# Anonymous Network Formation*

Itai Arieli†, Leonie Baumann‡ and Wade Hann-Caruthers§

January 26, 2026

**Abstract**

Social connections provide various benefits, such as access to information, support, and collaboration, motivating individuals to form networks. However, in many social settings—like academic conferences or networking events—participants are initially strangers, making the networking process inherently anonymous and random. This paper incorporates anonymity into the canonical non-cooperative connections model (Bala and Goyal (2000)) to explore symmetric, mixed-strategy equilibria in network formation. We show that, for any trembling-hand perfect equilibrium, strategies can be interpreted as socialization effort and yield a random network, closely related to but distinct from classical Erdős–Rényi graphs. This provides a strategic microfoundation for random graphs. We fully characterize these equilibria and efficient networks for large populations as a function of connection costs.

## 1 Introduction

Social connections hold many benefits like opportunities to access new information, receive support, socialize and collaborate. These benefits are commonly known and are an important motivation for forming social connections. Often, individuals that are initially

*Leonie Baumann acknowledges financial support from FRQSC grant 2023-NP-310915 and SSHRC grant 430-2022-00601. Itai Arieli acknowledges financial support for the ISF grant no: 2071717.

†Department of Economics, University of Toronto Mississauga and Technion - Israel Institute of Technology. itai.arieli@utoronto.ca.

‡Department of Economics, McGill University. leonie.baumann@mcgill.ca.

§University of Michigan. whanncar@gmail.com.

strangers to one another find themselves in a situation that is meant to build connections among them. Consider an introductory event for new assistant professors at a university, a social mixer for newcomers to a city, a professional networking event or an academic conference where participants have not known each other before. In such situations, an individual may not base her networking decision on the identity of others. However, she can still decide how many connections to form or, in other words, how much effort to invest into socializing.

When individuals cannot differentiate others based on their identity, strategic network formation possesses an innate randomness due to uncertainty about which individuals meet. We show that incorporating anonymity into a canonical strategic network formation model naturally connects strategic and random network formation – two topics that have been studied widely but mostly independently of each other. In particular, we provide a microfoundation for a random network formation process similar to but not the same as Erdős–Rényi through a canonical strategic network formation model with discrete link choices in large but finite populations. This identifies a new family of networks – in this case, random – that can be rationalized by symmetric, mixed-strategy equilibria of the same model which gives rise to deterministic networks like the circle and the star.

We investigate symmetric anonymous Nash equilibria of the canonical non-cooperative connections model with decay by Bala and Goyal (2000). In the model, an individual derives a benefit from every individual they are (directly or indirectly) connected to, that is decreasing in the distance $d$ of that connection as determined by a decay function $b(d)$, and they have to pay a constant cost $c$ for each link they form. With the anonymity assumption, an individual's strategy set of choosing whom to link to reduces to choosing a distribution over their out-degree. In equilibrium, individuals choose this distribution so as to maximize their expected utility from the resulting network. We characterize the asymptotic behavior of a sequence of equilibria (one equilibrium for each $n$) when the population size $n$ approaches infinity.

Every nontrivial equilibrium is of one of the following three types: 1) *mixed-degree* where individuals mix between two adjacent degrees, 2) *all-or-nothing* where individuals mix between degree zero and $n-1$, or 3) *more-than-half* where individuals mix over degrees that are larger than half of the population. Of these three types, only mixed-degree equilibria can be robust, in the sense that they can satisfy trembling-hand perfection, and we focus our analysis on them.

Mixed-degree equilibria are fully described by one real number which represents the mixture over two adjacent out-degrees and is equal to the expected out-degree. This equilibrium parameter can be interpreted as a socialization effort.[1] Any such equilibrium gives rise to a random network formation process, first discussed in the Scottish book under the name $\mathcal{G}_{k-out}$ (see Bollobás (2001), p. 41). An essential property of the random process is that every node has out-degree $k$ with links formed uniformly at random. Importantly, this implies that edges do not form independently, unlike in an Erdős–Rényi random graph.[2]

The nature of mixed-degree equilibria for large populations varies across three linking cost regimes. For low costs ($c < b(1) - b(2)$), in equilibrium agents link to a constant fraction of the population, $k \approx \alpha n$, and in the resulting random graph, the distance between any two agents is at most 2. For intermediate costs ($b(1) - b(2) < c < b(1)$), $k \approx \sqrt{n}$. In this case, agents link to a vanishing fraction of the population, but still the distance between any two agents is at most 2. For high costs ($c > b(1)$), there exist two types of equilibria. The first is a high-degree equilibrium, where $k \approx \sqrt{n}$ as with intermediate costs. The second is a low-degree equilibrium, where agents link to a bounded number of other agents, and this bound is independent of $n$. Despite the fact that the resulting random graph is connected, from each agent's perspective the graph is tree-like up to a far distance. In particular, the size of the smallest cycle an agent is contained in goes to infinity. The bounds of our regimes exactly correspond to the bounds in the pure equilibrium analysis of Bala and Goyal for large populations.[3] We will discuss the relation to their paper in more detail in Section 6.

We also study asymptotic efficiency within the class of symmetric anonymous strategy profiles. We find that anonymous equilibria are not efficient. We show that, in the low-

[1]Such a socialization effort has been used in Golub and Livne (2010) and Cabrales et al. (2011) to investigate a random and weighted network formation process, respectively. We will elaborate on our relation to these papers later.

[2]In Section 4.1, we discuss in more detail the relationship between our random network formation process and Erdős–Rényi graphs.

[3]Bala and Goyal find three bounds where the equilibrium behavior changes: $b(1) - b(2)$, $b(1)$, and $b(1) + (n-2)b(2)$. Since we study equilibria as the population size $n$ approaches infinity, this last bound disappears.

cost case, a strategy that links to a larger fraction of agents compared to the equilibrium strategy achieves higher social welfare. In the intermediate- and high-cost ranges, the efficient strategy profiles are *all-or-nothing* strategy profiles. The probability that is assigned to linking to all agents is vanishingly small, so that the resulting network is the join of a nonempty but small set of stars. We show that these profiles achieve the full social welfare relative to efficient pure networks, so anonymity is not a barrier to efficiency in these ranges.

Finally, we extend our analysis to incorporate homophily and show how strategic network formation can give rise to community structures. Building on the island model of Jackson and Rogers (2005) we consider a setting where link costs depend on type similarity. We characterize the resulting semi-anonymous equilibria and show that they feature dense intra-group and sparse inter-group links. This provides a strategic microfoundation for modular networks shaped by homophily, connecting rational behavior with empirically observed features of social networks.

**Related literature.** Numerous papers have studied network formation in environments where players have perfect knowledge about others and their strategies. In such environments, players choose their utility-maximizing actions considering every other player's choice and a network deterministically emerges from action profiles. Seminal papers on such settings of strategic network formation of deterministic networks are Jackson and Wolinsky (1996) and Bala and Goyal (2000). Many papers have followed to investigate various issues in such settings. In our setting, where players do not perfectly know others and their strategies and do not target particular other individuals, network formation possesses a random element in that there exists uncertainty about who ends up forming links with whom.

Randomness in network formation has traditionally been treated as a mechanical process without any strategic choices by agents. Examples are Watts (1999) and Watts and Strogatz (1998) that produce random graphs with high clustering and a small average between two nodes. Price (1976), Barabási and Albert (1999), and Cooper and Frieze (2003) generate random graphs where degrees follow a power-law distribution. Jackson and Rogers (2007) are able to explain several key features of large social networks, including small diameters, high clustering, fat-tailed degree distributions, and positive assortativity. They combine a random meeting process with network-based meetings. In their model, agents are mechanically connected to others either uniformly at random or by searching locally through the current structure of the network (e.g., meeting friends of friends). The main distinction to our paper is that our agents are strategic and a random network formation process is induced from agents' strategic choices, and not from a mechanical process with non-strategic agents.

Some papers have investigated random network formation with strategic agents. Currarini et al. (2009) assume a mechanical random matching process between agents during which friendships are formed. Individuals strategically choose how much time to spend matching (and thus forming relationships) with others. The probabilities of matching with different types depend on the types' endogenous participation ratios in the matching process. They use this model to analyze the contribution of chance and preferences towards homophily in networks – the tendency of individuals to form connections with others of their own type.

Cabrales et al. (2011) are the first to study a network formation process where linking strategies do not depend on others' identities and reflect the idea of network formation among strangers. In their model, agents' strategic choice is a socialization level and a productive effort. A profile of socialization levels deterministically leads to a weighted network where two agents' interaction (link) intensity depends on their chosen socialization efforts. This intensity can be interpreted as a linking probability. Agents' utilities depend only on their own productive effort and the productive efforts of their direct neighbors. In contrast to Cabrales et al. (2011), we consider an agent's only choice to be the number of links he forms, with each other agent having the same probability of receiving a given link from that agent, and utility arising from the number of (in)direct connections decreasing with their distance. Our strategy space gives rise to a random network formation process with discrete links. A mixed-degree equilibrium arises endogenously in our model, and the equilibrium strategy can be interpeted as a socialization effort. As this paper, Cabrales et al. (2011) analyze the limit of Nash equilibria when the population size goes to infinity. They identify a low- and high-activity equilibrium, as we do for mixed-degree equilibria under high costs.

Golub and Livne (2010) develop a theory of social network formation in which individuals make strategic decisions about how much effort to invest in socializing under uncertainty. In a first stage of network formation, links between agents are formed independently, with a probability that depends on chosen socialization parameters. A second stage of network formation ensues in which agents are linked mechanically to friends of their friends with an (independent) exogenous probability. In this paper, anonymity ensures that, ex-ante, all links are formed with the same probability; however, link formation is not independent. Conditional on a link being formed, all other links' probabilities of being formed have decreased. This corresponds to the intuition that an individual who spends effort on one connection cannot spend this effort on another connection. An individual does not interact with all other individuals with the same intensity, but rather a group of others to interact with is chosen uniformly. Furthermore, Golub and Livne (2010) assume that an agent's utility from the final network is given by the value of his direct connections only, meaning the utility of connections they form in the first networking stage and an expected utility from forming links with agents that are friends of friends. In contrast, we consider distance-based utility from both direct and indirect connections. Under certain parametric values, Golub and Livne identify that equilibrium networks can be either sparse or dense, again similar to our findings in the high-cost regime for mixed-degree equilibria.

# 2 Model

We consider a network formation game with *distance-based* utility à la Bala and Goyal (2000). The set of players is $N = \{1, \dots, n\}$, and the strategy set of player $i \in N$ is $S_i = 2^{N \setminus i}$, the collection of all subsets of players other than $i$.[4] Every strategy profile $\mathbf{s} = (s_1, \dots, s_n)$ determines a directed network $g = g(\mathbf{s})$, where for $i, j \in N$, $ij \in g$ if $j \in s_i$. We denote by $\hat{g}$ the undirected network that is formed by $g$ where $ij \in \hat{g}$ if $ij \in g$ or $ji \in g$.

For $i \in N$, we let $N_i(g)$ be the set of agents $j \in N$ that can be accessed from $i$ via a directed path in $g$. Similarly, let $N_i(\hat{g})$ be the set of agents $j \in N$ that can be accessed

[4]Note that the standard approach is to take the strategy space to be the set of $0-1$ vectors with coordinates corresponding to agents other than $i$. Taking the strategy space to be the set of subsets of players other than $i$ is equivalent and for our purposes is notationally more convenient.

from $i$ via a path in $\hat{g}$ or, in other words, the set of nodes that are in the same component as $i$.

We consider the two-way flow model, meaning a directed link confers benefits in both directions. Agent $i$ pays a cost of $c \cdot |s_i|$, where $c > 0$ is a positive constant representing the cost of an edge, and her benefit from a network $g$ is

$$B_i(g) = \sum_{j \in N_i(\hat{g}) \setminus \{i\}} b(d_{\hat{g}}(i, j)),$$

where $b : \mathbb{N} \to \mathbb{R}_+$ is a decreasing function and $d_{\hat{g}}(i, j)$ is the length of the shortest path from $i$ to $j$ in $\hat{g}$. This defines a utility in the game where

$$u_i(\mathbf{s}) = B_i(g(\mathbf{s})) - c \cdot |s_i|.$$

**Anonymity.** We next introduce anonymity in the baseline network formation game. Let $X_i = \Delta(S_i)$ be the set of mixed strategies of agent $i \in N$. Say that a mixed strategy $x_i \in X_i$ is *anonymous* if the probability it assigns to any subset of agents in $N \setminus i$ depends only on its size and not the identity of the agents. Equivalently, $x_i$ is anonymous if there exists a function $f_i : \{0, \ldots, n-1\} \to [0, 1]$ such that the probability agent $i$ assigns to any subset of agents $T \subseteq N \setminus i$ is $x_i(T) = \frac{1}{\binom{n-1}{|T|}} \cdot f_i(|T|)$. Note that for any $k$, $f_i(k)$ is the probability that agent $i$ chooses out-degree $k$, i.e. the probability that $|s_i| = k$.

This construction assumes that agents treat all other agents uniformly when forming links, reflecting the anonymity assumption. We will say that the anonymous strategy $x_i$ is a *degree strategy* if there is a $k$ such that $f_i(k) = 1$; when there is no ambiguity, we will refer to this strategy as simply $k$. The support of an anonymous strategy, defined as $Supp(x_i) = \{k : f_i(k) > 0\}$, represents all possible numbers of agents that the agent decides to link to with positive probability. Note that every anonymous strategy is effectively a mixture over degree strategies. In particular, if an anonymous strategy is a best response, then all degree strategies in its support must also be best responses.

This formulation aligns with the practical interpretation of anonymous strategies in network formation. Agents, constrained by the lack of distinguishing information, make strategic choices based on their expectations regarding the overall level of networking activity rather than individual efforts.

We call a profile of anonymous strategies $\mathbf{x} = (x_1, \ldots, x_n)$ *symmetric* if there exists a function $f$ such that $f_i = f$ for all $i \in N$, in which case we say that $\mathbf{x}$ is a symmetric anonymous profile. This ensures that all agents adopt the same probabilistic approach to forming links, further reinforcing the symmetry inherent in the model.

We will abuse notation and write $u_i(\mathbf{x})$ to refer to agent $i$'s expected utility when the mixed strategy profile $\mathbf{x}$ is used:

$$u_i(\mathbf{x}) = \mathbb{E}_{\mathbf{s}\sim\mathbf{x}}[u_i(\mathbf{s})].$$

We analyze the Nash equilibria of the game. We say an equilibrium is *symmetric anonymous* if its strategy profile is symmetric anonymous. For the remainder of the paper, unless otherwise stated, "equilibrium" will mean symmetric anonymous equilibrium.

For $c \geq b(1)$, there always exists the trivial equilibrium of the empty network, where no agent links to any other agent. In the following, we will focus on non-trivial equilibria where agents use strategies with positive (expected) out-degree.

# 3 Equilibrium structure

In this section, we will provide a preliminary characterization of symmetric anonymous equilibria. Throughout this section, we fix the number of agents to be $n$.

Denote by $u(k; x)$ the expected utility of an agent if they use the degree strategy $k$ when all other agents use the anonymous strategy $x$. We will say that $u(\cdot; x)$ is *concave at $k$* if $u(k+1; x) - u(k; x) \leq u(k; x) - u(k-1; x)$, and we will say that $u(\cdot; x)$ is *concave* if it is concave at every $k \in \{1, \ldots, n-2\}$. We will say that $u(\cdot; x)$ is *strictly* concave if this inequality is strict.

**Proposition 1.** For any anonymous strategy $x$, $u(\cdot; x)$ is concave. Moreover, if $x$ has full support, then $u(\cdot; x)$ is strictly concave.

This follows from the more general fact that $u_i(s_{-i}, s_i)$ is submodular in $s_i$; see Appendix A. Concavity implies a useful structural property for equilibrium analysis. When the utility function $u(\cdot; x)$ is concave, agents face diminishing returns when increasing their out-degree. This condition simplifies the equilibrium characterization by limiting the range of candidate equilibrium strategies. Moreover, if $u(\cdot; x)$ is strictly concave, then at most two degrees that are adjacent maximize $u(\cdot; x)$: either $\arg\max u(\cdot; x) = \{k\}$ or $\arg\max u(\cdot; x) = \{k, k+1\}$ for some $k$. We will say that an anonymous strategy is a *mixed-degree* strategy if it is a degree strategy $k$ or it is supported on two adjacent degree strategies $k$ and $k+1$. If $x$ has full support, it follows that any best response to all other agents using the strategy $x$ must be a mixed-degree strategy. As an immediate corollary, any trembling hand perfect equilibrium must be mixed-degree.

**Corollary 1.** If $x$ corresponds to a trembling hand perfect equilibrium, then $x$ is a mixed-degree strategy.

Mixed-degree strategies can be naturally identified with real numbers $\kappa \in [0, n-1]$, where for $k \in \{0, \ldots, n-2\}$ and $\rho \in [0,1]$, $\kappa = k + (1-\rho)$ corresponds to the strategy that mixes between degree $k$ and $k+1$ with probability $\rho$ and $1-\rho$, respectively. Note moreover that the expected out-degree of the mixed-degree strategy $\kappa$ is exactly $\kappa$. The identification of mixed-degree strategies with a real number allows us to interpret such strategies as socialization efforts. In Section 4, we will investigate mixed-degree equilibria in detail. Mixed-degree equilibria exist for all costs and they feature different properties for different cost regimes. For high costs, all equilibria are mixed-degree, and for all costs, mixed-degree equilibria are the only robust ones, in the sense that no other equilibria are trembling hand perfect.

For intermediate and small costs, equilibria can also take two other forms. We will say that an anonymous strategy is *all-or-nothing* if it is supported on the degree strategies 0 and $n-1$. We will say that an anonymous strategy is *more-than-half* if it is supported on degree strategies $k \geq (n-1)/2$.

Our first main result provides a complete characterization of all symmetric anonymous equilibria provided that $n$ is sufficiently large.

**Theorem 1.** For any $c > 0$ and a sufficiently large $n$, there exists a nontrivial mixed-degree equilibrium. In addition, the non-mixed-degree equilibria are characterized as follows:

a) If $c \geq b(1)$, there are no non-mixed-degree equilibria.

b) If $c < b(1)$, there exists a unique all-or-nothing equilibrium.

c) If $c < \frac{b(1)-b(2)}{2}$, then any more-than-half strategy profile with expected out-degree $\left(1 - \frac{c}{b(1)-b(2)}\right)(n-1)$ is an equilibrium.

We outline here the ideas used to prove Theorem 1; a formal proof is in Appendix E. While it can be shown that standard results imply the existence of mixed-degree equilibria, they do not imply the existence of a nontrivial one. To establish a nontrivial mixed-degree equilibrium, suppose that all agents are using the mixed-degree strategy $\kappa$. The marginal benefit of an additional edge is very large (approaching $\infty$) if $k \approx \sqrt{n}$ and is close to 0 if $\kappa \approx n-1$. Since the marginal benefit of an additional edge is continuous (in a suitable sense) in $\kappa$, if $n$ is large, then there is some $\kappa$ between $\sqrt{n}$ and $n-1$ for which the marginal benefit is $c$.

As for the remaining characterizations, observe that if $x$ corresponds to an equilibrium which is not mixed-degree, then $u(\cdot; x)$ cannot be strictly concave. Characterizing equilibria which are not mixed-degree reduces to understanding failures to be strictly concave.

As we show, if $x$ is an anonymous strategy and $k$ is in the support of $x$ for some $1 \leq k < (n-1)/2$, then $u(\cdot; x)$ is strictly concave at all $k' \in \{1, \ldots, n-1-k\}$. In particular, if $k$ is a maximizer of $u(\cdot; x)$, then the set of maximizers must be either $\{k-1, k\}$, $\{k\}$, or $\{k, k+1\}$. Hence, if $x$ corresponds to an equilibrium whose support contains some $1 \leq k < (n-1)/2$, then $x$ must be mixed-degree.

In a similar spirit, we show that if the support of $x$ contains 0 and $k$ for some $(n-1)/2 \leq k < n-1$, then $u(\cdot; x)$ is strictly concave at 1. But this would imply that 0 and $k$ cannot both maximize $u(\cdot; x)$, so such an $x$ cannot correspond to an equilibrium.

Together, these observations imply that every equilibrium which is not mixed-degree must be all-or-nothing or more-than-half. In the remainder of this section, we will provide a brief analysis of non-mixed-degree equilibria. The next section will focus on mixed-degree equilibria and their corresponding random graphs.

### 3.1 All-or-nothing equilibrium

Let $x$ be the all-or-nothing strategy which puts probability $p$ on the degree strategy $n-1$ and $1-p$ on the degree strategy 0, and consider agent $i$'s payoff if all other agents use the strategy $x$. If agent $i$ forms links with all other agents, then she pays a cost of $c$ for each link and gets a benefit of $b(1)$ from each other agent, so $u(n-1; x) = (n-1) \cdot (b(1) - c)$. Now suppose agent $i$ forms no links. If another agent $j$ forms $n-1$ links, then $i$'s distance to $j$ is 1, and this occurs with probability $p$. If $j$ forms no links and at least one other agent forms $n-1$ links, then $i$'s distance to $j$ is 2, and this occurs with probability $(1-p)\cdot(1-(1-p)^{n-2})$. With the remaining probability, the network is empty. Hence,

$$u(0;x) = (n-1)\cdot[p\cdot b(1) + (1-p)\cdot(1-(1-p)^{n-2})\cdot b(2)].$$

A necessary condition for $x$ to correspond to an equilibrium is that 0 and $n-1$ must both be maximizers of $u(\cdot;x)$. In particular, 0 and $n-1$ must both give the same utility. Setting the expressions for $u(0;x)$ and $u(n-1;x)$ equal and rearranging yields the following equation, which $p$ must solve in order for $x$ to correspond to an equilibrium:

$$(1-p)\cdot(b(1)-b(2)) + (1-p)^{n-1}\cdot b(2) = c.$$

As we show in Appendix C.1, this condition is also sufficient. This equation has a unique solution for $c < b(1)$ and no solution for $c > b(1)$. If $c < b(1)-b(2)$, then the solution is $p = 1 - \frac{c}{b(1)-b(2)} + o(1)$, and if $b(1)-b(2) < c < b(1)$, then the solution is $p = -\frac{1}{n}\cdot\log\left(\frac{c-(b(1)-b(2))}{b(2)}\right) + o(\frac{1}{n})$.

## 3.2 More-than-half-equilibria

Let $x$ be a more-than-half strategy. If an agent $i$ uses the degree strategy $k \geq (n-1)/2$ when all other agents use the strategy $x$, then by the pigeonhole principle, the distance to any other agent is at most 2 with probability 1. Hence, the benefit agent $i$ gets from another agent $j$ is $b(1)$ if $i$ and $j$ are directly connected and $b(2)$ otherwise, so

$$\begin{aligned} u(k;x) &= \sum_{j\neq i}\{\mathbb{P}[ij\in g]\cdot b(1) + \mathbb{P}[ij\notin g]\cdot b(2)\} - c\cdot k \\ &= \sum_{j\neq i}\{b(2) + \mathbb{P}[ij\in g]\cdot(b(1)-b(2))\} - c\cdot k. \end{aligned}$$

Denote by $\tilde{k}$ the expected out-degree under the strategy $x$. The probability that $j$ forms a link to $i$ is $\frac{\tilde{k}}{n-1}$ and the probability that $i$ forms a link to $j$ is $\frac{k}{n-1}$. Since these events are independent,

$$\mathbb{P}[ij\in g] = \mathbb{P}[i\in s_j] + \mathbb{P}[i\notin s_j]\cdot\mathbb{P}[j\in s_i] = \frac{\tilde{k}}{n-1} + (1-\frac{\tilde{k}}{n-1})\cdot\frac{k}{n-1},$$

so

$$u(k;x) = (n-1)\cdot b(2) + \tilde{k}\cdot(b(1)-b(2)) + \left((1-\frac{\tilde{k}}{n-1})\cdot(b(1)-b(2)) - c\right)\cdot k.$$

If $x$ is an equilibrium and $(1 - \frac{\tilde{k}}{n-1}) \cdot (b(1) - b(2)) - c \neq 0$, then $u(\cdot; x)$ has a unique maximizer among $k \geq (n-1)/2$, so $x$ must be mixed-degree. It follows that if $x$ is not mixed-degree, then

$$\tilde{k} = (1 - \frac{c}{b(1) - b(2)}) \cdot (n - 1).$$

In fact, this condition is necessary and sufficient. That is, if $x$ is a more-than-half strategy, then it is an equilibrium if and only if its expected out-degree is equal to $(1 - \frac{c}{b(1)-b(2)}) \cdot (n-1)$. Note that since $x$ is more-than-half, the expected out-degree is at least $(n-1)/2$, so this is only possible for $c \leq \frac{1}{2} \cdot (b(1) - b(2))$.

# 4 Mixed-degree equilibria

## 4.1 Random network formation process

Erdős–Rényi graphs are a well-studied random graph model that often serves as a baseline for understanding the structure and dynamics of more complex social networks. While our $\mathcal{G}_{\text{k-out}}$ graphs are strongly related to Erdős–Rényi graphs, there are some fundamental distinctions.

The Erdős–Rényi (ER) model $G(n, p)$ is constructed by starting with $n$ nodes and independently including each possible (undirected) edge between pairs of nodes with a fixed probability $p$. This independence between edges makes $G(n, p)$ particularly simple to analyze, as many of its properties, such as degree distribution and connectivity thresholds, follow directly from classical probability theory. One key feature of $G(n, p)$ is that the degree distribution of nodes follows a binomial distribution, which approximates a Poisson distribution for large $n$ and small $p$. The Erdős–Rényi model is often used to study the emergence of connectivity and giant components in random graphs.

In the $\mathcal{G}_{\text{k-out}}$ model, there are $n$ agents, and each agent independently selects exactly $k$ other agents uniformly at random from the remaining $n - 1$ agents. These selections are directed: when agent $i$ selects agent $j$, it represents an outgoing link from $i$ to $j$. The resulting undirected network, denoted by $\hat{g}$, is then constructed by placing an undirected edge between agents $i$ and $j$ if either $i$ selects $j$ or $j$ selects $i$, or both.[5]

While the Erdős–Rényi model and the $\mathcal{G}_{\text{k-out}}$ model give rise to different forms of random graphs, they share some common features. To illustrate this, fix $k > 0$ and compare $G_{k\text{-out}}$ to $G(n, p)$ with $p = \frac{2k}{n}$ for large values of $n$. This is the natural choice of $p$, as it leads to a network with the same expected degree $2k$ asymptotically. Approximately, the degree distribution of the ER graph is Pois$(2k)$, while the degree distribution of the $G_{k\text{-out}}$ graph is Pois$(k) + k$, since each agent selects $k$ other agents and is selected by Pois$(k)$ other agents. $G_{k\text{-out}}$ is connected with probability that approaches 1 for $k > 1$ as $n$ grows (see Bollobás (1998) Theorem 7.34). In contrast, for large values of $n$, the ER graphs have a giant component with constant fractions of agents but also have a constant fraction of isolated nodes (see, e.g., Durrett (2010) Chapter 2).

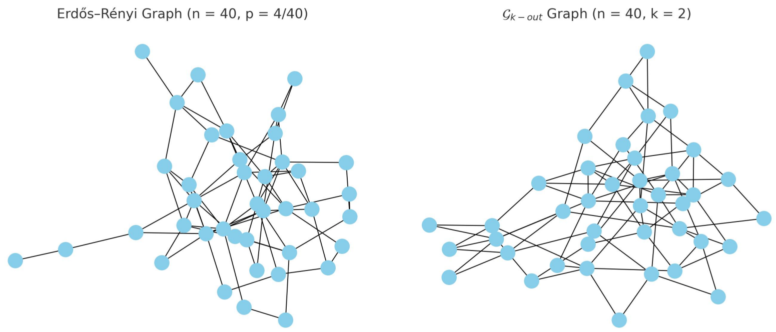


Figure 1: Comparison of random graphs

## 4.2 Characterization of mixed-degree equilibrium

Our main result provides the characterization of mixed-degree equilibria for all cost regimes. More precisely, we determine the properties of mixed-degree equilibrium sequences when the population grows large. We show that for linking costs less than the benefit of a direct connection, $b(1)$, the limit behavior of any equilibrium sequence is unique. For linking costs larger than $b(1)$, we show the existence of low-degree equilibria and characterize the behavior of high-degree equilibria.

We will denote by $E(c, n)$ the set of equilibrium values $\kappa$ when the cost is $c$ and there

[5] Our variant allows agents to choose any non-negative real number $\kappa$ (rather than only integer values $k$). In this case, each agent $i$ selects a random subset of $\lfloor \kappa \rfloor$ or $\lceil \kappa \rceil$ other agents, with probabilities chosen to ensure expected out-degree $\kappa$.

are $n$ agents. If $c > b(1)$, then $0 \in E(c, n)$ for all $n$, and if $c < b(1)$, then $0 \notin E(c, n)$ for any $n$. In Theorem 2, we characterize the limit of non-trivial equilibria where $\kappa_n > 0$.

**Theorem 2.** Let $\kappa_n \in E(c, n)$ for each $n$.

a) If $c < b(1) - b(2)$, then $\lim_{n\to\infty} \frac{\kappa_n}{n} = 1 - \frac{c}{b(1)-b(2)}$.

b) If $c \in (b(1)-b(2), b(1))$, then $\kappa_n > \sqrt{n}$ for all sufficiently large $n$ and $\lim_{n\to\infty} \frac{\log(\kappa_n)}{\log(n)} = \frac{1}{2}$.

c) If $c > b(1)$, then for all sufficiently large $n$, there is at least one low-degree equilibrium, $0 < \kappa_n < \sqrt{n}$ and at least one high-degree equilibrium, $\kappa_n > \sqrt{n}$. Moreover, every sequence of low-degree equilibria is bounded, and for every sequence of high-degree equilibria, $\lim_{n\to\infty} \frac{\log(\kappa_n)}{\log(n)} = \frac{1}{2}$.

The proof of Theorem 2 is in Appendix F. The realized networks exhibit distinct properties across the three cost regimes. In the low-cost regime ($c < b(1)-b(2)$), networks are highly connected, with agents forming links to a nonzero fraction of the population. This ensures that the average path length between any two agents is less than 2, implying a dense network structure. In the intermediate-cost regime ($b(1) - b(2) < c < b(1)$), the networks become sparser, as agents form a vanishing fraction of links relative to the population size. Despite this sparsity, the average path length is exactly 2, maintaining a balance between connectivity and cost efficiency. For the high-cost regime ($c > b(1)$), the equilibrium networks take on two distinct forms: low-degree or high-degree. A low-degree equilibrium results in a sparse, tree-like structure where the number of links is bounded. From the perspective of any given agent, the local network topology appears branching and acyclic. Conversely, in a high-degree equilibrium, the network is dense and highly connected, similar to the intermediate-cost regime but achieved at a higher individual cost. This is in sharp contrast to the intuition that a higher cost implies a less connected network.

We next outline the main ideas behind the proof of Theorem 2. Note that while for every population size $n$ the equilibrium is not necessarily unique, the asymptotic equilibrium behavior is often unique. The nature of equilibria is critically dependent on the relationship between the linking cost $c$ and the benefit from an agent being distance one or two away, $b(1)$ and $b(2)$. Three distinct regimes emerge:

1. Low-Cost Regime ($c < b(1) - b(2)$):

In this regime, the cost of forming a link is relatively low. Consequently, if two agents are not linked, it is mutually beneficial for one to link to the other. However, if agent $i$ is already linked to agent $j$, the marginal benefit for $j$ to reciprocate the link is zero, as the link is redundant. When the average degree, $\kappa$, is small relative to the network size (i.e., $\kappa/n$ is small), the probability of establishing a redundant link is low. Thus, increasing the degree is generally advantageous. Conversely, when the average degree is large relative to the network size ($\kappa/n$ is large), the likelihood of redundancy increases and reducing the degree becomes beneficial. An equilibrium is attained when the probability of forming a redundant link is just high enough that the expected marginal benefit of an additional link is precisely offset by the cost, $c$.

2. Intermediate-Cost Regime ($b(1) - b(2) < c < b(1)$):

In this regime, the cost of forming a link is moderate. The fraction of agents that are at distance 2 from a given agent becomes a salient factor in determining equilibrium behavior. If the average degree, $\kappa$, is large relative to $\sqrt{n}$, the network exhibits small-world properties, with the average path length between agents being at most 2. Consequently, the marginal benefit to an agent from additional links is small as most agents are already in close proximity. In fact, if $\kappa > n^{\frac{1}{2}+\varepsilon}$, it is profitable for any agent to switch to forming no links.

Now, if $n^{\varepsilon} < \kappa < n^{\frac{1}{2}-\varepsilon}$, most agents are a distance at least 3 from a given agent $i$. With high probability, adding an additional link results in a new direct connection with some agent $j$, as well as changing the distance from $i$ to $j$'s approximately $2\kappa$ neighbors from at least 3 to 2. Hence, the benefit of an additional link is at least approximately $(b(2) - b(3)) \cdot 2\kappa$, which is much larger than $c$ for large $n$. On the other hand, if $\kappa < n^{\varepsilon}$ for some small $\varepsilon$, then most agents are far away from $i$. In particular, for $m$ such that $b(1) - b(m) > c$ and $\varepsilon < \frac{1}{m}$, almost all agents are at least distance $m$ away from $i$. It follows that the benefit of an additional link is at least approximately $b(1) - b(m)$, so adding an additional link is profitable. Putting these cases together, it follows that in equilibrium, $\kappa$ must be between $n^{\frac{1}{2}-\varepsilon}$ and $n^{\frac{1}{2}+\varepsilon}$ for large enough $n$.

3. High-Cost Regime ($c > b(1)$):

In this regime, the cost of forming a link is substantial and two distinct types of equilibria can emerge: low-degree equilibria, characterized by a bounded number of links per agent, and high-degree equilibria, where the number of links per agent grows with network size.

**High-Degree Equilibria:** Despite the high linking cost, one can show the existence of a high-degree equilibrium. The behavior of these high-degree equilibria closely mirrors that of the intermediate-cost regime, with each agent forming approximately $\sqrt{n}$ links.[6]

**Low-Degree Equilibria:** If the number of links $\kappa_n$ is unbounded as the network size grows but remains less than $\sqrt{n}$, then the benefit of an additional link is at least proportional to $\kappa_n$. Consequently, such a sequence can only correspond to equilibria for a finite number of $n$, so any sequence of low-degree equilibria must be bounded.

A necessary condition for equilibrium is that an agent is indifferent between forming $k$ and $k+1$ links. For networks with a growing but less than $\sqrt{n}$ number of links, the substantial benefit of an additional edge precludes this. Using a continuity argument, we show that there is always some low-degree strategy where this indifference holds.

# 5 Welfare

In this section, we ask what the welfare-maximizing symmetric anonymous strategy profile is. We consider a social planner who can choose the agents' strategy profile but is restricted to choosing from among the symmetric anonymous strategy profiles. The planner's goal is to choose the degree distribution that maximizes the sum of agents' utilities. We compare the maximum welfare achieved by a symmetric anonymous profile to the maximum welfare attainable using pure strategies, as characterized by Bala and Goyal (2000). Note that the maximum welfare under pure strategies is the highest possible welfare achievable in this model if we did not restrict ourselves to symmetric anonymous strategies and thus is an upper bound to the maximum welfare with a symmetric anonymous strategy profile.

For a population of size $n$ and an anonymous strategy $x$, denote by $W_n(x)$ the social welfare (sum of agents' utilities) when all agents use the strategy $x$. Recall that in general, an anonymous strategy can be identified with a distribution over $\{0, 1, \ldots, n-1\}$. Let

[6] To be precise, each agent forms $n^{\frac{1}{2}+o(1)}$ links.

$W_n^* = \max_x W_n(x)$ where the maximum is taken across all anonymous strategies. We are interested in the asymptotic behavior and growth rate of $W_n^*$.

**Proposition 2.** The asymptotic optimal welfare among symmetric anonymous strategy profiles is achieved by:

1. for $c < 2(b(1) - b(2))$, the mixed-degree strategy $\frac{k_n}{n} = 1 - \frac{c}{2(b(1)-b(2))}$, and results in welfare which is a constant fraction $< 1$ of the optimal welfare attainable using pure strategies, and

2. for $2(b(1) - b(2)) < c$, the all-or-nothing strategy where the agent links to all others with probability $\frac{\log(n)}{n}$ and results in asymptotic welfare that has fraction of 1 of the welfare achieved by the optimal star network.

Note that for $c < 2(b(1) - b(2))$, the efficient anonymous network is achieved with a mixed-degree strategy. One can show in fact that any anonymous strategy with expected degree equal to $n(1 - \frac{c}{2(b(1)-b(2))})$ achieves the same asymptotic welfare.

Compared to equilibria in the range $c < 2(b(1) - b(2))$, the expected degree is higher in the efficient network. The standard reasoning applies here: the efficient profile takes into account the positive externality that one agent's links have on the connectivity of other agents. On the other hand, for the case $c > 2(b(1) - b(2))$, the efficient anonymous random network is obtained by the all-or-nothing strategy where $n - 1$ is chosen with probability $\frac{\log(n)}{n}$. The resulting network is a "multi-center star" network, which has at least one center (i.e. is not the empty network) with high probability. At the same time, the expected fraction of center agents vanishes as $n \to \infty$. Note that in this case, the expected degree in the efficient network is much lower than in the high-degree equilibrium.

Clearly, restricting attention to anonymous strategies implies that in the realized network two individuals may both link to each other. This results in a welfare loss that cannot be avoided with non-anonymous strategies due to their random nature. Surprisingly, in the range $2(b(1) - b(2)) < c$, the loss is not significant, and a carefully chosen all-or-nothing strategy achieves a fraction 1 of the optimal asymptotic welfare.

# 6 Relation to the deterministic network formation process (Bala and Goyal (2000))

We start by comparing our results on mixed-degree equilibria to the pure equilibria of Bala and Goyal (2000). In their analysis of pure equilibria, Bala and Goyal identify four cost domains:

- **Lower domain** ($c < b(1) - b(2)$): The unique equilibrium is a complete network, with each link between a pair sponsored by exactly one of the agents.
- **Intermediate domain** ($(b(1) - b(2)) < c < b(1)$): The star network forms an equilibrium, with each link sponsored by either the center or the periphery agent.
- **Upper domain** ($b(1) < c < b(1) + \frac{(n-2)}{2}b(2)$): The unique equilibria consist of peripherally sponsored stars.
- **Infeasible domain** ($b(1) + \frac{(n-2)}{2}b(2) < c$): Only the empty network is stable.

Note first that interestingly the anonymous equilibria analysis essentially identifies the same cost domains as Bala and Goyal (2000) where the last cost domain does not appear in our analysis since we take the population size $n$ to infinity. In the lower-cost domain, we identify that the anonymous equilibrium has the property that agents choose to link uniformly to a fraction $1 - \frac{c}{b(1)-b(2)}$ of the population. Thus, instead of a complete network, we have a linear out-degree network. The complete network does not arise as an equilibrium in our model because implementing it with an anonymous symmetric strategy profile results in double investment into links.

In the intermediate and upper cost domain, our analysis of the mixed-degree equilibrium shows that in anonymous equilibrium networks, agents link to approximately $\sqrt{n}$ of the population of size $n$. The equilibrium star network of Bala and Goyal minimizes the total cost across all networks with a diameter 2. Similarly, the $\sqrt{n}$ threshold has the property that it is the minimal mixed-degree strategy to get a network such that each pair of agents lie within distance 2 of each other with high probability.

Finally, our finite degree equilibrium in the upper cost domain has no equivalent in the pure equilibrium analysis. In particular, the finite degree equilibrium may exhibit nonmonotonic behavior of the degree with respect to the cost.

As for welfare-maximizing networks, the pure strategy analysis identifies the following networks as a function of the cost:

1. For $c < 2(b(1) - b(2))$, the complete network is efficient.

2. For $2(b(1) - b(2)) < c < 2b(1) + (n-2)b(2)$, stars are efficient.

3. For $2b(1) + (n-2)b(2) < c$, the empty network is efficient.

It follows from Proposition 2 that, as in the equilibrium analysis, the cost ranges across which the structure of the welfare-optimizing network is constant are identical in the anonymous case and the pure case identified by Bala and Goyal (2000). For the low cost range, the efficient mixed-degree strategy identified in Proposition 2 approximates the complete network while avoiding costly double investment. As discussed above, the efficient network in the higher cost range approximate the star network.

# 7 Strategic Origins of Community Networks

One of our main contributions is to bridge the traditionally separate fields of strategic and random network formation. In particular, we show that anonymous equilibrium behavior induces a random $\mathcal{G}_{k-\text{out}}$ network formation process that is similar to the Erdős-Rényi model. This establishes a novel link between strategic interaction and a class of random networks, demonstrating that random network structures can emerge from rational decision-making in anonymous settings.

Erdős-Rényi (ER) graphs provide a simple model that allows researchers to derive analytical results on key network properties such as degree distribution, clustering, and connectivity. Their tractability makes ER graphs a widely used, valuable baseline model for understanding more complex network structures.

Despite their mathematical appeal, however, Erdős-Rényi ($G(n,p)$) and $\mathcal{G}_{k-\text{out}}$ graphs fail to capture several key structural features of real-world social networks. One such feature is the presence of pronounced *community structure*, a property extensively documented in the empirical and theoretical literature (see, e.g., Zachary (1977); Girvan and Newman (2002); Newman and Girvan (2004)).

A network exhibits community structure if its nodes can be partitioned into groups (or "communities") such that links within groups are significantly denser than links between groups. This modular organization reflects the empirical observation that social ties tend to concentrate within social circles, clubs, or workplaces, forming cohesive subgroups embedded in a larger sparse network.

Community structure in social networks is often attributed to *homophily*, which is the tendency of individuals to form ties with others who share similar attributes, such as age, profession, education level, or cultural background. Since ER and $\mathcal{G}_{k-\text{out}}$ graphs assume edges form independently of node attributes, they fail to generate the assortative mixing patterns and modular clustering observed in real social systems. These limitations underscore the importance of network formation models that incorporate both strategic interactions and homophily-driven group formation.

## 7.1 Anonymous Network Formation with Homophily

We illustrate how our anonymity idea can be extended to generate social networks that exhibit realistic community structure by explicitly incorporating homophily. Our approach builds upon the island model introduced in Jackson and Rogers (2005).

Let $\ell \in \mathbb{N}$ be the number of *communities* and consider a set of agents $N = \{1, 2, \ldots, \ell n\}$. Community $m$ is given by the set of agents

$$G_m = \{(m-1)n + 1, \ldots, mn\}, \quad m = 1, 2, \ldots, \ell.$$

We assume that all members of the community $G_m$ share the same homophily type, denoted $\theta_m$.

As in our model, agents may form costly links with one another. The cost of forming a link depends on the homophily types involved. If agents $i$ and $j$ share type $\theta_m$, the cost is $\underline{c}$, whereas if they belong to different types, the cost is $\overline{c}$, with $0 < \underline{c} < b(1) - b(2) < b(1) < \overline{c}$. This condition ensures that within–type links are always less costly than across–type links, reflecting homophily. We consider a semi-anonymous strategy in which agents condition their linking decisions only on whether the other agent is of the same type. Concretely, each type–$\theta_m$ agent adopts a mixed strategy with expected degrees $(k_{\text{in}}, k_{\text{out}})$, where

- $k_{\text{in}}$ denotes the expected number of links formed with agents of the same type, and
- $k_{\text{out}}$ denotes the expected number of links formed with agents from each other type.

We note that in this setting, the existence of a mixed-degree strategy equilibrium is not guaranteed. However, the existence of a semi-anonymized equilibrium defined above follows standard fixed-point considerations.

Our analysis yields the following result.

**Proposition 3.** Let $\left(k_{\text{in},n},\, k_{\text{out},n}\right)$ be the expectation of an equilibrium strategy profile in the game defined above. Then

$$\lim_{n\to\infty} \frac{k_{\text{in},n}}{n} \;=\; 1 - \frac{\underline{c}}{b(1) - b(2)}.$$

In addition, there exist $c_1, c_2 > 0$ such that $c_1 \log(n) \leq k_{\text{out},n} \leq c_2 \log(n)$.

In equilibrium, agents balance the benefits of connectivity against the variation in link formation costs across types. The resulting network features dense links within groups, as $k_{\text{in},n}$ scales linearly with group size, and relatively sparse links between groups, as $k_{\text{out},n}$ grows only logarithmically. This contrast in link density within and across groups provides a strategic foundation for the emergence of community structure, with cohesive clusters forming around shared types and limited interaction across groups. The model thus offers a microfoundation for modular social networks shaped by homophily and strategic behavior.

# 8 Conclusion

This paper studies the strategic formation of networks under the constraint of anonymity, meaning individuals do not distinguish between potential partners. This seemingly small change fundamentally shifts the nature of network creation, moving away from deterministic, identity-based links to a world where random network structures emerge from strategic choices. A core contribution of our work is demonstrating how strategic decisions under anonymity can lead to specific kinds of random networks. Instead of being a purely mechanical phenomenon, we show that network randomness arises naturally in certain settings through strategic choices about socialization.

Our analysis has unveiled how network structure changes dramatically depending on the cost of forming links. At low link costs, agents naturally form denser, highly connected networks. Intermediate costs lead to networks that are sparser but still have small-world properties, with nodes in the graph within close range to each other on average. In contrast, for high link costs, we observed two qualitatively different equilibria: a “low-degree” scenario resulting in networks that are tree-like with large diameters and a “high-degree” scenario where nodes link to a number of other nodes that is approximately the square root of the population size. The analysis reveals a crucial duality where the underlying features of networks diverge significantly depending on equilibrium. We also note that our strategic approach shares regime change points with the classical pure strategy literature in large populations. Mixed-degree strategies within the model can be understood as representing an agent’s ”socialization effort,” which can have real-world interpretations for those contexts where individuals invest effort to be more social (see, e.g. Cabrales et al. 2011).

Furthermore, our study explored the efficiency of these networks, revealing a key tension between equilibrium choices and collectively optimal outcomes. We find that in all cost regimes, equilibrium connectedness differs from the efficient one. This highlights a gap where there could be scope for interventions to increase the social benefit produced by the network.

This research opens several new avenues for further exploration. We provide results for the two-way flow model, but we expect that the qualitative results regarding the emergence of random networks and the strategic underpinnings of those results will likely hold under a one-way flow model as well, since the underlying social mechanism in which players derive a social benefit from being connected to other agents remains the same. Another important direction is exploring anonymous models where agents face random edge costs. For example, agents could face costs that are independently drawn from some specified distribution. These costs would allow for interim heterogeneity and introduce a new layer of stochasticity in the network formation process. Exploring these avenues will provide a deeper understanding of the interplay between random social networks, link costs, and strategic individual behavior.

# A Concavity

In this section, we prove the concavity properties of $u(\cdot; x)$. As the following result shows, $u_i$ is submodular in agent $i$'s action $s_i$.

**Lemma 1.** For any profile $s_{-i}$ and $T, T' \in 2^{N\setminus i}$,

$$u_i(s_{-i}, T) + u_i(s_{-i}, T') \geq u_i(s_{-i}, T \cap T') + u_i(s_{-i}, T \cup T').$$

Moreover, if $d_{g(s_{-i},T\cap T')}(i,j) \neq \max(d_{g(s_{-i},T)}(i,j), d_{g(s_{-i},T')}(i,j))$ for some $j$, then this inequality is strict.

*Proof.* Let $j \in N \setminus i$. For convenience of notation, we will take $d_g(i,j) = \infty$ if $j \notin N_i(g)$ and $b(\infty) = 0$. Since a path from $i$ to $j$ occurs in $g(s_{-i}, T \cup T')$ if and only if it appears in either $g(s_{-i}, T)$ or $g(s_{-i}, T')$,

$$d_{g(s_{-i},T\cup T')}(i,j) = \min\{d_{g(s_{-i},T)}(i,j), d_{g(s_{-i},T')}(i,j)\},$$

and since a path from $i$ to $j$ occurs in $g(s_{-i}, T \cup T')$ if and only if it appears in both $g(s_{-i}, T)$ and $g(s_{-i}, T')$,

$$d_{g(s_{-i},T\cap T')}(i,j) \geq \max\{d_{g(s_{-i},T)}(i,j), d_{g(s_{-i},T')}(i,j)\}.$$

Hence,

$$\begin{aligned} b(d_{g(s_{-i},T\cup T')}(i,j)) + b(d_{g(s_{-i},T\cap T')}(i,j)) &\leq b(\min\{d_{g(s_{-i},T)}(i,j), d_{g(s_{-i},T')}(i,j)\}) \\ &\quad + b(\max\{d_{g(s_{-i},T)}(i,j), d_{g(s_{-i},T')}(i,j)\}) \\ &= b(d_{g(s_{-i},T)}(i,j)) + b(d_{g(s_{-i},T')}(i,j)). \end{aligned}$$

Summing over $j$, it follows that

$$B_i(g(s_{-i}, T \cap T')) + B_i(g(s_{-i}, T \cup T')) \leq B_i(g(s_{-i}, T)) + B_i(g(s_{-i}, T')).$$

Hence,

$$\begin{aligned} u_i(s_{-i}, T) + u_i(s_{-i}, T') &= B_i(g(s_{-i}, T)) + B_i(g(s_{-i}, T')) - c \cdot (|T| + |T'|) \\ &\geq B_i(g(s_{-i}, T \cap T')) + B_i(g(s_{-i}, T \cup T')) - c \cdot (|T \cap T'| + |T \cup T'|) \\ &= u_i(s_{-i}, T \cap T') + u_i(s_{-i}, T \cup T'). \end{aligned}$$

□

This implies that an agent's utility for degree strategies is always weakly concave, and also provides some conditions under which the utility is strictly concave.

**Proposition 4.** For any mixed strategy profile $x_{-i}$ and any $0 \le k < n-2$,

$$u_i(x_{-i}, k+2) - u_i(x_{-i}, k+1) \le u_i(x_{-i}, k+1) - u_i(x_{-i}, k).$$

Moreover, if $x_{-i}$ is an anonymous strategy profile and there is some $0 < m < n-2$ such that for all $j \neq i$, $m$ in the support of $x_j$, then this inequality is strict for $0 \le k \le n-m-2$.

*Proof.* Let $\pi$ be a uniformly randomly drawn permutation of $N \backslash i$, and let $T = \{\pi_1, \ldots, \pi_{k+1}\}$ and $T' = \{\pi_1, \ldots, \pi_k, \pi_{k+2}\}$. Then by Lemma 1 for any $s_{-i}$,

$$\begin{aligned} 2u_i(s_{-i}, k+1) &= \mathbb{E}_\pi[u_i(s_{-i}, T)] + \mathbb{E}_\pi[u_i(s_{-i}, T')] \\ &\ge \mathbb{E}_\pi[u_i(s_{-i}, T \cap T')] + \mathbb{E}_\pi[u_i(s_{-i}, T \cup T')] \\ &= u_i(s_{-i}, k) + u_i(s_{-i}, k+2). \end{aligned}$$

Hence,

$$\begin{aligned} u_i(x_{-i}, k+2) - u_i(x_{-i}, k+1) &= \mathbb{E}_{s_{-i} \sim x_{-i}}[u_i(s_{-i}, k+2) - u_i(s_{-i}, k+1)] \\ &\le \mathbb{E}_{s_{-i} \sim x_{-i}}[u_i(s_{-i}, k+1) - u_i(s_{-i}, k)] \\ &= u_i(x_{-i}, k+1) - u_i(x_{-i}, k). \end{aligned}$$

It follows immediately that if there is some $s_{-i}$ in the support of $x_{-i}$ and some $T$ and $T'$ such that

$$u_i(s_{-i}, T) + u_i(s_{-i}, T') > u_i(s_{-i}, T \cap T') + u_i(s_{-i}, T \cup T'),$$

then this inequality is strict. We prove the last claim by showing that if the condition holds, then there are always such $s_{-i}$, $T$, and $T'$.

For convenience, take $i = n$. Consider first the case $m > 1$. For $1 \le j \le m$, let $s_j = \{1, \ldots, m+1\} \setminus j$, and for $m+1 \le j \le n-1$, let $s_j = \{1, \ldots, m\}$. Let $T = \{1\} \cup \{m+1, \ldots, m+k\}$ and $T' = \{2\} \cup \{m+1, \ldots, m+k\}$. Since $|s_j| = m$ for $j \neq n$, $s_{-n}$ is in the support of $x_{-n}$. Moreover,

$$d_{g(s_{-n}, T)}(n, n-1) = d_{g(s_{-n}, T')}(n, n-1) = 2$$

and

$$d_{g(s_{-n},T\cap T')}(n,n-1)=3.$$

Hence, strict inequality follows from Lemma 1.

Now, if $m=1$, let $s_j=\{j+1\}$ for $j\in\{1,\ldots,n-5\}\cup\{n-3,n-2\}$, and let $s_{n-4}=\{1\}$ and $s_{n-1}=\{n-3\}$. Let $T=\{1,\ldots,k\}\cup\{n-3\}$ and $T'=\{1,\ldots,k\}\cup\{n-2\}$. Then $|s_j|=1$ for $j\neq n$, so $s_{-n}$ is in the support of $x_{-n}$,

$$d_{g(s_{-n},T)}(n,n-1)=d_{g(s_{-n},T')}(n,n-1)=2$$

and

$$d_{g(s_{-n},T\cap T')}(n,n-1)=\infty,$$

and so the result again follows from Lemma 1. □

Proposition 1 follows easily from Proposition 4.

*Proof of Proposition 1.* Concavity follows immediately from the first part of Proposition 4. If $x$ has full support, then strict concavity follows from the second part of Proposition 4 by taking $m=1$. □

# B Equilibrium characterization

In this section, we prove that every symmetric anonymous equilibrium must be one of the three types identified in Section 3.

**Proposition 5.** If $x$ corresponds to a symmetric anonymous equilibrium, then $x$ is either mixed-degree, all-or-nothing, or more-than-half.

We will use the following result, which states that the maximizers of $u(\cdot; x)$ must be an interval. This is a straightforward implication of concavity.

**Lemma 2.** For any anonymous strategy $x$ there exist $\underline{k}$ and $\bar{k}$ such that $k \in \arg\max u(\cdot; x)$ if and only if $\underline{k} \leq k \leq \bar{k}$.

*Proof.* Let $\underline{k} = \min \arg\max u(\cdot; x)$ and $\bar{k} = \max \arg\max u(\cdot; x)$. If $\underline{k} = \bar{k}$, then the claim is immediate, so assume $\underline{k} < \bar{k}$, and suppose towards a contradiction that the claim does not hold. Then there is some $\underline{k} < k < \bar{k}$ such that $u(k+1; x) > u(k; x)$. But then by Proposition 1, $u(i+1; x) - u(i; x)$ is decreasing in $i$, so

$$u(k+1; x) - u(\underline{k}; x) = \sum_{i=\underline{k}}^{k} u(i+1; x) - u(i; x) \geq \sum_{i=\underline{k}}^{k} u(k+1; x) - u(k; x) > 0,$$

and hence $u(k+1; x) > u(\underline{k}; x) = \max u(\cdot; x)$, contradiction. □

*Proof of Proposition 5.* Let $x$ be an anonymous strategy which corresponds to a symmetric equilibrium. Suppose first that 0 and $m$ are in the support of $x$ for some $1 \leq m < n-1$. Then by Proposition 4, $u(\cdot; x)$ is strictly concave at 1, so

$$u(2; x) - u(1; x) < u(1; x) - u(0; x) \leq 0,$$

and so $u(2; x) < u(1; x) \leq u(0; x)$. By Lemma 2, it follows that $Supp(x) \subseteq \{0, 1\}$. Thus, if $x$ is not all-or-nothing and contains 0 in its support, then $x$ must be mixed-degree.

Next, suppose that 0 is not in the support of $x$, and let $m = \min Supp(x)$. If $m \geq (n-1)/2$, then $x$ is more-than-half, so assume $m < (n-1)/2$. By Proposition 4, $u(\cdot; x)$ is strictly concave at $m+1$, so

$$u(m+2; x) < 2 \cdot u(m+1; x) - u(m; x) \leq u(m; x).$$

Hence, by Lemma 2, $Supp(x) \subseteq \{m, m+1\}$, so $x$ is mixed-degree. □

# C Non-mixed-degree equilibria

In this section, we provide proofs for the characterization of all-or-nothing and more-than-half equilibria.

## C.1 All-or-nothing

As we will show, when all other agents are playing the all-or-nothing strategy which mixes between degree strategies $n-1$ and 0 with probability $p$ and $1-p$, respectively, the marginal benefit of making an additional link is

$$\alpha(p) = (1-p) \cdot [b(1) - (1-(1-p)^{n-2}) \cdot b(2)].$$

In an all-or-nothing equilibrium, this marginal benefit must be offset by the marginal cost of an additional link, and as the following result shows, this turns out to be essentially necessary and sufficient.

**Proposition 6.** For any $c$ and $b$, there is a unique all-or-nothing strategy $x$ which corresponds to a symmetric anonymous equilibrium. If $c \geq b(1)$, then $x$ is the degree strategy 0 (i.e. the trivial equilibrium). If $c < b(1)$, then $x$ assigns nonzero probability $p^*$ to the degree strategy $n-1$, where $\alpha(p^*) = c$.

*Proof.* $x$ corresponds to a symmetric anonymous equilibrium if and only if 0 and $n-1$ maximize $u(\cdot; x)$, so by Lemma 2, $x$ corresponds to a symmetric anonymous equilibrium if and only if $u(\cdot; x)$ is constant.

Let $p$ be the probability that $x$ assigns to $n-1$. If all agents apart from, say, agent $n$ use the strategy $x$ and agent $n$ uses the strategy $s_n = T$, then the benefit agent $n$ gets from agent $i$ is $b(1)$ if $i \in T$ or $|s_i| = n-1$, $b(2)$ if $i \notin T$, $|s_i| = 0$, and $|s_j| = n-1$ for some $j \neq i$, and 0 otherwise, so

$$\begin{aligned}
\mathbb{E}[u_n(s_{-i}, T)] &= (|T| + p \cdot (n-1-|T|)) \cdot b(1) \\
&\quad + (n-1-|T|) \cdot [(1-p) \cdot (1-(1-p)^{n-2})] \cdot b(2) - |T| \cdot c \\
&= (\alpha(p) - c) \cdot |T| + (b(1) - \alpha(p)) \cdot (n-1)
\end{aligned}$$

and thus

$$u(k; x) = (\alpha(p) - c) \cdot k + (n-1) \cdot \alpha(p).$$

Hence, $u(\cdot;x)$ is constant if and only if $\alpha(p) = c$.

Now, since $\alpha(p)$ is a product of strictly decreasing functions, $\alpha(p)$ is strictly decreasing on $[0, 1]$. Since $\alpha(0) = b(1)$ and $\alpha(1) = 0$, this has a unique (and nonzero) solution if $c < b(1)$. If $c \geq b(1)$, this has no nonzero solution, and in this case the degree strategy 0 corresponds to a symmetric anonymous equilibrium, since 0 maximizes $u(k; 0) = k \cdot (b(1) - c)$. $\square$

## C.2 More-than-half

**Proposition 7.** A more-than-half strategy $x$ which is not mixed-degree corresponds to a symmetric anonymous equilibrium if and only if the expected out-degree under $x$ is $(1 - \frac{c}{b(1)-b(2)}) \cdot (n - 1)$.

*Proof.* Denote by $p_k$ the probability of choosing the degree strategy $k$ under $x$ and by $\tilde{k} = \sum_k p_k \cdot k$ the expected out-degree. By the pigeonhole principle, if agent $n$ uses a degree strategy $k \geq (n-1)/2$ when all other agents use the strategy $x$, then the distance to any other agent is at most two, so the benefit agent $n$ gets from agent $i$ is $b(1)$ if $i \in s_n$ or $n \in s_i$ and $b(2)$ otherwise. Note that

$$\mathbb{P}[n \notin s_i] = \sum_k \mathbb{P}[|s_i| = k] \cdot \mathbb{P}[n \notin s_i \mid |s_i| = k] = \sum_k p_k \cdot (1 - \frac{k}{n-1}) = 1 - \frac{\tilde{k}}{n-1},$$

so

$$\begin{aligned} u(k; x) &= (n-1) \cdot (b(1) + (1 - \frac{k}{n-1}) \cdot (1 - \frac{\tilde{k}}{n-1}) \cdot (b(2) - b(1))) - k \cdot c \\ &= u(n-1; x) + (1 - \frac{k}{n-1}) \cdot (b(1) - b(2)) \cdot \left[\tilde{k} - (1 - \frac{c}{b(1) - b(2)}) \cdot (n-1)\right]. \end{aligned}$$

Suppose first that $\tilde{k} = (1 - \frac{c}{b(1)-b(2)}) \cdot (n-1)$. Then $u(\cdot;x)$ is constant for $k \geq (n-1)/2$. Furthermore, since $u(n-1;x) \geq u(n-2;x)$, by Proposition 1, $u(\cdot;x)$ is increasing. Hence, $k$ maximizes $u(\cdot;x)$ for all $k \geq (n-1)/2$, so $x$ corresponds to a symmetric anonymous equilibrium.

Now, suppose that $x$ corresponds to a symmetric anonymous equilibrium. If $\tilde{k} \neq (1 - \frac{c}{b(1)-b(2)}) \cdot (n - 1)$, then $u(\cdot;x)$ is either strictly increasing or strictly decreasing, so $x$ must be either the degree strategy $\lceil (n-1)/2 \rceil$ or $n - 1$. But by assumption this is impossible, since $x$ is not mixed-degree. $\square$

# D Mixed-degree cost correspondence

For this section, fix $b$. Recall that $E(c,n)$ is the set of $\kappa \in [0, n-1]$ such if there are $n$ agents and the cost is $c$, then the symmetric anonymous strategy profile where all agents play the mixed-degree strategy $\kappa$ is an equilibrium.

Denote by $b(k;\kappa)$ the expected benefit of playing the degree strategy $k$ when all other agents play the mixed-degree strategy $\kappa$. Note that $b(k;\kappa)$is continuous in $\kappa$. For $\kappa \in [0, n-1] \setminus \{0, \ldots, n-1\}$, define

$$C(\kappa) = b(\lceil \kappa \rceil; \kappa) - b(\lfloor \kappa \rfloor; \kappa).$$

For $k \in \{1, \ldots, n-2\}$, denote $C(k^-) = \lim_{\kappa \to k^-} C(\kappa)$ and $C(k^+) = \lim_{\kappa \to k^+} C(\kappa)$.

**Proposition 8.** Let $\kappa \in [0, n-1]$, $c \geq 0$. Then $\kappa \in E(c,n)$ if and only if one of the following holds:

- $\kappa \notin \{0, \ldots, n-1\}$ and $c = C(\kappa)$
- $\kappa \in \{1, \ldots, n-2\}$ and $C(\kappa^+) \leq c \leq C(\kappa^-)$
- $\kappa = 0$ and $c \geq b(1)$
- $\kappa = n-1$ and $c = 0$

*Proof.* If $\kappa \notin \{0, \ldots, n-1\}$, let $k = \lfloor \kappa \rfloor$. Then $\kappa \in E(c,n)$ if and only if $k, k+1 \in \arg\max u(\cdot;\kappa)$. By Proposition 1, $u(\cdot;\kappa)$ is concave, so this holds if and only if $u(k;\kappa) = u(k+1;\kappa)$. Since

$$u(k+1;\kappa) - u(k;\kappa) = [b(k+1;\kappa) - (k+1)\cdot c] - [b(k;\kappa) - k \cdot c] = C(\kappa) - c,$$

this holds if and only if $c = C(\kappa)$.

If $\kappa \in \{1, \ldots, n-2\}$, then by Proposition 1, $u(\cdot;k)$ is concave, so $k \in E(c,n)$ if and only if $k \in \arg\max u(\cdot;k)$ if and only if $u(k;k) \geq \max(u(k-1;k), u(k+1;k))$. Now,

$$u(k+1;k) - u(k;k) = b(k+1;k) - b(k;k) - c = C(k^+) - c$$

and

$$u(k;k) - u(k-1;k) = b(k;k) - b(k-1;k) - c = C(k^-) - c,$$

so this holds if and only if $C(k^+) - c \leq 0$ and $C(k^-) - c \geq 0$.

If $\kappa = 0$, then $u(k; 0) = k \cdot (b(1) - c)$, so $0 \in \arg\max u(\cdot; 0)$ if and only if $c \geq b(1)$.

If $\kappa = n - 1$, then $u(k; n-1) = (n-1) \cdot b(1) - k \cdot c$, so $n - 1 \in \arg\max u(\cdot; n-1)$ if and only if $c = 0$. $\square$

We use the next result for proving existence of mixed-degree equilibria.

**Lemma 3.** Let $\kappa_1, \kappa_2 \in [0, n-1]$ with $\kappa_1 < \kappa_2$ and let $c_1, c_2 \geq 0$. If $\kappa_1 \in E(c_1, n)$ and $\kappa_2 \in E(c_2, n)$, then for every $c$ with $\min(c_1, c_2) \leq c \leq \max(c_1, c_2)$ there is a $\kappa$ with $\kappa_1 \leq \kappa \leq \kappa_2$ such that $\kappa \in E(c, n)$.

*Proof.* If $\kappa_1 \in E(c, n)$ or $\kappa_2 \in E(c, n)$, then the result is immediate, so assume this is not the case. Suppose first that that $c_1 < c_2$. Since $\kappa_1, \kappa_2 \notin E(c, n)$,

$$\lim_{\tilde{k} \to \kappa_1^+} C(\tilde{\kappa}) < c < \lim_{\tilde{k} \to \kappa_2^-} C(\tilde{\kappa}).$$

Let

$$\kappa = \sup\{\tilde{\kappa} \in (\kappa_1, \kappa_2) \setminus \{0, \ldots, n-1\} \,:\, C(\tilde{\kappa}) < c\}.$$

Observe that $\kappa < \kappa_2$, since

$$\lim_{\tilde{\kappa} \to \kappa^-} C(\tilde{\kappa}) \leq c < c_2 \leq \lim_{\tilde{\kappa} \to \kappa_2^-} C(\tilde{\kappa}).$$

Hence, $\lim_{\tilde{\kappa} \to \kappa^-} C(\tilde{\kappa}) \geq c$, since otherwise $\lim_{\tilde{\kappa} \to \kappa^+} C(\tilde{\kappa}) < c$, contradicting the definition of $\kappa$. It follows that $\lim_{\tilde{\kappa} \to \kappa^-} C(\tilde{\kappa}) = c$, so $\kappa \in E(c, n)$.

An analogous argument shows that if $c_1 > c_2$ and

$$\kappa = \sup\{\tilde{\kappa} \in (\kappa_1, \kappa_2) \setminus \{0, \ldots, n-1\} \,:\, C(\tilde{\kappa}) > c\},$$

then $\kappa \in E(c, n)$. $\square$

# E Proof of Theorem 1

Throughout this section, we prove Theorem 1 using the results established in the previous Appendices. To begin, the following result establishes the existence of nontrivial equilibria for large $n$.

**Claim 3.** For any $c > 0$ and sufficiently large $n$, there exists a $\kappa > \sqrt{n}$ such that $\kappa \in E(c, n)$.

*Proof.* By Lemma 3, it is sufficient to show that for all large enough $n$, there are $\kappa_1, \kappa_2 > \sqrt{n}$ such that $C(\kappa_1) > c$ and $C(\kappa_2) < c$, since this implies that there is some $\kappa > \sqrt{n}$ such that $\kappa \in E(c, n)$. By Proposition 8, $n - 1 \in E(0, n)$, so we may take $\kappa_2 = n - 1$.

Now, let $\alpha, \beta$ as in Lemma 5. Then for all sufficiently large $n$,

$$C(\sqrt{n}^+) = \lim_{\kappa \to \sqrt{n}^+} C(\kappa) \geq \alpha \cdot e^{-\beta} \cdot \sqrt{n} > c.$$

In particular, there exists some $\kappa_1 > \sqrt{n}$ with $C(\kappa_1) > c$. □

Now, by Proposition 5, every symmetric anonymous equilibrium is either mixed-degree, all-or-nothing, or more-than-half. By Proposition 6, for any $c$ there is a unique all-or-nothing symmetric anonymous equilibrium, which is nontrivial if $c < b(1)$ and trivial otherwise. By Proposition 7, a more-than-half strategy corresponds to a symmetric anonymous equilibrium if and only if the expected out-degree is $(1 - \frac{c}{b(1)-b(2)}) \cdot (n - 1)$. Parts $(b)$ and $(c)$ of Theorem 1 follow immediately from these results. The following result shows that part $(a)$ also follows from these results.

**Claim 4.** If $c \geq b(1)$, there is no non-mixed-degree symmetric anonymous equilibrium.

*Proof.* Since $c \geq b(1)$, the only all-or-nothing symmetric anonymous equilibrium is for all agents to choose $k = 0$, which is mixed-degree. Moreover, $(1 - \frac{c}{b(1)-b(2)}) \cdot (n - 1) < 0$, so there is no more-than-half strategy with this expected degree. □

# F Proof of Theorem 2

Throughout this section, we prove Theorem 2, making use of the results established in the previous Appendices, as well as several bounds which we provide proofs for in the next Appendix.

## F.1 Low cost regime

For part $(a)$, we make use of the following result.

**Lemma 4.** Let $\kappa \in [0, n-1] \setminus \{0, \ldots, n-1\}$. Then

$$C(\kappa) \geq \left(1 - \frac{\kappa}{n-1}\right) \cdot (b(1) - b(2)).$$

Moreover, if $\kappa > 10$, then

$$C(\kappa) \leq \left(1 - \frac{\kappa}{n-1}\right) \cdot (b(1) - b(2)) + n^3 \cdot e^{-\frac{1}{2}\frac{\kappa^2}{n}}.$$

**Claim 5.** Suppose $c < b(1) - b(2)$, and let $\kappa_n \in E(c, n)$ for each $n$. Then

$$\lim_{n \to \infty} \frac{\kappa_n}{n} = 1 - \frac{c}{b(1) - b(2)}.$$

*Proof.* Let $\rho = 1 - \frac{c}{b(1)-b(2)}$. If $\kappa < \rho \cdot (n-1)$, then by Lemma 4,

$$C(\kappa) > \left(1 - \frac{\rho \cdot (n-1)}{n-1}\right) \cdot (b(1) - b(2)) = c,$$

so $\kappa_n > \rho \cdot (n-1)$.

Now, fix $\varepsilon > 0$. For all sufficiently large $n$, if $\kappa > (1+\varepsilon) \cdot \rho \cdot (n-1)$, then $\kappa > 10$, so by Lemma 4,

$$\begin{aligned} C(\kappa) &\leq \left(1 - \frac{(1+\varepsilon) \cdot \rho \cdot (n-1)}{n-1}\right) \cdot (b(1) - b(2)) + n^3 \cdot e^{-\frac{1}{2}\frac{((1+\varepsilon)\cdot\rho\cdot(n-1))^2}{n}} \\ &= c - \varepsilon \cdot \rho \cdot (b(1) - b(2)) + n^3 \cdot e^{-\frac{1}{2}\frac{((1+\varepsilon)\cdot\rho\cdot(n-1))^2}{n}} \\ &< c, \end{aligned}$$

so $\kappa_n \leq (1+\varepsilon) \cdot \rho \cdot (n-1)$. □

## F.2 Intermediate cost regime

For part $(b)$, we will need the following bounds.

**Lemma 5.** There are $\alpha, \beta > 0$ such that if $n > 100$, $2 < \kappa < \frac{1}{5}n$, and $\kappa \in E(c, n)$, then $c \geq \alpha \cdot e^{-\beta \frac{\kappa^2}{n}} \cdot \kappa$.

**Lemma 6.** For $\kappa \in [0, n-1] \setminus \{0, \ldots, n-1\}$,

$$C(\kappa) \geq \left(1 - \frac{\log n}{n-1}\right) \cdot (b(1) - b\left(\left\lfloor \frac{\log \log n}{\log 2(\kappa+1)} \right\rfloor\right))$$

**Claim 6.** Suppose $c \in (b(1) - b(2), b(1))$, and let $\kappa_n \in E(c, n)$ for each $n$. Then

$$\lim_{n \to \infty} \frac{\log(\kappa_n)}{\log(n)} = \frac{1}{2}.$$

*Proof.* Let $\alpha, \beta$ as in Lemma 5, and fix some integer $K > \frac{c}{\alpha e^{-\beta}}$. Then for all sufficiently large $n$, if $K < \kappa < \sqrt{n}$ and $\kappa \in E(c', n)$, then

$$c' > \alpha \cdot e^{-\beta} \cdot K > c,$$

so either $\kappa_n \leq K$ or $\kappa_n \geq \sqrt{n}$. Moreover, by Lemma 6, for $n$ sufficiently large, if $\kappa \leq K$ and $\kappa \in E(c', n)$, then

$$c' \geq \inf_{\kappa \in [0,K] \setminus \{0,\ldots,n-1\}} C(\kappa) \geq \left(1 - \frac{\log n}{n-1}\right) \cdot (b(1) - b\left(\left\lfloor \frac{\log \log n}{\log 2(\kappa+1)} \right\rfloor\right)) > c,$$

so $\kappa_n \geq \sqrt{n}$. Finally, by Lemma 4, for any $\varepsilon > 0$, if $n$ is sufficiently large and $\kappa > n^{\frac{1}{2}+\varepsilon}$, then

$$C(\kappa) < b(1) - b(2) + n^3 e^{-\frac{1}{2} n^{2\varepsilon}} < c,$$

so $\kappa_n \leq n^{\frac{1}{2}+\varepsilon}$. □

Now, fixing $\varepsilon \in (0, \frac{1}{2})$, it follows that $\kappa_n > n^{\frac{1}{2}-\varepsilon}$ for all sufficiently large $n$. By Lemma 5, for large $n$ and $\kappa \in [n^{\frac{1}{2}-\varepsilon}, \sqrt{n}]$, if $\kappa \in E(c, n)$, then $c \geq \alpha \cdot e^{-\beta} \cdot n^{\frac{1}{2}-\varepsilon} > b(1) - b(2)$, so $\kappa_n > \sqrt{n}$.

## F.3 High cost regime

For every $n$, choose $c_n$ such that $\sqrt{n} \in E(c_n, n)$. By Proposition 8 and Lemma 5, for all sufficiently large $n$,

$$c_n \geq C(\sqrt{n}^+) \geq \alpha \cdot e^{-\beta} \cdot \sqrt{n} > c.$$

Moreover, by Proposition 8, $0 \in E(b(1), n)$ and $n - 1 \in E(0, n)$. Thus, by Theorem 3, there are $\kappa, \kappa' \in E(c, n)$ with $\kappa < \sqrt{n}$ and $\kappa' > \sqrt{n}$.

By Lemma 5, if $n > 100$ and $\kappa_n$ is a low-degree equilibrium with $\kappa_n > 2$, then $\kappa_n \leq \frac{c}{\alpha \cdot e^{-\beta}}$. Hence, if $(\kappa_n)$ is a sequence of low-degree equilibria, then $\kappa_n \leq \max(2, \frac{c}{\alpha \cdot e^{-\beta}})$ for all $n > 100$, so $(\kappa_n)$ is bounded. The asymptotic characterization of high-degree equilibria is a straightforward application of Lemma 4.

**Claim 7.** If $c > b(1)$ and $\kappa_n \in E(c, n)$ is a high-degree equilibrium for every sufficiently large $n$, then

$$\lim_{n \to \infty} \frac{\log(k_n)}{\log(n)} = \frac{1}{2}.$$

*Proof.* By Lemma 4, for any $\varepsilon > 0$, if $n$ is sufficiently large and $\kappa > n^{\frac{1}{2}+\varepsilon}$, then

$$C(\kappa) < b(1) - b(2) + n^3 e^{-\frac{1}{2}n^{2\varepsilon}} < c.$$

Hence, for all sufficiently large $n$, $\sqrt{n} < \kappa_n \leq n^{\frac{1}{2}+\varepsilon}$. □

# G Proofs of bounds

We will make use of the following bound several times.

**Lemma 7.** Let $\kappa \in [0, n-1] \setminus \{0, \dots, n-1\}$. If all agents apart from $n$ use the strategy $\kappa$ and agent $n$ uses the strategy $k = \lfloor \kappa \rfloor$, then

$$\mathbb{P}[d_G(n,1) < M] \leq \frac{(2(\kappa+1))^M}{n-1}.$$

*Proof.* The distance from $n$ to 1 is less than $M$ if and only if there is a path of length $m$ from $n$ to 1 for some $m < M$. Let $n = v_0, v_1, \dots, v_{m-1}, v_m = 1$ be a sequence of distinct vertices, and for $0 \leq i < m-1$, let $E_i$ be the event that $v_i \in s_{v_{i+1}}$ or $v_{i+1} \in s_{v_i}$. Then this is a path in $G$ if and only all of the $E_i$ occur. Now, $\mathbb{P}[E_0] \leq 2\frac{k+1}{n-1}$ and for $i > 0$,

$$\mathbb{P}[E_i \mid E_j \text{ for } j < i] = \mathbb{P}[E_i \mid E_{i-1}] \leq 2\frac{k+1}{n-1},$$

so the probability that this is a path in $G$ is bounded above by $\left(2\frac{k+1}{n-1}\right)^m$. Hence, by the union bound, the probability that there is a path of length $m$ from $n$ to 1 is at most $(n-1)^{m-1} \cdot \left(2\frac{k+1}{n-1}\right)^m = \frac{1}{n-1} \cdot (2(k+1))^m$, so the probability that there is a path of length less than $M$ from $n$ to 1 is at most

$$\sum_{m=0}^{M-1} \frac{1}{n-1} \cdot (2(k+1))^m = \frac{1}{n-1} \cdot \frac{(2(k+1))^M - 1}{(2(k+1)) - 1} \leq \frac{(2(\kappa+1))^M}{n-1}.$$

□

*Proof of Lemma 4.* Let $k = \lfloor \kappa \rfloor$, and let $T = \{1, \dots, k\}$ and $T' = \{1, \dots, k+1\}$. Then assuming all agents other than $n$ use the strategy $\kappa$,

$$C(\kappa) = \mathbb{E}[b_n(s_{-n}, T') - b_n(s_{-n}, T)].$$

Let $G = g(s_{-n}, T)$, and let $D$ be the event that $d_G(i,j) \leq 2$ for all $i, j$. If $d_G(n, k+1) \geq 2$, then the benefit to adding a link to $k+1$ is at least $b(1) - b(2)$, so

$$b_n(s_{-n}, T') - b_n(s_{-n}, T) \geq \mathbb{1}(d_G(n, k+1) \geq 2) \cdot (b(1) - b(2)).$$

On the other hand, the benefit is at most $nb(1)$, and if $D$ holds, then the benefit is $b(1) - b(2)$ if $d_G(n, k+1) = 2$ and 0 otherwise, so

$$\begin{aligned} b_n(s_{-n}, T') - b_n(s_{-n}, T) &\leq \mathbb{1}(d_G(n, k+1) \geq 2, D) \cdot (b(1) - b(2)) + \mathbb{1}(\neg D) \cdot nb(1) \\ &\leq \mathbb{1}(d_G(n, k+1) \geq 2) \cdot (b(1) - b(2)) + \mathbb{1}(\neg D) \cdot n. \end{aligned}$$

Hence,

$$\mathbb{P}[d_G(n,k+1) \geq 2] \cdot (b(1)-b(2)) \leq C(\kappa) \leq \mathbb{P}[d_G(n,k+1) \geq 2] \cdot (b(1)-b(2)) + \mathbb{P}[\neg D] \cdot n.$$

Since $k+1 \notin T$,

$$\mathbb{P}[d_G(n,k+1) \geq 2] = \mathbb{P}[n \notin s_{k+1}] = 1 - \frac{\kappa}{n-1},$$

so

$$C(\kappa) \geq \left(1 - \frac{\kappa}{n-1}\right) \cdot (b(1)-b(2)).$$

Now, if $D$ does not hold, then there are some $i,j$ such that $d_G(i,j) > 2$. This is only possible if $\{i,j\} \not\subseteq s_m$ for all $m \neq i,j$, so

$$\mathbb{P}[d_G(i,j) > 2] \leq \mathbb{P}[\{i,j\} \not\subseteq s_m \text{ for all } m \neq i,j] \leq \left(1 - \frac{k \cdot (k-1)}{(n-1)\cdot(n-2)}\right)^{n-2}.$$

If $\kappa > 10$, then $\frac{k\cdot(k-1)}{(n-1)} > \frac{1}{2}\frac{\kappa^2}{n}$, so the last term is bounded by $e^{-\frac{1}{2}\frac{\kappa^2}{n}}$, and hence by the union bound, $\mathbb{P}[\neg D] \leq n^2 \cdot e^{-\frac{1}{2}\frac{\kappa^2}{n}}$. Thus,

$$C(\kappa) \leq \left(1 - \frac{\kappa}{n-1}\right) \cdot (b(1)-b(2)) + n^3 \cdot e^{-\frac{1}{2}\frac{\kappa^2}{n}}.$$

□

*Proof of Lemma 5.* Assume first that $\kappa \notin \{0,\dots,n-1\}$. Let $s_{-1}$ be drawn according to the strategy profile where all agents aside from 1 use the mixed-degree strategy $\kappa$, and let $s_1$ be drawn according to agent 1 using the degree strategy $k = \lfloor \kappa \rfloor$. Observe that $b(k+1;\kappa) - b(k;\kappa)$ is equal to the expected increased benefit from agent 1 adding one more link uniformly at random. Observe that if for some agent $m$, $d_{\hat{g}(s)}(1,m) > 2$, then the probability that agent 1's benefit with respect to $m$ increases by at least $b(2)-b(3)$ is at least $\frac{k}{n-1}$, since $m$ has at least $k$ neighbors. Hence,

$$\begin{aligned} b(k+1;\kappa) - b(k;\kappa) &\geq \sum_{i=2}^{n} \mathbb{P}[d_{\hat{g}(s)}(1,i) > 2] \cdot \frac{k}{n-1} \cdot (b(2)-b(3)) \\ &= (b(2)-b(3)) \cdot \mathbb{P}[d_{\hat{g}(s)}(1,m) > 2] \cdot k. \end{aligned}$$

The result will follow from a lower bound on the probability that agents 1 and $m$ are distance more than two apart. Let $A$ be the event that $1 \notin s_m$, $m \notin s_1$, and $s_1 \cap s_m = \emptyset$, and let $B$ be the event that

- $\forall i \in s_1$, $m \notin s_i$,
- $\forall i \in s_m$, $1 \notin s_i$, and
- $\forall i \notin s_1 \cup s_m \cup \{1, m\}$, $\{1, m\} \not\subseteq s_i$.

Agents 1 and $m$ are distance more than two apart if and only if both $A$ and $B$ occur. Since $\mathbb{P}[1 \notin s_m] \geq 1 - \frac{k+1}{n-1}$, $\mathbb{P}[m \notin s_1] \geq 1 - \frac{k+1}{n-1}$, and

$$\mathbb{P}[s_1 \cap s_m = \emptyset \mid s_m, 1 \notin s_m, m \notin s_1] = \frac{\binom{n-1-(|s_m|+1)}{k}}{\binom{n-1}{k}} \geq \left(1 - \frac{2(k+1)}{n-1}\right)^{k+2},$$

we have

$$\mathbb{P}[A] \geq \left(1 - \frac{2(k+1)}{n-1}\right)^{2(k+1)}$$

and thus

$$\mathbb{P}[A] \geq e^{-8\frac{(k+1)^2}{n-1}},$$

where the last inequality follows from the fact that $1 - x \geq e^{-2x}$ for $0 \leq x \leq \frac{1}{2}$.

Now, for $i \in s_1$, $\mathbb{P}[m \notin s_i \mid A] \geq 1 - \frac{k+1}{n-1}$, for $i \in s_m$, $\mathbb{P}[1 \notin s_i \mid A] \geq 1 - \frac{k+1}{n-1}$, and for $i \notin s_1 \cup s_m \cup \{1, m\}$, $\mathbb{P}[\{1, m\} \not\subseteq s_i \mid A] \geq 1 - \left(\frac{k+1}{n-1}\right)^2$. Since these events are all independent conditional on $A$,

$$\begin{aligned}\mathbb{P}[B \mid A, s_1, s_m] &> \left(1 - \frac{k+1}{n-1}\right)^{|s_1|} \cdot \left(1 - \frac{k+1}{n-1}\right)^{|s_m|} \cdot \left(1 - \left(\frac{k+1}{n-1}\right)^2\right)^{n-(|s_1|+|s_m|+2)} \\ &> \left(1 - \frac{k+1}{n-1}\right)^{2(k+1)} \cdot \left(1 - \left(\frac{k+1}{n-1}\right)^2\right)^{n-1},\end{aligned}$$

so

$$\mathbb{P}[B \mid A] > e^{-4\frac{(k+1)^2}{n-1}} \cdot e^{-2\frac{(k+1)^2}{n-1}} = e^{-6\frac{(k+1)^2}{n-1}}.$$

Thus,

$$\mathbb{P}[d_{\hat{g}(s)}(1, m) > 2] = \mathbb{P}[A] \cdot \mathbb{P}[B \mid A] \geq e^{-14\frac{(k+1)^2}{n-1}},$$

so

$$C(\kappa) = b(k+1; \kappa) - b(k; \kappa) \geq (b(2) - b(3)) \cdot e^{-14\frac{(k+1)^2}{n-1}} \cdot k.$$

Since $\frac{(k+1)^2}{n-1} \leq 8 \cdot \frac{\kappa^2}{n}$ and $k \geq \frac{1}{2}\kappa$, taking $\alpha = \frac{b(2)-b(3)}{2}$ and $\beta = 112$ gives the result.

Finally, if $\kappa = k \in \{0, \ldots, n-1\}$, observe that

$$\frac{c}{\alpha \cdot e^{-\beta \cdot \frac{k^2}{n}} \cdot k} \geq \frac{C(k^+)}{\alpha \cdot e^{-\beta \cdot \frac{k^2}{n}} \cdot k} = \lim_{\kappa \to k^+} \frac{C(\kappa)}{\alpha \cdot e^{-\beta \cdot \frac{\kappa^2}{n}} \cdot \kappa} \geq 1.$$

□

*Proof of Lemma 6.* Let $k = \lfloor \kappa \rfloor$, and let $G$ be the random graph when agent $n$ uses the degree strategy $k$ and all other agents use the strategy $\kappa$. For any $M \in \mathbb{Z}_{>0}$, the expected benefit with respect to agent $i$ for agent $n$ of adding an additional link is bounded below by

$$\frac{1}{n-1} \cdot \mathbb{P}[d_G(n,i) \geq M] \cdot (b(1) - b(M)).$$

Summing over $i$, agent $n$'s expected benefit from an additional edge is bounded below by $\mathbb{P}[d_G(n,1) \geq M] \cdot (b(1) - b(M))$, and by Lemma 7, $\mathbb{P}[d_G(n,1) \geq M] \leq 1 - \frac{(2(\kappa+1))^M}{n-1}$.

Taking $M = \left\lfloor \frac{\log \log n}{\log 2(\kappa+1)} \right\rfloor$, it follows that

$$\mathbb{P}[d_G(n,1) < M] \leq \frac{\log n}{n-1}.$$

□

# H Proof of Proposition 2

*Proof of Proposition 2.* Consider the first assertion. Consider any sequence of anonymous strategies $\{x_n\}_n$ of the network formation games with $n$ agents such that the profile $(x_n, \ldots, x_n)$ achieves $W_n$. Let $y_n$ be the expected fraction of agents that is chosen by any agent according to $x_n$. In other words, $y_n$ equals $\frac{1}{n-1}$ times the expected outdegree of any agent. We claim that

$$W_n \leq \binom{n}{2}\Big(2b(1)(1-(1-y_n)^2) + 2b(2)(1-y_n)^2 - c[2y_n(1-y_n) + 2y_n^2]\Big) \qquad (1)$$

To understand equation (1) consider a pair of distinct agents $i, j$ out of the $n$ agents. The probability that the realized network includes the realized edge $\{i,j\}$ equals $1-(1-y_n)^2$. If the edge is realized, then the benefit of $i$ and $j$ from its existence is $2b(1)$. Therefore $2b(1)(1-(1-y_n)^2)$ represents the expected welfare from the existence of the edge $\{i,j\}$. The term $2b(2)(1-y_n)^2$ represents an upper bound for the utility $b(d_G(i,j))$ of $i$ and $j$

conditional on the absence of the edge $\{i, j\}$. Since this is true for any pair $\{i, j\}$, the term $\binom{n}{2}\Big(2b(1)(1-(1-y_n)^2)+2b(2)(1-y_n)^2\Big)$ bounds the positive expected welfare from the realized network.

Similarly $c[2y_n(1-y_n)+2y_n^2]$ is the negative utility $\{i, j\}$ receives from the existence of the edge $\{i, j\}$. With probability $2y_n(1-y_n)$ only one of the agents pays for the edge, where with probability $y_n^2$ both of the agents pay for the edge. Thus we have shown that equation (1) bounds $W_n$ from above.

Applying F.O.C with respect to $y_n$ the maximum of the expression in (1) is obtained when $y_n = 1 - \frac{c}{2(b(1)-b(2))}$. We next claim that this utility is achievable by choosing $k_n = (1 - \frac{c}{2(b(1)-b(2))})n$. To see this note that the only term that we bound from above in (1) rather than calculating the exact expectation is the term $\binom{n}{2}2b(2)(1-y_n)^2$. We implicitly assume that if agents $\{i, j\}$ are not directly connected, then there exists another agent $k$ such that both $i$ and $j$ are connected to $k$. Since the degree $k_n = (1-\frac{c}{2(b(1)-b(2))})n$ this property indeed holds with probability one as we have shown in the proof of Theorem 2.

The optimal asymptotic welfare is thus obtained by letting $y_n = 1 - \frac{c}{2(b(1)-b(2))}$ and the optimal welfare is thus by (1) becomes

$$\binom{n}{2}\left(2b(1) - 2c + \frac{b(2)c^2 - b(1)c^2 - 3c^3}{2(b(1)-b(2))^2} + \frac{2c^2}{(b(1)-b(2))}\right)$$

This yields a constant fraction of $\binom{n}{2}\big(2b(1)-c\big)$ which is the optimal social welfare in the pure case.

We next consider the second assertion. Note that the expected number of agents $i$ with degree $n-1$ is $\log(n)$. In addition. The probability that at least one agent with degree $n$ approaches $1 - \frac{1}{n}$ when $n$ grows. Therefore for large values of $n$ the expected social welfare from the aforementioned strategy can be bound from below using the following expression:

$$2(n-1)b(1) + (n-2)(n-1)b(2) - log(n)nc$$

To see this consider the case where there are $m$ "stars." I.e., $m$ agents with degree $n-1$. The welfare in this case is bounded from below by

$$(m+1)(n-1)b(1) + (n-m)(n-m-1)b(2) - m(n-1)c.$$

The first expression bounds from below the welfare that is achieved by a direct links from the agents. The second expression represents the indirect utility of the entire $n-m$ agents from each other. The last expression represents the cost. Note that $m = 1$ minimizes the first two expressions and therefore the welfare can be bound by

$$2(n-1)b(1) + (n-1)(n-2)b(2) - m(n-1)c.$$

Since the expectation of $m$ is $log(n)$ we have reached the desired bound.

Note that the welfare in the case $m = 1$ which is the optimal welfare in the pure case is:

$$2(n-1)b(1) + (n-1)(n-2)b(2) - (n-1)c.$$

We clearly have that

$$\lim_{n\to\infty} \frac{2(n-1)b(1) + (n-1)(n-2)b(2) - (n-1)c}{2(n-1)b(1) + (n-2)(n-1)b(2) - log(n)nc} = 1.$$

This demonstrates that the proposed strategy extracts the full welfare asymptotically.

□

# I Proof of Proposition 3

*Proof of Proposition 3.* The fact that $\lim_{n\to\infty} \frac{k_{\text{in},n}}{n} = 1 - \frac{c}{b(1)-b(2)}$ follows exactly as in the proof of Theorem 2.

We first claim that $k_{\text{out},n}$ grows sublinearly, that is, $\limsup_n \frac{k_{\text{out},n}}{n} = 0$. To see this, assume by way of contradiction that $k_{\text{out},n}$ grows linearly along some equilibrium subsequence, then there exists a constant $\beta > 0$ such that agent $i$ forms a link with agent $j$ with probability at least $\beta$ for any two agents $i$ and $j$, and all sufficiently large $n$. This implies that the probability that $i$ lies within distance 2 of $j$ goes to 1 for any pair $i$, $j$, regardless of $i$'s strategy. To see this, note that any node $k \neq i, j$ is connected to both with probability at least $1 - (1-\beta)^2$ for all sufficiently large $n$. Since these events are independent across $k$, it follows that with probability approaching 1, there exists such an agent $k$. In such a case, it follows from $\overline{c} > b(1)$ that deviating and playing $k_{\text{out},n} = 0$ is profitable for player $i$ for all sufficiently large $n$. A contradiction.

Next, given that $k_{\text{out},n}$ grows sublinearly, we try to estimate the ~~marginal~~ benefit $\pi_n^k$ of a given agent $i$ from having $k = o(n)$ edges to any other homophily group for sufficiently large $n$. Let $\beta = 1 - \frac{c}{b(1)-b(2)}$.

Fix any agent $j$ in a different community from $i$. Observe that if there at least one agent $j'$ in the same community as $j$ such that $i$ links to $j'$ and $j'$ links to $j$, then the distance from $i$ to $j$ is at most two, and this occurs with probability at least $1-(1-\beta)^k$. Moreover, if $i$ links to some $j'$ that is within distance two of $j$, then the distance from $i$ to $j$ is at most three, and this occurs with probability close to 1. Since the probability that agents $i$ and $j$ are directly linked is close to 0, it follows that the expected utility $i$ gets from $j$ is approximately

$$(1-(1-\beta)^k)b(2)+(1-\beta)^k b(3).$$

Thus, the marginal expected benefit from all agent in $j$'s community if agent $i$ chooses $k+1$ instead of $k$ edges is approximately

$$\Delta(k)=n(1-\beta)^k(b(2)-b(3))\beta.$$

Therefore, for $k_n=o(\log(n))$ we have that the marginal benefit from adding an edge is arbitrarily large as $n$ goes to infinity. To see this, define $\lambda := -\ln(1-\beta)>0$. Since $(1-\beta)^k=\exp\big(k\ln(1-\beta)\big)=\exp(-\lambda k)$, we conclude

$$\Delta(k)=n\beta\big(b(2)-b(3)\big)\exp(-\lambda k).$$

Hence the marginal benefit of adding another edge diverges whenever $k_n=o(\log n)$. This implies that the support of the out group strategy contains only points $k=O(\log(n))$.

Assume the sequence $k=k_n$ satisfies $k_n=\omega(\log n)$; that is, $\dfrac{k_n}{\log n}\xrightarrow[n\to\infty]{}\infty$. Write $k_n=\alpha_n\log n$ with $\alpha_n\to\infty$. Then

$$\begin{aligned}\Delta(k_n)&=n\,\beta\big(b(2)-b(3)\big)\exp\big(-\lambda\alpha_n\log n\big)\\&=\beta\big(b(2)-b(3)\big)\,n^{\,1-\lambda\alpha_n}.\end{aligned}$$

Because $\alpha_n\to\infty$ we have $1-\lambda\alpha_n\to-\infty$, hence $n^{1-\lambda\alpha_n}\to 0$. Therefore

$$\Delta(k_n)\xrightarrow[n\to\infty]{}0\quad\text{whenever } k_n=\omega(\log n).$$

For this reason, we must have that the support of the out-group degree contains only points $k=\Theta(\log(n))$ as desired. $\square$

# References


V. Bala and S. Goyal. A noncooperative model of network formation. *Econometrica*, 68 (5):1181–1229, 2000.

A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *science*, 286 (5439):509–512, 1999.

B. Bollobás. *Random graphs*. Springer, 1998.

B. Bollobás. *Random Graphs*. Cambridge Studies in Advanced Mathematics. Cambridge University Press, 2 edition, 2001.

A. Cabrales, A. Calvó-Armengol, and Y. Zenou. Social interactions and spillovers. *Games and Economic Behavior*, 72(2):339–360, 2011.

C. Cooper and A. Frieze. A general model of web graphs. *Random Structures & Algorithms*, 22(3):311–335, 2003.

S. Currarini, M. O. Jackson, and P. Pin. An economic model of friendship: Homophily, minorities, and segregation. *Econometrica*, 77(4):1003–1045, 2009.

R. Durrett. *Random graph dynamics*, volume 20. Cambridge university press, 2010.

M. Girvan and M. E. Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

B. Golub and Y. Livne. Strategic random networks. *Available at SSRN 1694310*, 2010.

M. O. Jackson and B. W. Rogers. The economics of small worlds. *Journal of the European Economic Association*, 3(2-3):617–627, 2005.

M. O. Jackson and B. W. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, 2007.

M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996. ISSN 0022-0531. doi: https://doi.org/10.1006/jeth.1996.0108. URL `https://www.sciencedirect.com/science/article/pii/S0022053196901088`.

M. E. Newman and M. Girvan. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113, 2004.

D. d. S. Price. A general theory of bibliometric and other cumulative advantage processes. *Journal of the American society for Information science*, 27(5):292–306, 1976.

D. J. Watts. Networks, dynamics, and the small-world phenomenon. *American Journal of sociology*, 105(2):493–527, 1999.

D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world'networks. *nature*, 393(6684):440–442, 1998.

W. W. Zachary. An information flow model for conflict and fission in small groups. *Journal of anthropological research*, 33(4):452–473, 1977.