Social dilemmas with manifest and unknown networks

ARMANDO RAZO*

Preprint: November 2019

ABSTRACT

Scholarly consensus that social ties resolve social dilemmas is largely predicated on common knowledge of networks. But what happens when people do not know all relevant social ties? Does network uncertainty translate into worse outcomes? I address these concerns by advancing the notion of a Network Estimation Bayesian Equilibrium (NEBE) to examine cooperative behavior under different epistemic conditions. When networks are common knowledge, I find that all possible outcomes of an original cooperation game can be realized in equilibrium, albeit with a higher likelihood of defection for more connected players. Variable knowledge of the network also has a distributional impact. With incomplete network knowledge, it's possible to observe reversed equilibrium behavior when more connected players actually cooperate more often than less connected ones. In fact, aggregate network uncertainty in some social contexts incentivizes more mutual cooperation than would be the case with common knowledge of all social ties.

CONTENTS

LIST OF FIGURES

* *Department of Political Science and IU Network Science Institute, Indiana University, Bloomington, IN, USA*

1 INTRODUCTION

There's ample consensus among many social scientists that "networks matter" due to their established impact on a wide range of individual and aggregate outcomes (Burt, 2000; Pescosolido, 2007; Granovetter, 2005, Victor et al., 2018). To understand enabling mechanisms, social scientists have also developed increasingly richer theories of strategic decisions connect networks to significant aggregate outcomes (Coleman, 1990; Chwe, 1999; Goyal, 2007; Jackson, 2008a). Notwithstanding this consensus and theoretical advances, we currently lack a systematic examination of the informational requirements of networked strategic situations. Is it necessary for participants to have complete knowledge of a network to obtain desirable outcomes? If not, what are the implications of incomplete network knowledge for individual and collective choices? This paper examines these general questions in the context of two-person cooperation games with participants who have variable knowledge of an exogenous network that shapes cooperative behavior. This initial focus on cooperation games paves the way for new research on how aggregate network uncertainty affects the ability of societies to resolve social dilemmas.

Indeed, broad interdisciplinary knowledge suggests that social context enables necessary collective action to resolve social dilemmas. Collective action generally refers to the ability of groups to cooperate and coordinate (Olson 1965). Social dilemmas entail various situations that pit the interests of individuals against the public interest (Shepsle, 2010). Among other social dilemmas, free-riding is a pernicious and pervasive phenomenon in many societies that has received a lot of attention through the lens of cooperation or so-called prisoner's dilemma games. Theoretical predictions for these games go a long way in explaining the high incidence of free-riding. Yet there is ample evidence that people can cooperate even when it is not apparently in their best interest. Indeed, laboratory experiments have consistently shown that some people cooperate when they should not (Sell 2007).[1]. These contrary findings are typically explained by appealing to social preferences or players' broader concerns that attenuate tensions between individual and public interests (Fehr and Fischbacher, 2002).[2] These arguments further call for greater consid-

* *Department of Political Science and IU Network Science Institute, Indiana University, Bloomington, IN, USA*

1 Similar results of higher than expected cooperative behavior appear in public goods experiments that also exhibit a tension between individual and public interests (Ostrom et al., 1994)

2 The theory of repeated games offers alternative explanations for cooperative behavior based on trigger strategies to reward cooperation and punish defection over time (Axelrod, 1984) This paper focuses on incentives to cooperate without the prospect

eration of social ties as determinants of altruistic and cooperative behavior (Brañas et al., 2010).

Networks are especially relevant because these are supposed to embed social norms that encourage pro-social behavior (Elster, 2007; Ostrom, 1990; Ostrom, 1998). Societies with extensive social ties are best positioned to resolve collective action problems under a wide range of mechanisms that include both instrumental and normative considerations (Gould, 1993). More generally, Granovetter (2005) argues that dense networks not only internalize cooperative norms but also generate trust that encourages other beneficial activities. Along these lines, an emphasis on the positive contributions of networks mirrors and reinforces related arguments about the benefits of social capital (Lin, 2001). Sobel (2002) notes that this is a contestable term, but there is nonetheless widespread agreement that social capital is inherently relational because it is *embodied* in the relations among persons" (Coleman, 1990).

Networks and social capital especially intersect through the lens of collective action problems. Scholz et al. (2008) argue that collective action requires solutions to two distinct problems: establishing credible commitments to cooperate, and researching solutions to novel problems. And these collective action problems are directly impacted by networks. So-called *bonding* social capital mitigates risks and free-riding and is facilitated by dense networks. In turn, *bridging* capital calls for expansive ties and network-mediated searches to find useful information for coordinated problem-solving activities (Berardo and Scholz, 2010).

It makes sense to consider networks because social interactions don't generally occur in a vacuum, devoid of any social structure (Granovetter, 2005). And it's plausible that networks can enable collective action through norms and other behavioral mechanisms. However, the implicit requirement regarding network and normative knowledge requires further examination.[3] To be sure, some scholars are careful to delineate required knowledge for social structures that have limited and local scope (Goyal, 2007; Manski, 2000; Young, 1998). However, it's more common to advance arguments about dense networks and social capital with global scope that must be either well-known or readily inferred. Indeed, the promotion and creation of social capital is contingent on relevant audiences knowing the exact

---

of future interactions, but I discuss implications for repeated networked interactions in concluding comments.

3 Many of these arguments have policy implications that call for the creation of beneficial ties, so there is sometimes more emphasis on endogenous rather than the exogenous networks in this article. Regardless, acknowledgment of incipient network structures that require improvements is contingent on complete knowledge of existing patterns.

social structures that enable desirable outcomes (Keefer and Knack, 2005; Stiglitz, 2000).

Overall, extant theory and policy advice suggest that contextual factors–especially networks–are self-evident to all participants (Osborne and Rubinstein, 1994). Modern notions of networks conceive them as relational structures (Wasserman and Faust, 1994). From this perspective, assuming known networks implies that agents possess complete knowledge of relevant patterns of social ties throughout society. But this is not a trivial informational requirement because it's well known that the number of possible patterns increases exponentially with the size of society.

Of course, the actual size of society matters, but a simple example will illustrate that we face a general informational concern that spans all scales. Consider a small society of 5 individuals with 10 distinct pairs (dyads) that can be either connected or disconnected. These simple conditions already translate into 1,024 possible patterns of connections or distinct networks for what is arguably a fairly small group. Adding a single person raises the number of distinct networks by a factor of 2^5 to 32,768.[4] It's not reasonable to expect that any member of society will have exact knowledge of overarching relational patterns under general conditions, especially within large contemporary societies.

Does it really matter whether people know the networks that shape collective behavior? We lack a clear answer to this question. On the one hand, there is some indication that network uncertainty is not always problematic. Some scholars have directly addressed this concern within the larger economic literature on social learning (Chamley, 2004). Societies are represented by a countably infinite set of agents making sequential choices one person at a time (thus, social interactions take place over an infinite time horizon). Agents have informative private signals to aid their decision, but are not completely sure that a particular choice will generate the highest payoff under an unknown state. Sequential choices nonetheless offer the possibility of learning from the past behavior of other (presumably better informed) actors. A big concern in this literature is whether "information aggregation" is possible; namely, whether agents converge to the best action. In that context, Acemoglu et al. (2011) examine social learning mediated by random social connections. Actual network realizations condition exposure to other actors, and can thus induce inefficient learning. However, all relevant information is confined to independent neighborhoods, so it's still possible for agents to learn

4 For a society of size n, the number of distinct dyads equals the Binomial coefficient $\binom{n}{2} = n(n-1)/2$; and each dyad can be in one of two states: connected or disconnected.

the best action as long as network-mediated observations are not confined to a small set of uninformative decisions.

Lobel and Sadler (2015) further examine social learning in a setting with correlated neighborhoods where an agent can derive additional information from knowing whether observed neighbors are themselves related.[5] Aggregate network uncertainty induces different beliefs about which actors are better examples to emulate, thus complicating social learning but not entirely precluding it. With sufficient connectivity and beliefs that minimize what they call "network distortion", information can still diffuse and enable agents to learn up to some expert level.

On the other hand, the positive results of social learning under aggregate network uncertainty stem from a collection of agents who invariably want to choose the same correct actions with higher utilities. There's neither disagreement regarding this common goal nor are there distributional consequences to social learning. Indeed, (evolving) information is a public good that is only partially limited insofar as networks condition exposure to previous choices. Overall, this situation stands in stark contrast to the nature of social dilemmas where not all individuals agree on desirable outcomes, suggesting that the impact of network uncertainty plays a different role under more general strategic conditions.

Indeed, there are strong reasons to believe that network uncertainty is more likely to have negative consequences in the context of social dilemmas. Herreros (2004), for instance, discusses how social capital is affected by mistrust caused by uncertainty about networked participants. Chwe (1999, 2000) demonstrates that uncertainty about neighbors' expected behavior limits the ability of groups to engage in collective action, especially in coordinating large-scale responses to resolve perceived social problems.[6]

Although we lack a general answer about the impact of network uncertainty, it is clear that we need to be explicit about the exact ways through which networks impact strategic behaviors. Along those lines, this paper argues that explanations about the impact of social networks on collective action must explicitly model the implications of an added network structure to an existing social dilemma. Otherwise, it might not be clear whether participants have requisite contex-

5 Required information effectively calls for knowledge of ego networks. These are social structures built around a focal node (a so-called *ego*) and patterns of relationships among its neighbors (so-called *alters*). See Perry et al 2018 for more details.

6 There are also statistical approaches that model uncertainty about underlying social structures that appear similar (Hoff et al., 2002). However, the uncertainty is qualitatively diferent, involving a researcher's inability to account or measure a latent social structure that conditions individual behavior. Presumably, observed behavior was still conditioned by social structures that were known to participants but are hidden from researchers.

tual knowledge, especially in modern societies where agents operate within multiple networks or fairly large network structures (e.g., social media networks where privacy features preclude full knowledge of the underlying social network).

We are a long way from a general framework that adds network uncertainty to all social dilemmas, but this paper makes two critical contributions that move us in that direction. First, I add an explicit social network dimension to canonical games of cooperation, which effectively alters the original interaction with the possibility of multiple games with variable numbers of participants. In my framework, linked games are derived from existing social connections that introduce extra rewards and punishments.[7] I find that complete knowledge of the network does not guarantee mutual cooperation, but networks do enable more equilibrium outcomes with cooperation from at least one player. However, networks are also more likely to produce asymmetric equilibria where highly connected nodes themselves do not cooperate, but their connections nonetheless compel other players to cooperate.

My second contribution entails a direct examination of network uncertainty, modeled with an underlying network generating process (NGP) that randomly creates ties across society. Players have some private network knowledge about their own ties, and they rely on separate NGP beliefs to estimate unknown network features. I find that equilibrium outcomes largely depend on whether players share common prior NGP beliefs. When they share such beliefs, I obtain symmetric outcomes that result in either mutual cooperation or mutual defection, depending on a society's expected connectivity. In other epistemic scenarios, players need to form beliefs based on private network information and partial knowledge of other parts of the network.

Overall, my models indicate that network uncertainty does not inhibit cooperation in its own right if other favorable contextual conditions are present. However, network uncertainty does impact individual behavior because I find that more connected players are more likely to cooperate when they have limited rather than complete knowledge of the network. Relatedly, my findings indicate that network uncertainty affects the distribution of potential gains, thus

7 There is some related work on networks and collective action that focuses on endogenous networks. They show that evolving social structures can affect a society's ability to cooperate or coordinate in the long run (Ahn et al., 2009). Raub et al. (2013) also argue that endogenous networks that exchange information facilitate cooperative outcomes. These results suggest that network uncertainty is not always an impediment, but it's worth noting that their analysis involving a qualitatively different type of uncertainty (not knowing future networks) as opposed to a structure that affects current payoffs.

having distributional implications that are absent from extant arguments about the positive impact of social networks.

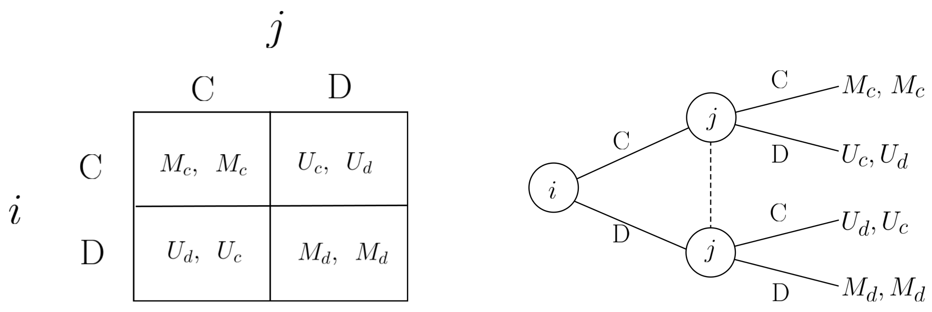
I derive these findings in the rest of this paper as follows. In the next section, I describe relevant social context that can alter a two-player cooperation game with the introduction of two network mechanisms that respectively reward collaboration and punish defection. I first examine these mechanisms under conditions of complete network knowledge. A subsequent section introduces a Bayesian game and a corresponding Network Estimation Bayesian Equilibrium (NEBE) to examine the implications of aggregate network uncertainty on cooperative behavior. I then proceed to examine NEBE predictions under different epistemic scenarios with and without common prior beliefs of an underlying network generating process. A final section concludes.

2 HOW DO NETWORKS IMPACT COOPERATION GAMES?

I show here that social networks matter for the study of cooperation by transforming an otherwise one-on-one interaction between two players into a new strategic situation with one-on-many social interactions. To that effect, I first review the logic of cooperation dilemmas with a baseline two-player cooperation game. I then introduce relevant social context that augments this baseline game with additional players and strategic considerations regarding potential network rewards and retaliations. I conclude by deriving a reduced-form networked cooperation game for continued use in subsequent sections.

2.1 Baseline cooperation games without networks

The baseline interaction involves a game between two players i and j with identical strategy sets $S_i = S_j = \{C, D\}$. Different simultaneous choices to cooperate (C) or defect (D) produce the payoff combinations $U : S_i \times S_j \mapsto \mathbb{R}$ shown in Figure 1(a). Mutual cooperation and mutual defection produce identical payoffs of M_c and M_d, respectively. In combinations where one player cooperates and the other defects, players get respective payoffs for unilateral cooperation (U_c) and unilateral defection (U_d). I focus on the class of such games that obey the following restriction on individual payoffs: $U_d > M_c > M_d > U_c$. This payoff structure identifies a canonical formulation known as the *Prisoner's Dilemma* that I henceforth identify as a *Coop*$\langle i, j\rangle$ game (Fudenberg and Tirole, 1991).



(a) Normal form (b) Extensive form

Figure 1: Basic cooperation game

A classic result for *Coop*$\langle i, j\rangle$ games is a unique Nash equilibrium prediction of mutual defection, $\{D, D\}$, due to incentives to deviate from any cooperative options. If one player were expected to cooperate, the other player would rather defect to gain a positive extortion premium $e = U_d - M_c$. If one player were to defect, the other would behave similarly to avoid extortion and gain $a = M_d - U_c$. Given the symmetric payoff structure, defection is thus the unique best response–indeed a dominant strategy–for each player.

Predicted behavior for *Coop*$\langle i, j\rangle$ games reveals a social dilemma with major implications for collective action. On the one hand, both players can generate a higher social utility of $2 \cdot M_c$ if they cooperate. On the other hand, individuals have private incentives e and a that encourage defection. In the end, individual interests prevail over the public interest, so the socially (Pareto) optimal outcome is unrealized. Not only do players generate a lower social utility of $2 \cdot M_d$, but neither player obtains the highest individual payoff U_d.

2.2 Social context and the impact of networks

2.2.1 Social context

In general, *Coop*$\langle i, j\rangle$ games occur under different social contexts or states of the (social) world. A given state of the world includes more agents in society besides i and j, along with some existing pattern of social connections among them.

To define these underlying states, I first introduce a finite set of agents $I_n = \{1, 2, ..., n\}$ to represent a society of size $n > 2$. I use a *dyad* (a pair of distinct agents or so-called *nodes* $x, y \in I_n$) as the basic unit of analysis to capture patterns of connections. We can thus build up a *network* (structure) succinctly defined by the collection of dyadic ties among nodes in I_n, where n also stands for the size of the network.

To simplify the presentation, I will henceforth use conventional matrix representations of network structures (Wasserman and Faust, 1994).

Definition 1. A **network** (state) ω is an element from the set Ω_n that includes all possible patterns of connections for a society of size n. Any $\omega \in \Omega_n$ can be represented as an $n \times n$ adjacency matrix or sociomatrix ω, where each entry is a 1/0 binary variable. If any distinct agents x and y are connected, then $\omega_{xy} = \omega_{yx} = 1$.[8]

A particular *Coop*$\langle i, j \rangle$ game draws two distinct agents i and j from I_n. I refer to these players either as a *focal* dyad or *focal* nodes to distinguish them from the rest of society denoted by the set $I_{n-ij} \equiv I_n \setminus \{i, j\}$ with size n_{-ij}. Given a particular network (matrix) ω, the entry ω_{ij} determines whether the focal nodes are themselves connected. To capture connections with other agents, I identify a separate set of extraneous *alters* or neighbors for each focal node: $A_i^{\omega} = \{x \in I_{n-ij} | \omega_{ix} = 1\}$ and $A_j^{\omega} = \{y \in I_{n-ij} | \omega_{jy} = 1\}$. These alter sets will have $n_i^A = |A_i^w| \geqslant 0$ and $n_j^A = |A_j^w| \geqslant 0$ elements, respectively.[9]

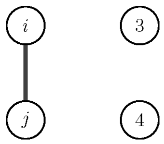
2.2.2 *Collaborative Synergy*

In the context of *Coop*$\langle i, j \rangle$ games, there are several reasons why cooperation between connected parties might confer additional benefits. As noted in the introduction, scholars generally emphasize several benefits to social connections. Additionally, players can share a language, know-how, or other mutual attributes with a synergistic impact upon successful collaborations. Collaboration can also be a rewarding experience in its own right. I capture these potential benefits with a single *synergy bonus* $b_{\omega} \geqslant 0$ that amplifies the payoffs of mutual cooperation as follows: $M_c(1 + w_{ij} \cdot b_{\omega})$. This is a network effect with a narrow social context restricted to a focal dyad.

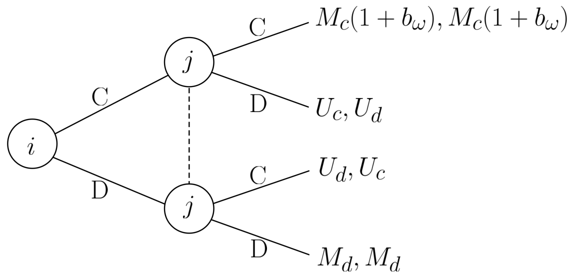
A network with $\omega_{ij} = 1$ alters the original *Coop*$\langle i, j \rangle$ payoffs, so focal nodes will already be playing a slightly different game in that case. Additionally, this single connection can drastically alter the qualitative nature of their interaction. If $b_{\omega} M_c$ were to exceed the extortion

8 This is an *undirected* network that only accounts for the existence of a dyadic tie. Therefore, w will be a symmetric matrix with $\binom{n}{2} = n(n-1)/2$ distinct dyads, each of which can take a value in $\{0, 1\}$. The set Ω_n itself includes all combinations of these dyadic values, ranging from the *empty* network ω_{ϵ} (where all nodes are disconnected) to the *complete* network ω_{all} (where all nodes are connected). A common measure of connectivity is the (relative)*density* of a network with support $[0, 1]$, equal to the ratio of actual ties to the total number of dyads. The extreme worlds ω_{ϵ} and ω_{all} have density values of 0 and 1, respectively.

9 I use conventional terminology for ego networks, given the focus on two original players and their respective connections. However, the network effects that I posit for cooperation games are contained within players' neighborhoods, so it's unnecessary to present a full characterization of ego networks that also includes alter-to-alter ties.



(a) Network ω_1



(b) Networked cooperation game

Figure 2: Collaborative synergy with a connected focal dyad

premium, then the *Coop*$\langle i, j\rangle$ game would be effectively transformed into a coordination game with multiple equilibria that include both mutual cooperation and mutual defection. To retain exclusive focus on cooperation dilemmas, I henceforth assume the following bonus restriction: $0 \leqslant b_\omega < e/M_c$.

To illustrate the collaborative synergy effect, consider a small society with four actors, $I_4 = \{i, j, 3, 4\}$. The related Figure 2a includes an alternative representation of networks as *sociograms*, where nodes are depicted by labeled circles, and dyadic ties by connecting lines. Here, the network ω_1 illustrates the general possibility of isolated focal dyads. Additionally, Figure 2b shows an extensive-form representation of ω_1's corresponding networked cooperation game. These game representations enable direct comparisons against the baseline *Coop*$\langle i, j\rangle$ case in Figure 1.

As i and j are the only connected nodes, the two extraneous agents are irrelevant, so these are not included in the corresponding extensive form. Payoffs for outcomes involving defection are almost identical to *Coop*$\langle i, j\rangle$ game counterparts. However, mutual cooperation payoffs have changed from M_c to $M_c(1 + b_\omega)$, so even a single tie within ω_1 has indeed altered the original interaction, which I capture with the following definition.

Definition 2. A **collaborative synergy transformation** changes the individual payoffs for $\{C, C\}$ in a *Coop*$\langle i, j\rangle$ game from M_c to $M_c(1 + \omega_{ij} \cdot b_\omega)$.

2.2.3 *Solidary Retaliation*

A second network effect stems from an extant logic of solidarity that can increase the number of players, depending on focal nodes' connections. Here, I assume that all members of society are motivated by a notion of solidarity that entails responsibilities to stand together with connected agents that face a negative situation. Specifically, alters are required to punish defection whenever a connected focal node is affected, similar to a schoolyard fight scenario with one students chooses to cooperate (not fight) but another defects (throws a punch). Typical depictions include affected students surrounded by bystanders, some of whom might be inclined to participate to defend a friend.

Solidary retaliation is not constrained to small groups, however, having wide applicability to larger-scale and more frequent social dilemmas. For instance, Bates (2010) offers a sweeping historical account of many societies that have suffered from long-lasting feuds driven by solidary retaliation. I denote the call for additional assistance a *solidary retaliation* requirement that is only relevant when at least one of the alter sets is non-empty. If those set size conditions are satisfied, then the total number of participants can change from the original two to a maximum of $2 + n_i^A + n_j^A$ players.

Assuming at least one alter, solidary retaliation will be both selective and variable. To explain selectivity, which determines participation requirements, I first define the set $H = \{h_{CD}, h_{CD}, h_{DC}, h_{DD}\}$ that includes the possible outcomes from a *Coop*$\langle i, j \rangle$ game, with elements indexed by the respective choices of players i and j. All available alters can observe the behavior of focal nodes, an element of H, so they know the history of play and assumed responsibilities.[10]

In a subsequent stage, if A_i^ω is not empty, player i's alters target$j$ after histories h_{CD} and h_{DD}. If A_j^ω is not empty, then player j's own alters simultaneously target i following histories h_{DC} and h_{DD}. Cumulative penalties targeted at focal nodes will be then be determined by the simultaneous choice of all affected alters, which occurs under conditions of perfect information.

It's possible for A_i^ω and A_j^ω to overlap, in which case those common alters would intervene after all histories excepting h_{CC}. In the case of mutual defection, they will target both i and j. Not doing so would undermine the logic of solidarity in two major respects: (1) if they didn't intervene at all, they would disregard all social responsibilities; and (2) if they only targeted one focal node, they would be showing favoritism.

Alters are themselves strategic actors with actual behavior that depends on their own–and only on their own–circumstances. Each alter k chooses a strategy $r_k \in \{0, 1\}$, where $r_k = 1$ stands for retaliation. By retaliating, each alter k receives ν_k that represents an idiosyncratic value from being individually responsible. However, this action entails some effort, imposing a similar cost c that is common knowledge. Although alters move together, their respective pay-

10 If not through direct observation, alters could readily learn the outcome from focal nodes who would be assumed to broadcast truthful reports to their own connected agents.

offs π_k are independent in the following sense: for any $k \in I_{i-j}$, $\pi_k(r_k = 1, r_{-k}|\forall h \in H) = \nu_i - c$ and $\pi_k(r_k = 0, r_{-k}|\forall h \in H) = 0$. A given alter k will only choose $r_k = 1$ under favorable circumstances if $\pi_k - c \geqslant 0$.[11]

Solidary retaliation thus accumulates in additive fashion, so the identity of third parties is irrelevant. We just need a count.

Definition 3. A **solidary retaliation transformation** takes as inputs a *Coop*$\langle i, j\rangle$ game, and changes the sequence of the game and payoffs as follows:

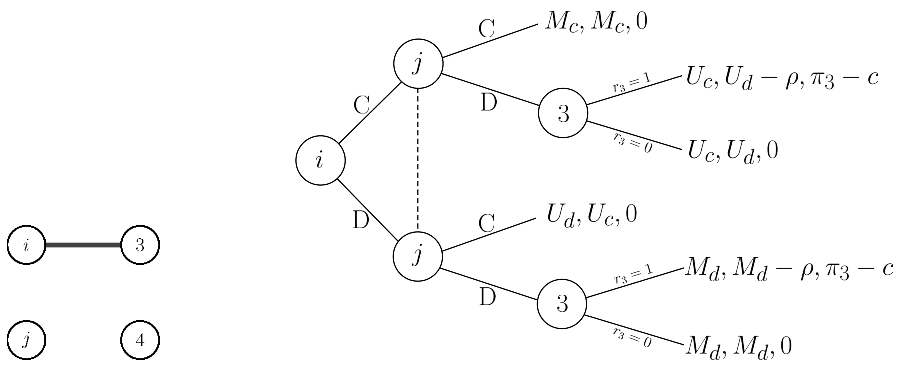
- If $A_i^\omega \neq \emptyset$, then cooperation-stage histories $\{h_{CD}, h_{DD}\}$ extend the game with the simultaneous retaliation choices of i's alters that generate a penalty $R_i(\omega) = \sum_{k \in A_i^\omega} r_k$. Player j's payoffs in outcomes stemming from history h_{CD} equal $U_D - R_i(\omega)$. In outcomes stemming from history h_{DD}, player j's payoffs equal $M_d - R_i(\omega)$.

- Additionally, if $A_j^\omega \neq \emptyset$, then histories $\{h_{DC}, h_{DD}\}$ further extend the game with the simultaneous retaliation choices of j's alters that generate a penalty $R_j(\omega) = \sum_{k \in A_j^\omega} r_k$. Player i's payoffs in outcomes stemming from history h_{DC} equal $U_D - R_j(\omega)$. In outcomes stemming from history h_{DD}, player i's payoffs equal $M_d - R_j(\omega)$.

- In addition to revised payoffs for focal nodes, each outcome will include a payoff vector $(u_k)_{k \in A_i^\omega \cup A_j^\omega}$ that collects payoffs for linked alters, ordered by their corresponding index in I_n. Alters with exclusive ties to either i or j get $u_k = \max(\pi_i - c, 0)$. Players that are connected to both i and j will have a $u_k = \max(2(\pi_k - c), 0)$.

To illustrate this second network transformation, Figure 3a shows a network ω_2 that produces $A_i = \{3\}$ and $A_j = \emptyset$ for the focal dyad.

11 There is a longstanding debate in the sociological literature regarding social solidarity that lies beyond the scope of this paper, but nonetheless merits a brief discussion because there are multiple perspectives (Hechter, 2015; Buchanan, 2019; Kornter 2004). On the one hand, there are rationalist notions of solidarity that suggest that assuming common responsibilities is instrumental to obtain future assistance. Relatedly, people in society might express solidary behavior because their intervention precludes future risks and losses. On the other hand, there is a constitutive or ideational notion that predicates compliance on a shared sense of identity irrespective of costs and benefits. The assumption of a direct cost (and conditional participation on positive net benefits) places the logic of solidarity in this paper within the rationalist camp. However, the subjective nature of π_3 is itself consistent with both motivations because the games analyzed here lack a future to assess instrumental considerations. Moreover, it's generally possible for instrumental and identity considerations to be concurrently relevant (De Deken et al., 2006).



(a) Network ω_2 (b) Networked cooperation game

Figure 3: Sample solidary retaliation with one exclusive neighbor

Under ω_2, a second stage is only enabled by histories where j has chosen to defect (hence agent 3 can directly observe the joint behavior of focal nodes). Notably, agent 3 does not react to i's own uncooperative behavior. Rather, it targets agent j with a penalty $R_i(w) = \rho$.

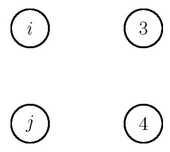
Agent 3's intervention changes the baseline *Coop*$\langle i, j \rangle$ game in two significant ways. First, the original *Coop*$\langle i, j \rangle$ game becomes the first stage of a sequential game with an additional player as shown in Figure 3b. Second, these added game features can change equilibrium behavior. In particular, an applicable Subgame Perfect Nash equilibrium (SPNE) depends on agent 3's parameters that are common knowledge.

In one case where $c > \pi_3$, agent 3 irresponsibly chooses not to retaliate to obtain a zero payoff in retaliation subgames. Player j anticipates no retaliation and thus chooses to defect. Defection is still the best response for player i who is unaffected by agent 3's behavior. In the other case where $\pi_3 > c$, agent 3 retaliates, so j anticipates a penalty. With sufficiently high values such that $\rho > e$ and $\rho > a$, player j is induced to cooperate. Player i faces similar incentives to defect, so we get a SPNE with different choices for focal nodes: i defects, j cooperates, and agent 3 retaliates following j's defections.

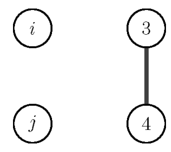
In contrast to ω_1, the network ω_2 induces some cooperative behavior under the right conditions. However, we can interpret this outcome as an extortion equilibrium because i takes advantage of j as evidenced by their corresponding equilibrium payoffs of $\{U_d, U_c\}$.

2.3 Normal-form representation of networked cooperation games

The above examples already demonstrate that distinct networks can affect the number of participants, payoffs, and the nature of the game itself (from simultaneous to sequential choices). However, we are



(a) Empty network ω_ϵ



(b) Network ω_3 with extraneous ties

Figure 4: Not all networks change the baseline cooperation game

barely scratching the surface because despite the small society size, there are 64 possible networks in the set of states Ω_4 per note 8. It's plausible that other network configurations will generate cooperative behavior, but not always. Consider, for example, the networks shown in Figure 4. The empty network ω_ϵ has a null effect on the original interaction. With only two participants–the original players–we replicate the mutual defection equilibrium of a *Coop*$\langle i, j\rangle$ game. Moreover, not all existing patterns of connections matter because we get a similar equilibrium outcome under network ω_3 as for the original interaction.

In the end, whether or not we continue to expect mutual defection becomes an empirical question requiring exact knowledge of the underlying social network. Although extensive-form game representations clearly reveal differences from applicable network transformations, this is an impractical approach beyond fairly small societies. As a less cumbersome alternative, I next derive a reduced normal-form representation that can readily map collaborative synergy and solidary retaliation transformations onto the original 2x2 normal form matrix for any society size. This reduced form is motivated by reconsideration of *focal nodes'* beliefs about potential retaliation.

2.3.1 How relevant is social solidarity?

Although possible in small societies, it's unreasonable to expect that any focal node knows whether other agents–including own alters–are willing retaliators.[12] This situation prompts major concerns about the social relevance of solidary retaliation, which warrants further explanation. To address this concern, I start from the most problematic epistemic scenario for focal nodes. Suppose that focal nodes find themselves in a situation of incomplete information regarding private retaliation benefits for linked alters. However, it's common knowledge that these private types are independently drawn from the same density function $f(v)$ with support $[0, v_{max}]$. This separate knowledge

12 In fact, there is evidence that solidarity does not guarantee desired participation. See Opp (2012) and Abell (1989).

can result from long-term observations in a given society regarding the actual manifestation of solidary behavior and proclaimed benefits.

This scenario is amenable to the formulation of a sequential Bayesian game that generates the following equilibrium strategies that we can carry forward as exogenous contextual conditions. Given singleton information sets from observable histories, each activated alter-type v_k chooses $r_k = 1$ if $v_k \geqslant c$. Focal nodes need to incorporate this unknown retaliation when deciding whether to cooperate. To that effect, focal nodes can use the common prior distribution to readily calculate a variable $\alpha = \Pr(v \geqslant c) = 1 - \int_0^c f(v)dv$, to form beliefs about potential retaliation from any private type v_k. Because payoffs are ultimately affected by accumulated penalties, focal nodes can derive the expected value of potential retaliation as a function of a contextual factor α as follows:

$$\hat{R}_i(\omega) = \rho \sum_{k \in A_i^\omega} \hat{r}_k = \rho \cdot (\alpha n_i^A) \tag{1}$$

$$\hat{R}_j(\omega) = \rho \sum_{k \in A_j^\omega} \hat{r}_k = \rho \cdot (\alpha n_j^A) \tag{2}$$

Expected retaliation will thus determine whether focal nodes cooperate, depending on the underlying social context that now includes the social relevance of a logic of solidarity.[13] First, in societies where $v_{max} \leqslant c$, solidary retaliation is simply impossible, so focal players won't be deterred by potential retaliation. Given a corresponding $\alpha = 0$, they will both choose to defect. Second, if solidarity is somewhat relevant such that $\alpha > 0$, whether focal nodes cooperate will still depend on the combination of individual penalties ρ and alter network sizes n_i^A and n_j^A. If any of these values is low, expected retaliation might not be high enough to deter defection.

2.3.2 Collecting network effects

The piecemeal approach thus far enables separate identification of distinct network effects, which is a practical modeling strategy to gradually accumulate the many impacts that networks can have on social dilemmas. Once defined, I summarize the overall impact with a more general formulation that embeds *Coop*$\langle i, j\rangle$ games within a particular social context and its applicable effects. In particular, my formulation below will generate a simpler depiction of networked cooperation games in terms of normal rather than extensive-form representations.

Definition 4. Let G be the set of all possible *Coop*$\langle i, j\rangle$ games in a society, with elements distinguished by the focal dyad $\{i, j\}$. A **networked**

13 I revisit the uncertain nature of this contextual parameter in concluding comments regarding extensions to this study.

		j	
		C	D
i	C	$M_c(1+\omega_{ij}\cdot b_\omega),\ \ M_c(1+\omega_{ij}\cdot b_\omega)$	$U_c,\ \ U_d-\rho\alpha n_i^A$
	D	$U_d-\rho\alpha n_j^A,\ \ U_c$	$M_d-\rho\alpha n_j^A,\ \ M_d-\rho\alpha n_i^A$

Figure 5: Reduced-form networked cooperation game generated by $\phi(\{i,j\},\omega,\alpha)$

cooperation game generator is a function $\phi : G \times \Omega_n \times \alpha \mapsto \Gamma$ that applies collaborative synergy and solidary retaliation transformations to a *Coop*$\langle i,j\rangle$ game. A networked cooperation game $\gamma \in \Gamma$ includes the original players and strategies along with context-specific payoffs as shown in Figure 5.[14]

The set Γ is restricted to include only those games that result from applying network transformations to a *Coop*$\langle i,j\rangle$ game, so $\phi(\cdot)$ covers this whole range by design. However, we can quickly show that ϕ will not be a one-to-one mapping by reviewing previous examples, holding constant the value of α. Let γ_{base} denote the original *Coop*$\langle i,j\rangle$ game in Figure 1. This outcome occurs whenever focal nodes are isolates, which is possible under multiple network configurations such as ω_ϵ and ω_3 in Figure 4 (the only possibilities in a society with four actors). Here, both $\phi(\{i,j\},w_\epsilon,\alpha)$ and $\phi(\{i,j\},\omega_3,\alpha)$ map onto the same element $\gamma_{base} \in \Gamma$.

Before closing this section, it's important to note that $\phi(\cdot)$ wraps a network around a *particular* dyad. We have seen that $\phi(\{i,j\},\omega_3,\alpha) = \gamma_{base}$, but this same network produces a different outcome for other players. If agents 3 and 4 were the focal nodes, then $\phi(\{3,4\},\omega_3,\alpha)$ generates a different networked game equivalent to Figure 2b.

3 NETWORKED COOPERATION WITH MANIFEST NETWORKS

Suppose now that the underlying network ω is common knowledge, an epistemic scenario of manifest networks. Given an interest on cooperative behavior, I proceed to derive conditions that reverse the dominant strategies of a *Coop*$\langle i,j\rangle$ game. To that effect, let $R_{min}(\omega) =$

14 To simplify the analysis in following sections, this reduced form does not show the payoffs of retaliators because their equilibrium behavior is already accounted for through the contextual parameter α. Their background behavior will generate payoffs according to the detailed solidary retaliation transformation on page 13.

$\max(e, a)$ be the minimum penalty that makes cooperation a dominant strategy under all ω_{ij} values. With that retaliation threshold, we can derive a minimum number n^A_{min} that satisfies $R_{min} = \alpha\rho n^A_{min}$ per equations 1 and 2. Solving for n^A_{min}, and rounding up to obtain a sufficiently high integer value, we get the following *cooperation threshold* in terms of alter set sizes:

$$n^A_{min} = \left\lceil \frac{R_{min}}{\alpha\rho} \right\rceil \tag{3}$$

With this cooperation threshold, we can readily identify the following **Manifest Networks Nash equilibrium**:

- If $n^A_j \geqslant n^A_{min}$ and $n^A_i \geqslant n^A_{min}$, then $\{s^*_i, s^*_j\} = \{C, C\}$
- If $n^A_j \geqslant n^A_{min}$ and $n^A_i < n^A_{min}$, then $\{s^*_i, s^*_j\} = \{C, D\}$
- If $n^A_j < n^A_{min}$ and $n^A_i \geqslant n^A_{min}$, then $\{s^*_i, s^*_j\} = \{D, C\}$
- Otherwise, $\{s^*_i, s^*_j\} = \{D, D\}$.

Note that this equilibrium depends on network-related variables, α and ρ through the n^A_{min}cooperation threshold. This threshold goes to infinity if either parameter approaches zero. But these parameters do not have to be that low to preclude cooperation. If the value of equation 3 exceeds $n - 2$, then mutual cooperation is impossible even if the network were fully connected because that society is not sufficiently large to produce requisite penalties. This situation occurs if either solidary retaliation is not socially relevant (low values of α) or defection penalties ρ are too small relative to deviation incentives e and a.

Assuming $n^A_{min}(\alpha, \rho) < n - 2$, then we can visualize the equilibrium as a function of known alter set sizes, as shown in Figure 6. The underlying context reflected by α and ρ is still relevant because higher values decrease the cooperation threshold for both players, thus increasing the prospects for mutual cooperation. Depending on the actual location of the cooperation threshold, it is now possible for all outcomes of the *Coop*$\langle i, j\rangle$ game to be in equilibrium. Note that when the network is common knowledge, equilibrium behavior will depend on the relative size of alter sets; and other things equal, similar behavior will require similar social structures. As n^A_{min} gets close to $n - 2$, the focal nodes will need to have similar numbers to encourage cooperation. It's possible that the focal nodes are similarly disconnected, in which case they will not be deterred from mutual defection. With uneven number of ties, the behavioral symmetry disappears: it is more likely that there will be an extortion equilibrium that favors the player with more connections.

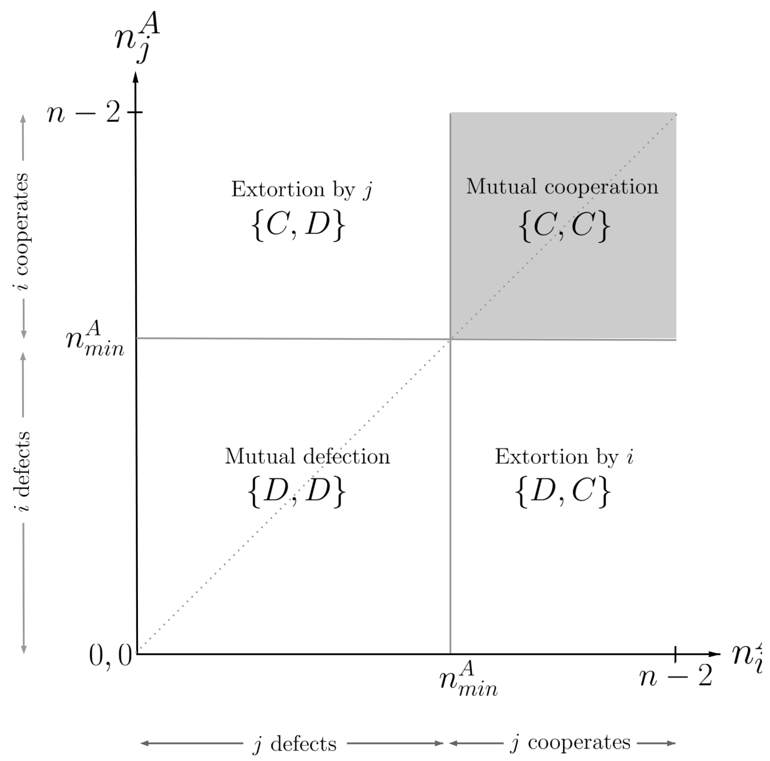


Figure 6: Nash equilibrium with known networks as a function of cooperation thresholds

4 EQUILIBRIUM BEHAVIOR UNDER AGGREGATE NETWORK UNCERTAINTY

We explored before how a manifest network ω can alter a $Coop\langle i, j\rangle$ game by introducing known collaborative synergy and solidary retaliation effects. We have not seen, however, how the network gets established in the first place. Neither have we inquired how participants learn requisite social context information. To investigate these concerns, it will be helpful to reconsider the underlying sequence behind networked cooperation in terms of this multi-stage process:

1. *Social Context Determination* stage: In a preliminary stage, a non-strategic actor Nature chooses ω and α. The value α is revealed to all members of society, so it will be common knowledge.[15] However, knowledge of ω will vary under several epistemic conditions that will be defined below.

15 Technically, Nature draws retaliator types using the density function $f(\nu)$. But agents can readily derive–as if it were revealed– a summary variable α equal to the probability of individual retaliation, which can be treated as a contextual variable that is commonly known by all participants.

2. *Cooperation* stage: Next, players i and j decide whether to cooperate within the corresponding networked cooperation game $\phi(\{i, j\}, \omega, \alpha)$. If both focal nodes have empty alter sets, the game ends; otherwise, it proceeds on to the next stage.

3. *Solidary Retaliation* stage: Linked alters from the rest of society decide whether to retaliate. Their decisions end the game.

By construction, our networked cooperation mapping $\phi(\{i, j\}, \omega, \alpha)$ condenses the cooperation and retaliation stages into a reduced-form game that already incorporates equilibrium beliefs about the optimal behavior of potential retaliator types. We can therefore focus attention on further specifications of the preliminary stage that incorporate possible scenarios where agents might not be fully informed about the underlying network ω. To that effect, I first discuss connections between alter sets and the underlying network to motivate plausible epistemic scenarios. I proceed to discuss how agents handle network uncertainty in order to form beliefs about opponents' types. Finally, I present a specialized Bayesian Nash Equilibrium notion that requires separate beliefs to estimate unknown features of the underlying social network ω.

4.1 Network Generation Processes

Solidary retaliation enlarges the set of players by activating the focal nodes' respective alter sets, A_i^ω and A_j^ω. By definition, this participation is mediated by some ω, so players require network information to anticipate potential retaliation. If the network is not common knowledge, then it will be crucial for players to have information about how the network was established, which could then be used to forecast retaliation. This informational requirement invites consideration of network generating processes, which could be fairly specialized to particular societies. I model this potential heterogeneity with a random process that defines a probability measure on the set of social states Ω_n as follows:

Definition 5 (Network Generation Process or NGP)**.** A **NGP** for a society of size n is a vector of $\binom{n}{2}$ probabilities $\{\theta_{ij}\}_{i \neq j \in I_n}$ that generates values for non-diagonal entries of a network (adjacency matrix) ω, where $\theta_{ij} = \Pr(\omega_{ij} = 1)$.[16]

Some readers might question that the unstructured collection in Definition 5 is both too general and impractical, by requiring a number of parameters that increases exponentially with the size of society. While acknowledging these concerns, this level of generality allows

16 Unspecified diagonal entries ω_{ii} have a constant zero value.

me to make three points about the general difficulties of reasoning about unknown networks.

First, agents indeed face a formidable task when they lack network information if they need to reason about fairly complicated NGP processes. Second, it's not that analysts lack relevant information about how people go about forming ties, but network formation mechanisms vary widely across applications and disciplines. And it's one thing to incorporate known mechanisms and another for agents to make decisions under conditions of network uncertainty. For example, models of preference attachment that are popular in the study of large networks (implicitly) assume that new nodes know the existing structure or otherwise have requisite knowledge to readily identify highly connected nodes (Barabasi and Albert, 1999). It's neither clear how agents might acquire requisite knowledge in all cases, nor is it clear whether this or other particular NGPs would apply to all social dilemmas.

Third, it is indeed possible to assume additional NGP structure to incorporate substantive knowledge such as assortative connections (homophily) and other network formation mechanisms (Lusher and Robins, 2013). In fact, it is common practice–especially involving statistical analysis–to rely instead on joint probability models that impose additional structure to generate networks with a smaller number of parameters. Besides node attributes, it's possible that new ties are conditioned on existing network structure (Lusher et al., 2013). However, some but not all statistical or generative models have a clear behavioral foundation (Jackson, 2008b).

All in all, we lack a baseline behavioral model to examine reasoning under network uncertainty under general conditions. To move forward, I advance below two reasons to develop a baseline around a specialized NGP where all dyadic probabilities are identical.

Definition 6 (NGP with identical, independent ties)**.** A **NGP** (θ) process generates independent ties if it satisfies these two properties: (1) independent ties: $\Pr(\theta_{ij}|\theta_{ik}) = \Pr(\theta_{ij}) \cdot \Pr(\theta_{ik})$ for all distinct $i, j, k \in I$; and (2) identical tie probabilities, $\Pr(\theta_{ij}) = \theta$ for all distinct dyads in I_n.[17]

4.2 What network knowledge do players have?

I make two general assumptions that hold for the rest of the paper.

Epistemic Assumption 1 (Private network knowledge). All nodes $x \in I_n$ know their corresponding row $\omega_{x.}$ in ω.

17 As defined, this NGP(θ) is also known as an Erdős-Rényi random graph model, where each possible tie is drawn from a Bernoulli distribution parameterized by $\theta \in [0, 1]$.

Private network knowledge is a basic condition that will generally hold under all epistemic scenarios.[18] In fact, this assumption is automatically satisfied when the network is common knowledge. And it's otherwise reasonable to think that agents will know their neighbors even if they lack knowledge of extraneous dyads.

Network uncertainty in this paper is therefore residual, extending beyond an agent's private network knowledge. The second epistemic assumption handles that residual uncertainty by assuming that all agents know the family of NGP processes that generated the unknown network ω that underlies a society. In particular, agents will hold similar *qualitative* beliefs that all ties are independent random variables. But, as will be explained below, they need not have identical *quantitative* beliefs regarding the true parameter θ.

Epistemic Assumption 2 (General NGP(θ) knowledge). It is common knowledge that ω was derived from the family of single-paramter NGP(θ) processes.

This second epistemic assumption is defensible for several reasons. First, not only is it easier for agents to reason about one dyad at a time, but it is arguably more natural than thinking about more complicated structures involving collections of dyads–including the whole network–at a time. Reasoning about networks in terms of dyads is consistent with the presentation of dyads as the unit of analysis to determine social context. Second, even if (some) agents had the capacity to do more complicated calculations involving other social structures, we could argue that decisions in *Coop*$\langle i, j\rangle$ games could not be delayed to perform required calculations. As either a time-saving or heuristic device, agents could simply try to estimate particular ties.

4.3 Network beliefs and player types

By epistemic assumption 2, focal nodes have an underlying "network belief" NGP(θ) that all ties are independent and identically distributed (i.i.d.) random variables drawn from some Bernoulli(θ) distribution. These beliefs are useful to reconstruct the rest of the network one tie at a time provided that agents know (or can estimate) the key parameter θ.

Whether agents need to reconstruct the whole network will depend on the underlying strategic interaction. In particular, the networked cooperation games in this paper entail two effects that operate through direct connections. These localized effects greatly simplify

18 I will later address this epistemic condition in terms of an agent x knowing her own connections, thus the "private network" label. But, in fact, x also knows that she is not connected to some other agent y if $\omega_{xy} = 0$. Acemoglu et al (2011) and Lobel and Sadler (2015) make similar assumptions for what they call neighborhoods or network topologies.

the task for focal nodes who only need to estimate their opponents' alter sets, rather than the whole network. These alter sets A_i^ω and A_j^ω are effectively the players' unknown types, each representing a collection of connected dyads.

Under general conditions, agents would use their NGP beliefs to predict particular ties between an opponent and other actors. The nature of solidary retaliation lends itself, however, to an additive formulation that facilitates reasoning for the focal nodes. In practical terms, focal nodes can derive beliefs about the number of alters that other players have, regardless of their particular identities. Although these numbers depend on knowledge of alter sets, it will be more convenient to substitute a set-theoretic formulation with a derived type space based on the size of alter sets.

By definition, both n_i^A and n_j^A will take on an integer value from a set $T = \{0, 1, ..., n-2\}$. Specifically, given subjective beliefs θ_i and θ_j, agents can reason that alter set sizes will correspond to i.i.d. random variables with the following conditional probability distributions:[19]

$$B_i(n_j^A|\theta_i, n_{-ij}) \sim \mathrm{Binomial}(n_{-ij}, \theta_i) \tag{4}$$

$$B_j(n_i^A|\theta_j, n_{-ij}) \sim \mathrm{Binomial}(n_{-ij}, \theta_j) \tag{5}$$

These derived beliefs mean that the type of focal nodes can be estimated as the number of "successes" (actual ties) in $n_{-ij} = n-2$ trials, where the probability of success is either θ_i or θ_j. For player i, the expected alter size of j will be $\hat{n}_j^A(B_i) = \mathbb{E}[n_j^A|B_i(n_j^A|\theta_i, n_{-ij})] = \theta_i \cdot n_{-ij}$. The corresponding value for player j is $\hat{n}_i^A(B_j) = \mathbb{E}[n_i^A|B_j(n_i^A|\theta_j, n_{-ij})] = \theta_j \cdot n_{-ij}$.

4.4 Network Estimation Bayesian Equilibrium (NEBE)

Networked cooperation under aggregate network uncertainty derives into a strategic interaction with the following components:

- Two players: $i, j \in I_n$
- a common action space $A = \{1, 0\}$ for both players, where 1 stands for cooperation.
- Private types n_i^A and n_j^A drawn from a common type space $T = \{0, 1, ..., n-2\}$
- Alter set beliefs $B_i(n_j^A|\theta_i, n_{-ij}), B_j(n_i^A|\theta_j, n_{-ij})$ as defined in equations 4 and 5, respectively. If players receive information about

19 Note that these shown beliefs are not conditioned on a player's own type because the underlying NGPs also produce uncorrelated types.

the underlying NGP process, we will further require that they update their basic network beliefs θ_i and θ_j using Bayes' Rule.

- Payoffs $u_i, u_j : A \times T$ are shown in Figure 5.[20].

A strategy for each player collects cooperation choices from each type: $\sigma_i(n_i^A) \in A$ and $\sigma_j(n_j^A) \in A$, respectively. The basic task for players is to estimate unknown features of the underlying network ω using basic network beliefs regarding the underlying NGP. They then derive beliefs about alter sizes to maximize expected utilities for each player's private type as follows.

Definition 7. In the Networked Cooperation Game $\gamma(\{i, j\}, \omega, \alpha)$, a collection of strategies $\{(\sigma_i^*(n_i^A))_{n_i^A \in T}, (\sigma_j^*(n_j^A))_{n_j^A \in T}\}$ is a **Network Estimation Bayesian Equilibrium** (NEBE) if for each $n_i^A \in T$,

$$\sigma_i^*(n_i^A) = \arg\max_{a_i \in A_i} \sum_{n_j^A \in T} u_i(a_i, \sigma_j^*(n_j^A); n_i^A) \cdot B_i(n_j^A | \theta_i)$$

and for each $n_j^A \in T$,

$$\sigma_j^*(n_i^A) = \arg\max_{a_j \in A_j} \sum_{n_i^A \in T} u_j(a_j, \sigma_i^*(n_i^A); n_j^A) \cdot B_j(n_i^A | \theta_j)$$

As defined, $\gamma(\{i, j\}, \omega, \alpha)$ is a Bayesian game. By requiring player-types to maximize expected utility given their beliefs, NEBE defines a specialized Bayes Nash equilibrium that is guaranteed to exist, perhaps as a mixed strategy equilibrium. As was the case with manifest networks, contextual parameters affect possible equilibrium outcomes. If the cooperation threshold exceeds the maximum number of alters, $n - 2$, then the only possible NEBE is one where all types choose to defect. Notably, player's payoffs are unaffected by their own types, so we will be able to construct a NEBE where all types choose the same strategy.

Assuming that $n_{min}^A \leqslant n - 2$, I now construct a pooling strategy equilibrium based on the cooperation threshold n_{min}^A. To that effect, let $\mathbb{1}[E]$ be an indicator function that equals 1 when the argument E is true, and consider the strategy profile σ_{pooled} consisting of $(\sigma_i(n_i^A) = \mathbb{1}[\hat{n}_j^A(B_i) \geqslant n_{min}^A])_{n_i^A \in T}$ and $(\sigma_j(n_j^A) = \mathbb{1}[\hat{n}_i^A(B_j) \geqslant n_{min}^A])_{n_j^A \in T}$.

20 In general, payoffs can depend on all player-types. Typically, these payoffs only incorporate own payoffs, but here player's alter sets do not impact own payoffs. Instead, payoffs will depend on another player's types due to the nature of solidary retaliation. An alternative interpretation is consistent with typical assumptions if we think of the unknown retaliations as a player's own type. For example, player i's type would be defined by n_j^A, and vice versa for player j. Under this interpretation, however, focal nodes would not know their own types. However, they can still use the same beliefs to predict their own type.

Claim 1. The strategy profile σ_{pooled} is a Network Estimation Bayesian Equilibrium.

Proof. See appendix. □

I explore this pooling equilibrium below under different epistemic scenarios. To simplify the notation, I henceforth identify the collection of type-specific optimal strategies by a single pair of representative strategies for each player's respective types:

$$\{\sigma_i^*(n_i^A), \sigma_j^*(n_j^A)\} = \left\{\mathbb{1}\left[\hat{n}_j^A(B_i) \geqslant n_{min}^A\right], \mathbb{1}\left[\hat{n}_i^A(B_j) \geqslant n_{min}^A\right]\right\} \quad (6)$$

5 NETWORKED COOPERATION UNDER DIFFERENT EPISTEMIC SCENARIOS

I examine cooperative behavior under the following scenarios that gradually relax knowledge of the underlying network structure:

- Common NGP prior: NGP(θ) is common knowledge.
- Uninformative NGP priors: The actual NGP(θ) is not common knowledge, but agents have prior beliefs about θ.
- Uninformative NGP priors with κ-supplementary network knowledge: In addition to prior beliefs about θ, agents can learn a fraction κ of extraneous dyads beyond their private network knowledge.

5.1 NEBE with common priors

I previously assumed that participants knew the underlying network ω. That assumption might hold in fairly small societies, but not under more general conditions with variable network sizes. I examine here an epistemic scenario where the network itself is no longer common knowledge.

Consider a situation in which players know the NGP(θ) that generated that society's network, ω. This scenario captures a situation where there is ample historical evidence to reveal the overall connectivity or potential for (dyadic) ties in society, so the parameter θ is common knowledge.

We can readily substitute the common parameter $\theta = \theta_i = \theta_j$ into conditional belief equations 4 and 5 to derive identical estimates $\hat{n}_j^A(B_i) = \hat{n}_j^A(B_i) = \theta \cdot (n-2)$. In turn, these estimates can be plugged into the equilibrium strategy conditions of equation 6. Rearranging terms within the indicator functions results in this equivalent equilibrium formulation in terms of the known θ:

$$\{\sigma_i^*(n_i^A), \sigma_j^*(n_j^A)\} = \left\{\mathbb{1}\left[\theta \geqslant \frac{n_{min}^A}{n-2}\right], \mathbb{1}\left[\theta \geqslant \frac{n_{min}^A}{n-2}\right]\right\} \tag{7}$$

Under this epistemic scenario, we get a symmetric equilibrium where both players choose the same strategy. As was the case with manifest networks, if $n_{min}^A > (n-2)$ then there is insufficient retaliation within the whole network to induce cooperation, so we predict mutual defection. However, if n_{min}^A is not too large, then both players will have an incentive to cooperate in societies with a sufficiently high connectivity parameter such that $\theta \geqslant n_{min}^A/(n-2)$. If $n_{min}^A = n-2$, then only a complete network with density 1 will work. With low-density networks with few connections, reflected in a common prior θ below that ratio, then both focal nodes will defect.

5.2 Social Projection NEBE with uninformative priors

A more interesting case arises when the actual NGP(θ) is unknown. In this case, players will need to form beliefs about the underlying key parameter. This analysis will use a Beta-Binomial model to derive updated priors about θ from available information.

To this effect, I refine the preliminary *Social Context* stage with an additional step for Nature. Previously, I noted that Nature determined $\omega \in \Omega_n$ without further specification. I now model that state selection by letting Nature make a random draw of θ from a uniform distribution with support $[0, 1]$. This is equivalent to saying that $\theta \sim \mathrm{Beta}(1, 1)$, a case of uninformative priors because all possible values of θ are equally likely. Without additional information, the best prediction for players is to assume that for any extraneous tie ω_{xy}, with $x, y \notin \{i, j\}$, the events $\omega_{xy} = 0$ and $\omega_{xy} = 1$ are equally likely a priori: $\hat{\theta}_{i,0} = \hat{\theta}_{j,0} = \mathbb{E}(\theta) = 1/2$.

However, per the first epistemic assumption, any agent x gains $n^{NEW} = n-1$ pieces of information regarding the actual values of w_{xk} involving other actors $k \in I_n$. Suppose that this agent also sees t_x ties, which will be useful to get an updated belief $\theta_{x,1}$. Using Bayes' rule to update the $\mathrm{Beta}(1, 1)$ prior distribution with these new observations derives into a Beta-Binomial model that generates the following posterior beliefs about θ for any agent x:[21]

$$\theta_{x,1}|t_x, n^{NEW} \sim \mathrm{Beta}(t_i + 1, n^{NEW} - t_i + 1) \tag{8}$$

The posterior mean from the derived probability distribution 8 equals $\mathbb{E}[\theta_{x,1}|t_x, n^{NEW}] = (t_x+1)/(n^{NEW}+2)$. Focal nodes observe ω_{ij} in ad-

21 See Gelman et al. (2003), chapter 2.

dition to their respective n_i^A and n_j^A actual ties. They can therefore update their previous estimates as follows:

$$\hat{\theta}_{i,1} = \frac{t_i + 1}{n^{NEW} + 2} = \frac{\omega_{ij} + n_i^A + 1}{n + 1} \tag{9}$$

$$\hat{\theta}_{j,1} = \frac{t_j + 1}{n^{NEW} + 2} = \frac{\omega_{ij} + n_j^A + 1}{n + 1} \tag{10}$$

Substituting these posterior beliefs for a previous common prior results in this NEBE:

$$\{\sigma_i^*(n_i^A), \sigma_j^*(n_j^A)\} = \left\{ \mathbb{1}\left[\hat{\theta}_{j,1} \geqslant \frac{n_{min}^A}{n-2}\right], \mathbb{1}\left[\hat{\theta}_{j,1} \geqslant \frac{n_{min}^A}{n-2}\right] \right\} \tag{11}$$

Note that each representative player-type strategy depends on own connections. I call this outcome a "social projection" equilibrium because players rely on their own ties to estimate the underlying θ parameter. Also, in contrast to the case of common priors where dissimilar alter set sizes did not matter ($n_i^A \neq n_j^A$), this second epistemic scenario will introduce asymmetric responses that are more likely to result in extortion, depending on the required n_{min}^A. Other things equal, more connected players will produce higher estimates that encourage cooperation ($\partial\hat{\theta}_{i,1}/\partial n_i^A > 0$, $\partial\hat{\theta}_{j,1}/\partial n_j^A > 0$). In contrast, less connected players will tend to think that opponents lack enough solidary retaliators to preclude defection, and act accordingly with their own defection.

5.3 NEBE with supplementary network information

I now examine a scenario in which players obtain additional information about the rest of the network. I parameterize this extra knowledge with a variable $\kappa \in [0, 1]$ that enables some player y to see the connections of a subset $K_y(\omega) \subset I_{n-ij}$ with $K_y^{obs} = \kappa(n-2)$ agents.[22] For example, player i can readily check how many elements of $K_i(\omega)$ have ties to player j, a quantity denoted by $n_j^{K_i}$. This scenario entails availability of a separate mechanism or space that enables agents to make additional observations. For example, players i and j might interact in a densely populated area where they can readily confirm some extraneous ties (or lack thereof). The case of $\kappa = 0$ is a special case where that extraneous information is unavailable.

Players will update prior beliefs similarly to the previous epistemic scenario. We just need to account for relevant changes in acquired

22 The variable κ is continuous, so the resulting set size K_y^{obs} is rounded up to the nearest integer.

information. Player i now gains access to $n_{i,\kappa}^{NEW} = n - 1 + K_i^{obs}$ observations that result in $t_{i,\kappa} = \omega_{ij} + n_i^A + n_j^{K_i}$ manifest ties. Player i still has to estimate a residual number of ties among $K_i^{rest} = n - 2 - K_i^{obs}$ agents not included in $K_i(\omega)$. Player i can estimate this residual quantity by pooling all available information to derive this new posterior distribution: $\theta_{i,1}|t_i \sim Beta(t_{i,\kappa} + 1, n_\kappa^{NEW} + 2)$ with posterior mean $(t_{i,\kappa} + 1)/(n_{i,\kappa}^{NEW} + 2)$. Conditional on known quantities, player i's final estimate for j's total number of connections will then be:

$$\hat{n}_j^A(K_i(\omega)) = n_j^{K_i} + \hat{\theta}_{i,1} \cdot K_i^{rest} \tag{12}$$

Player j incurs similar calculations to derive an estimate $\hat{\theta}_{j,1} = (t_{j,\kappa} + 1)/(n_\kappa^{NEW} + 2)$. Her final estimate regarding i's alter set is:

$$\hat{n}_i^A(K_j(\omega)) = n_i^{K_j} + \hat{\theta}_{j,1} \cdot K_j^{rest} \tag{13}$$

This time we characterize the equilibrium in terms of these estimates as follows:

$$\{\sigma_i^*, \sigma_j^*\} = \left\{ \mathbb{1}\left[\hat{n}_i^A(K_i(\omega)) \geqslant n_{min}^A\right], \mathbb{1}\left[\hat{n}_i^A(K_j(\omega)) \geqslant n_{min}^A\right] \right\} \tag{14}$$

If $K_i^{rest} = 0$, then player i just needs to compare known $n_j^{K_i}$ to the threshold n_{min}^A as in the case of manifest networks. Similarly, If $K_j^{rest} = 0$, then player j just compares $n_i^{K_j}$ to the threshold to decide whether to cooperate.

Beyond the case of manifest networks, supplementary network information reduces uncertainty, so players will be better positioned to assess the credibility of solidary retaliation. Whether or not players cooperate will largely depend on the actual connectivity of the social network. In highly connected societies, players are more likely to find opponents with sufficiently large alter set sizes, so more knowledge will confirm the need to cooperate to avoid punishment. Under sparse networks, however, players will readily know that they can get away with defection.

5.4 Discussion on the impact of manifest and unknown networks

I discuss here the implications of different epistemic scenarios, starting with a quick review of networked cooperation with manifest networks. In that setting, the actual NGP(θ) was irrelevant for strategic purposes because players possessed all relevant information. However, the underlying state does impact the possibility of cooperative outcomes with important lessons for our understanding of social dilemmas. Given a known NGP(θ), we can derive the conditional probability of individual cooperation as $p_c = Pr(\sigma_x^* = 1|n_{min}^A) = Pr(B_x(n -$

$2, \theta) \geqslant n^A_{min})$.[23]. Since alter sizes are independent, the joint probability of mutual cooperation therefore equals p_c^2. Similarly, the joint probability of unilateral cooperation and defection that occurs with mismatched choices equals $2 \cdot p_c \cdot p_d$. Finally, both players defect with joint probability p_d^2.

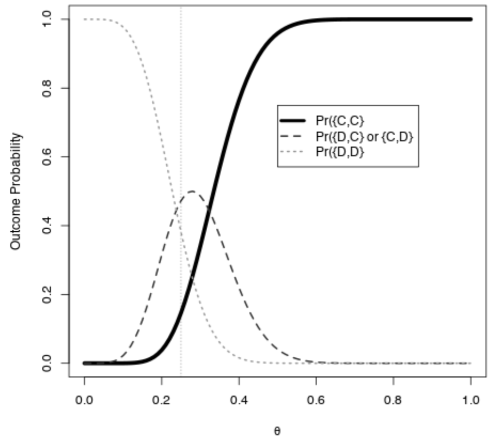
Figure 7 illustrates two general implications for probable equilibria with three sample scenarios from a society where two focal nodes co-exist with 20 additional agents.[24] First, as we move across panels, we see an intuitive result that higher thresholds require more connected societies (with correspondingly higher θ's). Second, a more interesting result stems from the vertical lines that identify θ values that would generate expected alter sizes equal to the applicable threshold. In general, these enabling θ values do not guarantee mutual cooperation because that separate function requires more connectivity for optimal outcomes to be the only option. What is more likely to happen within a wide neighborhood surrounding enabling θ values is for one of the players to defect while the other cooperates (this outcome is reflected by the higher probability for either $\{C, D\}$or$\{D, C\}$ outcomes). These observations reiterate the notion that networks matter, but closer attention to the underlying NGP further reveals that actual connectivity will condition which possible outcomes are most likely to occur.

The impact of aggregate network uncertainty depends on particular epistemic conditions. In the first scenario, which is commonly assumed in the social learning literature, uncertainty need not be an impediment to cooperation. As long as agents know the common prior, they will be incentivized to cooperate if associated beliefs suggest sufficient connectivity. In fact, agents will ignore their own private knowledge because their own ties might actually mislead them by suggesting fewer or more connections than would be expected from common priors. I relaxed this assumption in subsequent scenarios where players could disagree on regarding underlying NGP beliefs. The implications of these scenarios is illustrated by Figure 8 that plots belief updating equations for the same sample society of two focal players and 20 additional aters.

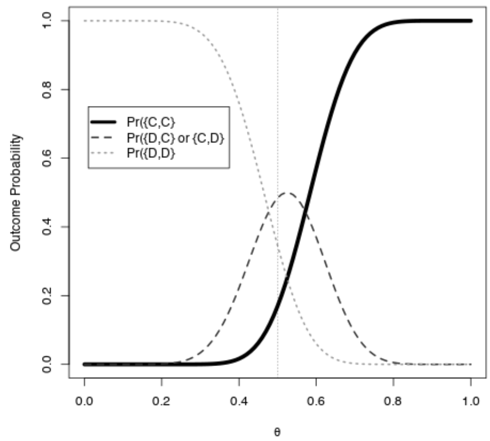
This figure shows posterior beliefs as a function of supplementary network knowledge. The line for $\kappa = 0$ corresponds to revised beliefs that underlie the social projection NEBE based on private network knowledge of $n - 1$ ties. As network knowledge increases, there is the possibility of observing up to $2n - 3$ dyads, so the slope of belief update functions changes accordingly, becoming less steep as shown

23 $\Pr(B_x(n-2, \theta) \geqslant n^A_{min})$ itself equals $1 - \sum_{k=0}^{n^A_{min}-1} \binom{n-2}{k} \theta^k (1-\theta)^{n-2-k}$.

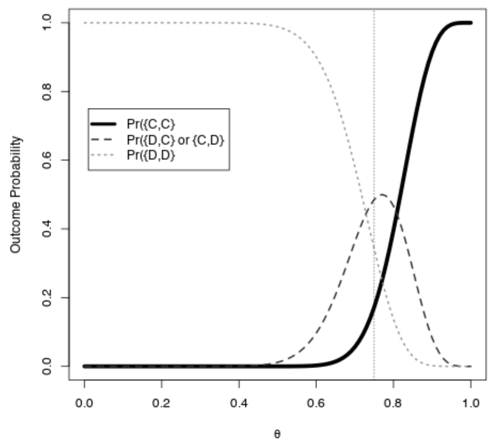
24 Each panel corresponds to a different cooperation threshold. The horizontal axis captures the overall connectivity in society as determined by the parameter θ. The vertical axis measures the conditional probabilities of various outcomes.



(a) $n^{A}_{min} = 5$



(b) $n^{A}_{min} = 10$



(c) $n^{A}_{min} = 15$

Figure 7: Outcome Probabilities with common NGP(θ) priors that vary along the horizontal axis. Each panel shows a different threshold. The darkest curve tracks the probability of mutual cooperation.

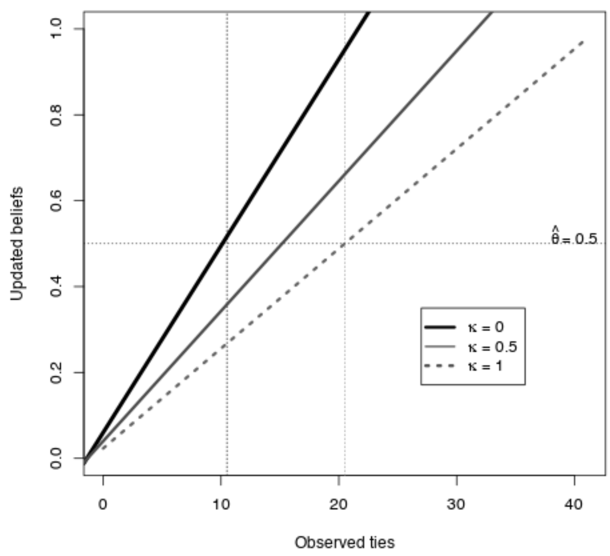


Figure 8: Network belief updates as a function of partial network knowledge that varies along the horizontal axis. Darker lines represent partial knowledge. The dotted line indicates complete network knowledge.

for the cases of $\kappa = 0.5$ and $\kappa = 1$.[25] Each new observed tie receives a lower weight for belief updating purposes.

Overall, this pattern indicates that more knowledge will generally make it more difficult for agents to produce high θ values when both players have similar partial information, a common κ. Reducing network uncertainty thus inhibits cooperation. However, we can also incorporate some heterogeneity where agents differ in terms of κ values. Under more general epitesmic conditions, we would again expect more informed players to be less cooperative (unless they operate in highly connected societies). In contrast, less informed agents will be more sensitive to any observed ties–especially their own–and thus more readily form beliefs that encourage cooperative behavior. These informational asymmetries in sparse networks will generate extortion equilibria where more informed players take advantage of less informed ones. As was the case with variable connections under manifest networks, variable network knowledge will itself have distributional consequences.

25 This extreme case with $\kappa = 1$ means that an agent already has complete knowledge of the actual state ω, so derived posterior beliefs do not affect strategic decisions. Regardless, this information improves estimates of the actual NGP that generated the observed network ω.

6 CONCLUSION

This paper examined the impact of manifest and unknown networks on social dilemmas through the lens of cooperation games. I developed a general framework to embed these games within a network structure that alters the original interaction with revised payoffs and more players induced by networked-mediated collaboration gains and penalties. I first examined the impact of networks when these are common knowledge, finding results that deviate from conventional analyses of cooperation insofar as social networks can have both positive and negative consequences. One positive consequence was the increase in social utility with respect to a situation of mutual defection. However, a negative consequence is that social networks add structural heterogeneity that allows connected players to exploit disconnected players.

I then examined the consequences of network uncertainty with a Bayesian game formulation that motivated a Network Exploration Bayesian Equilibrium (NEBE) notion requiring players to make optimal decisions based on beliefs regarding the underlying network generating processes that created unknown networks. In general, I found that network uncertainty is not an impediment to cooperation. To the contrary, more knowledgeable actors can find it more difficult to cooperate if they realize that they operate in societies that have sparse networks. I also found that network uncertainty raises equity concerns regarding the distribution of potential gains because it allows connected players with limited network knowledge to be exploited by more knowledgeable actors.

The results discussed in this paper are specific to a particular social dilemma game subject to specific network transformations. However, future work can readily extend this study in many fruitful directions. For example, we could further explore the consequences of relaxing the assumption that collaborative synergy cannot guarantee cooperation on its own, which introduces novel strategic considerations that include multiple equilibria. Additionally, the approach in this paper that collects multiple effects in algorithmic fashion can be readily extended to other social dilemmas beyond free-riding that incorporate other network effects.

Another relevant extension entails modeling repeated networked interactions to directly engage the literature on repeated games that suggests other conditions to induce cooperation. Such an extension would require consideration regarding the collective memory of networks and how these transmit useful information to incentivize ongoing cooperation. Additionally, adding dynamics enables reconsideration of the contextual parameter α that captures the social relevance of solidarity retaliation. With repeated interactions, there might be opportunities to update beliefs about extant solidarity itself, thus paving the way for an enhancing understanding of the impact of networks on social dilemmas that addresses two distinct types of uncertainty: (1) what is the actual network (the subject of this paper); and (2) how networks actually shape behavior. That is, network mechanisms need not be common knowledge, so how can players know the exact networked game they are playing? How do they know whether certain norms will apply if these are mediated by unknown networks?

REFERENCES

Abell P. (1989) Games in networks: A sociological theory of voluntary associations. *Rationality and society* 1(2): 259-282.

Acemoglu D, Dahleh MA, Lobel I, et al. (2011) Bayesian learning in social networks. *The Review of Economic Studies* 78(4): 1201-1236.

Ahn TK, Esarey J and Scholz JT. (2009) Reputation and Cooperation in Voluntary Exchanges. *The Journal of Politics* 71(02): 398-413.

Axelrod RM. (1984) *The evolution of cooperation*, New York: Basic Books.

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

Berardo R and Scholz JT. (2010) Self-organizing policy networks. *American Journal of Political Science* 54(3): 632-649.

Brañas-Garza P, Cobo-Reyes R, Espinosa MP, et al. (2010) Altruism and social integration. *Games and Economic Behavior* 69(2): 249-257.

Buchanan DR. (2019) Public Health Interventions. *The Oxford Handbook of Public Health Ethics*: 77.

Burt R. (2000) The network structure of social capital. In: Staw BM and Sutton RI (eds) *Research in Organizational Behavior*. New York: Elsevier, 345-423.

Chamley CP. (2004) *Rational Herds*, Cambridge, UK: Cambridge University Press.

Chwe MS. (1999) Structure and strategy in collective action. *American Journal of Sociology* 105(1): 128-156.

Chwe MS. (2000) Communication and Coordination in Social Networks. *Review of Economic Studies* 67: 1-16.

Coleman JS. (1990) *Foundations of social theory*, Cambridge, MA: Harvard University Press, 304 and 590-600.

De Deken JJ, Ponds E and Van Riel B. (2006) Social Solidarity. In: Clark GL, Munnell AH, Williams K, et al. (eds) *The Oxford handbook of pensions and retirement income*. Oxford: Oxford Handbooks, 141-159.

Elster J. (2007) *Explaining social behavior*, Cambridge; New York: Cambridge University Press, chapter 22.

Fehr E and Fischbacher U. (2002) Why social preferences matter-the impact of non-selfish motives on competition, cooperation and incentives. *Economic Journal*: 1-33.

Fudenberg D and Tirole J. (1991) *Game theory*, Cambridge, MA: MIT Press, 9-10.

Gelman A, Carlin JB, Stern HS, et al. (2003) *Bayesian Data Analysis*, Boca Raton: Chapman & Hall/CRC.

Gould RV. (1993) Collective action and network structure. *American Sociological Review*: 182-196.

Goyal S. (2007) *Connections*, Princeton: Princeton University Press.

Granovetter M. (2005) The impact of social structure on economic outcomes. *Journal of economic perspectives* 19(1): 33-50.

Hechter M. (2015) Sociology of Solidarity. *International Encyclopedia of the Social & Behavioral Sciences*: Second Edition. Elsevier Inc., 6-9.

Herreros F. (2004) *The problem of forming social capital: why trust?*, New York: Palgrave Macmillan.

Hoff P, Raftery A and Handcock M. (2002) Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97(460): 1090-1098.

Jackson MO. (2008a) *Social and economic networks*, Princeton, NJ: Princeton University Press.

Jackson MO. (2008b) Observing and measuring social interaction. *Social and economic networks*. Princeton, NJ: Princeton University Press, 434-458.

Keefer P and Knack S. (2005) Social capital, social norms, and the new institutional economics. In: Menard C and Shirley MM (eds) *Handbook of New Institutional Economics*. Dordrecht, Netherlands: Springer, 701-726.

Knack S. (1992) Civic norms, social sanctions, and voter turnout. *Rationality and society* 4(2): 133-156.

Komter AE. (2005) *Social solidarity and the gift*: Cambridge University Press.

Lin N. (2001) *Social capital: a theory of social structure and action*, New York: Cambridge University Press.

Lobel I and Sadler E. (2015) Information diffusion in networks through social learning. *Theoretical Economics* 10(3): 807-851.

Lusher D, Koskinen J and Robins G. (2013) *Exponential random graph models for social networks*, New York: Cambridge University Press.

Lusher D and Robins G. (2013) Formation of Social Network Structure. In: Lusher D, Koskinen J and Robins G (eds) *Exponential random graph models for social networks*. New York: Cambridge University Press, 16-28.

Manski CF. (2000) Economic analysis of social interactions. *The Journal of Economic Perspectives*: 115-136.

Olson M. (1965) *The logic of collective action*, Cambridge, MA: Harvard University Press.

Opp K-D. (2012) Collective identity, rationality and collective political action. *Rationality and society* 24(1): 73-105.

Osborne MJ and Rubinstein A. (1994) *A course in game theory*, Cambridge, MA: MIT Press, 73-74.

Ostrom E. (1990) *Governing the commons*, Cambridge, UK; New York: Cambridge University Press.

Ostrom E. (1998) A behavioral approach to the rational choice theory of collective action. *American political science review* 92(1): 1-22.

Ostrom E, Gardner R and Walker J. (1994) *Rules, games, and common-pool resources*, Ann Arbor: University of Michigan Press.

Perry BL, Pescosolido BA and Borgatti SP. (2018) *Egocentric network analysis*, Cambridge, UK: Cambridge University Press.

Pescosolido BA. (2007) The Sociology of Social Networks. In: Bryant CD and Peck DL (eds) *21st Century Sociology: A Reference Handbook*. Thousand Oaks, CA: Sage Publications, Inc., 208-217.

Raub W, Buskens V and Frey V. (2013) The rationality of social structure. *Social Networks* 35(4): 720-732.

Scholz JT, Berardo R and Kile B. (2008) Do networks solve collective action problems? *The Journal of Politics* 70(2): 393-406.

Sell J. (2007) Social Dilemma Experiments in Sociology, Psychology, Political Science and Economics. In: Webster M and Sell J (eds) *Laboratory experiments in the social sciences*. Amsterdam ; Boston: Academic Press/Elsevier, 459-479.

Shepsle K. (2010) *Analyzing Politics*, New York and London: W.W. Norton, chapters 8-10.

Sobel J. (2002) Can we trust social capital? *Journal of Economic Literature* 40(1): 139-154.

Stiglitz J. (2000) Formal and informal institutions. In: Dasgupta P and Serageldin I (eds) *Social Capital: a multifaceted perspective*. Washington, DC: World Bank, 59-68.

Victor JN, Lubell M and Montgomery AH. (2018) *Oxford Handbook of Political Networks*. Oxford: Oxford University Press.

Wasserman S and Faust K. (1994) *Social network analysis*, Cambridge, UK: Cambridge University Press.

Young HP. (1998) *Individual strategy and social structure*, Princeton, N.J.: Princeton University Press, chapter 6.

ACKNOWLEDGEMENTS

The author thanks Larry Moss for useful comments on an early draft. Two anonymous reviewers provided detailed feedback to guide this final revision.

FUNDING

The author received no financial support for the research, authorship, and/or publication of this article.

APPENDIX

6.1 Proof for Claim 1

Proof. Take any $n_i^A \in T$ to derive an optimal strategy from player i's perspective. By assumption, player j's types are themselves playing a pooled strategy, so either $\sigma_i^*(n_j^A) = 0$ or $\sigma_i^*(n_j^A) = 1$ for all $n_j^A \in T$. We want to show that n_i^A will cooperate under both scenarios if the expected value $\hat{n}_j^A(B_i)$ exceeds the cooperation threshold.

Case 1: $\sigma_i^*(n_j^A) = 0$

In the first case, $\mathbb{E}[u_i(\sigma_i(n_i^A) = 1)|B_i(n_j^A|\theta_i, n_{-ij})] = U_c$. Given linear payoffs in terms of alter set sizes, $\mathbb{E}[u_i(\sigma_i(n_i^A) = 0)|B_i(n_j^A|\theta_i, n_{-ij})] = M_d - \alpha\rho \cdot \hat{n}_j^A(B_i)$. Let $a = M_d - U_c$. Player-type n_i^A will then cooperate if:

$$\hat{n}_j^A(B_i) \geqslant a/\alpha\rho \tag{15}$$

Case 2: $\sigma_i^*(n_j^A) = 1$

In this case $\mathbb{E}[u_i(\sigma_i(n_i^A) = 1)|B_i(n_j^A|\theta_i, n_{-ij})] = M_c(1 + \omega_{ij}b_\omega)$ and $\mathbb{E}[u_i(\sigma_i(n_i^A) = 0)|B_i(n_j^A|\theta_i, n_{-ij})] = U_d - \alpha\rho \cdot \hat{n}_j^A(B_i)$. Player-type n_i^A will therefore cooperate if $\hat{n}_j^A(B_i) \geqslant [U_d - M_c(1 + \omega_{ij}b_\omega)]/\alpha\rho$ or in terms of the extortion parameter e:

$$\hat{n}_j^A(B_i) \geqslant [e - \omega_{ij}b_\omega M_c]/\alpha\rho \tag{16}$$

By the definition of R_{min} and equation 3, $n^A_{min} \geqslant \max(e, a)/\alpha\rho$, so both cooperation conditions 15 and 16 are satisfied. Therefore, all n^A_i types choose $\sigma^*_i(n^A_i) = 1$ if $\hat{n}^A_j(B_i) \geqslant n^A_{min}$.

Given symmetric payoff structures, a similar logic applies to n^A_j types who choose an optimal strategy $\sigma^*_j(n^A_j) = 1$ if $\hat{n}^A_i(B_j) \geqslant n^A_{min}$, and $\sigma^*_j(n^A_j) = 0$ otherwise.

$\square$