Relationship Externalities

Selman Erol* Rakesh Vohra†

October 2022

Abstract

We propose a model of network formation where agent's payoffs depend on the connected component they belong to in a way that is specific enough to be tractable yet general enough to accommodate a number of economically relevant settings. Among them are formation in the presence of contagion via links and collaboration with spillovers. A key feature of this setting is that the externalities stem from *links* rather than nodes. We characterize stable and efficient networks. Under negative externalities, disjoint cliques are stable and efficient. Under positive externalities complete networks and star networks are stable. Efficient networks feature a mix: pineapple networks which consist of one large clique and a star network appended to each other.

JEL classification: D62, D85

Keywords: Network formation, Stability, Strong Stability, Externalities

*Tepper School of Business, Carnegie Mellon University (email: erol@cmu.edu).

†Department of Economics and Department of Electrical and Systems Engineering, University of Pennsylvania (email: rvohra@seas.upenn.edu).

1 Introduction

We propose a model of network formation where an agent's payoff depends on the connected component they belong to in a way that is general enough to accommodate a number of economically relevant settings, yet specific enough to be tractable. An agent's payoff depends on own degree *and* the number of links in the component it belongs to. Formally, the utility of an agent is denoted $U(d, e)$ where d is its degree in the network and e is the number of links in the component that contains it.

U increasing in d but decreasing in e captures negative link-externalities. Links incident to oneself are beneficial but links incident to others are not. This would hold in a model of club formation where friends of friends invited into the club create negative externalities because of the competition for scarce club resources.[1] Suppose an agent agent can only join one club. The benefit of doing so arises from a direct connection to other individuals in the club. All agents in the same component are part of the same club. A link between a pair represents the pairwise activities/events that benefit the pair. Each such activity consumes club resources. Unused club resources are used for public goods that benefit the whole club. For example, if b is the total club budget, c the cost of each activity, r the benefit to the parties from the activity/event, then, the payoff function is $U(d, e) = rd + V(b - ec)$ where $V(\cdot)$ is the benefit from the public good provision within the club. A similar trade-off is present in the formation of file sharing networks Neglia et al. (2007).

Our formulation also encodes, in reduced form, the effect of contagion through link rather than node failure.[2] Link failure is responsible for performance degradation in the power grid and communication networks. An overloaded link can lead to capacity loss and result in cascading breakdowns. Damage to optical fiber cables can partially overload data delivery, resulting in a regional interruption of Internet services. In an economic context suppose firms enjoy a benefit from each direct link to other firms via a joint venture. The venture is risky and can fail which may cause the counterparties financial distress and force them into default. Even if a venture does not fail, a defaulting counterparty to the venture may cause one to default. Therefore, each firm bears a default risk from each direct and indirect linkage to other firms.

[1]Initial social interactions between individuals often occur within social clubs such as fraternities.

[2]Existing contagion models, for example Kempe et al. (2003), focus on contagion that starts with nodes. In these models links serve only as a conduit for externalities.

Whenever, one of the links in the component containing the firm is hit by a negative shock, then all firms in the same connected component face a default risk (see Erol and Vohra (2022) for a detailed specification). More generally, the institutions that underpin contracts and facilitate business transactions decline in quality. This makes the associated bilateral relationship more prone to a disruption which can trigger a cascade.

We also consider the 'opposite' case where U is decreasing in d but increasing in e. This captures positive link-externalities. Incident links are costly, but other's links are beneficial. This captures research collaborations with knowledge spillovers. Collaborating on a project is costly and the cost rises with their number. However, knowledge transfer from other projects via chains of collaborators complements ones own productivity. For more details see the example before Proposition 6.

We characterize strongly stable networks under both negative and positive link externalities. The characterization does not rely on a particular functional form for $U(d, e)$ which will make it useful for applications. In the presence of negative link-externalities, strongly stable networks consist of disjoint cliques of a particular size that depends on $U(\cdot, \cdot)$. Under positive link-externalities, we find that strongly stable networks are either star networks or complete networks. The center of the star network serves as a public good provider and receives a significantly lower payoff than the other agents. In all cases, strongly stable networks are Pareto efficient, but not necessarily efficient. We identify various instances where strong stability implies efficiency.

These results hold even with heterogenous payoff functions (see Section 7 for details), i.e., the payoff to agent i is $U_i(d_i, e_i, f_i)$ where f_i is the number of nodes in the component of node i (which allows for node externalities), U_i is strictly decreasing/increasing in e_i (the number of links in i's component), weakly decreasing/increasing in f_i, and depends arbitrarily on d_i, node i's degree. This last feature allows one to encode link costs.

The next section provides an overview of the important differences with prior work. More detailed discussions of differences and similarities appear in proximity to the formal statements. The model and notation appear in Section 3. Section 4 discusses the case of negative externalities while Section 5 focuses on positive externalities. Section 6 discusses efficiency and examples while Section 7 describes some extensions of our model.

2 Related Literature

A general way to model payoffs is to use a function that depends on the *entire* network. There are a large number of papers on network formation within such a framework (see, for example, sections 8.2.1. and 8.2.2. in Vannetelbosch and Mauleon (2016)). They provide characterizations under *weaker* notions of stability with explicit functional forms for agent payoffs which focus on node-externalities. We characterize stable and strongly stable networks under qualitative assumptions on the payoff functions that emphasize link-externalities.

Goyal and Joshi (2006) and Buechel and Hellmann (2012) also consider link externalities with an important difference. In these papers, the utility of an agent depends on the aggregate number of *all* links. Thus, an agent is subject to an externality by agents to which it is not connected. In our case, only the links in the connected component that contain the agent matters. In Morrill (2011) externalities are limited to the number of links neighbors have with others. The precise difference in outcomes is discussed in Section 4 and 5.

Our negative externality case is roughly related to models of contagion via nodes rather than links. Network formation outcomes result in disjoint cliques as in our model. Blume et al. (2013) show, in a model of independent contagion, that pairwise Nash stable networks exist and consist of disjoint cliques. Erol (2019) shows that disjoint cliques are the unique strongly stable outcome under threshold contagion but his model features node-externalities rather than link-externalities.

Regarding our positive externality results, Galeotti et al. (2006), and Hojman and Szeidl (2008) find star networks as Nash equilibria using the non-cooperative connections model of Bala and Goyal (2000). The connections model does not map on to notions of stability we examine. The closest to ours is the connections model in Jackson and Wolinsky (1996) who find that star networks are pairwise stable for some range of parameters that feature positive node-externalities that decay with distance. To the best of our knowledge, there are no characterization results under link-externalities even for weaker solution concepts than strongly stable.

3 Model

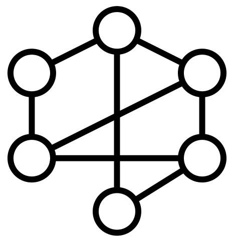
Preliminaries A *network* G is a pair (N, E) where $N = \{1, 2, ..., n\}$ is a finite set of *nodes* and $E \subset [N]^2$ is a set of *links*.[3] Throughout this paper we assume $n \geq 2$. Call two nodes $i, j \in N$ *neighbors* in G if $\{i, j\} \in E$. A link $\{i, j\} \in E$ is *incident* to i and j in G. For $i \in N$, denote by $N_{i,G} = \{j \in N_G | \{i, j\} \in E\}$ the set of neighbors of i in G and call $d_{i,G} = |N_{i,G}|$ the *degree* of i in G. A *path* from i_1 to i_k in G is a sequence of distinct nodes $(i_1, i_2, ..., i_k)$, all in N, such that $\{i_t, i_{t+1}\} \in E$ for all t. Two nodes i and j are *connected* in G if there is path in G between them. Network $G' = (N_{G'}, E_{G'})$ is a *subnetwork* of network G if $N_{G'} \subset N$ and $E_{G'} \subset E \cap [N_{G'}]^2$ denoted $G' \preceq G$. Note that G is a subnetwork of itself. A subnetwork $G' \preceq G$ is *connected* if every pair of nodes in $N_{G'}$ is connected in G' or if $|N_{G'}| = 1$. The *component* of a node i in G is the largest connected subnetwork of G that includes node i with respect to the order $\preceq$. Denote by $e_{i,G}$ the number of links in the component of i in G. The *payoff* to node i from network G is denoted $u_{i,G} = U(d_{i,G}, e_{i,G})$ where $U : \mathbb{Z}^2 \to \mathbb{R}$.

Pairwise Stability The conventional solution concept is pairwise stability due to Jackson and Wolinsky (1996). A network is *pairwise stable* if no pair of nodes wants to establish a new link and no individual node wants to sever a link. It is a weak concept because nodes can only add or delete links and never at the same time. While the existence of pairwise stable networks is easy to establish, it is a permissive concept. For example, suppose $U(d_{i,G}, e_{i,G}) = d_{i,G} - ce_{i,G}$ where $c \in (0.5, 1)$ for all i (this example appears again in Proposition 3). Then, *any* partition of the nodes into disjoint cliques of size at least 2 would be pairwise stable.[4]

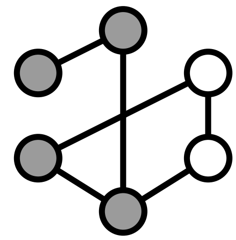
Stability Consider a candidate network $G = (N, E)$ and a subset of nodes $D \subset N$ called deviators. A feasible deviation from G by D allows nodes in D to *simultaneously* add any set of absent links within $[D]^2$ and delete any set of links incident to

[3] $[X]^2$ is the mathematical notation for the set of all 2-element subsets of a set X.

[4] Nodes can have negative payoffs in a pairwise stable solution which is problematic if nodes have a non-negative reservation payoff. To see why, consider the clique on three nodes and assume $c \in (\frac{2}{3}, 1)$. To verify pairwise stability, observe that no more links can be added. If any node deletes a link, its payoff becomes $1 - 2c < 2 - 3c$. But the payoff of each node is $2 - 3c < 0$. Accommodating reservation payoffs or allowing for cutting multiple links at once require modifying the notion of pairwise stability.



Candidate network G



A feasible deviation by the grey nodes.

Figure 1: A feasible deviation

any node in D. Formally, network $G' = (N, E') \neq G$ is a *feasible deviation* from G by $D \subset N$ if $E' \cap [N \backslash D]^2 = E \cap [N \backslash D]^2$ and $E' \cap ([N]^2 \backslash [D]^2) \subset E \cap ([N]^2 \backslash [D]^2)$. A feasible deviation is illustrated in Figure 1. A *profitable deviation* from G by D is a feasible deviation from G by D that strictly improves the payoff of each member of D: $u_{i,G'} > u_{i,G}$ for all $i \in D$. G is called $k-$*stable* if there is no profitable deviation from G by any $D \subset N$ with $|D| \leq k$. We refer to 2-stable networks as *stable networks*, and $\infty-$stable networks as *strongly stable networks*.[5]

Our notion of strong stability coincides with Dutta and Mutuswami (1997) whose focus is to find value functions and sharing rules such that efficient and strongly stable networks exist. This endeavor is categorically different from that of characterizing strongly stable networks for given payoff functions.

The notion of strong stability is weaker than that in Jackson and Van den Nouweland (2005) because they also require immunity against Pareto improving deviations. To distinguish it from strong stability, call it Pareto Strongly Stable (PSS). Jackson and Van den Nouweland (2005) show by example that existence of strongly stable networks does not imply the existence of PSS networks. If transfers are possible among deviators, PSS is perhaps the more suitable solution concept. In our setting, there are no transfers.

Other definitions that will be used throughout follow next. We collect them in one place for the convenience of the reader. For each integer k, let $B(k) = \frac{1}{2}k(k-1)$. A *connected network* is a network that is a connected subnetwork of itself. Call a subnetwork $G' \preceq G$ *the subnetwork induced by* $N_{G'}$ if $E_{G'} = E \cap [N_{G'}]^2$. Notice that

[5] 1-stable networks are Nash networks.

A clique A star

Figure 2: A clique and a star

components are induced subnetworks. The *order* of a subnetwork is the number of nodes in it. A *clique* is a subnetwork in which all nodes are neighbors of each other. A *disjoint clique* is a clique which is also a component. A $k-$clique is a disjoint clique of order k. A *complete network* is a network which is a clique. A *star* is a component in which one node called the *center*, is neighbors with all other nodes, called *leaves* wherein no two leaves are neighbors of each other. A *star network* is one where the network itself is a star component. A clique and a star are illustrated in Figure 2. A component is called *non-trivial* if there are at least two links in it, a *link-component* if there is exactly one link in it, and a *node-component* if there are no links in it. A *cycle* in G is a sequence of nodes $(i_1, i_2, ..., i_k)$ where $\{i_t, i_{t+1}\} \in E$ for all t and it holds that $i_t = i_{t'}$ iff $\{t, t'\} = \{1, k\}$. A node in a non-link component is called SC if it lies on a cycle or is adjacent to a leaf. Call it WC otherwise. Call a node i *central (C)* if $d_i = n - 1$ and *non-central (NC)* if $d_i \neq n - 1$.

4 Negative externalities

Suppose U is weakly increasing in d and strictly decreasing in e, namely *negative link-externalities*.

Proposition 1. *Under negative link-externalities, any stable network consists of disjoint cliques.*

The proof relies on a simple observation. If there are two nodes i and j who are not neighbors, but share a common neighbor k, then i and j can each cut their link with k and add the link $\{i, j\}$. This decreases the number of links in the component containing i and j, leaving the degrees of i and j unchanged. Therefore, the payoffs of i and j both strictly increase. One can increase the degrees of nodes in $D \subset N$ by either adding links inside D or by adding links between D and $N \backslash D$. Conditional on the desired increment to the degrees of nodes in D, the latter option entails twice the amount of negative externalities than the former option. High clustering minimizes the negative externalities imposed on a set of nodes conditional their degrees.

We contrast this result with Belleflamme and Bloch (2004), Goyal and Joshi (2006) and Morrill (2011). The first two examine the consequences of negative externalities that depend on the *total* number of links and not just the number of links in one's connected component.[6]

Clique size is not pinned down by stability, therefore there can be multiple stable networks. See Section 6 for an example. As a selection, we use strong stability. Denote the expected payoff of a node in a $(d+1)$-clique by $V(d) = U(d, B(d+1))$. Let $d^* = \arg\max_d V(d)$. We assume that d^* is well defined, which is generically true.

Theorem 1. *Suppose that n is divisible by d^*+1. Under negative link-externalities, strongly stable networks exist. A network is strongly stable iff it consists of (d^*+1)-cliques.*

For example, consider $U(d, e) = d - ce$ where $c > 0$. Then, $V(d) = d - \frac{c}{2}d(d+1)$. Take $\frac{1}{c} + \frac{1}{2} \in \mathbb{N}$ for simplicity. Then, a strongly stable network consists of disjoint $(\frac{1}{c} + \frac{1}{2})$-cliques. We explore this example in more detail in Section 6.

The proof uses a method we call the min-cut deviation. Suppose a profitable deviation exists. Among all such deviations, choose one which minimizes the number of links between the deviators and the non-deviators that remain after the deviation. The

[6] Buechel and Hellmann (2012) does the same, but allows for transfers which we do not.

In Belleflamme and Bloch (2004), payoffs have a specific functional form arising from Cournot competition. Pairwise stable outcomes are also disjoint cliques of varying size. However, their result relies on log convexity of firm profit functions. We impose no such requirement in payoffs (only monotonicity) and we also characterize strongly stable outcomes are also disjoint cliques but of the same size (see below).

In Goyal and Joshi (2006), payoffs are increasing in own degree and decreasing in the total number of links. They show that any symmetric network is pairwise stable. In our case stable networks can only be disjoint cliques.

In Morrill (2011) payoffs increase with own degree but decline in the degree of neighbors only. Pairwise stable networks are not limited to cliques but can also be d-regular graphs for suitable d.[7]

key is to show that a deviating node with smallest degree after this deviation, say d', cannot achieve a payoff exceeding $V(d')$, which is weakly less than $V(d^*)$.

As Jackson and Van den Nouweland (2005) point out, strong stability is a demanding solution concept. A situation wherein strongly stable networks exist is when all nodes can jointly achieve the maximal payoff they could achieve across all networks, as in Erol (2019). In our framework, nodes may not desire more links because they are strictly harmed by their neighbors having more links. That is, each node would strictly prefer to be at the center of a star network of appropriate order. But this is impossible. Thus, the incentives of nodes are misaligned and they cannot simultaneously achieve maximal payoffs. This conflict between nodes in terms of who gets a higher degree militates against the existence of strongly stable networks. Interestingly, this same force implies the deviators face the same conflict among themselves. The worst-off deviator gets exposed to high negative externalities by the better-off deviators. Therefore, under negative link-externalities, notions of farsightedness are not required to discipline the plausibility of deviations. It is remarkable that strongly stable networks exist in our framework because negative externalities are proportional to the number of edges whereas the benefits of links are proportional to degrees, and the former grows exponentially compared to the latter.

5 Positive externalities

Next we study the case when U is weakly decreasing in d and strictly increasing in e, namely *positive link-externalities.*

Proposition 2. *Under positive link-externalities, a stable network consists of either one non-trivial component and some node-components, or some link-components and some node-components. Call the former non-trivial stable networks and the latter trivial stable networks.*

Roughly stated, the main idea is that under positive externalities separate non-trivial components can be combined to boost link-externalities. This isn't the case if the cost of links are too high compared to the positive externalities so that nodes want to belong to trivially small components. A characterization of trivial stable networks is contained in Proposition 7 in the Appendix. We now focus on non-trivial stable networks.

In Goyal and Joshi (2006), when each agent's payoff increases in the total number of links, a stable network is either empty or consists of isolated nodes and a single clique. In our case, the single connected component need not be a clique.

To obtain a sharper characterization of stable networks in the positive externality case we impose more structure on U. If $U(d,e) < U(d+1,e+1)$ for all $d \leq e$, then, all pairs would add all the missing links. The unique stable network is the complete network, i.e., single clique. If $U(d,e) < U(d-1,e-1)$ for all $d \leq e$, the unique stable network is the empty network. To see why, recall that a node in a non-link components is called SC if it belongs to a cycle or is adjacent to a leaf . Every non-link component has an SC node (see Lemma 1 in the Appendix). But, an SC node i can cut a single link and increase its payoff to $U(d_i - 1, e_i - 1)$ with a unilateral deviation. Link-components also benefit from cutting their links. Such cases of global monotonicity yield uninteresting results. The problem is more interesting if there are non-trivial tradeoffs between d and e. These can be incorporated by constraining the 'cross-partials' of $U(d,e)$ with respect to d and e. Call $U(d,e)$ *strictly quasi-convex* if $U(d,e) < \max\{U(d-1,e-1), U(d+1,e+1)\}$ for all $1 \leq d \leq e$. See Section 6 for a setting where quasi-convexity would arise naturally.

To interpret quasi-convexity, suppose a node could add/delete an incident link without changing the connectivity of its component. Then, d and e both change by 1. Suppose also that in this case, the node's payoff is monotone with the addition of an *incident* link. Strict quasi-convexity asserts that adding an *incident* link is always beneficial in the case of negative externalities and always harmful in the case of positive externalities. Thus incident links act as a counterweight to the externalities imposed by distant links. Quasi-convexity differs from monotonicity of U in d in that e is not assumed fixed while changing d.

Theorem 2. *Under positive link-externalities and strict quasi-convexity, non-trivial stable networks are either a complete network or a star network.*

The complete network is stable iff $U(n-1, B(n)) \geq \max\{U(1, 1+B(n-1)), U(0,0)\}$*. A stable complete network is strongly stable.*[8]

[8]In fact, a stable complete network is strongly stable for any U with positive externalities even if U is not quasi-convex.

A star network is stable iff $U(1, n-1) \geq U(2, n)$ *and* $U(n-1, n-1) \geq U(0, 0)$. *A stable star network is strongly stable iff* $U(1, n-1) \geq U(n-1, B(n))$.

That the complete network is stable is unsurprising given positive link externalities. If the incremental benefit of increasing the number of links in a connected component is sufficiently large, one will want to add as many links as possible. It is less obvious that a star network will be stable. If the incremental cost of increasing one's degree, is large enough it will discourage nodes from adding links. However, in the absence of links, nodes will not enjoy the benefits of being in a connected component. A star network resolves this tension by making one node shoulder all the burden. We illustrate with an example. Consider the additive form $U(d, e) = e - cd$ where $c < 1$. It is strictly quasi-convex and $U(d + 1, e + 1) = U(e, d) + 1 - c > U(e, d)$. Thus, the incremental benefit of adding a link is positive. In this case, the unique stable network is the complete network which is also strongly stable.

Next, consider the multiplicative form $U(d, e) = \frac{e}{d+\epsilon}$ where $1 > \epsilon > 0$. It is strictly quasi-convex and $U(1, n-1) \geq U(n-1, B(n))$. The unique stable network is the star network. It is also strongly stable.

Strongly stable star networks are interesting because the center of the star is providing a public good by connecting all nodes to each other. The leaves enjoy higher payoffs than the center. Recall that in the case of negative link-externalities, all nodes would strictly prefer to be at the center of a star component, but this poses a conflict for the agents. Under positive link-externalities, each node strictly prefer to be a leaf of the star network, yet one node will occupy the center if the benefit from the number of links in the relevant connected component exceed the costs associated with an increase in degree. In this sense the center of the star is provider of a public good.

Star networks emerge as an equilibrium outcome in König et al. (2014) and Hiller (2017b). However, payoffs in these papers depend on the sum of discounted 'walk lengths' to other nodes and not the number of links in one's connected component.[9]

[9] Hiller (2017a) is related but not comparable. Node payoffs in this paper are function of incident links only and their 'sign'.

6 Examples, Stability, and Efficiency

In this section we give three examples to illustrate the difference between negative and positive and negative externalities, the different notions of stability and their connection to efficiency. The first two are chosen for their simplicity to make the opposing effects of d and e on payoffs as transparent as possible. The third, can be interpreted as model of collaboration with spillovers.

The first example illustrates the difference between stable and strongly stable networks in the case of negative externalities.

Proposition 3. *Take $U(d, e) = d - ce$ for some $c > 0$. Under the appropriate parity assumptions, the following holds. A network is stable if and only if it consists of disjoint cliques with orders $k_1 \leq k_2 \leq \ldots \leq k_s$ where $k_s \leq \frac{2}{c} + 1$ and $k_2 \geq \frac{1}{2}(\sqrt{\frac{8}{c} - 7} - 1)$. A network is strongly stable if and only if it consists of disjoint $(\frac{1}{c} + \frac{1}{2})$-cliques. A network is efficient if and only if it consists of disjoint $(\frac{1}{c} + \frac{1}{2})$-cliques*

All stable networks must consist of disjoint cliques according to Proposition 1. The conditions to preclude profitable deviations by pairs reduces to $k_s \leq \frac{2}{c} + 1$ and $k_2 \geq \frac{1}{2}(\sqrt{\frac{8}{c} - 7} - 1)$. So, in general, cliques in a stable network can be twice as large or *exponentially* smaller than the strongly stable cliques. This illustrates the weakness of stability. Nodes can get "stuck" at small cliques. When only pairs are allowed to add links in a feasible deviation, nodes from separate cliques can enjoy the benefit of having one extra link, but they expose themselves to large negative externalities due to existing links in the other node's existing clique. In contrast, strong stability allows for joint deviations. Two separate cliques can include all of their missing links and form a new and larger clique. Each deviator is still exposed to large negative externalities due to the pre-existing links in the other clique, but the benefit is larger since many links are being added at once for each deviator.

Stable networks need not be efficient in the sense of maximizing total payoff. Consider the instance described in Proposition 3. Observe that for a given $k < \frac{2}{c}$, it is efficient to have all links in the component. Also observe that it is inefficient to have any components with order larger than $\frac{2}{c}$. So the efficient network consists of disjoint cliques. Then efficiency requires maximizing the average payoff which coincides with strong stability.

Efficiency and strong stability don't coincide in general because strong stability implies Pareto efficiency, not efficiency. This can be seen in a simple example. Suppose that $n = 4$ and U satisfies

$$3U(3,4) + U(1,4) > U(1,1) > \max\{U(0,0), U(2,3), U(3,6)\}.$$

This is consistent with negative link-externalities. The unique strongly stable network is given by two links by Theorem 1. Yet, the network that consists of a triangle with the extra node attached to it, i.e. $E = \{\{1,2\},\{2,3\},\{3,1\},\{1,4\}\}$, gives higher total payoff than the strongly stable network.

By restricting the curvature of $U(d,e)$ one can ensure that strong stability will coincide with efficiency. To illustrate, suppose $U(d,e)$ is additive in d, as is the case in Proposition 3. Then, strong stability and efficiency coincide.

Proposition 4. *Consider negative link-externalities. Suppose that n is divisible by $d^* + 1$ and $U(d,e)$ is additive in d. Further, assume $U(2x,x)$ is increasing in x and $U(0,0) > 0$. Then, efficient networks and strongly stable networks coincide.*

$U(0,0) > 0$ can be seen as a normalization. If we drop the assumption that $U(2x,x)$ is increasing, then strongly stable networks are still "roughly" efficient. Each component must still be close to a clique in the sense that a component with k nodes cannot have less than $B(k-1)+1$ links in the efficient network. Then, the optimal number of nodes in each component is close to d^* if U is sufficiently "smooth."

Now we consider a positive link-externalities example that is the 'mirror image' of the one in Proposition 3.

Proposition 5. *Take $U(d,e) = e - cd$ where $c > 0$. For $c > \frac{n}{2}$, the unique stable network, the unique strongly stable network, and the unique efficient network is the empty network. For $\frac{n}{2} > c > 1$, the unique stable network is the empty network. The unique efficient network is the complete network. There are no strongly stable networks. For $c < 1$, the unique stable network, the unique strongly stable network, and the unique efficient network is the complete network.*

For $c < 1$ we have already determined that the unique non-trivial stable and strongly stable network is the complete network. Clearly, this is also the efficient network. For

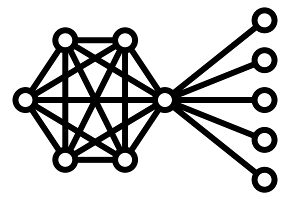


Figure 3: A pineapple graph, $P(11, 6)$

$1 < c$, the unique stable network is the empty network. But this is neither strongly stable nor efficient unless $c > \frac{n-2}{2}$ which is unlikely to be true for large networks.

Next, we consider a more interesting case of positive externalities. It generates a strongly stable network that is a pineapple graph, $P(n, k)$, found by appending $n-k$ nodes to one single node in a clique of order k. Figure 3 illustrates a pineapple graph.

To motivate the payoffs in this example, suppose each of n agents must decide how many collaborations or joint projects it should participate in. The more projects she is a part of, the more responsibilities she has. Thus, the marginal cost to agent i of exerting effort x_i will increase not just with x_i but also with the number of collaborations she is a part of. If d_i denotes the number of such collaborations, her cost of effort is $0.5x_i^2(d_i+\epsilon)$. However, being a part of a network of collaborators allows her to enjoy the benefits of spillovers (say from knowledge transfer) from projects she is not a part of as long as she is connected to, via a chain of collaborators, to participants in other projects. If e is the total number of projects in the connected component of which she is a member, her benefit from effort x_i is $\sqrt{e}x_i$. Thus, the spillovers complement her own efforts, but holding x_i fixed, they have diminishing returns. Holding e and d_i fixed, agent i will choose x_i to maximizes $\sqrt{e}x_i - 0.5x_i^2(d_i + \epsilon)$. The optimal effort choice will be $x_i = \frac{\sqrt{e}}{d_i+\epsilon}$. Agent i's payoff will be $\frac{0.5e}{d_i+\epsilon}$.

Proposition 6. *Take $U(d, e) = \frac{e}{d+\epsilon}$ for some arbitrarily small $\epsilon > 0$. Then, the star network is the unique strongly stable network, and the unique non-trivial stable network. The unique efficient network is given by $P(n, k_n)$ where $k_n \approx \frac{2n}{3}$.*

By Theorem 2, the unique non-trivial stable network is the star network, which is also strongly stable. For efficiency, note that in a given component $G' \preceq G$ with nodes N', the sum of payoffs is $\frac{1}{2}(\sum_{i\in N'} d_i)(\sum_{i\in N'} \frac{1}{d_i})$. The term $\sum_{i\in N'} \frac{1}{d_i}$ is called the inverse degree of the graph which is an object of interest in Chemical Graph Theory and we make use of an existing result from Hu et al. (2007) in the proof of efficiency.

7 Extensions

Here we summarize various ways in which our results can be extended. Theorems and proofs are available upon request.

One can account for payoffs that feature node externalities and non-monotonicity in d. Denote by f_i the number of nodes in the component of i and suppose that the payoff from a network is $U(d_i, e_i, f_i)$. If U is strictly decreasing in e_i and weakly decreasing in f_i with no restriction on how it depends on d_i, Proposition 1 and Theorem 1 (under appropriate parity assumptions in the spirit of $\frac{n}{d^*+1} \in \mathbb{Z}$) still hold.

If U is strictly increasing in e_i and weakly decreasing in f_i with no restriction on how it depends on d_i, Proposition 2 continues to hold. The characterization of stability in Theorem 2 fails, because there can be stable networks that consist of one disjoint clique and some node-components. This happens because of the weakness of $2-$stability in that it does not allow larger coalitions of nodes to add links at the same. The characterization of strongly stable networks is unaffected.

8 Conclusion

In the study of network externalities, links are often modeled to be only a transmitter of externalities. In various scenarios, the links themselves create externalities and not just transmit them. For example, collaborations with spillovers. Joint projects can fail and lead to cascading defaults of real or financial firms. We study a simple form of this type of externality and characterize stable and strongly stable networks. Negative externalities, such as contagion via links, lead to highly clustered network, in particular disjoint cliques. Positive externalities lead to either complete networks or star networks depending on the costs and benefits of links. When a star network is stable, the center of the star is a public good provider, transmitting positive externalities across the network although it is costly for the center to maintain a large number of links. Efficient networks tradeoff these two forces: a pineapple network emerges wherein a central node is still public good provider being connected to all other nodes, but there is also a large clique that further increases the extent of positive externalities which also benefits the center node.

References

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

Belleflamme, Paul and Francis Bloch, "Market sharing agreements and collusive networks," *International Economic Review*, 2004, *45* (2), 387–411.

Blume, Lawrence, David Easley, Jon Kleinberg, Robert Kleinberg, and Éva Tardos, "Network formation in the presence of contagious risk," *ACM Transactions on Economics and Computation (TEAC)*, 2013, *1* (2), 1–20.

Buechel, Berno and Tim Hellmann, "Under-connected and over-connected networks: the role of externalities in strategic network formation," *Review of Economic Design*, 2012, *16* (1), 71–87.

Dutta, Bhaskar and Suresh Mutuswami, "Stable networks," *Journal of economic theory*, 1997, *76* (2), 322–344.

Erol, Selman, "Network Hazard and Bailouts," *Working Paper*, 2019.

_ **and Rakesh Vohra**, "Network Formation and Systemic Risk," *European Economic Review*, 2022.

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

Goyal, Sanjeev and Sumit Joshi, "Unequal connections," *International Journal of Game Theory*, 2006, *34* (3), 319–349.

Hiller, Timo, "Friends and enemies: A model of signed network formation," *Theoretical Economics*, 2017, *12* (3), 1057–1087.

_ , "Peer effects in endogenous networks," *Games and Economic Behavior*, 2017, *105*, 349–367.

Hojman, Daniel A and Adam Szeidl, "Core and Periphery in Networks," *Journal of Economic Theory*, 2008, *139* (1), 295–309.

Hu, Yumei, Xueliang Li, Yongtang Shi, and Tianyi Xu, "Connected (n, m)-graphs with minimum and maximum zeroth-order general Randić index," *Discrete applied mathematics*, 2007, *155* (8), 1044–1054.

Jackson, Matthew O and Anne Van den Nouweland, "Strongly Stable Networks," *Games and Economic Behavior*, 2005, *51* (2), 420–444.

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

Kempe, David, Jon Kleinberg, and Éva Tardos, "Maximizing the Spread of Influence Through a Social Network," in "Proceedings of the Ninth Acm Sigkdd International Conference on Knowledge Discovery and Data Mining" ACM 2003, pp. 137–146.

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

Morrill, Thayer, "Network formation under negative degree-based externalities," *International Journal of Game Theory*, 2011, *40* (2), 367–385.

Neglia, Giovanni, Lo Giuseppe Presti, Honggang Zhang, and Don Towsley, "A Network Formation Game Approach to Study BitTorrent Tit-for-Tat," in "Network Control and Optimization," Vol. 4465 Springer Berlin Heidelberg 2007, pp. 13–22.

Vannetelbosch, Vincent and Ana Mauleon, "Network formation games," in "The Oxford Handbook of the Economics of Networks," Oxford University Press, 2016.

9 Proofs

Proof of Proposition 1

Suppose, for a contradiction, there exists a stable network with at least one component containing two non-neighbors i and j. As i and j are connected, each has at least one neighbor. Delete one link incident to i and another incident to j, and insert the link $\{i, j\}$. The degrees of i and j are unchanged but the number of links in their component strictly decreases. Hence, this is a profitable deviation by i and j. Contradiction.

Proof of Theorem 1

Step 1. First we prove that if G is strongly stable it must consist of disjoint cliques of the same size. Then, we show that a strongly stable network exists. Suppose G is strongly stable. Then, it is stable and so by Proposition 1 it consists of disjoint cliques. Now, the payoff to $i \in N$ is $V(d_i)$. If there exists a a disjoint clique with order $d+1 > d^*+1$ in the realized network, d^*+1 members would deviate by forming a disjoint (d^*+1)-clique and get $V(d^*) > V(d)$. If there exists various cliques with order less than d^*+1 but the sum of their orders is larger than equal to d^*+1, d^*+1 many nodes in these cliques would form a disjoint d^*+1 clique. Combining the two, by the parity assumption that d^*+1 divides n, the unique candidate for a strongly stable network is one that consists of disjoint (d^*+1)-cliques. It remains to show that a graph consisting of disjoint (d^*+1)-cliques is strongly stable.

Step 2. For any profitable deviation G' by N', define $\chi(G')$ to be the number of links between N' and N/N' in G'. Let the minimum of χ be attained at G^*. Denote by N^* the set of deviators in G^*, $E^* = E_{G^*}$, $N_1^* = \{i \in N^* : \exists j \notin N^* \, s.t. \, \{i,j\} \in E^*\}$, and $N_0^* = N^*/N_1^*$. Take the/a node $i^* \in N^*$ such that $d_{i^*,G^*} = \min\{d_{i,G^*} | i \in N^*\}$. Denote $d_* = d_{i^*,G^*}$. Denote $a = |N_{i^*,G^*} \cap (N \backslash N^*)|$ and $b = |N_{i^*,G^*} \cap N_0^*|$. Then $|N_{i^*,G^*} \cap N_1^*| = d_* - a - b$.

Step 3. Suppose that there exists $i \in N_0^*$ such that $d_i = d_*$. By definition of N_0^*, $N_{i,G^*} \subset N^*$. By definition of i^*, all $i \in N^*$ have $d_{i,G^*} \geq d_*$. Therefore, $e_{i,G^*} \geq d_i + \frac{1}{2} d_i d_* = B(d_*+1)$. Thus $u_{i,G^*} = U(d_*, e_{i,G^*}) \leq U(d_*, B(d_*+1)) = V(d_*) \leq V(d^*)$ which is a contradiction. Therefore, $d_i \geq d_*+1$ for all $i \in N_0^*$. In particular, $i^* \in N_1^*$.

Step 4. Consider any $i, j \in N_1^*$ that are connected in G^*. If $\{i, j\} \notin E^*$, consider the deviation G' by N^* such that $E_{G'} \backslash E^* = \{i, j\}$ and $E^* \backslash E_{G'} = \{\{i, i'\}, \{j, j'\}\}$ for some $i', j' \notin N^*$. Notice that G' is a profitable deviation since G^* is whereas $\chi(G') < \chi(G^*)$. This is a contradiction so $\{i, j\} \in E^*$. Therefore, all nodes in N_1^* that are connected in G^* are neighbors of each other, and so make up a clique (not disjoint clique). In particular, there are at least $B(d_* - a - b + 1)$ links in within N_1^* in the component of i^* in G^*.

Step 5. Take any $i \in N_0$ that is connected to i^* in G^*. By Step 3, $d_i \geq d_* + 1$. By Step 4, $N_{i,G^*} \cap N_1^* \subset \{i^*\} \cup N_{i^*,G^*}$. Thus $|N_{i,G^*} \cap N_1^*| \leq d_* - a - b + 1$. Then $|N_{i,G^*} \cap (N \backslash N_1^*)| \geq a + b$. By definition of N_0^*, $N_{i,G^*} \cap (N \backslash N^*) = \emptyset$ and so $|N_{i,G^*} \cap N_0^*| \geq a + b$.

Accordingly, if $b \geq 1$, then $N_0 \neq \emptyset$, so there exists $j \in N_0$ such that $|N_{j,G^*} \cap N_0^*| \geq a+b$. Notice that all nodes in $N_{j,G^*} \cap N_0^*$ are connected to i^* meaning that they all have at least $a+b$ neighbors in N_0^*. Therefore, there are at least $\frac{1}{2}(a+b+1)(a+b) = B(a+b+1)$ links within in N_0^* that are in the component of i^* in G^*.

Step 6. Take any $i \in N_{i^*,G^*} \cap N_1^*$. By definition of i^*, $d_{i,G^*} \geq d_*$. By Step 4, $N_{i,G^*} \cap N_1^* \subset \{i^*\} \cup N_{i^*,G^*} \backslash \{i\}$. Thus, $|N_{i,G^*} \cap N_1^*| \leq d_* - a - b$. Then $|N_{i,G^*} \cap (N \backslash N_1^*)| \geq a + b$. This holds for all $i \in N_{i^*,G^*} \cap N_1^*$ and $|N_{i^*,G^*} \cap N_1^*| = d_* - a - b$. Thus the number of links between N_1^* and $N \backslash N_1^*$ in the component of i^* in G^* is at least $(d_* - a - b + 1)(a + b)$.

Step 7. By feasibility of the deviation G^*, links between N^* and N/N^* in G^* exist in G as well. Then $N_{i^*,G^*} \cap (N \backslash N^*) \subset N_{i^*,G} \cap (N \backslash N^*)$. Since G consists of disjoint cliques, all nodes in $N_{i^*,G} \cap (N \backslash N^*)$ are neighbors of each other in G. Then all nodes in $N_{i^*,G^*} \cap (N \backslash N^*)$ are neighbors of each other in G. Since G^* is a deviation by N^*, by the feasibility of the deviation, all nodes in $N_{i^*,G^*} \cap (N \backslash N^*)$ are neighbors of each other in G^*. Therefore, there are at least $B(a)$ many links within N/N^* in the component of i^* in G^*.

Combining all steps we find $e_{i^*,G^*} \geq B(d_* - a - b + 1) + \mathbb{1}_{b \geq 1} B(a+b+1) + (d_* - a - b + 1)(a+b) + B(a) \geq B(d_* + 1)$. Accordingly, $u_{i,G^*} \leq U(d_*, B(d_* + 1)) = V(d_*) \leq V(d^*)$ which is a contradiction. There are no profitable deviations from G.

Proof of Proposition 2

Recall that a node in a non-link component is called SC if it lies on a cycle or is adjacent to a leaf.

Lemma 1. *In any network, every non-link component has at least one SC node. A non-link component has exactly one SC node iff it is a star component.*

Proof. Any non-link component, the neighbor of a leaf is obviously SC. If the non-link components has no leaves, then, every node must be on a cycle, thus every node is SC.

It is clear that a star network has only one SC node. Consider any non-link components that has exactly one SC node. If it contains a cycle, then, all nodes in the cycle are SC, which implies at least three nodes. So, the component must be a tree. As all parents of all leaves are SC nodes, there can be only one such parent. The only tree in which there is one common parent for all leaves is the star network. □

Step 1. Suppose that there exist two non-link components. By Lemma 1, each of them has at least one SC node. Pick one SC node from each, say i and j. Let i and j cut one link each with their component, and add the link $\{i, j\}$. This leaves their degrees unchanged. e_i becomes $e_i - 1 + 1 + (e_j - 1) \geq e_i + 1$. Same for j. Thus, this is a profitable deviation for i and j. This means there cannot be two non-link components.

Step 2. Denote $e^* = \max\{e \in \mathbb{Z} : U(1,1) \geq U(2,e)\}$. Take a non-link component with e links. Suppose that $e' < e^*$. Take one SC node from the component, say i. Note that SC nodes have degree at least 2. Cut all links of i but one. After this unilateral deviation, i's payoff is at least $U(1,1)$ whereas $U(1,1) \geq U(2,e^*) > U(2,e') \geq U(d_i,e')$. As this is a profitable deviation, we must have $e' \geq e^*$.

Step 3. Suppose that there exists a link-component and a non-link component. Then, consider the following deviation. Take one SC from the non-link component, say i, and one node from the link-component, say j. Cut one link of i and form a link between i and j. Node i's payoff becomes $U(d_i, e'+1) > U(d_i, e') \geq U(d_i, e^*)$ by Step 2. Node j's payoff becomes $U(2, e'+1) > U(1,1)$ by Step 2 and the definition of e^*. So this is a profitable deviation. Thus a non-link component and a link-component cannot coexist. Steps 1 and 3 together, complete the proof.

Characterization of trivial stable networks

Proposition 7. *Under negative link-externalities, trivial stable networks are characterized as follows. A network of multiple link-components and multiple node-components is stable if and only if* $U(1,1) = U(0,0) \geq U(2,3)$*. A network of multiple link-components and at most one node-component is stable if and only if* $U(1,1) \geq \max\{U(0,0), U(2,3)\}$*. A network of one link-component and multiple node-components is stable if and only if* $U(1,1) = U(0,0) \geq U(2,2)$*. A network of multiple node-components is stable if and only if* $U(0,0) \geq U(1,1)$*. (Straightforward.)*

Proof of Theorem 2

Lemma 2. *Under positive-link externalities, a connected network is stable iff it is a 1-stable network, and no pair of nodes can strictly improve their payoff by adding a link without cutting any links.*

Proof. "Only if" part is true by definition. For the "if" part, take a 1-stable network and suppose that a pair i and j has a profitable deviation to G'. Denote by S_i the set of links, except the link $\{i,j\}$, that i cuts in the deviation from G to G'. Define S_j analogously.

Suppose that the deviation to G' does not involve adding the link $\{i,j\}$, i.e. either $\{i,j\} \notin E_G, E_{G'}$ or $\{i,j\} \in E_G$. Then G'' where $E_{G''} = E_{G'} \cup S_j$ is a unilateral feasible deviation by i from G. Notice that $d_{i,G'} = d_{i,G''}$ and $e_{i,G'} \geq e_{i,G''}$. Thus $u_{i,G''} \geq u_{i,G'} > u_{i,G}$. This is a contradiction because G is $1-$stable.

Next, suppose that the deviation to G' involves adding the link $\{i,j\}$, i.e. $\{i,j\} \in E_{G'} \backslash E_G$. Since no pair of nodes can strictly improve their payoff by adding a link without cutting any links in G, we cannot have $S_i = S_j = \emptyset$. Then w.l.o.g. take $S_i \neq \emptyset$. If $|S_i| = 1$ then $u_{i,G} = u_{i,G'}$ which contradicts G' being a profitable deviation. So $|S_i| \geq 2$.

Since the network G is connected and $\{i,j\} \notin G$, there is a path from i to j in G that does not have the link $\{i,j\}$. Then, there exists a path from some $k \in \{i\} \cup \{i' : \{i,i'\} \in S_i\}$ to j in G that does not use any of the links in S_i.

If $k \neq i$, consider the unilateral deviation by i from G to G'_1 in which i cuts links $S_i \backslash \{i,k\}$. (Note that this is indeed a deviation because $|S_i \backslash \{i,k\}| \geq 1$ is implied by $|S_i| \geq 2$.) Clearly $d_{i,G'} = d_{i,G'_1}$. Since there is path from k to j in G that does not include any links in $S_i \backslash \{i,k\}$, k is connected to j in G'_1. Since $\{i,k\} \in E_{G'_1}$, i is connected to j in G'_1. Thus $e_{i,G'_1} \geq e_{i,G'}$. Therefore, $u_{i,G'_1} \geq u_{i,G'} > u_{i,G}$. Contradiction with $1-$stability.

If $k = i$, consider the unilateral deviation by i from G to G'_2 in which i cuts all links in S_i but one. Clearly $d_{i,G'} = d_{i,G'_2}$. Since there is a path from $k = i$ to j that does not involve any links in S_i, i is connected to j in G'_2. Hence $e_{i,G'_2} \geq e_{i,G'}$. Thus $u_{i,G'_2} \geq u_{i,G'} > u_{i,G}$ which contradicts $1-$stability. □

Lemma 3. *Under positive link-externalities and strict quasi-convexity, a star network with n' nodes is stable iff $U(n'-1, n'-1) \geq U(0,0)$ and $U(1, n'-1) \geq \max\{U(2,n'), U(0,0)\}$ and a complete network with n' nodes is stable iff $U(n'-1, B(n')) \geq \max\{U(1, 1+B(n'-1)), U(0,0)\}$.*

Proof. Call the unique non-leaf node in the star the center. According to Lemma 2, the relevant deviations are type 1– the center cuts some links, type 2– a leaf cuts its link with the center, and type 3– two leaves add a link between them. By quasi-convexity, the center does not have a profitable type 1 deviation iff $U(n'-1, n'-1) \geq U(0,0)$. There is no type 2 profitable deviation iff $U(1, n'-1) \geq U(0,0)$. There is no profitable type 3 deviation iff $U(1, n'-1) \geq U(2, n')$.

Suppose that G' is complete. According to Lemma 2, the only relevant deviations are those in which a node cuts some links. By quasi-convexity, such a deviation is not profitable iff $U(n'-1, B(n')) \geq \max\{U(1, 1+B(n'-1)), U(0,0)\}$ □

Lemma 4. *In any network, if there exists a non-central node, there exists at least two non-central nodes. (Straightforward.)*

Lemma 5. *In any network, if there is a central node, all WC nodes must have degree 1. (Straightforward.)*

Lemma 6. *Under positive link-externalities and strict quasi-convexity, a connected stable network is either a star network or a complete network.*

Proof. Consider a stable network with n' nodes and e links.

Step 1. An SC node i can cut one link and get $U(d_i - 1, e - 1)$ because the component it belongs to has exactly one less link. So all SC nodes i must have $U(d_i, e_i) \geq U(d_i - 1, e - 1)$.

Two NC nodes can add a link. Then for at least one of them, the deviation must be non-profitable. This holds for every NC pair. Thus, for all but at most one NC nodes i we must have $U(d_i, e) \geq U(d_i + 1, e + 1)$.

Notice, by strict quasi-convexity of U, we have

$$U(d-1, e-1) + U(d+1, e+1) > 2U(d, e).$$

Then, all but at most one of the NC nodes must be WC. Otherwise, they want to cut a link unilaterally.

Step 2. Suppose that there exists a C node and an NC node. Then, there exists at least two NC nodes by Lemma 4. Since all but one NC nodes are WC (by Step 1), there exists a WC node. We know that there is a C node and a WC node. Recall Lemma 5: whenever there is a C node, all WC nodes must have degree 1. If there were at least two C nodes, then, no node can have degree 1. Therefore, there is exactly one C node and all nodes but one are NC, which means all but at most two are WC by Step 1. Thus, all but at most two nodes have degree 1. In sum, there is one node with degree $n' - 1$, and $n' - 2$ nodes with degree 1. This implies that the remaining node also has degree 1. Therefore, the network is a star if there exists a C node and an NC node.

Step 3. If there is no NC node, by definition, the network is complete.

Suppose that there is no C node and so all nodes are NC. Then, by Step 1, all but at most one node are WC. Hence, there is at most one SC node. By Lemma 1, the network is a star network since there is a unique SC node. But then, it is a C node, since the network is a star which is a contradiction.

Combining Steps 3 and 4, we find that the network is either a star or a complete network. □

Lemma 7. *If a network is stable, then all of its subnetworks are stable. (Straightforward.)*

Stability part. Now we are ready to complete the proof of the stability part. Take a non-trivial stable network G and denote by G' the non-trivial component in G. G' is a connected network by definition. By Lemma 7, G' is stable. By Lemma 6, G' is either a star or complete. Denote by n' the order of G'. Suppose that $n' < n$. This means there are some node-components in G along with the non-trivial component G'.

If G' is a star, by Lemma 3, we have $U(n'-1, n'-1) \geq U(0,0)$ and $U(1, n'-1) \geq \max\{U(2, n'), U(0,0)\}$. Strict quasi-convexity implies $U(n', n') > U(n'-1, n'-1) \geq U(0,0)$. So the center of the star strictly improves its utility by forming a link with a node-component. By positive link-externalities, $U(1, n') > U(1, n'-1) \geq U(0,0)$, hence a node-component strictly improves its utility by forming a link with the center of the star. This contradicts the stability of G. Thus, $n' = n$ and so G is a star. By 3, it is stable iff $U(n-1, n-1) \geq U(0,0)$ and $U(1, n-1) \geq \max\{U(2, n), U(0,0)\}$

If G' is complete, by Lemma 3, we have $U(n'-1, B(n')) \geq \max\{U(1, 1+B(n'-1)), U(0,0)\}$. By strict quasi-convexity, we have $U(n', B(n')+1) > U(n'-1, B(n')) \geq U(1, 1+B(n'-1))$. So, any node in the the non-trivial component strictly improves its utility by forming a link with a node in a node-component. By positive link-externalities, $U(1, B(n')+1) \geq U(n', (B(n')+1) > U(n'-1, B(n')) \geq U(0,0)$. Therefore, any node in a node-component strictly improves its utility by forming a link with a node in the non-trivial component. This contradicts the stability of G. Thus $n' = n$ and so G is a complete network. By Lemma 3, it is stable iff $U(n-1, (B(n)) \geq \max\{U(1, 1+B(n-1)), U(0,0)\}$. This completes the proof of the stability part.

Strong stability part. Consider a stable star network G. Denote the center node by i_0.

Step 1. First, consider feasible deviations that involve i_0. Denote one such feasible deviation by G_1. If G_1 is a unilateral deviation by i_0, it cannot be profitable, by stability, so G_1 involves some leaves, say $D \subset N \backslash \{i_0\}$. Consider the feasible deviation G_2 by D such that $E_{G_2} \cap [D \cup \{i_0\}]^2 = E_{G_1} \cap [D \cup \{i_0\}]^2$ and $E_G \backslash [D \cup \{i_0\}]^2 \subset E_{G_2}$. This is a feasible deviation by D because 1) nodes in D are neighbors with i_0 in G, and so G_1 can not involve adding a non-existing link between i_0 and D, and 2) if some links between i_0 and D were cut in G_1, these links can still be cut by D. Clearly $d_{i,G_1} = d_{i,G_2}$ and $e_{i,G_2} \geq e_{i,G_1}$ for all $i \in D$. Therefore, if G_1 is a profitable deviation, then G_2 is also a profitable deviation. That is, if there is a profitable deviation that involves the center, then there is a profitable deviation that doesn't involve the center. So, there exists a profitable deviation iff there exists a profitable deviation by a subset of the leaves. Call the latter *leaf-deviations*.

Step 2. A deviation by some leaves that involves at least one leaf cutting the link with i_0 and not adding any links can't be profitable, by stability. So, in any profitable deviation by some leaves, all deviators who cut their link with i_0 must have added some links to other deviators.

Consider the/a leaf-deviation in which the minimum number of links between the deviators and the center are cut, if leaf-deviations exists. Denote this deviation by G_1. If there exists a deviator, say i_1, that cut its link with i_0, it must have added at least one link with another deviator, say i_2, by Step 1. Then, consider the feasible deviation G_2 by N_{G_1} given by $E_{G_2} \backslash E_{G_1} = \{i_0, i_1\}$ and $E_{G_1} \backslash E_{G_2} = \{i_1, i_2\}$. It is clear that G_2 is a profitable deviation since G_1 is. But G_2 involves cutting fewer links with i_0. This contradicts the definition of G_1. Thus, there are no deviators in G_1 that cut the link with i_0.

In sum, there exists a leaf-deviation iff there exists a leaf-deviation in which no deviators cut their link with i_0. Call the latter *connected-leaf-deviations*.

Step 3. Consider the/a connected-leaf-deviation in which the deviators have the largest number of edges in their component (which is the whole network) after the deviation, if any connected-leaf-deviation exists. Call this deviation G_1. Denote by e the number of links in G_1. Take one deviator, say i, that has the largest degree among all deviators after the deviation. It holds that $U(1, n-1) < U(d_{i,G_1}, e)$ since this is a profitable deviation. If there exists two deviators j, j' such that $d_{j,G_1}, d_{j',G_1} < d_i$, then consider the deviation G_2 by N_{G_1} given by $E_{G_2} = E_{G_1} \cup \{j, j'\}$. The number of links in the component of deviators is now $e+1$. So for $j'' \in N_1 \backslash \{j, j'\}$, $u_{j'',G_2} > u_{j'',G_1} > u_{j'',G}$. As for j and j', $u_{j/j',G_2} = U(d_{j/j',G_2}, e+1) = U(d_{j/j',G_1} + 1, e+1) \geq U(d_{i,G_1}, e+1) > U(d_{i,G_1}, e) > u_{i,G} = u_{j/j',G}$. Thus, G_2 is a connected-leaf-deviation, but G_2 has more links than G_1, which is a contradiction. Therefore, the degree of at most one deviator is altered after the deviation G_1.

Let all but one of the deviators in G_1 have degree d, and one other, say j''_0 has degree $d' \leq d$. Suppose that $d' \leq d-2$. We first claim that two deviators $j''_1, j''_2 \notin N_{j'',G_1}$ can't be neighbors of each other in G_1. Otherwise, the deviation G_2 by N_{G_1} given by $E_{G_2} \backslash E_{G_1} = \{\{j'', j''_1\}, \{j'', j''_2\}\}$ and $E_{G_1} \backslash E_{G_2} = \{\{j''_1, j''_2\}\}$ would be a connected-leaf-deviation that involves more links than G_1. Therefore, deviators in $N_{G_1} \backslash N_{j''_0,G_1}$ are not neighbors of each other. But then, they can't have larger degree than j'', which is a contradiction. So, we must have $d' \geq d-1$. (Note that $d' = d-1$ iff $(d-1)|N_{G_1}|$

is odd.) Then, notice that $e_{G_1} = n - 1 + \lfloor (0.5)(d-1)|N_{G_1}| \rfloor$.

Given these, suppose that $N_{G_1} \neq N$. The payoffs of all but one of the deviators are $U(d, e_{G_1}) > U(1, n-1)$. Take some $i' \notin N_{G_1}$. Consider a deviation G_3 by $N_{G_1} \cup \{i'\}$ wherein one deviator has degree $d-1$ or d, and all other deviators have degree d.[10] Then, the deviators have payoff at least $U(d, n-1+\lfloor (0.5)(d-1)(|N_{G_1}|+1) \rfloor) \geq U(d, e_{G_1}) > U(1, n-1)$. Thus G_3 is a connected-leaf-deviation which involves more links than G_1. This is a contradiction so we must have $N_{G_1} = N$.

So, there exists a connected-leaf-deviation iff there exists a d and connected-leaf-deviation by all leaves in which all but one of the deviators have degree d and the other has degree d or $d-1$ (depending on the parity of $(d-1)(n-1)$) after the deviation. Call the latter a *$d-$regular-connected-leaf-deviation.*

Step 4. After a $d-$regular-connected-leaf-deviation, the payoff of deviators is $U(d, n-1+\lfloor (0.5)(d-1)(n-1) \rfloor)$. The potential residual deviator has higher payoff so his incentive constraint does not bind. Therefore, combining all steps, we find that G is strongly stable iff $U(1, n-1) \geq U(d, n-1+\lfloor (0.5)(d-1)(n-1) \rfloor)$ for all d. By strict quasi-convexity, this holds iff $U(1, n-1) \geq U(n-1, B(n))$.

Next, consider a stable complete network. By $1-$stability, $U(d, B(n)-(n-1-d)) \leq U(n-1, B(n))$ for all d. In any feasible deviation by any subset of nodes after which a deviators degree becomes d, the number of links in the component is at most $B(n)-(n-1-d)$. So there are no profitable deviations.

Proof of Proposition 3

The part about stability follows after checking relevant deviations and some algebra. For efficiency, consider any network. Index components by t. Let e_t and k_t be the number of links and nodes in component t. The sum of payoffs is $\sum_t e_t(2-ck_t)$. Thus, it is inefficient to have a component with $k_t > \frac{2}{c}$. For $k_t < \frac{2}{c}$, it is efficient to have all links possible in the component. Thus the efficient network must consist of disjoint cliques. Then, the objective is to maximize the average payoff in a network that consists of disjoint cliques, which coincides, by definition, with strongly stable networks.

[10] For any k and $1 \leq d'' \leq k-1$, if kd'' is even, there exists a d''regular network on k nodes. If kd'' is odd, there exists a network where one node has degree $d''-1$ and all the others have degree d''. This is a consequence of the Erdos-Gallai Theorem.

Proof of Proposition 4

Take a network and index its components by t. Denote by e_t and k_t the number of links and nodes in the t^{th} component. By additivity, the sum of payoffs in component t is $U(2e_t, e_t)$ which depends only on the number of links. Thus, $e_t = B(k_t)$ is efficient. Therefore, efficiency requires maximizing $\sum_t k_t U(k_t - 1, B(k_t))$ subject to $\sum k_t = n$. Then the solution is $k_t = d^* + 1$ for all t as $d^* = \arg\max U(d, B(d+1))$.

If $U(2x, x)$ is not an increasing function of x, it is still efficient to make components "dense" in the sense that if $e_t < B(k_t)$, then less than k_t nodes can be used to form a component with e_t links. The residual nodes can be used elsewhere or alone to increase total payoff as $U(0, 0) > 0$. Thus, e_t must be between $B(k_t - 1)$ and $B(k_t)$. That is, the components must be "close" to a clique with perhaps at most $k_t - 2$ missing links. Therefore, efficiency roughly requires maximizing $\sum_t k_t U(k_t - 1, B(k_t))$

Proof of Proposition 5

For $c < 1$, every pair would add their missing link. So the unique stable network is the complete network. By Theorem 2, this is strongly stable. For $c > 1$, every pair would cut their existing link. So the unique stable network is the empty network. For $c < \frac{n}{2}$, this is not strongly stable because all agents would jointly deviate to the complete network. So there are no strongly stable networks. For $c > \frac{n}{2}$, consider a deviation from the empty network and take a component with k nodes and e links. The sum of payoffs is $ke - 2ce < 0$ meaning that at least one deviator has negative payoff. So this can not be a profitable deviation. So, the empty network is strongly stable. Finally, consider efficiency. The sum of payoffs in a component is $e(k - 2c)$. If $n < 2c$, this cannot be made positive, so the unique stable network is the empty network. If $n > 2c$, there can be components with more than $2c$ nodes in the efficient network. But combining them into the complete network clearly increases the total payoff.

Proof of Proposition 6

First, we show that an efficient network must be connected. If not, take two components. If either are node-components, take two neighbors i_1, i_2 from one component and two neighbors j_1, j_2 from the other. Cut links $\{i_1, i_2\}$ and $\{j_1, j_2\}$, add links $\{i_1, j_1\}$ and $\{i_2, j_2\}$. Then, the degrees of all nodes are unchanged. The sum of payoffs clearly increase as the components are combined. If one is a node-component and the other G' is not, take i from the node-component and j from the other, and add the link $\{i, j\}$. The change in the sum of payoffs is

$$\frac{1}{2}(\sum_{j'\in N'} d_{j'} + 2)(\sum_{j'\in N'} \frac{1}{d_{j'}} + 1 - \frac{1}{d_{j'}(d_{j'}+1)}) - \frac{1}{2}(\sum_{j'\in N'} d_{j'})(\sum_{j'\in N'} \frac{1}{d_{j'}}) > 0.$$

So, the network must be connected. Now we are interested in finding a degree sequence $\{d_i\}_{i=1}^n$ for a connected graph to maximize $(\sum_{i=1}^n d_i)\sum_{i=1}^n \frac{1}{d_i}$. For the moment, fix the number of links in the graph to be m and consider the following problem:

$$\max m \sum_{i=1}^n \frac{1}{d_i}$$

subject to

$$\sum_i d_i = m$$

$$d_1 \geq d_2 \geq \ldots \geq d_n$$

$$\sum_{i=1}^r d_i \leq r(r-1) + \sum_{i=r+1}^n \min(d_i, r) \ \forall r$$

$$n - 1 \geq d_i \geq 1 \ \forall i$$

This is the problem of finding among all m-edge graphs with n nodes, the one with the smallest $\sum_{i=1}^n \frac{1}{d_i}$. Theorem 3.5 in Hu et al. (2007) tells us that we can describe the optimal graph for the problem just stated using two parameters k and $a \in [0, k]$. The optimal graph consists of a single clique of size k, one of the vertices in the clique will have degree $n-1$; call it vertex 1. There will be $n-k-1$ degree one vertices adjacent to vertex 1 and a single vertex with degree a whose neighbors are in the clique and include vertex 1. We can describe the degree sequence of the graph in the following way.

1. $d_1 = n - 1$

2. $d_j = k$ for $2 \leq j \leq a$

3. $d_j = k - 1$ for $a + 1 \leq j \leq k$

4. $d_{k+1} = a$

5. $d_j = 1$ for $k+2 \leq j \leq n$

This means that $m = k(k-1)/2 + a + n - k + 1$. Observe that if $a = 0, 1, k$, this graph is a pineapple.

Now lets make m a variable. The goal will be to show that for an optimal choice of m we choose $a \in \{0, 1, k\}$. Our objective function is

$$f(a) = [k(k-1) + 2a + 2(n-k) + 2][\frac{1}{n-1} + \frac{a-1}{k} + \frac{k-a}{k-1} + \frac{1}{a} + n - k - 1]$$

Let $A = k(k-1) + 2(n-k) + 2$ and $B = \frac{1}{n-1} - \frac{1}{k} + \frac{k}{k-1} + n - k - 1 = \frac{1}{n} + \frac{1}{k(k-1)} + n - k$. Then, our objective function becomes:

$$(A + 2a)(B - \frac{a}{k(k-1)} + \frac{1}{a}).$$

Now

$$B - \frac{a}{k(k-1)} + \frac{1}{a} \geq B - \frac{1}{k-1} + \frac{1}{k}$$

$$= \frac{1}{n-1} - \frac{1}{k} + \frac{k}{k-1} + n - k - 1 - \frac{1}{k-1} + \frac{1}{k} = \frac{1}{n-1} + n - k$$

Hence $f(a) \geq (A + 2a)(\frac{1}{n-1} + n - k)$ which is increasing in a. We show that f is convex.

$$f'(a) = 2(B - \frac{a}{k(k-1)} + \frac{1}{a}) - (A + 2a)(\frac{1}{k(k-1)} + \frac{1}{a^2})$$

$$f''(a) = -\frac{2}{k(k-1)} - \frac{1}{a^2} - \frac{2}{k(k-1)} + \frac{2A}{a^3} + \frac{2}{a^2}$$

$$= -\frac{4}{k(k-1)} + \frac{2A}{a^3} + \frac{1}{a^2}$$

For $a \leq k-1$ we have that

$$f''(a) \geq -\frac{4}{k(k-1)} + \frac{2A}{k(k-1)^2} + \frac{1}{k(k-1)}$$

$$= -\frac{3}{k(k-1)} + 2\frac{k(k-1) + 2(n-k) + 2}{k(k-1)^2}$$

$$= -\frac{3}{k(k-1)} + \frac{2}{k-1} + 4\frac{(n-k)+1}{k(k-1)^2} \geq 0$$

As f is convex and lies above a linear function that is increasing, it follows that $f(a)$ cannot have a maximum at some value of a strictly between 1 and k.

Therefore, the efficient network must be a pineapple graph, say with m nodes in the clique. The remaining $n-m$ nodes are neighbors only with the center node. The value of this network is

$$\left(\binom{m}{2} + n - m\right)\left(\frac{n}{n-1} + n - m\right).$$

An analytical solution exists as this is a cubic in m but the expression is extremely long and hard to interpret. For large n let $m = n\alpha$ where $\alpha \in [0, 1]$. In the limit with respect to n,

$$\begin{aligned}
&\left(\binom{m}{2} + n - m\right)\left(\frac{n}{n-1} + n - m\right) \\
\approx&\left(\frac{n^2}{2}\alpha^2 + n(1-\alpha)\right)\left(\frac{n}{n-1} + n - n\alpha\right) \\
=&n^3\left(\frac{1}{2}\alpha^2 + \frac{1-\alpha}{n}\right)\left(\frac{1}{n-1} + 1 - \alpha\right) \\
\approx&n^3\left(\frac{1}{2}\alpha^2\right)(1-\alpha)
\end{aligned}$$

which is maximized for $\alpha = \frac{2}{3}$.