Public Goods in Endogenous Networks*

Markus Kinateder§
Luca Paolo Merlino¶

August 4, 2016

Abstract

We study a local public good game in an endogenous network with heterogeneous players. The source of heterogeneity affects the gains from a connection and hence equilibrium networks. When players differ in the cost of producing the public good, active players form pyramidal complete multipartite graphs; yet, better types need not have more neighbors. When players differ in the valuation of the public good, nested split graphs emerge in which production need not be monotonic in type. In large societies, few players produce a lot; furthermore, networks dampen inequality under cost heterogeneity and increase it under heterogeneity in valuation.

JEL classifications: C72, D00, D85, H41.

*We would like to thank Juan Pereyra, Estelle Cantillon, Karl Schlag, Andrea Galeotti, Patrick Legros, Marco Giani, Gani Aldashev, 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, BiNoMa 2016, Vienna, Southampton, Universidad del País Vasco and Universidad 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 grants ECO2014-55236-R and ECO2015-64330-P 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.

Economic agents often have to decide how much to rely on their own effort and how much on the support of their social contacts. For example, to choose between alternatives whose advantages are not known, they can acquire information either personally or through their peers.

Hence, effort and peers are often jointly determined. To date, however, most of the literature has focused either on games played in fixed networks (Bramoullé, Kranton and D'Amours, 2014) or on network formation games (Jackson and Wolinsky, 1996; Bala and Goyal, 2000).

Galeotti and Goyal (2010), henceforth G&G, were the first to combine the two approaches in a local public good game, i.e., when a player's effort has positive externalities on her neighbors while crowding out their effort. This assumption captures many applications, such as consumers who decide how much information about alternative products to acquire or farmers who learn to use new fertilizers. The main result of G&G is the *law of the few*: in large societies, a few players produce most of the public good.

While G&G consider homogeneous players, empirical studies find that individual characteristics are relevant. For example, market mavens are those consumers who like shopping the most (Feick and Price, 1987). Similarly, influential farmers are the most experienced ones (Conley and Udry, 2010). Heterogeneity determines who links with whom and who produces what. This is a challenge for the estimation of the impact of social networks on behavior (Jackson, 2008, p. 437). Theoretical guidance is then needed to understand to what extent correlation in behavior is due to the unobservable common characteristics that drive the link formation process.

Our paper provides a first step in this direction by studying how individual characteristics affect a player's decision on public good provision and networking in G&G's framework. Our analysis yields important insights for empirical applications by pointing out that the relationship between investment, number of neighbors, and type depends on the source of heterogeneity. This framework also allows us to answer some natural questions such as: is free riding pervasive in an economy with heterogeneous players? does the network dampen or increase inequality?

In the model, players simultaneously choose public good production and links. Links are established unilaterally, but once two players are linked, they access each other's public good production. We relax G&G's

assumption of homogeneous players and allow them to differ either in the (linear) cost of producing or in the (concave) valuation of consuming the public good. We define better types as those who optimally acquire more public good in isolation and label them with lower indexes starting from 1.

When the network is exogenous, only a player's type and her position in the network determine the individual production of the public good. When the network is endogenous, however, the actual source of heterogeneity affects the decision to establish a link by determining how much each player gains from a connection. This yields different equilibrium outcomes in the two frameworks we study.

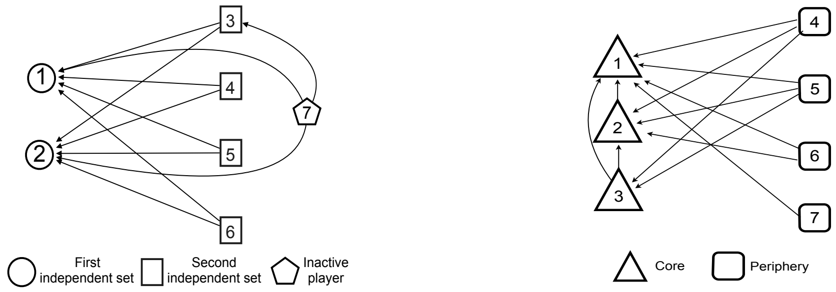
In the model with cost heterogeneity, better types are more efficient in producing the public good. Hence, they are less willing to create costly links to access the public good produced by others. As a result, better types establish few or no links and produce a lot, while worse types free ride on them. Social hierarchies with a pyramidal structure emerge, as in the example of Figure 1(a). In these networks, called *complete multipartite graphs*, active players are ordered in independent sets, which are groups of unconnected similar types. Better types produce more and belong to sets that comprise fewer players. Active players link to all better types in other independent sets. Inactive players instead are pure free riders and might link only to some players in an independent set, such as player 7 in Figure 1(a). This implies that the number of neighbors need not be monotonic in player type.

In complete multipartite graphs, highly connected players tend not to be neighbors themselves. Our model suggests that this property, which emerges in several situations,[1] arises because more efficient players have lower gains from a connection. Therefore, they sponsor few links, acquire a lot of information on their own, and receive many in-links from bad types.

In the model with heterogeneity in valuation, players with a higher valuation of the public good benefit more from free riding than worse types. Therefore, better types have more links. They also are the most active since they need to acquire a greater amount of the public good. Since

[1]For example, market mavens do not rely on other market mavens' information (Feick and Price, 1987). Similarly, the most experienced farmers do not learn from each other, but rather inexperienced farmers learn from them (Conley and Udry, 2010).

additionally a player always links to the players producing the most, equilibrium networks are nested split graphs in which one's neighborhood is a subset of the neighborhoods of better types (see Figure 1(b)). Hence, the best types are connected. Periphery players sponsor links, but do not receive any. Hence, they complement the public good they receive from their neighbors to obtain their desired amount. As a result, the level of production of periphery players need not be monotonic in type.



(a) Heterogeneity in Production Cost

(b) Heterogeneity in Valuation

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

These results show that better types need neither have more links nor produce more. Hence, empirical studies should be careful in interpreting centrality or production as a measure of (often unobservable) individual characteristics.

However, both models predict *negative assortativity*: very connected players, i.e., the best types who produce and consume more public good, tend to have a large number of links with poorly connected players. This is a property of many real-world networks (Newman, 2002).

We show that a similar result as the law of the few holds when players are heterogeneous. Indeed, as the population size increases, there cannot be too many players producing a lot of public good. Otherwise, some players would consume an infinite amount of it. However, this is not possible: since we assume decreasing marginal returns, these players would then become inactive or delete some links. Thus, if there are many active players, most of them produce an infinitesimal amount of the public good.

In a context with heterogeneous players, it is natural to ask whether the network exacerbates the initial inequality in payoffs. In general, it is not possible to answer this question because both good and bad types benefit from the network by accessing each other's production of the public good once they are linked. In large societies, however, the law of the few implies that most people free ride on a few players. Depending on the source of heterogeneity, free riders gain more or less from a connection. Hence, networks dampen inequality for most players under cost heterogeneity and increase it under heterogeneity in valuation.

Efficient networks are stars from which the worst or best types (but 1) are excluded depending on the source of heterogeneity. Indeed, this structure minimizes the linking cost. Furthermore, the planner precludes some players from free riding if they are either too efficient or value the public good too little to justify paying the linking cost. Hence, equilibrium networks are over-connected.

The model described so far is very stylized. Yet our results are robust to several empirically relevant extensions, such as *(i)* the introduction of both types of heterogeneity at the same time, *(ii)* the indirect flow of spillovers, *(iii)* decay, *(iv)* imperfect substitutability between one's own effort and that of others, and *(v)* heterogeneity in linking cost.

Finally, we assume unilateral link formation because often those who initiate communication bear the associated cost. However, in some situations, mutual consent is needed to create a link. In that case, we show that the same networks are still equilibria as long as players can make transfers.

Related literature. The few papers in the literature that address games played in endogenous networks mostly model network formation as a generic socialization effort (Cabrales, Calvó-Armengol and Zenou, 2011; Merlino, 2014) or as a not fully strategic decision (König, Tessone and Zenou, 2014).

The most significant exception is G&G, in which homogeneous players provide a local public good and establish links. Strict Nash equilibria are core-periphery networks. However, some complementarity in neighbors' actions or decay in the information flow invalidates these results. By characterizing the Nash outcomes with an arbitrary degree of heterogeneity among players,[2] we get richer network structures and more robust results. In Section 3, we discuss in detail how our work relates to G&G's.

[2]G&G consider a very limited form of heterogeneity, i.e., when one player has a lower production cost than the others.

When players' efforts are strategic complements, either nested split graphs or complete multipartite graphs emerge when the value function is convex (Hiller, 2012) or concave (Baetz, 2015), respectively. For a game of strategic substitutes, we show that these structures emerge depending on how the gains from a connection vary with types.

We model network formation non-cooperatively as in Bala and Goyal (2000). When there is no strategic interaction beyond link creation, heterogeneity in valuation plays a minor role (Galeotti, Goyal and Kamphorst, 2006).[3] Yet, it significantly affects the equilibria when linking costs are also heterogeneous (Billand, Bravard and Sarangi, 2011) to a point that Nash networks might not even exist (Haller, Kamphorst and Sarangi, 2007). In our model, the players' benefit from linking is determined not only exogenously by their type, but also endogenously by their choice of effort.

Bramoullé and Kranton (2007) study public good provision in fixed networks: there always exist specialized equilibria in which active players belong to an independent set and their direct neighbors are inactive. These results have been generalized to any degree of substitutability by Bramoullé, Kranton and D'Amours (2014). However, links are often endogenous. In this case, fewer effort profiles are equilibria because players establish a link only if it yields them access to enough public good.

The paper proceeds as follows. Section 1 introduces the model. Section 2 characterizes the equilibria and their welfare properties. Section 3 discusses the comparison with the homogeneous agents' case and several extensions. Section 4 concludes. The proofs are in the Appendix.

1 Model

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

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

Network. We denote the directed network g by an adjacency matrix in

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

which each line represents player i's links 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 $g_{ij} = 1$. Similar to a phone call, the linking cost is paid by the player who initiates the communication. Hence, linking decisions are one-sided: the player proposing a link pays k and the link is established. Since in our game direct spillovers are never negative, incoming links are always accepted.

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.

While network g is a directed graph, its closure $\bar{g}$ is an undirected network such that $\bar{g}_{ij} = \max\{g_{ij}, g_{ji}\}$, 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 players such that there is a path connecting every two players in the set and no path to players outside the set. A network $\bar{g}$ is *connected* if there is a unique component encompassing all players, and *minimally connected* if it is connected and there exists only one path between every pair of players. We denote the set of isolated players by $\mathcal{I}(\bar{g}) = \{i \mid \bar{g}_{ij} = 0 \text{ for all } j \in N\}$.

In a *core-periphery graph*, there are two groups of players, the *periphery* $\mathcal{P}(\bar{g})$ and the *core* $\mathcal{C}(\bar{g})$, 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 players' neighbors are nested is a *nested split graph*: for any pair of players 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*, players can be partitioned into a number S of independent sets $\mathcal{H}_s(\bar{g}^*)$, $s = 1, ..., S$, such that every player shares a link with all players outside her own set.

A network is *negative assortative* if the average degree of a player's neighbors decreases with her own degree.

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. Player i's payoffs under strategy profile (x, g) are:

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

where $c_i > 0$ is i's cost of producing the public good, $k > 0$ 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 $lim_{x \to \infty} f_i'(x) = m_i < c_i$. Since a player's benefits depend on the sum of her investment in the public good and that of her direct neighbors, the investments of connected players are perfect strategic substitutes. Hence, the game exhibits positive local externalities.

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., players 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, richer consumers, farmers with more land and firms with bigger market shares value information more.

Under both specifications, players' types capture the amount of public good they would optimally collect in isolation and $a_1 > a_2 > ... > a_n$. We refer to lower-indexed players 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 by i, 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 vary 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 heterogeneous valuations, better types benefit more from spillovers, while the players' production costs are identical.

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 players, small perturbations of the valuation or the production costs are enough to break eventual ties. Hence, we focus on *strict* equilibria in which the inequalities in the above definition are strict for all players.

Social Welfare. A strategy profile s is efficient if it maximizes the sum of individual payoffs, i.e., if $\sum_{i \in N} U_i(s) \geq \sum_{i \in N} U_i(s')$ for all $s' \in S$.

2 Main Analysis

In this section, we provide a characterization of the equilibria of this local public good game and derive its implications for large societies and inequality. Moreover, we obtain the efficient solution.

2.1 Equilibrium Analysis

First, we state some general results that help us characterize the equilibria of both models. Lemma 2 shows that, in equilibrium, active players always collect exactly the amount of public good they would in isolation.

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 straightforwardly extends a result in G&G to heterogeneous players. Indeed, given other players' strategies, an active player accessing more than the amount of public good she would optimally collect in isolation can increase her payoff by producing less due to the strict concavity of the valuation function.

Since each player optimally collects a different amount of public good in isolation, Lemma 2 implies that there cannot be two active players that consume the same amount of public good. Take now any two active players. Then, either they are not linked, in which case one cannot access the public good produced by the other; or they are linked and, for them to access different amounts of public good, there needs to be at least one active player who is connected to only one of them.

Lemma 3 *Take two active players i and $j \in N$. If $\bar{g}_{ij}^* = 1$, there exists at least one active player $z \in N$ who is neighbor of only one of them.*

This lemma implies that any complete core-periphery network, in which all players in the core are linked and have the same neighbors, can never be an equilibrium, except for stars, where there is only one player in the core.

2.1.1 Heterogeneity in Production Cost

When better types have a lower cost of producing the public good, they have lower gains from a connection. The payoff function is 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 relates public good provision and type in this model and characterizes the resulting network outcome.[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 players form a complete multipartite graph in which better types produce more and are in independent sets that comprise fewer players.*

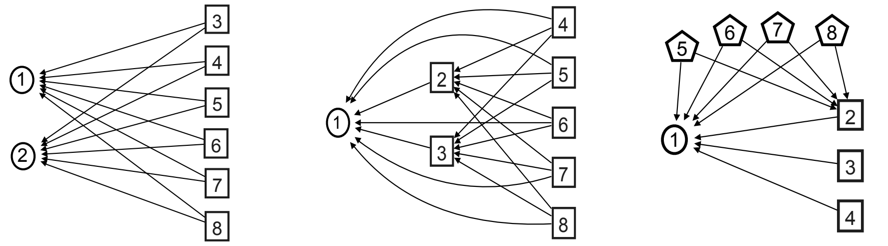
The characterization of Theorem 1 is very sharp. This is surprising since this game typically has many equilibria, some of which emerge only for certain parameters. Figure 2 presents some examples.

Let us now illustrate some general implications of the theorem.

First, focusing on active players, *any better type produces more* and thus has (weakly) more in-links than any worse type. These properties emerge because: *(i)* any player first links to the one producing the most, *(ii)* better types gain more from a connection (Lemma 1), and *(iii)* active players of better types consume more public good (Lemma 2). Intuitively, consider a player i who is more efficient than player j but $x_i < x_j$. Then, i cannot have more in-links than j, since it is optimal to link to players who produce the most. This implies that i needs to acquire more public good than j by linking to other active players. However, since $c_i < c_j$, if i links to someone, so would j. Hence, this leads to a contradiction.

In fact, Theorem 1's proof is more involved: to link might not be profitable for j if thereby she becomes inactive. Suppose this is the case. Then, checking the amount of public good that each player consumes, together with Lemma 2, yields that j has to be similar to the players i links to;

[4] When linking is sufficiently costly, the unique equilibrium network is empty. As k decreases, player n, who benefits the most from linking to 1, eventually links to 1. Below this linking cost threshold, a periphery-sponsored star with player 1 as the hub producing a_1 is always an equilibrium. This argument 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 $f(x, g) = 2\sqrt{x_i + \sum_{j \in N_i(\bar{g})} x_j}$.

thus, j herself should receive some links. We can iterate this argument and, when all players are taken into account, the situation is analogous to the one described in the previous paragraph. Then, a contradiction arises.

Since best types produce the most, less efficient players free ride on them. In particular, free riders link to players producing the most as long as the gains from connections are greater than the linking cost. As a result, players with similar production costs sponsor links to the same players and receive links only from worse players. Hence, similar types who sponsor links are not connected themselves.

This yields a *complete multipartite graph* in which active players are ordered in independent sets, or tiers, according to their type. They sponsor out-links to all better players in other independent sets and receive in-links from all active players in independent sets with worse types. Since better players are in higher tiers, this structure yields *vertical clustering*: a player's neighbors are likely to be neighbors as well when they are sufficiently different.

Lower tiers comprise more players, yielding a *pyramidal structure* among active players. Indeed, take two players i and j in two different tiers such that $i < j$. Since players in the same tier are not linked, i does not access the public good produced by players in her tier, but rather she accesses that produced by players in the same tier as j. Since better players need to acquire more public good, the total production of the players in j's tier must be larger than any player's production in i's tier.

Given that equilibrium hierarchies are pyramidal, better active players have many more links than worse ones. Moreover, since everybody links to the most efficient players (and similar types are not linked), the average degree of a player's neighbors decreases with her own degree, i.e., *equilibrium networks are negative assortative.*

Furthermore, if more than one of the connected players is active, then *the network is connected.* Clearly, only very efficient players with very low gains from linking can be isolated. However, this would require player 1 to be the only active player in the component. Indeed, isolated players should be more efficient than 1's neighbors, because they are not willing to link to 1. At the same time, isolated players collect less public good than 1's neighbors, who should then be inactive.[5]

Best types are either neighbors, and then belong to different tiers of the hierarchy, or they are not neighbors when they all belong to the top tier. In this case, a core does not emerge because best players' gains from a connection are so low that linking to each other is not profitable for them. This happens for example in the network depicted in Figure 2(a): while players 3 to 8 link to 1, player 2 is so efficient that she is not willing to pay the linking cost k to link to 1. Hence, highly connected active individuals are not themselves neighbors.

While Theorem 1 characterizes the networks formed by active players, the predictions regarding inactive players' behavior are less sharp. Again, worse types always link to the most active players and sponsor more links because they gain more from a connection. However, inactive players might link only to some players in an independent set, such as 5 to 8 in example 2(c). In this case, inactive players might have more links than the players to which they do not link and who are more efficient than them, such as

[5]If 1's neighbors are active, they access the optimal amount of public good they would acquire in isolation, so that this amount is greater than x_1^*. Isolated players instead produce more than i's neighbors and less than x_1^* since they do not receive links. This leads to a contradiction.

3 and 4 in example 2(c). Therefore, in general, a player's type and her number of neighbors are not related, as 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*.*

Hence, when there are many inefficient players that are inactive, *the number of neighbors might not provide information about a player's type.*

2.1.2 Heterogeneity in the Valuation of the Public Good

When better types value the public good more, the payoff function is

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

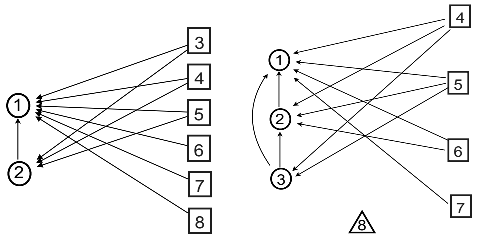
In this case, the gains from a connection are higher for better types. We now characterize the equilibrium network structure for this case.[6]

Theorem 2 *Under heterogeneity in the valuation of the public good, if* $k \leq f_2(a_1) - f_2(a_2) + ca_2$*, in a strict Nash equilibrium,* $\bar{g}^*$ *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.*

While again there might be multiple equilibria depending on the parameters of the model, Theorem 2 gives a surprisingly sharp characterization of all equilibrium networks. We now discuss the structures implied by Theorem 2 through two particular examples presented in Figure 3.

The best players need to consume more public good, but also gain more from a connection. As a result, they are the most active and receive in-links. Furthermore, the fact that all players' costs of linking and producing the public good are identical implies that best players are linked among

[6]When linking is sufficiently costly, the unique equilibrium network is empty. As k decreases, player 2, who benefits the most from linking to 1, eventually links to 1. Below this linking cost threshold, a periphery-sponsored star with player 1 as the hub producing a_1 is always an equilibrium. This argument guarantees equilibrium existence.



(a) (b)

				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 $c = .3$.

themselves. For example, consider players 2 and 3 in Figure 3(b), who both receive links. Given that the linking cost is sufficiently low to make free riding on their production profitable, that is, $cx_i > k$ for $i = 2, 3$, it is also profitable for 2 and 3 to link to each other.[7] Hence, *a core emerges* and it is constituted of the players who need more public good.[8]

Worse players do not produce enough to receive in-links. Thus, they form a periphery of free riders who link to the most active players. However, all players in the periphery do not share the same neighbors. Indeed, a player who would become inactive after establishing a link might not gain enough from this link to be willing to pay the associated linking cost.

These facts imply that *equilibrium networks are nested split graphs* in which the neighborhoods of worse types are subsets of the neighborhoods of better types. Hence, *degree centrality is always higher for better types.*

This property together with the fact that better types need to consume more public good implies that *equilibrium networks are negative assortative.*

The characterization of Theorem 3 neither pins down the production level of periphery players nor implies that the network is connected.

First, better types in the periphery need not produce more since a player who has one link more than a worse type might exert less effort in order to reach her optimal amount of public good. For example, in Figure 3(a), players 3 to 5 produce less than 6, who does not link to 2.

[7]Suppose that 3 does not link to 2. This implies that $f_3(a_3) - cx_3 > f_3(a_3 - x_3 + x_2) - k$, which cannot hold since $a_3 - x_3 + x_2 \geq a_3$ and $cx_3 > k$.

[8]Note that it is undetermined who proposes a link between players in the core since both parties would sponsor it. Yet, the different directed graphs have the same closure.

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

However, this non-monotonicity does not arise in the core. Indeed, take two players i and j in the core such that i values the public good more than j. Both are connected to all other players in the core, but i needs to access strictly more public good than j. Hence, i has to receive more links than j. This requires i to produce more than j.

Second, worst types are isolated if they do not benefit enough from linking to anyone, such as player 8 in example 3(b).

Our characterization of equilibrium networks with heterogeneous players shows that, in general, a player's type and investment 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.

2.2 Large Societies

The law of the few was formally derived by G&G. It predicts that as the number of players increases, the proportion of active players in non-empty strict equilibrium networks goes to zero. We now show that a similar result also holds when players are heterogeneous.

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

Proposition 1 *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, $\lim_{n\to\infty} \mathcal{A}_{IN}(x^*, g^*)/n = 0$.*

In a model with heterogeneous players, the number of active players need not be bounded: some players might be isolated and some peripheral players might complement the amount of public good received from players in the core. However, peripheral players as a whole cannot produce but a limited amount of public good. Otherwise, the players receiving in-links would produce too little to make linking to them beneficial.

Therefore, the number of players that produce more than an infinitesimal amount of public good and the number of players that receive links are bounded as the population size increases. For large societies, this implies that a small group of players produces a significant amount of public good, while the others mostly free ride or are isolated.

2.3 Inequality

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

In the examples reported in Table 1, however, 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 the other players.

	$f(x,g)$ 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 players who receive links. Hence, it is not possible to free ride on most free riders, and what only matters is how the gains from a connection vary with type.

We now state this result formally. 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 2 *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, while it increases under heterogeneity in valuation.*

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

2.4 Efficiency

The following proposition characterizes the efficient allocation of production and links in both models.

Proposition 3 *The efficient network is a star whose hub produces* y *such that* $\sum_{i:\bar{g}_{i1}=1} f_i'(y) = c_1$*. Under heterogeneity in production cost, the hub is* 1 *and there is* $\underline{n}$ *such that players* $1 < i < \underline{n}$ *are isolated. Under heterogeneity in valuation, there is* $\bar{n} \leq n$ *such that players* $i > \bar{n}$ *are isolated.*

Since linking is costly and free riding is limited to direct neighbors, minimizing linking costs maximizes social welfare. 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 players is very different in the two models: under cost heterogeneity, the most efficient players but 1 might be isolated, while under heterogeneity in valuation, the players that value the public good the least might be isolated. Hence, if the equilibrium network is non-empty and not a star, then it is over-connected.

Moreover, non-empty equilibrium networks entail under-investment since the hub does not internalize the value of her production to her neighbors.

3 Discussion

Homogeneous players. For homogeneous players, G&G show that strict equilibria are complete core-periphery structures in which periphery players are inactive. In non-strict equilibria instead, there are low producing players that link to high producing players, but not among themselves. The resulting networks can be core-periphery structures, complete multipartite graphs, or there can even be more than one component.

In comparison, our analysis reveals some striking features. When players are heterogeneous, we can focus on strict equilibria because small perturbations are enough to break ties. Furthermore, each player's optimal amount of public good is different. Hence, all active players cannot be connected and share all their neighbors (Lemma 3). In other words, either a core does not emerge or active periphery players do not all connect to the same core players. Hence, complete core-periphery structures are not equilibria, except for a star with player 1 as the hub.

These differences imply that the characterization of the set of strict equilibria changes when we introduce some heterogeneity across players in the production cost or in the valuation of the public good. This is summarized in the following remark.

Remark 1 *In the model with heterogeneous players, the only non-empty equilibrium network that has a complete core-periphery structure is a star.*

Indeed, starting from a complete core-periphery structure that G&G obtain for homogeneous players, some links need to be established or deleted to get a multipartite graph (Theorem 1) or a nested split graph (Theorem 2). For example, consider the equilibrium of the model with heterogeneous players depicted in Figure 3(a). This is no longer an equilibrium if $\bar{a}_1$ converges to $\bar{a}_2$: as players 1 and 2 become more homogeneous, x_2^* goes to zero. At some point, players 3 to 5 break their links with 2. Eventually, the star with 1 as the hub is the only equilibrium.

However, there are other networks that are equilibria for homogeneous and heterogeneous players. Quite surprisingly though, these are Nash equilibria of the model with homogeneous agents that are *non-strict*.

Remark 2 *There are complete multipartite networks that are strict equilibria in the model when players are heterogeneous in the cost of producing the public good and non-strict equilibria when players are homogeneous.*

We illustrate this by means of an example. Consider the network in Figure 2(a) and let the economy converge to the homogeneous players' case as follows: let $a_1 = a_2 = \bar{a}$ and $a_i = \underline{a}$ for $i = 3, ..., 8$, i.e., there are two types of players. Then, the network presented in Figure 2(a) is an equilibrium as both $\bar{a}$ and $\underline{a}$ converge to .527 if $\bar{a} \geq \underline{a}$ and $\underline{a} \in [(11k/12 + \sqrt{(11k/12)^2 + \bar{a}/6}, 11k\sqrt{\bar{a}}/6 + \bar{a}/6]$.

When eventually $\underline{a} = \bar{a} = .527$, i.e., for homogeneous players as in G&G, the network is still a Nash equilibrium, although not a strict one. Indeed, 2 is indifferent between producing and linking to 1 (and vice versa).

The focus on strict Nash networks is natural due to their resilience when a link is deleted or established by mistake. However, Remark 1 stresses that the only strict Nash network which is robust to introducing heterogeneity is the star. Furthermore, Remark 2 shows that there are non-strict Nash networks which are robust to introducing cost heterogeneity. Hence, when players are homogeneous, strict Nash networks might be less robust than non-strict ones to trembles on players' payoffs affecting the amount of public good that players would optimally acquire in isolation.

Finally, given homogeneous players, small complementarities in neighbors' actions or decay in the information flow change the properties of equilibrium networks. We now show that, instead, the characterization of the model with heterogeneity is robust to decay and many other extensions.

Robustness Analysis. Although our benchmark model is quite stylized, we now show that the results we have derived are robust to several extensions. 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, while

– $\varepsilon_1 \in \mathbb{R}^N$ and $\varepsilon_2 \in \mathbb{R}^N$ introduce **both types of heterogeneity** at the same time, since players are often heterogeneous along several dimensions;

– $\varepsilon_3 \in \mathbb{R}^N$ introduces **heterogeneity in linking costs**; for example, some individuals have cheaper phone rates;[9]

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

– $\varepsilon_4 \in [0,1]$ 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 $(1-\varepsilon_4)x_j$;
– $\varepsilon_5 \in [0,1]$ introduces **indirect spillovers** since the public good is often also shared among indirect neighbors; in that case, information is discounted by ε_5 for each link it travels in the network;[10]
– $\varepsilon_6 \in [0,\infty)$ 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 a player 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 *For each strict equilibrium network g^*, there exist shocks ε such that g^* is an equilibrium network of the perturbed game if $|\varepsilon| < \bar{\varepsilon}$.*

In strict equilibria, all inequalities representing players' optimal linking decisions are strict. Moreover, each player's payoffs are continuous in the shocks. Hence, given the shocks, the players' effort levels can be adjusted in a consistent way without violating the initial equilibrium conditions. Furthermore, when the shocks are such that inactive players remain inactive, the law of the few holds in the perturbed game.

Two-Sided Link Formation and Transfers. Some situations captured by our model are 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, players might ask for compensation to share the public good they produce. In what follows, we study the impact of this different network formation protocol on equilibrium properties.[11]

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, $\tau_{ij} + \tau_{ji} \geq k$.

[10] Since ε_5 converges to zero from above, it is natural to define $\lim_{\varepsilon_5 \to 0^+} {\varepsilon_5}^0 = 1$. Thus, our benchmark model with only direct spillovers corresponds to the case $\varepsilon_5 = 0$. If instead $\varepsilon_5 = 1$, equilibrium networks are minimally connected with possibly some isolated players.

[11] Without transfers, other equilibria arise because players might refuse some links.

A strategy profile $s = (x, \tau)$ specifies investments x and transfers $\tau = \{\tau_1, ..., \tau_n\}$. The payoff function is then defined as

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

A link between two players is formed only if it is profitable for both of them. Formally, 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} (Jackson and Wolinsky, 1996; Bloch and Jackson, 2007).

Proposition 5 *Given (x^*, g^*), 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 an equilibrium in the two-sided model with transfers that induces $\bar{g}^*$.*

Hence, each equilibrium network under one-sided linking is an equilibrium under two-sided link formation with transfers. The proof is omitted since it suffices to note that the player proposing a link under one-sided linking can always find transfers so that the link is accepted by the other party.

4 Conclusion

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

The source of heterogeneity determines how the gains from a connection differ across types. This affects the relationship between public good provision, players' types, and their position in the network. Hence, our analysis uncovers to what extent correlation in behavior is due to individual characteristics that determine network position. In this sense, our results are relevant beyond the theoretical literature on networks.

Surprisingly, the network structures we single out also arise under strategic complements when the value function is either convex (Hiller, 2012) or concave (Baetz, 2015). Hence, future research should investigate whether more general results can be obtained.

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 that $c_i = c$ 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 not. Since i and j are active, by Lemma 2, they access a_i and a_j respectively. But i and j have the same neighbors and $\bar{g}_{ij} = 1$, which implies that $a_i = a_j$, contradicting $a_i \neq a_j$. ■

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* that this is not the case. Without loss of generality, consider first best type j and worst type $i < j$ such that $x_i^* < x_j^*$. We show in a series of Lemmata that a contradiction arises.

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 produce 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* $L = \{l : x^*_l > 0, \bar{g}^*_{lj} = 1, g^*_{li} = 1, \bar{g}^*_{lp} = 0 \text{ for any } p \in P\}$.

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 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 $x_p^* + x_j^*$ and i does not, while i receives x_i^* and 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 6.*(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 contradiction and 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 players that 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 players 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 players 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 players 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 that $x^*_j > x^*_l$. 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 players 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. If g^* is empty, the statement follows trivially. If the network is non-empty, for all players j such that $\eta^{OUT}_j(g^*) > 0$, $g^*_{j1} = 1$ 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 players 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 2. Let $n \to \infty$ and consider $i < j$ which both receive no in-links. Then $|U_i(x^*, g^*) - U_j(x^*, g^*)| \leq \underline{U_i} - \underline{U_j}$. 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 $c_i(a_i - x_i) - k\eta_i^{OUT}(g^*) \leq c_j(a_j - x_j) - k\eta_j^{OUT}(g^*)$, which means $c_i \sum_{z:\bar{g}^*_{zi}=1} x_z \leq c_j \sum_{z:\bar{g}^*_{zj}=1} x_z$. 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 $\eta_i(g^*) = \eta_j(g^*)$, $U_i(x^*, g^*) - U_j(x^*, g^*) \leq \underline{U_i} - \underline{U_j}$ can be rewritten as $f(a_i - x^*_i) - c_i(a_i - x^*_i) \geq f(a_j) - c_j a_j$. 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 $\eta_j^{OUT}(g^*) > \eta_i^{OUT}(g^*)$, $U_i(x^*, g^*) - U_j(x^*, g^*) \leq \underline{U_i} - \underline{U_j}$ can be rewritten as $f(\sum_{z:g_{jz}=1} x_z) - k[\eta_j^{OUT}(g^*) - \eta_i^{OUT}(g^*)] \geq f(a_j) - c_j a_j$. 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 $f(\sum_{z:g_{ji}=1} x_z) - k[\eta_j^{OUT}(g^*) - \eta_i^{OUT}(g^*)] \geq f(a_i - x^*_i) - c_i(a_i - x^*_i)$. This concludes the proof of the first part of Proposition 2.

Under heterogeneity in valuation, i and j can be active or inactive. Suppose i and j are active and $\eta_i^{OUT}(g^*) = \eta_j^{OUT}(g^*)$. Then, $U_i(x^*, g^*) - U_j(x^*, g^*) \geq \underline{U_i} - \underline{U_j}$ 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, $U_i(x^*, g^*) - U_j(x^*, g^*) \geq \underline{U_i} - \underline{U_j}$ implies $c(a_i - x_i) - k\eta_i^{OUT}(g^*) \geq c(a_j - x_j) - k\eta_j^{OUT}(g^*)$. Since all neighbors but z are common, this implies $cx_z \geq k$, which needs to hold because i links to z. The same argument holds when i has at least two more neighbors 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 2. ■

Proof of Proposition 3. In any component only one player produces to minimize linking costs.

Under heterogeneity in production cost, it is efficient that only the most efficient player, 1, produces while, for low enough k, all others link to 1. Hence, a star with 1 as the hub is the efficient network. Note next 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

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

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 players are isolated; (A-1) is constant in k when all players 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 $y(\underline{n})$ such that $(n - \underline{n} + 2)f'(y(\underline{n})) = c_1$.

Under heterogeneity in valuation, suppose there are several components. Since players 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, for low enough k, the efficient solution is a star with only one active player. Note next 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

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

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 players is given by

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

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 players are isolated; the function is constant in k when all players 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 players 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 $y(\bar{n})$ such that $\sum_{i=1}^{\bar{n}} f_i'(y(\bar{n})) = c$. 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 player i such that $x_i^* = 0$, $a_i < \sum_{j \in N} \bar{g}_{ij}^* x_j^*$ by Lemma 2 while for any active player i, $a_i = x_i^* + \sum_{j \in N} \bar{g}_{ij}^* x_j^*$. Defining the adjacency matrix of links among active players 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 players, 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

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.

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.

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

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

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.

Jackson, Matthew O., and Asher Wolinsky. 1996. "A Strategic Model of Social and Economic Networks." *Journal of Economic Theory*, 71(1): 44–74.

König, Michael D., Claudio J. Tessone, and Yves Zenou. 2014. "Nestedness in networks: A theoretical model and some applications." *Theoretical Economics*, 9(3): 695–752.

Merlino, Luca Paolo. 2014. "Formal and informal job search." *Economics Letters*, 125(3): 350–352.

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

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.