguish the following cases:

(a) Assume that W_l does not contain the link (i, j) or the link (i, k). Then each such walk W_l in G' is also contained in G'', since G' and G'' differ only in the links (i, j) and (i, k).

(b) Consider the graph G' and a walk W_l starting at agent i and proceeding to agent j. For each walk W_l in G', there exists a walk $\tilde{W}_l$ in G'' that is identical to W_l except that instead of proceeding from i to j, it proceeds from i to k and then to the neighbor of j that is visited after j in W_l. This is always possible since the neighbors of j are also neighbors of k.

(c) Consider a walk W_l in G' that starts at i but first takes a detour, returning to i before proceeding from i to j. Using the same argument as in (ii), it follows that for each such walk W_l in G', there exists a walk of the same length in G''.

(d) Consider a walk W_l in G' that starts at agent i and at some point in its sequence of agents and links proceeds from agent j to agent i. For each such walk W_l in G', there exists a walk $\tilde{W}_l$ in G'' that is identical to W_l except that it does not proceed from a neighbor of j to j and then to i; it proceeds from a neighbor of j to k and then to i.

The above cases take into account all possible walks in G' and G'' of an arbitrary length l and show that in G'' there are at least as many walks of length l starting from agent i as there are in G'.

Now consider the walks of length 2, W_2, in G' starting at agent i and proceeding to agent j. Then there are $|\mathcal{N}_j|$ such walks in G'. However, there are $|\mathcal{N}_k| > |\mathcal{N}_j|$ such walks in G'' of length 2 that start at agent i.

The Bonacich centrality $b_i(G(t), \lambda)$ is computed by the number of all walks in $G(t)$ starting from i, where the walks of length l are weighted by their geometrically decaying factor λ^l. We have shown that for each l, the number of walks in G'' is larger than or equal to the number of walks in G', and for $l = 2$, it is strictly larger. Thus, the Bonacich centrality of agent i in G'' is higher than in G'.

(iii) An analogous argument as for the creation of a link holds for the removal of a link for agent i from $j, k \in \mathcal{N}_i$ with $d_k > d_j$. Since the number of walks starting from i is reduced more by removing the link (i, k) than by removing the link (i, j) (there are at least as many walks from i passing through k than there are through j), we must have that $b_i(G \ominus (i, k), \lambda) < b_i(G \ominus (i, j), \lambda)$.

Finally, note that all agents in a nested split graph are at most two links separated from each other (if there exists any walk between them). Thus, the agent with the highest degree is also the agent with the highest degree among the neighbors' neighbors. From this discussion, we see that in a nested split graph $G(t)$, the best response of an agent i is the agents with the highest degrees in i's second-order neighborhood. □

Proof of Proposition 2. It has to be clear that this proof only holds for the unperturbed dynamics when the noise vanishes, that is, when $\zeta = 0$. This is the case of the stochastically stable states. As a result, we are making a claim about the stochastically stable states ($\zeta = 0$), but not the case of $\zeta > 0$.

$$\mathbf{A}' = \begin{pmatrix} 0&1&1&1&1&1&1&1&1&1\\ 1&0&1&1&1&1&1&1&1&1\\ 1&1&0&1&1&1&1&1&0&0\\ 1&1&1&0&1&1&0&0&0&1\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 1&1&0&1&0&0&0&0&0&0 \end{pmatrix} \xleftarrow[G\oplus(4,10)]{} \mathbf{A} = \begin{pmatrix} 0&1&1&1&1&1&1&1&1&1\\ 1&0&1&1&1&1&1&1&1&1\\ 1&1&0&1&1&1&1&1&0&0\\ 1&1&1&0&1&1&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0 \end{pmatrix} \xrightarrow[G\oplus(4,7)]{} \mathbf{A}'' = \begin{pmatrix} 0&1&1&1&1&1&1&1&1&1\\ 1&0&1&1&1&1&1&1&1&1\\ 1&1&0&1&1&1&1&1&0&0\\ 1&1&1&0&1&1&1&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0 \end{pmatrix}$$

FIGURE 13. Two possible positions for the creation of a link from agent 4, either to agent 7 (right) or to agent 10 (left). Agent 7 has degree 3, while agent 10 has degree 1. Creating a link to an agent with higher degree results in higher equilibrium payoffs. Thus, the best response of agent 4 is agent 7 and not agent 10.

Let us now start with the proof. We give a proof by induction. The induction basis is trivial. We start at $t = 0$ from an empty network $G(0) = \bar{K}_n$, which has a trivial stepwise adjacency matrix (see also Definition 5). Since there are no links present in $\bar{K}_n$, we can omit the removal of a link. Consider a small time increment $\Delta t > 0$. During that time interval, a one step transition with positive probability can only involve the creation of a link by an isolated agent. All other isolated agents are best responses of this agent. The formation of the link creates a path of length 1 whose adjacency matrix is stepwise. This is true because we can always find a simultaneous columns and rows permutation that makes the adjacency matrix stepwise. Thus, $G(\Delta t)$ has a stepwise adjacency matrix.

Next we consider the induction step of a one step transition from $G(t)$ to $G(t + \Delta t)$. By the induction hypothesis, $G(t)$ is a nested split graph with a stepwise adjacency matrix. First, we consider the creation of a link (i, j). Let agent j be a best response of agent i, that is, $j \in \mathcal{B}_i(G(t))$. Using Proposition 1, this means that agent i must be the agent with the highest degree not already connected to j. From the stepwise adjacency matrix $\mathbf{A}(G(t))$ of $G(t)$ (see Definition 5), we find that adding the link (i, j) to the network $G(t)$ such that j has the highest degree among all agents not already connected to i results in a matrix $\mathbf{A}(G(t) \oplus (i, j))$ that is stepwise. Therefore, the network $G(t) \oplus (i, j)$ is a nested split graph.

We give an example in Figure 13. Let the agents be numbered by the rows, respectively, columns, of the adjacency matrix. We assume that agent 4 receives a link creation opportunity. Two possible positions for the creation of a link from agent 4, either to agent 7 or to agent 10, are indicated with boxes. Since, in a stepwise matrix, the agent in the best response set has the highest degree, agent 7 is a best response of agent 4, while agent 10 is not. It further holds that agent 4 is also a best response of agent 7, since agent 4 is the agent with the highest degree not already connected to agent 7. Finally, we observe that creating the link 47 preserves the stepwise form of the adjacency matrix (see also Definition 5).[41]

For the decay of a link, a similar argument can be applied as in the preceding discussion. Disconnecting from the agent with the smallest degree decreases the Bonacich centrality and equilibrium payoffs the least, and, hence, this will be the link that decays

[41]The adjacency matrix is uniquely defined up to a permutation of its rows and columns. Applying such a permutation, we can always find an adjacency matrix that is stepwise.

as $\zeta = 0$. From the properties of the stepwise matrix $\mathbf{A}(G(t))$, it then follows that the matrix $\mathbf{A}(G(t) \ominus (i, j))$ is stepwise.

Thus, at any time $t \geq 0$ in the network formation process $(G(t))_{t\in\mathbb{R}_+}$, $G(t)$ is a nested split graph with an associated stepwise adjacency matrix $\mathbf{A}(G(t))$. Let Ψ denote the set of nested split graphs on n nodes. It can be shown that $|\Psi| = 2^{n-1}$ (Mahadev and Peled 1995). We thus have shown that for the unperturbed dynamics in the limit of vanishing mistakes (noise), when $\zeta = 0$, the network $G(t)$ is a nested split graph almost surely, which is to say that $\mathbb{P}(G(t) \in \Psi | G(0) = \bar{K}_n) = 1$ for all $t \geq 0$. □

Proof of Corollary 1. In Proposition 2, we have shown that $G(t)$ generated by $(G(t))_{t\in\mathbb{R}_+}$ is a nested split graph for all times t. In a nested split graph, any node in the connected component is directly connected to the node(s) with maximum degree. Thus, there exists a path of at most length 2 from any node to any other node in the connected component. It follows that $G(t)$ consists of a connected component and possible isolated nodes. □

Proof of Proposition 3. First, we show that $(G(t))_{t\in\mathbb{R}_+}$ is a Markov chain. Since the transition rate $q^\zeta(G, G')$ governing the transition from a network G to a network G' depends only on the current network G, the following Markov property holds:

$$\mathbb{P}\big(G(t+s) = G' | G(s) = G, \{G(u) : 0 \leq u < s\}\big) = \mathbb{P}(G(t+s) = G' | G(s) = G)$$

for all $t \geq 0$, $s \geq 0$, and $G, G' \in \Omega$. The number of possible networks $G(t)$ is finite for any time $t \geq 0$ and the transition rates depend on the state $G(t)$ but not on the time t. Therefore, $(G(t))_{t\in\mathbb{R}_+}$ is a finite state, continuous time, homogeneous Markov chain. Further, note that the transition rates are bounded.

Next, we show that the Markov chain is irreducible. Consider two networks $G, G' \in \Omega$. $(G(t))_{t\in\mathbb{R}_+}$ is irreducible if there exists a positive probability to pass from any G to any other G' in Ω. This means that there exists a sequence of networks $G_1, G_2, \ldots, G_n$ with the property that $q^\zeta(G, G_1)q^\zeta(G_1, G_2)\cdots q^\zeta(G_n, G') \neq 0$. We say that G' is accessible from G. For $\zeta > 0$, the logistic function in the transition rates implies that such a sequence always exists and irreducibility follows. We then have that a unique invariant distribution μ^ζ exists.

Next, we consider the case of $\zeta = 0$. Let Ψ be the set of nested split graphs and denote $\bar{\Psi} = \Omega \setminus \Psi$. In the following discussion, we show that the networks in $\bar{\Psi}$ are transient. Observe that for any network $G \in \bar{\Psi}$ and $\alpha_i > 0$, there exists a positive probability that in a finite number of consecutive transitions in the Markov chain, links are removed and no links are created until the empty network $\bar{K}_n \in \Omega$ is reached. Let $T < \infty$ be the time when this happens starting from $G \notin \Psi$. Note that $\bar{K}_n \in \Psi$ and, therefore, Proposition 2 implies that all networks $G(t)$, $t > T$, visited by the chain will be in Ψ. A state G is transient if $\int_0^\infty \mathbb{P}(G(t+s) = G | G(t) = G)\, ds < \infty$ (see, e.g., Grimmett and Stirzaker 2001, Chapter 6). We have that $\int_0^\infty \mathbb{P}(G(t+s) = G | G(t) = G)\, ds = \mathbb{E}(\int_0^\infty \mathbb{1}_G(G(s))\, ds | G(t) = G) \leq \mathbb{E}(T) < \infty$. Therefore, all networks that are not nested split graphs are transient and they have vanishing probability in the stationary distribution, i.e., $\mu(\bar{\Psi}) = 0$.

In the following discussion, we show that the set Ψ is a communicating class. Similar to our previous analysis, it holds that for any $G \in \Psi$ and $\alpha_i > 0$, there exists a positive probability that in all consecutive transitions in the Markov chain, links are created and no links are removed until the complete network $K_n \in \Omega$ is reached. Then for $\beta_i > 0$, there exists a positive probability that from K_n only those links decay such that the network G' remains.[42] Therefore, there exists a positive probability to pass from any network G to any other network G' with positive probability, as long as $G, G' \in \Psi$. Similarly, one can show that G is accessible from G'. States G and G' in Ψ are accessible from one another. We say that they communicate and Ψ is a communicating class.

Thus, in the case of $\zeta = 0$, the state space Ω can be partitioned into a communicating class Ψ and a set of transient states $\bar{\Psi}$. The long run behavior of the chain is determined by the states in recurrent class Ψ and we have a unique invariant distribution with $\mu(\Psi) = 1$ (see, e.g., Ethier and Kurtz 1986). □

Before we proceed with the proof of Proposition 4, we introduce the sampled-time Markov chain $(G(t))_{t\in\mathcal{T}}$, $\mathcal{T} \equiv \{0, \Delta t, 2\Delta t, \ldots\}$, associated with the continuous time Markov chain $(G(t))_{t\in\mathbb{R}_+}$ in the limit of $\zeta = 0$ on the same measure space $(\Omega, \mathcal{F})$ (see, e.g., Gallager 1996, Chapter 6). In the sampled-time Markov chain $(G(t))_{t\in\mathcal{T}}$, transitions occur only at discrete times $t \in \mathcal{T}$ separated by (small) increments of size Δt.

Lemma 1. *The continuous time Markov chain $(G(t))_{t\in\mathbb{R}_+}$ and the sampled time Markov chain $(G(t))_{t\in\mathcal{T}}$, $\mathcal{T} \equiv \{0, \Delta t, 2\Delta t, \ldots\}$, $\Delta t \geq 0$, have the same stationary distribution μ on Ω.*

Proof. To see this, consider a probability measure $\mu : \Omega \to [0, 1]$. The stationary distribution of the sampled-time Markov chain satisfies

$$\mu(G) = \sum_{G'\in\Omega} p(G', G)\mu(G') = \sum_{G'\neq G} q(G', G)\Delta t\mu(G') + (1 - q(G, G)\Delta t)\mu(G),$$

which implies the system of equations determining the stationary distribution of the continuous time Markov chain $\mu(G)q(G, G) = \sum_{G'\neq G} q(G', G)\mu(G')$ or, equivalently, $\mu\mathbf{Q} = 0$. □

Hence, so as to investigate the states in the support of the stationary distribution of $(G(t))_{t\in\mathbb{R}_+}$, it suffices to study the stationary distribution of the discrete time Markov chain $(G(t))_{t\in\mathcal{T}}$. Moreover, note that in the limit of $\Delta t \downarrow 0$, also the sample paths of the two chains agree (see, e.g., Gallager 1996, Chapter 6).

One can show that the sampled-time Markov chain on the nested split graphs Ψ (it is enough to require that $G(0) \in \Psi$ such that $G(t) \in \Psi$ for all $t > 0$) is irreducible and aperiodic, and, hence, is ergodic. Moreover, it has a primitive transition matrix $\mathbf{P}$ defined by $(\mathbf{P})_{ij} = \mathbb{P}(G(t + \Delta t) = G_j | G(t) = G_i)$ for any $G_i, G_j \in \Psi$.

[42]From the adjacency matrix associated with the complete network, there exists a sequence of stepwise matrices in which a link to a neighbor with the smallest degree is removed such that any other stepwise matrix can be obtained.

Proof of Proposition 4. In the following discussion, we consider the sampled-time Markov chain $(G(t))_{t\in\mathcal{T}}$ with $\alpha \equiv \alpha_i = 1-\beta_i$ for all $i \in N$. Due to Lemma 1, the stationary distribution of this chain is equivalent to the continuous time Markov chain of Definition 2 as ζ converges to 0. Moreover, because of ergodicity from Proposition 3, we can assume without loss of generality (w.l.o.g.) that $G(0) \in \Psi$. It then follows that $G(t) \in \Psi$ for all $t > 0$ and, therefore, we restrict the state space Ω to the set of nested split graphs Ψ.

At every step $t \in \mathcal{T}$ in the sampled-time Markov chain, a link is created with probability α and a link is removed with probability $1-\alpha$. Further, we consider the complementary chain $(G'(t))_{t\in\mathcal{T}}$ on the same state space Ω, where in every period t, a link is created with probability $\alpha' = 1-\alpha$ and a link is removed with probability $1-\alpha' = \alpha$.[43] This means that a link is removed in $G'(t)$ whenever a link is created in $G(t)$ and a link is created in $G'(t)$ whenever a link is removed in $G(t)$.

As an example, consider the network G represented by the adjacency matrix $\mathbf{A}$ in Figure 13. The complement $\bar{G}$ has an adjacency matrix $\bar{\mathbf{A}}$ obtained from $\mathbf{A}$ by replacing each 1 element in $\mathbf{A}$ by 0 and each 0 element by 1, except for the elements on the diagonal. Let H be the network obtained from G by adding the link 47 (setting $a_{47} = a_{74} = 1$ in $\mathbf{A}$). The probability of this link being created and, thus, the probability of reaching H after the process was in G is $3\alpha/n$, either by selecting one of the two nodes with degree 3 or the node with degree 5 to create a link. Observe that this is identical to the probability of reaching the network $\bar{H}$ from $\bar{G}$ if either the two nodes with degree 7 or the node with degree 4 in $\bar{G}$ are selected to remove a link (with probability $\alpha' = 1-\alpha$).

In general, we can say that for any $G_1, G_2 \in \Omega$, we have that

$$\mathbb{P}(G(t+\Delta t) = G_2 | G(t) = G_1) = \mathbb{P}(G'(t+\Delta t) = \bar{G}_2 | G'(t) = \bar{G}_1). \tag{5}$$

Next consider the stationary distribution μ of $(G(t))_{t\in\mathcal{T}}$ and the corresponding transition matrix $\mathbf{P}$. Similarly, consider the stationary distribution μ' of $(G'(t))_{t\in\mathcal{T}}$ and the corresponding transition matrix $\mathbf{P}'$. Further, consider an ordering of states $G_1, G_2, \ldots$ in Ω and the transition matrix $\mathbf{P}$ with elements $(\mathbf{P})_{ij}$ giving the probability of observing G_j after the Markov chain $(G(t))_{t\in\mathcal{T}}$ was in G_i. Similarly, consider an ordering of states $\bar{G}_1, \bar{G}_2, \ldots$ in Ω and the transition matrix $\mathbf{P}'$ with elements $(\mathbf{P}')_{ij}$ giving the probability of observing $\bar{G}_j$ after the Markov chain $(G'(t))_{t\in\mathcal{T}}$ was in $\bar{G}_i$. Equation (5) implies that $\mathbf{P} = \mathbf{P}'$. Moreover, for the stationary distributions, it must hold that $\mu\mathbf{P} = \mu$ and $\mu'\mathbf{P}' = \mu'$. Since $\mathbf{P}$ is primitive, $\mathbf{P}$ has a unique positive eigenvector and, therefore, $\mu' = \mu$. It follows that for any network $G \in \Omega$ with probability $\mu(G)$, we can take the complement $\bar{G} = G'$ and assign it the probability $\mu(G)$ to get the corresponding probability in μ', i.e., $\mu(G) = \mu'(G')$. □

Before we proceed with the proof of Proposition 5, we state two useful lemmas.

[43]Two nodes of $G'(t)$ are adjacent if and only if they are not adjacent in $G(t)$. Note that the complement of a nested split graph is a nested split graph as well (Mahadev and Peled 1995). In particular, the networks $G'(t)$ are nested split graphs in which the number of nodes in the dominating subsets corresponds to the number of nodes in the independent sets in $G(t)$, and the number of nodes in the independent sets in $G'(t)$ corresponds to the number of nodes in the dominating subsets in $G(t)$. Thus, $(G'(t))_{t\in\mathcal{T}}$ has the same state space Ω as $(G(t))_{t\in\mathcal{T}}$, namely the space Ψ consisting all unlabeled nested split graphs on n nodes.

Lemma 2. *Consider the sampled-time Markov chain* $(G(t))_{t\in\mathcal{T}}$, $\mathcal{T} \equiv \{0, \Delta t, 2\Delta t, \ldots\}$, $\Delta t \equiv 1/n$, *with* $\alpha \equiv \alpha_i = 1 - \beta_i$ *for all* $i \in \mathcal{N}$, $0 < \alpha \leq \frac{1}{2}$, *and restrict the state space* Ω *to the set* Ψ *of all nested split graphs on* n *nodes. For any* $0 \leq d \leq n-1$, *let* X *denote the set of states in* Ω *in which there is exactly one node with degree* $d+1$ *and let* Y *denote the set of states where there is no node with degree* $d+1$. *Denote by* μ_X *the probability of the states in* X *in the stationary distribution* μ *of* $(G(t))_{t\in\mathcal{T}}$ *and by* μ_Y *the probability of states in* Y. *If the number* N_d *of nodes with degree* d *in* Y *is* $\Theta(n)$ *such that* $\lim_{n\to\infty} N_d/n > 0$, *then* $\lim_{n\to\infty} \mu_Y = 0$.[44]

Proof. Let $N(X, Y, y)$ be the expected number of times states in X occur before the process reaches Y (not counting the process as having immediately reached Y if $y \in Y$) when the process starts in y. Then the following relation holds (see Theorem 6.2.3 in Kemeny and Snell 1960 and also Ellison 2000):

$$\frac{\mu_X}{\mu_Y} = N(X, Y, y).$$

Let p_{YX} denote a lower bound on the probability that a state in X occurs after the process is in a state in Y and, conversely, let p_{XY} denote the probability that a state in Y occurs after the process is in a state in X. This probability is the same for all states in X, since from the properties of the Markov chain $(G(t))_{t\in\mathcal{T}}$, it follows that $p_{XY} = 2(1-\alpha)/n$, because there exist two possibilities to remove the link of the node with degree $d+1$ and the probability to select a node for link removal is $(1-\alpha)/n$. Observe that this probability vanishes for large n and that $\lim_{n\to\infty} p_{XY} = 0$. Moreover, we have that

$$\begin{aligned} N(X, Y, y) &\geq p_{YX} p_{XY} + 2 p_{YX}(1-p_{XY}) p_{XY} + 3 p_{YX}(1-p_{XY})^2 p_{XY} + \cdots \\ &= p_{YX} p_{XY} \sum_{i=1}^{\infty} i(1-p_{XY})^{i-1} = \frac{p_{YX}}{p_{XY}}. \end{aligned}$$

The right hand side of the above inequality takes into account the fact that states in X can be reached once, twice, etc. before a state in Y is reached and assigns the corresponding probabilities to compute the expected value.

By assuming that there exists a number N_d of nodes with degree d that is $\Theta(n)$, we have that $p_{YX} \geq \alpha N_d/n$ and $\lim_{n\to\infty} p_{YX} > 0$. It then follows that

$$\frac{\mu_X}{\mu_Y} = N(X, Y, y) \geq \frac{p_{YX}}{p_{XY}} = \frac{\alpha}{2(1-\alpha)} N_d \to \infty \quad \text{as } n \to \infty. \tag{6}$$

Since μ_X is a probability with $\mu_X \leq 1$, (6) implies that $\lim_{n\to\infty} \mu_Y = 0$. $\square$

Lemma 3. *Consider the sampled-time Markov chain* $(G(t))_{t\in\mathcal{T}}$, $\mathcal{T} \equiv \{0, \Delta t, 2\Delta t, \ldots\}$, $\Delta t \equiv 1/n$, *with* $\alpha \equiv \alpha_i = 1 - \beta_i$ *for all* $i \in \mathcal{N}$, *and state space* Ψ *consisting of all nested split graphs on* n *nodes. Then for* $0 < \alpha \leq \frac{1}{2}$, *the asymptotic expected proportion of isolated nodes in the limit of large* n *is given by* $P(0) = (1-2\alpha)/(1-\alpha)$.

[44] By $f = \Theta(g)$ we mean that $0 < \liminf_{n\to\infty} |f(n)/g(n)| \leq \limsup_{n\to\infty} |f(n)/g(n)| < \infty$. In particular, $f = \Theta(1)$ implies that $0 < \lim_{n\to\infty} f(n) < \infty$.

Proof. We consider the expected change in the number of links $m(t)$ in $G(t)$ from t to $t+\Delta t$.[45] The number of links increases by 1 if any node that does not have the maximum degree $n-1$ is selected for creating a link. This happens with probability $\alpha(n-N_{n-1}(t))/n$. The number of links decreases whenever a node with degree higher than 0 is selected for removing a link. This happens with probability $(1-\alpha)(n-N_0(t))/n$. Putting the above contributions together, we can write for the expected change in the total number of links from t to $t+\Delta t$,

$$\mathbb{E}(m(t+\Delta t)|\mathbf{N}(t))-m(t)=\frac{\alpha}{n}(n-N_{n-1}(t))-\frac{1-\alpha}{n}(n-N_0(t)).$$

Taking expectations on both sides of the above equation and denoting $\bar{P}_t(d)\equiv\mathbb{E}(N_d(t)/n)$, we obtain

$$\mathbb{E}(m(t+\Delta t))-\mathbb{E}(m(t))=\alpha(1-\bar{P}_t(n-1))-(1-\alpha)(1-\bar{P}_t(0)). \tag{7}$$

Let ρ denote the initial distribution of states, with $\rho_i=1$ if $G_i=\bar{K}_n$ and 0 otherwise. Further, let $\mathbf{m}$ be the column vector whose jth coordinate, m_j, is the number of links of network $G_j\in\Omega$ and let $G_i=\bar{K}_n$. Then we can write

$$\begin{aligned}\mathbb{E}(m(t))&=\mathbb{E}(m(t)|G(0)=G_i)=\sum_{G_j\in\Omega}\mathbb{P}(G(t)=G_j|G(0)=G_i)m_j=\sum_{G_j\in\Omega}(\mathbf{P}^\top)_{ij}m_j\\&=(\mathbf{P}^\top\mathbf{m})_i=\rho\mathbf{P}^\top\mathbf{m}.\end{aligned}$$

For large times t, the expectation is computed over the invariant distribution μ. In particular, $\lim_{t\to\infty}\rho\mathbf{P}^\top=\mu$ and, therefore, $\lim_{t\to\infty}\mathbb{E}(m(t))=\lim_{t\to\infty}\rho\mathbf{P}^\top\mathbf{m}=\mu\mathbf{m}=\lim_{t\to\infty}\mathbb{E}(m(t+\Delta t))$. Thus, we can set the left hand side of (7) to 0, in the limit of large t, and obtain a relationship between the asymptotic expected proportion of nodes of degree 0 and 1, respectively,

$$1-2\alpha=(1-\alpha)\bar{P}(0)-\alpha\bar{P}(n-1), \tag{8}$$

where we have denoted $\bar{P}(d)=\lim_{t\to\infty}\bar{P}_t(d)$. Next, we consider the chain $(G'(t))_{t\in\mathcal{T}}$, which is constructed from $(G(t))_{t\in\mathcal{T}}$ by taking the complement of each network $G(t)$ in every period t (see also the proof of Proposition 4). In the following discussion, denote the asymptotic expected number of links, $\lim_{t\to\infty}\mathbb{E}(m(t))$, of $(G(t))_{t\in\mathcal{T}}$ by $\bar{m}$ and of $(G'(t))_{t\in\mathcal{T}}$ by $\bar{m}'$. By construction, we must have that $\bar{m}=n(n-1)/2-\bar{m}'$. From Proposition 4, we know that the Markov chain $(G'(t))_{t\in\mathcal{T}}$ has the same stationary distribution μ' as the chain $(G(t))_{t\in\mathcal{T}}$ for a link creation probability of $\alpha'=1-\alpha$. For $\alpha=\frac{1}{2}$, the two processes are identical and we must have that also their expected number of links is the same. This implies that for $\alpha=\frac{1}{2}$, $\bar{m}=\bar{m}'=n(n-1)/4$. The only nested split graph with this number of links, for which the complement has the same number of links as the original graph, is the one in which each independent set is of size 1 and also each dominating subset has size 1 (except possibly for the set corresponding to the $(\lfloor k/2\rfloor+1)$th partition). Thus, for $\alpha=\frac{1}{2}$, it must hold that $\bar{P}(0)=\bar{P}(n-1)=1/n$.

[45] We have that $2m(t)=\sum_{d=0}^{n-1}N_d(t)d$.

Moreover, we know that for $\alpha < \frac{1}{2}$, the expected number of maximally connected nodes (with degree $n-1$) is at most as large as the expected number for $\alpha = \frac{1}{2}$, since the probability of links being created strictly decreases, while the probability of links being removed increases for values of α below $\frac{1}{2}$ (and the probability of a maximally connected node losing a link strictly increases). Thus, $\bar{P}(n-1) \leq 1/n$ for $\alpha \leq \frac{1}{2}$ and for large n, we can write (8) as $1 - 2\alpha = (1-\alpha)\bar{P}(0)$. This is equivalent to

$$\bar{P}(0) = \frac{1-2\alpha}{1-\alpha}.$$

For $\alpha = 0$, no links are created and all nodes are isolated, that is, $\bar{P}(0) = 1$, while for $\alpha = \frac{1}{2}$, the asymptotic expected number of isolated nodes vanishes in the limit of large n. □

With these two lemmas in hand, let us now prove Proposition 5.

Proof of Proposition 5. For the proof of the proposition, it is enough to consider the sampled-time Markov chain $(G(t))_{t\in\mathcal{T}}$ with $\alpha_i = 1 - \beta_i = \alpha$ for all $i \in \mathcal{N}$. Due to Lemma 1, it has the same stationary distribution as the continuous time Markov chain of Definition 2 when ζ converges to 0. Moreover, because of ergodicity from Proposition 3, we can assume w.l.o.g. that $G(0) \in \Psi$ and $G(t) \in \Psi$ for all $t > 0$. We can then restrict the state space Ω to the nested split graphs $\Psi \subset \Omega$. We further assume w.l.o.g. that the step size is given by $\Delta t = 1/n$, which becomes arbitrarily small as n grows.

Note that $G(t)$ is completely determined by $\mathbf{N}(t)$ and vice versa. Thus, it follows that $\{\mathbf{N}(t)\}_{t\in\mathcal{T}}$ is a Markov chain. Denote by $\bar{P}_t(d) \equiv \mathbb{E}(N_d(t)/n)$ the expected proportion of nodes with degree d at time t and let us denote $\bar{P}(d) = \lim_{n\to\infty}\bar{P}_t(d)$; $\bar{P}(d)$ is determined by the invariant distribution μ in the limit of large times t. Lemma 3 shows that (3) holds for $d = 0$. In the following discussion, we show by induction that, given that (3) holds for $\bar{P}(d-1)$ and $\bar{P}(d)$, as n becomes large, also $\bar{P}(d+1)$ satisfies (3) for all $0 \leq d < d^*$, in the limit of large n. For this purpose, we consider (a) the expected number of isolated nodes $\mathbb{E}(N_0(t+\Delta t)|\mathbf{N}(t))$ and (b) the expected number of nodes with degree $d = 1, \dots, d^*$, $\mathbb{E}(N_d(t+\Delta t)|\mathbf{N}(t))$ at time $t + \Delta t$, conditional on the current degree distribution $\mathbf{N}(t)$.

(a) Consider a particular network $G(t)$ in period t generated by $(G(t))_{t\in\mathbb{R}_+}$ and its associated degree distribution $\mathbf{N}(t)$. Figure 14 (left) shows an illustration of the corresponding stepwise matrix. In the following discussion, we compute the expected change of the number $N_0(t)$ of isolated nodes in $G(t)$.

The expected change of $N_0(t)$ due to the creation of a link has the following contributions. An agent with the highest degree k in $N_k(t)$ can create a link to an isolated agent and, thus, decreases the number of isolated agents by 1. The expected change from this link is $-\alpha N_k(t)/n$. On the other hand, if an isolated agent creates a link, then the expected change in the number of isolated agents is $-\alpha N_0(t)/n$.

Moreover, the removal of links can affect $N_0(t)$ if there is only one agent with maximal degree, i.e., $N_k(t) = 1$. In this case, if the agent with the highest degree

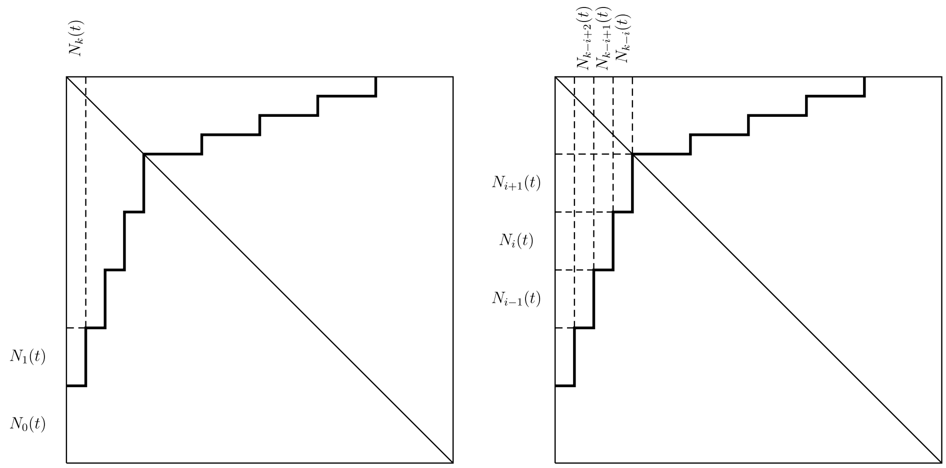


FIGURE 14. (Left) Representation of the stepwise matrix $\mathbf{A}$ of a nested split graph G and some selected degree partitions. The step function separating the 0 entries in the matrix from the 1 entries is shown with a thick line. (Right) Representation of the stepwise matrix $\mathbf{A}$ of a nested split graph G. The step function separating the 0 entries in the matrix from the 1 entries is shown with a thick line.

removes a link, then an additional isolated agent is created, yielding an expected increase in $N_0(t)$ of $(1-\alpha)N_k(t)/n$. Next, if an agent with degree 1 in $N_1(t)$ removes a link, then the number of isolated agents increases. Note that in a nested split graph, $N_1(t) > 0$ implies that $N_k(t) = 1$ and vice versa. This gives an expected change of $N_0(t)$ given by $(1-\alpha)N_1(t)/n$.

Putting the above contributions together, the expected change in the number of isolated nodes at time $t + \Delta t$, conditional on $\mathbf{N}(t)$, is given by the expression[46]

$$\mathbb{E}(N_0(t+\Delta t)|\mathbf{N}(t)) - N_0(t) = -\frac{\alpha}{n}(N_0(t) + N_k(t)) + \frac{1-\alpha}{n}(N_1(t)+1)\delta_{N_k(t),1}. \quad (9)$$

We can take expectations on both sides of (9). For large times t, the expectation is computed on the basis of the invariant distribution μ, and similarly to the proof of Lemma 3, after taking expectations, we can set the left hand side of (9) to 0 for large times t. Note that from Lemma 3, we know that the asymptotic expected proportion $\bar{P}(0)$ of isolated nodes is $\Theta(1)$ for n large. Thus, we can apply the result of Lemma 2 that tells us that the networks in which there does not exist a node with degree 1 have vanishing probability in μ for large n. Since the existence of a node with degree 1 implies that $N_k(t) = 1$, in the limit of large n we can set $\delta_{N_k(t),1} = 1$. We then obtain from (9),

$$\bar{P}(1) = \frac{\alpha}{1-\alpha}\bar{P}(0).$$

This shows that also $\bar{P}(1)$ satisfies (3). Together with Lemma 3, this proves the induction basis.

[46] $\delta_{i,j}$ denotes the usual Kronecker delta, which is 1 if $i = j$ and is 0 otherwise.

(b) We give a proof by induction on the number $N_d(t)$ of nodes with degree $0 < d < d^*$ in a network $G(t)$ in the support of the stationary distribution μ. In the following discussion, we compute the expected change in $N_d(t)$ due to the creation or the removal of a link. An illustration can be found in Figure 14 (right).

Let us investigate the creation of a link. With probability α/n, a link is created from an agent in $N_{k-d}(t)$ to an agent in $N_d(t)$. This yields a contribution to the expected change of $N_d(t)$ of $-\alpha N_{k-d}(t)/n$. If a link is created from an agent in $N_{k-d+1}(t)$ to an agent in $N_d(t)$, then the expected change is α/n if $N_{k-d+1}(t)$ contains only a single agent. Similarly, if a link is created from an agent in $N_{d-1}(t)$ to an agent in $N_d(t)$, then the expected change of $N_d(t)$ is $\alpha N_{d-1}(t)/n$ if $N_{k-d+1}(t) = 1$. Moreover, if an agent in $N_d(t)$ is selected for link creation, then we get an expected decrease of $-\alpha N_d(t)/n$.

Now we consider the removal of a link. If a link is removed from the agent in $N_{k-d+1}(t)$ to an agent in $N_d(t)$, then the expected change of $N_d(t)$ is $-(1-\alpha)N_{k-d+1}(t)/n$. If a link is removed from an agent in $N_{k-d}(t)$ to an agent in $N_{d+1}(t)$, then the expected increase of $N_d(t)$ is $(1-\alpha)/n$ if $N_{k-d}(t) = 1$. Moreover, if an agent in $N_{d+1}(t)$ is selected for removing a link, then we get an expected increase of $(1-\alpha)N_{i+d}(t)/n$ if $N_{k-d}(t) = 1$. Finally, if an agent in $N_d(t)$ is selected for removing a link, then we get an expected change of $-(1-\alpha)N_d(t)/n$.

Putting the above contributions together, the expected change in $N_d(t)$ is given by

$$\begin{aligned}
&\mathbb{E}(N_d(t+\Delta t)|\mathbf{N}(t)) - N_d(t) \\
&\quad = \frac{\alpha}{n}\big(-N_d(t) + (N_{d-1}(t)+1)\delta_{N_{k-d+1}(t),1} - N_{k-d}(t)\big) \\
&\qquad\qquad + \frac{1-\alpha}{n}\big(-N_d(t) + (N_{d+1}(t)+1)\delta_{N_{k-d}(t),1} - N_{k-d+1}(t)\big).
\end{aligned} \tag{10}$$

We can take expectations on both sides of (10), and similarly to part (a) of this proof, we can set the left hand side of (10) as t becomes large. For large times t, the above expectation is computed on the basis of the invariant distribution μ. By the induction assumption, the asymptotic expected proportion $\bar{P}(d-1)$ of nodes with degree $d-1$ is $\Theta(1)$ in the limit of large n (as follows from (3)). Thus we can apply Lemma 2 and neglect the networks in which there does not exist a node with degree d since they have vanishing probability in μ for large n. Similarly, we know from the induction assumption that the asymptotic proportion $\bar{P}(d)$ of nodes with degree d is $\Theta(1)$ and, by virtue of Lemma 2, we know that the networks in which there does not exist a node with degree $d+1$ have vanishing probability in μ for large n. Thus, in the limit of large n, we can set $\delta_{N_{k-d+1}(t),1} = \delta_{N_{k-d}(t),1} = 1$, since the existence of nodes with degrees d and $d+1$ implies that $N_{k-d+1}(t) = N_{k-d}(t) = 1$ in the limit of large t and n. Therefore, we get from (10) the relationship

$$\bar{P}(d+1) = \frac{1}{1-\alpha}\bar{P}(d) - \frac{\alpha}{1-\alpha}\bar{P}(d-1). \tag{11}$$

Inserting the expressions for $\bar{P}(d-1)$ and $\bar{P}(d)$ from (3) into (11) yields

$$\bar{P}(d+1) = \frac{1}{1-\alpha}\frac{1-2\alpha}{1-\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{d} - \frac{\alpha}{1-\alpha}\frac{1-2\alpha}{1-\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{d-1} = \frac{1-2\alpha}{1-\alpha}\left(\frac{\alpha}{1-\alpha}\right)^{d+1}.$$

Thus, (3) also holds for $\bar{P}(d+1)$. This proves the induction step.

Finally, we have that the degree distribution must be normalized to 1, i.e., $\sum_{d=0}^{n-1}\bar{P}(d)=1$. We know that the number of agents in the dominating subsets with degrees larger than d^* is d^* (since each set contains only one node and there are d^* such sets).[47] Adding this to the number of agents in the independent sets with degree $d=0,\dots,d^*$ yields $n\sum_{d=0}^{d^*}\bar{P}(d)+d^*=n$. Further, inserting (3), we can derive the number d^* of independent sets as a function of n and α:

$$d^*(n,\alpha) = \frac{\ln\left(\frac{2(1-\alpha)}{(1-2\alpha)n}\right)}{\ln\left(\frac{\alpha}{1-\alpha}\right)}.$$

d^* is a monotonic decreasing function of n for a fixed value of α. Conversely, for a fixed value of n, we get the limits $\lim_{\alpha\to 0} d^* = 0$ and $\lim_{\alpha\to 1/2} d^* = n/2$.

We finish the proof with the following observation, showing that the empirical degree distribution concentrates around its expected value in the limit of large n when $\Delta t = 1/n$. More precisely, for any $\varepsilon > 0$, we have that $\mathbb{P}(|P_t(d)-\mathbb{E}(P_t(d))|\geq\varepsilon)\leq 2e^{-\varepsilon^2 n^2\Delta t/(8t)}$. To see this, let us define the random variable $Y_d(s)\equiv\mathbb{E}(N_d(t)|\mathbf{N}(s))$, $s\in\mathcal{T}$. Since $\{\mathbf{N}(t), t\in\mathcal{T}\}$ is a Markov chain, the sequence $\{Y_d(s), s\in\mathcal{T}, s\leq t\}$ is a Martingale with respect to $\{\mathbf{N}(t), t\in\mathcal{T}\}$.[48] Moreover, the change in the number of nodes with degree d per period t is bounded by 2, i.e. $|N_d(t)-N_d(t-\Delta t)|\leq 2$, since at most one link is added or removed in every period t and this can change the degrees of at most two nodes. Therefore, we can apply Hoeffding's inequality (see, e.g., Theorem 3, Section 12.2 in Grimmett and Stirzaker 2001), which states that for any $0<s\leq t$ with $|Y(s)-Y(s-\Delta t)|\leq c$ and any $\varepsilon>0$, $\mathbb{P}(|Y(t)-Y(0)|>\varepsilon)\leq 2e^{-\varepsilon^2\Delta t/(2tc^2)}$. With $c=2$, $Y(t)=\mathbb{E}(N_d(t)|\mathbf{N}(t))=N_d(t)$, $Y(0)=\mathbb{E}(N_d(t)|\mathbf{N}(0))=\mathbb{E}(N_d(t))$, and $\Delta t=1/n$, it then follows that

$$\begin{aligned}\mathbb{P}\left(\left|\frac{N_d(t)}{n}-\mathbb{E}\left(\frac{N_d(t)}{n}\right)\right|\geq\varepsilon\right) &= \mathbb{P}\left(\left|N_d(t)-\mathbb{E}(N_d(t))\right|\geq n\varepsilon\right)\leq 2e^{-\varepsilon^2 n^2\Delta t/(8t)}\\ &= 2e^{-\varepsilon^2 n/(8t)}\to 0\end{aligned} \tag{12}$$

as $n\to\infty$. This implies that the empirical proportion $N_d(t)/n$ of nodes with degree d converges in probability to its expected value $\mathbb{E}(N_d(t)/n)$ as n becomes large.

Since $(\Omega,\mathcal{F},\mathbb{P})$ is a discrete probability space, this also implies convergence almost surely. To see this, let $\mathcal{A}_n\equiv\{G\in\Omega:|\frac{N_d(t)}{n}-\mathbb{E}(\frac{N_d(t)}{n})|\geq\varepsilon\}$. By (12), we have that

[47]Note that since networks in which there does not exist a node with degree $0\leq d\leq d^*$ in the corresponding independent set can be neglected, the structure of nested split graphs implies that all dominating subsets have size 1.

[48]We have that $\mathbb{E}(Y_d(s)|\mathbf{N}(s-\Delta t))=\mathbb{E}(\mathbb{E}(N_d(t)|\mathbf{N}(s))|\mathbf{N}(s-\Delta t))=\mathbb{E}(N_d(t)|\mathbf{N}(s-\Delta t))=Y_d(s-\Delta t)$. Further, one can show that the first and second moments of $\{Y_d(s), s\leq t\}$ are bounded. Thus, $\{Y_d(s), s\leq t\}$ is a Martingale with respect to $\{\mathbf{N}(t)\}$ for $s,t\in\mathcal{T}$ (see, e.g., Grimmett and Stirzaker 2001, Chapter 12).

$\lim_{n\to\infty}\mathbb{P}_t(\mathcal{A}_n)=0$. Then there exists and $n_0\in\mathbb{N}$ such that $\mathbb{P}_t(\mathcal{A}_n)<\mathbb{P}_t(G)$ for all $G\in\Omega$ with $\mathbb{P}_t(G)>0$ and $n>n_0$. Hence, for all $n>n_0$, we have that $\{G\in\Omega:\mathbb{P}_t(G)>0\}\notin\mathcal{A}_n$, $\{G\in\Omega:\mathbb{P}_t(G)>0\}\cap\mathcal{A}_n=\varnothing$ and, therefore, $\mathbb{P}_t(\bigcap_{n=1}^{\infty}\bigcup_{m=n}^{\infty}\mathcal{A}_m)=0$. $\square$

The proof of Corollary 2 follows directly from the functional form of $d^*(n,\alpha)$ in Proposition 5.

For the proof of Corollary 3, see Theorem 1.2.4 in Mahadev and Peled (1995).

Appendix B: Network definitions and characterizations

A *network* (*graph*) G is the pair $(\mathcal{N},\mathcal{E})$ consisting of a set of *nodes* (*vertices*) $\mathcal{N}=\{1,\dots,n\}$ and a set of edges (*links*) $\mathcal{E}\subset\mathcal{N}\times\mathcal{N}$ between them. A link (i,j) is *incident* with nodes i and j. The *neighborhood* of a node $i\in\mathcal{N}$ is the set $\mathcal{N}_i=\{j\in\mathcal{N}:(i,j)\in\mathcal{E}\}$. The *degree* d_i of a node $i\in\mathcal{N}$ gives the number of links incident to node i. Clearly, $d_i=|\mathcal{N}_i|$. Let $\mathcal{N}_i^{(2)}=\bigcup_{j\in\mathcal{N}_i}\mathcal{N}_j\setminus(\mathcal{N}_i\cup\{i\})$ denote the second-order neighbors of node i. Similarly, the kth order neighborhood of node i is defined recursively from $\mathcal{N}_i^{(0)}=\{i\}$, $\mathcal{N}_i^{(1)}=\mathcal{N}_i$, and $\mathcal{N}_i^{(k)}=\bigcup_{j\in\mathcal{N}_i^{(k-1)}}\mathcal{N}_j\setminus(\bigcup_{l=0}^{k-1}\mathcal{N}_i^{(l)})$. A *walk* in G of length k from i to j is a sequence $\langle i_0,i_1,\dots,i_k\rangle$ of nodes such that $i_0=i$, $i_k=j$, and $i_p\neq i_{p+1}$, and i_p and i_{p+1} are (directly) linked, that is, $i_pi_{p+1}\in\mathcal{E}$ for all $0\leq p\leq k-1$. Nodes i and j are said to be *indirectly linked* in G if there exists a walk from i to j in G containing nodes other than i and j. A pair of nodes i and j is connected if they are either directly or indirectly linked. A node $i\in\mathcal{N}$ is *isolated* in G if $\mathcal{N}_i=\varnothing$. The network G is said to be *empty* (denoted by $\bar{K}_n$) when all its nodes are isolated.

A *subgraph*, G', of G is the graph of subsets of the nodes, $\mathcal{N}(G')\subseteq\mathcal{N}(G)$, and links, $\mathcal{E}(G')\subseteq\mathcal{E}(G)$. A graph G is *connected* if there is a path connecting every pair of nodes. Otherwise G is disconnected. The *components* of a graph G are the maximally connected subgraphs. A component is said to be *minimally connected* if the removal of any link makes the component disconnected.

A *dominating set* for a graph $G=(\mathcal{N},\mathcal{E})$ is a subset $\mathcal{S}$ of $\mathcal{N}$ such that every node not in $\mathcal{S}$ is connected to at least one member of S by a link. An *independent set* is a set of nodes in a graph in which no two nodes are adjacent. For example, the central node in a star $K_{1,n-1}$ forms a dominating set while the peripheral nodes form an independent set.

In a *complete* graph K_n, every node is adjacent to every other node. The graph in which no pair of nodes is adjacent is the empty graph $\bar{K}_n$. A *clique* $K_{n'}$, $n'\leq n$, is a complete subgraph of the network G. A graph is *k-regular* if every node i has the same number of links $d_i=k$ for all $i\in\mathcal{N}$. The complete graph K_n is $(n-1)$-regular. The cycle C_n is 2-regular. In a *bipartite graph*, there exists a partition of the nodes into two disjoint sets V_1 and V_2 such that each link connects a node in V_1 to a node in V_2. V_1 and V_2 are independent sets with cardinalities n_1 and n_2, respectively. In a complete bipartite graph K_{n_1,n_2}, each node in V_1 is connected to each other node in V_2. The *star* $K_{1,n-1}$ is a complete bipartite graph in which $n_1=1$ and $n_2=n-1$.

The *complement* of a graph G is a graph $\bar{G}$ with the same nodes as G such that any two nodes of $\bar{G}$ are adjacent if and only if they are not adjacent in G. For example, the complement of the complete graph K_n is the empty graph $\bar{K}_n$.

Let $\mathbf{A}$ be the symmetric $n \times n$ *adjacency matrix* of the network G. The element $a_{ij} \in \{0, 1\}$ indicates whether there exists a link between nodes i and j such that $a_{ij} = 1$ if $(i, j) \in \mathcal{E}$ and $a_{ij} = 0$ if $(i, j) \notin \mathcal{E}$. The kth power of the adjacency matrix is related to walks of length k in the graph. In particular, $(\mathbf{A}^k)_{ij}$ gives the number of walks of length k from node i to node j. The *eigenvalues* of the adjacency matrix $\mathbf{A}$ are the numbers $\lambda_1, \lambda_2, \ldots, \lambda_n$ such that $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{v}_i$ has a nonzero solution vector $\mathbf{v}_i$, which is an *eigenvector* associated with λ_i for $i = 1, \ldots, n$. Since the adjacency matrix $\mathbf{A}$ of an undirected graph G is real and symmetric, the eigenvalues of $\mathbf{A}$ are real, $\lambda_i \in \mathbb{R}$ for all $i = 1, \ldots, n$. Moreover, if $\mathbf{v}_i$ and $\mathbf{v}_j$ are eigenvectors for different eigenvalues, $\lambda_i \neq \lambda_j$, then $\mathbf{v}_i$ and $\mathbf{v}_j$ are orthogonal, i.e., $\mathbf{v}_i^\top \mathbf{v}_j = 0$ if $i \neq j$. In particular, $\mathbb{R}^n$ has an orthonormal basis consisting of eigenvectors of $\mathbf{A}$. Since $\mathbf{A}$ is a real symmetric matrix, there exists an orthogonal matrix $\mathbf{S}$ such that $\mathbf{S}^\top\mathbf{S} = \mathbf{S}\mathbf{S}^\top = \mathbf{I}$ (that is $\mathbf{S}^\top = \mathbf{S}^{-1}$) and $\mathbf{S}^\top\mathbf{A}\mathbf{S} = \mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix of eigenvalues of $\mathbf{A}$ and the columns of $\mathbf{S}$ are the corresponding eigenvectors. The *Perron–Frobenius eigenvalue* $\lambda_{\mathrm{PF}}(G)$ is the *largest real eigenvalue* of $\mathbf{A}$ associated with G, i.e., all eigenvalues λ_i of $\mathbf{A}$ satisfy $|\lambda_i| \leq \lambda_{\mathrm{PF}}(G)$ for $i = 1, \ldots, n$ and there exists an associated nonnegative eigenvector $\mathbf{v}_{\mathrm{PF}} \geq 0$ such that $\mathbf{A}\mathbf{v}_{\mathrm{PF}} = \lambda_{\mathrm{PF}}(G)\mathbf{v}_{\mathrm{PF}}$. For a connected graph G, the adjacency matrix $\mathbf{A}$ has a unique largest real eigenvalue $\lambda_{\mathrm{PF}}(G)$ and a positive associated eigenvector $\mathbf{v}_{\mathrm{PF}} > 0$. There exists a relation between the number of walks in a graph and its eigenvalues. The number of closed walks of length k from a node i in G to herself is given by $(\mathbf{A}^k)_{ii}$ and the total number of closed walks of length k in G is $\mathrm{tr}(\mathbf{A}^k) = \sum_{i=1}^n (\mathbf{A}^k)_{ii} = \sum_{i=1}^n \lambda_i^k$. We further have that $\mathrm{tr}(\mathbf{A}) = 0$, $\mathrm{tr}(\mathbf{A}^2)$ gives twice the number of links in G and $\mathrm{tr}(\mathbf{A}^3)$ gives six times the number of triangles in G.

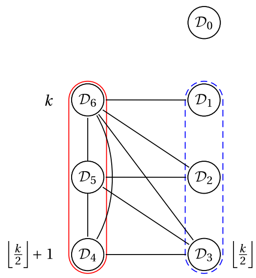
Appendix C: Topological properties of nested split graphs

In this appendix, we discuss in more detail the topological properties of nested split graphs that arise from our network formation process. We first derive several network statistics for nested split graphs. We compute the degree distribution, the clustering coefficient, average nearest neighbor connectivity, and the characteristic path length in a nested split graph. In particular, we show that connected nested split graphs have small characteristic path length, which is at most 2. We then analyze different measures of centrality in a nested split graph.[49] From the expressions of these centrality measures, we then can show that degree, closeness, eigenvector, and Bonacich centrality induce the same ordering of nodes in a nested split graph. If the ordering is not strict, then this holds also for betweenness centrality. Finally, for all statistics derived in this section, we show that they are all completely determined by the degree partition in a nested split graph.

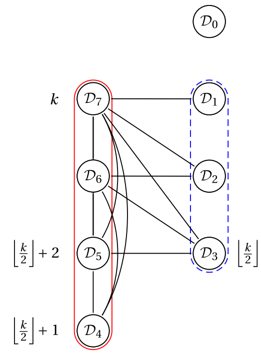
C.1 *Network statistics*

In the following sections we will compute the degree connectivity, the clustering coefficient, assortativity and average nearest neighbor connectivity, and the characteristic path length in a nested split graph G as a function of the degree partition $\mathcal{D}$ (introduced in Definition 3).

[49] See Wasserman and Faust (1994, Chapter 5.2) for an overview of different measures of centrality.



$$\mathbf{A} = \begin{pmatrix} 0&1&1&1&1&1&1&1&1&1\\ 1&0&1&1&1&1&1&1&1&1\\ 1&1&0&1&1&1&1&1&0&0\\ 1&1&1&0&1&1&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0 \end{pmatrix}$$



$$\mathbf{A} = \begin{pmatrix} 0&1&1&1&1&1&1&1&1&1\\ 1&0&1&1&1&1&1&1&1&0\\ 1&1&0&1&1&1&1&1&0&0\\ 1&1&1&0&1&0&0&0&0&0\\ 1&1&1&1&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&1&0&0&0&0&0&0&0\\ 1&1&0&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0 \end{pmatrix}$$

Figure 15. Representation of nested split graphs and their degree partitions $\mathcal{D}$ with corresponding adjacency matrices $\mathbf{A}$. A line between $\mathcal{D}_i$ and $\mathcal{D}_j$ indicates that every node in $\mathcal{D}_i$ is adjacent to every node in $\mathcal{D}_j$. The partitions included in the solid frame ($\mathcal{D}_i$ with $\lfloor k/2 \rfloor + 1 \leq i \leq k$) are the dominating subsets, while the partitions in the dashed frame ($\mathcal{D}_i$ with $1 \leq i \leq \lfloor k/2 \rfloor$) are the independent sets. The figure on the left considers the case of $k = 6$ (even) and the figure on the right considers the case of $k = 7$ (odd). The illustration follows Mahadev and Peled (1995, p. 11).

C.1.1 *Degree connectivity* The nested neighborhood structure of a nested split graph allows us to compute the degrees of the nodes according to a recursive equation that is stated in the next corollary.

Corollary 3. *Consider a nested split graph $G = (\mathcal{N}, \mathcal{E})$ and let $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ be the degree partition of G. Then $d_u = 0$ if $u \in \mathcal{D}_0$ and for each $u \in \mathcal{D}_i$, $v \in \mathcal{D}_{i-1}$, $i = 1, \ldots, k$, we get*

$$d_u = \begin{cases} d_v + |\mathcal{D}_{k-i+1}| & \text{if } i \neq \left\lfloor \dfrac{k}{2} \right\rfloor + 1 \\ d_v + |\mathcal{D}_{k-i+1}| - 1 & \text{if } i = \left\lfloor \dfrac{k}{2} \right\rfloor + 1 \end{cases} \tag{13}$$

or, equivalently,

$$d_u = \begin{cases} \displaystyle\sum_{j=1}^{i} |\mathcal{D}_{k+1-j}| & \text{if } 1 \leq i \leq \left\lfloor \dfrac{k}{2} \right\rfloor \\ \displaystyle\sum_{j=1}^{i} |\mathcal{D}_{k+1-j}| - 1 & \text{if } \left\lfloor \dfrac{k}{2} \right\rfloor + 1 \leq i \leq k. \end{cases} \tag{14}$$

Equation (13) shows that the neighborhoods of the agents in a nested split graph are nested (see also Definition 4). The degrees of the agents in ascending order of the graph in Figure 15, left, are 2, 3, 4, 5, 7, 9, while in the graph in Figure 15, right, they are 1, 2, 3, 4, 7, 8, 9.

C.1.2 *Clustering coefficient* The clustering coefficient $C(u)$ for an agent u is the proportion of links between the agents within her neighborhood $\mathcal{N}_u$ divided by the number of links that could possibly exist between them (Watts and Strogatz 1998). It is given by

$$C(u) \equiv \frac{|\{vw : v, w \in \mathcal{N}_u \wedge vw \in \mathcal{E}\}|}{d_u(d_u-1)/2}.$$

In a nested split graph, the clustering coefficient can be derived from the degree partition, as the following corollary shows.

Corollary 4. *Consider a nested split graph $G = (\mathcal{N}, \mathcal{E})$ and let $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \dots, \mathcal{D}_k)$ be the degree partition of G. Denote by $S^i_{\mathcal{D}} = \sum_{j=i}^{k} |\mathcal{D}_j|$. Then for each $u \in \mathcal{D}_i$, $i = 0, \dots, k$, and $d_u \geq 2$, the clustering coefficient is given by*

$$C(u) = \begin{cases} 0 & \text{if } i = 0 \\ 1 & \text{if } 1 \leq i \leq \left\lfloor \frac{k}{2} \right\rfloor \\ \frac{1}{d_u(d_u-1)}(S^{\lfloor k/2 \rfloor+1}_{\mathcal{D}} - 1)[(S^{\lfloor k/2 \rfloor+1}_{\mathcal{D}} - 2) + 2|\mathcal{D}_{\lfloor k/2 \rfloor}|] & \text{if } i = \left\lfloor \frac{k}{2} \right\rfloor + 1,\ k \text{ even} \\ \frac{1}{d_u(d_u-1)}(S^{\lfloor k/2 \rfloor+1}_{\mathcal{D}} - 1)(S^{\lfloor k/2 \rfloor+1}_{\mathcal{D}} - 2) & \text{if } i = \left\lfloor \frac{k}{2} \right\rfloor + 1,\ k \text{ odd} \\ \frac{1}{d_u(d_u-1)}\left[(S^{\lfloor k/2 \rfloor+1}_{\mathcal{D}} - 1)(S^{\lfloor k/2 \rfloor+1}_{\mathcal{D}} - 2) \right. & \\ \qquad \left. + 2\sum_{j=k-i+1}^{\lfloor k/2 \rfloor} |\mathcal{D}_j|(S^{k-j+1}_{\mathcal{D}} - 1)\right] & \text{if } \left\lfloor \frac{k}{2} \right\rfloor + 2 < i \leq k, \end{cases} \tag{15}$$

where d_u is given by (14).

Proof. Note that for all agents in the independent sets, $u \in \mathcal{D}_i$ with $1 \leq i \leq \lfloor k/2 \rfloor$, the clustering coefficient is 1, since their neighbors are all connected among each other. Next, we consider the agents $u \in \mathcal{D}_i$ with $\lfloor k/2 \rfloor + 1 \leq i \leq k$ and degree $d_u = \sum_{j=1}^{i} |\mathcal{D}_{k+1-j}| - 1$. The neighbors of agent u in the dominating subsets are all connected among each other with a total of $\frac{1}{2}(\sum_{j=\lfloor k/2 \rfloor+1}^{k} |\mathcal{D}_j| - 1)(\sum_{j=\lfloor k/2 \rfloor+1}^{k} |\mathcal{D}_j| - 2)$ links, excluding agent u from the dominating subset. The neighbors of u in the independent sets are not connected. Finally, we consider the links between neighbors for which one neighbor is in a dominating subset and one neighbor is in an independent set. If k is even, we get $\sum_{j=k-i+1}^{\lfloor k/2 \rfloor} |\mathcal{D}_j|(\sum_{l=k-j+1}^{k} |\mathcal{D}_l| - 1)$ links, excluding agent u in the dominating subset (see Figure 15 (left)). If k is odd, there is no such contribution for the agents in the set $\mathcal{D}_{\lfloor k/2 \rfloor+1}$ (see Figure 15 (right)). Putting these contributions together, we obtain the clustering coefficient of an agent $u \in \mathcal{D}_i$ for all $i = 1, \dots, k$, as given by (15). □

The total clustering coefficient is the average of the clustering coefficients over all agents, $C \equiv (1/n)\sum_{u \in \mathcal{N}} C(u)$. The clustering coefficients of the agents in ascending order of the graph in Figure 15 (left) are $\frac{5}{12}$, $\frac{5}{12}$, $\frac{13}{21}$, $\frac{9}{10}$, 1, 1, 1, 1, 1, 1, with a total clustering coefficient of $C=0.84$. In the graph in Figure 15 (right) it is $\frac{13}{36}$, $\frac{13}{28}$, $\frac{4}{7}$, 1, 1, 1, 1, 1, 1, 1, with a total clustering of $C=0.74$.

C.1.3 *Assortativity and nearest neighbor connectivity* There exists a measure of degree correlation called *average nearest neighbor connectivity* (Pastor-Satorras et al. 2001). More precisely, the average nearest neighbor connectivity $d_{nn}(u)$ is the average degree of the neighbors of an agent with degree d_u. It is defined by

$$d_{nn}(u) \equiv \frac{1}{d_u} \sum_{v \in \mathcal{N}_u} d_v.$$

In a nested split graph, the average nearest neighbor connectivity is determined by its degree partition.

Corollary 5. *Consider a nested split graph $G=(\mathcal{N},\mathcal{E})$ and let $\mathcal{D}=(\mathcal{D}_0,\mathcal{D}_1,\dots,\mathcal{D}_k)$ be the degree partition of G. Denote $S^i_{\mathcal{D}}=\sum_{j=1}^i |\mathcal{D}_{k+1-j}|$. Then for each $u\in\mathcal{D}_i$, $i=0,\dots,k$,*

$$d_{nn}(u)=\begin{cases} 0 & \text{if } i=0 \\ \dfrac{1}{S^i_{\mathcal{D}}}\displaystyle\sum_{j=1}^{i}|\mathcal{D}_{k+1-j}|(S^{k+1-j}_{\mathcal{D}}-1) & \text{if } i=1,\dots,\left\lfloor \dfrac{k}{2}\right\rfloor \\ \dfrac{1}{S^{\lfloor k/2\rfloor+1}_{\mathcal{D}}-1}\Bigg[\displaystyle\sum_{j=\lfloor k/2\rfloor+2}^{k}|\mathcal{D}_j|(S^j_{\mathcal{D}}-1) & \\ \qquad +(|\mathcal{D}_{\lfloor k/2\rfloor+1}|-1)(S^{\lfloor k/2\rfloor+1}_{\mathcal{D}}-1)+|\mathcal{D}_{\lfloor k/2\rfloor}|S^{\lfloor k/2\rfloor}_{\mathcal{D}}\Bigg] & \text{if } i=\left\lfloor \dfrac{k}{2}\right\rfloor+1,\ k \text{ even} \\ \dfrac{1}{S^{\lfloor k/2\rfloor+1}_{\mathcal{D}}-1}\Bigg[\displaystyle\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|(S^j_{\mathcal{D}}-1)\Bigg]-1 & \text{if } i=\left\lfloor \dfrac{k}{2}\right\rfloor+1,\ k \text{ odd} \\ \dfrac{1}{S^i_{\mathcal{D}}-1}\Bigg[\displaystyle\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|(S^j_{\mathcal{D}}-1)+\sum_{j=k-i+1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|S^j_{\mathcal{D}}\Bigg]-1 & \text{if } i=\left\lfloor \dfrac{k}{2}\right\rfloor+2,\dots,k. \end{cases}$$

Proof. First, consider an agent $u\in\mathcal{D}_i$ with $i=1,\dots,\lfloor k/2\rfloor$ corresponding to the independent sets. We know that the number of neighbors (degree) of agent u is given by $\sum_{j=1}^i|\mathcal{D}_{k+1-j}|$. The neighbors of agent u are the agents in the dominating subsets with degrees given in (14). Thus, the number of neighbors of the neighbors of u in the sets $\mathcal{D}_{k+1-j}$ is $\sum_{l=1}^{k+1-j}|\mathcal{D}_{k+1-l}|-1$. Putting the above results together, we obtain for the average nearest neighbor connectivity of agent $u\in\mathcal{D}_i$, $i=1,\dots,\lfloor k/2\rfloor$, the expression

$$d_{nn}(u)=\frac{1}{\sum_{j=1}^i|\mathcal{D}_{k+1-j}|}\sum_{j=1}^{i}|\mathcal{D}_{k+1-j}|\left(\sum_{l=1}^{k+1-j}|\mathcal{D}_{k+1-l}|-1\right).$$

Next, we consider an agent u in the set $\mathcal{D}_i$ with $\lfloor k/2\rfloor+2\le i\le k$ corresponding to the dominating subsets. The number of neighbors of agent u is given by $\sum_{j=1}^i|\mathcal{D}_{k+1-j}|-1$. The number of neighbors of an agent $v\in\mathcal{D}_j$, $\lfloor k/2\rfloor+1\le j\le k$, in the dominating subsets is given by $\sum_{l=1}^{j}|\mathcal{D}_{k+1-l}|-1$. Since agent u is connected to all other agents in the dominating subsets, we can sum over all their neighborhoods with a total of $\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|(\sum_{l=1}^{j}|\mathcal{D}_{k+1-l}|-1)$ neighbors. Note, however, that we have to subtract agent u herself from this sum. Moreover, the number of neighbors of an agent $w\in\mathcal{D}_j$, $1\leq j\leq\lfloor k/2\rfloor$ in the independent sets is given by $\sum_{l=1}^{j}|\mathcal{D}_{k+1-l}|$. Thus, the average nearest neighbor connectivity of agent $u\in\mathcal{D}_i$, $\lfloor k/2\rfloor+2\leq i\leq k$, is given by

$$d_{nn}(u)=\frac{1}{\sum_{j=1}^{i}|\mathcal{D}_{k+1-j}|-1}\left[\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|\left(\sum_{l=1}^{j}|\mathcal{D}_{k+1-l}|-1\right)+\sum_{j=k-i+1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|\sum_{l=1}^{j}|\mathcal{D}_{k+1-l}|\right]-1.$$

In a similar way we can consider the cases $i=\lfloor k/2\rfloor+1$ for both k even and k odd. □

When the average nearest neighbor connectivity is a monotonic increasing function of the degree d, then the network is assortative, while if it is monotonic decreasing with d, it is dissortative (Pastor-Satorras et al. 2001, Newman 2002). Nested split graphs are dissortative, since for $i<j$ and $d_u\in\mathcal{D}_i<d_v\in\mathcal{D}_j$, it follows that $d_{nn}(u)>d_{nn}(v)$. This is because the higher is the degree of an agent in a dominating subset, the more neighbors she has from the independent sets with low degrees, which decreases her average nearest neighbor connectivity. For example, the average nearest neighbor connectivities of the agents in the graph in Figure 15 (left) in ascending order are $\frac{13}{3}$, $\frac{13}{3}$, $\frac{37}{7}$, $\frac{33}{5}$, $\frac{15}{2}$, $\frac{15}{2}$, $\frac{25}{3}$, $\frac{25}{3}$, 9, 9, while in the graph in Figure 15 (right), they are $\frac{35}{9}$, $\frac{35}{8}$, $\frac{34}{7}$, 7, 7, 8, 8, 8, $\frac{17}{2}$, 9.

C.1.4 *Characteristic path length* The characteristic path length is defined as the number of links in the shortest path between two agents, averaged over all pairs of agents (Watts and Strogatz 1998). This can be written as

$$\ell(G)\equiv\frac{1}{n(n-1)/2}\sum_{u\neq v\in G}d(u,v),\tag{16}$$

where $d(u,v)$ is the geodesic (shortest path) between agent u and agent v in $N\setminus\mathcal{D}_0$.[50] Then the characteristic path length in a nested split graph is given by the following corollary.

Corollary 6. *Consider a nested split graph $G=(\mathcal{N},\mathcal{E})$ and let $\mathcal{D}=(\mathcal{D}_0,\mathcal{D}_1,\dots,\mathcal{D}_k)$ be the degree partition of G. Then the characteristic path length of G is given by*

$$\ell(G)=\frac{1}{n(n-1)/2}\left[\frac{1}{2}\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|\left(\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|-1\right)+\sum_{j=1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|\left(\sum_{j=1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|-1\right)+\sum_{l=1}^{\lfloor k/2\rfloor}|\mathcal{D}_l|\left(\sum_{j=k-l+1}^{k}|\mathcal{D}_j|+2\sum_{j=\lfloor k/2\rfloor+1}^{k-l}|\mathcal{D}_j|\right)\right].$$

[50]Note that we do not consider the isolated agents in the set $\mathcal{D}_0$ because the characteristic path length $\ell(G)$ is not defined for disconnected networks G.

Proof. We first consider all pairs of agents in the dominating subsets. All these agents are adjacent to each other and thus the shortest path between them has length 1. Moreover, there are $\frac{1}{2}\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|(\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|-1)$ pairs of agents in the dominating subsets.

Next, we consider all pairs of agents in the independent sets. From (17), we know that all of them are at a distance of two links separated from each other. Moreover, there are $\frac{1}{2}\sum_{j=1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|(\sum_{j=1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|-1)$ pairs of agents in which both agents stem from an independent set.

Further, consider the pairs of agents in which one agent is in the independent set $\mathcal{D}_1$ and the other is in a dominating subset. Then there are $|\mathcal{D}_1||\mathcal{D}_k|$ pairs of agents with shortest path 1 and $|\mathcal{D}_1|\sum_{j=\lfloor k/2\rfloor+1}^{k-1}|\mathcal{D}_j|$ pairs of agents with shortest path 2. Similarly, we can consider the pairs in which one agent is in the set $\mathcal{D}_2$. Then we have $|\mathcal{D}_2|(|\mathcal{D}_k|+|\mathcal{D}_{k-1}|)$ pairs of agents with shortest path 1 and $|\mathcal{D}_2|\sum_{j=\lfloor k/2\rfloor+1}^{k-2}|\mathcal{D}_j|$ pairs of agents with shortest path 2. Finally, if one agent is in the set $\mathcal{D}_{\lfloor k/2\rfloor}$, then we have $|\mathcal{D}_{\lfloor k/2\rfloor}|\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|$ pairs of agents with distance 1 and none with distance 2 if k is even (see Figure 15 (left)). If k is odd (see Figure 15 (right)) and we have one agent in the set $\mathcal{D}_{\lfloor k/2\rfloor}$, then we have $|\mathcal{D}_{\lfloor k/2\rfloor}|\sum_{j=\lfloor k/2\rfloor+2}^{k}|\mathcal{D}_j|$ pairs of agents with distance 1 and $|\mathcal{D}_{\lfloor k/2\rfloor}||\mathcal{D}_{\lfloor k/2\rfloor+1}|$ pairs with distance 2.

Therefore, the average path length $\ell(G)$ defined in (16) is given by the equation

$$\frac{n(n-1)}{2}\ell(G)=\frac{1}{2}\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|\left(\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|-1\right)+2\frac{1}{2}\sum_{j=1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|\left(\sum_{j=1}^{\lfloor k/2\rfloor}|\mathcal{D}_j|-1\right)$$
$$+\sum_{l=1}^{\lfloor k/2\rfloor}|\mathcal{D}_l|\left[\sum_{j=k-l+1}^{k}|\mathcal{D}_j|+2\sum_{j=\lfloor k/2\rfloor+1}^{k-l}|\mathcal{D}_j|\right]. \qquad \square$$

Considering the graph in Figure 15 (left), the characteristic path length is $\ell(G)=\frac{22}{15}$, while in the graph in Figure 15 (right), we get $\ell(G)=\frac{68}{45}$.

Note that by taking the inverse of the shortest path length, one can introduce a related measurement—the network efficiency,[51] $\epsilon(G)\equiv(1/(n(n-1)))\sum_{u\neq v\in G}1/d(u,v)$—that is also applicable to disconnected networks. Finally, we find that in a connected nested split graph, agents are at most two links separated from each other and, thus, these graphs are characterized by a short characteristic path length.

C.2 *Centrality*

In the next sections we analyze different measures of centrality in a nested split graph G. We derive the expressions for degree, closeness, and betweenness centrality as a function of the degree partition of G. Finally, we show that these measures are similar in the sense that they induce the same ordering of the nodes in G based on their centrality values.

[51]The network efficiency must not be confused with the efficiency of a network. The first is related to short paths in the network, while the latter measures social welfare, that is, the efficient network maximizes aggregate payoff.

C.2.1 *Degree centrality* The degree centrality of an agent $u \in \mathcal{N}$ is given by the proportion of agents that are adjacent to u (Wasserman and Faust 1994). We obtain the normalized degree centrality simply by dividing the degree of agent u with the maximum degree $n-1$. This yields the following corollary.

Corollary 7. *Consider a nested split graph $G=(\mathcal{N},\mathcal{E})$ and let $\mathcal{D}=(\mathcal{D}_0,\mathcal{D}_1,\ldots,\mathcal{D}_k)$ be the degree partition of G. Then for each $u \in \mathcal{D}_i$, $i=0,\ldots,k$, the degree centrality is given by*

$$C_d(u)=\begin{cases} \dfrac{1}{n-1}\displaystyle\sum_{j=1}^{i}|\mathcal{D}_{k+1-j}| & \text{if } 1 \le i \le \left\lfloor \dfrac{k}{2} \right\rfloor \\ \dfrac{1}{n-1}\left(\displaystyle\sum_{j=1}^{i}|\mathcal{D}_{k+1-j}|-1\right) & \text{if } \left\lfloor \dfrac{k}{2} \right\rfloor + 1 \le i \le k. \end{cases}$$

The result follows directly from Corollary 3.

We observe that degree centrality as well as the degree are increasing with increasing index i of the set $\mathcal{D}_i$ to which agent u belongs. Degree centralities for the graphs shown in Figure 15 can be derived from the degrees given in Appendix C.1.1 by dividing the degrees with $n-1$.

C.2.2 *Closeness centrality* Excluding the isolated nodes in G, closeness centrality of agent $u \in \mathcal{N} \setminus \mathcal{D}_0$ is defined as

$$C_c(u)=\frac{n-1}{\sum_{v\neq u\in G} d(u,v)},$$

where $d(u,v)$ measures the shortest path between agent u and agent v in $\mathcal{N}\setminus\mathcal{D}_0$. For a nested split graph, we obtain the following corollary.

Corollary 8. *Consider a nested split graph $G=(\mathcal{N},\mathcal{E})$ and let $\mathcal{D}=(\mathcal{D}_0,\mathcal{D}_1,\ldots,\mathcal{D}_k)$ be the degree partition of G. Then for each $u \in \mathcal{D}_i$, $i=0,\ldots,k$, the closeness centrality is given by*

$$C_c(u)=\begin{cases} \dfrac{n-1}{\sum_{j=k-i+1}^{k}|\mathcal{D}_j|+2\sum_{j=1}^{k-i}|\mathcal{D}_j|-2} & \text{if } 1 \le i \le \left\lfloor \dfrac{k}{2} \right\rfloor \\ \dfrac{n-1}{\sum_{j=k-i+1}^{k}|\mathcal{D}_j|+2\sum_{j=1}^{k-i}|\mathcal{D}_j|-1} & \text{if } \left\lfloor \dfrac{k}{2} \right\rfloor + 1 \le i \le k. \end{cases}$$

Proof. For both agents in the independent sets, $u \in \mathcal{D}_i$ with $1 \le i \le \lfloor k/2 \rfloor$, and in the dominating subsets, $u \in \mathcal{D}_i$ with $\lfloor k/2 \rfloor + 1 \le i \le k$, we can compute the length of the shortest paths as

$$d(u,v)=\begin{cases} 1 & \text{for all } v \in \displaystyle\bigcup_{j=k-i+1}^{k}\mathcal{D}_j \\ 2 & \text{for all } v \in \displaystyle\bigcup_{j=1}^{k-i}\mathcal{D}_j. \end{cases} \tag{17}$$

To compute the closeness centrality we have to consider all pairs of agents in the graph and compute the length of the shortest path between them, which is given in (17). We obtain for any agent $u \in \mathcal{D}_i$, $i = 1, \ldots, k$, the expression

$$C_c(u) = \begin{cases} \dfrac{n-1}{\sum_{j=k-i+1}^{k} |\mathcal{D}_j| + 2\sum_{j=1}^{k-i} |\mathcal{D}_j| - 2} & \text{if } 1 \leq i \leq \left\lfloor \dfrac{k}{2} \right\rfloor \\ \dfrac{n-1}{\sum_{j=k-i+1}^{k} |\mathcal{D}_j| + 2\sum_{j=1}^{k-i} |\mathcal{D}_j| - 1} & \text{if } \left\lfloor \dfrac{k}{2} \right\rfloor + 1 \leq i \leq k. \end{cases}$$

Note that we have subtracted 1 and 2 in the denominator, respectively, since the sums would otherwise include the contribution of agent u herself. $\square$

We have that closeness centrality is identical for all agents in the same set. Also note that $C_c(u) = 1$ for $u \in \mathcal{D}_k$. Moreover, closeness centrality is increasing with increasing degree. The closeness centralities of the agents in descending order for the graph in Figure 15 (left) are $1, 1, \frac{9}{11}, \frac{9}{13}, \frac{9}{14}, \frac{9}{14}, \frac{9}{15}, \frac{9}{15}, \frac{9}{16}, \frac{9}{16}$, while in the graph in Figure 15 (right), they are $1, \frac{9}{10}, \frac{9}{11}, \frac{9}{14}, \frac{9}{14}, \frac{9}{15}, \frac{9}{15}, \frac{9}{15}, \frac{9}{16}, \frac{9}{17}$.

C.2.3 *Betweenness centrality* Betweenness centrality is defined as (Freeman 1977)

$$C_b(u) = \sum_{u \neq v \neq w \in G} \frac{g(v, u, w)}{g(v, w)},$$

where $g(v, w)$ denotes the number of shortest paths from agent v to agent w and $g(v, u, w)$ counts the number of paths from agent v to agent w that pass through agent u.

The betweenness centrality for a nested split graph can be derived from its degree partition as follows.

Corollary 9. *Consider a nested split graph $G = (\mathcal{N}, \mathcal{E})$ and let $\mathcal{D} = (\mathcal{D}_0, \mathcal{D}_1, \ldots, \mathcal{D}_k)$ be the degree partition of G. Then $c_b(u) = 0$ if $u \in \mathcal{D}_i$, $i = 0, \ldots, \lfloor k/2 \rfloor$, and for each $u \in \mathcal{D}_i$, $v \in \mathcal{D}_{i-1}$, $i = \lfloor k/2 \rfloor + 1, \ldots, k$, the betweenness centrality is given by*

$$C_b(u) = \begin{cases} 0 & \text{if } i = \left\lfloor \dfrac{k}{2} \right\rfloor + 1,\ k \text{ odd} \\ \dfrac{|\mathcal{D}_{\lfloor k/2 \rfloor}|(|\mathcal{D}_{\lfloor k/2 \rfloor}| - 1)}{\sum_{j=\lfloor k/2 \rfloor + 1}^{k} |\mathcal{D}_j|} & \text{if } i = \left\lfloor \dfrac{k}{2} \right\rfloor + 1,\ k \text{ even} \\ c_b(v) + \dfrac{|\mathcal{D}_{k-i+1}|(|\mathcal{D}_{k-i+1}| - 1)}{\sum_{j=i}^{k} |\mathcal{D}_j|} & \\ \qquad + \dfrac{2|\mathcal{D}_{k-i+1}| \sum_{j=k-i+2}^{i-1} |\mathcal{D}_j|}{\sum_{j=i}^{k} |\mathcal{D}_j|} & \text{if } \left\lfloor \dfrac{k}{2} \right\rfloor + 2 \leq i \leq k. \end{cases}$$

Proof. In this proof, we follow closely Hagberg et al. (2006). The agents in the independent sets $\mathcal{D}_i$, $0 \leq i \leq \lfloor k/2 \rfloor$, do not lie on any shortest path between two other agents in the network and, thus, their betweenness centrality vanishes. For the agents in the dominating subsets, we have that the betweenness centrality of the agent $u \in \mathcal{D}_{\lfloor k/2 \rfloor + 1}$

vanishes if k is odd and is given by $|\mathcal{D}_{\lfloor k/2\rfloor}|(|\mathcal{D}_{\lfloor k/2\rfloor}|-1)/\sum_{j=\lfloor k/2\rfloor+1}^{k}|\mathcal{D}_j|$ if k is even. The latter result is due the shortest path between agents that are both in $\mathcal{D}_{\lfloor k/2\rfloor}$. Next, consider an agent $u\in\mathcal{D}_i$ and $v\in\mathcal{D}_{i-1}$, with $\lfloor k/2\rfloor+2\le i\le k$. Then the betweenness centrality of agent u is given by the recursive relationship

$$C_b(v)+\frac{|\mathcal{D}_{k-i+1}|(|\mathcal{D}_{k-i+1}|-1)}{\sum_{j=i}^{k}|\mathcal{D}_j|}+\frac{2|\mathcal{D}_{k-i+1}|\sum_{j=k-i+2}^{i-1}|\mathcal{D}_j|}{\sum_{j=i}^{k}|\mathcal{D}_j|}. \tag{18}$$

The first term in (18) is due to the fact that all shortest paths through lower dominating nodes $v\in\mathcal{D}_{i-1}$ have the same length as through $u\in\mathcal{D}_i$. The second term in (18) represents the contribution of paths between nodes in $\mathcal{D}_{k-i+1}$, divided by the number of shortest paths passing through the agents in the dominating subsets $\mathcal{D}_j$, $i\le j\le k$. The third term in (18) represents all paths between an agent in $\mathcal{D}_{k-i+1}$ and the other being in $\mathcal{D}_j$, $k-i+2\le j\le i-1$, divided by the number of shortest paths passing through the agents in the dominating subsets $\mathcal{D}_j$, $i\le j\le k$. □

From Corollary 9, we find that the agents in the independent sets $\mathcal{D}_i$ with $1\le i\le \lfloor k/2\rfloor$ have vanishing betweenness centrality. From the above equation, we also observe that the betweenness centrality is increasing with degree such that the agents in $\mathcal{D}_k$ have the highest betweenness centrality, the agents in $\mathcal{D}_{k-1}$ have the second highest betweenness centrality, and so on. Thus, the ordering of betweenness centralities follows the degree ordering for all agents in the dominating subsets, while the agents in the independent sets have vanishing betweenness centrality. For the betweenness centralities of the agents in the graph in Figure 15 (left), we obtain in descending order $\frac{109}{6}$, $\frac{109}{6}$, $\frac{31}{6}$, $\frac{1}{2}$, 0, 0, 0, 0, 0, 0, while in the graph in Figure 15 (right), they are 28, 12, 6, 0, 0, 0, 0, 0, 0, 0.

C.2.4 *Eigenvector centrality* There is a central property that holds for nested split graphs in relation to Bonacich centrality, namely that the agents with higher degree also have higher Bonacich centrality. Similar to part (i) of Proposition 1, we can give the following corollary.[52]

Corollary 10. *Let* $\mathbf{v}$ *be the eigenvector associated with the largest real eigenvalue* $\lambda_{\mathrm{PF}}(G)$ *of the adjacency matrix* $\mathbf{A}$ *of a nested split graph* $G=(\mathcal{N},\mathcal{E})$*. For each* $i\in\mathcal{N}$*, let* v_i *be the eigenvector centrality of agent* i*. Consider a pair of agents* $i,j\in\mathcal{N}$*. If and only if agent* i *has a higher degree than agent* j*, then* i *has a higher eigenvector centrality than* j*, i.e.,* $d_i>d_j\Leftrightarrow v_i>v_j$.

The proof of Corollary 10 is identical to the proof of part (i) of Proposition 1.

C.2.5 *Centrality rankings* Putting together the results for different centrality measures derived in the previous sections, we can make the following observation of the rankings of agents for different centrality measures in a nested split graph.

[52] A similar result can be found in Grassi et al. (2007).

Corollary 11. *Consider a nested split graph $G = (\mathcal{N}, \mathcal{E})$. Let $\mathbf{C}_d$, $\mathbf{C}_c$, $\mathbf{C}_b$, and $\mathbf{C}_v$ denote the vectors of degree, closeness, betweenness, and eigenvector centrality in G. Then for any $l, m \in \{d, c, v\}$, $l \neq m$, and $i, j \in \mathcal{N}$, we have that $C_l(i) \geq C_l(j) \Leftrightarrow C_m(i) \geq C_m(j)$ and $C_l(i) \geq C_l(j) \Rightarrow C_b(i) \geq C_b(j)$.*

The proof is a direct application of Corollaries 7, 8, and 9, and Proposition 1. If and only if an agent i has the kth highest degree centrality, then i is the agent with the kth highest closeness and eigenvector centrality. This result also holds for Bonacich centrality (see Proposition 1). Moreover, if an agent i has the kth highest degree centrality, then she also has the kth highest betweenness centrality and this also holds for closeness, eigenvector, and Bonacich centrality, respectively. The ordering induced by degree, closeness, eigenvector, and Bonacich centrality coincide, and these orderings also apply in a weak sense for betweenness centrality.

Appendix D: Interpreting the model for financial and trade networks

In this section, we discuss two stylized applications of our payoff function introduced in (1) that will be useful for our empirical analysis. Appendix D.1 introduces networks of banks operating in loan markets, while Appendix D.2 discusses networks of trade relationships between countries.

D.1 *Networks of banks*

Consider a population of banks $\mathcal{N} = \{1, 2, \ldots, n\}$ and a network $G \in \Omega$ representing links between them. Links in this context can be defined in a variety of ways. In the banking networks that we study, the links represent the presence of an *interbank loan*. In this context, we consider a model of quantity choice based on competition in quantities of lending à la Cournot between banks with a single product (a loan). Each bank $i \in \mathcal{N}$ provides a quantity $x_i \geq 0$ of loans. As in Cohen-Cole et al. (2011), we assume the following inverse linear demand function for the price (interest rate) of the loans of bank i,

$$p_i = 1 - \theta x_i, \tag{19}$$

where $\theta > 0$. Equation (19) implies that banks offer different types of loans that cannot be substituted and, therefore, operate in independent loan markets.[53] The marginal cost of each bank i is $c_i(G) \geq 0$. The profit function $\pi_i : \mathbb{R}_+^n \times \Omega \to \mathbb{R}$ of bank i in a network $G \in \Omega$ is given by

$$\pi_i(\mathbf{x}, G) = p_i x_i - c_i(G) x_i = x_i - \theta x_i^2 - c_i(G) x_i.$$

[53] This is an extreme assumption where customers are typically not able to substitute one type of loan for another and each type of loan constitutes an independent market. Adams et al. (2002) provide a related empirical study of the market power of banks offering different loan types. See Cohen-Cole et al. (2011), where banks operate in a single loan market and loans are perfectly substitutable.

We consider an interrelated cost function defined as (see Cohen-Cole et al. 2011)

$$c_i(G) = c_0 - \lambda \sum_{j=1}^{n} a_{ij} x_j, \tag{20}$$

where $c_0 > 0$ represents a bank's marginal cost when it has no links and $\lambda > 0$ is the cost reduction induced by each link formed by a bank. Equation (20) means that the marginal cost of each bank i is a decreasing function of the quantities produced by all banks $j \in \mathcal{N}_i$ that have a direct link with bank i. As stated above, this is because the operational costs of a trading floor or treasury operation decline per dollar of loan as loan size increases. We further assume that c_0 is large enough such that $c_i(G) \geq 0$ for all $i \in \mathcal{N}$ and $G \in \Omega$. The profit function for bank i is then given by

$$\pi_i(\mathbf{x}, G) = x_i - \theta x_i^2 - c_i(G) x_i = a x_i - \theta x_i^2 + \lambda \sum_{j=1}^{n} a_{ij} x_i x_j, \tag{21}$$

where we have denoted $a = 1 - c_0$. In the following discussion, we normalize $\theta = 1$. We are in the framework of our model since the utility functions in (1) and (21) are equivalent. It is then straightforward to show that the equilibrium loan quantities are given by $q_i^* = a b_i(G, \lambda)$, where $b_i(G, \lambda)$ is the Bonacich centrality of bank i in the network G, and equilibrium profits are $\pi_i^* = (q_i^*)^2 = a^2 b_i(G, \lambda)^2$ (Ballester et al. 2006).

D.2 *International trade networks*

Consider a set of countries $\mathcal{N} = \{1, 2, \dots, n\}$ and a network $G \in \Omega$ representing links between them. A link in this context indicates the presence of an (import or export) *trade* relationship between two countries. Each country i provides a volume $x_i \geq 0$ of trade. Countries are local monopolists and the inverse demand function for country $i \in \mathcal{N}$ is given by (19) with a parameter $\theta > 0$. This means that we assume that products produced by different countries are not substitutable. The marginal cost of production of each country i is $c_i(G) \geq 0$. The profit function $\pi_i : \mathbb{R}_+^n \times \Omega \to \mathbb{R}$ of country i in a trade network $G \in \Omega$ is given by (21), where x_i is the quantity produced by country i. We consider the marginal cost function defined in (20), where $c_0 > 0$ represents a country's marginal cost when it has no links and $\lambda > 0$ is the cost reduction induced by each trade relationship formed by a country. Production costs decrease with the volume of trade of the trading partner due to technology spillovers (see, e.g., Coe and Helpman 1995, Grossman and Helpman 1991). We further assume that c_0 is large enough such that $c_i(G) \geq 0$ for all $i \in \mathcal{N}$ and $G \in \Omega$. The profit function for country i is then given by (21), where we denote $a = 1 - c_0$ and normalize $\theta = 1$. It is clear that, as for the banking network, we are again in the framework of our model. Note also that in the context of trade, the influence of centrality has been documented in De Benedictis and Tajoli (2011).

Appendix E: Introducing link formation costs

Consider the network formation process of Definition 2 with one modification: The agent who wants to create a link needs to pay a cost c and creates the link only of it increases her payoff. The following proposition gives a bound on the linking cost c such that *link monotonicity* holds,[54] that is, marginal payoffs from forming a link are always positive.

Proposition 6. *Consider the network formation process $(G(t))_{t\in\mathbb{R}_+}$ in Definition 2. Assume that there is a cost $c \geq 0$ of creating a link for the agent who initiates that link. Further assume that agents create links only if it increases their payoff. Then, if c is smaller than $\lambda(2-\lambda)/(2(1-\lambda)^2)$, link monotonicity holds and the emerging network will always be a nested split graph.*

Proof. For notational simplicity, we drop the time index and denote $G \equiv G(t)$. We consider the network $G \oplus (i,j)$ obtained by adding the link $(i,j) \notin G$. The marginal payoff from forming a link (i,j) for agent $i \in \mathcal{N}$ is given by

$$\begin{aligned}\pi_i^*(G\oplus(i,j),\lambda) - \pi_i^*(G,\lambda) &= \tfrac{1}{2}\big(b_i(G\oplus(i,j),\lambda)^2 - b_i(G,\lambda)^2\big) \\ &= \tfrac{1}{2}\big(b_i(G\oplus(i,j),\lambda) - b_i(G,\lambda)\big)\big(b_i(G\oplus(i,j),\lambda) + b_i(G,\lambda)\big).\end{aligned}$$

Note that the Bonacich centrality of agent $i \in \mathcal{N}$ can be written as $b_i(G,\lambda) = 1 + \sum_{j\in N_i} b_j(G,\lambda)$. The change in the Bonacich centrality from forming the link (i,j) is given by

$$\begin{aligned}b_i(G\oplus(i,j),\lambda) - b_i(G,\lambda) &= \underbrace{\sum_{k\in N_i\backslash\{j\}}\big(b_k(G\oplus(i,j),\lambda) - b_k(G,\lambda)\big)}_{>0} + \lambda b_j(G\oplus(i,j),\lambda) \\ &\geq \lambda b_j(G\oplus(i,j),\lambda) \geq \lambda \min_{k\in\mathcal{N}} b_k(G\oplus(i,j),\lambda).\end{aligned}$$

In the first line above, we have used the fact that the number of walks emanating at i is increasing with the addition of a link and so is the Bonacich centrality.

The smallest Bonacich centrality in a nonempty graph G (after the creation of a link, the graph is always nonempty) is obtained in a path of length 2 (dyad), P_2, for which $b_i(P_2,\lambda) = 1/(1-\lambda)$. Hence, we have that $b_i(G\oplus(i,j),\lambda) - b_i(G,\lambda) > \lambda/(1-\lambda)$ and $b_i(G\oplus(i,j),\lambda) + b_i(G,\lambda) > (2-\lambda)/(1-\lambda)$, so that the marginal payoff of agent i from forming a link (i,j) is bounded from below by $\pi_i^*(G\oplus(i,j),\lambda) - \pi_i^*(G,\lambda) > \lambda(2-\lambda)/(2(1-\lambda)^2)$. This bound might seem crude, but note that if the linking cost is higher than $\lambda(2-\lambda)/(2(1-\lambda)^2)$, the empty graph is a stable network, irrespective of how we allow agents to remove links. Hence, we find that if the cost c of a link is lower than $\lambda(2-\lambda)/(2(1-\lambda)^2)$, a link will always be formed and the networks generated in our network formation process will all be nested split graphs. □

The above proposition shows that nested split graphs can also arise even when links are costly to be formed, as long as the costs are not too large.

[54] Link monotonicity requires that $\pi_i^*(G\oplus(i,j),\lambda) > \pi_i^*(G,\lambda)$ for all agents $i,j\in\mathcal{N}$ and links $(i,j)\notin G$ (see, e.g., Dutta et al. 2005).

References

Adams, Robert M., Lars-Hendrik Röller, and Robin C. Sickles (2002), "Market power in outputs and inputs: An empirical application to banking." Finance and Economics Discussion Series 2002–52, U.S. Board of Governors of the Federal Reserve System. [745]

Åkerman, Anders and Anna Larsson (forthcoming), "The global arms trade network 1950–2007." *Journal of Comparative Economics*. [696, 719]

Bala, Venkatesh and Sanjeev Goyal (2000), "A noncooperative model of network formation." *Econometrica*, 68, 1181–1229. [699]

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006), "Who's who in networks. Wanted: The key player." *Econometrica*, 74, 1403–1417. [696, 699, 700, 701, 702, 746]

Barabási, Albert-László and Réka Albert (1999), "Emergence of scaling in random networks." *Science*, 286, 509–512. [696]

Bonacich, Phillip (1987), "Power and centrality: A family of measures." *American Journal of Sociology*, 92, 1170–1182. [696, 701]

Boss, Michael, Helmut Elsinger, Martin Summer, and Stefan Thurner (2004), "Network topology of the interbank market." *Quantitative Finance*, 4, 677–684. [718]

Bramoullé, Yann and Rachel Kranton (2007), "Public goods in networks." *Journal of Economic Theory*, 135, 478–494. [699]

Bramoullé, Yann, Rachel Kranton, and Martin D'Amours (2014), "Strategic interaction and networks." *American Economic Review*, 104, 898–930. [699]

Bramoullé, Yann, Dunia López, Sanjeev Goyal, and Fernando Vega-Redondo (2004), "Network formation and anti-coordination games." *International Journal of Game Theory*, 33, 1–19. [699]

Brooks, Stephen P. and Gareth O. Roberts (1998), "Assessing convergence of Markov chain Monte Carlo algorithms." *Statistics and Computing*, 8, 319–335. [721]

Brualdi, Richard A. and Alan J. Hoffman (1985), "On the spectral radius of $(0,1)$-matrices." *Linear Algebra and Its Applications*, 65, 133–146. [706]

Cabrales, Antonio, Antoni Calvó-Armengol, and Yves Zenou (2011), "Social interactions and spillovers." *Games and Economic Behavior*, 72, 339–360. [699]

Calvó-Armengol, Antoni and Yves Zenou (2004), "Social networks and crime decisions: The role of social structure in facilitating delinquent behavior." *International Economic Review*, 45, 939–958. [699]

Chib, Siddhartha (2001), "Markov chain Monte Carlo methods: Computation and inference." In *Handbook of Econometrics*, Vol. 5 (J. J. Heckman and E. Leamer, eds.), 3569–3649, Elsevier Science, Amsterdam. [720, 721]

Coe, David T. and Elhanan Helpman (1995), "International R&D spillovers." *European Economic Review*, 39, 859–887. [746]

Cohen-Cole, Ethan, Eleonora Patacchini, and Yves Zenou (2011), "Systemic risk and network formation in the interbank market." Discussion Paper 8332, CEPR. [745, 746]

Cvetković, Dragoš and Peter Rowlinson (1990), "The largest eigenvalue of a graph: A survey." *Linear and Multilinear Algebra*, 28, 3–33. [697, 701, 705]

De Benedictis, Luca and Lucia Tajoli (2011), "The world trade network." *World Economy*, 34, 1417–1454. [696, 746]

Dutta, Bhaskar, Sayantan Ghosal, and Debraj Ray (2005), "Farsighted network formation." *Journal of Economic Theory*, 122, 143–164. [699, 747]

Ellison, Glenn (2000), "Basins of attraction, long-run stochastic stability, and the speed of step-by-step by evolution." *Review of Economic Studies*, 67, 17–45. [729]

Ethier, Stewart N. and Thomas G. Kurtz (1986), *Markov Processes*. Wiley, New York. [727]

Freeman, Linton C. (1977), "A set of measures of centrality based on betweenness." *Sociometry*, 40, 35–41. [743]

Freeman, Linton C. (1979), "Centrality in social networks: Conceptual clarification." *Social Networks*, 1, 215–239. [717]

Galeotti, Andrea and Sanjeev Goyal (2010), "The law of the few." *American Economic Review*, 100, 1468–1492. [699]

Galeotti, Andrea, Sanjeev Goyal, Matthew O. Jackson, Fernando Vega-Redondo, and Leat Yariv (2010), "Network games." *Review of Economic Studies*, 77, 218–244. [699]

Gallager, Robert G. (1996), *Discrete Stochastic Processes*. Springer, Berlin. [727]

Goyal, Sanjeev and Sumit Joshi (2003), "Networks of collaboration in oligopoly." *Games and Economic Behavior*, 43, 57–85. [699, 709, 717]

Goyal, Sanjeev, Marco van der Leij, and José Luis Moraga-González (2006), "Economics: An emerging small world." *Journal of Political Economy*, 114, 403–412. [709, 716]

Goyal, Sanjeev and Fernando Vega-Redondo (2005), "Network formation and social coordination." *Games and Economic Behavior*, 50, 178–207. [699]

Grassi, Rosanna, Silvana Stefani, and Anna Torriero (2007), "Some new results on the eigenvector centrality." *Journal of Mathematical Sociology*, 31, 237–248. [744]

Grimmett, Geoffrey and David Stirzaker (2001), *Probability and Random Processes*. Oxford University Press, Oxford. [726, 734]

Grossman, Gene M. and Elhanan Helpman (1991), *Innovation and Growth in the Global Economy*. MIT Press, Cambridge. [746]

Hagberg, Aric, Pieter Swart, and Daniel Schult (2006), "Designing threshold networks with given structural and dynamical properties." *Physical Review E*, 74, 056116. [705, 743]

Hagedoorn, John (2002), "Inter-firm R&D partnerships: An overview of major trends and patterns since 1960." *Research Policy*, 31, 477–492. [703]

Harrigan, Katherine R. (1988), "Strategic alliances and partner asymmetries." *Management International Review*, 28, 53–72. [703]

Jackson, Matthew O. and Brian W. Rogers (2007), "Meeting strangers and friends of friends: How random are social networks?" *American Economic Review*, 97, 890–915. [696, 699]

Jackson, Matthew O. and Alison Watts (2002a), "On the formation of interaction networks in social coordination games." *Games and Economic Behavior*, 41, 265–291. [699]

Jackson, Matthew O. and Alison Watts (2002b), "The evolution of social and economic networks." *Journal of Economic Theory*, 106, 265–295. [704]

Jackson, Matthew O. and Asher Wolinsky (1996), "A strategic model of social and economic networks." *Journal of Economic Theory*, 71, 44–74. [699]

Jackson, Matthew O. and Yves Zenou (2014), "Games on networks." In *Handbook of Game Theory*, Vol. 4 (H. Peyton Young and Shmuel Zamir, eds.), 91–157, Elsevier Science, Amsterdam. [699]

Kandori, Michihiro, George J. Mailath, and Rafael Rob (1993), "Learning, mutation, and long run equilibria in games." *Econometrica*, 61, 29–56. [697]

Kemeny, John G. and Laurie J. Snell (1960), *Finite Markov Chains*. Van Nostrand, New York. [729]

Kogut, Bruce, Pietro Urso, and Gordon Walker (2007), "Emergent properties of a new financial market: American venture capital syndication, 1960–2005." *Management Science*, 53, 1181–1198. [703]

König, Michael D., Claudio J. Tessone, and Yves Zenou (2009), "A dynamic model of network formation with strategic interactions." Discussion Paper 09-006, CCSS. [700]

König, Michael D., Claudio J. Tessone, and Yves Zenou (2010), "From assortative to dissortative networks: The role of capacity constraints." *Advances in Complex Systems*, 13, 483–499. [717]

König, Michael D. and Claudio J. Tessone (2011), "Network evolution based on centrality." *Physical Review E*, 84, 056108. [699, 700, 715]

Liggett, Thomas M. (2010), *Continuous Time Markov Processes: An Introduction*. American Mathematical Society, New York. [704]

Liu, Xiaodong, Eleonora Patacchini, Yves Zenou, and Lung-Fei Lee (2012), "Criminal networks: Who is the key player?" Discussion Paper 8772, CEPR. [699]

Mahadev, Nadimpalli V. R. and Uri N. Peled (1995), *Threshold Graphs and Related Topics*. North-Holland, Amsterdam. [697, 705, 706, 710, 726, 728, 735, 737]

Marjoram, Paul, John Molitor, Vincent Plagnol, and Simon Tavaré (2003), "Markov chain Monte Carlo without likelihoods." *Proceedings of the National Academy of Sciences of the USA*, 100, 15324–15328. [719]

Mele, Angelo (2010), "Segregation in social networks: Theory, estimation and policy." Working Paper 10-16, NET Institute. [699]

Minoiu, Camelia and Javier A. Reyes (2011), "A network analysis of global banking: 1978–2009." Working paper, IMF. [718]

Newman, Mark E. J. (2002), "Assortative mixing in networks." *Physical Review Letters*, 89, 208701. [716, 740]

Park, Seung H. and Michael V. Russo (1996), "When competition eclipses cooperation: An event history analysis of joint venture failure." *Management Science*, 42, 875–890. [703]

Pastor-Satorras, Romualdo, Alexei Vàzquez, and Alessandro Vespignani (2001), "Dynamical and correlation properties of the Internet." *Physical Review Letters*, 87, 258701. [716, 739, 740]

Rodríguez-Gironés, Miguel A. and Luis Santamaría (2006), "A new algorithm to calculate the nestedness temperature of presence–absence matrices." *Journal of Biogeography*, 33, 924–935. [719]

Sandholm, William H. (2010), *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge. [697, 705]

Snijders, Tom A. B. (2001), "The statistical evaluation of social network dynamics." *Sociological Methodology*, 31, 361–395. [703]

Snijders, Tom A. B., Johan Koskinen, and Michael Schweinberger (2010), "Maximum likelihood estimation for social network dynamics." *The Annals of Applied Statistics*, 4, 567–588. [703]

Sokal, Alan D. (1996), "Monte Carlo methods in statistical mechanics: Foundations and new algorithms." In *Functional Integration: Basics and Applications* (C. DeWitt-Morette, P. Cartier, and A. Folacci, eds.), 131–192, Plenum, New York. [721]

Soramaki, Kimmo, Morten Bech, Jeffrey Arnold, Robert Glass, and Walter Beyeler (2007), "The topology of interbank payment flows." *Physica A: Statistical Mechanics and Its Applications*, 379, 317–333. [696]

Staudigl, Mathias (2011), "Potential games in volatile environments." *Games and Economic Behavior*, 72, 271–287. [703]

Uzzi, Brian (1996), "The sources and consequences of embeddedness for the economic performance of organizations: The network effect." *American Sociological Review*, 61, 674–698. [695]

Vega-Redondo, Fernando (2007), *Complex Social Networks*. Cambridge University Press, Cambridge. [698]

Wasserman, Stanley and Katherine Faust (1994), *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge. [696, 707, 736, 742]

Watts, Alison (2001), "A dynamic model of network formation." *Games and Economic Behavior*, 34, 331–341. [699]

Watts, Duncan J. and Steven H. Strogatz (1998), "Collective dynamics of 'small-world' networks." *Nature*, 393, 440–442. [696, 717, 738, 740]

Young, Peyton H. (2001), *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, Princeton. [705]

Submitted 2012-9-25. Final version accepted 2013-6-13. Available online 2013-6-13.