Enforcing Cooperation in Networked Societies

S. Nageeb Ali and David A. Miller*

November 20, 2013

Abstract

Which social norms and networks maximize cooperation in bilateral relationships? We study a network of players in which each link is a repeated bilateral partnership with two-sided moral hazard. The obstacle to community enforcement is that each player observes the behavior of her partners in their partnerships with her, but not how they behave in other partnerships. We introduce a new metric for the rate at which information diffuses in a network, which we call *viscosity*, and show that it provides an operational measure for how conducive a network is to cooperation. We prove that clique networks have the lowest viscosity and that the optimal equilibrium of the clique generates more cooperation and higher average utility than any other equilibrium of any other network. This result offers a strategic foundation for the perspective that tightly knit groups foster the most cooperation.

---

*Ali: University of California, San Diego. Miller: University of Michigan. We thank Joyee Deb, Matt Elliott, Itay Fainmesser, Ben Golub, Joe Harrington, Matt Jackson, Navin Kartik, Fahad Khalil, Markus Möbius, Kalle Moene, Paul Niehaus, Marcin Peski, Andres Santos, Joel Sobel, Max Stinchcombe, Adam Szeidl, Gaute Torsvik, Joel Watson, Alex Wolitzky, Muhamet Yildiz, Bill Zame, and numerous seminar and conference participants for helpful comments. We thank Aislinn Bohren for excellent research assistance. This research is financially supported by NSF grant SES–1127643. In addition, Ali gratefully acknowledges financial support from NSF grant NetSe–0905645, as well as financial support and hospitality from Harvard and Microsoft Research; Miller gratefully acknowledges financial support and hospitality from Microsoft Research, Yale, and the Cowles Foundation.

Contents

1 Introduction

Motivation: A large and recent literature emphasizes the role of networks in informal enforcement, describing how the network influences the extent to which a firm can trust its workers to exert effort or deliver goods, a trader can trust a partner to be honest, or a community can share risk. The longstanding view is that groups that are more "socially connected" or "tightly knit" have more powerful social incentives and generate the greatest cooperation.[1] The logic is intuitive: a tightly knit network diffuses information quickly and effectively so that if a player shirks on one of her partners, her other partners come to learn of it. Faster diffusion generates stronger social incentives, and thus each bilateral partnership can use this "social collateral" to enhance cooperation. Yet this powerful logic has lacked a strategic framework in which this mechanism is transparent. Our motivation is to develop strategic foundations for the *informational* role of networks in supporting cooperation in bilateral relationships.

We study a networked society in which each link is an ongoing bilateral partnership with two-sided moral hazard. Each partnership meets at exponentially distributed arrival times to play a prisoners' dilemma at stakes that it chooses. Higher stakes increase cooperation payoffs for the partnership but increase the temptation to shirk even faster—and thus require stronger incentives to preserve cooperation. Two partners can cooperate at modest stakes using bilateral grim trigger strategies between them, but involving other members of the community in their enforcement arrangement enables them to cooperate at higher stakes, using rewards and punishments in other relationships to enforce cooperation in their own. This is the essence of community enforcement.

Yet, community enforcement must overcome the obstacle of *private monitoring*: each player knows what transpires within her own partnerships, but does not observe how her partners interact with others, and so she lacks direct access to information that would en-

[1] Within sociology, see Festinger, Schachter, and Back (1948), Granovetter (1985), Coleman (1988), Raub and Weesie (1990), and Granovetter (2005) for a survey. Within economics, the impact of social connectedness and networks on economic behavior is highlighted by Glaeser, Laibson, and Sacerdote (2002), and features prominently in discussions of risk-sharing (Besley, Coate, and Loury 1993; Udry 1994; Bloch, Genicot, and Ray 2008; Ambrus, Möbius, and Szeidl 2010), and trading without enforcement (McMillan and Woodruff 1999; Dixit 2003).

able her to reward or punish her partners for what they have done in other relationships. The social network can mitigate this difficulty, since each link is not only a productive relationship, but also a conduit for information. For a fixed level of patience, which networks support more cooperation and why?

Preview of results: We offer a partial answer to this difficult question. Restrict attention to all networks in which each player has at most d partnerships, and to all Nash equilibria on these networks in which behavior is stationary on the equilibrium path (such equilibria may feature rich non-stationary punishments). Our main result is that the optimal network is composed of disjoint islands of $d+1$ players that are completely connected; i.e., *cliques of degree d*.

More precisely, we prove in Theorem 1 that there exists a weak perfect Bayesian equilibrium on the clique of degree d that attains a weakly higher utilitarian average payoff than any Nash equilibrium of any network in which no player has more than d partnerships. The comparison is sharper if we restrict attention to equilibria in which players always work on the equilibrium path, i.e. *mutual effort equilibria*: the equilibrium that we construct on the clique is unanimously preferred to any mutual effort Nash equilibrium of any network within this class. An implication of our result is that if each player can have degree no greater than d, then it is optimal to organize society into cliques of degree d (ignoring remainder issues).

We prove analogous results in which we incorporate an explicit cost of linking rather than a constraint on degrees in the network. The complete network is optimal if linking costs are linear or concave, and an optimally sized clique is Pareto efficient if linking costs are strictly convex. All of these network comparisons apply for every level of patience and frequency of interaction.

Our approach: At first glance, one may envision that these results involve a "two-stage" optimization process: for a fixed discount rate, find an optimal equilibrium for each network, and then optimize over the set of networks. Despite being conceptually straightforward, such an approach would be technically challenging at both stages. Given our monitoring assumptions, even finding an optimal equilibrium on a single asymmetric network for a fixed discount rate typically requires imposing functional form assumptions and resorting to numerical methods to solve a nonlinear convex optimization problem. Scaling to a large set of possible networks and every possible discount rate is simply infeasible.

We bypass these difficulties by directly connecting the speed of information diffusion with cooperation in Nash equilibria, and then comparing that speed across networks. Consider the incentives that Ann faces in deciding whether to work in her partnership with Bob in a particular network. The key question that she must ask herself is, were she to shirk, would she be able to take advantage of her other partners before they punished her? At the core of her incentives is the speed at which information propagates through this network from her partnership with Bob back to her other partnerships. Were Ann to interact with everyone in the network—as in the random matching environment studied by Kandori (1992) and Ellison (1994)—then a direct "contagion-infection" measure that describes the travel speed of infection to a random node (e.g., Golub and Jackson 2012) would be appropriate. But in our setting, Ann interacts only with those with whom she has partnerships. Thus her strategic calculations hinge on the number, length, and arrangement of paths within the network from Bob that *return* to her other partners. To account for these incentives, we derive a new network measure, *viscosity*, which measures the discounted probability that Ann will be able to take advantage of another one of her partners (say, Carol) before Carol can punish Ann.

Although it is intuitive that the speed of transmission should connect to incentives in any network and in any equilibrium of the repeated game on that network, we face the challenge that the one-shot deviation principle does not apply in our setting. Accordingly, it would be difficult to assess what are the equilibria for a fixed network, and any bounds that apply to them. We sidestep this challenge by restricting attention to equilibria that are stationary on the path of play and considering the simplest possible deviation: shirking forever on every partner. In any such Nash equilibrium, such a deviation cannot be strictly profitable. Our first key lemma (Lemma 1) formalizes this incentive constraint and shows that it can be used to generate a fundamental connection between viscosity and incentives.

The constraint expressed above is loose for many networks. Our second key lemma (Lemma 2) derives a setting for which it is tight: on a clique there exists an equilibrium in which the most profitable deviation is to shirk forever on every player, and moreover this is the best Nash equilibrium on this network. This equilibrium can be constructed so as to be not only a Nash equilibrium, but also a perfect Bayesian equilibrium. In particular, we implement it using *contagion strategies* (Kandori 1992; Ellison 1994) in which all players shirk on all their partners once they are off the equilibrium path. On a clique, for all parameters and without making functional form assumptions, these contagion equilibria exist; Pareto dominate all other mutual effort equilibria; and, if the stage game satisfies strategic complementarity, maximize the utilitarian average across equilibria.

Our third key lemma (Lemma 3) shows that the viscosity of any network is bounded below by the viscosity of the clique with the same maximal degree. Computing viscosity for each network is infeasible, so instead we construct a more fundamental coupling argument to uniquely map each path that propagates punishment on any arbitrary networks with a path on the corresponding clique that propagates it through (weakly) fewer links. The approach transparently mirrors the powerful intuition that the indirect paths between Bob and each of Ann's other partners on the clique are shorter than they could be on any other network with maximal degree d.

Our main result connects these lemmas to establish that there exists an equilibrium on the clique that outperforms all equilibria (that are stationary on the equilibrium path) on non-clique networks bounded by the same degree. We prove that the equilibrium we identify Pareto dominates all mutual effort equilibria on the non-clique networks by using the connection between speed and incentives uncovered in the earlier lemmas, and that it has higher utilitarian average than any non-mutual effort equilibrium (assuming strategic complementarity) by considering and constructing an "aggregated incentive constraint" satisfied by any Nash equilibrium (including those in which players may shirk on the path of play).

Broader contributions: Our results on community enforcement and cliques crystallize features of networks and norms that foster cooperation. In most settings, one may envision that the members of a partnership are the only ones privy to its details, or at least have the best knowledge of whether each has been cooperating. Even if these relationships lack payoff interlinkages, strategic interlinkages can be leveraged towards more cooperation. Our setting models how these strategic interlinkages and the formation of a new relationship benefits the community if it completes a cycle, and thereby serves as a conduit for information.

Our logic builds upon the foundations for community enforcement proposed by Kandori (1992) in his study of cooperation in an anonymous random matching environment. In applying this logic to our framework (repeated prisoners' dilemmas on networks with variable stakes), we emerge with several new conclusions. First, on a clique (or any complete graph), contagion equilibria generally exist and are the best equilibria within a large class for a fixed discount rate. Second, and more importantly, contagion equilibria can be shown to dominate equilibria of other networks with a (weakly) lower maximal degree. Thus, the usefulness of contagion as a tool for understanding the limits of community enforcement is not predicated on anonymity in a random matching environment, but can apply even when players have identities, and interact in an intricate network topology.

In developing these strategic foundations for cooperation in networks, our work offers a concrete measure for how conducive a network is to cooperation. The rich literature on networks has resorted to an array of network measures—including clustering coefficients, number of common partners, length of shortest path, differences in eigenvector centrality—many of which are motivated by questions of information diffusion but lack foundations in strategic play over time. Viscosity presents a step in the direction of identifying a global measure that accounts for how the structure of paths in a network between two partners influences their level of cooperation.

Finally, we hope that our model serves as a tractable framework to study community enforcement. We formulate a variable-stakes framework in which individuals select the level of cooperation. Apart from the inherent realism in many applications, this environment generates a convenient metric to compare equilibria and networks, and permits analytical comparisons of equilibria across these settings. Were we to adopt the standard approach that fixes stage game payoffs, we would have to indirectly compare networks by the intervals of discount factors for which cooperative equilibria exist.[2] We have found it convenient to use a similar framework to study the incentive compatibility of commu-

[2]Variable stakes have been used in prior work (Ghosh and Ray 1996; Kranton 1996; Watson 1999) but usually towards a different end: building cooperation over time helps screen out myopic players.

nication in our companion paper (Ali and Miller 2013), and we believe that a number of community enforcement questions can be tractably posed and answered in this setting.

Related literature: We build on ideas in both repeated games and networks. We have already discussed how we build on the idea of contagion first developed by Kandori (1992), Ellison (1994), and Harrington (1995). Other papers on community enforcement have also seen these papers to be a point of origin, but have taken different directions: Takahashi (2010) constructs folk theorems when players are anonymous but limited information about them is available exogenously, and Deb (2011) and Deb and González-Díaz (2011) construct a folk theorems when anonymous players engage in interactions that do not take the form of prisoners' dilemmas. Our work is orthogonal to these important results, and deriving a folk theorem in our setting is considerably less interesting: since each partnership in our model shares a common history, if two partners were arbitrarily patient they could simply use their bilateral relationship alone to sustain high payoffs, so every feasible and individually rational payoff in each bilateral relationship can be sustained. Thus, in our setting, the important issue is not in sustaining cooperation at the limits of patience but to augment bilateral relationships when players are not patient enough to achieve high payoffs on their own.

We now contrast our work with the recent literature on network-based cooperation. In principle, networks may serve two roles: they determine who can punish a player for shirking, and how quickly they can find out whom to punish. Most papers focus on the first of these roles, by studying environments in which monitoring is perfect and information diffusion has no role (Karlan, Möbius, Rosenblat, and Szeidl 2009; Jackson, Rodriguez-Barraquer, and Tan 2012). Our work complements these results by focusing on information propagation as the source of social collateral, in a model where the history of each partnership is privately observed by those partners.

The closest paper is Lippert and Spagnolo (2011), who study how network-based cooperation can pool incentive constraints when relationships that have slack in their bilateral incentive constraints can be used to subsidize other relationships, analogous to the logic of multimarket collusion (Bernheim and Whinston 1990). In their setting, community enforcement can help some but not all relationships. They describe the benefits of information propagation, formalizing the insight that indirect paths foster cooperation through the channel of information propagation. However, in their environment, where stakes are fixed and all pairs of partners meet simultaneously, the main focus is on whether a network has cycles or takes a tree-like structure. We establish that the insight of information propagation is considerably more general and powerful: community enforcement can benefit all relationships, and networks can be ranked by the speed with which information propagates back to a player's local neighborhood.

Another strand of the literature studies local interaction environments in which each player takes a single action that affects all of his neighbors, rather than interacting with each of them bilaterally. Haag and Lagunoff (2006) study optimal network design in this setting and find that cliques optimally separate impatient players from those who are more patient. Because a player's action is observed by all of her neighbors in a local interaction environment, the force is complementary to information propagation. Recently, Wolitzky (2012) uses contagion equilibria to support public good provision with private monitoring, and Nava and Piccione (2012) construct "temporary" contagion equilibria for local interaction games in which players are uncertain about the network structure. Both papers focus on synchronous local interaction rather than asynchronous bilateral interactions, and offer results and insights that are better suited towards public and community good applications. By contrast, in our model, each relationship is independent unless the players introduce strategic interdependence through their own actions, and players are potentially heterogeneous in the number and scale of their different relationships. Accordingly, our results speak to the role of network structures in informal contracting, trade, and bilateral risk-sharing arrangements.

2 Model

Network: A society is a finite set of players, $N \equiv \{1, \ldots, n\}$, connected by an undirected *network* G, which is a set of cardinality-2 subsets of N. The network is commonly known by the players, and is fixed throughout the game. We use $\{ij\}$ to indicate a link, and define $|G|$ to be the number of links in G. Much of our analysis concerns the incentives of each player on the link: accordingly, we use ij to signify "player i on link $\{ij\}$" as distinct from

ji ("player j on link $\{ij\}$").

In network G, player i's *neighborhood* N_i is the set of players to whom player i is linked: $N_i \equiv \{j \in N : \{ij\} \in G\}$. The cardinality of N_i is player i's *degree*, denoted by d_i. A *path* from player i to j is a sequence of nodes $i_1, \ldots, i_Z$ such that $\{i_z i_{z+1}\} \in G$ for each $z \in \{1, ..., Z-1\}$, $i_1 = i$, $i_Z = j$, and each node in the sequence is distinct. A *cycle* is a sequence of nodes $i_1, \ldots, i_{Z-1}, i_1$ such that $i_1, \ldots, i_{Z-1}$ is a path, and $\{i_1 i_{Z-1}\} \in G$. A *component* G' is a maximal connected subnetwork; i.e., if $\{ij\} \in G'$ then $\{k\ell\} \in G'$ if and only if there exists a path in G that contains both $\{ij\}$ and $\{k\ell\}$.

Social interactions: Time is continuous, and players discount payoffs realized at time t in $\mathbb{R}_+$ by the common discount rate $r > 0$. Each link in the network is governed by an independent Poisson recognition process with the common rate $\lambda > 0$. Whenever link $\{ij\}$ is recognized, players i and j engage in a two-stage interaction that occurs in that instant:

1. *Stake selection stage*: Players i and j simultaneously propose the *stakes* at which they should interact. Player i's proposal is $\hat{\phi}_{ij} \in \mathbb{R}_+$. The selected stakes of the $\{ij\}$ relationship, ϕ_{ij}, are a function of the proposals $\mathcal{F} : \mathbb{R}^2_+ \to \mathbb{R}_+$ with the property that for every $\phi \in \mathbb{R}_+$, $\mathcal{F}(\phi, \phi) = \phi$. We use $\phi \in \mathbb{R}_+$ as a generic stakes parameter when the identities of the players along the link are unimportant.
2. *Action stage*: Each player simultaneously chooses an action from $A \equiv \{\text{work}, \text{shirk}\}$. Their stakes determine the payoffs; higher stakes increase the payoffs from mutual effort but strengthen the temptation to shirk. Specifically, given stakes $\phi = \min\{\hat{\phi}_{ij}, \hat{\phi}_{ji}\}$ they face the prisoners' dilemma in Figure 1.

		Player j	
		Work	Shirk
Player	Work	ϕ, ϕ	$-V(\phi), T(\phi)$
i	Shirk	$T(\phi), -V(\phi)$	$0, 0$

FIGURE 1. The prisoners' dilemma of stakes ϕ

The "temptation reward" T and the "victim's penalty" V are smooth functions satisfying $T(0) = V(0) = 0$, as well as $V(\phi) > 0$ and $T(\phi) > \phi$ for all $\phi > 0$. Thus if the stakes are positive, shirking is the strictly dominant action in the stage game for each player. Throughout the paper we assume that the temptation reward is increasing in the following manner:

Assumption 1 (Increasing Temptation)**.** *T is strictly increasing and strictly convex, with $T'(0) = 1$ and $\lim_{\phi\to\infty} T'(\phi) = \infty$.*

The important implication from Assumption 1 is that $T(\phi)/\phi$—the ratio of the payoffs from shirking vs. working while one's partner works—is close to 1 at low stakes but increases without bound as the stakes increase. As a consequence, the players require proportionally stronger incentives to work at higher stakes. We use the specification $T(\phi) = \phi + \phi^2$ in examples. Allowing players to set the stakes of their relationships is a key feature of our framework, on which we comment in Section 6.

For some of our results, we restrict attention to prisoners' dilemmas in which the incremental gain from working in the stage game is higher when one's partner works: this is a condition of supermodularity on the stage game, also referred to as strategic complementarity.

Definition 1. *The stage game satisfies **strategic complementarity** if $V(\phi) > T(\phi) - \phi$ for all $\phi > 0$.*

Monitoring and equilibrium: Monitoring is pairwise: as play unfolds, each player observes only what transpires along his own links, and observes neither the meeting times nor the behavior along any other links. Each meeting between two players is an *interaction*, characterized by the link $\{ij\}$ that was recognized, the time t at which it was recognized, the stakes $(\hat{\phi}_{ij}, \hat{\phi}_{ji})$ that players i and j proposed in the stake selection stage, and the actions (a_{ij}, a_{ji}) that they chose in the action stage. For a player i, when one of his links is recognized, his *private history* h_i^t is an ordered list of all his interactions up to (but not including) time t, along with the identity of the partner he is interacting with at time t.[3]

[3] Since player i acts only when meeting a partner, there is no need to define his private history at any other times. Also, since at most one link is ever recognized at a time (almost surely), we do not define histories for simultaneous meetings.

We denote the set of player i's private histories ending with a recognition of link $\{ij\}$ at time t as H_{ij}^t, and write $H_{ij} \equiv \bigcup_{t\in[0,\infty)} H_{ij}^t$, and $H_i \equiv \bigcup_{j\in N_i} H_{ij}$.

A (behavior) strategy for player i is a function $\sigma_i = (\sigma_i^{\mathrm{S}}, \sigma_i^{\mathrm{A}})$ such that $\sigma_i^{\mathrm{S}} : H_i \to \Delta[0,\infty)$ is his stake-selection strategy and $\sigma_i^{\mathrm{A}} : H_i \times \mathbb{R}_+^2 \to \Delta A$ is his action strategy. We restrict attention to equilibria that are *stationary* on-path: along each link, the partners' choices lead to the same distribution of stakes and actions at every equilibrium history.

Definition 2. *A strategy profile σ is* ***stationary*** *if for every $\{ij\} \in G$*

1. *there exists $\overline{\sigma}_{ij}^{\mathrm{S}} \in \Delta[0,\infty)$ such that $\sigma_i^{\mathrm{S}}(h) = \overline{\sigma}_{ij}^{\mathrm{S}}$ for every equilibrium path history h in H_{ij};*
2. *there exists $\overline{\sigma}_{ij}^{\mathrm{A}} : \mathbb{R}_+^2 \to \Delta A$ such that $\sigma_i^{\mathrm{A}}(h, \hat{\phi}_{ij}, \hat{\phi}_{ji}) = \overline{\sigma}_{ij}^{\mathrm{A}}(\hat{\phi}_{ij}, \hat{\phi}_{ji})$ for every equilibrium path history h in H_{ij} and proposals $(\hat{\phi}_{ij}, \hat{\phi}_{ji})$ in $\mathrm{Supp}(\overline{\sigma}_{ij}^{\mathrm{S}}) \times \mathrm{Supp}(\overline{\sigma}_{ji}^{\mathrm{S}})$.*

We study Nash equilibria in which behavior is stationary on the path of play, and henceforth, we refer to these as equilibria. We describe the importance of this restriction in Section 6 after presenting our results. Note that the set of feasible deviations for a player is unrestricted.

For some of our efficiency results, it is useful to distinguish a particular class of equilibria that is often focal in applications: those in which players work on the equilibrium path.

Definition 3. *A stationary strategy profile σ is a* ***mutual effort profile*** *if $\overline{\sigma}_{ij}^{\mathrm{A}}(\hat{\phi}_{ij}, \hat{\phi}_{ji})$ assigns probability* 1 *to* work *for all $(\hat{\phi}_{ij}, \hat{\phi}_{ji})$ in $\mathrm{Supp}(\overline{\sigma}_{ij}^{\mathrm{S}}) \times \mathrm{Supp}(\overline{\sigma}_{ji}^{\mathrm{S}})$.*

A strategy profile σ *Pareto dominates* another strategy profile $\tilde{\sigma}$ if no player is worse off with σ and at least one player is strictly better off. The *utilitarian value* of a strategy profile is the average of players' expected payoffs that it delivers on the path of play.

3 An Example

We begin with an example that highlights the essence of our approach. Consider a society in which each of Ann, Bob, and Carol is connected to the other two players. Suppose that $T(\phi) = \phi + \phi^2$, and consider equilibria in which all pairs coordinate on the same stakes at every on-path history.

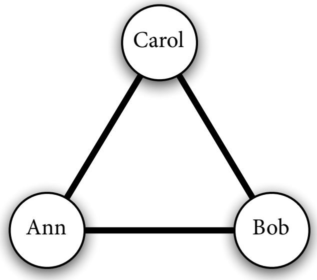

FIGURE 2. A triangle network

Bilateral enforcement: Consider the benchmark of *bilateral strategies*, in which behavior is strategically independent across links. Effectively, each pair plays an infinitely repeated prisoners' dilemma in isolation. Nash reversion is an optimal punishment in this class of strategies: consider strategies in which both players in a partnership work if and only if neither of them has ever deviated. Under these strategies, when Ann meets Bob, her incentive constraint to work along the equilibrium path is

$$\phi + \phi^2 \leq \phi + \int_0^\infty e^{-rt} \lambda \phi \, dt.$$

The highest stakes at which working is incentive compatible is λ/r.

Instantaneous public monitoring: In contrast to bilateral enforcement, suppose that everyone in society observes all the meetings, stakes announcements, and actions along every link in real time. Then if Ann shirks on Bob, it immediately becomes common knowledge among Ann, Bob, and Carol that continuation play is off the equilibrium path. In this alternative environment, consider an equilibrium in which once anyone shirks, everyone subsequently shirks perpetually. Ann's incentive constraint when she meets Bob along the equilibrium path is:

$$\phi + \phi^2 \leq \phi + 2 \int_0^\infty e^{-rt} \lambda \phi \, dt.$$

The highest stakes at which working is incentive compatible are $2\lambda/r$, doubling what is attainable under bilateral enforcement. Ann is willing to cooperate with Bob at higher stakes because of the immediate punishment that she receives from Carol if she shirks on Bob. This benchmark is infeasible in our environment because each player observes only the activity along his or her own links, so Carol cannot instantaneously learn that Ann should be punished.

Contagion strategies: In *contagion strategies*, a player works if all of his partners have always worked in the past; otherwise he shirks. If Ann shirks on Bob, Bob will shirk on Carol at their next interaction, and from then on Carol will shirk on both Ann and Bob. Ann's only chance for further gain is to meet Carol before Carol becomes "infected." According to her strategy, Ann should then shirk in her next interaction with Carol, so her cooperation phase incentive constraint is

$$\phi + \phi^2 + \int_0^\infty e^{-rt} e^{-\lambda t} \lambda e^{-\lambda t} (\phi + \phi^2)\, dt \le \phi + 2 \int_0^\infty e^{-rt} \lambda \phi\, dt.$$

Here, $e^{-\lambda t}\lambda$ is the density of Ann's first meeting with Carol, and $e^{-\lambda t}$ is the probability that at that first meeting Carol will not yet have met Bob. The highest stakes at which working is incentive compatible are $\left(\frac{r+4\lambda}{r+3\lambda}\right)\frac{\lambda}{r}$, strictly greater than what is attainable under bilateral enforcement.

To verify that contagion is sequentially rational, we must show that each of Ann and Bob wish to shirk on Carol after Ann shirks on Bob. Consider the interaction in which Bob meets Carol and is unsure if Carol has already been "infected."[4] If Carol hasn't been infected, an upper-bound for Bob's payoff from working forever with Carol is $\phi + \frac{\lambda}{r}\phi$, which is outweighed by the immediate payoff of $\phi + \phi^2$ from shirking today. Therefore, working forever cannot be sequentially rational for Bob once Ann has shirked; discounting guarantees that Bob is then best off by shirking immediately on Carol rather than delaying it. The intuition is simple: once Ann has shirked on Bob and continues to do so in the future, her presence is no longer a carrot-and-stick that can be used to encourage Bob to cooperate with Carol. The same logic, of course, applies to Ann when she meets Carol after having shirked on Bob.[5]

[4] If Ann was the first to shirk, she may have shirked on Carol as well as Bob. Another possibility is that Carol was the first to shirk, and Ann was spreading the contagion to Bob.

[5] Notice that there is an additional force: Bob believes that even if he works with Carol, she will be

The transmission of information through the network, as crystallized by contagion, exploits the strategic interdependence of networked relationships. When Ann cheats Bob, she forfeits the opportunity to also cheat Carol if Bob should meet Carol first. This uncertainty dampens her motive to shirk and thereby enables her to cooperate at higher stakes. The diffusion of information from Bob to Carol permits the relationship between Ann and Carol to become "social collateral" for the relationship between Ann and Bob.

This force emerges from the combination of asynchronous interaction and cycles in the network. Were all behavior constrained to instead be synchronous, Ann could simultaneously shirk on Bob and Carol, and so analogous to the multimarket logic of Bernheim and Whinston (1990), no Nash equilibrium could support anything more than the stakes of bilateral grim trigger. Similarly, if the network were a tree in which Ann was linked to both Bob and Carol (so Bob and Carol were not linked to each other), then no equilibrium could support mutual effort at stakes greater than under bilateral enforcement—shirking on Bob would not cause Carol to suffer any consequences in her relationship with Ann. A network must have cycles to transmit punishments.

Triangle vs. circles: Shorter cycles transmit punishment more quickly. Consider a society comprising $n \geq 3$ individuals connected in a circle. Under contagion strategies, after Ann shirks on Bob, Bob will shirk on Carol, Carol will shirk on Dante, and the contagion will spread around the circle. Ann's only chance to capture another temptation reward is to meet her other neighbor before he becomes contagious. We show that Ann's incentive constraint is

$$\phi + \phi^2 + (\phi + \phi^2) \sum_{z=1}^{n-2} \left(\frac{\lambda}{r + 2\lambda} \right)^z \leq \phi + 2 \int_0^\infty e^{-rt} \lambda \phi dt.$$

Evidently, Ann's other neighbor is less likely to be contagious for greater n, and so the maximal stakes on each link are attained for $n = 3$. At the same time, each player has the same degree in the circle as in the triangle. Therefore, if n is divisible by 3 then rearranging the network into $\frac{n}{3}$ triangles induces shorter indirect paths, and hence faster

eventually infected by Ann. However, this force adds slack to off-path incentives: it suffices for Bob to believe that Ann will no longer work with him for him to have a strict incentive to shirk on Carol.

punishments, larger stakes, and higher payoffs. This force pushes optimal network design towards cliques.

4 Main Result

Our main result compares Nash equilibria across networks (recalling that we restrict attention to equilibria that are stationary on the path of play). We call a completely connected network in which there are $d+1$ players the *clique of degree d*; we denote such a network by $\overline{G}(d)$. Our main result shows that the clique of degree d is optimal, in a precise sense, among networks with degree bounded by d.

Theorem 1. *For every interaction rate λ and discount rate r, there exists a symmetric perfect Bayesian equilibrium on the clique of degree d yielding each player a payoff of $\overline{u}(d)$, such that for any network G in which no player has more than d partnerships:*

(a) Every mutual effort Nash equilibrium on network G yields each player a payoff no greater than $\overline{u}(d)$.

(b) If the stage game satisfies strategic complementarity, then every Nash equilibrium on network G has utilitarian value no greater than $\overline{u}(d)$.

Intuitively, cliques maximize the level of cooperation because they maximize the rate at which information about a player's defection diffuses to his neighbors. We show that on any clique there exists a "binding contagion" equilibrium that exploits the diffusion of information to its fullest extent, while satisfying sequential rationality off the equilibrium path. Hence our result applies even when limiting attention to community enforcement mechanisms that are sequentially rational.

If we restrict attention to mutual effort equilibria on arbitrary networks, then a clique of degree d "Pareto dominates" every other network in which degrees are bounded by d: each player in the clique is at least as well off, and some are strictly better off.[6] If we allow

[6] This comparison ignores any remainder that might arise if the population of players is not divisible by $d+1$. Formally, we could make Pareto comparisons between two networks by replicating them both to form two "replica networks," each containing the least common multiple of the original populations.

for equilibria in which players shirk on the equilibrium path, then non-clique networks are not necessarily Pareto dominated. Nonetheless, if the game satisfies strategic complementarity then the binding contagion equilibrium on the clique is utilitarian optimal.

A major challenge in ranking networks is that it is difficult to know which equilibria exist and are efficient on any given network (particularly asymmetric networks), for any given interaction rate, and for any given discount rate. Our argument uses three preliminary lemmas to sidestep this challenge. Lemma 1 connects incentives in a Nash equilibrium to the speed of information transmission on an arbitrary network. Every equilibrium must satisfy the constraint that no player should prefer to shirk forever on all his partners. This incentive constraint features a key coefficient that we term *viscosity*, because its reciprocal describes the maximal discounted rate at which information can diffuse. Viscosity depends only on the network, the interaction rate, and the discount rate, but not on the details of any particular equilibrium. Lemma 2 offers a tighter characterization for cliques, producing a closed form expression for viscosity and constructing a binding contagion equilibrium. Lemma 3 uses a coupling argument to show that an arbitrary network with maximal degree d has higher viscosity than the clique of degree d. We then use these Lemmas to prove both parts of Theorem 1.

4.1 Three Key Lemmas

Connecting viscosity to Nash equilibria across networks: We begin by describing on-path behavior in a pure strategy mutual effort strategy profile. On the path of play, whenever link $\{ij\}$ is recognized partners i and j always select stakes ϕ_{ij} and work. Thus on-path behavior for each player i is summarized by an *individual stakes profile*: $\Phi_i = (\phi_{ij})_{j \in N_i}$, which is a profile of stakes in all of player i's relationships. The *collective stakes profile*, $\Phi = (\Phi_i)_{i \in N}$, is an element of the subset of $\mathbb{R}_+^{2|G|}$ in which $\phi_{ij} = \phi_{ji}$ for all $\{ij\} \in G$.

When link $\{ij\}$ is recognized at time t, if player i has never observed any deviation then her payoff from following her strategy is $\phi_{ij} + \frac{\lambda}{r} \sum_{k \in N_i} \phi_{ik}$. Now consider her gain from deviating. Since no one other than player j observes her deviation, it is unclear what player i's most profitable deviation might be, nor does it suffice to consider one-shot deviations. Instead, we study the following simple deviation: with each partner k in

N_i, player i proposes ϕ_{ik} and then shirks. What is the worst that could happen to player i if she deviates in this way?

Suppose that player i's first deviation from on-path play is to shirk on partner j at time t. Then, whenever player j meets another player j' after time t, suppose the equilibrium calls for player j to "communicate" to j' the bad news that a deviation has occurred. Such communication could either be encoded through stakes selection on link $\{jj'\}$, or player j shirking on partner j' as in contagion (Kandori 1992; Ellison 1994); at this stage of the argument, it does not matter how bad news spreads. What matters is that player j conveys the information about the deviation to each of his partners when he meets them, they convey the same to their partners, and so on. No process could spread bad news faster.

Now suppose that if link $\{ik\}$ is recognized at time $t+\tau$, player k works only if he has not heard the bad news about player i's deviations; otherwise he shirks. Given the contagious process by which bad news spreads, let $x_{ijk}(\tau)$ be the probability that, immediately after her first deviation on link $\{ij\}$ at time t, player i assigns to player k being willing to work on link $\{ik\}$ at time $t+\tau$. Now we can write player i's payoff from this deviation as

$$T(\phi_{ij}) + \sum_{k \in N_i \setminus \{j\}} T(\phi_{ik}) \int_0^\infty e^{-r\tau} \lambda\, x_{ijk}(\tau)\, d\tau.$$

The first term is what player i earns immediately by shirking on player j, and the second term is a lower bound for what she expects to earn in the future by shirking on her other neighbors. We combine the effects of x_{ijk}, λ, and r on the left hand side into a single term, the *ijk viscosity factor* $X_{ijk} \equiv \int_0^\infty e^{-r\tau} \lambda x_{ijk}(\tau)\, d\tau$. Since a player's payoff from shirking on every partner cannot exceed her cooperation payoff, an individual stakes profile is compatible with a pure strategy mutual effort equilibrium only if

$$T(\phi_{ij}) + \sum_{k \in N_i \setminus \{j\}} T(\phi_{ik}) X_{ijk} \leq \phi_{ij} + \frac{\lambda}{r} \sum_{k \in N_i} \phi_{ik}\,. \qquad (\mathrm{IC}_{ij}^{\mathrm{Coop}})$$

Our first result generalizes this insight to all pure and mixed Nash equilibria. Since we restrict attention to equilibria that are stationary on the path of play, let μ_{ij}^{S} be the distribution of equilibrium path stakes on link $\{ij\}$, $p_{ij}^{\mathrm{ww}}(\phi)$ be the on-path probability of mutual effort when stakes ϕ are realized, $p^{\mathrm{ws}}_{ij}(\phi)$ be the on path probability that player i works while player j shirks, and $p^{\mathrm{ss}}_{ij}(\phi)$ be the on-path probability of mutual shirking. Player i's expected equilibrium stage game payoff from link $\{ij\}$ being recognized is

$$u_{ij} \equiv \int_0^\infty \left(p^{\mathrm{ww}}_{ij}(\phi)\phi + p^{\mathrm{ws}}_{ji}(\phi)T(\phi) - p^{\mathrm{ws}}_{ij}(\phi)V(\phi)\right) d\mu^{\mathrm{S}}_{ij}.$$

Consider player i's payoff from following the stake proposal strategy but deviating at the action stage to shirking regardless of the realized stakes:[7]

$$w_{ij} \equiv \int_0^\infty T(\phi)\left(p^{\mathrm{ww}}_{ij}(\phi) + p^{\mathrm{ws}}_{ji}(\phi)\right) d\mu^{\mathrm{S}}_{ij}.$$

Our first key lemma argues that a counterpart of $\mathrm{IC}^{\mathrm{Coop}}_{ij}$ holds: in every equilibrium each player i must prefer to follow her equilibrium strategy than to shirk on all her partners forever.

Lemma 1. *For every Nash equilibrium σ, every player i in N, and every neighbor j in N_i,*

$$w_{ij} + \sum_{k\in N_i\setminus\{j\}} w_{ik}X_{ijk} \le u_{ij} + \frac{\lambda}{r}\sum_{k\in N_i} u_{ik}. \tag{1}$$

Proof. Consider an equilibrium path history in which players i and j meet at time s. The strategy profile σ is a Nash equilibrium only if the expected payoff from following the equilibrium strategy, denoted on the RHS, is at least that of every deviation. We argue that the LHS is a lower bound for player i's payoff when following the stake proposal strategy and deviating at the action stage to shirking on each partner. The payoff w_{ij} is obtained immediately. The probability with which player k observes off-path behavior at time $s + t$ is no greater than $1 - x_{ijk}(t)$, and so, a lower bound on the payoff that player i obtains from link $\{ik\}$ is $w_{ik}X_{ijk}$, leading to the constraint in (1). □

Contagion and viscosity on cliques: Lemma 1 states a necessary condition for Nash equilibrium on any network. For a clique of degree d, we can show that this condition

[7] With mixed stake proposal strategies, player i's best multi-shot deviation may also involve deviating in stake proposal stages. But for Lemma 1 we need only a necessary condition, not a sufficient condition, for equilibrium, so it suffices to consider deviations only in action stages.

binds for a particular perfect Bayesian equilibrium—contagion—and characterize viscosity in closed form. Later we will show that is the optimal "binding contagion" equilibrium identified in Theorem 1.

Lemma 2. *For the clique of degree d, and every pair of links $\{ij\}$ and $\{ik\}$,*

$$X_{ijk} = \overline{X}(d) \equiv \frac{1}{d-1}\sum_{m'=2}^{d}\left(\frac{1}{m'}\prod_{m=2}^{m'}\frac{\lambda m(d-m+1)}{r+\lambda m(d-m+1)}\right). \qquad \text{(Clique Viscosity)}$$

In the clique of degree d, there exists a pure strategy mutual effort weak perfect Bayesian equilibrium in which the equilibrium path stakes on each link, $\overline{\phi}(d)$, solve

$$\frac{T(\phi)}{\phi} = \frac{1+d\frac{\lambda}{r}}{1+(d-1)\overline{X}(d)}. \tag{2}$$

and each player's expected equilibrium path payoff is $\overline{u}(d) \equiv d\frac{\lambda}{r}\overline{\phi}(d)$.

Proof. We derive $(d-1)X_{ijk}$ by considering the expected payoff from the following "contagion process." Players i and j are contagious at time 0, and at each stage, whenever the link between a contagious player and an uninfected player is recognized, the latter is infected and becomes contagious. If player i meets an uninfected partner, she obtains 1, and otherwise she obtains 0. Recurse on the number of contagious neighbors to obtain player i's expected payoff from this process: suppose that there are $m-1$ of player i's neighbors that are currently contagious. Then there are $m(d+1-m)$ links by which the contagion spreads to an uninfected neighbor of player i; of these $d+1-m$ are links of player i. Therefore, $(d-1)\overline{X}(d)$ equals $\chi(2)$, where, for $m \geq 2$,

$$\begin{aligned}\chi(m) &= \int_0^\infty e^{-rt}e^{-\lambda m(d-m+1)t}\lambda m(d-m+1)\left(\frac{1}{m}+\chi(m+1)\right)dt\\ &= \frac{\lambda m(d-m+1)}{r+\lambda m(d-m+1)}\left(\frac{1}{m}+\chi(m+1)\right).\end{aligned}$$

The recursion is initialized by setting $\chi(d+1) = 0$, since all player obtain zero when everyone is contagious. This generates the Clique Viscosity expression in Lemma 1.

Now define $\overline{\phi}(d)$ as in (2). We first argue that (2) has a unique non-zero solution.

Notice that for a player k to not be contagious by time t in the process described above, she must have met neither player i nor player j. Therefore, $x_{ijk}(t) \le e^{-2\lambda t} < e^{-\lambda t}$, so

$$\overline{X}(d) < \int_0^\infty e^{-rt} e^{-\lambda t} \lambda \, dt = \frac{\lambda}{r+\lambda} < \frac{\lambda}{r}.$$

Therefore, the fraction on the righthand side of (2) is strictly greater than 1. By Assumption 1, the equation has a unique non-zero solution.

We now describe the weak perfect Bayesian equilibrium that implements mutual effort at stakes $\overline{\phi}(d)$. Consider a strategy profile $\overline{\sigma}(d)$ such that for each link $\{ij\}$ in G and each history h in H_{ij}, player i plays according to which of the two phases she is in:

1. Cooperation phase: $\sigma_i^{\mathrm{S}}(h) = \overline{\phi}(d)$, and $\sigma_i^{\mathrm{A}}(h, \hat{\phi}_{ij}, \hat{\phi}_{ji})$ = work if and only if $\hat{\phi}_{ij} = \hat{\phi}_{ji} = \overline{\phi}(d)$.
2. Contagion phase: $\sigma_i^{\mathrm{S}}(h) = \overline{\phi}(d)$, and $\sigma_i^{\mathrm{A}}(h, \hat{\phi}_{ij}, \hat{\phi}_{ji})$ = shirk for all $\hat{\phi}_{ij}, \hat{\phi}_{ji}$.

Each player begins in the cooperation phase. If player i is in the cooperation phase at history h, then she stays in the cooperation phase if and only if both players i and j announced stakes $\overline{\phi}(d)$ and worked; otherwise, she transitions to the contagion phase. The contagion phase is absorbing.

To prove that $\overline{\sigma}(d)$ is a weak perfect Bayesian equilibrium, we have to show that incentives are satisfied in both cooperation and contagion phases, given appropriate beliefs. By construction, cooperation phase incentives bind: setting stakes equal to $\overline{\phi}(d)$ on each link makes $\mathrm{IC}_{ij}^{\mathrm{Coop}}$ on each link $\{ij\}$ bind. Lemma 5 in Appendix A proves that if players are indifferent between working and shirking on the equilibrium path, then the players strictly prefer to shirk off the equilibrium path—regardless of their beliefs—because their incentive to work strictly declines in the number of contagious players. The proof exploits the players' ability to select their stakes, and adapts Lemma 1 of Ellison (1994) to this setting. Therefore, $\overline{\sigma}(d)$ is a weak perfect Bayesian equilibrium. □

Comparing viscosity across networks: Our next result shows that the clique of degree d minimizes viscosity among all networks in which no player has more than d partners.

Lemma 3. *If the maximal degree in network G is no more than d, then for every pair of links $\{ij\}, \{ik\}$ in G,*

$$X_{ijk} \geq \overline{X}(d).$$

The inequality is strict if and only if the component of G that contains players i, j, and k is not a clique of degree d.

We develop some notation to describe paths by which contagion spreads. For a path ζ, let Z_ζ be its length, and for every $z \in \{1, ..., Z_\zeta\}$, let $\zeta(z)$ be the zth node in the path. Given a generic realization of the link recognition process on the time period $[0, \infty)$, let $(l_z)_{z=1}^{\infty}$ be the list of links in their order of recognition. (We exclude the non-generic set of realizations for which two links ever meet simultaneously.) We say that $(l_z)_{z=1}^{\infty}$ contains a path ζ if ζ is a subsequence of $(l_z)_{z=1}^{\infty}$.

Proof. Consider a network G in which the maximal degree is d, and fix a triple $\{i, j, k\}$ such that $\{ij\}, \{ik\} \subset G$. Let G_{-i} be the network that results from deleting all of player i's links.

First suppose there is no path between players j and k in G_{-i}. Then player k becomes contagious in a contagion profile only through meeting player i on link ik, which occurs at rate λ. Hence $x_{ijk}(t) = e^{\lambda t}\lambda$ so $X_{ijk} = \int_0^\infty e^{-r\tau}e^{\lambda t}\lambda \, d\tau = \frac{\lambda}{\lambda + r}$, which is shown to be greater than $\overline{X}(d)$ in the proof of Lemma 2.

Now suppose instead that there is at least one path between players j and k in G_{-i}. Consider all the paths in G_{-i} from player j to player k: let ζ be a generic such path, and let S be the set of all such paths. We consider a partition of S, such that two paths are in the same partition element if and only if $\zeta(2) = \zeta'(2)$ and $\zeta(Z_\zeta - 1) = \zeta'(Z_{\zeta'} - 1)$. In other words, the second and second-to-last players in the two paths coincide. Since player j and player k each has at most $(d-1)$ neighbors in G_{-i}, there are at most $(d-1)^2$ partition elements. We denote a partition element by S_{uv} if $\zeta(2) = u$ and $\zeta(Z_\zeta - 1) = v$ for every $\zeta \in S_{uv}$.

Now, consider an arbitrary triple $\{\bar{i}, \bar{j}, \bar{k}\}$ in $\overline{G}(d)$. Let $\overline{N}_m$ be player m's neighborhood in $\overline{G}(d)$. Consider injective functions $g : N_j \to \overline{N}_{\bar{j}}$ and $h : N_k \to \overline{N}_{\bar{k}}$ such that (i) if $\{jk\} \in G$, then $g(k) = \bar{k}$ and $h(j) = \bar{j}$; (ii) $g(v) = h(v)$ for every $v \in N_j \cap N_k$; and (iii) $\bar{i} = g(i) = h(i)$.

We couple the link recognition processes on G and $\overline{G}(d)$ as follows. Given a sequence of link recognitions $(l_z)_{z=1}^{\infty}$ on G,

1. Player $\bar{j}$ meets player $g(u) \in \overline{N}_{\bar{j}}$ whenever player j meets player $u \in N_j$.
2. Player $\bar{k}$ meets player $h(v) \in \overline{N}_{\bar{k}}$ whenever player k meets player $v \in N_k$.
3. For any path $\zeta \in S_{uv}$ contained in $(l_z)_{z=1}^{\infty}$ for which $Z_\zeta \geq 4$, player $g(u) \in \overline{N}_{\bar{j}}$ meets player $h(v) \in \overline{N}_{\bar{k}}$ when player u meets player $\zeta(3)$ in ζ.

Consider any sequence of link recognitions $(l_z)_{z=1}^{\infty}$ on G such that at time 0 the set of contagious players is $\{i, j\}$, and such that player k becomes contagious before link $\{ik\}$ is recognized. Such a sequence contains some path ζ that is completed (all its links are recognized) before link $\{ik\}$ is recognized. By considering all possible such paths, we argue that in the coupled link recognition process on $\overline{G}(d)$, player $\bar{k}$ must become contagious before meeting $\bar{i}$. The most straightforward case is that of $Z_\zeta = 2$: it follows that link $\{jk\} \in G$, and the unique path $(\{jk\}) \in S_{kj}$ is completed before link $\{ik\}$ meets; by the coupling, player $\bar{j}$ must meet player $\bar{k}$ (and so player $\bar{k}$ is contagious) before link $\{\bar{i}\bar{k}\}$ is recognized. The next case is that of $Z_\zeta = 3$: it follows that there exists a player $u \in N_j \cap N_k \backslash \{i\}$, and the path is $\zeta \in S_{uu}$; by the coupling, first player $\bar{j}$ must meet player $g(u) \neq \bar{i}$, and then player $g(u)$ must meet player $\bar{k}$ (and so player $\bar{k}$ is contagious), both before link $\{ik\}$ is recognized. Finally, consider any path $\zeta \in S_{uv}$ with $Z_\zeta \geq 4$: by the coupling, first player $\bar{j}$ must meet player $g(u) \neq \bar{i}$, next player $g(u)$ must meet player $h(v) \neq \bar{i}$, then player $h(v)$ must meet player $\bar{k}$ (and so player $\bar{k}$ is contagious), all before link $\{ik\}$ is recognized.

Thus, for any sequence of link recognitions $(l_z)_{z=1}^{\infty}$ on G in which link $\{ik\}$ meets only after player k has already infected by a path from player j, there is a coupled sequence of link realizations on $\overline{G}(d)$ in which the analogous event occurs. Note that the density of link realizations on $\overline{G}(d)$ is at least as high as that in G.

The other form of link realization by which player k may be infected before $\{ik\}$ is recognized is if player i meets a neighbor $m \in N_i \backslash \{j, k\}$, and then a path from player m infects player k before link $\{ik\}$ is recognized. The proof applies *mutatis mutandi* with player m taking the role of player j. Since player i has at most d neighbors in network G, the density of these link realizations is no greater in network G than in $\overline{G}(d)$.

Finally, we argue that the inequality is strict if the component of G that contains players i, j, and k differs from $\overline{G}(d)$. If $|N_j \cap N_k| < d$, or if $\{jk\} \notin G$, then $\overline{G}(d)$ has additional paths by which player $\overline{k}$ may become infected before link $\{\overline{i}\overline{k}\}$ is recognized, and similarly if $|N_i \cap N_k| < d$. □

4.2 Proof of Theorem 1

We use the lemmas of Section 4.1 to prove our main result. Consider a network G with maximal degree d.

Mutual effort Nash equilibria: Consider a mixed strategy mutual effort equilibrium in which μ_{ij}^S is the distribution of equilibrium path stakes on link $\{ij\}$. Using Lemma 1, the equilibrium path incentive constraint (the analogue of (1)) is

$$\int_0^\infty \phi_{ij}\, d\mu_{ij}^S + \frac{\lambda}{r} \sum_{k \in N_i} \int_0^\infty \phi_{ik}\, d\mu_{ik}^S \geq \int_0^\infty T(\phi_{ij})\, d\mu_{ij}^S + \sum_{k \in N_i \setminus \{j\}} \left(\int_0^\infty T(\phi_{ik})\, d\mu_{ik}^S \right) X_{ijk}.$$

Let $\overline{\phi}_{ij} = \int_0^\infty \phi_{ij}\, d\mu_{ij}^S$, and consider the individual stakes profile $\overline{\Phi}_i = (\overline{\phi}_{ij})_{j \in N_i}$. It follows from Jensen's Inequality (since T is convex) and Lemma 3 that

$$\overline{\phi}_{ij} + \frac{\lambda}{r} \sum_{k \in N_i} \overline{\phi}_{ik} \geq T(\overline{\phi}_{ij}) + \sum_{k \in N_i \setminus \{j\}} T(\overline{\phi}_{ik}) X_{ijk} \geq T(\overline{\phi}_{ij}) + \sum_{k \in N_i \setminus \{j\}} T(\overline{\phi}_{ik}) \overline{X}(d). \tag{3}$$

Since the above inequality is true for each of player i's links, there are d_i inequalities of this form for player i. Therefore, adding and averaging across them

$$\begin{aligned}
\frac{1}{d_i} \sum_{j \in N_i} \overline{\phi}_{ij} + \frac{\lambda}{r} \sum_{k \in N_i} \overline{\phi}_{ik} &\geq \frac{1}{d_i} \sum_{j \in N_i} \left(T(\overline{\phi}_{ij}) + \sum_{k \in N_i \setminus \{j\}} T(\overline{\phi}_{ik}) \overline{X}(d) \right) \\
&= (1 + (d_i - 1)\overline{X}(d)) \left(\frac{1}{d_i} \sum_{j \in N_i} T(\overline{\phi}_{ij}) \right) \\
&\geq (1 + (d_i - 1)\overline{X}(d))\, T\left(\frac{\sum_{j \in N_i} \overline{\phi}_{ij}}{d_i} \right),
\end{aligned}$$

where the final inequality follows from Jensen's Inequality. Using $\phi_i' \equiv \frac{1}{d_i}\sum_{j\in N_i}\overline{\phi}_{ij}$, the above inequality implies

$$\frac{T(\phi_i')}{\phi_i'} \leq \frac{1+d_i\frac{\lambda}{r}}{1+(d_i-1)\overline{X}(d)}.$$

Our aim is to show that $\phi_i' \leq \overline{\phi}(d)$. If $d_i = d$, then this follows from Assumption 1 since the RHS corresponds to $T(\overline{\phi}(d))/\overline{\phi}(d)$. Otherwise, if $d_i < d$ it suffices to show that

$$\frac{1+d_i\frac{\lambda}{r}}{1+(d_i-1)\overline{X}(d)} < \frac{1+d\frac{\lambda}{r}}{1+(d-1)\overline{X}(d)},$$

which is equivalent to $\overline{X}(d) < \frac{\lambda}{r+\lambda}$, already established in the proof of Lemma 2. Therefore, it follows that $\phi_i' \leq \overline{\phi}(d)$. Since player i's payoff in equilibrium σ is $\frac{d_i\lambda}{r}\phi_i'$, it follows from $d_i \leq d$ and $\phi' \leq \overline{\phi}(d)$ that her payoff in equilibrium σ is less than $\overline{u}(d)$. Since player i is arbitrary, every player's expected payoff from equilibrium σ in network G is less than $\overline{u}(d)$. Observe that if the component that contains player i is not $\overline{G}(d)$, then the first inequality in (3) is strict, which translates into every subsequent inequality in this line of reasoning being strict.

Non-mutual effort Nash equilibria: Proving that $\overline{u}(d)$ bounds the utilitarian value of an equilibrium that may involve shirking on the equilibrium path involves a series of steps.

Step 1: Derive an aggregate incentive constraint. The "aggregate expected equilibrium utility" of an equilibrium with shirking on the equilibrium path can be written as:

$$U(\sigma) \equiv \frac{\lambda}{r}\sum_{i\in N}\sum_{j\in N_i} u_{ij} = \frac{\lambda}{r}\sum_{\{ij\}\in G}(u_{ij}+u_{ji}) = \sum_{i\in N}\int_0^\infty e^{-rt}e^{-d_i\lambda t}\lambda\sum_{j\in N_i}\Big(u_{ij}+\frac{\lambda}{r}\sum_{k\in N_i}u_{ik}\Big)\,dt.$$

Applying Lemma 1 and Lemma 3 implies that for each player i and link $\{ij\}$,

$$w_{ij}+\sum_{k\in N_i\setminus\{j\}} w_{ik}\overline{X}(d) \leq w_{ij}+\sum_{k\in N_i\setminus\{j\}} w_{ik}X_{ijk} \leq u_{ij}+\frac{\lambda}{r}\sum_{k\in N_i}u_{ik}. \tag{4}$$

It follows that for every i,

$$\begin{aligned}\frac{\lambda}{r}\sum_{j\in N_i} u_{ij} &= \int_0^\infty e^{-rt}e^{-d_i\lambda t}\lambda\sum_{j\in N_i}\Big(u_{ij}+\frac{\lambda}{r}\sum_{k\in N_i}u_{ik}\Big)\,dt\\ &\geq \int_0^\infty e^{-rt}e^{-d_i\lambda t}\lambda\sum_{j\in N_i}\Big(w_{ij}+\sum_{k\in N_i\setminus\{j\}}\overline{X}(d)w_{ik}\Big)\,dt\\ &= \frac{\lambda}{r+d_i\lambda}\big(1+(d_i-1)\overline{X}(d)\big)\sum_{j\in N_i}w_{ij},\\ &\geq \frac{\lambda}{r+d\lambda}\big(1+(d-1)\overline{X}(d)\big)\sum_{j\in N_i}w_{ij},\end{aligned}$$

where the first inequality follows from (4) and the second inequality follows from $\overline{X}(d) < \frac{\lambda}{r+\lambda}$. Define

$$W(\sigma) = \frac{\lambda}{r+d\lambda}\big(1+(d-1)\overline{X}(d)\big)\sum_{\{ij\}\in G}(w_{ij}+w_{ji}).$$

Adding across each of the i players implies that $U(\sigma) \geq W(\sigma)$, an aggregated incentive constraint.

Step 2: Find a dominating mutual effort profile that satisfies the aggregate constraint. We construct a mutual effort strategy profile $\tilde{\sigma}$ on G such that $U(\tilde{\sigma}) \geq U(\sigma) \geq W(\sigma) \geq W(\tilde{\sigma})$. Consider the function $f(\phi) \equiv \max\big\{0, \frac{1}{2}\big(T(\phi) - V(\phi)\big)\big\}$, and for a set $A \subset \mathbb{R}_+$, define $f^{-1}(A) \equiv \big\{\phi \in \mathbb{R}_+ : f(\phi) \in A\big\}$. If the stage game satisfies strategic complementarity, then $f(\phi) < \phi/2$. Therefore

$$T(f(\phi)) < T\left(\frac{\phi}{2}\right) < \frac{T(\phi)}{2}, \tag{5}$$

where the first inequality follows from T being strictly increasing, and the second inequality from T being strictly convex.

Using f, we construct a new distribution of stakes ρ^{S}_{ij} from μ^{S}_{ij}. For every measurable subset A of $\mathbb{R}_+$, let

$$\rho^{\mathrm{S}}_{ij}(A) = \int_{\phi\in A} p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_{\phi\in A}\int_{\hat\phi\in f^{-1}(\{\phi\})} \left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi + \mathbb{1}(0\in A)\int_{\phi} p^{\mathrm{ss}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij}.$$

Consider a mutual effort profile in which if $i<j$, for every history $h\in H_{ij}$, $\tilde\sigma^{\mathrm{S}}_i(h) = \rho^{\mathrm{S}}_{ij}$, and if $i>j$, $\tilde\sigma^{\mathrm{S}}_i(h) = \sup_{\phi\in\mathrm{Supp}(\rho^{\mathrm{S}}_{ij})}\phi$. Such a stake proposal strategy profile implements the distribution ρ^{S}_{ij} in each link $\{ij\}$. Since $\tilde\sigma$ is a mutual effort profile, players work on the equilibrium path.

To argue that $U(\tilde\sigma)\geq U(\sigma)$, let $\tilde u_{ij}$ and $\tilde w_{ij}$ be the analogues of u_{ij} and w_{ij}. By construction of f, u, and $\tilde u$,

$$\begin{aligned}
\tilde u_{ij} + \tilde u_{ji} &= \int_0^\infty 2\phi\, d\rho^{\mathrm{S}}_{ij}\\
&= \int_0^\infty 2\phi p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_0^\infty 2\phi \int_{\hat\phi\in f^{-1}(\{\phi\})}\left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi\\
&= \int_0^\infty 2\phi p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_0^\infty \int_{\hat\phi\in f^{-1}(\{\phi\})} 2f(\hat\phi)\left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi\\
&\geq \int_0^\infty 2\phi p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_0^\infty \int_{\hat\phi\in f^{-1}(\{\phi\})} \left(T(\hat\phi) - V(\hat\phi)\right)\left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi\\
&= u_{ij} + u_{ji}.
\end{aligned}$$

Since this holds for every $\{ij\}$ in G, it follows that $U(\tilde\sigma)\geq U(\sigma)$.

We take the analogous steps for W: to prove that $W(\tilde\sigma)\leq W(\sigma)$, it suffices to establish that for every $\{ij\}$ in G, $\tilde w_{ij} + \tilde w_{ji} \leq w_{ij} + w_{ji}$:

$$\begin{aligned}
\tilde w_{ij} + \tilde w_{ji} &= \int_0^\infty 2T(\phi)\, d\rho^{\mathrm{S}}_{ij}\\
&= \int_0^\infty 2T(\phi) p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_0^\infty 2T(\phi) \int_{\hat\phi\in f^{-1}(\{\phi\})}\left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi\\
&= \int_0^\infty 2T(\phi) p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_0^\infty \int_{\hat\phi\in f^{-1}(\{\phi\})} 2T(f(\hat\phi))\left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi\\
&\leq \int_0^\infty 2T(\phi) p^{\mathrm{ww}}_{ij}(\phi)\, d\mu^{\mathrm{S}}_{ij} + \int_0^\infty \int_{\hat\phi\in f^{-1}(\{\phi\})} T(\hat\phi)\left(p^{\mathrm{ws}}_{ij}(\hat\phi) + p^{\mathrm{ws}}_{ji}(\hat\phi)\right) d\mu^{\mathrm{S}}_{ij}\, d\phi\\
&= w_{ij} + w_{ji},
\end{aligned}$$

where the inequality follows from (5), and the rest from construction.

Step 3: Prove that the binding contagion equilibrium on the clique dominates $\tilde{\sigma}$.
Consider

$$\phi' \equiv \frac{\sum_{\{ij\}\in G} \int_0^\infty \phi_{ij} d\rho_{ij}^S}{|G|}$$

as the average on-path path stakes in the strategy profile $\tilde{\sigma}$. Consider a pure strategy mutual effort strategy profile σ' in which the stakes on each link $\{ij\}$ in G are ϕ'. It follows by construction that $U(\sigma') = U(\tilde{\sigma})$. Observe that

$$\begin{aligned} W(\sigma') &= \frac{\lambda}{r + d\lambda}(1 + (d-1)\overline{X}(d))2|G|T(\phi') \\ &\leq \frac{\lambda}{r + d\lambda}(1 + (d-1)\overline{X}(d))2 \sum_{\{ij\}\in G} \int_0^\infty T(\phi_{ij}) d\rho_{ij}^S = W(\tilde{\sigma}), \end{aligned}$$

in which the inequality follows from Jensen's Inequality. Therefore, $U(\sigma') \geq W(\sigma')$, which by substitution implies that

$$\frac{2|G|\lambda}{r}\phi' \geq \frac{2|G|\lambda}{r + d\lambda}\big(1 + (d-1)\overline{X}(d)\big)T(\phi') \qquad \Longrightarrow \qquad \frac{T(\phi')}{\phi'} \leq \frac{r + d\lambda}{r + r\overline{X}(d)(d-1)} = \frac{T(\overline{\phi}(d))}{\overline{\phi}(d)},$$

which by Assumption 1 implies that $\phi' \leq \overline{\phi}(d)$. Summarizing, the utilitarian value of σ is

$$\frac{U(\sigma)}{n} \leq \frac{|G|}{n}\left(\frac{\lambda}{r}\overline{\phi}(d)\right) \leq \frac{d\lambda}{r}\overline{\phi}(d) = \overline{u}(d), \tag{6}$$

in which the second inequality follows from $|G|/n \leq d$. Notice that if G includes a component that is not $\overline{G}(d)$, the first inequality in (4) is strict for some player i and link $\{ij\}$, and thus the first inequality in (6) is also strict. ∎

5 Costly Linking

Establishing relationships can be costly, in which case an optimal network must balance the benefits of linking with its costs. The previous section analyzed the setting with a constraint on the maximal degree. This section applies those insights when linking costs take a more general form.

Suppose that if player i has d_i links, she pays a linking cost $c(d_i)$ at time 0. Her *net payoff* is the sum of her expected equilibrium payoff from interaction (henceforth *interaction payoffs*) minus the linking costs incurred at time 0. An equilibrium's *net value* is the average net payoff in society.

The linking cost function c is non-decreasing and, as a normalization, satisfies $c(0) = 0$. We study linking costs that belong to one of two categories below: for $d \geq 1$, linking costs are

1. *Concave* if $c(d)/d$ is non-increasing.
2. *Strictly convex* if $c(d)/d$ is strictly increasing.

A special case of concave linking costs is that in which linking costs are linear.

In this environment there are two benefits to forming a link: it creates a new relationship, and it indirectly benefits other relationships if it completes a cycle (ignoring potential off-path complications). The first benefit is internalized by the partners who form the link, while the latter is a positive externality. Since the marginal linking cost is either flat or decreasing, an extreme solution dominates in which if it is worthwhile to link to anyone then it is better to link to everyone. Since it is symmetric, the complete network has a binding contagion equilibrium in which off-path incentives are guaranteed. In this equilibrium, each player earns a net payoff of $\overline{u}(n-1) - c(n-1)$; by symmetry this is also the equilibrium's net value.

Theorem 2. *Suppose that linking costs are concave.*

1. *Consider any mutual effort equilibrium on any incomplete network. If a player has a non-negative net payoff, then her net payoff is strictly less than* $\overline{u}(n-1) - c(n-1)$.

2. *Consider any equilibrium on any incomplete network, and suppose the stage game satisfies strategic complementarity. If the equilibrium's net value is non-negative, then its net value is strictly less than $\overline{u}(n-1) - c(n-1)$.*

Thus the complete network Pareto dominates other networks if players follow a mutual effort equilibrium, and is utilitarian optimal even when shirking is allowed on the equilibrium path.

Proof of Theorem 2 on p. 27. Consider a non-empty incomplete network G in which some player i obtains interaction payoff u_i in a mutual effort equilibrium, and $u_i \geq c(d_i)$. From the argument in Theorem 1, it follows that

$$\frac{u_i}{d_i} < \frac{\overline{u}(n-1)}{n-1},$$

and, because linking costs are concave, $c(d_i)/d_i \geq c(n-1)/(n-1)$. Combining these two inequalities and multiplying by $n-1$ yields

$$\overline{u}(n-1) - c(n-1) > \left(\frac{n-1}{d_i}\right)(u_i - c(d_i)) \geq u_i - c(d_i).$$

Now consider an equilibrium in which shirking occurs on the equilibrium path. From the argument in Theorem 1, it follows that the average interaction payoff is strictly less than $\overline{u}(n-1)$. Because linking costs are concave, it also follows that the average linking cost is at least $c(n-1)/(n-1)$, and therefore, it follows as above that the net value is strictly less than $\overline{u}(n-1) - c(n-1)$. □

Comparisons across networks are more subtle if $c(\cdot)$ is convex. Now a player may not find it in her own interest to link to all other players, but others always benefit from her doing so. Since a player's preference is no longer aligned with others' preferences, we weaken network comparisons to that of Pareto efficiency. We prove that that there exists a Pareto efficient clique: if any player does better on another network G in a mutual effort equilibrium, then there must be another player who is worse off in G than in the clique. We find this Pareto efficient clique by considering the "best clique size." If linking costs are sufficiently convex, there exists some (generically unique) *optimal clique* $\overline{G}(d^*)$ such that d^* uniquely maximizes $\overline{u}(d) - c(d)$, and if there is no interior optimum, then the optimal clique is the complete network.

Theorem 3. *Suppose that linking costs are strictly convex. For every network G and every mutual effort equilibrium, if there exists a player whose net payoff is strictly greater than $\overline{u}(d^*) - c(d^*)$, then there exists another player whose net payoff is strictly less than $\overline{u}(d^*) - c(d^*)$.*

Proof. Consider a network G and a mutual effort equilibrium in which some player earns a strictly higher net payoff than $\overline{u}(d^*) - c(d^*)$. Necessarily, the component containing this player cannot be a clique since by Theorem 1 no clique yields any player a net value greater than $\overline{u}(d^*) - c(d^*)$. Within this component, consider the player with the highest degree, d'. By Theorem 1, her interaction payoff is strictly less than $\overline{u}(d')$, the interaction payoff she would obtain in a clique of her own degree, $\overline{G}(d')$; meanwhile, her linking costs in G and $\overline{G}(d')$ are identical. It follows that her net payoff is strictly less than $\overline{u}(d') - c(d')$, which by definition is no greater than $\overline{u}(d^*) - c(d^*)$. □

Stronger comparisons emerge when contrasting contagion on cliques to smaller classes of networks. Consider the class of networks that are *regular*, i.e., those in which all players share the same degree. Since the linking costs on a regular network are identical to those on the clique with the same degree, Theorem 1 has a direct corollary.

Corollary 1. *No mutual effort equilibrium on any regular network can yield any player a payoff that exceeds $\overline{u}(d^*) - c(d^*)$. Moreover, if the stage game satisfies strategic complementarity, no equilibrium on any regular network attains a higher net value than $\overline{u}(d^*) - c(d^*)$.*

6 Discussion

This paper characterizes networks that optimally sustain cooperation. Our main result (Theorem 1) compares a perfect Bayesian equilibrium on a clique of degree d to Nash equilibria on all networks that have maximal degree d. The perfect Bayesian equilibrium on the clique takes the form of a "binding contagion" equilibrium. It Pareto dominates all mutual effort equilibria on these other networks, and is utilitarian optimal even among equilibria that may not feature mutual effort. In the remainder of this section, we comment on various features of our framework.

Variable stakes: We represent the level of cooperation in a mutual effort equilibrium by the endogenously selected stakes at which cooperation is incentive compatible. In our view, permitting individuals to select the stakes of their relationship is a realistic formulation of partnerships because it allows partners to choose the terms on which they cooperate. Such cooperative arrangements are ubiquitous: in risk-sharing arrangements, individuals choose how much self-insurance they can attain; in trading and employer-employee relationships, the seller of a good or service chooses how much effort to exert, and the the buyer chooses how much to pay.

Ghosh and Ray (1996) and Kranton (1996) were the first to note the relevance of variable stakes in community enforcement, but to elucidate a different force: building cooperation over time helps screen out myopic players and deters patient players from shirking and re-matching with a new partner.[8] Our stylized framework for stakes and stake selection departs from theirs, but identical results would hold for many different formulations of variable stakes, including theirs. In terms of the stake selection protocol, all that is needed is that it be sufficiently permissive: for every $\phi > 0$, there must be some strategy profile of the partnership that implements ϕ with probability 1.[9] It also would suffice for players to select stakes only at time 0, rather than at each interaction. Moreover, similar results would also hold if, as in Ghosh and Ray (1996), players in each partnership simultaneously chose actions from a continuum in which higher actions benefit the partner but come at a greater cost.

Apart from realism, the inherent flexibility of variable stakes simplifies analysis and exposition considerably. In contrast to the standard repeated games approach of fixing the stage game payoffs and then identifying sets of discount factors for which cooperation

[8]See Watson (1999, 2002) and Athey, Calvano, and Jha (2010) for related insights. Haag and Lagunoff (2007) and Wolitzky (2012) use continuous action sets to study local interaction settings, where a player takes a single action with respect to her entire neighborhood.

[9]For example, similar results would hold if players were required to both propose ϕ for ϕ to be selected, or the average of two proposals were selected.

arises, we can identify the maximal level of cooperation given a fixed level of patience and then directly compare payoffs across equilibria and networks at the same discount rate. Were the stakes ϕ fixed, we would be compelled to distinguish networks and equilibria by the sets of parameters for which incentive conditions are satisfied, which is both indirect and less transparent. Moreover, in fixed stakes environments it is particularly challenging to verifying both equilibrium path incentives and the credibility of punishments off the equilibrium path.

Stationary behavior on the path of play: An important assumption that we make to tractably compare networks is restrict attention to equilibria that are stationary on the equilibrium path. This restriction is *with* loss of generality, and rules out the following kind of behavior: Carol sets higher stakes with Ann if she has recently interacted with Bob, and lower stakes otherwise. Such behavior facilitates cooperation between Ann and Bob in the following way: by choosing to work with Bob, Ann benefits from the prospect of future cooperation with Carol at higher stakes if Bob and Carol meet in between, but if she chooses to shirk on Bob then she can successfully shirk on Carol only if Bob and Carol have not met, in which case the stakes must be low. Unlike a setting in which behavior on each link is publicly observed, such non-stationary behavior potentially offers greater incentives for cooperation.

We abstract from this kind of behavior for two reasons. First, it is difficult to stitch together this form of complex behavior into a perfect Bayesian equilibrium, because it would have to be that Ann would not wish to shirk on Carol at the higher stakes described above, and thus, some non-stationary behavior with Bob is needed for incentives at that later stage. Thus, the mere construction of such an equilibrium for a non-trivial network appears challenging. Second, even if we could construct such equilibria, existing techniques would not permit us to compare sets of equilibria across networks for fixed parameters and patience.

Nash equilibrium and communication: Although we construct a perfect Bayesian equilibrium in "binding contagion" strategies that is optimal among Nash equilibria, there also exist other Nash equilibria that attain the same payoffs and implement the same behavior on the equilibrium path. A particularly interesting class of equilibria is enabled if we add a communication stage to each interaction, prior to the stake selection stage. In the communication stage both partners simultaneously choose which interactions from their private history to reveal; for simplicity we assume that they can conceal interactions, but not fabricate or falsify them. In this environment the following is a Nash equilibrium:

- In every communication stage, reveal your private history truthfully;
- In every stake selection stage, propose stakes $\overline{\phi}$;
- In every action stage, work if and only if you have never seen or heard of your current partner deviate.

Under these strategies, players who have deviated are "ostracized" permanently. The best deviation is to shirk on all your partners, while doing your best to conceal your own past deviations. A deviator faces exactly the same punishment as in the binding contagion equilibrium.

Restricting attention to Nash equilibria, one can enrich the game with further opportunities to communicate. Suppose that players also meet at rate $\mu > 0$ purely to communicate (in addition to meeting at rate λ to interact). We have verified that Theorem 1 still holds, with one small change: there exists a *Nash* equilibrium on the clique of degree d that Pareto dominates all mutual effort equilibria and is utilitarian optimal among all equilibria on all networks with maximal degree d.

The key distinction between this result and Theorem 3 is that this Nash equilibrium is not a perfect Bayesian equilibrium. The challenge is with the incentive to communicate truthfully off the equilibrium path. In Ali and Miller (2013) we show that if sequential rationality is required then no "permanent ostracism" equilibrium can outperform mere bilateral enforcement. The problem is not with the equilibrium path incentive constraints; instead the problem arises from the requirement that the "victim" of a deviator must prefer to reveal the truth off the equilibrium path. In that paper we discuss how more complex "temporary ostracism" equilibria can outperform bilateral enforcement.

Applications: We have framed our results in the context of a prisoners' dilemma with variable stakes, but our findings are relevant for other settings that feature relationships with moral hazard. In an earlier working version of this paper (available on our websites), we showed that the same results apply to "favor-trading" networks where each interaction is a one-sided opportunity for a player to do a favor for her partner. A clique of degree d generates a greater volume of favors than any other network of maximal degree d. We also studied two-sided "networked markets," where links can form only between buyers and sellers. Our main result extends to networked markets as follows: for any given pair of maximal degrees (d_S, d_B) among buyers and among sellers the utilitarian optimal network is the "bipartite clique" comprising d_B sellers each with degree d_S and d_S buyers each with degree d_B. This setting requires a different coupling argument than Lemma 2 to compare viscosities across bipartite graphs, but once that is in place, our other results are easily adapted.

References

S. Nageeb Ali and David A. Miller. Ostracism. Working paper, 2013.

Attila Ambrus, Markus Möbius, and Adam Szeidl. Consumption risk-sharing in social networks. Working paper, February 2010.

Susan Athey, Emilio Calvano, and Saumitra Jha. A theory of community formation and social hierarchy. Working paper, 2010.

B. Douglas Bernheim and Michael Whinston. Multimarket contact and collusive behavior. *Rand Journal of Economics*, 21(1):1–26, 1990.

Timothy Besley, Stephen Coate, and Glenn Loury. The economics of rotating savings and credit associations. *American Economic Review*, 83(4):792–810, 1993.

Francis Bloch, Garance Genicot, and Debraj Ray. Informal insurance in social networks. *Journal of Economic Theory*, 143(1):36–58, November 2008.

James S. Coleman. Social capital in the creation of human capital. *American Journal of Sociology*, 94(S1):S95–S120, 1988.

Joyee Deb. Cooperation and community responsibility: A folk theorem for repeated matching games with names. Working paper, May 2011.

Joyee Deb and Julio González-Díaz. Community enforcement beyond the prisoner's dilemma. Working paper, July 2011.

Avinash K. Dixit. Trade expansion and contract enforcement. *Journal of Political Economy*, 111(6):1293–1317, 2003.

Glenn Ellison. Cooperation in the prisoner's dilemma with anonymous random matching. *Review of Economic Studies*, 61(3):567–588, July 1994.

Leon Festinger, Stanley Schachter, and Kurt Back. *Social Pressures in Informal Groups*. MIT Press, Cambridge, M.A., 1948.

Parikshit Ghosh and Debraj Ray. Cooperation in community interaction without information flows. *Review of Economic Studies*, 63(3):491–519, 1996.

Edward L. Glaeser, David Laibson, and Bruce Sacerdote. An economic approach to social capital. *Economic Journal*, 112(483):F437–F458, 2002.

Benjamin Golub and Matthew O. Jackson. How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*, 127(3):1287–1338, August 2012.

Mark S. Granovetter. Economic action and social structure: The problem of embeddedness. *American Journal of Sociology*, pages 481–510, 1985.

Mark S. Granovetter. The impact of social structure on economic outcomes. *Journal of Economic Perspectives*, 19(1):33–50, 2005.

Matthew Haag and Roger Lagunoff. Social norms, local interaction, and neighborhood planning. *International Economic Review*, 47(1):265–296, February 2006.

Matthew Haag and Roger Lagunoff. On the size and structure of group cooperation. *Journal of Economic Theory*, 135(1):68–89, 2007.

Joseph E. Harrington, Jr. Cooperation in a one-shot prisoners' dilemma. *Games and Economic Behavior*, 8:364–377, 1995.

Matthew O. Jackson, Tomas Rodriguez-Barraquer, and Xu Tan. Social capital and social quilts: Network patterns of favor exchange. *American Economic Review*, 102(5):1857–1897, 2012.

Michihiro Kandori. Social norms and community enforcement. *Review of Economic Studies*, 59(1):63–80, 1992.

Dean Karlan, Markus Möbius, Tanya Rosenblat, and Adam Szeidl. Trust and social collateral. *Quarterly Journal of Economics*, 124(3):1307–1361, August 2009.

Rachel E. Kranton. The formation of cooperative relationships. *Journal of Law, Eco-*

nomics, and organization, 12(1):214–233, 1996.

Steffen Lippert and Giancarlo Spagnolo. Networks of relations and word-of-mouth communication. *Games and Economic Behavior*, 72(1):202–217, May 2011.

John McMillan and Christopher Woodruff. Interfirm relationships and informal credit in Vietnam. *Quarterly Journal of Economics*, 114(4):1285–1320, 1999.

Francesco Nava and Michele Piccione. Efficiency in repeated games with local interaction and uncertain monitoring. Working paper, July 2012.

Werner Raub and Jeroen Weesie. Reputation and efficiency in social interactions: An example of network effects. *American Journal of Sociology*, 96(3):626–654, 1990.

Satoru Takahashi. Community enforcement when players observe partners' past play. *Journal of Economic Theory*, 145(1):42–62, January 2010.

Christopher Udry. Risk and insurance in a rural credit market: An empirical investigation in northern Nigeria. *Review of Economic Studies*, 61(3):495–526, 1994.

Joel Watson. Starting small and renegotiation. *Journal of Economic Theory*, 85(1):52–90, 1999.

Joel Watson. Starting small and commitment. *Games and Economic Behavior*, 38(1): 176–199, 2002.

Alexander Wolitzky. Cooperation with network monitoring. *Review of Economic Studies*, 2012.

Appendix A Binding contagion

In this section we show that the Nash equilibrium payoffs identified in **??** can be implemented in weak perfect Bayesian equilibrium using "binding contagion" strategies.

Let $\pi_i(M)$ be player i's continuation payoff in a contagion strategy profile when she believes that $M \subseteq N$ is the set of contagious players. In a contagion profile, if M is non-empty then it must include player i. A sufficient condition for player i to prefer to shirk on player k is

$$\phi_{ik} + \pi_i(M) \leq T(\phi_{ik}) + \pi_i(M \cup \{k\}). \qquad (\mathrm{IC}_{ij}^{\mathrm{Cont}})$$

If $M \subset N_i \cup \{i\}$, then this inequality embodies the incentives that player i faces when she knows (from her past history) that players in M are contagious, and believes that the remaining players are cooperative. If $\text{IC}_{ij}^{\text{Cont}}$ is satisfied, then player i prefers to shirk even if player k is not contagious.[10]

We first prove the analogue of Lemma 1 of Ellison (1994): in a contagion equilibrium, the marginal incentive to work decreases in the number of contagious players.

Lemma 4. *For every set of players $M \subseteq N$ with $i \in M$,*

$$\pi_i(M \backslash \{j\}) - \pi_i(M \cup \{j\}) \leq \pi_i(\{i\}) - \pi_i(\{i,j\}), \tag{7}$$

with strict inequality if M contains any player $j' \notin \{i,j\}$.

Proof. We establish this claim for every generic sequence of link recognitions (in which no two links meet simultaneously) and then take expectations over them. Let $\xi = (\tau_z, \ell_z)_{z=1}^{\infty}$ be a sequence of link recognitions that take place in $[0,\infty)$, where $(\tau_z)_{z=1}^{\infty}$ is the ordered list of link recognition times and $(l_z)_{z=1}^{\infty}$ is the list of links in their order of recognition.

Fix a player i and suppose that M_0 is the set of players who are contagious (including player i) at a time normalized to zero. If the subsequent sequence of link recognitions follows ξ, then the set of contagious players at time τ_z is

$$C_z(M_0,\xi) \equiv \begin{cases} M_o & \text{if } z = 0, \\ C_{z-1}(M_0,\xi) & \text{if } z > 0 \text{ and either } l_z \subseteq C_{z-1}(M_0,\xi) \text{ or } l_z \subseteq N \backslash C_{z-1}(M_0,\xi), \\ C_{z-1}(M_0,\xi) \cup l_z & \text{otherwise.} \end{cases} \tag{8}$$

When two players who are both cooperative or both contagious meet, no player changes phase; it is only when a contagious player meets a cooperative player that the latter also becomes contagious. Define $\pi_i(M_0|\xi)$ to be the equilibrium continuation payoff

[10] Generally, contagious players need not hold such optimistic beliefs about their partners, but beliefs that attribute greater probability to others being contagious create a stronger incentive to shirk. Establishing incentives to shirk under the most optimistic beliefs about others ensures that contagious behavior is incentive compatible for all beliefs.

of player i when players in M_0 (including player i) are in the contagion phase at time zero, the realization of recognition times is $\{\tau_z\}_{z=1}^{\infty}$. By calculation,

$$\begin{aligned}
&\pi_i(M_0|\xi) - \pi_i(M_0 \cup \{j\}|\xi) \\
&\quad= \sum_{z=1}^{\infty} e^{-r\tau_z} \sum_{k\in N_i} T(\phi_{ik})\, \mathbb{1}\big(\ell_z = \{i,k\} \text{ and } k \in C_z(M_0 \cup \{j\},\xi) \backslash C_z(M_0,\xi)\big) \\
&\quad\leq \sum_{z=1}^{\infty} e^{-r\tau_z} \sum_{k\in N_i} T(\phi_{ik})\, \mathbb{1}\big(\ell_z = \{i,k\} \text{ and } k \in C_z(\{i,j\},\xi) \backslash C_z(\{i\},\xi)\big) \\
&\quad= \pi_i(\{i\}|\xi) - \pi_i(\{i,j\}|\xi)
\end{aligned} \tag{9}$$

where $\mathbb{1}$ is the indicator function. The weak inequality follows from

$$C_z(M_0 \cup \{j\},\xi) \backslash C_z(M_0,\xi) \subseteq C_z(\{i,j\},\xi) \backslash C_z(\{i\},\xi), \tag{10}$$

since the set of players who catch contagion via a path through player j is decreasing in the number of other players through whom contagion can spread. Therefore, since (9) holds for every ξ, taking the expectation over ξ yields (7) with weak inequality.

Moreover, suppose there exists some $j' \in M_0 \backslash \{i\}$ and $k \notin M_0$, in which case Eq. 10 is strict for every sequence of link realizations starting with $\ell_1 = \{j'k\}$ and $\ell_2 = \{ik\}$.[11] Hence Eq. 10 is strict with probability at least $(n(n-1))^{-2}$, so taking expectations over ξ yields (7) with strict inequality. □

Working with contagion equilibria poses a well-known challenge (Kandori 1992; Ellison 1994): if the reward for working far exceeds the punishment for shirking, then a contagious player may prefer to delay infecting others and instead choose to work. We identify a class of contagion equilibria that solves this problem:

Definition 4. *A contagion profile is* ***binding*** *if IC_{ij}^{Coop} holds with equality for each neighbor j in N_i, for each player i in N.*

The virtue of making incentives bind on the path of play is that contagious players have an incentive to shirk. The logic is analogous to that of Lemma 1 of Ellison (1994):

[11] Indeed, if $j' = j$ then the lefthand side of Eq. 10 is the empty set and so Eq. 10 is strict for every sequence of link realizations starting with $\ell_1 = \{jk\}$.

the marginal gain from working is decreasing in the number of contagious players and so indifference on the equilibrium path implies that a player prefers to shirk off the equilibrium path.

Lemma 5. *Every binding contagion profile is an equilibrium, with a strict incentive to shirk at every history off the equilibrium path.*

Proof. Consider a binding contagion profile with collective stakes profile Φ. Since this profile satisfies cooperation phase incentive constraints by construction, it suffices to establish contagion phase incentives when $\{ik\}$ is recognized and the set of contagious players M contains player i and at least one other player.

If $\hat{\phi}_{ik}$ or $\hat{\phi}_{ki}$ differ from ϕ_{ik}, or if either player i or player k has shirked with the other in a previous interaction, player i has a strict incentive to shirk since player k will shirk. Instead, suppose as in $\text{IC}_{ij}^{\text{Cont}}$ on p. 35 that neither of these have occurred, and player i believes that players M are in the contagion phase, where $M \subseteq N \backslash \{k\}$ and $|M| \geq 2$. Let $\pi_i(\varnothing)$ represent player i's equilibrium continuation payoff. It follows from binding $\text{IC}_{ij}^{\text{Coop}}$ on p. 16 that

$$T(\phi_{ik}) - \phi_{ik} = \pi_i(\varnothing) - \pi_i(\{i,k\}) = \pi_i(\{i\}) - \pi_i(\{i,j\}) > \pi_i(M) - \pi_i(M \cup \{j\}), \tag{11}$$

where the first equality is an expression of binding $\text{IC}_{ij}^{\text{Coop}}$, the second inequality follows from all of player i's cooperation phase incentive constraints binding, and the inequality follows from Lemma 4. Adding $\phi_{ij} + \pi_i(M \cup \{j\})$ to each side yields $\text{IC}_{ij}^{\text{Cont}}$, so the incentive to shirk in the contagion phase is strictly satisfied at all off-path histories. $\square$