Public Goods in Endogenous Networks*

Markus Kinateder[§]
Luca Paolo Merlino[¶]

Abstract

In this paper, we study a local public good game in an endogenous network with heterogeneous agents. We consider two specifications in which different networks arise. When agents differ in the cost of acquiring the public good, active agents form hierarchical complete multipartite graphs; yet, better types need not have more neighbors. When agents' benefits from the public good are heterogeneous, nested split graphs emerge in which investment need not be monotonic in type. In large societies, few agents produce a lot and networks dampen inequality for most agents under cost heterogeneity and increase it under heterogeneity in benefits.

JEL classifications: C72, D00, D85, H41.

*We would like to thank Juan Pereyra, Estelle Cantillon, Karl Schlag, Andrea Galeotti, Patrick Legros, Anna Gunnthorsdottir, and (seminar) participants in PET 2013, SAEe 2013, CTN 2014, ESEM 2014, Information Transmission in Networks Conference (Harvard, 2015), SAET 2015, ASSET 2015, Vienna, U del País Vasco and U Pública de Navarra for their helpful comments. The usual disclaimers apply.

[§]Departamento de Economía, Edificio Amigos, Universidad de Navarra, 31009 Pamplona, Spain; email: mkinateder@unav.es. Financial support from the Spanish Ministry of Education and Science through grant ECO2014-55236-R is gratefully acknowledged.

[¶]Université Paris 1 Panthéon Sorbonne, Centre d'Economie de la Sorbonne, 106-112 Boulevard de l'Hôpital, 75647 Paris, France; email: lpmerlino@univ-paris1.fr. Merlino gratefully acknowledges financial support from the National Bank of Belgium, ERC through grant 208535 and the Wiener Wissenschafts-, Forschungs- und Technologiefonds (WWTF) under project fund MA 09-017. Merlino is a Postdoctoral Researcher of the Fonds National de la Recherche Scientifique-FNRS.

Consider consumers who decide how much information about alternative products to acquire or farmers who learn about new fertilizers. To choose between alternatives whose advantages they do not know, they acquire some information either personally or through their peers.

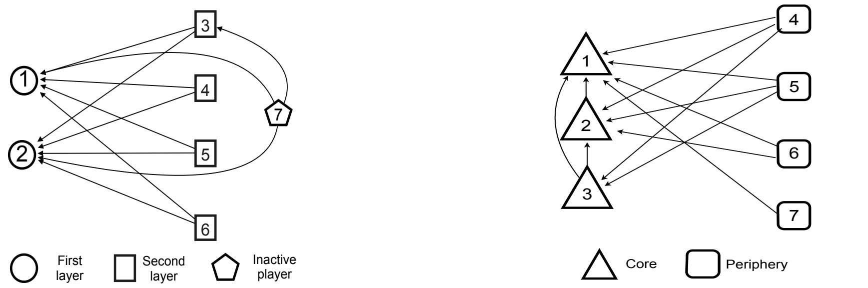
Since agents benefit from their neighbors' investment, the personal acquisition of information is a local public good. In these situations, as well as in many others, the network of interactions is, at least to a certain extent, endogenous. Since social structure often depends on the factors it affects, this poses a challenge to the estimation of the impact of social networks (Jackson, 2008, p. 437). Indeed, individual characteristics affect an agent's decision on public good provision and networking: influential consumers enjoy shopping more (Feick and Price, 1987) and farmers imitate more experienced neighbors (Conley and Udry, 2010). Yet, while games on fixed networks have been thoroughly studied (Bramoullé, Kranton and D'Amours, 2014), there is far less understanding of strategic interactions when networks are endogenous.

A relevant exception is Galeotti and Goyal (2010), henceforth G&G, in which homogeneous agents simultaneously choose public good provision and links, which are established unilaterally. They find that strict Nash equilibria are core-periphery networks in which few agents produce a significant amount of the public good, the so-called *law of the few*. However, some complementarity in neighbors' actions or decay in information flow invalidates these results. Yet, for a theory to be empirically relevant, it should have robust predictions. Furthermore, it is important to understand the role of heterogeneity in individual characteristics.

The aim of this paper is precisely to find robust predictions on who produces what amount and who links with whom when different sources of heterogeneity are introduced into the framework of G&G.

In particular, we study a framework in which agents differ in the (linear) cost or in the (concave) benefits of the public good. For example, some consumers enjoy shopping more and some farmers better assess the reaction of their crops to the fertilizer because they are more experienced. As a result, they differ in the marginal cost of collecting information. On the other hand, richer consumers and farmers with more land value the same piece of information more because they can exploit it better. Hence, they differ in the marginal utility of the public good. In both cases, better types are those who optimally invest more into the public good in isolation.

When the network is exogenous, the actual source of heterogeneity does not matter. However, it influences the decision of whether to establish a link by determining how much different types gain from a connection (Lemma 1). As a result, the network architectures that arise in equilibrium differ in *(i)* the relationship between investment, number of neighbors, and type, and *(ii)* whether a core of well-connected agents emerges. This is the "rich club phenomenon", first described in Zhou and Mondragon (2004), which we are the first to explain as a result of strategic interaction.



(a) Heterogeneity in Production Cost (b) Heterogeneity in Valuation

Agents are labeled from the best to the worst type starting from 1.

Figure 1: Nash Equilibria of Public Good Games in Endogenous Networks.

In the model with cost heterogeneity, players with a lower production cost find it more profitable to produce rather than to free ride, but worse types free ride on them. As a result, agents are more likely to be connected to players who are very different from them; in particular, the best types are not interconnected. Rather, social hierarchies with a pyramidal structure emerge (Figure 1(a)): active connected agents are ordered in independent sets, i.e. layers of similar types that are not connected, and form *complete multipartite graphs* (Theorem 1). Better types produce more and belong to smaller, higher layers. In such equilibria, centrality and type might not be monotonically related (Corollary 1) because of the links established by inactive agents (such as agent 7).

In the model with heterogeneity in benefits, better types gain more from a link: if one agent links, all those with higher valuations link as well.

Hence, the best players form a core of very active agents. Worse types are less willing to link and need less of the public good, and eventually are isolated. Equilibrium networks are then *nested split graphs* (Figure 1(b)) in which one's neighborhood is a subset of the neighborhoods of the better type agents (Theorem 2). Since active agents at the periphery with different links just produce enough to access their stand-alone optimal output, better types need not produce more (Corollary 2).

Efficient architectures are stars where the worst or the best types (but 1) are isolated depending on the source of heterogeneity (Proposition 1).

We extend the law of the few to a society with arbitrary degrees of heterogeneity in cost or in valuation of the public good in the local public good game proposed in the seminal paper of G&G (Proposition 2).[1] The networks we get are also negative assortative (Newman, 2002): few agents tend to have a large number of connections with poorly connected ones. These properties are relevant because they arise in many contexts.[2]

Networks dampen inequality for most agents under cost heterogeneity and increase it under heterogeneity in benefits (Proposition 3). Again, this stresses the importance of how the gains from links change with type.

The model we have described so far is very stylized. Yet our results are robust to many extensions relevant for empirical applications. In particular, we can introduce both types of heterogeneity at the same time, the indirect flow of spillovers, some decay, imperfect substitutability between one's effort and that of others, and heterogeneity in the linking cost (Proposition 4).

Often, those who initiate communication bear the associated cost, such as paying for a phone call or going to someone's farm. Therefore, in our model, links are established unilaterally. While this assumption is not always appropriate, the same networks arise when mutual consent is needed to create a link and agents can make transfers (Proposition 5).

We further discuss the relationship between G&G and our work at the end of Section 3. When agents' efforts are strategic complements instead,

[1] G&G consider a very limited form of heterogeneity, i.e. one agent has a lower production cost than the others, in which case this player is the hub of a star.

[2] Some examples are the networks observed in peer-to-peer exchanges (Adar and Huberman, 2000), intra-sector R&D (Tomasello et al., 2013), inter-bank linkages (Soramäki et al., 2007), and trade (De Benedictis and Tajoli, 2011).

either nested split graphs or complete multipartite graphs emerge when best replies are increasing and convex or concave, respectively (Hiller, 2012 and Baetz, 2015). We obtain these structures with strategic substitutes, i.e. decreasing best replies, depending on how the gains from a connection differ with types. Hence, we provide a unifying approach that is a first step towards a general theory of strategic interaction in endogenous networks.

Bramoullé and Kranton (2007) study public good provision in fixed networks. They show that there always exist specialized equilibria in which active agents are organized in an independent set and their direct neighbors are inactive. However, links are often at least to some extent endogenous. In these situations, studying the incentives to link matter. As first noted by G&G, fewer effort profiles are equilibria when the network is endogenous, because production and links are deeply related: players establish a link only if it gives them access to enough public good.

In particular, we model network formation non-cooperatively as in Bala and Goyal (2000) who show that center-sponsored stars are strict Nash equilibria of a model without effort choice, two-way flow of information, and homogeneous agents. Heterogeneity in benefits plays a minor role here: equilibrium networks are a collection of stars (Galeotti, Goyal and Kamphorst, 2006). It plays a major role instead when the gains from connections depend on agents' investment. When linking costs are also heterogeneous, Billand, Bravard and Sarangi (2011, 2012) give sufficient conditions for the existence of Nash networks, which do not necessarily exist (Haller, Kamphorst and Sarangi, 2007), an issue that does not emerge here.[3]

1 Model

We now introduce a local public good game, in which agents exert effort and establish costly connections to free ride on the effort exerted by others. All proofs are in the Appendix.

Players. There is a set of players $N = \{1, ..., n\}$; i denotes a typical player.

[3]Haller and Sarangi (2005) were the first to study heterogeneity in the model of Bala and Goyal (2000) with perfect indirect flow of information. In particular, they focus on heterogeneity in link failures. In our paper, links never fail.

Network. Player i's set of links is represented by a row vector $g_i = (g_{i1}, ..., g_{ii-1}, g_{ii+1}, ..., g_{in})$, where $g_{ij} \in \{0, 1\}$, for each $j \in N \setminus \{i\}$. Let $g_i \in G_i = \{0, 1\}^{n-1}$. We say that player i links to player j if $g_{ij} = 1$. The cost associated to a link is supported by those who initiate the communication, similarly to a phone call. Hence, linking decisions are one-sided: the agent proposing a link pays k and the link is established. Since in our game, direct spillovers are never negative, incoming links are always accepted.

The network g obtained from the players' linking strategies is a directed graph. We define $N_i^{OUT}(g) = \{j \in N : g_{ij} = 1\}$ as the set of players to which i links, and $\eta_i^{OUT}(g) = \left|N_i^{OUT}(g)\right|$ as the number of links that i sponsors.

The closure of g is an undirected network denoted by $\bar{g}$, where $\bar{g}_{ij} = \max\{g_{ij}, g_{ji}\}$, for each $i, j \in N$. That is, each directed link in g is replaced by an undirected one. Let $N_i(\bar{g}) = \{j \in N : \bar{g}_{ij} = 1\}$ be the set of players to which i is linked in the undirected graph $\bar{g}$, and let $\eta_i(\bar{g}) = |N_i(\bar{g})|$ be the number of i's neighbors in $\bar{g}$, or i's *degree.*

There is a path in $\bar{g}$ between i and j if either $\bar{g}_{ij} = 1$, or there are m different players $j_1, ..., j_m$ distinct from i and j, such that $\bar{g}_{ij_1} = \bar{g}_{j_1 j_2} = ... = \bar{g}_{j_m j} = 1$. The length of the path is 1 in the first case, and $m + 1$ in the second. A *component* of the network is a set of agents such that there is a path connecting every two agents in the set and no path to agents outside the set. A network $\bar{g}$ is *connected* if there is a unique component encompassing all agents, and *minimally connected* if it is connected and there exists only one path between every pair of players. We denote the set of isolated agents as $\mathcal{I}(\bar{g}) = \{i \mid \bar{g}_{ij} = 0 \text{ for all } j \in N\}$.

In a *core-periphery graph*, there are two groups of players, $\mathcal{P}(\bar{g})$, the *periphery*, and $\mathcal{C}(\bar{g})$, the *core*, such that for every $i, j \in \mathcal{P}(\bar{g})$, $\bar{g}_{ij} = 0$, while for every $l, m \in \mathcal{C}(\bar{g})$, $\bar{g}_{lm} = 1$; furthermore, for any $i \in \mathcal{P}(\bar{g})$, there exists $l \in \mathcal{C}(\bar{g})$ such that $\bar{g}_{il} = 1$. A *complete core-periphery* network is such that $N_i(\bar{g}) = \mathcal{C}(\bar{g})$ for all $i \in \mathcal{P}(\bar{g})$, and $N_l(\bar{g}) = N \setminus \{l\}$ for all $l \in \mathcal{C}(\bar{g})$. Nodes in $\mathcal{C}(\bar{g})$ are referred to as *hubs.* A core-periphery network with a single hub is referred to as a *star.* A core-periphery network in which the sets of agents' neighbors are nested is a *nested split graph*: for any pair of agents i and j, if $\eta_i(\bar{g}) > \eta_j(\bar{g})$, then $N_j(\bar{g}) \cup \{j\} \subset N_i(\bar{g}) \cup \{i\}$.

An *independent set* of $\bar{g}$ is a non-empty subset of players who are not linked. In a *complete multipartite graph*, agents can be partitioned into a number S of independent sets $\mathcal{H}_s(\bar{g}^*)$, $s = 1, ..., S$, such that every agent shares a link with all agents outside her own set.

A network is *negative assortative* if the average degree of one's neighbors is decreasing in one's own degree (Newman, 2002).

Effort. Player i's effort is denoted by $x_i \in X$, where $X = [0, +\infty)$. A player i is active if $x_i > 0$; otherwise i is inactive.

Strategies. Player i's set of strategies is $S_i = X \times G_i$, and the set of all players' strategies is $S = S_1 \times ... \times S_n$. A strategy profile $s = (x, g) \in S$ specifies investment $x = (x_1, ...x_n)$ and links $g = (g_1, ...g_n)$ for each player.

Payoffs. We consider a game of positive local externalities: direct neighbors' investments in the public good are perfect strategic substitutes. Hence, player i's payoffs under strategy profile (x, g) are:

$$U_i(x, g) = f_i\Big(x_i + \sum_{j \in N_i(\bar{g})} x_j\Big) - c_i x_i - \eta_i^{OUT}(g)k, \tag{1}$$

where $k > 0$ is the linking cost paid by the player who initiates a link and $f_i(x)$ is twice continuously differentiable in x and i. Furthermore, *(i)* $f_i(x)$ is a strictly concave and increasing function in x for all $i \in N$, and *(ii)* for all i, $f_i'(0) > c_i$, and $lim_{x \to \infty} f_i'(x) = m_i < c_i$.

Under these assumptions, there is a unique and non-negative optimal investment in the public good in isolation for every i denoted by

$$a_i = \arg\max_{x_i \in X} f_i(x_i) - c_i x_i.$$

We introduce *ex ante* heterogeneity in two ways.

(a) Differences in the cost of producing the public good: $f_i = f$ for all i, while $c_1 < c_2 < ... < c_n$, i.e. agents are heterogeneous in how efficient they are in producing the local public good. For example, some consumers enjoy shopping more and some farmers better assess the reaction of their crops to fertilizer because they are more experienced.

(b) Differences in the valuation of the public good: $c_i = c$ for all i and $\partial^2 f_i / \partial x \partial i < 0$, or $f_i'(x) > f_j'(x)$ for all $x > 0$, if $i < j$. For example, there could be richer consumers, farmers with more land and firms with bigger market shares. These agents would value information more, accessed directly or indirectly, because they can exploit it better.

Under both specifications, the players' types capture the amount of public good they would optimally collect in isolation such that $a_1 > a_2 > ... > a_n$. We refer to lower-indexed agents as *better types*. We assume that all inequalities are strict, i.e. there is one player per type. This assumption simplifies the analysis but does not substantially affect our results.

We define player i's *gain from a connection* to player z who produces $x_z \geq 0$ given a certain amount y of spillovers already received as

$$GC_i(x_z, y) = f_i(x' + x_z + y) - f_i(x_i + y) - c_i(x' - x_i),$$

where $x' = \arg\max_{x \geq 0} f_i(x + x_z + y) - c_i x$ is the effort that i exerts after accessing z's production of the public good. The following lemma describes how the gains from a connection change with type in both models.

Lemma 1 *Under heterogeneity in the cost of producing the public good, GC_i is increasing in i. Under heterogeneity in the valuation of the public good, GC_i is decreasing in i.*

In particular, under cost heterogeneity, players value the spillovers associated with an additional link identically, but more efficient players enjoy a lower reduction in production cost. Hence, they have lower gains from a connection. Conversely, under heterogeneity in the valuation of the public good, better types benefit more from spillovers, while producing the public good has the same cost for all players.

Equilibrium. A strategy profile $s^* = (x^*, g^*)$ is a *Nash equilibrium* if for all $s_i \in S_i$ and all $i \in N$, $U_i(s^*) \geq U_i(s_i, s^*_{-i})$, where $s = (s_i, s_{-i})$. For heterogeneous agents, small perturbations of the valuation or production costs are enough to break eventual ties, so that we focus on *strict* equilibria, in which the inequalities in the above definition are strict for all players.

Social Welfare. For any $s \in S$, social welfare is given by the sum of individual payoffs, $SW(s) = \sum_{i \in N} U_i(s)$. A strategy profile s^* is socially efficient if $SW(s^*) \geq SW(s)$, for every $s \in S$.

2 Main Analysis

In this section, we characterize the equilibria of this game, solve the social-planner problem, derive results for large societies, and study inequality.

The results are stated in terms of the closure of equilibrium networks $\bar{g}^*$, partly because, under heterogeneity in the valuation of the public good, there are equilibria in which both parties involved in a link could sponsor it, so that the corresponding closure of the directed graph is identical.

2.1 Equilibrium Analysis

The next lemma shows that, in equilibrium, active agents always collect exactly the level of public good they would in isolation. This is very helpful to characterize the equilibria of both models.

Lemma 2 *Given any Nash equilibrium,* $s^* = (x^*, g^*)$, $x_i^* + \sum_{j \in N_i(\bar{g})} x_j^* \geq a_i$, *for all* $i \in N$, *and if* $x_i^* > 0$, *then* $x_i^* + \sum_{j \in N_i(\bar{g})} x_j^* = a_i$.

The proof is omitted since it is a straightforward extension of a result in Bramoullé and Kranton (2007) to the case of heterogeneous players. Lemma 2 implies that the set of active players cannot be fully interconnected when agents differ in the amount of public good they optimally collect in isolation. Otherwise, they would have access to the same amount of public good, so that someone would have a profitable deviation. Hence, stars are the only complete core-periphery networks that can be equilibria.

Lemma 3 *In equilibrium, active players are not all connected.*

2.1.1 Heterogeneity in Production Cost

When better types are more efficient in producing the public good, they have lower gains from a connection. The payoff function is then given by:

$$U_i(x, g) = f\Big(x_i + \sum_{j \in N_i(\bar{g})} x_j\Big) - c_i x_i - \eta_i^{OUT}(g) k. \tag{2}$$

The following theorem characterizes the relationship between public good provision and type in this model.[4]

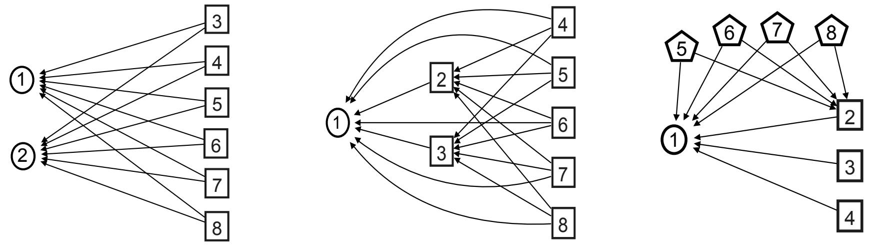
Theorem 1 *Under heterogeneity in the cost of producing the public good, if $k \leq f(a_1) - f(a_n) + c_n a_n$, in a strict Nash equilibrium, active agents form a complete multipartite graph in which better types produce more and are in independent sets that comprise fewer agents.*

Intuitively, when agents decide whether to establish a link, they compare the gain from free riding on someone's public good production with the cost of acquiring the same amount of public good. For the best and most efficient types, linking is relatively less attractive, so they might not be connected to each other. Yet, they produce a lot. As a consequence, they receive links from lower types that have higher gains from these connections given their higher production cost.

In equilibrium, there is a tight correspondence between type and investment since any player always produces more than a worse type. This is surprising in a framework where the value of connections is endogenous because, in principle, less efficient agents could produce more to attract many in-links. However, this is no equilibrium. Indeed, when $x_i^* < x_j^*$ for some $i < j$, both players need to have active in-links, j more so than i, and moreover, i needs more active out-links than j. The players linking to i and j in turn need to have active in-links themselves, some of them being distinct. Reiterating this argument recursively leads to a contradiction since the number of players is finite.

Since the profitability of a link depends on one's production cost, similar players have similar out-links and produce similar quantities. Hence, agents in the same independent set collect similar amounts of public good both directly and via their neighbors. However, better types need to have access to more public good, produce more, and receive many in-links from worse types. As a result, in equilibrium, active players are ordered in a hierarchy with a pyramidal structure in which the lower layers comprise more agents.

[4] When linking is sufficiently costly, the unique equilibrium is an empty network. When the linking cost decreases, player n, who has the highest benefit from linking to 1, eventually finds it profitable to link to 1. This yields a threshold in terms of the linking cost below which a non-empty network exists: a periphery-sponsored star with player 1 as the hub producing a_1. This guarantees equilibrium existence.



(a) (b) (c)

	example (a)					example (b)					example (c)				
i	c_i	a_i	x_i^*	U_i^*	$U_i^*(\emptyset)$	c_i	a_i	x_i^*	U_i^*	$U_i^*(\emptyset)$	c_i	a_i	x_i^*	U_i^*	$U_i^*(\emptyset)$
1	.6	2.777	.540	3.010	1.666	.695	2.070	.447	2.546	1.439	.6	2.777	.842	1.995	1.666
2	.601	2.769	.530	3.010	1.664	.774	1.670	.405	1.940	1.292	.8	1.562	.720	1.324	1.25
3	.83	1.452	.382	1.433	1.205	.775	1.665	.401	1.940	1.290	.83	1.452	.610	1.304	1.205
4	.831	1.448	.378	1.432	1.203	.831	1.448	.164	1.280	1.203	.831	1.448	.606	1.303	1.203
5	.832	1.417	.375	1.432	1.202	.832	1.445	.161	1.280	1.202	.840	1.448	0	1.3	1.190
6	.833	1.414	.371	1.432	1.200	.833	1.441	.157	1.280	1.200	.841	1.448	0	1.3	1.189
7	.834	1.411	.368	1.431	1.200	.834	1.438	.154	1.280	1.200	.842	1.448	0	1.3	1.188
8	.835	1.434	.364	1.430	1.198	.835	1.434	.150	1.280	1.198	.9	1.235	0	1.3	1.111
k	k∈[.33,.44]					.33					.6				

Figure 2: Examples of Nash equilibria under heterogeneity in the cost of producing the public good with $f(x,g) = 2\sqrt{x_i + \sum_{j \in N_i(\bar{g})} x_j}$.

Figure 2 exhibits three examples of possible equilibrium configurations that illustrate some general features. First, *equilibrium networks do not have a core* since the most efficient players need not link to each other, as in example 2(a). This property emerges in several real-world situations. For example, Feick and Price (1987) show that influential consumers (market mavens) enjoy shopping more than others, and do not rely on other market mavens' information. Similarly, Conley and Udry (2010) show that the most experienced farmers do not learn from each other, but rather inexperienced farmers learn from more experienced ones. Through the lens of our model, market mavens and experienced farmers do not free ride on others' information precisely because their cost of acquiring information is so low. Therefore, they acquire much information themselves, receive many in-links from bad types, and have a low gain from sponsoring a link.

Second, equilibrium networks display *vertical clustering*: someone's neighbors are likely to be neighbors as well when they are sufficiently different. This follows from the fact that similar types have similar gains from a link, and hence, are likely to be in the same independent set, as in examples 2(a) and 2(b). This is particularly surprising because, intuitively, one would think that players are more likely to cluster the more similar they are, a phenomenon which is known as *homophily*.

Third, *equilibrium networks are negative assortative*. Since equilibrium hierarchies are pyramidal, better players have many more links than worse types who link to them. Furthermore, similar types are not linked. Therefore, the average degree of one's neighbors decreases with one's own degree.

Fourth, if some active players are connected in equilibrium, then there is a *unique component encompassing all agents*. Very efficient players, except 1, can be isolated only if all other agents are inactive and link to 1. If a very efficient player j does not link to 1 and some periphery players are active, then all of them link to j as well. Otherwise, each of the periphery players linking to 1 would receive more public good than j.

So far, we have restricted our attention to active agents. Yet, the behavior of inactive agents reveals another important implication of the higher gains from a connection for worse types. It is stated in the next corollary.

Corollary 1 *Under heterogeneity in the cost of producing the public good, $\eta_i(\bar{g}^*)$ need not be monotonic in i.*

Expressed in words, the *number of neighbors or degree might not be monotonic in type*, as in example 2(c): although agents 5 to 8 are the most inefficient players, they have two links, i.e. one more than players 3 and 4. Indeed, worse types might have more links if they are inactive because they out-link more and need not link to all players in an independent set.

2.1.2 Heterogeneity in the Valuation of the Public Good

When better types value the public good more, the gains from a connection are higher for better types. In this case, the payoff function is given by

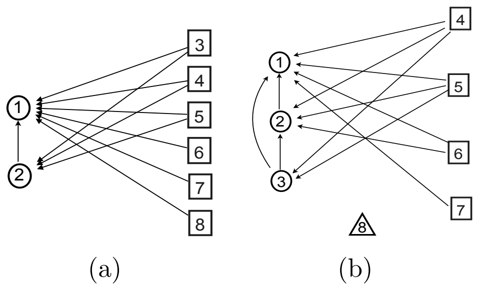
$$U_i(x, g) = f_i\Big(x_i + \sum_{j \in N_i(\bar{g})} x_j\Big) - cx_i - \eta_i^{OUT}(g)k. \tag{3}$$

We now show how this property affects the relationship between type and investment. First, we characterize the equilibrium network structure.[5]

[5]When linking is sufficiently costly, the unique equilibrium is an empty network. When the linking cost decreases, player 2, who has the highest benefit from linking to 1, eventually finds it profitable to link to 1. This yields a threshold in terms of the linking

Theorem 2 *Under heterogeneity in the valuation of the public good, if* $k \leq f_2(a_1) - f_2(a_2) + ca_2$, $\bar{g}^*$ *is a nested split graph in which better types have more links. Moreover, there exist* $\tilde{n}_1$ *and* $\tilde{n}_2$, $\tilde{n}_1 < \tilde{n}_2 \leq n$, *such that*
(i) $\mathcal{C}(\bar{g}^*) = \{i \in N : i \leq \tilde{n}_1, x_i^* > 0\}$ *is the core of active players;*
(ii) $\mathcal{P}(\bar{g}^*) = \{i \in N : \tilde{n}_1 < i \leq \tilde{n}_2\}$ *is the periphery;*
(iii) $\mathcal{I}(\bar{g}^*) = \{i \in N : i > \tilde{n}_2\}$ *is a set of isolated players.*

In equilibrium, the best players form a core because they not only need more public good, but also gain more from a connection. Then, there are players who do not produce enough to receive in-links, but who benefit from linking to some players in the core. These agents form a periphery of free riders that can be active, in which case both core players and free riders enjoy positive spillovers. Finally, there are isolated agents whose benefits from linking even to the top producers are too low to justify the linking cost.



				example (a)		example (b)	
i	b_i	a_i	$U_i^*(\emptyset)$	x_i^*	U_i^*	x_i^*	U_i^*
1	1.849	9.5	2.85	4.26	4.442	2.3	5.01
2	1.754	8.55	2.565	2.74	3.508	2.26	3.802
3	1.686	7.9	2.37	.9	2.87	2.16	2.783
4	1.627	7.35	2.205	.35	2.705	.6	2.28
5	1.621	7.3	2.19	.3	2.69	.55	2.265
6	1.370	5.21	1.563	.95	2.041	.65	1.631
7	1.082	3.25	0.975	0	1.433	.95	1.015
8	.657	1.2	.36	0	.557	1.2	.36
k		-		.8		.65	

Figure 3: Nash equilibria under heterogeneity in the valuation of the public good with $f_i(x,g) = b_i\sqrt{x_i + \sum_{j \in N_i(\bar{g})} x_j}$ and $c = .3$.

Figure 3 depicts two possible equilibrium configurations that illustrate some general properties. First, *equilibrium networks are nested split graphs with a core that contains the best types.* This structure emerges because *(i)* players who sponsor η links always link to the η players that invest more and from whom they receive no in-links, and *(ii)* better types have more incentives to link and produce more. In example 3(b), agents 1 to 3 constitute the core of connected agents who receive in-links, while agents 4 to 7 are at the periphery, i.e. they free ride by sponsoring links to core players, and the more so the better their type.

cost below which a non-empty network exists: a periphery-sponsored star with player 1 as the hub producing a_1. This guarantees equilibrium existence.

Overall, $N_7(\bar{g}^*) \cup \{7\} \subset N_6(\bar{g}^*) \cup \{6\} \subset ... \subset N_1(\bar{g}^*) \cup \{1\}$, i.e. the neighborhoods of worse types are subsets of the neighborhoods of better types. Hence, *degree centrality is always higher for better types.*

These two facts imply that equilibrium networks display *negative assortativity.* In example 3(b), the average degree of player 1's neighbors is 3, while player 7 is only linked to player 1, whose degree is 6.

By Lemma 2, active players produce enough to attain the optimal stand-alone effort given the spillovers from their neighbors. Hence, if a player has one link more than a worse type, the effort she exerts in order to achieve her optimal level might be lower. In Figure 3(a), players 3 to 5 produce less than 6 who does not link to 2. Hence, *investment in the public good need not be monotonic in type*, as summarized in the next corollary.

Corollary 2 *Under heterogeneity in the valuation of the public good,* x_i^* *need not be monotonic in* i.

Finally, *there can be isolated agents*, such as player 8 in Figure 3(b). In general, the worse an agent's type, the more likely she is to be isolated since she does not value the public good enough to pay linking cost k.

Our characterization of equilibrium networks shows that an agent's type and investment in general cannot be inferred from her position in the network. Hence, one must be careful in interpreting degree centrality as evidence of how good or important a player is.

The only architecture which can be a strict equilibrium in both models (and in a model with homogeneous agents) is a periphery-sponsored star with 1 as the hub. We show next that such structures are also efficient.

2.2 Efficiency

In the following proposition, the efficient allocations of production and links in the model where agents have different production costs and valuations of the public good are denoted by C and V, respectively.

Proposition 1 *The socially optimal network is a star such that:*
(a) under heterogeneity in production cost, there is $\underline{n} > 1$ *such that* $g_{i1} = 1$ *for all players* $i \geq \underline{n}$, *and the hub* 1 *produces* y^C *given by* $(n - \underline{n} + 2) f'(y^C) =$

c_1, while players $1 < i < \underline{n}$ are isolated;
(b) under heterogeneity in valuation, there is $\bar{n} \leq n$ such that all players $i \leq \bar{n}$ form a star with a hub that produces y^V given by $\sum_{i \leq \bar{n}} f_i'(y^V) = c$, while players $i > \bar{n}$ are isolated.

The social planer minimizes linking costs. Therefore, efficient networks are stars from which some players are excluded depending on the gains from the connection to the hub. Due to the different relationship between type and gains from a connection, the identity of isolated agents is very different in the two models: under cost heterogeneity, the most efficient agents but 1 are isolated, while under heterogeneity in valuation, the agents that value the public good the least are isolated.

Hence, decentralized non-empty equilibria always entail under-investment since no player internalizes the marginal value of her production for all other players. Moreover, if the equilibrium network is non-empty and is not a star, then it is over-connected.

2.3 Large Societies

The law of the few, formally derived first by G&G, predicts that as the number of players increases, the number of active players in the network increases at a lower rate. This result captures well many social and economic networks observed in reality. We now show that a similar result also holds when agents are heterogeneous.

Given an equilibrium (x^*, g^*), we define $\mathcal{A}(x^*, g^*, \varepsilon)$ as the number of *agents in the component* of g^* who produce at least ε and $\mathcal{A}_{IN}(x^*, g^*)$ as the number of (active) agents in g^* who receive at least one in-link.

Proposition 2 *Under heterogeneity in valuation or in production cost, given f_1 and c_1, for any $\varepsilon > 0$, $\lim_{n\to\infty} \mathcal{A}(x^*, g^*, \varepsilon)/n = 0$. Furthermore, $\lim_{n\to\infty} \mathcal{A}_{IN}(x^*, g^*)/n = 0$.*

The number of active agents need not be bounded for two reasons. First, some agents might be isolated producing their optimal amount of public good. Second, some peripheral agents might produce some public good to complement the amount received from their neighbors. However, periphery players as a whole cannot produce but a limited amount of public good, otherwise players to whom they link would receive too large an amount of public good. Furthermore, the amount of public good acquired via each link cannot fall below a certain threshold determined by the linking cost. Therefore, the number of agents that produce more than an infinitesimal amount of public good and the number of agents that receive links are bounded as the population size increases. Hence, in large societies, a small group of players produce a significant amount of public good, while most other players either only free ride or produce almost nothing.

2.4 Inequality

Networks can increase inequality, i.e. the difference in payoffs across agents in isolation versus in a network. This happens in particular when the best types access a large amount of public good from active free riders, as happens for some agents in the examples of Figures 2 and 3.

In the examples reported in Table 1 however, the best types do not benefit much from the network because free riders produce little or nothing. In these cases, the best type actually has the lowest equilibrium payoff, so that the network dampens inequality with respect to these agents.

	$f(x,g)$ as in Fig. 2 and $k=.1$					$f(x,g)$ as in Fig. 2 and $k=.6$					$f_i(x,g)$ as in Fig. 3 and $k=.65$				
i	c_i	a_i	x_i^*	U_i^*	$U_i^*(a,\emptyset)$	c_i	a_i	x_i^*	U_i^*	$U_i^*(a,\emptyset)$	b_i	a_i	x_i^*	U_i^*	$U_i^*(a,\emptyset)$
1	.747	1.792	1.103	1.853	1.339	.6	2.777	2.777	1.666	1.666	1.849	9.5	9.5	2.85	2.85
2	.774	1.669	.127	2.386	1.290	.8	1.562	0	2.777	1.25	1.754	8.55	0	2.565	4.757
3	.775	1.665	.123	2.385	1.205	.83	1.452	0	2.777	1.205	1.686	7.9	0	2.37	4.548
4	.831	1.448	.095	2.028	1.203	.831	1.448	0	2.777	1.203	1.627	7.35	0	4.364	2.205
5	.832	1.417	.091	2.028	1.202	.840	1.448	0	2.777	1.190	1.621	7.3	0	4.347	2.19
6	.833	1.414	.088	2.028	1.200	.841	1.448	0	2.777	1.189	1.370	5.21	0	3.571	1.563
7	.834	1.411	.084	2.028	1.200	.842	1.448	0	2.777	1.188	1.082	3.25	0	2.684	.975
8	.835	1.434	.081	2.028	1.198	.9	1.235	0	2.777	1.111	.657	1.2	0	1.376	.36

Table 1: The impact on inequality of star networks with 1 as the hub.

Overall, how networks affect inequality depends on how much free riders produce and who benefits more from a link. In large societies, the law of the few implies that there are few active agents that receive links. Hence, it is not possible to free ride on most free riders, and the only element which matters is how the gains from a connection vary with type.

Define $\underline{U_i} = f_i(a_i) - c_i a_i$ for all $i \in N$. We say that in a given network g, the inequality between any two players i and j such that $i < j$ decreases if $U_i(x^*, g^*) - U_j(x^*, g^*) \leq \underline{U_i} - \underline{U_j}$.

Proposition 3 *Given any g^* and any $i < j$ which receive no in-links, as $n \to \infty$, under heterogeneity in production cost, the inequality between i and j decreases and, under heterogeneity in valuation, it increases.*

In large populations, by the law of the few, the proportion of players that receive in-links is very small, so that this result applies to most players. Intuitively, the possibility of establishing links benefits those players that gain more from a connection. Under cost heterogeneity, these are the worst players, and thus, inequality decreases. Under heterogeneity in benefits, the best types gain most from each link, and thus, inequality increases.

3 Discussion

Robustness Analysis. The benchmark model described in Section 2 is very stylized. However, we now derive precise bounds on the robustness of our characterization to having both cost and valuation heterogeneity at the same time, indirect flow of public good, decay, imperfect substitutability, and (some) heterogeneity in the linking cost. To do so, we introduce the following payoffs

$$\begin{aligned} U_i(x,g,\varepsilon) &= (1+\varepsilon_{1,i}) f_i \left(\left(x_i^{1-\varepsilon_6} + (1-\varepsilon_4) \sum_{d=1}^{\infty} \varepsilon_5^{d-1} \sum_{j \in N_i^d(\bar{g})} x_j^{1-\varepsilon_6} \right)^{\frac{1}{1-\varepsilon_6}} \right) \\ & \quad -(c_i - \varepsilon_{2,i}) x_i - \eta_i^{OUT}(g)(k+\varepsilon_{3,i}) \end{aligned} \tag{4}$$

where $N_i^d(\bar{g}) = \{j \in N : d_{i,j}(\bar{g}) = d\}$ is defined as the set of neighbors that are connected to player i via a shortest path of length d. The shocks are:
– $\varepsilon_1 \in \mathbb{R}^N$ and $\varepsilon_2 \in \mathbb{R}^N$ introduce **both types of heterogeneity** at the same time; indeed, agents are often heterogeneous along several dimensions;
– $\varepsilon_3 \in \mathbb{R}^N$ introduces **heterogeneity in linking costs**; for example, some individuals prefer to talk more or have cheaper phone rates;[6]
– $\varepsilon_4 \in [0,1]$ introduces **decay**: some information is lost when transmitted to neighbors, either because communication is imperfect or some knowledge is tacit. Hence, if i is linked to j, the spillover she gets is only $(1-\varepsilon_4)x_j$;

[6] We do not study differences in the linking cost *per se* since they do not affect players' type—the optimal public good production in isolation—which is the focus of our paper.

– $\varepsilon_5 \in [0,1]$ introduces **indirect spillovers** since often the public good is also shared among indirect neighbors; in that case, information is discounted by ε_5 for each link it travels in the network.[7] For example, consumers who get information from market mavens might share it with others;
– $\varepsilon_6 \in [0,\infty)$ captures **imperfect substitutability** between individuals' efforts, for example because the information collected displays some content heterogeneity (as in Zhang and van der Schaar, 2012).

We denote an equilibrium of the game by $(x^*(\varepsilon), g^*(\varepsilon))$ and the optimal amount of the public good an agent would collect in isolation by $a_i(\varepsilon)$ to stress the dependence on the shocks ε. The following proposition describes how to determine precise bounds on ε for g^* to remain an equilibrium.

Proposition 4 *Under heterogeneity in valuation or in production cost, for each strict equilibrium network g^*, there exist shocks ε such that g^* is an equilibrium network of the perturbed game if $|\varepsilon| < \bar{\varepsilon}$.*

The key ingredients in obtaining this result are the following. First, each agent's payoffs are continuous in the shocks. Second, in strict equilibria, all inequalities representing agents' optimal linking decisions are strict. Hence, there is room to perturb payoffs. It is then enough to check that the effort level of agents can be adjusted in a consistent way, which in general is possible given that agents are heterogeneous. Furthermore, we can find shocks such that inactive agents remain inactive. In that case, the law of the few holds also for the perturbed game.

Two-sided Link Formation and Transfers. Some situations that can be captured by our model include bilateral R&D collaborations among firms or local constituencies that provide services and share them with nearby jurisdictions. In these cases, however, mutual consent is needed to share the public good. Furthermore, agents might ask for compensation to communicate the information they acquire. In what follows, we study the impact of this different network formation protocol on equilibrium properties.[8]

[7] For $\varepsilon_5 = 0$, our benchmark models arise. Since ε_5 converges to zero from above, it is natural to define $\lim_{\varepsilon_5 \to 0^+} {\varepsilon_5}^0 = 1$. If instead $\varepsilon_5 = 1$, equilibrium networks are minimally connected with possibly some isolated agents.

[8] Without transfers, other equilibria arise because players refuse some links once they have acquired the optimal amount of public good. Some examples are available upon request.

We denote the transfers proposed by player i by $\tau_i = \{\tau_{ij}\}_{j \in N}$, where $\tau_{ij} \in \mathbb{R}$ for all $j \in N$. We assume that $\bar{g}_{ij} = 1$ if, and only if, $\tau_{ij} + \tau_{ji} \geq k$. A strategy profile $s = (x, \tau)$ specifies investments x and transfers $\tau = \{\tau_1, ..., \tau_n\}$. The payoff function is then defined as

$$U_i(x, \tau) = f_i\Big(x_i + \sum_{j \in N_i(\bar{g})} x_j\Big) - c_i x_i - \sum_{j \in N} \bar{g}_{ij} \tau_{ij}. \tag{5}$$

In equilibrium, a link is formed if it is profitable and, if a link is not there, at least one of the two agents involved does not benefit from it. Formally, Bloch and Jackson (2007) and G&G define pairwise equilibrium as follows:

Definition 1 *A strategy s^* is a pairwise equilibrium if (1.) s^* is a Nash equilibrium, and (2.) for all $\tau^*_{ij} + \tau^*_{ji} < k$, if $U_i(x'_i, x'_j, \tau'_{ij}, \tau'_{ji}, x^*_{-ij}) > U_i(s)$, then $U_j(x'_i, x'_j, \tau'_{ij}, \tau'_{ji}, x^*_{-ij}) < U_j(s^*)$, for all $x'_i, x'_j \in X$ and for all τ'_{ij}, τ'_{ji}.*

Proposition 5 shows that for each strict equilibrium in the benchmark models, there is an equilibrium under two-sided link formation, such that the resulting network is identical.

Proposition 5 *Under heterogeneity in valuation or in production cost, take a strict equilibrium (x^*, τ^*) in the model with one-sided linking. For all i and j, let τ_{ij} be such that if $g^*_{ij} = 1$, then $\tau^*_{ij} = k + \varepsilon$, $\varepsilon > 0$, while if $g^*_{ij} = 0$, then $\tau^*_{ij} = -\varepsilon$. Then, (x^*, τ^*) is a strict equilibrium in the two-sided model with transfers that induces $\bar{g}^*$.*

The proof of this proposition is trivial and hence omitted. Intuitively, if under one-sided linking an agent is willing to sponsor a link, the other agent accepts this link if the proposing player bears most of its cost.

Homogeneous Agents (G&G). When agents are homogeneous, strict equilibria are complete core-periphery structures in which the law of the few holds. Moreover, there are at most two levels of production. Compared with these results, we get richer structures and some striking differences.

First, all active agents cannot be connected and share all their neighbors since each agent's optimal amount of public good is different (Lemmata 2 and 3). Hence, either a core does not emerge (Theorem 1) or active periphery agents do not all connect to the same core agents (Theorem 2).

Second, for strict equilibria, there is a discontinuity in the limit of the heterogeneous to the homogeneous population case since some links need to be established or deleted to get complete core-periphery structures.[9] There are two exceptions: *(i)* stars are equilibria in all the models mentioned; and *(ii)* some complete multipartite graphs, such as the one in Figure 2(a), are both strict equilibria with heterogeneous production cost and non-strict equilibria with homogeneous agents.[10] Therefore, it might be inappropriate to focus on strict equilibria when agents are homogeneous.

Finally, when agents are homogeneous, even infinitesimal complementarity in neighbors' actions or decay in information flow might destroy the equilibrium characterization. When agents are heterogeneous instead, equilibrium networks are robust to decay, as well as to many other extensions.

4 Conclusion

In this paper, we study a local public good game with an endogenous choice of neighbors among heterogeneous agents. Depending on the dimensions along which agents are heterogeneous (which in isolation is not relevant), we find that *(i)* active agents form either complete multipartite or nested split graphs, and *(ii)* the network reduces or increases inequality for most agents. In both models, the law of the few holds in large societies.

The source of heterogeneity matters because it determines how the gains from a connection differ across types. In equilibrium, this affects the relationship between outcomes or type and network statistics. In this sense, our results are relevant beyond the theoretical literature of networks.

Surprisingly, the network structures we single out also arise under strategic complements when players' best replies are either convex (Hiller, 2012) or concave (Baetz, 2015). Hence, future research should investigate whether more general results are obtainable.

[9] For example, consider the network in Figure 3(a). This is no longer an equilibrium if $\bar{a}_1 \to \bar{a}_2$ because $x_2^* \to 0$. Hence, as agents get more homogeneous, eventually only the star with 1 as the hub is an equilibrium. Yet, at the limit, i.e., when all agents are homogeneous, other complete core-periphery structures are equilibria.

[10] For example, consider the network in Figure 2(a). Let the economy converge to the homogeneous agents' case in the following way: first, $a_1 = a_2 = \bar{a}$ and $a_i = \underline{a}$ for $i = 3, ..., 8$, i.e. there are two types. Then, Figure 2(a) is an equilibrium as $\bar{a} \to \underline{a} \to .527$ if $\underline{a} \in [(11k/12 + \sqrt{(11k/12)^2 + \bar{a}/6}, 11k\sqrt{\bar{a}}/6 + \bar{a}/6]$. Eventually, it is a non-strict equilibrium when $\underline{a} = \bar{a} = .527$. The other equilibria are periphery sponsored stars with 1 or 2 as the hub; if $\underline{a} \neq \bar{a}$ both of them can be active under some conditions.

Appendix

Proof of Lemma 1. Suppose $f_i = f$ for all $i \in N$. Then, $GC_i(x_z, y) = f(x'+x_z+y) - f(x_i+y) - c_i(x'-x_i)$ becomes $c_i(x_i - x')$ if $x_i - x' > 0$, i.e., i is active when linking to z, or $f(x_z+y) - f(x_i+y) + c_i x_i$ otherwise. In both cases, GC_i is increasing in c_i and, hence, in i. Suppose instead $c_i = c$ for all $i \in N$. Then, $GC_i(x_z, y) = f_i(x' + x_z + y) - f_i(x_i + y) - c(x' - x_i)$ becomes $c(x_i - x')$ if i is active when linking to z, or $f_i(x_z + y) - f_i(x_i + y) + cx_i$ otherwise. Since $\partial^2 f_i/\partial x \partial i < 0$, GC_i is decreasing in i. ■

Proof of Lemma 3. Suppose that $s^* = (x^*, g^*)$ is a Nash equilibrium and that the active players are all connected among them. In other words, for all active $i, j \in N$, $\bar{g}_{ij} = 1$. Take two players $i, j \in N$, then $a_i \neq a_j$. If they are active, $a_i = x_i + \sum_{z \in N, z \neq i} x_z$ and $a_j = x_j + \sum_{z \in N, z \neq j} x_z$. But i and j have the same neighbors and $\bar{g}_{ij} = 1$, which implies $a_i = x_i + \sum_{z \in N, z \neq i} x_z = x_j + \sum_{z \in N, z \neq j} x_z = a_j$. At the same time, $a_i \neq a_j$, a contradiction. ■

Proof of Theorem 1. First we show that for any active i and j such that $i < j$, $x_i^* > x_j^*$. Suppose *ad absurdum* this is not the case. Without loss of generality, consider first the best type j and the worst type $i < j$ such that $x_i^* < x_j^*$. We show in a series of Lemmata that a contradiction emerges.

Lemma 4 *Suppose there exist players i, j and z such that $g_{iz}^* = 1$ but $\bar{g}_{jz}^* = 0$ and $x_z^* > x_j^* \geq 0$. Then, it holds that $x_z^* - x_j^* > a_i - a_j$.*

Proof of Lemma 4. The result is trivial if $a_i < a_j$ since $x_z^* > x_j^* \geq 0$. If $a_i > a_j$, define $\Delta = x_z^* - x_j^*$. In order to prove that $(a_j + \Delta) > a_i$, we suppose *ad absurdum* that $(a_j + \Delta) \leq a_i$. Then, the following inequalities arise: since $c_i < c_j$, it holds that $k - c_i x_j^* > k - c_j x_j^*$. Since j is not linked to z, it holds that $k - c_j x_j^* > f(a_j + \Delta) - f(a_j)$. Finally, since $(a_j + \Delta) \leq a_i$, by the concavity of f, $f(a_j + \Delta) - f(a_j) \geq f(a_i) - f(a_i - \Delta)$. Together this yields $k - c_i x_j^* > f(a_i) - f(a_i - \Delta)$. Thus, player i is strictly better off to break the link with z and to invest x_j instead, a contradiction. ■

Lemma 5 *Suppose that there exist i and j such that $i < j$ and $x_i^* < x_j^*$. Then, (i) the set of players $Z = \{z : x_z^* > 0, g_{iz}^* = 1, \bar{g}_{jz}^* = 0\}$ is non-empty; (ii) for any $z \in Z$, $x_z^* > x_j^*$ and $x_z^* - x_j^* > a_i - a_j$; (iii) the set* *of players* $P = \{p : x_p^* > 0, g_{pj}^* = 1, \bar{g}_{pi}^* = 0\}$ *is non-empty; (iv)* $\bar{g}_{ij}^* = 0$; *(v) for any* $p \in P$, $a_i > a_p$; *(vi) there exists a non-empty set of players* $L = \{l : x_l^* > 0, \bar{g}_{lj}^* = 1, g_{li}^* = 1, \bar{g}_{lp}^* = 0$ *for any* $p \in P\}$.

Proof of Lemma 5. If there are i and j such that $i < j$ and $x_i^* < x_j^*$, it must be the case that $\sum_{h \in N_i(\bar{g}^*)} x_h^* > \sum_{h \in N_j(\bar{g}^*)} x_h^*$ (since $a_i > a_j$). However, i does not receive more in-links than j (since $x_i^* < x_j^*$). Indeed, if there is a player l such that $g_{li}^* = 1$, then $\bar{g}_{lj}^* = 1$. (Obviously this holds if $g_{jl}^* = 1$. If $g_{li}^* = 1$, but $\bar{g}_{lj}^* = 0$, then l would profitably sever the link with i and link to j. This implies that $\bar{g}_{lj}^* = 1$ whenever $g_{li}^* = 1$.) Hence, given g^*, it holds that $\{l : g_{li}^* = 1\} \subseteq \{l : g_{lj}^* = 1\}$, and therefore that $\{l : g_{jl}^* = 1\} \subset \{l : g_{il}^* = 1\}$. Thus, there exists a non-empty set of players $Z = \{z : x_z^* > 0, g_{iz}^* = 1, \bar{g}_{jz}^* = 0\}$. This concludes the proof of part *(i)*.

To show part *(ii)*, pick any $z \in Z$. Then, $g_{iz}^* = 1$ implies that $c_i x_z^* > k$, and since $c_j > c_i$, it holds that $c_j x_z^* > k$, that is, it is cheaper for j to link to z rather than to produce x_z^* by herself. However, j does not link to z implying that by linking to z, j would stop producing and $f(a_j) - c_j x_j^* = f(x_j^* + \sum_{h \in N_j(\bar{g}^*)} x_h^*) - c_j x_j^* > f(x_z^* + \sum_{h \in N_j(\bar{g}^*)} x_h^*) - k$. Then by Lemma 4, $x_z^* > x_j^*$ and $x_z^* - x_j^* > a_i - a_j$. This concludes the proof of part *(ii)*.

To show part *(iii)*, suppose instead that $P = \emptyset$. Pick any $z' \in Z$. Then, the inequality $a_j - x_j^* + x_{z'}^* > a_i$, shown in part *(ii)*, is violated since

$$\sum_{l:g_{lj}^*=g_{li}^*=1} x_l^* + \sum_{t:g_{jt}^*=g_{it}^*=1} x_t^* + x_{z'}^* \leq \sum_{l:g_{lj}^*=g_{li}^*=1} x_l^* + \sum_{t:g_{jt}^*=g_{it}^*=1} x_t^* + \sum_{z:g_{iz}^*=1,\bar{g}_{jz}^*=0} x_z^* + x_i^*.$$

Then, j needs to have more active in-links than i in order for $a_j - x_j^* + x_{z'}^* > a_i$ to hold. Hence, P is non-empty. This concludes the proof of part *(iii)*.

To prove part *(iv)*, we need to show that $g_{ij}^* = 0$ and $g_{ji}^* = 0$. Suppose first that $g_{ij}^* = 1$. Pick any $z \in Z$. From part *(ii)* it follows that $f(a_j - x_j^* + x_z^*) - k > f(a_i) - k$. Suppose that i links to j paying k. Then, $f(a_i) - k > f(a_i) - c_i x_j^*$. Since $a_i > a_j$ and $c_i < c_j$, it holds that $f(a_i) - c_i x_j^* > f(a_j) - c_j x_j^*$. Finally, since j is not linked to z it holds that $f(a_j) - c_j x_j^* > f(a_j - x_j^* + x_z^*) - k$, a contradiction. Thus, player i does not link to j.

Furthermore, $g_{ji}^* = 0$ since $x_i^* < x_j^* < x_z^*$ and j does not link to z: if $g_{ji}^* = 1$, then j has a profitable deviation to sever the link with i and link to z instead, a contradiction. This concludes the proof of part *(iv)*.

To show part *(v)*, pick any $p \in P$ and suppose *ad absurdum* that $a_p > a_i$. Then, $x_p^* > x_i^*$ since we assumed that j is the best type and $i < j$ the worst type such that $x_i^* < x_j^*$. Pick some $z \in Z$. Then, either $x_p^* > x_z^*$ or $x_p^* < x_z^*$. In both cases a contradiction arises. In the first case, since $g_{iz}^* = 1$, also p and i are linked which contradicts that $\bar{g}_{pi}^* = 0$, and in the second, the same argument as in part *(iv)* applies (that is, Lemma 4 holds analogously for p and j), and thus, $\bar{g}_{jp}^* = 0$, which contradicts that $g_{pj}^* = 1$. This shows that $a_i > a_p$ and concludes the proof of part *(v)* of Lemma 5.

To show part *(vi)*, suppose *ad absurdum* that $L = \emptyset$. Consider now $p \in P$. Since $g_{pj}^* = 1$ and $x_p^* > 0$, p is also linked to any $z \in Z$ and to all other players to which i links. Thus, p receives $x_p^* + x_j^*$ which i does not, while i receives x_i^* which p does not. Since $x_j^* > x_i^*$, $x_p^* + x_j^* > x_i^*$. Hence, p receives strictly more public good than i. This contradicts $a_i > a_p$, as shown in part *(v)*, and $x_p^* > 0$ implies, by Lemma 2, that p accesses exactly a_p. Hence, there is a player l such that $x_l^* > 0$, $g_{li}^* = 1$ and $\bar{g}_{lp}^* = 0$. If $\bar{g}_{lp}^* = 1$, then, by the same argument, a contradiction would arise. This implies that $x_i^* > 0$ and $\bar{g}_{lj}^* = 1$ and concludes the proof of part *(vi)* of Lemma 5. ■

Lemma 6 *If there are i and j such that $i < j$ and $x_i^* < x_j^*$, given the sets of players P and L as defined above, then (i) $x_i^* > x_p^*$, for any $p \in P$; (ii) there exists a non-empty set of players $Q = \{q : x_q^* > 0, g_{qp}^* = 1, \bar{g}_{ql}^* = 0$ for any $p \in P$ and for any $l \in L\}$; (iii) for any $p \in P$ and for any $l \in L$, $x_p^* > x_l^*$; (iv) for any $l \in L$, $g_{lj}^* = 1$ and $g_{jl}^* = 0$; (v) there exists a non-empty set of players $R = \{r : x_r^* > 0, g_{rl}^* = 1, \bar{g}_{rp}^* = 1$ for any $p \in P$ and for any $l \in L\}$.*

Proof of Lemma 6. To show part *(i)*: since $g_{pj}^* = 1$ but $\bar{g}_{ij}^* = 0$, there is no player h such that $g_{ih}^* = 1$ but $g_{ph}^* = 0$, if not i (p) would profitably sever the link with h (j) and link to j (h) if $x_j^* > x_h^*$ ($x_j^* < x_h^*$). Suppose now *ad absurdum* that $x_p^* > x_i^*$. Then, for any e such that $g_{ei}^* = 1$, $g_{ep}^* = 1$. Therefore, p has at least as many in- and out-links as i. Furthermore, since $x_j^* > x_i^*$ and $g_{pj}^* = 1$, while $\bar{g}_{ij}^* = 0$, by Lemma 4 it holds that $a_i - x_i^* + x_j^* > a_p$. However, rewriting and simplifying this inequality we get a contradiction because $x_j^* < \sum_{q:g_{hp}^*=1 \wedge \bar{g}_{hi}^*=0} x_h^* + \sum_{h:g_{ph}^*=1 \wedge \bar{g}_{ih}^*=0} x_h^* + x_p^* + x_j^*$. Hence, $x_i^* > x_p^*$ is necessary for i to attract more active in-links than p. This concludes the proof of part *(i)* of Lemma 6.

To show part *(ii)*, note that for any $l \in L$ and $p \in P$, $x_i^* > x_p^*$ and $g_{li}^* = 1$ while $\bar{g}_{pi}^* = 0$. Hence, from Lemma 4, $a_p - x_p^* + x_i^* > a_l$. Suppose now *ad* *absurdum* that players of type p and l receive the same amount of public good via in-links. There is no player h such that $g^*_{ph} = 1$ but $g^*_{lh} = 0$, if not l (p) would profitably sever the link with i (h) and link to h (i) if $x^*_h > x^*_i$ ($x^*_h < x^*_i$). Hence, l receives more public good than p via out-links, at least from i. Finally, l produces x^*_l. Hence, $a_p - x^*_p + x^*_i < a_l$, a contradiction. This concludes the proof of part *(ii)* of Lemma 6.

Hence, $x^*_p > x^*_l > 0$ for all $l \in L$ and $p \in P$ if not any player $q \in Q$ would profitably deviate and link to l. This in turn implies that $g^*_{lj} = 1$ and $g^*_{jl} = 0$, if not j would have a profitable deviation to sever the link with l and establish one with i or some $z \in Z$ (since by part *(i)*, $x^*_i > x^*_p$, and as just shown $x^*_p > x^*_l$). This concludes the proof of Lemma 6.*(iii)* and *(iv)*.

To show *(v)*, suppose *ad absurdum* that $R = \emptyset$. Pick $p' \in P$. Then, since $x^*_{p'} > x^*_l$, $g^*_{qp'} = 1$, and $\bar{g}^*_{lp'} = 0$, *(i)* by Lemma 4, it holds that $a_l - x^*_l + x^*_{p'} > a_q$, and, *(ii)* q has more active out-links than l (to $p \in P$). Then,

$$\sum_{h:g^*_{lh}=g^*_{qh}=1} x^*_h + x^*_l - x^*_l + x^*_{p'} < \sum_{m:g^*_{mq}=1,\bar{g}^*_{ml}=0} x^*_m + \sum_{h:g^*_{lh}=g^*_{qh}=1} x^*_h + \sum_{p:g^*_{qp}=1,\bar{g}^*_{lp}=0} x^*_p + x^*_q,$$

or $x^*_{p'} < \sum_{m:g^*_{mq}=1\wedge\bar{g}^*_{ml}=0} x^*_m + \sum_{p:g^*_{qp}=1\wedge\bar{g}^*_{lp}=0} x^*_p + x^*_q$. This contradicts $a_l - x^*_l + x^*_{p'} > a_q$.This concludes the proof of part *(v)* of Lemma 6. ■

Lemma 7 *There exist $l \in L$ and $p \in P$ such that $\bar{g}^*_{lp} = 0$.*

Proof of Lemma 7. Suppose that $a_l > a_p$. Since $x^*_p > x^*_l$, the argument of part *(iv)* of Lemma 5 applies, implying that $\bar{g}^*_{lp} = 0$. Suppose instead that $a_p > a_l$. Since $x^*_i > x^*_l$ and $g^*_{pi} = 0$, $g^*_{pl} = 0$. Suppose *ad absurdum* that $g^*_{lp} = 1$ for all $l \in L$ and for all $p \in P$.

Then, i and p have the same active in-links. Compare the amount of public good that i and p receive, respectively: player i receives $x^*_i + \sum_{h:g^*_{hi}=1} x^*_h + \sum_{m:g^*_{im}=1} x^*_m$ and player p receives $x^*_p + x^*_j + \sum_{h:g^*_{hp}=1} x^*_h + \sum_{m\neq j:g^*_{pm}=1} x^*_m$. There is no player h such that $g^*_{ih} = 1$ but $g^*_{ph} = 0$, if not i (p) would profitably sever the link with h (j) and link to j (h) if $x^*_j > x^*_h$ ($x^*_j < x^*_h$). However, since $g^*_{pj} = 1$, $\bar{g}^*_{ij} = 0$ and $x^*_j > x^*_i$, Lemma 4 implies that $a_i - x^*_i + x^*_j > a_p$. Using $g^*_{lp} = 1$, this yields $0 < x^*_p + \sum_{m\neq j:g^*_{pm}=1\wedge\bar{g}^*_{im}=0} x^*_m$, where the sum are p's out-links to players other than j to which i is not linked. This is a contradictionand it concludes the proof of Lemma 7. ■

The results of Lemma 6 for players $p \in P$ and $l \in L$ apply analogously to players of type $q \in Q$ and $r \in R$ after relabeling p as q and l as r.

A recursive argument arises since, when someone produces more than a more efficient player, there are some active players that link to both of them and some that only link to the player producing more. In turn, these last players produce more and need to receive more active in-links. And so on and so forth. However, the set of players is finite, and eventually there are agents who have no further active in-links. A contradiction then arises, showing that in a strict equilibrium better active types produce more.

Suppose now that active agents do not form a complete multipartite graph. Then, there exist i and j such that $\bar{g}^*_{ij} = 0$ and, there is z such that $g^*_{zi} = 1$, $\bar{g}^*_{jz} = 0$ and $x^*_z > 0$. Clearly, $x^*_i > x^*_j$, if not z would rather link to j. This implies that $i < j$. For agents z and j, there is no player h such that $g^*_{jh} = 1$ but $\bar{g}^*_{zh} = 0$, if not z (j) would profitably sever the link with i (h) and link to h (i), if $x^*_h > x^*_i$ ($x^*_h < x^*_i$); i.e., z has no less out-links than j. Since $x^*_i > x^*_j$ and $g^*_{zi} = 1$ but $\bar{g}^*_{zj} = 0$, by Lemma 4, $a_j - x^*_j + x^*_i > a_z$, or $\sum_{l:g^*_{lj}=1} x^*_l + \sum_{h:g^*_{jh}=1} x^*_h + x^*_i > \sum_{l:g^*_{lz}=1} x^*_l + \sum_{h\neq i:g^*_{zh}=1} x^*_h + x^*_i + x^*_z$. This holds only if $\sum_{h:g^*_{lj}=1} x^*_l > \sum_{h:g^*_{lz}=1} x^*_l$. Hence, there exists $l^{(0)}$ such that $x^*_{l^{(0)}} > 0$, $g^*_{l^{(0)}j} = 1$ and $\bar{g}^*_{l^{(0)}z} = 0$, thus implying $x^*_j > x^*_z$ and $j < z$.

Now consider $l^{(0)}$ and z. There is no h such that $g^*_{zh} = 1$ but $\bar{g}^*_{l^{(0)}h} = 0$, if not z ($l^{(0)}$) would sever the link with h (j) and link to j (h) if $x^*_j > x^*_h$ ($x^*_j < x^*_h$). Since $x^*_j > x^*_z$ and $g^*_{l^{(0)}j} = 1$ but $\bar{g}^*_{l^{(0)}z} = 0$, by Lemma 4, $a_z - x^*_z + x^*_j > a_{l^{(0)}}$, or $\sum_{l:g^*_{lz}=1} x^*_l + \sum_{h:g^*_{zh}=1} x^*_h + x^*_j > \sum_{l:g^*_{ll^{(0)}}=1} x^*_l + \sum_{h\neq j:g^*_{l^{(0)}h}=1} x^*_h + x^*_j + x^*_{l^{(0)}}$. This holds only if $\sum_{l:g^*_{lz}=1} x^*_l > \sum_{l:g^*_{ll^{(0)}}=1} x^*_l$. Hence, there exists $l^{(1)}$ such that $g^*_{l^{(1)}z} = 1$ and $\bar{g}^*_{l^{(1)}l^{(0)}} = 0$. This implies $x^*_z > x^*_{l^{(0)}}$ and $z < l^{(0)}$.

Now consider $l^{(0)}$ and $l^{(1)}$. The same argument holds, and can be iterated for any couple of players $l^{(i)}$ and $l^{(i+1)}$, until we get at most to $l^{(n)}$, who have no more in-links than $l^{(n-1)}$ because there are no players left that can link only to $l^{(n)}$ but not to $l^{(n-1)}$. At that point, we reach a contradiction. Hence, active agents form a complete multipartite graph.

Finally, consider i and j belonging to the same independent set, i.e., $\bar{g}^*_{ji} = 0$, and a player $l > \{i, j\}$ such that $x^*_s > 0$ for $s = i, j, l$. Then $g^*_{li} = 1$. Suppose that $\eta_l(\bar{g}) > \eta_i(\bar{g}) = \eta_j(\bar{g})$. Without loss of generality, consider $i < j$. Then, $g^*_{lj} = 1$, and by Lemma 4, $a_j - x^*_j + x^*_i > a_l$. This is possible if and only if j receives more in-links than l, implying $x^*_j > x^*_l$. If $x^*_j > x^*_l$, then $j < l$. Then, the same holds for all players to which l links but i does not. If there are more players of type j than of type l, clearly l receives more public good than j, which by Lemma 2, leads to a contradiction with $x_l^*, x_j^* > 0$ and $j < l$. This concludes the proof of Theorem 1. ■

Proof of Theorem 2. First, we show parts *(i)* and *(ii)*. Note that if there are players j and i such that $g_{ji}^* = 1$, then there is no player z such that $x_z^* > x_i^*$ and $\bar{g}_{zi}^* = 0$. Suppose not. Then, since $g_{ji}^* = 1$, $k < cx_i^*$, player z could profitably reduce effort by x_i^* linking to i instead, a contradiction.
Therefore, any i receiving active in-links is connected to all players that produce more than x_i^*. This set of agents forms the core, $\mathcal{C}(\bar{g}^*)$. Since at least player 1 is in the core, $\mathcal{C}(\bar{g}^*) \neq \emptyset$, i.e. $\tilde{n}_1 \geq 1$. By Lemma 3, $\tilde{n}_1 < n$.
Next we show that if there is more than one player in $\mathcal{C}(\bar{g}^*)$, then there is a player in $\mathcal{P}(\bar{g}^*)$ exerting a positive amount of effort. Suppose not. Then, all players in $\mathcal{C}(\bar{g}^*)$ receive an identical amount of public good, a contradiction. Moreover, for any $i < j$ in the core, $x_i^* > x_j^*$ and $\eta_i(\bar{g}^*) > \eta_j(\bar{g}^*)$. Suppose not and that $x_i^* \leq x_j^*$. Then, i gets no more in-links than j from the periphery and $x_i^* + \sum_{z \in N_i(\bar{g}^*)} x_z \leq x_j + \sum_{z \in N_j(\bar{g}^*)} x_z$, a contradiction. Hence, $x_i^* > x_j^*$ and this implies that $\eta_i(\bar{g}^*) > \eta_j(\bar{g}^*)$.
Suppose that there is j, $1 < j < \tilde{n}_1$, who receives no in-links, i.e., $j \notin \mathcal{C}(\bar{g}^*)$. Since $a_j > a_{\tilde{n}_1}$ and player $\tilde{n}_1$ receives more public good than j via links, $x_j^* > x_{\tilde{n}_1}^*$. Then, the periphery player who links to $\tilde{n}_1$ can profitably deviate by linking to j instead, a contradiction. Hence, all players $1, ..., \tilde{n}_1$ belong to $\mathcal{C}(\bar{g}^*)$. This concludes the proof of part *(i)*.
Given that $\eta_i(\bar{g}^*) > \eta_j(\bar{g}^*)$ for any $i < j$ in the core, it follows immediately that, for any $l, m \in \mathcal{P}(\bar{g}^*)$ such that $l < m$, $\eta_l^{OUT}(g^*) \geq \eta_m^{OUT}(g^*)$.
Note next that $\tilde{n}_1 < \tilde{n}_2 \leq n$. Suppose that not all players $\tilde{n}_1 + 1, ..., \tilde{n}_2$ belong to $\mathcal{P}(\bar{g}^*)$. Then, there is j, $\tilde{n}_1 < j < \tilde{n}_2$, who is in $\mathcal{C}(\bar{g}^*)$ or isolated. If $j \in \mathcal{C}(\bar{g}^*)$, then j is active and gets more public good than $\tilde{n}_1 + 1$, a contradiction. If instead j is isolated, $a_j < x_{\tilde{n}_1}^*$. If not, since player $\tilde{n}_1$ receives an in-link, it would be profitable for j to link to $\tilde{n}_1$ and to produce $a_j - x_{\tilde{n}_1}^*$. Hence, suppose that $a_j < x_{\tilde{n}_1}^*$. Then, j does not link to player 1, if $k > f_j(x_1^*) - f_j(a_j) + ca_j$. By the envelope theorem, the derivative of this inequality's right-hand side with respect to type is ${}^{\partial f_j(x_1^*)}/_{\partial j} - {}^{\partial f_j(a_j)}/_{\partial j}$, which is negative since we assume that ${}^{\partial^2 f}/_{\partial i \partial x} < 0$ for all $x > 0$. Hence, if it is not profitable for j to link to 1, it is neither profitable for all players $i > j$, a contradiction. This concludes the proof of part *(ii)*.

The existence of a core in which players receive more in-links the better their type implies that the component of the network is a nested split graph. To show part *(iii)*, consider player n. If $\tilde{n}_2 = n$, then $n \in \mathcal{P}(\bar{g}^*)$, $x_n^* \geq 0$ and $g_{n1}^* = 1$. Hence, $\mathcal{I}(\bar{g}^*) = \emptyset$. If $\tilde{n}_2 < n$, then n sponsors no link and $f_n(a_n) - ca_n$ yields n a larger payoff than any other strategy (x_n, g_n). In this case, n is isolated and $\mathcal{I}(\bar{g}^*) \neq \emptyset$, if n receives no in-link. Suppose that n receives some in-link. Then, n belongs to the core and receives more public good than $\tilde{n}_1 + 1$, the player in $\mathcal{P}(\bar{g}^*)$ who wants more public good, a contradiction. This concludes the proof of Theorem 2. ■

Proof of Proposition 1. In any component only one agent produces to minimize linking costs. Under heterogeneity in production cost, it is efficient that only the most efficient agent, 1, produces while all others link to 1. Hence, a star with 1 as a hub is the efficient network. Under heterogeneity in valuation, suppose there are several components. Since agents are heterogeneous, different components produce different amounts of public good. Thus, players in less productive components would profitably link to the highest producing player. Hence, the efficient solution is a star with only one active agent.

To show part (a), note that $\bar{g}_{12}$ is the first to be severed as k increases since $f(a_j) - c_j a_j$ is smaller for higher j but linking to 1 yields any player $f(a_1)$. Hence, defining y such that $(n-1)f'(y) = c_1$, the social planner maximizes

$$\max_{x,m} mf(x) - c_1 x - (m-1)k + \sum_{j=2}^{n-m+1} \left[f(a_j) - c_j a_j \right]. \tag{A-1}$$

Given m, the objective function of the planner problem (A-1) is linearly decreasing in k with a slope equal to $-(m-1)$ and an intercept at $k = 0$ that is lower as more agents are isolated; (A-1) is constant in k when all agents are isolated. The objective function (A-1) is the upper envelope of all these linear functions, i.e., it is piece-wise decreasing in k. Therefore, the optimal m decreases as k increases, and for any k, there exists a threshold $\underline{n} > 1$ such that all players $i \geq \underline{n}$ connect to 1 and the others are isolated. The star's hub produces $y^C(\underline{n})$ such that $(n - \underline{n} + 2)f'(y^C(\underline{n})) = c_1$.

To show part (b), note that any player in the component can be the hub, denoted by h, since $c_i = c$ for all $i \in N$. When g_{jh} is severed, the linking cost k is saved, while the remaining social welfare changes by

$$f_j(a_j) - ca_j + \sum_{i \in N \setminus \{j\}} f_i(y) - cy - \big[\sum_{i \in N} f_i(x) - cx\big], \tag{A-2}$$

where x solves $\sum_{i \in N} f_i'(x) = c$ and y solves $\sum_{i \in N \setminus \{j\}} f_i'(y) = c$. By the envelope theorem, the derivative of (A-2) with respect to j is $\partial f_j(a_j)/\partial j - \partial f_j(x)/\partial j$, which is positive since $x > a_j$, $\partial f/\partial j < 0$ and $\partial^2 f/(\partial j \partial x) < 0$. Hence, worst types are isolated. The number of isolated agents is given by

$$\max_{x,m} \sum_{i=1}^{m} f_i(x) - cx - (m-1)k + \sum_{i=m+1}^{n} [f_i(a_i) - ca_i]\,. \tag{A-3}$$

Given m, the objective function of the planner problem (A-3) is linearly decreasing in k with a slope equal to $-(m-1)$ and an intercept at $k = 0$, which is lower as more agents are isolated; the function is constant in k when all agents are isolated. Since the objective function (A-3) is the upper envelope of all these linear functions, it is piece-wise decreasing in k. Therefore, the number of agents in the star decreases as k increases, and for any k, there exists a threshold $\bar{n} \leq n$ such that all players $i \leq \bar{n}$ are in the star and the others are isolated. The star's hub produces $y^V(\bar{n})$ such that $\sum_{i=1}^{\bar{n}} f_i'(y^V(\bar{n})) = c$. This concludes Proposition 1's proof. ■

Proof of Proposition 2. If g^* is empty, the statement follows trivially. If the network is non-empty, for all players j such that $\eta_j^{OUT}(g^*) > 0$, $g_{j1}^* = 1$ holds. In both models, by Lemma 2, player 1 produces at most $x_1^* = a_1 - \sum_{j:\bar{g}_{1j}^*=1} x_j^*$. For players j with $x_j^* > 0$ to link to 1, $x_1^* c_j \geq k$ must hold (with $c_j = c$ when $f_j \neq f$). Hence, $(a_1 - \sum_{j:\bar{g}_{1j}^*=1} x_j^*)c_j \geq k$. Suppose now that $\lim_{n\to\infty} |\{j : x_j^* > 0 \text{ and } g_{j1}^* = 1\}|/n > 0$. If $\lim_{n\to\infty} \sum_{j:\bar{g}_{1j}^*=1} x_j^* = \infty$, then since $a_1 < \infty$, by Lemma 2, player 1 is not active, a contradiction.
Re-label agents such that $n < n'$ if and only if $x_n > x_{n'}$. Given $0 < \bar{x} < \infty$, $\lim_{n\to\infty} \sum_{j:\bar{g}_{1j}^*=1} x_j^* = \bar{x}$ holds only if the series $\{x_n\}_n$ decreases in n. Even more, it must decrease faster than the series $\{1/n\}_n$ which does not converge to a finite value. This implies that for any n, the smallest element in the series is smaller than $1/(n-1)$. For any $\varepsilon > 0$, take $\bar{n}(\varepsilon)$ such that $\varepsilon \leq 1/(\bar{n}-1)$. Hence, there are at most $\bar{n}(\varepsilon)$ players who link to 1 and produce more than ε, so that, for any $\varepsilon > 0$, $\lim_{n\to\infty} \mathcal{A}(x^*, g^*, \varepsilon)/n = 0$.
The same arguments apply to all players receiving in-links. Suppose now that $\lim_{n\to\infty} \mathcal{A}_{IN}(x^*, g^*)/n > 0$. If players in this set link to 1, the same arguments as above apply. If instead they do not link to 1, there are $z \in N$

such that $c_z x_j^* > k > c_1 x_j^*$ and $g_{zj}^* = 1$ for all such j who are in the independent set of 1. This is possible only under cost heterogeneity. Clearly, $x_z = 0$ since $\lim_{n\to\infty} \mathcal{A}_{IN}(x^*, g^*)/n > 0$ implies $\lim_{\mathcal{A}_{IN}(x^*,g^*)\to\infty} \sum_{j:g_{zj}^*=1} x_j^* = \infty$, and $f(\sum_{j:g_{zj}^*=1} x_j^*) - f(\sum_{j:g_{zj}^*=1} x_j^* - \min_{j:g_{zj}^*=1} x_j^*) > k$. Yet, $f'' < 0$ implies $\lim_{\sum_{j:g_{zj}^*=1} x_j^* \to\infty} \left[f(\sum_{j:g_{zj}^*=1} x_j^*) - f(\sum_{j:g_{zj}^*=1} x_j^* - \min_{j:g_{zj}^*=1} x_j^*)\right] = 0$, a contradiction. Then, $\lim_{n\to\infty} \mathcal{A}_{IN}(x^*, g^*)/n = 0$. ■

Proof of Proposition 3. Let $n \to \infty$ and consider $i < j$ which both receive no in-links. Then $|U_i(x^*, g^*) - U_j(x^*, g^*)| \le \underline{U_i} - \underline{U_j}$. Note first that $\underline{U_i} > \underline{U_j}$ because, since there is no $z \in N$ such that $g_{zi}^* = 1$ or $g_{zj}^* = 1$, i can replicate j's strategy and get higher payoffs. Consider now the two models. Under heterogeneity in the production cost, i and j can be active or inactive. If they are both active, then inequality between i and j decreases if $c_i(a_i - x_i) - k\eta_i^{OUT}(g^*) \le c_j(a_j - x_j) - k\eta_j^{OUT}(g^*)$, which means $c_i \sum_{z:\bar{g}_{zi}^*=1} x_z \le c_j \sum_{z:\bar{g}_{zj}^*=1} x_z$. If i and j have the same neighbors, then the statement follows. If j has more out-links, by Theorem 1, j is in a lower independent set. But then there is z such that $g_{zi}^* = 1$, a contradiction. If i and j are inactive, they have the same neighbors. Hence, $U_i(x^*, g^*) - U_j(x^*, g^*) = 0$ while $\underline{U_i} - \underline{U_j} > 0$. Finally, if i is active while j is not, there are two cases. *(1)* If $\eta_i(g^*) = \eta_j(g^*)$, $U_i(x^*, g^*) - U_j(x^*, g^*) \le \underline{U_i} - \underline{U_j}$ can be rewritten as $f(a_i - x_i^*) - c_i(a_i - x_i^*) \ge f(a_j) - c_j a_j$. This clearly holds when $a_i - x_i^* = a_j$. If not, the left-hand-side is increasing in $a_i - x_i^*$ since $a_i - x_i^* < a_i$ given our assumptions on f. Hence, the statement follows. *(2)* If $\eta_j^{OUT}(g^*) > \eta_i^{OUT}(g^*)$, $U_i(x^*, g^*) - U_j(x^*, g^*) \le \underline{U_i} - \underline{U_j}$ can be rewritten as $f(\sum_{z:g_{jz}=1} x_z) - k[\eta_j^{OUT}(g^*) - \eta_i^{OUT}(g^*)] \ge f(a_j) - c_j a_j$. Since j has some more link, say to z, $f(\sum_{z:g_{jz}=1} x_z) - k\eta_j^{OUT}(g^*) > f(a_i - x_i) - k\eta_i^{OUT}(g^*)$, so that $f(\sum_{z:g_{ji}=1} x_z) - k[\eta_j^{OUT}(g^*) - \eta_i^{OUT}(g^*)] \ge f(a_i - x_i^*) - c_i(a_i - x_i^*)$. This concludes the proof of the first part of Proposition 3.

Under heterogeneity in benefits, i and j can be active or inactive. Suppose i and j are active and $\eta_i^{OUT}(g^*) = \eta_j^{OUT}(g^*)$. Then, $U_i(x^*, g^*) - U_j(x^*, g^*) \ge \underline{U_i} - \underline{U_j}$ holds because $a_i - x_i^* = a_j - x_j^*$. Suppose i and j are active and that i has an out-link more than j, to some player z. Then, $U_i(x^*, g^*) - U_j(x^*, g^*) \ge \underline{U_i} - \underline{U_j}$ implies $c(a_i - x_i) - k\eta_i^{OUT}(g^*) \ge c(a_j - x_j) - k\eta_j^{OUT}(g^*)$. Since all neighbors but z are common, this implies $cx_z \ge k$, which needs to hold because i links to z. The same argument holds when i has more than one neighbor more than j. Hence, $U_i(x^*, g^*) - U_j(x^*, g^*) \geq \underline{U_i} - \underline{U_j}$. When i and j are both inactive, suppose that $\eta_i^{OUT}(g^*) = \eta_j^{OUT}(g^*)$. Then, $|U_i(x^*, g^*) - U_j(x^*, g^*)| \geq |\underline{U_i} - \underline{U_j}|$ implies $f_i(\sum_{z:g^*_{iz}=1} x_z) - \eta_i^{OUT}(g^*)k - f_i(a_i) + ca_i \geq f_j(\sum_{z:g^*_{jz}=1} x_z) - \eta_j^{OUT}(g^*)k - f_j(a_j) + ca_j$. Then the condition follows if $f_i(y) - f_i(a_i) + ca_i$ is decreasing in i for $y > a_i$, i.e., higher for better types. By the envelope theorem, this depends on $\partial f_i(y)/\partial i - \partial f_i(a_i)/\partial i$, which is negative since $\partial f_i/\partial i < 0$ and $\partial^2 f_i/(\partial x \partial i) < 0$. Hence inequality increases. The same argument holds when i has more links than j since then, by optimality, $f_i(\sum_{z:g^*_{iz}=1} x_z) - \eta_i^{OUT}(g^*)k > f_i(\sum_{z:g^*_{jz}=1} x_z) - \eta_j^{OUT}(g^*)k$. This concludes the proof of Proposition 3. ∎

Proof of Proposition 4. Consider a strict equilibrium (x^*, g^*) under cost heterogeneity. Consider $\varepsilon_1 = (\varepsilon_{1,1}, ..., \varepsilon_{1,n}) \in \mathbb{R}^N$, while $\varepsilon_s = 0$ for all $s = 2, .., 6$. Clearly, $(x^*, g^*) = (x^*(\varepsilon_1 = 0), g^*(\varepsilon_1 = 0))$. For any agent i such that $x_i^* = 0$, $a_i < \sum_{j \in N} \bar{g}^*_{ij} x_j^*$ by Lemma 2 while for any active agent i, $a_i = x_i^* + \sum_{j \in N} \bar{g}^*_{ij} x^*_j$. Defining the adjacency matrix of links among active agents as $\bar{g}_A$, the vectors of their efforts and optimal efforts as $x^*_A(\varepsilon_1)$ and $a_A(\varepsilon_1)$, respectively, and the A-dimensional identity matrix as I_A,

$$a_A(\varepsilon_1) = x^*_A(\varepsilon_1)(I_A + \bar{g}^*_A). \tag{A-4}$$

This system has an interior solution for $\varepsilon_1 = 0$. By Cramer's rule, each $x_i^*(\varepsilon_1)$ is given by the ratio between the determinants of $(I_A + \bar{g}^*_A)$ with column i replaced by vector $a_A(\varepsilon_1)$ divided by the determinant of $(I_A + \bar{g}^*_A)$. Since, by Leibniz formula, this determinant is continuous in $a_i(\varepsilon_1)$, for small ε_1 the solution $x^*_A(\varepsilon_1)$ exists and is arbitrarily close to $x^*_A(\varepsilon_1 = 0)$ as $\varepsilon_1 \to 0$. Focusing on inactive agents, this implies that there is $\varepsilon_1 \in \mathbb{R}^N$ such that

$$a_i < \sum_{j \in N} \bar{g}^*_{ij} x^*_j(\varepsilon_1). \tag{A-5}$$

Then, for all $i \in N$, $U_i(x^*_i(\varepsilon_1), g^*_i)$ is continuous in ε_1 and $U_i(x^*, g^*) = U_i(x^*(\varepsilon_1 = 0), g^*(\varepsilon_1 = 0))$. Finally, in strict equilibria, for any $i \in N$, there exists $\bar{\varepsilon}_1 \in \mathbb{R}^N_+$ such that (A-5) and (A-4) are satisfied for all $|\varepsilon_1| < \bar{\varepsilon}_1$ and $U_i(x^*(\varepsilon_1), g^*) > U_i(x'_i, g'_i, x^*_{-i}(\varepsilon_1), g^*_{-i})$ for any any $(x'_i, g'_i) \in S_i \setminus \{(x^*_i(\varepsilon_1), g^*_i\}$. Hence, the same network structure is an equilibrium.

Analogousy, it follows immediately that for all ε_s, $s = 2, ..., 6$, there is $\bar{\varepsilon}_s \in \mathbb{R}^N_+$ such that for any $|\varepsilon_s| < \bar{\varepsilon}_s$, g^* is an equilibrium. ∎

References

Adar, Eytan, and Bernardo Huberman. 2000. "Free Riding on Gnutella." *First Monday*, 5(10).

Baetz, Oliver. 2015. "Social Activity and Network Formation." *Theoretical Economics*, 10: 315–340.

Bala, Venkatesh, and Sanjeev Goyal. 2000. "A Noncooperative Model of Network Formation." *Econometrica*, 68(5): 1181–1229.

Billand, Pascal, Christophe Bravard, and Sudipta Sarangi. 2011. "Strict Nash networks and partner heterogeneity." *International Journal of Game Theory*, 40(3): 515–525.

Billand, Pascal, Christophe Bravard, and Sudipta Sarangi. 2012. "Existence of Nash networks and partner heterogeneity." *Mathematical Social Sciences*, 64(2): 152–158.

Bloch, Francis, and Matthew O. Jackson. 2007. "The Formation of Networks with Transfers among Players." *Journal of Economic Theory*, 133(1): 83–110.

Bramoullé, Yann, and Rachel Kranton. 2007. "Public Goods in Networks." *Journal of Economic Theory*, 135(1): 478–494.

Bramoullé, Yann, Rachel Kranton, and Martin D'Amours. 2014. "Strategic Interaction and Networks." *The American Economic Review*, 104(3): 898–930.

Conley, Timothy G., and Christopher R. Udry. 2010. "Learning about a New Technology: Pineapple in Ghana." *The American Economic Review*, 100(1): 35–69.

De Benedictis, Luca, and Lucia Tajoli. 2011. "The World Trade Network." *The World Economy*, 34(8): 1417–1454.

Feick, Lawrence, and Linda Price. 1987. "The Market Maven: a Diffuser of Marketplace Information." *The Journal of Marketing*, 51: 83–97.

Galeotti, Andrea, and Sanjeev Goyal. 2010. "The Law of the Few." *The American Economic Review*, 100(4): 1468–1492.

Galeotti, Andrea, Sanjeev Goyal, and Jurjen Kamphorst. 2006. "Network Formation with Heterogeneous Players." *Games and Economic Behavior*, 54(2): 353–372.

Haller, Hans, and Sudipta Sarangi. 2005. "Nash networks with heterogeneous links." *Mathematical Social Sciences*, 50(2): 181 – 201.

Haller, Hans, Jurjen Kamphorst, and Sudipta Sarangi. 2007. "(Non-)existence and Scope of Nash Networks." *Economic Theory*, 31(3): 597–604.

Hiller, Timo. 2012. "Peer Effects in Endogenous Networks." mimeo.

Jackson, Matthew O. 2008. *Social and Economic Networks.* Princeton University Press.

Newman, M. E. J. 2002. "Assortative Mixing in Networks." *Physical Review Letters*, 89: 208701.

Soramäki, Kimmo, Morten L. Bech, Jeffrey Arnold, Robert J. Glass, and Walter E. Beyeler. 2007. "The Topology of Interbank Payment Flows." *Physica A*, 379(1): 317–333.

Tomasello, Mario V., Mauro Napoletano, Antonios Garas, and Frank Schweitzer. 2013. "The Rise and Fall of R&D Networks." mimeo.

Zhang, Yu, and Mihaela van der Schaar. 2012. "Information Production and Link Formation in Social Computing Systems." *Selected Areas in Communications, IEEE Journal on*, 30(11): 2136–2145.

Zhou, Shi, and Raul J. Mondragon. 2004. "The rich-club phenomenon in the Internet topology." *Communications Letters, IEEE*, 8(3): 180–182.