Supplement to "The formation of networks with local spillovers and limited observability"

(*Theoretical Economics*, Vol. 11, No. 3, September 2016, 813–863)

Michael D. König
Department of Economics, University of Zurich

Appendix C: Attachment kernel and link incentive function

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

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

where $K_t^{\beta}(j|\mathcal{S}_t, G_{t-1})$ is the probability, conditional on the sample $\mathcal{S}_t$ and the prevailing network G_{t-1}, that an agent j receives a link after the m draws (without replacement) by the entrant, and β is a parameter related to the distribution of the additive error term ε_{tj} from (1) (see below). Since the entrant forms links to the agents that maximize his link incentive function plus a random element, we need to consider the cases where agent j has the highest value among all agents in the sample, or the second highest, and so on. The corresponding probability can be written as[1]

$$\begin{aligned} &K_t^{\beta}(j|\mathcal{S}_t, G_{t-1}) \\ &\quad = \sum_{l=1}^{m} \sum_{i_1, i_2, \ldots, i_{l-1}} \prod_{r=1}^{l-1} \mathbb{P}_t\Big(f_t^{\delta}(G_{t-1}, i_r) + \varepsilon_{t,i_r} = \max_{k \in \mathcal{S}_t \setminus \{i_1, \ldots, i_r\}} f_t^{\delta}(G_{t-1}, k) + \varepsilon_{t,k}\Big) \\ &\qquad\qquad \times \mathbb{P}_t\Big(f_t^{\delta}(G_{t-1}, j) + \varepsilon_{t,j} = \max_{k \in \mathcal{S}_t \setminus \{i_1, \ldots, i_{l-1}\}} f_t^{\delta}(G_{t-1}, k) + \varepsilon_{t,k}\Big) \mathbb{1}_{\mathcal{S}_t}(j), \end{aligned} \tag{S1}$$

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

[1] I assume that the entrant does not update the link incentive functions while forming links but evaluates it only once after he has observed the sample. The first sum in (S1) considers the case that agent j receives a link in the lth round while the second sum takes into account all possible sequences of agents i_1, i_2, …, i_{l-1} that receive a link in the $l-1$ previous rounds.

DOI: 10.3982/TE1524

with indices $i_1 \in \mathcal{S}_t \setminus \{j\}$, $i_2 \in \mathcal{S}_t \setminus \{j, i_1\}$, $i_3 \in \mathcal{S}_t \setminus \{j, i_1, i_2\}, \ldots, i_{l-1} \in \mathcal{S}_t \setminus \{j, i_1, i_2, \ldots, i_{l-2}\}$, and $1 \leq l \leq m$. In the following discussion, I assume that the exogenous random terms ε_{tj} are identically and independently type I extreme value distributed (or Gumbel distributed) with parameter η.[2] This assumption is commonly made in random utility models in econometrics (see, e.g., McFadden 1981). Under this distributional assumption, the probability that an entering agent t chooses the passive agent $j \in \mathcal{S}_t$ for creating the link tj (in the first of the m draws of link creation) follows a multinomial logit distribution given by (cf. Anderson et al. 1992)[3]

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

where we have applied condition (LD) for the link incentive function $f_t^\delta(G_{t-1}, \cdot)$, dropped terms of $o(\delta^b)$, and denoted $\beta \equiv \eta \delta^b$.

Appendix D: Large observation radius

D.1 *Sampling of agents*

In the following discussion, we provide a lower bound on the observation radius n_s such that with high probability all agents in the network are observed by an entrant. Note that the probability that an agent $i \in \mathcal{P}_{t-1}$ does not enter the sample $\mathcal{S}_t$ is given by

$$
\begin{aligned}
\mathbb{P}_t(i \notin \mathcal{S}_t | G_{t-1}) &= \left(1 - \frac{1 + d^-_{G_{t-1}}(i)}{t-1}\right)\left(1 - \frac{1 + d^-_{G_{t-1}}(i)}{t-2}\right) \cdots \left(1 - \frac{1 + d^-_{G_{t-1}}(i)}{t - 1 - (n_s - 1)}\right) \\
&= \left(1 - \frac{1 + d^-_{G_{t-1}}(i)}{t}\right)^{n_s} + o\left(\frac{1}{t}\right).
\end{aligned}
\tag{S2}
$$

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

[3] Assuming instead that we have a multiplicative error term ε_{tk} that follows an inverse exponential distribution with parameter η, one can show that this probability can be written as $\mathbb{P}_t(f_t^\delta(G_{t-1}, j) \cdot \varepsilon_{tj} = \max_{k \in \mathcal{S}_t} f_t^\delta(G_{t-1}, k) \cdot \varepsilon_{tk}) = (f_t^\delta(G_{t-1}, j)^\eta)/(\sum_{k \in \mathcal{S}_t} f_t^\delta(G_{t-1}, k)^\eta)$, which corresponds to the ratio form of the contest success function (Jia 2008).

To see that this equality holds, note that when denoting $c \equiv 1 + d^-_{G_{t-1}}(i)$, we can write the above product as

$$\left(1-\frac{c}{t-1}\right)\left(1-\frac{c}{t-2}\right)\ldots\left(1-\frac{c}{t-1-(n_s-1)}\right)=\prod_{s=1}^{n_s}\left(1-\frac{c}{t-s}\right).$$

Further, note that

$$\left(1-\frac{c}{t-n_s}\right)^{n_s} \leq \prod_{s=1}^{n_s}\left(1-\frac{c}{t-s}\right) \leq \left(1-\frac{c}{t}\right)^{n_s}. \tag{S3}$$

Now we have that

$$\frac{\left(1-\frac{c}{t}\right)^{n_s}}{\left(1-\frac{c}{t-n_s}\right)^{n_s}} = \left(\frac{(t-c)(t-n_s)}{t(t-c-n_s)}\right)^{n_s},$$

and using the fact that

$$\lim_{t\to\infty}\frac{(t-c)(t-n_s)}{t(t-c-n_s)}=1,$$

it follows that both the lower and the upper bound in (S3) converge to the same limit as t becomes large. Hence, we can write

$$\prod_{s=1}^{n_s}\left(1-\frac{c}{t-s}\right)=\left(1-\frac{c}{t}\right)^{n_s}+o\left(\frac{1}{t}\right).$$

Applying Bonferroni's inequality and neglecting terms of $o(1/t)$ in (S2), we then find that the probability that at least one of the agents in the set $\mathcal{P}_{t-1}$ is not observed by the entrant is bounded by $\mathbb{P}_t(\bigcup_{i\in\mathcal{P}_{t-1}}\{i \notin \mathcal{S}_t\}|G_{t-1}) \leq \sum_{i=1}^{t-1}\mathbb{P}_t(i \notin \mathcal{S}_t|G_{t-1}) \approx \sum_{k=0}^{t-2}(1-(1+k)/t)^{n_s}P_t(k) \approx \sum_{k=0}^{t-2}(1-n_s(1+k)/t)P_t(k) = 1-n_s(1+m)/t$, where we have assumed that $k = o_p(t)$ and used the fact that the average in-degree $\sum_{k=0}^{t-2}kP_t(k)$ equals the out-degree m. Hence, if we require the probability of an agent not being sampled to be lower than $\epsilon > 0$, then we must have that $n_s > t(1-\epsilon)/(1+m)$.

D.2 *Attachment kernel*

The probability that an agent j with in-degree $d^-_{G_{t-1}}(j)$ receives a link in the $(k+1)$st draw, given that the agents $l_1,\ldots,l_k$ have received a link in the previous k draws, $1 \leq k \leq m$, is (cf. (2))

$$\begin{aligned}\frac{e^{\beta d^-_{G_{t-1}}(j)}}{\sum_{i\in\mathcal{P}_{t-1}\backslash\{l_1,\ldots,l_k\}}e^{\beta d^-_{G_{t-1}}(i)}} &\approx \frac{1+\beta d^-_{G_{t-1}}(j)}{\sum_{i\in\mathcal{P}_{t-1}\backslash\{l_1,\ldots,l_k\}}(1+\beta d^-_{G_{t-1}}(i))}\\ &= \frac{1+\beta d^-_{G_{t-1}}(j)}{(1+\beta m)t}\left(1+O\left(\frac{1}{t}\right)\right),\end{aligned}$$

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

Appendix E: Payoff functions

This appendix contains a discussion of various models in the economic literature that satisfy Assumptions 1 and 2 introduced in Section 2.1.[4]

E.1 *Information diffusion in networks*

Following Fafchamps et al. (2010) I consider agents that exchange information in a network G, where information that travels longer paths is discounted by a factor $\delta \in [0, 1]$. It is assumed that information can travel both directions of a link and so I consider the (undirected) paths in the closure $\overline{G}$ of G. The probability that an agent j transmits information along a given path in $\overline{G}$ is independent of the probability that the same agent j transmits the same information along another path. With this assumption, the probability that agent i receives the information over all distances $k \geq 1$, when there are $c^k_{ij}(\overline{G})$ (undirected) paths of length k connecting i to j, becomes

$$P^{\delta}_{ij}(G) \equiv 1 - \prod_{k=1}^{\infty}(1 - \delta^k)^{c^k_{ij}(\overline{G})}.$$

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

$$\begin{aligned}
\pi_i(G, \delta) &\equiv V \sum_{j \in \mathcal{N}} \left(1 - \prod_{k=1}^{\infty}(1 - \delta^k)^{c^k_{ij}(\overline{G})}\right) - c d^+_G(i) \\
&= V \sum_{j \in \mathcal{N}} (1 - (1 - c^1_{ij}\delta)(1 - c^2_{ij}\delta^2)) + O(\delta^3) - c d_G(i) \\
&= V \sum_{j \in \mathcal{N}} (1 - 1 + c^1_{ij}\delta + c^2_{ij}\delta^2 - c^1_{ij}c^2_{ij}\delta^3) + O(\delta^3) - c d_G(i) \\
&= V\left(\delta d_G(i) + \delta^2 \sum_{j \in \mathcal{N}_G(i)} d_G(j)\right) - c d^+_G(i) + O(\delta^3).
\end{aligned}$$

[4] All the models discussed here (which fall into our general framework) exhibit the property that the payoff of an agent is increasing with the number of collaborations, i.e., his degree. This characteristic has been found in empirical studies of co-authorship networks (e.g., Abbasi et al. 2011, Ductor 2015).

It then follows that the link incentive function is given by $f_i^\delta(G,j) = V\delta - c + V\delta^2 d_G(j) + O(\delta^3)$. Link monotonicity (LM) holds if $c < V\delta$ and linear differences (LD) holds for $g(x) = Vx$ and $\gamma = 2$, since $f_i^\delta(G,j) - f_i^\delta(G,k) = V\delta^2(d_G(j) - d_G(k)) + O(\delta^3)$. As our measure of welfare we consider aggregate payoff given by

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

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

E.2 *Two-way flow communication*

The two-way flow model with decay was introduced by Bala and Goyal (2000). In this model links are interpreted as lines of communication between two individuals. If i wants to communicate with j, then i must first pay a fee of $c \geq 0$ to open the channel. By creating this link, i not only gets access to j but also to all individuals that are approachable by j via an (undirected) path in the closure $\overline{G}$. Formally, the payoff function $\pi_i : \mathcal{G}(n) \times \mathbb{R}_+ \to \mathbb{R}$ of agent $i \in \mathcal{N}$ is given by[5]

$$\pi_i(G,\delta) \equiv 1 + \sum_{i\neq j} \delta^{\ell(i,j,\overline{G})} - cd_G^+(i) \tag{S4}$$

for some $\delta \in [0,1]$, which is interpreted as the *degree of friction* in communication. The number $\ell(i,j,\overline{G})$ is the length of the shortest path connecting agent i with j in the graph $\overline{G}$. If i and j are not connected we adopt the convention that $\ell(i,j,\overline{G}) = \infty$. The difference to the payoff function in Fafchamps et al. (2010) of the previous section and the one in (S4) is that in the latter only the shortest paths matter.

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

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

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

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

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

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

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

where $e(\overline{G})$ is the number of edges in $\overline{G}$, $n = |\mathcal{N}|$, and we have used the fact that for $\overline{G} \in \mathcal{T}(\mathcal{N})$ the number of edges is $e(\overline{G}) = n-1$. It follows that for small δ such that terms of $O(\delta^3)$ become negligible, maximizing aggregate payoffs becomes equivalent to maximizing the degree variance. Hence, condition (DC) holds for aggregate payoff when $\overline{G} \in \mathcal{T}[\mathcal{N}]$.[6]

E.3 *Public goods provision*

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

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

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

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

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

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

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

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

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

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

E.4 *A linear-quadratic complementarity game*

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

$$\pi_i(\mathbf{x}, G, \delta) \equiv x_i - \tfrac{1}{2}x_i^2 + \delta\sum_{j=1}^{n} a_{ij}x_i x_j, \tag{S5}$$

where $\delta \geq 0$ and $a_{ij} \in \{0, 1\}$, $i, j \in \mathcal{N} = \{1, \ldots, n\}$ are the elements of the symmetric $n \times n$ *adjacency matrix* $\mathbf{A}$ of $\overline{G}$. This payoff function is additively separable in the idiosyncratic effort component ($x_i - \frac{1}{2}x_i^2$) and the peer effect contribution ($\delta\sum_{j=1}^{n} a_{ij}x_i x_j$). Payoffs display strategic complementarities in effort levels, i.e., $(\partial^2\pi_i(\mathbf{x}, G, \delta))/(\partial x_i\,\partial x_j) = \delta a_{ij} \geq 0$. Ballester et al. (2006) have shown that if $\delta < 1/\lambda_{\mathrm{PF}}(G)$, then the unique interior Nash equilibrium solution of the simultaneous n-player move game with payoffs given by (S5) and strategy space $\mathbb{R}_+^n$ is given by the Bonacich centrality $x_i^* = b_i(G, \delta)$ for all $i \in \mathcal{N}$ (Bonacich 1987).[7] Moreover, the payoff of agent i in equilibrium is given by

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

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

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

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

Note that equilibrium payoff can be written as

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

and the link incentive function is then given by

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

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

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

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

Appendix F: The LF-MCMC algorithm

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

Definition S1. Consider the statistics $\mathbf{S}$ and denote by $\mathbf{S}^o$ the observed statistics. Furthermore, let $\Delta(\mathbf{S}_i^o,\mathbf{S}_i)$ be a measure of distance between the ith realized statistic $\mathbf{S}_i$ of the network formation process $(G_t)_{t=1}^{T}$ with parameter vector $\boldsymbol{\theta}$ and the ith observed statistic $\mathbf{S}_i^o$ for $i=1,\dots,K$. Then we consider the Markov chain $(\boldsymbol{\theta}_s)_{s=1}^{n}$ induced by the following algorithm:

(i) Given $\boldsymbol{\theta}$, propose $\boldsymbol{\theta}'$ according to the proposal density $q_s(\boldsymbol{\theta}\to\boldsymbol{\theta}')$.

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

(iii) Calculate

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

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

(iv) Accept $\boldsymbol{\theta}'$ with probability $h(\boldsymbol{\theta}, \boldsymbol{\theta}')$; otherwise stay at $\boldsymbol{\theta}$ and go to (i).

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

Proposition S1. *The stationary distribution $f : \mathbb{R}^K \to [0,1]^K$ of the Markov chain $(\boldsymbol{\theta}_s)_{s=1}^n$ is given by*

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

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

$$\frac{q_s(\boldsymbol{\theta}' \to \boldsymbol{\theta})}{q_s(\boldsymbol{\theta} \to \boldsymbol{\theta}')} \leq 1. \tag{S6}$$

Consider the distribution of the parameter vector $\boldsymbol{\theta}$, conditional on the event $\{\Delta(\mathbf{S}^o, \mathbf{S}) \leq \boldsymbol{\epsilon}\} \equiv \prod_{i=1}^K \mathbb{1}_{\{\Delta(\mathbf{S}_i,\mathbf{S}_i^o)<\epsilon_i^{\min}\}}$, that is, $f(\boldsymbol{\theta}|\Delta(\mathbf{S}^o, \mathbf{S}) \leq \boldsymbol{\epsilon}) = \mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \boldsymbol{\epsilon}|\boldsymbol{\theta}) / \mathbb{P}(\Delta(\mathbf{S}^o, \mathbf{S}) \leq \boldsymbol{\epsilon})$. We have that

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

where we have used the fact that $h(\boldsymbol{\theta}', \boldsymbol{\theta}) = 1$ if the inequality in (S6) is satisfied. It follows that $f(\boldsymbol{\theta}|\Delta(\mathbf{S}^o, \mathbf{S}) \leq \boldsymbol{\epsilon})$ satisfies a detailed balance condition and therefore is the stationary distribution of the Markov chain. □

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

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

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

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

During the "burn-in" phase (Chib 2001), I consider a monotonic decreasing sequence of thresholds given by $\epsilon_{i,s} \geq \epsilon_{i,s+1} \geq \cdots \geq \epsilon_i^{\min}$ with $\epsilon_{i,s+1} = \max\{(1-\gamma)\epsilon_{i,s}, \epsilon_i^{\min}\}$ and $\gamma = 0.05$. Similarly, I assume a decreasing sequence of variances $\sigma_{i,s}^2 \geq \sigma_{i,s+1}^2 \geq \cdots \geq (\sigma_i^{\min})^2$ with $\sigma_{i,s+1}^2 = \max\{(1-\gamma)\sigma_{i,s}^2, (\sigma_i^{\min})^2\}$ for the proposal distribution $q_s(\theta_i \to \theta_i')$. The maximum number of iterations, n, has been chosen such that reasonably high values of $p_\theta(n)$ were obtained. As a measure of distance I choose the Euclidean distance $\Delta(\mathbf{S}_i, \mathbf{S}_i^o) = \sqrt{\sum_{j=1}^T (S_{i,j} - S_{i,j}^o)^2}$. The parameter ranges are $n_s \in \{1,\ldots,100\}$, $p \in [0,1]$ and $\beta \in [0,100]$. The parameters $\epsilon_i^{\min}$ are choose sufficiently small after long experimentation with different starting values and burn-in periods.

Appendix G: Undirected links

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

Definition S2. For a fixed $T \in \mathbb{N} \cup \{\infty\}$ we define a network formation process $(G_t)_{t\in[T]}$ as follows. Given the initial graph $G_1 = \cdots = G_{m+1} = K_{m+1}$, for all $t > m+1$ the graph G_t is obtained from G_{t-1} by applying the following steps:

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

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

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

G.1 *Large observation radius*

We first consider the case of $\mathcal{S}_t = \mathcal{P}_{t-1}$. Let $k_j(t)$ denote the degree of agent j at time t. Considering only the leading terms in $O(1/t)$ we can write the probability that an agent

$j \in \mathcal{P}_{t-1}$ receives a link by the entrant t as

$$K_t^{\beta}(j|G_{t-1}) \approx \frac{m}{1+2\beta m}\frac{1+\beta d_{G_{t-1}}(j)}{t}. \tag{S7}$$

Using the recursive equation (11) with the attachment kernel in (S7) yields the following proposition.

PROPOSITION S2. *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^{\beta})_{t\in\mathbb{N}}$ introduced in Definition S2 with n_s large enough such that $\mathcal{S}_t = \mathcal{P}_{t-1}$ for every $t > m+1$. Then, for all $k \geq 0$, we have in the limit $\beta \to 0$ that $P_t(k) \to P^{\beta}(k)$, where*

$$P^{\beta}(k) = \frac{(1+2m\beta)\Gamma\left(k+\frac{1}{\beta}\right)\Gamma\left(3+\frac{1}{\beta}+\frac{1}{m\beta}\right)}{(1+m+2m\beta)\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(k+3+\frac{1}{\beta}+\frac{1}{m\beta}\right)}. \tag{S8}$$

Equation (S8) follows directly from the recursion in (11) and the attachment kernel in (S7).

From (S8) we find that the large k behavior of the degree distribution follows a power law as $P^{\beta}(k) \sim k^{-(3+1/(m\beta))}$. In the continuum approximation we can write for the dynamics of $k_s(t)$ using (S7) as

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

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

$$k_s(t) = \frac{1}{\beta}\left((1+\beta m)\left(\frac{t}{s}\right)^{(\beta m)/(1+2\beta m)} - 1\right), \tag{S9}$$

and we obtain for the degree distribution in the continuum approximation

$$P^{\beta}(k) = \frac{1+2\beta m}{m}(1+\beta m)^{2+1/(\beta m)}(1+\beta k)^{-(3+1/(m\beta))},$$

with $\int_0^{\infty} P^{\beta}(k)\, dk = 1$. This yields the same asymptotic behavior of the degree distribution as in (S8).

Next, we turn to the average nearest neighbor connectivity.

PROPOSITION S3. *Consider the network formation process $(G_t^{\beta})_{t\in\mathbb{R}_+}$ of Definition S2 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that (S9) holds. Then in the limit $\beta \to 0$ the nearest neighbor degree distribution is given by*

$$k_{\mathrm{nn}}(k) = \frac{1}{\beta^2 k}\left(1 + \frac{1+\beta k}{1+\beta m}\left(\beta^2 R_s(s) - 1 + (1+\beta m)^2 \ln\left(\frac{1+\beta k}{1+\beta m}\right)\right)\right), \tag{S10}$$

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

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

and $s = t((1+\beta m)/(1+\beta k))^{2+1/(m\beta)}$.

Asymptotically, only the last term in (S10) is relevant and we obtain

$$k_{\mathrm{nn}}(k) \sim \frac{1+\beta m}{\beta} \ln\left(\frac{1+\beta k}{1+\beta m}\right)$$

as $k \to \infty$.

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

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

where we have denoted $a = m/(1+2\beta m)$ and used the fact that $1 + \beta k_s(t) = (1+\beta m)(t/s)^{\beta a}$ from (S9) under the continuum approximation. The initial condition is given by

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

Using the fact that

$$1 + \beta k_j(s) = (1+\beta m)\left(\frac{s}{j}\right)^{\beta a}, \tag{S11}$$

we obtain

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

We then get

$$R_s(t) = \frac{1}{\beta^2}\left(1 + \left(a\beta(1+\beta m)^2\left(\frac{1}{s}H(s,2a\beta) + (1+m\beta)\ln\left(\frac{t}{s}\right)\right) - 1 + \beta^2 b\right)\left(\frac{t}{s}\right)^{a\beta}\right). \tag{S12}$$

Once again using (S11) and inserting into $k_{\mathrm{nn}} = R_s/k$ delivers (S10). □

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

Proposition S4. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition S2 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that (S9) holds. Then in the limit $\beta \to 0$ the clustering degree distribution is given by*

$$C(k) = \frac{2}{k(k-1)}\left(M_s + \frac{b}{s(1-2a\beta)}\left(d + a\beta s^{2a\beta-1}\left(1-\left(\frac{t}{s}\right)^{2a\beta-1}\right)H_s^{2\beta a} - \left(\frac{t}{s}\right)^{2a\beta-1}\left(d+\ln\left(\frac{t}{s}\right)^{a\beta}\right)\right)\right), \tag{S13}$$

where

$$s = t\left(\frac{1+m\beta}{1+k\beta}\right)^{2+1/(m\beta)}, \qquad a = \frac{a}{1+2\beta m}, \qquad b = \frac{m(m-1)(1+\beta m)^2}{\beta(1+2\beta m)},$$
$$c = \frac{\beta m + a\beta(1-\beta)(1-2m\beta)}{(1+\beta m)^2}, \qquad d = \frac{c + a\beta(1-2c)}{1-2a\beta},$$

the harmonic number is defined as $H_s^a \equiv \sum_{j=1}^s j^{-a}$, and the initial condition is given by

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

The large k behavior of the clustering coefficient is dominated by the second term in (S13), yielding

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

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

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

With $R_s(t)$ from (S12) and (S11) we obtain

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

where

$$a = \frac{a}{1+2\beta m}, \qquad b = \frac{m(m-1)(1+\beta m)^2}{\beta(1+2\beta m)}, \qquad c = \frac{\beta m + a\beta(1-\beta)(1-2m\beta)}{(1+\beta m)^2},$$

and the harmonic number is defined as $H_s^a \equiv \sum_{j=1}^s j^{-a}$. The solution is given by

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

where $d = (c + a\beta(1-2c))/(1-2a\beta)$. Similar to the derivation of (26), the initial condition is given by

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

Using (S11) we then arrive at the expression in (S13). □

G.2 *Small observation radius*

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

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

In the following discussion, we assume that $\mathcal{S}_t \approx n_s(\bar{d}+1)$, where the average degree is given by $\bar{d} = 2m$, so that $\mathcal{S}_t \approx n_s(2m+1)$. Note that this assumption is much stronger than the approximation we made in (6). The probability that an agent j receives a link from t is then given by

$$\begin{aligned} K_t^\beta(j|G_{t-1}) &= \frac{m}{|\mathcal{S}_t|}\frac{n_s(1+d_{G_{t-1}}(j))}{t} + O\left(\frac{1}{t^2}\right) \\ &\approx \frac{m}{n_s(2m+1)}\frac{n_s(1+d_{G_{t-1}}(j))}{t} + O\left(\frac{1}{t^2}\right) \\ &\approx \frac{m}{2m+1}\frac{1+d_{G_{t-1}}(j)}{t}. \end{aligned} \tag{S14}$$

An analysis following the recursive equation (11) with the attachment kernel in (S14) yields the following proposition.

Proposition S5. *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ of Definition S2 with $\beta = 0$. If $n_s > 1$ or $m > 1$, further assume that (S14) holds. Then, for all $k \geq 0$, we have $P_t(k) \to P(k)$, where*

$$P(k) = \frac{(1+2m)\Gamma\left(3+\frac{1}{m}\right)}{m\Gamma\left(3+k+\frac{1}{m}\right)}. \tag{S15}$$

Equation (S15) follows directly from the recursion in (11) and (S14).

From (S15) we find that the degree distribution follows a power law as $P(k) \sim k^{-(3+1/m)}$ for large k. For the dynamics of $k_s(t)$ in the continuum approximation we get, with (S14), the differential equation

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

with the solution

$$k_s(t) = (m+1)\left(\frac{t}{s}\right)^{m/(2m+1)} - 1. \tag{S16}$$

The degree distribution in the continuum approximation is then given by[8]

$$P(k) = \frac{2m+1}{m}(m+1)^{2+1/m}(1+k)^{-(3+1/m)}, \tag{S17}$$

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

Next we consider the average nearest neighbor degree.

PROPOSITION S6. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition S2 in the continuum approximation with n_s small enough and assume that (S16) holds. If $\beta = 0$, then the nearest neighbor degree distribution is given by*

$$k_{\mathrm{nn}}(k) = \frac{1}{k}\left(\left(\frac{t}{s+1}\right)^a (a(m+1)^2 s^{2a-1} H_s^{2a} - 1) + (m+1)\left(\frac{t}{s}\right)^a \ln\left(\frac{t}{s+1}\right)^a\right), \tag{S18}$$

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

PROOF. Let $R_s(t) = \sum_{j\in\mathcal{N}_{G_t}(s)} k_j(t)$ be the sum of the degrees of the neighbors of vertex s at time t. Denoting $a = m/(1+2m)$, we have up to leading orders in $O(1/t)$ that[9]

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

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

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

[8]Note that the approximation for the degree distribution in (S17) has also been obtained in Wang et al. (2009).

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

and the initial condition is given by

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

Using this equation to solve for C_s delivers (S18). □

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

PROPOSITION S7. *Consider the network formation process $(G_t^{\beta})_{t\in\mathbb{R}_+}$ of Definition S2 in the continuum approximation with n_s small enough and assume that (S16) holds. Let $a = m/(2m+1)$ and $b = (2a(m-1))/(n_s(2m+1)-1)$ with $a > b > 0$. If $\beta = 0$, then the average clustering coefficient of an agent with degree k is bounded by $\underline{C}(k) \leq C(k) \leq \overline{C}(k)$, where*

$$\underline{C}(k) = \frac{2}{(a-b)k(k-1)}\left(a - (a+mb)\left(\frac{1+k}{1+m}\right)^{b/a} + bk\right) \tag{S19}$$

and

$$\overline{C}(k) = \frac{2}{(a-b)k(k-1)}\left(a + \left(\binom{m}{2}(a-b) - (a+mb)\right)\left(\frac{1+k}{1+m}\right)^{b/a} + bk\right), \tag{S20}$$

and the property that $C(k) = O(1/k)$.

PROOF. We need to consider the cases we have encountered already in the proof of Proposition 8 for a vertex s to form an additional triangle by an entrant t (see Figure 11). The expected number of triangles associated with case (i) is given by

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

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

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

and for case (iii) we obtain

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

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

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

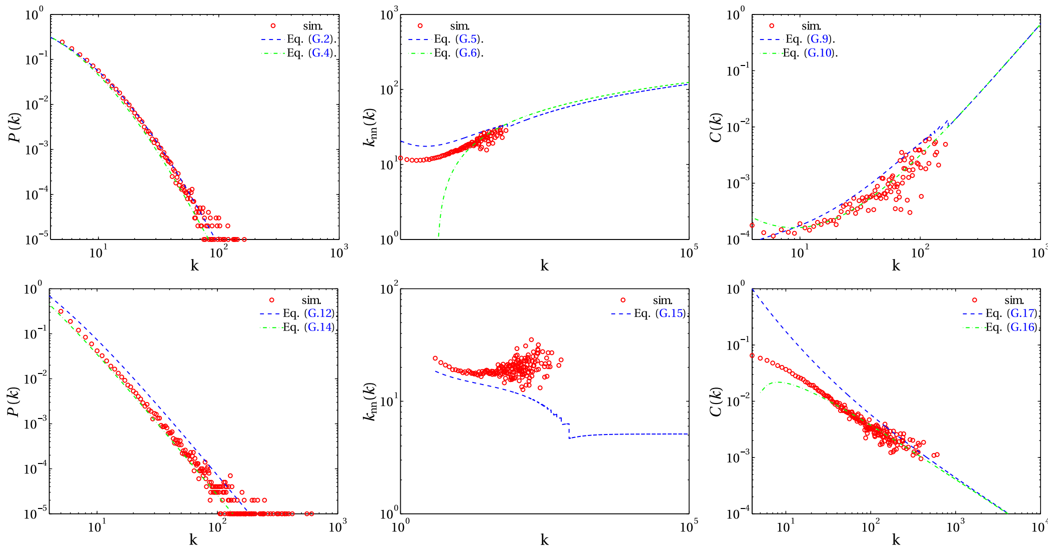

FIGURE S1. Top row: Comparison of simulation results with the theoretical predictions for $T = 10^5$, $\mathcal{S}_t = \mathcal{P}_{t-1}$, and $m = 4$, with $\beta = 0.1$ under the linear approximation to the attachment kernel. Bottom row: Comparison of simulation results for $T = 10^5$ and $n_s = m = 4$ ($\beta = 0$) with the theoretical predictions. Comparing the results of global and local information, we find that they differ mainly in the clustering degree distribution.

Using as a lower bound for the initial condition $M_s(s) \geq 0$ and an upper bound $M_s(s) \leq \binom{m}{2}$ as well as $s = ((1+k)/(1+m))^{-1/a}t$, we obtain the corresponding bounds for the clustering coefficient in (S19) and (S20). Both bounds decay as $(2b)/(a-b)(1/k)$ for large k and their difference vanishes for large k, implying that also $C(k) = O(1/k)$. $\square$

APPENDIX H: HETEROGENEOUS LINKING OPPORTUNITIES

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

DEFINITION S3. For a fixed $T \in \mathbb{N} \cup \{\infty\}$ we define a network formation process $(G_t)_{t \in [T]}$ as follows. Given the initial graph $G_1 = \cdots = G_{m+1} = K_{m+1}$, for all $t \in [T] \setminus \{1, \ldots, m+1\}$, the graph G_t is obtained from G_{t-1} by applying the following steps.

Growth: Given $\mathcal{P}_1$ and $\mathcal{A}_1$, for all $t > m$, if agent $t \in \mathcal{A}_{t-1}$, then the agent sets in period t are given by $\mathcal{P}_t = \mathcal{P}_{t-1} \cup \{t\}$ and $\mathcal{A}_t = \mathcal{A}_{t-1} \setminus \{t\}$, respectively; otherwise, set $\mathcal{P}_t = \mathcal{P}_{t-1}$ and $\mathcal{A}_t = \mathcal{A}_{t-1}$.

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

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

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

H.1 *Large observation radius*

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

$$K_t^{\beta}(j|G_{t-1}) \approx \frac{pm}{1+\beta pm} \frac{1+\beta d_{G_{t-1}}(j)}{t}. \tag{S21}$$

Following the recursive equation (11) with the attachment kernel in (S21) yields the following proposition.

Proposition S8. *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^{\beta})_{t\in\mathbb{N}}$ introduced in Definition S3 with n_s large enough such that $\mathcal{S}_t = \mathcal{P}_{t-1}$ for every $t > m+1$. Then, for all $k \geq m$, we have in the limit $\beta \to 0$ that $P_t^{\beta}(k) \to P^{\beta}(k)$ almost surely, where*

$$P^{\beta}(k) = \frac{1+\beta mp}{1+mp(1+\beta)} \frac{\Gamma\left(\frac{1}{\beta}+k\right)\Gamma\left(2+\frac{1+mp}{\beta mp}\right)}{\Gamma\left(\frac{1}{\beta}\right)\Gamma\left(2+\frac{1+mp}{\beta mp}+k\right)}. \tag{S22}$$

Equation (S22) follows directly from the recursion in (11) and the attachment kernel in (S21).

From the attachment kernel in (S21) we can write for the dynamics of the in-degree $k_s(t)$ of vertex s at time t in the continuum approximation

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

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

$$k_s(t) = \frac{1}{\beta}\left(\left(\frac{t}{s}\right)^{(\beta pm)/(1+\beta pm)} - 1\right), \tag{S23}$$

and we obtain for the degree distribution in the continuum approximation

$$P^{\beta}(k) = \frac{1+\beta mp}{mp}(1+\beta k)^{-(2+1/(\beta mp))}, \tag{S24}$$

with $\int_0^\infty P^\beta(k)\,dk = 1$. For $p=1$ we recover the distribution in (20). The degree distribution from (S22) and (S24) can be seen in Figure S2.

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

PROPOSITION S9. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition S3 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that (S23) holds. Then in the limit $\beta \to 0$, the nearest neighbor degree distribution is given by*

$$k_{\mathrm{nn}}^{-}(k) = \frac{1}{\beta^2 k}(1+(1+\beta k)(\ln(1+\beta k)-1)), \tag{S25}$$

and the average nearest neighbor out-degree is given by

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

where $a = (\beta mp)/(1+\beta mp)$ and $s = t(1+\beta k)^{-1/a}$.

Observe that (S25) is independent of p and identical to (23) from Proposition 5. From Proposition S9 we find that for large k, the average nearest in-neighbor connectivity grows logarithmically with k while the average nearest out-neighbor connectivity becomes independent of k and grows with the network sizes as $t^{(\beta mp)/(1+\beta mp)}$.

PROOF OF PROPOSITION S9. Let $R_s^-(t) = \sum_{j\in\mathcal{N}_{G_t}^-(s)} k_j(t)$. Up to leading orders in $O(1/t)$ we then have that

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

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

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

Using the fact that $t/s = (1+\beta k)^{1/a}$ from (S23), we obtain

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

With $k_{\mathrm{nn}}(k) = R_s^-/k$, the expression in (S25) follows.

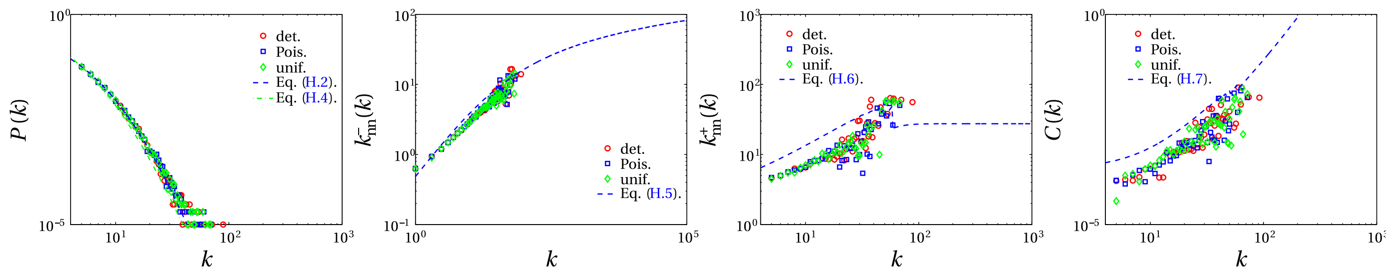

FIGURE S2. Comparison of simulation results with theoretical prediction of the link formation process in Definition S3 under global information with $p = 0.5$, $m = 4$, $\beta = 0.1$, and $T = 10^5$. Simulation results for the deterministic case (◯), a uniform distribution $X_m \sim \mathrm{U}\{1, 2m-1\}$ (⋄), and a Poisson distribution $X_m \sim \mathrm{Pois}(m)$ (□). both with expectation $\mathbb{E}(X_m) = m$ are shown.

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

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

where $a=(\beta pm)/(1+\beta pm)$. The solution is given by

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

The constant C_s is determined by the initial condition

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

We then obtain

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

with $s=t(1+\beta k)^{-1/a}$ from (S23) and $k_{\mathrm{nn}}^+ = R_s^+(k)/m$. $\square$

Moreover, we can derive the clustering degree distribution.

PROPOSITION S10. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition S3 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$ in the continuum approximation and assume that (S23) holds. Then, in the limit $\beta\to 0$, the clustering degree distribution is given by*

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

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

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

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

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

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

We see that this expression grows with k as a power law with exponent $2(1/a-2) = -2+2/(mp\beta)$.[10] Moreover, we find that the clustering coefficient is decreasing with the network size as $t^{-2(1-a)} = t^{-2/(1+mp\beta)}$.

Proof of Proposition S10. We need to consider the same cases as in the proof of Proposition 7. The probability associated with case (i) in Figure 10 is given by

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

Similarly, for the probability of case (ii) in Figure 10, we obtain

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

With R_s^+ and R_s^- given by (S25) and (S25), respectively, we obtain

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

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

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

where we have denoted $a = (\beta pm)/(1+\beta pm)$. The initial condition M_{s+1} together with (S26) delivers

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

Together with the initial condition, this is the expression in Proposition S10. □

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

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

PROPOSITION S11. *Let $N_s(t)$ denote the number of components of size s at time t. Consider the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ of Definition S3 with $\mathcal{S}_t = \mathcal{P}_{t-1}$ for all $t > m+1$. Assume that $m=1$ and $\beta=0$. If $p<1$, then there exists no giant component and the asymptotic (finite) component size distribution $P(s) = \lim_{t\to\infty} N_s(t)/t$ is given by*

$$P(s) = \frac{(1-p)\Gamma\left(\frac{1}{p}\right)\Gamma(s)}{p^2\Gamma\left(1+\frac{1}{p}+s\right)}. \tag{S26}$$

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

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

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

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

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

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

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

From this recursive equation, we obtain

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

which is (S26).

We next consider the generating function of the component size distribution $g(x) = \sum_{s=1}^{\infty} sP(s)x^s$. Observe that $g(1) = \sum_{s=1}^{\infty} sP(s)$, the fraction of nodes in finite components. In the absence of a giant component (that grows with t), we must have that $g(1)=1$. Inserting (S26) into $g(x)$ we find that $g(1)=1$ as long as $p<1$. Hence, the critical probability for the emergence of a giant component is $p=1$. □

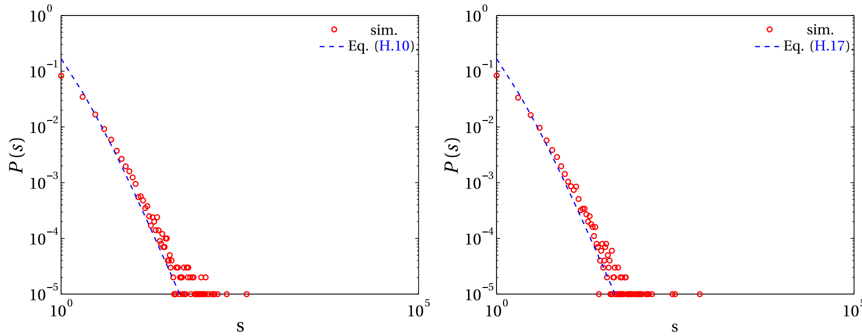

FIGURE S3. Comparison of simulation results with theoretical predictions for the component size distribution $P(s)$ of the link formation process in Definition S3 under global information with $p = 0.5$, $m = 1$, $\beta = 0$,, and $T = 10^5$ (left panel), and with $p = 0.5$, $n_s = 1$, $m = 4$, $\beta = 0$ and $T = 10^5$ (right panel).

From (S26) we find that the component size decays as a power law with exponent $1 + 1/p$, i.e.,

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

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

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

where we have used the approximation $\sum_{i \in H} k_i(t) \approx sp$. This is the same probability for the growth of a component of size s as in the case of $\beta = 0$ and, hence, we obtain the same component size distribution as in (S26).

H.2 *Small observation radius*

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

$$K_t^\beta(j|G_{t-1}) \approx \frac{pm}{1+m}\frac{d_{G_{t-1}}(j)+1}{t}. \tag{S27}$$

Using the recursive solution of (11), we can state the following proposition.

PROPOSITION S12. *Consider the sequence of degree distributions $\{P_t\}_{t\in\mathbb{N}}$ generated by an indefinite iteration of the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ of Definition S3 with $\beta = 0$.*

Further assume that (S27) holds. Then, for all $k \geq 0$, we have $P_t(k) \to P(k)$, where

$$P(k) = \frac{(1+m)k!\Gamma\big(2+\frac{m+1}{mp}\big)}{(1+m(1+p))\Gamma\big(2+\frac{m+1}{mp}+k\big)}. \tag{S28}$$

The proof of (S28) follows directly from the recursion in (11) and (S27).

With (S27) it follows for the dynamics of $k_s(t)$ in the continuum approximation that

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

with the solution

$$k_s(t) = \left(\frac{t}{s}\right)^{(pm)/(1+m)} - 1. \tag{S29}$$

The degree distribution in the continuum approximation is then given by

$$P(k) = \frac{1+m}{pm}(1+k)^{-(1+(1+m)/(pm))}, \tag{S30}$$

with $\int_0^\infty P(k)\,dk = 1$. For large k, (S28) and (S30) are equivalent. Moreover, for $p = 1$, we recover the distribution in (22). Next we turn to the analysis of the average nearest neighbor degree.

PROPOSITION S13. *Consider the network formation process $(G_t^\beta)_{t\in\mathbb{R}_+}$ of Definition S3 in the continuum approximation with n_s small enough and assume that (S29) holds. If $\beta = 0$, then the average nearest in-neighbor degree distribution is given by*

$$k_{\mathrm{nn}}^-(k) = \frac{1}{k}\big(1+(k+1)(\ln(k+1)-1)\big) \tag{S31}$$

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

$$k_{\mathrm{nn}}^+(k) = \frac{mp+1}{m+1}k + \frac{p}{m+1}t^{2a-1}(k+1)^{-(2a-1)/a}\zeta(t(k+1)^{-1/a}, 2a), \tag{S32}$$

where $a = (mp)/(1+m)$.

PROOF. So as to derive (S31), let us denote by $R_s^-(t)$ the sum of the in-neighbors' degrees of a vertex s at time t. We then have that

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

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

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

where we have used the fact that $s = t(k+1)^{-1/a}$ from (S29). Noting that $k^{-}_{\mathrm{nn}}(k) = R^{-}_s/k$, we readily obtain (S31).

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

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

The solution is given by $R^{+}_s(t) = -(m(1+mp))/(1+m) + C_s t^a$ and the initial condition is

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

so that we get

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

Inserting $s = t(k+1)^{-1/a}$ from (S29) and using the fact that $k_{\mathrm{nn}}(k) = R^{+}_s/m$ delivers (S32). □

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

Proposition S14. *Consider the network formation process $(G^{\beta}_t)_{t\in\mathbb{R}_+}$ of Definition S3 in the continuum approximation with n_s small enough and assume that (S29) holds. If $\beta = 0$, then the average clustering coefficient of an agent with degree k is given by Proposition 8, setting $a = (mp)/(m+1)$.*

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

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

For case (ii) we get

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

similarly, for case (iii) we get

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

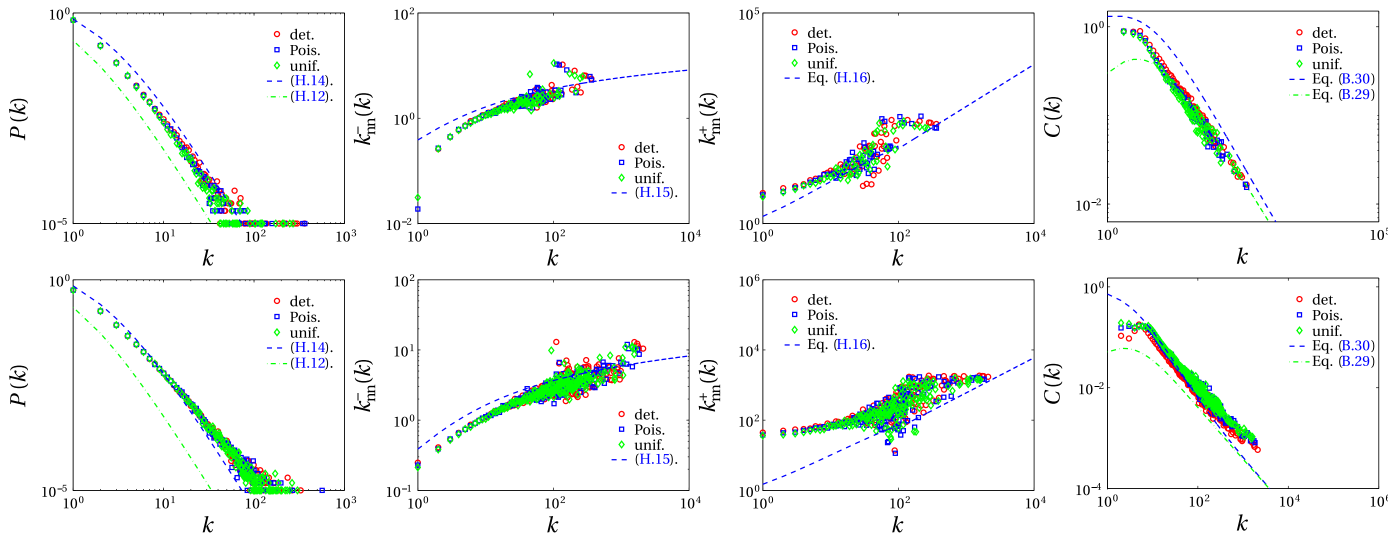

FIGURE S4. Comparison of simulation results with theoretical predictions of the link formation process in Definition S3 with $p = 0.5$, $n_s = 1$, $m = 4$, and $\beta = 0$, where the network size is $T = 10^5$ (top row) or $T = 2 \times 10^5$ (bottom row). We show simulations for the deterministic case (◯), a uniform distribution $X_m \sim \mathrm{U}\{1, 2m-1\}$ (⋄), and a Poisson distribution $X_m \sim \mathrm{Pois}(m)$ (□), both with expectation $\mathbb{E}[X_m] = m$.

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

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

with $a = (mp)/(m+1)$. This differential equation is identical to (27) and, hence, we obtain the same result as in Proposition 8. □

In the following text, we study the connectivity of the emerging networks in the network formation process introduced in Definition S3. We restrict our analysis to the case of $n_s = 1$. Observe that the probability that a component of size s grows by 1 unit due to the attachment of an entrant t is equivalent to the event that t observes one of the nodes in the component when constructing the sample $\mathcal{S}_t$. The probability of this event is $\frac{ps}{t}$. Hence, we obtain the same component size distribution as in Proposition S11. We then can state the following proposition.

Proposition S15. *Let $N_s(t)$ denote the expected number of components of size s at time t. Consider the network formation process $(G_t^\beta)_{t\in\mathbb{N}}$ of Definition S3 with $n_s = 1$. Then the asymptotic component size distribution $P(s) = \lim_{t\to\infty} N_s(t)/t$ is given by*

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

The proof follows the proof of Proposition S11.

References

Abbasi, Alireza, Jörn Altmann, and Liaquat Hossain (2011), "Identifying the effects of co-authorship networks on the performance of scholars: A correlation and regression analysis of performance measures and social network analysis measures." *Journal of Informetrics*, 5, 594–607. [4]

Anderson, Simon P., Andre de Palma, and Jacques-Francois Thisse (1992), *Discrete Choice Theory of Product Differentiation*. The MIT Press, Cambridge, Massachusetts. [2]

Bala, Venkatesh and Sanjeev Goyal (2000), "A noncooperative model of network formation." *Econometrica*, 68, 1181–1229. [5]

Ballester, Coralio, Antoni Calvó-Armengol, and Yves Zenou (2006), "Who's who in networks. Wanted: The key player." *Econometrica*, 74, 1403–1417. [7]

Bloch, Francis (1997), "Non-cooperative models of coalition formation with spillovers." In *New Directions in the Economic Theory of the Environment* (Carlo Carrao and Domenico Siniscalco, eds.), 311–352, Cambridge University Press, New York. [6]

Bonacich, Phillip (1987), "Power and centrality: A family of measures." *American Journal of Sociology*, 92, 1170. [7]

Chib, Siddhartha (2001), "Markov chain Monte Carlo methods: Computation and inference." In *Handbook of Econometrics* (James J. Heckman and Edward Leamer, eds.), 3569–3649, Elsevier Science, Amsterdam, The Netherlands. [10]

Debreu, Gerard and Israel Nathan Herstein (1953), "Nonnegative square matrices." *Econometrica*, 21, 597–607. [7]

Ductor, Lorenzo (2015), "Does co-authorship lead to higher academic productivity?" *Oxford Bulletin of Economics and Statistics*, 77, 385–407. [4]

Fafchamps, Marcel, Marco J. Van der Leij, and Sanjeev Goyal (2010), "Matching and network effects." *Journal of the European Economic Association*, 8, 203–231. [4, 5]

Goyal, Sanjeev and Sumit Joshi (2006), "Unequal connections." *International Journal of Game Theory*, 34, 319–349. [6]

Jackson, Matthew O. and Asher Wolinsky (1996), "A strategic model of social and economic networks." *Journal of Economic Theory*, 71, 44–74. [5]

Jia, Hao (2008), "A stochastic derivation of the ratio form of contest success functions." *Public Choice*, 135, 125–130. [2]

Marjoram, Paul, John Molitor, Vincent Plagnol, and Simon Tavaré (2003), "Markov chain Monte Carlo without likelihoods." *Proceedings of the National Academy of Sciences*, 100, 15324–15328. [9]

McFadden, Daniel (1981), "Econometric models of probabilistic choice." In *Structural Analysis of Discrete Data With Econometric Applications* (Charles F. Manski and Daniel McFadden, eds.), 198–272, The MIT Press, Cambridge, Massachusetts. [2]

Wang, Li-Na, Jin-Li Guo, Han-Xin Yang, and Tao Zhou (2009), "Local preferential attachment model for hierarchical networks." *Physica A: Statistical Mechanics and its Applications*, 388, 1713–1720. [15]

Co-editor Nicola Persico handled this manuscript.

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