Substitutability of Favors and Bilateral Enforcement of Cooperation

Oğuzhan Çelebi*
MIT

October 21, 2022

Abstract

I introduce a favor exchange model that allows players to rely on multiple partners to obtain favors (*i.e.*, cooperation is substitutable) and study bilateral enforcement of cooperation. Without substitutability, there is either no cooperation or universal cooperation, while under substitutability, each additional relationship is less valuable than the previous one and intermediate levels of cooperation are observed. I show that transfers facilitate cooperation but may exacerbate inequality when players are heterogeneous. I extend the model to allow for community and legal enforcement and characterize when each enforcement mechanism is optimal. In applications, I demonstrate how my model can offer insights into the stratification of social networks in post-Soviet states and the adoption of different enforcement mechanisms by medieval traders.

*Email: ocelebi@mit.edu

I am grateful to Daron Acemoglu, Daniel Barron, Roberto Corrao, Glenn Ellison, Joel Flynn, Drew Fudenberg, Robert Gibbons, Ali Kakhbod, Giacomo Lanzani, Parag Pathak, Karthik Sastry, Alp Simsek, Birger Wernerfelt, Alexander Wolitzky and participants at the MIT Theory and Organizational Economics Lunches for helpful comments and suggestions. First Draft: November 2021.

1. Introduction

In many settings, players depend on informal mechanisms to sustain cooperation. Examples include favor exchange networks in villages in rural India (Jackson, Rodriguez-Barraquer, and Tan, 2012) and Soviet Union (Ledeneva (1998), see also Section 2.1), traders in medieval era (Greif (1994), see also Section 2.2) and firms that interact frequently (Hendley and Murrell (2003)). In these settings, writing down and enforcing explicit contracts is either hard or impossible due to high costs or the lack of or infrastructure, and cooperative behavior enforced by informal mechanisms is important. Since the study of Maghribi traders' coalition (Greif, 1993), economists have studied the enforcement of cooperation extensively, usually concentrating on multilateral (in other words, community) enforcement, and analyzing how coalitions of individuals can punish those who behave in uncooperative ways.[1]

However, multilateral enforcement is not ubiquitous. Bilateral mechanisms play an important role in enforcing cooperation in many settings ranging from favor exchange relationships in Soviet Union to trade/sales agreements between Genoese traders in medieval era and modern firms.[2]

Thus motivated, this paper contributes to the literature on social cooperation in three ways. The first is the analysis of cooperation under bilateral enforcement. When applied to the previously studied standard models of favor exchange (see Footnote 21 for examples), my focus on bilateral enforcement has two important consequences: (i) a simpler and more tractable characterization of networks that can sustain cooperation and (ii) trivial and rather unrealistic predictions of either no cooperation or universal cooperation. The main reason behind (ii) is that the previous models implicitly assume that each favor can be performed by a unique individual. For example, if *Bob* can perform a favor for *Alice*, another friend of *Alice*, say, *Carol*, cannot perform that same favor. I call this *monopolistic cooperation* and extend the model to allow for substitutability of cooperation, which is the second main contribution of this paper. Considering substitutability not only is more realistic in many contexts[3] but also introduces important changes in how network structure affects the values individuals attach to their relationships, which is at the core of the analysis. Third, I extend the model to allow for community enforcement, where subsets of players can form *communities* and observe interactions between members of their community, and bilateral

[1]See Wolitzky (2021) for a recent survey of this literature, as well as other important early contributions from other disciplines such as Granovetter (1985) and Coleman (1994).

[2]Section 2 describes these settings in detail.

[3]It is reasonable to think that many favors can be performed by different individuals. For example, if an individual needs a neighbor to watch their kids for a short amount of time, a worker needs to change the time for their shift for a day or a firm needs to partner with another to execute a transaction, possibly there are many different individuals or firms who can help them in these situations.

enforcement with courts, where individuals in the society can create court systems that can punish those who deviate to complement bilateral enforcement. I characterize when community or bilateral enforcement (with or without courts) is the efficient mechanism.

To fix ideas, consider three individuals, *Alice*, *Bob*, and *Carol*. Assume that there are three different types of favors denoted by a, b, and c. In each period, each player is able to provide a given type of favor with some probability. In each period, a player requires a favor (of random type) with some probability and can only receive that favor from players who are able to provide it in that period. First, consider monopolistic cooperation where favor a can only be provided by *Alice*, favor b can only be provided by *Bob* and favor c can only be provided by *Carol*. In this case, *Alice* and *Bob* interact whenever *Alice* requires favor b and *Bob* is able to perform it in that period, or vice versa. As a result, the probability that *Alice* and *Bob* interact at any given period does not depend on whether they have a relationship with *Carol* or not. Moreover, if cooperation between *Alice* and *Bob* ends for some reason, *Alice* cannot obtain favor b from *Carol* even if they have a cooperative relationship. Thus, if *Bob* refuses to provide favors to *Alice* after a defection, *Alice* loses any benefit that she can obtain from that favor type. Therefore, the value of a relationship with *Bob* does not depend on how many other players *Alice* can interact with.

Conversely, consider the case where all players can provide all types of favors with the same probability, and the player who provides the favor is randomly selected if more than one player is able to. Then the probability that *Alice* and *Bob* interact now depends on whether they are linked with *Carol* or not.[4] Moreover, if the cooperation between *Alice* and *Bob* breaks down for some reason, and if they have a cooperative relationship with *Carol*, they can rely on her to obtain favors and losing their relationship is less important. Therefore, the value of their relationship is endogenous and depends on whether they have a relationship with *Carol* or not.

I consider a model with N players and F types of favors. To study bilateral enforcement, I define a refinement of Perfect Bayesian Equilibrium, called Bilateral Equilibrium, where the action of player i when playing with player j can only depend on the previous play between i and j, but not others. Bilateral equilibrium rules out community enforcement but allows players to punish each other by refusing to perform favors or removing relationships if their partner fails to cooperate at any given time. Analyzing a game where players can exchange

[4] In particular, if players need a favor with probability α and perform a favor with probability p, in monopolistic cooperation *Alice* asks a favor type b with probability $\alpha/3$ and *Bob* can provide it with probability p. When cooperation is substitutable, *Alice* needs a favor with probability α and *Bob* will provide it if (i) *Bob* is the only other player who can provide the favor in that period, which happens with probability $p(1-p)$ or (ii) both *Bob* and *Carol* can provide the favor, but *Bob* ends up providing it, which happens with probability $p^2/2$.

costly favors over time, I characterize *stable* networks that can be supported under bilateral equilibria where all favors are performed and all links are kept on the equilibrium path.

Under monopolistic cooperation, there is a sharp discontinuity on the predicted levels of cooperation in bilateral equilibria. In particular, there is a cut-off value for the discount factor such that if players are patient enough, then the complete network is stable, while otherwise, the empty network is the unique stable network (Proposition 1). As in many settings we observe neither full cooperation nor no cooperation, a model with bilateral enforcement does not make empirically valid or reasonable predictions.[5]

Next, I turn to the case where favors are fully substitutable, in other words, all players are equal in terms of their ability of favor provision for all favor types. When favors are substitutable and all favors are provided on the equilibrium path, an additional relationship affects the provision of a given favor when it is pivotal, that is, that player is the only player who can provide that favor in that period. When agents have many partners, the probability that any given partner is pivotal and hence the (present, discounted) value of a relationship decreases and is dominated by the cost of providing the favor. As a result, there is an upper bound B^* such that no player can have more than B^* relationships in any stable network, while the network where all players have B^* relationships is stable and efficient (Theorem 1).

Even though stability characterizes networks that can be sustained when reached, it does not take a position on which relationships will be formed.[6] Motivated by this, I focus on a refinement of stability, strong stability, that requires any mutually profitable relationship to be formed, and in Theorem 2, derive an upper bound on the number of individuals who do not have B^* relationships in any strongly stable network. Moreover, this bound does not depend on the size of the society and in large societies, the fraction of players who do not attain the cooperation bound converges to zero. Theorems 1 and 2 are important since they show that when favors are substitutable, intermediate levels of cooperation are observed in equilibrium and form the basis for the other results in the paper. In section Section 4.4, I extend the model to allow for agents to make transfers and show that the agents who can make transfers attain the cooperation bound characterized in Theorem 2 in any strongly stable network and conclude that transfers facilitate cooperation.

[5]For example, in the favor exchange networks in Soviet Union described in Section 2.1, most individuals have a small set of others that they rely to obtain goods and services. Additionally, in the favor exchange relationship data Banerjee, Chandrasekhar, Duflo, and Jackson (2013) collects in 72 Indian villages (with an average population above 900), most individuals have at least one such relationship, while more than 95 % of the individuals have fewer than 5. This is also observed when players are firms: Malcomson (2010) describes how some firms cultivate long term relationships with a subset of their suppliers.

[6]For example, the empty network is trivially stable since there are no favors to ask or relationships to remove.

Section 5 extends the model to allow for heterogeneity. I first show that intermediate cooperation generalizes to the case where players have heterogeneous probabilities of favor provision. Next, I leverage the flexibility and tractability of the model to gain insight into the transformation of favor exchange networks in Post Soviet states. Section 5.3 describes the importance and structure of networks in Soviet Union and Post-Soviet Russia and Kyrgyzstan, where increased inequality following the break-up of the union (i) caused a shrinkage of the networks of the poor, (ii) led to a stratified network structure between the rich and the poor and (iii) allowed the use of money to supplement connections by the rich. Comparing a homogeneous society with no transfers to a heterogeneous one (where some rich individuals have lower costs or higher ability to provide favors, as well as ability to use transfers), I show how my model can be used to explain the shrinkage of the networks of the poor, the growth of the networks of the rich to the efficient levels as well as the break-up of the relationships between the two groups.

Section 6 incorporates community enforcement by allowing members of a community to observe interactions between others in their community and punish those who deviate. I show that community enforcement crowds out bilateral enforcement: whenever community enforcement increases cooperation, it prevents establishment of bilateral links with those outside of the community. Next, I compare three different enforcement mechanisms, community enforcement, where players can invest in sharing information and use multilateral punishment strategies, pure bilateral enforcement, where the cooperation is enforced only by bilateral strategies and bilateral enforcement with courts, where players can invest in a legal system that punishes players who deviate others and characterize when each of them is optimal.

Related Literature The literature on social cooperation builds on the contributions of Kandori (1992) and Ellison (1994) and focuses on community enforcement. The two most related papers that concentrate on the structure of the networks that can sustain cooperation are Jackson et al. (2012) and Ali and Miller (2013). Both papers focus on multilateral enforcement and monopolistic cooperation. Jackson et al. (2012) analyzes the favor exchange game and focuses on renegotiation-proof networks that are "robust to social contagion", which means that any breakdown in cooperation between two players spreads only to their mutual neighbors. They show that these two features imply the network is a "social quilt", which is a union of cliques. Ali and Miller (2013) focus on the other role of networks, information propagation and show that the efficient networks are of the form cliques and cooperation is sustained through social contagion. An important feature of their analysis is that assuming any player has at most d links, cliques of d is efficient. My results on the extent of cooperation complements both by focusing on the size, rather than the structure of the social networks.

Barron, Guo, and Reich (2022) analyze a favor exchange model where agents can accumulate wealth and show that richer agents do not have incentives to participate in favor exchange and leave the community, which they call "too-big-for-their-boots" effect. My analysis complements theirs by showing that the favor exchange networks become stratified when agents are heterogeneous (*e.g.*, rich and poor) and demonstrates inequality has an effect even when agents stay in the community.

Bloch, Dutta, and Manea (2019) study how partnerships are formed in a network. Starting with a fixed network, they study a model where agents search for partners by asking favors from others and whenever a favor is provided, two agents form an exclusive long-term partnership and leave the network. They characterize efficient partnership networks and show that they are reached following a decentralized search for partners. Even though the games studied and lack of community enforcement is common between this paper and Bloch et al. (2019), the possibility and importance of multiple relationships allows me to focus on the size networks that can sustain cooperation, while their focus is the characterization of efficient exclusive relationships. Bendor and Mookherjee (1990) study bilateral enforcement in a model where all players play with each other every period, allowing players to deviate in all their relationships in a single period.[7]

Acemoglu and Wolitzky (2020) study specialized enforcement where agents specialized in the enforcement of cooperation may exert coercive punishment. In their model, the incentives of the specialized enforcers to carry out costly punishments are central and determine whether specialized enforcement or a mix of community and specialized enforcement is optimal. I model specialized enforcement as courts in a non-strategic way, where a punishment can be exerted upon the deviating agents. Unless the court system is very efficient, the use of bilateral enforcement always improves outcomes. My results on the comparison of different enforcement mechanisms complement theirs by focusing on different trade-offs such as the cost of forming communication networks and population size.

Several other papers study networks and their effect on social cooperation through communication (Raub and Weesie (1990),Wolitzky (2013); Lippert and Spagnolo (2011); Balmaceda and Escobar (2017); Ali and Miller (2020); Sugaya and Wolitzky (2021)) or risk sharing and informal insurance (Karlan, Mobius, Rosenblat, and Szeidl (2009), Ambrus, Mobius, and Szeidl (2014), Bloch, Genicot, and Ray (2008)). Wolitzky (2021) presents a

[7]For example, they show that multilateral enforcement cannot improve upon bilateral enforcement when payoffs are separable and symmetric, which is not true when players are interacting with one of their partners in each period.

recent and detailed survey of this literature.

This paper is also spiritually related to the large literature in relational contracts (see Malcomson (2010) for a survey of this literature). In particular, the prediction that intermediate cooperation can be sustained by bilateral enforcement when favors are substitutable is reminiscent of the insider and outsider firms in Board (2011). In his model, faced with relationship-specific investments and hold-up problem, the principal divides agents to insiders and outsiders, while in my model, agents can sustain intermediate levels of cooperation due to diminishing marginal value of relationships when favors are substitutable.

The results in this paper complement the previous literature on social cooperation, which has mainly focused on two possible uses of networks. First, a network may reflect how players communicate with each other to exchange information about how others behave (*e.g.* Wolitzky (2013)) in order to enforce cooperation through multilateral punishments. Second, it may reflect how and when players interact with each other. In this second case, players play with all players they are linked with every period with additively separable payoffs (Lippert and Spagnolo, 2011) or each link is active (in the sense that players play interact in that period) with fixed probability in each period (Jackson et al., 2012; Ali and Miller, 2013) and focus on understanding the role of the network on community enforcement through who can punish who as well as communication between these individuals. These assumptions imply that the value players derive from a relationship is constant, does not depend on the rest of the network and is completely lost after the relationship ends, ruling out analysis of interesting forces such as the ability of a player to rely on the rest of their network to compensate for lost relationships.

2. Background for Bilateral Enforcement

2.1. Favor Exchange (Blat) Networks in Soviet Union

Blat is defined as the use of personal networks and informal contacts to obtain goods and services in short supply. It is a reciprocal relationship in which people exchange favors over time and was prevalent during the Soviet Union, where perpetual conditions of shortage and lack of access to diverse goods and services through the markets necessitated the use of such relationships.[8] Blat networks were necessary to acquire many services ranging

[8]A similar institution, *Guanxi*, is present in China. As in blat relationships, guanxi is characterized by an informal and personal connection between two individuals who abide by an implicit psychological contract to maintain a long-term relationship based on interactions following dynamic reciprocity and long-term equity principles (See Chen and Chen (2004) for more details on Guanxi, as well as the bilateral nature of the relationships).

from basic necessities and quality healthcare to leisure activities like traveling or attending concerts.[9] Ledeneva (1998), based on 56 interviews of individuals who are involved in these relationships, describes this phenomenon as

> Blat was oriented to different needs in different historical periods (it was already flourishing in the 1930's) but it always - directly through obtaining goods and services or indirectly through obtaining jobs and status - related to personal consumption and thus the distribution of material welfare: "Blat was simply a necessity for a decent life. You couldn't eat or wear what you bought in the shops, everything was in short supply, queuing and bad quality of services were appalling. To live normally, one had to have acquaintances and informal access to every sphere where needs arose" many respondents remembered.

Therefore, many individuals used their position in society to be useful to others who could reciprocate. Blat was not only available to those with high positions. Indeed, "Even an ordinary worker was in position of privileged access to something; for example, by being entitled to sturdy work-boots, or by virtue of handling scarce tools and materials, which formed a basis for his ability to promise reciprocity when claiming favours" (Ledeneva, 1998). Blat relationships are personal and are based on reciprocity. An interviewee whose mother worked on a Soviet farm describes this reciprocal nature of his father's blat relationships as

> *City gasmen gave him more of the bigger cylinders because his wife would sell good meat to them... Contacts, old contacts. They allocated him cut wood for free or for a few kopecks. And he gave them fresh eggs or dung.*

An important aspect of these relationships is that the main enforcement mechanism was bilateral.[10] They depend on the expectation of future interactions between parties. The continuation of cooperation is primarily enforced by the threat of denial of future favors. Many interviewees who used blat emphasized this nature of the enforcement:

> *"I do dentistry and never make money out of it. But I make my contacts, I know that my patients will help me, if I need something. I keep these contacts but this does not mean that a long-term relationship cannot be developed from them. If my request gets refused the relationship breaks, but if the contact is good and reliable, we may become friends."*

[9] For example, the weekly goods and services a Soviet lawyer needs to obtain by blat includes "food, dry cleaning, toilet paper, concert tickets and flowers" (Simis, 1982).

[10] It is possible, and indeed true, that in some relationships, there are other considerations such as friendship and family. We discuss these at the end of this section.

> *"The exchange is fully dependent on the interest which each side has in the 'other' and previous exchanges."*
>
> *". . . request had to be made in such a way that it would be fulfilled or, in the case of refusal, would not jeopardise the whole relationship."*
>
> *". . . the request is not adequate, one runs the risk of being refused or even losing the relationship"*

Clearly, many blat relationships are based on mutual trust and expectation of future reciprocal exchanges. The relationship continues as long as both parties benefit from it, and is enforced by the threat of losing future exchanges. These observations are summarized in Ledeneva (1998) as:

> "Because blat tends to be repetitive and often operated with known partners, and because of the absence of any sanctions outside the relationship, it is possible to speak of balance in blat relations."

Even though in many settings the enforcement mechanism seems bilateral, it is possible that such relationships are maintained by different mechanisms. Two individuals might have many common acquaintances, such as business partners or family members. Indeed, Ledeneva (1998) distinguishes between different *regimes* in these relationships, regime of equivalence, which is outlined above, and regime of affection, where personal ties are much more important. While in the first regime the exchange is the basis of the relationship, in the second, the relationship provides a basis for the exchange.[11] In the affective regime, it is possible that individuals communicate and relationships are enforced by a group.[12] Section 4 analyzes the favor exchange networks under bilateral enforcement, motivated by the blat relationships under regime of equivalence, while section 6, extends the model to incorporate community enforcement and compares the two enforcement mechanisms. In Section 5, I

[11] One interviewee describes his relationships as *"I have got many contacts who are, in fact,my friends: a car mechanic, all kinds of specialists in medicine, . . . many friends are from the same plant I worked at long ago, some are in business now . . . None of these people are business contacts, even though we all help each other in many ways. We spend our leisure time together, go to sauna and for summer picnics, Our wives are friendly as well . . . For these friends I repair everything for free, with a quality as for myself. As they do for me."*

[12] Ledeneva (1998) marks this possibility by writing "As was mentioned above, blat relations are self-regulating - that is, parties are forced to act fairly by the relationship itself. But in cases where someone is considered untrustworthy, sanctions can be inflicted. The cooling down of relationships is a signal for the person to realise it is his turn and to make his move. If this does not succeed, his reputation for untrustworthiness may spread, and relations break, especially if one lets the other down". Even in this case, she notes the self-regulating nature of the relationship and cooling down of the relationship as the main enforcement mechanism.

provide details about the effect of economic inequality on favor exchange networks following the dissolution of the Soviet Union and show how my model can explain the stratification these networks have experienced.

2.2. Maghribi and Genoese Traders

Maghribi traders are a group of Jewish traders in the Mediterranean in the eleventh century, who used a reputation based institution to deal with the contractual problems inherent in the merchant-agent transactions (See Greif (2006) for a detailed description of the institutions Maghribi traders have used). Maghribi traders mainly relied on a multilateral reputation mechanism, where traders share information about each others' behavior with other traders and use a collective punishment strategy of not trading with agents who have previously cheated others. However, another group operating in the same era, the Genoese traders, have adopted a different enforcement mechanism that does not rely on collective punishments. Greif (1994) summarizes these differences as:

> *Collectivist cultural beliefs were a focal point among the Maghribis, and individualist cultural beliefs were a focal point among the Genoese. Does the historical evidence indicate the existence of the related societal organizations? That is, was there high investment in information and collective punishment among the Maghribis and low investment in information and individualist punishment among the Genoese? The historical evidence indicates that the Maghribis invested in sharing information and the Genoese did not. Each Maghribi corresponded with many other Maghribi traders by sending informative letters to them with the latest available commercial information and "gossip," including whatever transpired in agency relations among other Maghribis. Important business dealings were conducted in public, and the names of the witnesses were widely publicized.*
>
> *Although, most likely, not every Maghribi trader was familiar with all the others, belonging to the Maghribis was easily verifiable through common acquaintances, an extensive network of communication, a common religion, and a common language.*

Unlike Maghribi traders, Genoese traders have relied on court enforcement and a bilateral reputation mechanism. This mechanism is described in Greif (2006).

> *Thus for agency relations to be established in an individualistic society, an external mechanism such as a legal system backed by the state is needed to limit agents*

ability to embezzle merchants capital. A legal system complements an institution based on individualistic cultural beliefs; it does not replace the associated bilateral reputation institution. Where a legal system has only a limited ability to restrict cheating (e.g., from misreporting profit expenses), a reputation mechanism still has to be used. The extensive writing of agency contracts suggests that this was indeed the case among the Genoese.

Therefore, cooperation is enforced by the possibility of punishments from courts and the threat of losing the bilateral relationship.[13] The analysis of bilateral enforcement in section 4 allows for a punishment from an external mechanism (such as a legal system) thus speaks to the relationships between Genoese traders. Moreover, motivated by the different enforcement mechanisms used by these similar groups, in Section 6, I allow for community enforcement and compare the efficiency of different enforcement mechanisms.

2.3. Firms and Modern Societies

Bilateral enforcement mechanisms play an important role in enforcing cooperation between firms. There is a large literature in organizational economics that studies how firms can sustain cooperation through relational contracts (See Malcomson (2010) for a review of relational incentive contracts). Hendley and Murrell (2003) provides evidence that use of bilateral enforcement mechanisms is common in a study of 254 Romanian firms. They measure the importance of different mechanisms for supporting sales agreements between firms and find that 56% of the weight was on bilateral mechanisms while enforcement that relied on third parties accounted for less than 11%. Murrell (2003) analyzes under what conditions bilateral enforcement is preferred and finds that bilateral enforcement is used when the frequency of interactions is higher and quality of the product is important but unverifiable.

Finally, consider groups such as students in a school, workers in a division of a firm, or staff in a university department. In these settings, many individuals exchange small favors or cooperate in joint projects over time. Most people in such settings have a set of friends that they can ask for small favors like covering a shift, helping with a particular project, or recovering a forgotten item. It is reasonable to assume most of these relationships are based on bilateral reciprocity rather than complicated multilateral punishment strategies observed in collectivistic environments such as village societies.

[13] "…among the Genoese an agent was induced to be honest by the fear that, if he was not, his relations with a particular merchant (who could have represented a family or even a clan) would be terminated (Greif, 2002). "

3. Model

There is a finite set $N = \{1, 2, ..., n\}$ of players, connected on a network described by an unweighted and undirected graph g, represented by the set of its links. I use ij to represent the link $\{i, j\}$, so $ij \in g$ indicates that i and j are linked under the network g. $g - ij$ and $g + ij$ denote the networks obtained from g by deleting and adding the link ij, respectively. The neighbors of player i are denoted by $N_i(g) = \{j | ij \in g\}$ and the degree of player i in the network g is the number of their neighbors, $d_i(g) = |N_i(g)|$.

Time proceeds in discrete periods indexed by $t \in \{0, 1, \ldots\}$ and g_t denotes the network in the beginning of period t. In any period, any player needs a favor with probability α.[14] F denotes the finite set of different favor types. Whenever a player requires a favor, the type of the required favor is determined randomly from the uniform distribution over F. At any period, player i is able to provide favor type $f \in F$ with probability p_{if}. Thus, j can provide favor f to i if i needs a favor of type f and j is able to provide favor f in that period. These probabilities are collected in the *favor provision matrix* M where

$$M = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1|F|} \\ p_{21} & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ p_{n1} & \cdots & \cdots & p_{n|F|} \end{bmatrix}$$

In period t, if i requires a favor and some $j \in N_i(g_t)$ is able to provide it, then i and j play a normal form game $G = \langle \{i, j\}, \{S_1, S_2\}, \{u_1, u_2\} \rangle$, where i is player 1 with strategy space S_1 and utility function $u_1(s_1, s_2)$ and j is player 2 with strategy space S_2 and utility function $u_2(s_1, s_2)$. If multiple players in $N_i(g)$ can provide favor (of type f), then the player who plays the game is selected randomly. Therefore, a society can be described by $(N, \alpha, M, G, \delta)$.

At $t = 0$, players start the game with the network g_0, which is common knowledge.[15] In each period t, the timing is as follows:

1. Players decide whether to remove any links or not.
2. The player who requires the favor (player i) and type of the required favor (f) is determined randomly.
3. The players who can provide f is determined according to $\{p_{jf}\}_{j \neq i}$.
4. If multiple players in $N_i(g_t)$ can provide favor f, one of them (player j) will be selected randomly (with uniform probability) to provide the favor.

[14] For simplicity, I assume at most one player requires a favor at a given period. This does not affect the results.

[15] The results hold as long as the players know the identity and the degree of their neighbors at g_0.

5. Players i (as player 1) and j (as player 2) play the game G.

Players observe their opponent, their actions in every period they play and $N_i(g_t)$, as well as the initial network g_0, but not g_t. Let $h_{t,i} \in H_{t,i}$ denote the period t history of player i and $h_{t,ij} \in H_{t,ij}$ denote the part of the history between players i and j.[16] The strategies are denoted by $\sigma_i : H_{t,i} \times \{N \setminus i\} \to \Delta(\{K, R\} \times S_1 \times S_2)$, where $\sigma(h_{t,i}, j)$ is a probability distribution over i's decisions on her interaction with j: whether they keep (K) or remove (R) the link with j and which (mixed) action to play if they are playing at period t. To study cooperation under bilateral enforcement, I define the following special class of strategies.

Definition 1. *A strategy profile σ is* **measurable with respect to bilateral histories** *if $\sigma_i(h_{t,i}, j) = \sigma_i(h'_{t,i}, j)$ whenever $h_{t,ij} = h'_{t,ij}$ for all i and j.*

If σ is measurable with respect to bilateral histories, then the play of i when playing with j only depends on the past interaction between them, and not on their past interaction with other players, which rules out any kind of community enforcement.[17] I will analyze *Bilateral Equilibrium*, which is a refinement of Perfect Bayesian Equilibrium (PBE) that requires players' strategies to be measurable with respect to bilateral histories.

Definition 2. (g, σ) *is a* **bilateral equilibrium** *if:*

1. *σ is measurable with respect to bilateral histories.*
2. *There exists a PBE $(\sigma, \{\mu(h_{t,i})\}_{h_{t,i} \in H_{t,i}})$ with strategies σ, beliefs $\{\mu(h_{t,i})\}_{h_{t,i} \in H_{t,i}}$ and $g_0 = g$.*[18]

For most of the paper, I will concentrate on the following favor exchange game (G_f), and its modifications.[19] An alternative formulation where players play a prisoners dilemma is analyzed in Appendix B.

	C	D
A	$(v, -c)$	$(0, -\gamma)$

G_f is a one player game where player 1 asks for a favor from player 2 who has two options, either cooperate and provide the favor (C), or defect and refuse to provide the favor (D). If player 2 cooperates, player 1 gets v, the value of the favor and player 2 gets $-c$, the cost

[16] $h_{t,ij}$ includes the periods and the action profiles of all past interactions between i and j.

[17] Section 6 relaxes this assumption to allow for community enforcement and compare the two enforcement mechanisms.

[18] A perfect Bayesian equilibrium is a strategy profile and belief system in which, for every player i and private history $h_{t,i}$, player i's continuation strategy is optimal given her belief μ about the vector of private histories $(h_{t,j})_{j \in N}$ and these beliefs are updated using Bayes rule whenever possible.

[19] This game corresponds to the stage game studied in Jackson et al. (2012) when $\gamma = 0$.

of providing the favor. I assume $v > c > 0$, hence doing favors is costly but it is pareto efficient for players to exchange favors over time. If player 2 refuses, then the favor is not provided and player 1 gets 0, while player 2 gets $-\gamma$, where $\gamma \geq 0$. This $-\gamma$ term denotes any additional (expected) penalty that could be imposed on a deviating player, possibly by a legal system. In some settings, such as the blat networks, $\gamma = 0$ is more appropriate since the interaction has no legal implications, while in some other settings, such as the court system of the Genoese traders, $\gamma > 0$ is more reasonable.[20] Clearly, if $\gamma \geq c$, then courts alone can sustain any cooperation in the absence of any bilateral or community enforcement. Therefore, we assume $\gamma < c$. This is a reasonable assumption in our settings (especially for Genoese trader's) since even in the modern countries that has a well-functioning legal system there is a large sphere for relational contracts (See Section 2.2 for a discussion). Until Section 6, I treat γ as fixed and refer this as bilateral enforcement. In Section 6, I allow the society to choose γ by paying the cost $C(\gamma)$ of maintaining a legal system with expected punishment γ and differentiate between pure bilateral enforcement ($\gamma = 0$) and bilateral enforcement with courts ($\gamma > 0$) cases.

It is easy to see that if all favors are performed, the complete network is the most preferred network for all players as it maximizes favor provision. However, players have incentives to refuse to deliver favors due to the costs they entail and delays in rewards. The main goal of this paper is to characterize *stable* networks, where cooperation can be sustained.

Definition 3. *A bilateral equilibrium* (g, σ) *is* **stable** *if all favors are provided and all links are kept on the equilibrium path.*

A network g is stable if there exists σ_g such that (g, σ_g) is a stable bilateral equilibrium. If a network is stable, then the payoffs on the equilibrium path are uniquely determined by the network structure without any reference to the strategies.

4. Substitutability of Favors and Stable Networks

This section analyzes bilateral equilibrium under different assumptions about players' favor provision abilities. Section 4.1 considers the previously studied polar case where each favor type can be provided by a unique player and shows that the equilibrium level of cooperation exhibits an unrealistic discontinuity between full or no cooperation. Section 4.2 considers the opposite case of an homogeneous society where all players can provide each favor with the same probability and derives a tight upper bound on the extent of cooperation

[20]If players are traders or firms, then not providing a favor would correspond to cheating the other trader or delivering low quality goods.

in stable networks. Section 4.3 presents a refinement of stability where mutually beneficial relationships must be formed and shows that most players in the society will attain the upper bound. Section 4.4 augments the model with transfers and illustrates how they might facilitate cooperation.

4.1. Monopolistic Cooperation

Consider the setting where $|N| = |F|$ and $M = pI \equiv M_m(p)$, where I is the identity matrix and $p \in (0, 1)$. Under $M_m(p)$, player i can only provide favor k if and only if $i = k$, in other words, each player has monopoly over their type of favor. Under $M_m(p)$, which I refer as *monopolistic cooperation*, two linked players i and j interact if i (j) requires favor j (i), which happens with probability α/N, and j (i) is able to provide the favor that period, which happens with probability p, so i and j interact with probability $2\hat{p}$ where $\hat{p} = \frac{\alpha p}{N}$, which is independent from the rest of their networks.[21] My first result explains why earlier literature concentrated on multilateral enforcement rather than bilateral enforcement.

Proposition 1. *Let $\hat{p} = \frac{\alpha p}{N}$. Under monopolistic cooperation:*

- *If $\dfrac{c-\gamma}{\hat{p}(v-c)} \leq \dfrac{\delta}{1-\delta}$, then complete network is stable and efficient.*
- *If $\dfrac{c-\gamma}{\hat{p}(v-c)} > \dfrac{\delta}{1-\delta}$, then empty network is the unique stable network.*

In terms of networks that can be supported by bilateral enforcement, there is a sharp discontinuity under monopolistic cooperation: either the complete network is stable, thus universal cooperation can be sustained in equilibrium, or empty network is the unique stable network, hence any cooperation is impossible. Which state prevails is determined by comparing the discount factor to the ratio of the immediate cost of the performing the favor to the expected net benefit of the relationship. Proposition 1 shows that if one does not consider substitutability of favors, bilateral enforcement would seem like a trivial model of social cooperation that makes unrealistic predictions of no cooperation or universal cooperation. In the next section, I show how incorporating substitutability enriches the model and results in the intermediate level of cooperation in equilibrium.

[21] Jackson et al. (2012) assumes that in each period, any two linked players can interact with an exogenously given probability, Lippert and Spagnolo (2011) assumes that all players play with each other in all periods, while Ali and Miller (2013) assumes each link is active with an exogenously Poisson rate. These models are examples of monopolistic cooperation.

4.2. Substitutable Cooperation

In contrast to monopolistic cooperation, for the rest of Section 4, I analyze another special case where all players can provide all favors with the same probability. More general substitutable favor provision matrices are analyzed in section 5.1.

Definition 4. *The favors are fully substitutable if* $p_{ij} = p_{jk} = p$ *for all* i, j *and* k*.*

$M_s(p)$ denotes a fully substitutable favor provision matrix with probabilities p. Under full substitutability, assuming all favors are performed, players have the following expected utility for a period with network g.

$$u_i(g) = \alpha v \underbrace{\left(1 - \overbrace{(1-p)^{d_i(g)}}^{\substack{\text{Probability that} \\ \text{no one is available}}}\right)}_{\substack{\text{Probability that} \\ \text{favor is performed}}} - \sum_{j \in N_i(g)} \alpha c \underbrace{\left(\frac{1-(1-p)^{d_j(g)}}{d_j(g)}\right)}_{\substack{\text{Probability that favor} \\ \text{is performed by } i \text{ to } j}} \quad (1)$$

First term is the benefit having $d_i(g)$ neighbors: i will need a favor worth v with probability α and there will be at least one player who can provide the favor with probability $1 - (1-p)^{d_i(g)}$. Second term is the cost of having neighbors with degrees $d_j(g)$. For any $j \in N_i(g)$, that player receives a favor with probability $\alpha(1-(1-p)^{d_i(g)})$. As the player performing the favor is selected randomly and all players are homogeneous in their favor provision probabilities, there is a $\alpha \frac{1}{d_j(g)} \left(1-(1-p)^{d_i(g)}\right)$ chance that player i will provide it, costing her c.

Two features of u_i are direct consequences of substitutability and instrumental for the analysis. First, a relationship affects the provision of a favor when it is pivotal, *i.e.* when all others are unable to provide it. A relationship is pivotal less frequently when a player has more links. Therefore, the marginal value of new relationships is decreasing.[22] This is fundamentally different (and in most cases, more realistic) than the monopolistic cooperation, where the marginal value of new relationships is constant and does not depend on how many links the players have. As a result, the network structure is an important determinant of the value each player attaches to their relationships, and relationships are more valuable for a player when they are scarce. Second, players prefer their neighbors to have more links.[23] When a neighbor is well-connected, the number of other players who can provide a favor to her is higher, which reduces the probability that any given player provides it. I now characterize the set of stable networks based on $u_i(g)$.

[22]Formally, if $d_g(j) = d_g(k)$, $ij \notin g$ and $ik \notin g$ then $u_i(g+ij) - u_i(g) > u_i(g+ij+ik) - u_i(g+ij)$.
[23]Formally, if $ij \in g$ and $jk \notin g$, then $u_i(g+jk) > u_i(g)$.

Definition 5. *A relationship ij is* ***sustainable*** *at g if for $x \in \{i, j\}$,*

$$\frac{\delta}{1-\delta}u_x(g) - \frac{\delta}{1-\delta}u_x(g-ij) \geq c - \gamma \tag{2}$$

Proposition 2. *A network g is* **stable** *if and only if all $ij \in g$ are sustainable at g.*

A relationship is sustainable whenever the cost of providing the favor today and keeping the link ij is preferred by both players to not providing the favor but losing the link ij. Proposition 2 shows that sustainable relationships characterize stable networks. That is, there exists a bilateral equilibrium (g, σ_g) such that all favors are provided and all links are kept on the equilibrium path if and only if equation 2 is satisfied for all relationships in g.

The LHS of equation 2 is the marginal (expected, discounted) value of a relationship at g, which is decreasing when the favors are substitutable, while RHS of 2 is the immediate cost of providing a favor (net of punishment associated to not providing the favor), which is constant. As a result, players provide favors to each other only if their degree is low enough, which limits the extent of cooperation in any stable network and results in the following theorem.

Theorem 1. *For any $(N, \alpha, M_p, G_f, \delta)$, there is a $B(\alpha, M_p, G_f, \delta) \equiv B^*$ such that if there exists j with $d_j(g) > B^*$ then g is not stable. If $d_i(g) = B^*$ for all i, then g is stable.*

I refer to B^* as the cooperation bound. In order to prove Theorem 1, note that from Footnote 23, a player can support the maximum number of neighbors if all those neighbors have maximum number of neighbors themselves. The following equation represents the main trade-off for a player who has n neighbors who all have n neighbors:

$$\underbrace{\delta\alpha v(1-p)^{n-1}p}_{\substack{\text{Benefit from being}\\ \text{helped in future}}} - \underbrace{(1-\delta)(c-\gamma)}_{\substack{\text{(Net) Cost of}\\ \text{helping today}}} - \underbrace{\delta\alpha c\frac{1-(1-p)^n}{n}}_{\substack{\text{Cost of}\\ \text{helping in future}}} \geq 0 \tag{3}$$

LHS of equation 3 is strictly decreasing in n, thus there exists a B^* such that inequality in equation 3 holds for all $n \leq B^*$ and not for any $n > B^*$. If there is a player with more than B^* links in a network g, then the player who has the most links will not provide a favor if asked, and g is not stable.[24] Moreover, $u_i(g)$ is increasing in $d_i(g)$ and $d_j(g)$ for all $j \in N_i(g)$. Thus, networks where all players have B^* links are the optimal stable networks. Formally, a network g is *constrained-efficient* if for all stable g', $u_i(g) \geq u_i(g')$ for all $i \in N$. Then we have the following corollary:

[24] The reason for this is intuitive: let i denote the player who has most links in g. Then i would prefer not to provide a favor to any of her neighbors since $d_i(g) > B^*$ and all neighbors of i have at most $d_i(g)$ links.

Corollary 1. *A network g is constrained efficient if and only if $d_i(g) = B^*$ for all i.*

While stability characterizes all possible networks that can be sustained when reached, it does not take a position on link formation. In particular, the empty network is stable as there are no relationships to end or favors to perform. One approach in the literature is to concentrate on the efficient networks that can be sustained in equilibrium. Under that approach, if favors are substitutable, we expect to see intermediate levels of cooperation in equilibrium even in the absence of multilateral enforcement and any kind of monitoring technology.

4.3. Strong Stability

This section introduces *strong stability*, a refinement of stability. While stable networks are robust to breaking of links, strongly stable networks further require that any sustainable and profitable links to be formed. When it is easy to identify and form mutually beneficial relationships, it is reasonable to expect a strongly stable network to emerge in equilibrium.

Definition 6. *A network g' is obtainable from g via deviations by $\{i, j\}$ if*

- $kl \in g'$ *and* $kl \notin g \implies \{k, l\} = \{i, j\}$
- $kl \in g$ *and* $kl \notin g' \implies \{k, l\} \cap \{i, j\} \neq \emptyset$

Definition 6 characterizes all networks that i and j can create forming a link between themselves and potentially removing their links with other players.

Definition 7. *A stable network g is strongly stable if for every $\{i, j\} \subset N$, and g' that is obtainable from g via deviations by $\{i, j\}$, either (i) $ij \in g'$ and ij is not sustainable at g' or (ii) $u_i(g) \geq u_i(g')$ and $u_j(g) \geq u_j(g')$.*

A network is *strongly stable* if for all $\{i, j\}$ pairs, any alternative network they can create by adding a sustainable link between them is not mutually beneficial.[25] The following theorem characterizes the strongly stable networks:

Theorem 2. *For the B^* in Theorem 1, the following are true:*

- *If $d_i(g) = B^*$ for all i, then g is strongly stable.*
- *In any strongly stable network and $k < B^*$, there can be at most $k + 1$ players with k links.*

[25] Condition (i) is necessary since two players with same number of links can always add an unsustainable link and satisfy (ii).

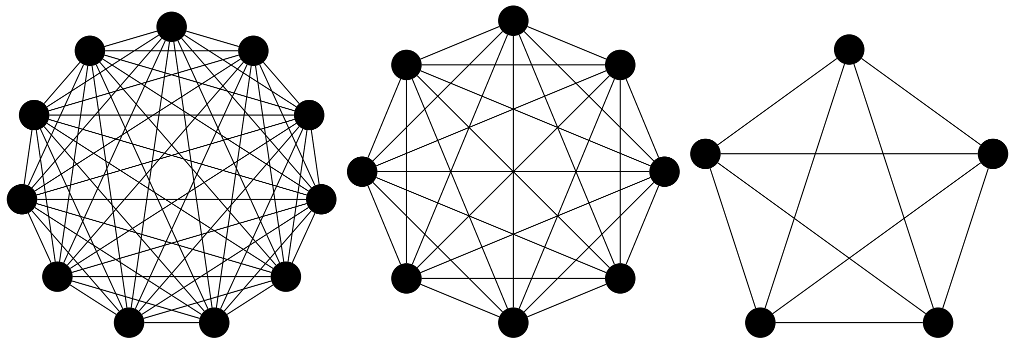

Fig. 1. Strongly Stable Network

The first part shows that constrained-efficient networks are strongly stable. Therefore, they are within the predictions of the model when profitable relationships can be identified and formed. However, there can still be some players who have fewer relationships than the efficient level. These players who do not reach the bound are in some sense undesirable partners; they require favors more frequently than they provide them.

In order to see this mechanism at work, let $p = 0.25$, $v = 6.5$, $c = 1$, $\alpha = 0.15$, $\gamma = 0$ and $\delta = 0.99$. It is useful re-write equation 2 to denote whether a link between i (with $m = d_i(g)$) and j (with $n = d_j(g)$) is sustainable or not:

$$h(m,n) = -c + \frac{\alpha\delta}{1-\delta}\left(v\left((1-p)^m - (1-p)^{m+1}\right) - \frac{c}{n+1}\left(1-(1-p)^{n+1}\right)\right) \tag{4}$$

A link between i and j is sustainable at g if $h(d_i(g), d_j(g)) \geq 0$. Under the given parameters, $B^* = 10$, while and $h(8,5) = -0.04$. Therefore, the network in Figure 1 is strongly stable. In this network, the first group of players have 10 relationships, which is the maximum number they can support. Each member of the second group has 7 and the third group has 4 relationships. Since $h(8,5) < 0$, any links between these two groups would not be sustainable and the network is strongly stable.

This example illustrates how networks can exacerbate inequality, even in symmetric models. Having a low degree makes players worse off not only because they are receiving favors less frequently, but also because it makes it harder for them to form new, sustainable relationships. Moreover, they have relationships with others who themselves have low degrees, which further reduces their payoffs. We will revisit this mechanism again in Section 5 and show that it causes stratification of social networks with heterogeneous players.

The second part of the proposition bounds the number of players who do not have B^* links. The reason behind this result is that, if two players, say i and j, have $n < B^*$ links, then they must be linked in any strongly stable network since $g + ij$ is sustainable and makes both player better off. Since B^* does not depend on N, we obtain the following corollary that shows in large societies, almost all players attain the cooperation bound B^* when mutually profitable links can be identified and formed.

Corollary 2. *As $N \to \infty$, the fraction of players who do not attain the cooperation bound vanishes.*

Given the importance of the value B^* for the results, it is instructive to study how it depends on the primitives of the model. As expected, cooperation becomes easier if the value of the favor is higher, the cost of the favor is lower, players require favors more frequently and are more patient. B^* is not monotone in p, since with very low values of p, no cooperation can be sustained and with very high values of p, only 1 neighbor will be enough to provide almost all favors and players do not have incentive to cooperate with more partners. However, it can be established that B^* is quasi-concave in p.

Proposition 3. *$B(\alpha, p, G_f, \delta)$ is increasing in δ, v, α and γ, decreasing in c and quasiconcave in p.*

4.4. Transfers

In this section, I allow players to make transfers to each other. In many settings, players can compensate each other in other ways than by performing required favors, such as buying a dinner or a small gift for a co-worker who has agreed to cover a shift. Transfers are also commonplace in professional settings. If the players represent firms, and favors represent a service provided by the firms, then parties can immediately compensate each other through transfers, as would be the case with Maghribi and Genoese traders. It is important to note that, even in such settings, the quality of the provided service may be imperfectly observable or non-contractible, so the firms can still deviate by providing a sub-par service, a low quality product or refusing to share the profits of a trade, which will correspond to not providing favor in this setting.

To incorporate transfers, I augment the network g with a transfer scheme matrix T such that $t_{ij} \geq 0$ denotes the transfer i pays for a favor from j. Let $\mathcal{T}$ denote the set of all possible transfer schemes. Moreover, let $\mathcal{T}_{\tilde{N}}$ denote the set of all transfers schemes where only players in $\tilde{N} \subseteq N$ have strictly positive transfers. This represents the inability of some of the players to make transfers, potentially due to liquidity constraints they face. For the rest of this section, I fix $\tilde{N} \subseteq N$ and assume $|\tilde{N}| > B^*$.

In the favor exchange game with transfers, G_{ft}, player 1 chooses whether to pay the transfer to player 2, while player 2 chooses whether to provide the favor simultaneously. The payment of the transfers is denoted by C and the refusal of payment is denoted by D, thus $S_1 = \{C, D\}$, $S_2 = \{C, D\}$ and payoffs are given by the following matrix.

	C	D
C	$(v - t_{12}, t_{12} - c)$	$(-t_{12}, t_{12} - \gamma)$
D	$(v - \gamma, -c)$	$(-\gamma, -\gamma)$

Let $\hat{u}_i(g, T)$ denote the per period expected utility of the players at network g.[26] Since T affects payoffs, I will include it in the definition of equilibria and say (T, g, σ) is a bilateral equilibrium if the conditions in the Definition 2 are satisfied. Moreover, for a bilateral equilibrium to be stable, all transfers must be paid in addition to all favors being provided and all links being kept on the equilibrium path. The definitions for sustainable relationships and stable networks are updated as follows.

Definition 8. *A relationship ij is sustainable at T, g if for all $x \in \{i, j\}$*

$$\frac{\delta}{1-\delta}\hat{u}_x(g', T') - \frac{\delta}{1-\delta}\hat{u}_x(g' - xy, T') \geq \max\{c, t_{xy}\} - \gamma \tag{6}$$

Definition 9. *A stable network g is strongly stable if there exists a $T \in \mathcal{T}_{\tilde{N}}$ such that for every $\{i, j\}$ and g' that is obtainable from g via deviations by $\{i, j\}$ and $T' \in \mathcal{T}_{\tilde{N}}$ where $t'_{xy} = t_{xy}$ whenever $\{x, y\} \neq \{i, j\}$, either (i) $ij \in g'$ and ij is not sustainable at (g, T') or (ii) $\hat{u}_i(g, T') \geq \hat{u}_i(g', T')$ and $u_j(g, T') \geq \hat{u}_j(g', T')$*

Definition 9 allows players to agree on a bilateral transfer scheme (t'_{ij} and t'_{ji}) as well as formation of a link (ij) and requires that in any alternative network i and j can create by themselves, either their link is not sustainable or they prefer the previous network to the alternative. The following proposition shows that transfers facilitate cooperation and allow players who can use them to reach to the cooperation bound.

Proposition 4. *In a strongly stable network, all players in $\tilde{N}$ have B^* links (the same value in Theorem 1).*

The intuition behind this results is that when a player can make transfers, she can compensate others for the favors they provide. This allows the player to avoid situations where new relationships are hard to form due to lack of initial connections. Therefore,

[26] $\hat{u}_i(g, T)$ is given by:

$$\hat{u}_i(g, T) = \alpha\left(1 - (1-p)^{d_i(g)}\right)\left(v - \sum_{j \in d_i(g)} \frac{t_{ij}}{d_i(g)}\right) - \alpha \sum_{j \in N_i(g)} \left(\frac{1 - (1-p)^{d_j(g)}}{d_j(g)}\right)(c - t_{ji}) \tag{5}$$

when all players can make transfers, in any strongly stable network, all players attain their cooperation bound.

Corollary 3. *Let $\tilde{N} = N$. A network is strongly stable if and only if all players have B^* links.*

Until now, I have focused on bilateral enforcement and homogeneous players (with the exception of transfers). In the next two sections, I leverage the flexibility of the model to study heterogeneous players and alternative forms of enforcement.

5. Heterogeneity and Social Networks in Post-Soviet States

This section extends the model in different directions to allow for heterogeneous players. First, I show that the bounded (intermediate) levels of cooperation observed under full substitutability (where all players can provide all favors with a given probability) generalize to arbitrary favor provision matrices. Next, I allow for heterogeneity in values, costs and discount factor and characterize strongly stable networks. Finally, I give a short historical background on the changes the favor exchange networks experienced after the dissolution of the Soviet Union and show that my model can explain the stratification and polarization that arose following inequality in the Post-Soviet era.[27]

5.1. Bounded Cooperation under General Favor Provision Matrices

Let F denote the finite set of different favor types and M denote the favor provision probability matrix where $p_{if} \in 0 \cup [\underline{p}, 1]$ for some $\underline{p} > 0$, in other words, the probability each player can provide each favor is either 0, or bounded below by some number. The following proposition shows that the bounded nature of cooperation is due to the substitutability of the favors and not symmetry of the fully substitutable model.

Proposition 5. *For any $(N, \alpha, M, G_f, \delta, \underline{p})$, there is a $B(\alpha, \underline{p}, G_f, \delta)$ such that if there exists j with $d_j(g) > B(\alpha, \underline{p}, G_f, \delta)$ then g is not stable.*

[27] Even though I will mainly focus on the experience of Kyrgyzstan, main mechanisms I study are present in other countries. Ledeneva (1998) describes how market reforms led to richer individuals cutting ties with poorer ones in Post-Soviet Russia. El-Said and Harrigan (2009) note how abrupt market reforms and increased poverty in Jordan caused poorer individuals to withdraw from social networks, which contributed to social and economic polarization.

Proposition 5 extends Theorem 1 to the case with a general favor provision matrix. Moreover, $B(\alpha, \underline{p}, G_f, \delta)$ does not depend on M or N. Therefore, cooperation in the society is bounded as long as the favors are substitutable.

5.2. Heterogeneity in Values, Costs and Discount Factor

I now allow players to have different values for v_i, c_i and δ_i. Formally, player i discounts future payoffs by δ_i and whenever players i and j are playing (where i is player 1 and j is player 2), they play the following game G_f^{ij}:

	C	D
C	$(v_i - t_{ij}, t_{ij} - c_j)$	$(-t_{ij}, t_{ij})$
D	$(v_i, -c_j)$	$(0, 0)$

I refer (v_i, c_i, δ_i) as the type of player i. For any i, let $B(\alpha, p, v_i, c_i, \delta_i)$ denote the cooperation bound in a society formed by players of type $(v_i, c_i, \delta_i) \in \mathcal{V}$, where $\mathcal{V}$ is finite. We say that i's type is lower than j's type if $B(\alpha, p, v_i, c_i, \delta_i) < B(\alpha, p, v_j, c_j, \delta_j)$ and i's type and j's type are equivalent if $B(\alpha, p, v_i, c_i, \delta_i) = B(\alpha, p, v_j, c_j, \delta_j)$.[28] Theorem 2 can be extended to this setting.

Proposition 6. *For any $(\alpha, p, \{G_f^{ij}\}_{(i,j)\in N^2}, \{\delta_i\}_{i\in N})$, in any strongly stable network with transfers, all players attain their cooperation bound and no player is linked with a player with lower type. Without transfers, the fraction of players who attain their cooperation bound goes to 1 and the fraction of players who are linked with lower type players goes to* 0 *as* $N \to \infty$.

Proposition 6 shows that the intuition in the homogeneous case generalizes to the case with heterogeneity and most of the players reach the cooperation bound of their own type. Moreover, it shows that the heterogeneity over the players have implications for the network structure and players payoffs. In particular, since higher type players prefer each other to lower type players, there are no relationships between these two groups in strongly stable networks when transfers are possible. When transfers are not possible, this result is holds in large societies

In the next section, I first give some historical background on the stratification and polarization in the favor exchange relationships in Post-Soviet countries after the dissolution of the union and show that my model can explain these changes.

[28] I also assume that if there exits i such that $B(\alpha, p, v_i, c_i, \delta_i) = B$ there are at least B other players with equivalent types.

5.3. Favor Exchange Networks in Post-Soviet Countries

The favor exchange networks have experienced changes in the market economy that followed the dissolution of the Soviet Union. Two respondents in Ledeneva (1998) describe the effect of the increased inequality as

> *"People separate when there is a material barrier. All my friends are now businessmen, and they turned their backs on me. Not at once, they gradually distanced themselves."*
>
> *"Those who have became wealthy have dropped out of my circle."*

Kuehnast and Dudwick (2004) study the favor exchange (blat) networks in Kyrgyz Repuclic.[29] Interestingly, the authors argue that the egalitarian conditions played a role in the formation of the relationships and the size of the networks.[30] However, the end of Soviet Union and the increased inequality that came with it had important effects on the structure of these social networks. Even though social networks continue to be an integral part of everyday life in post-socialist Kyrgyz society, Kuehnast and Dudwick (2004) emphasize three main impacts:

1. The size of networks and frequency of social encounters have significantly decreased among the poor, leading to greater economic, geographic, and social isolation. Simultaneously, the non-poor have become more reluctant to provide support to poor relatives.
2. Connections are still the primary currency for gaining access to public services, jobs, and higher education. The non-poor, however, are able to use cash to supplement or even substitute for connections.
3. There is increasing differentiation in the form and function of social networks of the poor and the non-poor. The polarization of these networks reflects the increasing socioeconomic stratification of the population.

Motivated by these observations, in the next sections I consider applications with two types of players, the rich and the poor who are heterogeneous in their abilities and costs of providing the favors.

[29] Their description of these relationships are in line with Ledeneva (1998): "...informal social networks were the most important mechanisms for getting things done, obtaining access to deficit goods and services, acquiring accurate information about events and opportunities, circumventing regulations and, in combination with bribes, gaining access to elite education, quality health care, and positions of power. This network-based economy of reciprocal favors ... was an important feature of the centralized socialist economy that helped people to compensate for failures of the state. ... the relatively egalitarian conditions of Soviet society enabled most people to establish far-reaching networks.

[30] Indeed, in my model, when players are homogeneous, any two players can have a favor exchange relationship in a stable network.

5.3.1. Heterogeneous Costs

I consider a case where the rich and the poor have different costs for providing a favor, c_r for the rich and c_p for the poor. N_r denotes the set of rich players. The rich have more resources at their disposal and it is easier for them to grant favors. Therefore I assume $c_r < c_p$. $B^*(c) = B(\alpha, M_p, v, c, \delta)$ denotes the cooperation bound of a society with cost c. Let $\tilde{c}_r = \inf_{c_r \in [0, c_p]} : B^*(c_r) = B^*(c_p)$ denote the minimum cost level under which rich and poor players have equivalent types.[31]

If the difference between c_r and c_p is large enough, then deviation is relatively more profitable for the poor compared to the rich, and it is harder for poorer players to maintain large networks. As players prefer having relationships with players who have more links, this limits the number of relationships between poor and rich players. Therefore, when players can end their relationships and form new ones, we expect limited interaction between rich and poor when the gap between them is large enough.

Proposition 7. *If $c_r > \tilde{c}_r$, all networks where all players have B^* links are strongly stable. If $c_r \leq \tilde{c}_r$, in strongly stable networks,*

- *All poor players have at most $B^*(c_p) < B^*(c_r)$ links.*
- *If $\tilde{N} = N_r$, then all rich players have $B^*(c_r)$ links and none is linked with a poor player.*
- *Without transfers ($\tilde{N} = \emptyset$), as $N \to \infty$, the fraction of rich players who have $B^*(c_r)$ links, the fraction of poor players who have $B^*(c_p)$ links and the fraction of players who only have links with their own group goes to* 1.

The proposition shows that although small differences in costs do not have an effect on the stability of constrained efficient networks where all players reach the cooperation bound, larger differences make it harder to maintain links between different groups. Therefore, my model predicts larger networks for the rich and high polarization, in line with the observed effects of inequality after the break-up of Soviet Union. Poor players are worse off through three different channels. First, the higher cost of favors directly reduces the benefits of favor exchange compared to rich players. Second, these higher costs make deviations more profitable and decrease the number of relationships poor players support. With fewer relationships, favors are performed less frequently for the poor. It is important to note that this is directly related to substitutability and would not be obtained in a model with monopolistic cooperation. Third, poor players are linked to other poor players, who have fewer relationships and depend on each of their neighbors more. Since it is better to have players with more connections, not only the networks of poor players are smaller, but they are also

[31] $\tilde{c}_r$ is characterized in the proof of Proposition 7

composed of worse partners compared to rich players. These final two channels show how inequality can result in stratification of the networks, which can exacerbate inequality in society when relationships are substitutable.

5.3.2. Heterogeneous Favor Provision

I now allow players to have different probabilities of favor provision. Rich players can provide a favor with probability p_r while poor players can only provide it with probability p_p, where $p_r > p_p$. I also assume that $\tilde{N} = N_r$ and $B^*_p = B(\alpha, \bar{p}_r, G_f, \delta) \leq \max_{p\in[0,1]} B(\alpha, p, G_f, \delta)$ *i.e.*, the rich is left of the peak of $B(\cdot, p)$, which exists by quasi concavity of B in p and in line with the observations in the application.[32]

Proposition 8. *In any strongly stable network, there are no links between rich and poor players. All rich players are linked with exactly B^*_r other rich players and attain the highest payoff they can get in any stable network. Poor players have at most B^*_p relationships.*

Contrasting this result with a homogeneous society with favor provision probability $p_h \in (p_p, p_r)$, where any two players can be linked with each other and most players have B^*_h relationships, the model explains how the use of transfers and heterogeneity in ability results in more polarized networks where rich has more relationships while poor has fewer.

Unlike the setting with different costs, there is immediate separation between the two groups. The reason is, rich players can accommodate more links since deviation is less profitable and this causes them to ask favors less frequently, which makes them better neighbors. In this setting, they are immediately better without adding more links since they can provide favors more frequently. Clearly, this immediate separation (and the separation right after the cut-off) is due to the definition of strong stability, where any mutually profitable link must be formed. In reality, we might expect to see this separation in a less knife edge way, especially if finding and forming new links is costly, or there are some non-monetary costs attached to breaking existing relationships. These can be included in the definition of strong stability.

These predictions are in line with the main observations of Kuehnast and Dudwick (2004). During the Soviet era, the society was egalitarian and most individuals had networks where they exchanged favors, which is the equilibrium of the model in the homogeneous case. However, following the collapse, society become more unequal, social networks become more polarized and the size of the networks of the poor has decreased, exhibiting the changes Propositions 7 and 8 predict.

[32]Remember that one of the main take-aways of Kuehnast and Dudwick (2004) is the growth of the networks of the richer players.

6. Community and Bilateral Enforcement

In this section, I relax the notion of bilateral equilibrium and allow community enforcement. I assume there are *communities*, which form a partition of players.[33] Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ denote a partition of players. I denote the community of a player i by $\phi(i)$.

The timing of the game is almost the same, with the only difference being that players can observe the outcomes of the interactions between members of their community. Formally, let $\tilde{\mathcal{H}}$ denote the set of all histories and let $\tilde{h}_{t,ij}$ denote the history between i and j. If $\phi(i) = \phi(j)$, then $\tilde{h}_{t,ij}$ includes the periods and the action profiles of all past interactions between i, j and any $k \in \phi(i)$. Otherwise, $\tilde{h}_{t,ij}$ only includes the periods and the action profiles of all past interactions between i and j. A society is described by $(N, \alpha, M, G, \delta, \Phi)$. The definitions 1 and 2 are updated by replacing $h_{t,ij}$ with $\tilde{h}_{t,ij}$. Given the updated definition of bilateral history, in this section, I call the equilibria *community equilibrium* instead of bilateral equilibrium, as the strategies of members of a community can depend on the past interactions within the community. The interpretation for a community is a group of players who have invested in monitoring and/or communication so that information about the interactions between community members can be monitored and credibly communicated within the community. The coalition of Maghribi traders described in Section 2.2 is a good example of a community as they have invested in sharing information and used collective punishments to enforce cooperation.

First, I will treat communities as exogenous and characterize stable networks with transfers. It turns out that to be effective, *i.e.*, increase cooperation, communities must be larger than the number of relationships that can be enforced bilaterally. This shows that these two enforcement methods cannot complement each other and the players who benefit from the higher cooperation from community enforcement cannot sustain bilateral relationships. This result explains the observation of lack of relationships between Maghribi traders (who engage community enforcement) and other traders (Greif, 1989). Next, I will introduce costs for maintain community links and court punishments, allowing the society to invest in courts to increase γ and study the efficiency of different enforcement mechanisms.

6.1. Exogenous Communities

Given Φ, a community ϕ is a *large community* if $|\phi| > B^*$. $\overline{\Phi} \subseteq \Phi$ denotes the set of large communities. These communities are important as they allow members to cooperate at a higher level than what can be done under bilateral enforcement. In a community

[33] A community can be a singleton, which means that the player is not part of any community.

equilibrium, a community attains *full cooperation* if all community links are active.[34] The following lemma shows that all large families can enforce cooperation in a strongly stable network.

Lemma 1. *The networks where all large communities attain full cooperation and all other players have B^* relationships are strongly stable.*

Moreover, full cooperation is necessary for a member of a large community to obtain their optimal payoff and whenever community enforcement increases cooperation, it prevents the establishment of bilateral relationships.

Proposition 9. *A member of a large community gets their optimal payoff within all strongly stable networks if and only if their community attains full cooperation. If a large community attains full cooperation at g, then $ij \in g$ if and only if $\phi(i) = \phi(j)$. If i is not in a large community, then $d_i(g) \leq B^*$.*

To gain intuition for this result, observe that if i and j are not in the same community, then their relationship must be enforced bilaterally. This implies that neither player can have more than B^* relationships, since in that case they will strictly prefer to withhold provision of a favor when they are called to help each other and cannot sustain cooperation.

There are two main implications of this result. First, players who rely on community enforcement via their community do not have other bilateral relationships. If $\phi(i) \in \overline{\Phi}$, then i is already cooperating with more than B^* players from its own community and thus, will prefer to not to provide a favor to any j who is not in the same community. This explains why players who operate through their communities do not have connections outside of the community. Indeed, Greif (1989) notes that "Evidence of business association between Maghribi traders and non- Maghribi traders (Jewish or Muslim) is rare". It is important to note that this prediction is a consequence of substitutability of favors and would not be obtained under monopolistic cooperation. If cooperation is monopolistic, the marginal benefit of establishing a bilateral relationship with a player outside of the community does not depend on cooperation within the community, and this relationship may be sustainable. Second, communities need to be large enough to have an effect on the level of cooperation. Any player whose community is not large enough needs bilaterally enforced relationships (and can sustain them), and therefore cannot have more than B^* links.

[34] Formally, for a community equilibrium (g, σ), ϕ attains full cooperation if $\phi(i) = \phi(j)$ implies $ij \in g$.

6.2. *Efficiency of Different Enforcement Mechanisms*

I now analyze the efficiency of community enforcement. First, if the monitoring and communication in the community networks is costless, then clearly a single community of the whole society is optimal. However, as argued in previous sections, it is plausible that there are costs attached to maintaining such a network. I will compare three different regimes:

1. *Pure Bilateral Enforcement:* Players play the favor exchange game with $\gamma = 0$.
2. *Community Enforcement:* Players are part of communities and play he favor exchange game with $\gamma = 0$. Any player i pays a per period cost κ for each j with $j \in \phi(i)$.
3. *Bilateral Enforcement with Courts*: Any deviating player gets $-\gamma$ after a deviation. Each player pays $C(\gamma)/N$ per period to maintain this legal system. C is convex, $\lim_{\gamma \to c} C(\gamma) = \infty$.[35]

Pure bilateral enforcement represents cases where cooperation cannot be enforced by a legal system or multilateral punishments. As outlined in section 2.1, many blat relationships fall under this category. Community enforcement is the main enforcement mechanism in Maghribi trader's coalition and κ denotes the costs attached to maintaining communication and monitoring in the community network. Bilateral enforcement with courts is a good model for Genoese traders and modern firms, especially when quality of the goods and services are important and not easily verifiable.

I now characterize efficient outcomes in both community and legal enforcement cases. In community enforcement, given κ, there exists an optimal community size that maximizes the utility of players if cooperation is sustained by community enforcement. Let $U(|\phi|, \kappa)$ denote the payoff of a community member when cooperation is enforced by a community of size $|\phi|$.

Proposition 10. *There exists a $B(\kappa)$ such that $U(|\phi|, \kappa)$ is maximized at $|\phi| = B(\kappa)$. $B(\kappa)$ is decreasing in κ. There is a $\overline{\kappa}$ such that*

- *If $\kappa \geq \overline{\kappa}$, B^* links supported by bilateral enforcement results in higher payoff than cooperating with community members for any community size.*
- *If $\kappa < \overline{\kappa}$, then a community of size $B(\kappa)$ results in a higher payoff than bilateral cooperation. Moreover, $B(\kappa) > B^*$.*

The optimal community size is a direct consequence of substitutability. When cooperation is enforced by communities, the (per-period) cost of adding a new member is κ, while the

[35] I assume that the cost of maintaining the legal system does not depend on the number of players, but the results will hold as long as there exists $\overline{N}$ such that for all $N > \overline{N}$, $\frac{\partial C(\gamma,N)}{\partial N} < 1$.

benefit decreases as the community grows. Therefore, it is optimal for players to organize themselves as communities of $B(\kappa)$. Clearly, for a community to increase the pay-off of its members, it must be larger than the number of relationships that can be sustained by pure bilateral enforcement, which happens when the cost of maintaining the community links is low enough.

Remark 1. *If there are $B(\kappa)$ players without communities, then they are strictly better off forming a community.*

Remark 2. *For any community with more than $B(\kappa)$ members, a subset of the community are strictly better of by seceding and forming a new, smaller community.*

Therefore, if it is easy for players to form new communities, we would expect the society to organize themselves as communities of size $B(\kappa)$, while pure bilateral enforcement is otherwise optimal.

I now turn attention to bilateral enforcement with courts. The following result follows from Theorem 2, Proposition 3 and Corollary 1.

Corollary 4. *For each γ, there is a $B^*(\gamma)$ such that in any strongly stable network, all players have $B^*(\gamma)$ links. $B^*(\gamma)$ is increasing in γ. Moreover, these networks are the constrained-efficient networks.*

I now derive derive the optimal level of investment in legal system and compare its efficiency to the two other enforcement mechanisms.

Proposition 11. *For each N, there exists $\gamma^*(N)$ that maximizes the utility under bilateral enforcement with courts. If $\gamma^*(N)$ and $\gamma^*(N')$ are maximizers and $N > N'$, then $\gamma^*(N) \geq \gamma^*(N')$. There exists $\overline{N}$ and $\overline{N}(\kappa)$ such that*

- *Bilateral enforcement with courts gives players higher payoffs than pure bilateral enforcement if and only if $N \geq \overline{N}$.*
- *Bilateral enforcement with courts gives players higher payoffs than community enforcement if and only if $N \geq \overline{N}(\kappa)$.*

This proposition shows that the legal system functions better in larger societies due to higher investment and dominates other forms of enforcement. The main trade-off is between cost of implementing a legal system and population size. Bilateral enforcement with courts is better than community enforcement in large populations, echoing Greif's observation about the size of Genoese and Maghribi traders, as well as the collectivism and individualism, being a determinant of these different choices. Maghribi traders share a common religion, language, belong to a collectivistic culture and lived in a smaller society (which would correspond to lower κ), while Genoese traders lived in a larger, individualistic society. Finally, pure bilateral enforcement is optimal when maintaining communication in a community network is hard and the society is small or establishing legal institutions is impossible.

7. Conclusion

This paper introduces a novel framework to study favor exchange by decomposing the process that determines which players interact at any given period to two: the realization of the need for a favor and whether a given player can provide that favor or not. This framework allows favors to be substitutable, that is, different individuals can perform a given favor. When favors are substitutable, the network structure becomes important to determine how frequently individuals interact with each other and affects the value of their relationship. By focusing on this new role of the network, I uncover an important relationship between substitutability of relationships and bilateral enforcement, and show that when players can rely on the rest of their network after losing a relationship, intermediate levels of cooperation is observed under bilateral enforcement. Moreover, when mutually profitable relationships can be identified and formed, most, while if transfers are also possible, all individuals reach this maximum level of cooperation.

I extend the model to allow for transfers, heterogeneous players and community enforcement. First, I apply the results to favor exchange networks in Post-Soviet States. Motivated by studies on changes after the break-up of the Soviet Union, I study a society which is divided to two groups, rich, and poor. The predictions of the model match the consequences emphasized in these studies: the networks of the rich grow while the networks of the poor contract and the two groups do not have relationships with each other. Second, I study the efficiency of different enforcement mechanisms where players can invest in communication and use multilateral punishments or a court system that punishes players who deviate. I show that community enforcement crowds out bilateral enforcement, dominates enforcement by courts when it is easy to form the communication channels, while enforcement by courts is better than community enforcement for larger societies. These results echo the mechanisms adopted by Genoese and Maghribi traders described in Greif (1994).

References

Acemoglu, D., Wolitzky, A., 2020. Sustaining cooperation: Community enforcement versus specialized enforcement. Journal of the European Economic Association 18, 1078–1122.

Ali, S. N., Miller, D. A., 2013. Enforcing cooperation in networked societies. Unpublished paper.[282] .

Ali, S. N., Miller, D. A., 2020. Communication and cooperation in markets. arXiv preprint arXiv:2005.09839 .

Ambrus, A., Mobius, M., Szeidl, A., 2014. Consumption risk-sharing in social networks. American Economic Review 104, 149–82.

Balmaceda, F., Escobar, J. F., 2017. Trust in cohesive communities. Journal of Economic Theory 170, 289–318.

Banerjee, A., Chandrasekhar, A. G., Duflo, E., Jackson, M. O., 2013. The diffusion of microfinance. Science 341.

Barron, D., Guo, Y., Reich, B., 2022. Wealth dynamics in communities. Working Paper .

Bendor, J., Mookherjee, D., 1990. Norms, third-party sanctions, and cooperation. JL Econ & Org. 6, 33.

Bloch, F., Dutta, B., Manea, M., 2019. Efficient partnership formation in networks. Theoretical Economics 14, 779–811.

Bloch, F., Genicot, G., Ray, D., 2008. Informal insurance in social networks. Journal of Economic Theory 143, 36–58.

Board, S., 2011. Relational contracts and the value of loyalty. American Economic Review 101, 3349–67.

Chen, X.-P., Chen, C. C., 2004. On the intricacies of the chinese guanxi: A process model of guanxi development. Asia Pacific Journal of Management 21, 305–324.

Coleman, J. S., 1994. Foundations of social theory. Harvard university press.

El-Said, H., Harrigan, J., 2009. you reap what you plant: Social networks in the arab worldthe hashemite kingdom of jordan. World development 37, 1235–1249.

Ellison, G., 1994. Cooperation in the prisoner's dilemma with anonymous random matching. The Review of Economic Studies 61, 567–588.

Granovetter, M., 1985. Economic action and social structure: The problem of embeddedness. American journal of sociology 91, 481–510.

Greif, A., 1989. Reputation and coalitions in medieval trade: evidence on the maghribi traders. The journal of economic history 49, 857–882.

Greif, A., 1993. Contract enforceability and economic institutions in early trade: The maghribi traders' coalition. The American economic review pp. 525–548.

Greif, A., 1994. Cultural beliefs and the organization of society: A historical and theoretical reflection on collectivist and individualist societies. Journal of political economy 102, 912–950.

Greif, A., 2002. Economic history and game theory. Handbook of game theory with economic applications 3, 1989–2024.

Greif, A., 2006. Institutions and the path to the modern economy: Lessons from medieval trade. Cambridge University Press.

Hendley, K., Murrell, P., 2003. Which mechanisms support the fulfillment of sales agreements?: Asking decision-makers in firms. Economics Letters 78, 49–54.

Jackson, M. O., Rodriguez-Barraquer, T., Tan, X., 2012. Social capital and social quilts: Network patterns of favor exchange. American Economic Review 102, 1857–97.

Kandori, M., 1992. Social norms and community enforcement. The Review of Economic Studies 59, 63–80.

Karlan, D., Mobius, M., Rosenblat, T., Szeidl, A., 2009. Trust and social collateral. The Quarterly Journal of Economics 124, 1307–1361.

Kuehnast, K. R., Dudwick, N., 2004. Better a hundred friends than a hundred rubles?: social networks in transition–the Kyrgyz Republic, vol. 41181. World Bank Publications.

Ledeneva, A., 1998. Russia's economy of favours: Blat, networking and informal exchange, vol. 102. Cambridge University Press.

Lippert, S., Spagnolo, G., 2011. Networks of relations and word-of-mouth communication. Games and Economic Behavior 72, 202–217.

Malcomson, J., 2010. Relational incentive contracts .

Murrell, P., 2003. Firms facing new institutions: transactional governance in romania. Journal of Comparative Economics 31, 695–714.

Raub, W., Weesie, J., 1990. Reputation and efficiency in social interactions: An example of network effects. American journal of sociology 96, 626–654.

Simis, K. M., 1982. Ussr: the corrupt society : the secret world of soviet capitalism .

Sugaya, T., Wolitzky, A., 2021. Communication and community enforcement. Journal of Political Economy 129, 2595–2628.

Wolitzky, A., 2013. Cooperation with network monitoring. Review of Economic Studies 80, 395–427.

Wolitzky, A., 2021. Cooperation in large societies. Working Paper .

Appendices

A. Omitted Proofs

A.1. Proof of Proposition 1

To prove the first part, assume $c - \gamma \leq \dfrac{\delta \hat{p}(v-c)}{1-\delta}$. We will show the following strategy-belief pairs are a bilateral equilibrium under the complete network:

$$\sigma_i^*(h_{t,i}, j) = \begin{cases} \{K, C\} & \text{if } ij \in N_i(g_t) \text{ and all favors are performed between } i \text{ and } j \text{ at } h_{t,i} \\ \{R, D\} & \text{otherwise} \end{cases} \tag{7}$$

and μ_i^* puts positive probability only to the private histories of other players where all previous favors have been performed and links have been kept apart from the ones i have already observed otherwise.

Let σ_i denote another strategy that is different from σ_i^*. In what follows, we will show that, for each $h_{t,i}$ such that σ_i deviates from σ_i^* and $j \neq i$, replacing $\sigma_i(h_{t,i}, j)$ with $\sigma_i^*(h_{t,i}, j)$ weakly increases the payoff of i, thus σ_i is not a profitable deviation. Let $i \neq j$ denote another player.

Case 1: There was a previous deviation in $h_{t,ij}$: In this case, the link $\{ij\}$ cannot be active in period t as $\sigma_j^*(h_{t,j}, i) = \{R, D\}$. As a result, replacing $\sigma_i(h_{t,i}, j)$ with $\{R, D\}$ does not change the utility of player i.

Case 2: There was not a previous deviation in $h_{t,ij}$:

Let ω_R denote the probability that R is played under $\sigma_i(h_{t,i}, j)$ and let ω_D denote the probability that $\{K, D\}$ is played under $\sigma_i(h_{t,i}, j)$. Note that the difference in the expected continuation payoff between σ^* and σ is given by

$$\omega_R \left(\frac{\alpha p}{N}(v-c) + \frac{\delta}{1-\delta}\frac{\alpha p}{N}(v-c) \right) + \omega_D \frac{\alpha p}{N} \left(-c + \gamma + \frac{\delta}{1-\delta}\frac{\alpha p}{N}(v-c) \right) > 0 \tag{8}$$

where the first term is positive since $v - c > 0$ and second is positive from assumption above and the inequality follows. Thus, in both cases, σ^* does better than σ and σ is not a profitable deviation, proving the first part of the result.

To prove the second part, assume $c - \gamma > \dfrac{\delta \hat{p}(v-c)}{1-\delta}$ and that there is a stable network g such that $ij \in g$. Thus, there is a bilateral equilibrium strategies σ such that when all players follow σ, all favors are provided on the equilibrium path. Clearly, σ_i assigns $\{K, C\}$

to both players in all $h_{t,ij}$ with no previous deviation, as otherwise g would not be stable. Define σ' by replacing $\sigma_i(h_{t,i}, j) = \{K, C\}$ with $\{K, D\}$ and leaving the rest of the strategy unchanged. The difference in expected continuation payoff between σ and σ' is given by:

$$\frac{\alpha p}{N}\left(-c+\gamma+\frac{\delta}{1-\delta}\frac{\alpha p}{N}(v-c)\right)=\hat{p}\left(-c+\gamma+\frac{\delta}{1-\delta}\hat{p}(v-c)\right)<0 \tag{9}$$

where the last inequality follows from our assumption above. As a result, σ' is a profitable deviation and there cannot exits such an equilibrium.

A.2. *Proof of Proposition 2*

First, I prove the statement in Footnote 23

Lemma 2. *if $ij \in g$ and $jk \notin g$, then $u_i(g + jk) > u_i(g)$.*

Proof. To prove the first part:

$$\begin{aligned}
u_i(g+ij)-u_i(g)-u_i(g+ij+ik)+u_i(g+ij) &= \alpha v\Big((1-p)^{d_i(g)}-(1-p)^{d_i(g)+1}\Big)\\
&\quad -\alpha v\Big((1-p)^{d_i(g)+1}-(1-p)^{d_i(g)+2}\Big)\\
&= \alpha v(1-p)^{d_i(g)}\Big(1-2(1-p)+(1-p)^2\Big)\\
&= \alpha v(1-p)^{d_i(g)}p^2\\
&> 0
\end{aligned} \tag{10}$$

□

To prove the if part, let g be a network such that the inequality in equation 2 holds. We will show that the following strategy-belief pairs constitute a bilateral equilibrium:

$$\sigma_i(h_{t,i}, j) = \begin{cases} \{K, C\} & \text{if } ij \in N_i(g_t) \text{ and all favors are performed between } i \text{ and } j \text{ at } h_{t,i} \\ \{R, D\} & \text{otherwise} \end{cases} \tag{11}$$

and μ_i^* puts positive probability only to the private histories of other players where all previous favors have been performed and links have been kept apart from the ones i have already observed otherwise.

Let σ_i denote another strategy that deviates from σ_i^* at a history $h_{t,i}$. In what follows, we will show that, for each $j \neq i$, replacing $\sigma_i(h_{t,i}, j)$ with $\sigma_i^*(h_{t,i}, j)$ weakly increases the payoff of i, thus σ_i is not a profitable deviation. Let $i \neq j$ denote another player.

Case 1: There was a previous deviation in $h_{t,ij}$: In this case, the link $\{ij\}$ cannot be active in period t as $\sigma_j^*(h_{t,j}, i) = \{R, D\}$. As a result, replacing $\sigma_i(h_{t,i}, j)$ with $\{R, D\}$ does not change the utility of player i.

Case 2: There was not a previous deviation in $h_{t,ij}$:

Let $\mathcal{D}(h_{t,i})$ denote the set of all k such that there was a previous deviation in $h'_{t,ik}$, where h' is a subhistory of h. Let $\tilde{g}$ denote the network obtained by removing all links ik for $k \in \mathcal{D}(h_{t,i})$. Note that $\mu_i^*(h_{t,i})$ puts positive probability to histories where all favors are provided and all links kept between players l and l' unless $ll' \in D(h_{t,i})$. Let ω_R denote the probability that R is played under $\sigma_i(h_{t,i}, j)$ and ω_D denote the probability that $\{K, D\}$ is played under $\sigma_i(h_{t,i}, j)$. The difference in the expected continuation payoff between σ^* and σ is given by:

$$\begin{aligned}
&\frac{\omega_R \alpha}{d_j(\tilde{g})}\left[\frac{\delta}{1-\delta}u_i(\tilde{g}) + \frac{\delta}{1-\delta}u_i(\tilde{g}-ij)\right] + \frac{\omega_D \alpha}{d_j(\tilde{g})}\left[-c+\gamma+\frac{\delta}{1-\delta}u_i(\tilde{g}) + \frac{\delta}{1-\delta}u_i(\tilde{g}-ij)\right] \\
\geq &\frac{\omega_R \alpha}{d_j(g)}\left[\frac{\delta}{1-\delta}u_i(g) + \frac{\delta}{1-\delta}u_i(g-ij)\right] + \frac{\omega_D \alpha}{d_j(g)}\left[-c+\gamma+\frac{\delta}{1-\delta}u_i(g) + \frac{\delta}{1-\delta}u_i(g-ij)\right] \\
\geq &0
\end{aligned} \tag{12}$$

where the first inequality follows from Lemma 2 and second from our assumption above. Thus σ^* has higher expected payoff compared to σ.

To prove the only if part, assume that the inequality in equation 2 does not hold and g is a stable network. As in any bilateral equilibrium all favors are performed on the equilibrium path, the expected utility under any bilateral equilibrium is same as the expected utility under σ. Let i, j denote a pair that violates the assumption. Let $\tilde{\sigma}$ denote an alternative strategy such that $\tilde{\sigma}_i(h_{t,i}, j) = \{K, D\}$. Then the expected utility difference between two strategies is given by

$$\frac{\alpha}{d_j(\tilde{g})}\left[\frac{\delta}{1-\delta}u_i(\tilde{g}) + \frac{\delta}{1-\delta}u_i(\tilde{g}-ij)\right] > \frac{\alpha}{d_j(g)}\left[\frac{\delta}{1-\delta}u_i(g) + \frac{\delta}{1-\delta}u_i(g-ij)\right] \tag{13}$$

where the first inequality follows from Lemma 2 and second from our assumption. Thus $\tilde{\sigma}_i$ is a profitable deviation and g is not a stable network.

A.3. Proof of Theorem 1

Let i be a player who has n neighbors (who all have n neighbors) at g. For g to be stable, from Proposition 2,

$$-c+\gamma+\frac{\delta}{1-\delta}u_i(g)\geq\frac{\delta}{1-\delta}u_i(g-ij)\text{ for }j\in N_i(g)\tag{14}$$

Rearranging, we obtain

$$\begin{aligned}-c+\gamma+\frac{\delta}{1-\delta}\Big[\alpha v\Big(1-(1-p)^n\Big)-n\frac{\alpha}{n}c\Big(1-(1-p)^n\Big)\Big]&\\ \geq&\\ \frac{\delta}{1-\delta}\Big[\alpha v\Big(1-(1-p)^{n-1}\Big)-(n-1)\frac{\alpha}{n}c\Big(1-(1-p)^n\Big)\Big]&\end{aligned}\tag{15}$$

The equation highlights the trade-off between keeping the neighbor and incurring the cost and not incurring the cost but having less continuation payoff. Rearranging we obtain

$$-c+\gamma+\frac{\delta}{1-\delta}\Big(\alpha v\left((1-p)^{n-1}-(1-p)^n\right)-\frac{\alpha c}{n}\left(1-(1-p)^n\right)\Big)\geq 0\tag{16}$$

Equation 16 can be written this as a cutoff on c, denoting the highest possible value of c to cooperate when a player has n neighbors (who all have n neighbors):

$$c(n)=\frac{\delta\alpha v(1-p)^{n-1}p}{1-\delta+\frac{\alpha}{n}\delta[1-(1-p)^n]}\tag{17}$$

First, notice that $\lim_{n\to\infty}c(n)=0$, so the number of people a player can cooperate is bounded for $c>0$. Now, we show $c(n)>c(n+1)$ for all n:

$$\begin{aligned}c(n)&=\frac{\delta\alpha v(1-p)^{n-1}p}{1-\delta+\frac{\alpha}{n}\delta[1-(1-p)^n]}+\gamma\\ &=\frac{\delta\alpha v(1-p)^{n-1}p}{1-\delta+\frac{\alpha}{n}\delta[p+p(1-p)+p(1-p)^2+\ldots+p(1-p)^{n-1}]}\frac{1-p}{1-p}+\gamma\\ &>\frac{\delta\alpha v(1-p)^{n}p}{1-\delta+\frac{\alpha}{n}\delta[p(1-p)+p(1-p)^2+\ldots+p(1-p)^{n}]}+\gamma\\ &>\frac{\delta\alpha v(1-p)^{n}p}{1-\delta+\frac{\alpha}{n+1}\delta[p+p(1-p)+p(1-p)^2+\ldots+p(1-p)^{n}]}+\gamma\\ &>\frac{\delta\alpha v(1-p)^{n}p}{1-\delta+\frac{\alpha}{n+1}\delta[1-(1-p)^{n+1}]}+\gamma=c(n+1)\end{aligned}\tag{18}$$

Where first inequality holds as we only partially multiply the denominator with $(1-p)<1$

and second inequality is true as $p > (1-p)^k p$ for all $k \geq 1$. As $c(n)$ monotonically decreases to 0, there is a B^* such that $c(n) > c$ for all $n \leq B^* = c(n) < c$ for all $n > B^*$.

To prove the first part, assume there is a stable network g where at least one player have more than B^* links. Let i be the player who has the most links at g. Then $d_i(g) = k > B^*$. Let $j \in N_i(g)$ and $N_j(g) = k' \leq k$. Then

$$
\begin{aligned}
u_i(g) - u_i(g - ij) &= -c + \gamma + \frac{\delta}{1-\delta}\left(\alpha v\left((1-p)^{k-1} - (1-p)^k\right) - \frac{\alpha c}{k'}\left(1 - (1-p)^{k'}\right)\right) \\
&\leq -c + \gamma + \frac{\delta}{1-\delta}\left(\alpha v\left((1-p)^{k-1} - (1-p)^k\right) - \frac{\alpha c}{k}\left(1 - (1-p)^k\right)\right) \\
&< 0
\end{aligned} \tag{19}
$$

where first inequality follows from $\frac{\alpha c}{k'}\left(1-(1-p)^{k'}\right)$ is decreasing in k' and second from the fact that $k > B^*$. Thus, by Proposition 2, g is not stable.

To prove the second part, assume at g all players have exactly B^* links. Then $-c + u_i(g) \geq u_i(g - ij)$ for all j and by Proposition 2 g is a stable network.

A.4. *Proof of Theorem 2*

First part follows from the fact that if there exists j with $d_j(g) > B^*$, then g is not stable, thus not strongly stable. Next, let g denote the network where all players have B^* links. Let $\{i, j\} = S$ and g' denote a network obtainable from g via deviations by S. First, note that $d_i(g') \leq B^*$ and $d_j(g') \leq B^*$ as equation condition (i) of strong stability cannot be satisfied otherwise. Moreover, as the only possible link that exists in g' but does not exist in g is $\{ij\}$, for all $k \in N_i(g') \cup N_j(g')$, $d_i(g') \leq B^*$. Thus, $u_i(g) \geq u_i(g')$ and g is strongly stable.

Lemma 3. *In any strongly stable network and $k < B^*$, there can be at most $k+1$ players with k links.*

Proof. Let g denote a strongly stable network. For a contradiction, assume for some $n < B^*$, there are $n' > n$ players with n number of links. Then at least two of those players are not linked, let i and j denote those players. Then $g' = g \cup \{ij\}$ does not satisfy conditions (i) and (ii) of strong stability and thus g is not strongly stable. □

As a result, at any strongly stable network, for any $n < B^*$ there can be at most n players with n links, which proves that total number of players with fewer than B^* can be at most $\frac{(B^*)^2 + B^*}{2}$, which is independent of N. To prove the corollary, note that, as $N \to \infty$, the fraction of players who do not attain the cooperation bound goes to 0.

A.5. *Proof of Proposition 3*

Comparative statics with respect to δ, α, v and c are immediate from differentiating equation 17. To show quasi-concavity with in consider $p \in (p_1, p_2)$. We need to show that $B(\cdot, p) \geq \min\{B(\cdot, p_1), B(\cdot, p_2)\} \equiv \hat{b}$. Note that if $\hat{b} = 0$, the result is immediate. If $\hat{b} = 0$, then Thus, first, assume $\hat{b} \geq 2$. Rewriting the Equation 17 as

$$c(n,p) = \frac{\delta\alpha v(1-p)^{n-1}p}{1-\delta+\frac{\alpha}{n}\delta[1-(1-p)^n]} + \gamma \tag{20}$$

we see that $c(\hat{b}, p_1) \geq c$ and $c(\hat{b}, p_2) \geq c$. Thus, to show $c(\hat{b}, p) \geq c$, it is enough to show the quasi-concavity of $c(n,p)$ in p. Since $\frac{\partial c(n,p)}{p}|_{p=0} > 0$, $\frac{\partial c(n,p)}{p}|_{p=1} < 0$ and $c(n,p)$ is continuously differentiable, it showing that $\frac{\partial c(n,p)}{p} = 0$ at only single $p \in (0,1)$ would prove the quasi-concavity of $c(n,p)$ for $n \geq 2$. Differentiating Equation 20 and re-arranging,

$$\frac{\partial c(n,p)}{\partial p} \propto \alpha\delta\left((1-p)^p + np - 1\right) + (1-\delta)n(np-1) \equiv h(n,p) \tag{21}$$

Note that whenever $n \geq 2$, $h(n,0) < 0$, $h(n,1) > 0$ and $h(n,p)$ is monotonically increasing in p. Thus, since $\frac{\partial c(n,p)}{p} = 0$ at a single point, c is quasi-concave. Finally, assume $\hat{b} = 1$. Then

$$\frac{\partial^2 c(n,p)}{\partial p^2}|_{n=1} = -\frac{2\alpha\delta\left(\alpha\delta p + (1-\delta)\right)}{\left(\alpha\delta p + (1-\delta)\right)^2} < 0 \tag{22}$$

Hence, $c(n,p)$ is strictly concave, thus quasi-concave in p, yielding the result.

A.6. *Proofs of Proposition 4 and Corollary 3*

Lemma 4. *g is not stable if there exists $ij \in g$ with $-c + \gamma + \frac{\delta}{1-\delta}\hat{u}_i(g') - \frac{\delta}{1-\delta}\hat{u}_x(g' - ij) < 0$*

Proof. Assume there exists such an ij and g is stable. Then there exists σ such that σ is measurable with respect to bilateral histories and at σ, all links are kept, all transfers are paid and all favors are performed on the equilibrium path. Let σ' denote the strategy profile where player i chooses D whenever she plays with player j but the rest of the strategy profile is the same. Since σ is measurable with respect to bilateral histories, the difference in payoffs between σ and σ' is

$$\alpha\frac{(1-(1-p)^{d_j(g)})}{d_j(g)}\frac{\delta}{1-\delta}\left(c + \gamma - \frac{\delta}{1-\delta}\hat{u}_i(g') + \frac{\delta}{1-\delta}\hat{u}_x(g' - ij)\right) > 0 \tag{23}$$

where the inequality holds by our assumption. Since σ'_i is a profitable deviation for player i, g is not stable. □

Lemma 5. *If g is strongly stable, $d_i(g) > B^*$ and $d_j(g) > B^*$, then $ij \notin g$.*

Proof. For a contradiction, assume $ij \in g$. Without loss of generality, assume $t_{ij} \geq t_{ji}$. Then

$$-c+\gamma+\frac{\delta}{1-\delta}\hat{u}_i(g')-\frac{\delta}{1-\delta}\hat{u}_x(g'-ij) \leq -c+\gamma+\frac{\delta}{1-\delta}u_i(g')-\frac{\delta}{1-\delta}u_x(g'-ij) < 0 \quad (24)$$

where the first inequality follows from $t_{ij} \geq t_{ji}$ and second from $d_i(g) > B^*$ and Theorem 1. By Lemma 4 g is not stable. □

Lemma 6. *If g is strongly stable, $d_i(g) \leq B^*$.*

Proof. For a contradiction, assume $d_i(g) > B^*$. From Lemma 5, for all $j \in N_i(g)$, $d_j(g) \leq B^*$. Then there exists j and k such that $\{j,k\} \subset N_i(g)$ but $jk \notin g$.[36] Next, note that $t_{ji} > t_{ij}$ and $t_{ki} > t_{ik}$, as otherwise g would not be stable. But then, $g' = g + jk - ij - ik$ and T' where $t_{jk} = t_{kj} = c$ satisfies all conditions of Definition 9 with $S = \{j,k\}$ and thus g is not strongly stable. □

This lemma establishes that transfers cannot increase the upper bound on the social cooperation. The following lemma finishes the proof of Proposition 4.

Lemma 7. *If g is strongly stable, then $d_i(g) = B^*$.*

Proof. We have already shown that $d_i(g) \leq B^*$. For a contradiction, assume $d_i(g) < B^*$. Let j and k denote two players with $jk \in g$, $ij \notin g$ and $ik \notin g$. If $d_j < B^*$, it is clear that $g' = g + ij$ with $t'_{ij} = t'_{ji} = c$ satisfies all conditions of Definition 9 with $S = \{ij\}$ and thus g is not strongly stable. Same is true if $d_k < B^*$. Thus, $d_j(g) = d_k(g) = B^*$.

Without loss of generality, assume $t_{jk} \geq t_{kj}$. Consider $g' = g + ij - jk$ with $t'_{ij} - c + \epsilon$ and $t'_{ji} = c$ for small enough ϵ. Then g' and T' satisfies all conditions of Definition 9 with $S = \{ij\}$ and thus g is not strongly stable. □

The following lemma proves Corollary 3:

Lemma 8. *If $d_i(g) = B^*$ for all i, then g is strongly stable under $t_{ij} = 0$ for all $\{i,j\} \subset N$.*

Proof. Assume the conditions in Definition 9 are satisfied for some S and g'. First, note that $S = \{i\}$ is impossible since i is worse of by removing any link and cannot form any other links in g'. Second, assume $S = \{i,j\}$ for some $\{i,j\} \subset N$. There are four cases

Case 1: Both players have more than B^* links at g': Without loss of generality, let $'t_{ij} \geq t'_{ji}$. Then condition (i) of Definition 9 is violated.

[36] To see why, note that if all such jk are in g, then $N_l(g) > B^*$ for all $l \in N_i(g)$.

Case 2: One player has fewer than B^* links at g': Clearly, condition (ii) of Definition 9 is violated for that player.

Case 3: One player has B^* links, the other has more than B^* links at g': Let $d_i(g') > B^*$ and assume $t_{ji} = t_{ij} = c$. Then at g', the updated transfers must have $t'_{ji} > t'_{ij}$ as otherwise condition (i) of Definition 9 is violated. However, if $t'_{ji} > t'_{ij}$, then $u_j(g) > u_j(g')$ and condition (ii) of Definition 9 is violated.

Case 4: Both players has B^* links at B^*: If $t'_{ij} > t'_{ji}$ or $t'_{ji} > t'_{ij}$, at least one player violates condition (ii) of Definition 9. If $t'_{ij} = t'_{ji}$, both players are indifferent and condition (ii) is again violated.

As a result, at any strongly stable network, it must be that $d_i(g) = B^*$. □

A.7. Proof of Proposition 5

First, note that, from Proposition 2, if g is a stable network, then

$$\frac{\delta}{1-\delta} u_x(g) - \frac{\delta}{1-\delta} u_x(g - ij) \geq c \tag{25}$$

Let $r = |F|$. Let

$$\overline{b}(n) = \frac{\delta v}{1-\delta} \left(1 - \left(1 - (1 - \underline{p}^n)\right)\right) - c \tag{26}$$

Rearranging, we obtain

$$\overline{b}(n) = \frac{\delta v}{1-\delta} \left((1 - \underline{p})^n\right) - \frac{c}{r} \tag{27}$$

Note that $\lim_{n\to\infty} b(n) < 0$ and $b(n)$ is strictly decreasing, thus there exists a b^* such that $b(n) \geq 0$ for all $n \leq b^*$ and $b(n) < 0$ for all $n > b^*$. We have the following lemma

Lemma 9. *If g is stable and $ij \in g$, there is at least one favor type f such that j can provide f with positive probability and the number of other players who can provide f is weakly less than b^*.*

Proof. Assume for a contradiction that the statement in the lemma is not correct. Then for each favor type j can provide, there are at least $b^* + 1$ other players who are linked with i and can provide that favor with positive probability. Then

$$\frac{\delta}{1-\delta} u_x(g) - \frac{\delta}{1-\delta} u_x(g - ij) < \overline{b}(b^* + 1) < 0 \tag{28}$$

where first inequality is holds since $p_{jf} \leq 1$ for all favor types f and second follows from the definition of b^*. Thus, g is not stable, which is a contradiction. □

The proposition then follows from Lemma 9 as if $d_i(g) > r(b^*)$, then g is not stable.

A.8. Proof of Proposition 6

First, note that Lemma 4 does not depend on M, hence applies in this setting too.

Lemma 10. *Let $ij \in g$ and i has higher type than j . Then g is not strongly stable.*

Proof. For a contradiction, assume g satisfies the conditions of the lemma and is strongly stable. There are two cases, either there exists another player with equivalent type to i who has fewer than B_r^* links or all such players have B_r^* links.

Case 1: There exists a player with equivalent type who has fewer than B_r^* links: Let k denote this player. Consider the following network $g' = g + ik$ and transfer scheme T' where $t'_{ik} = t'_{ki} = c$ and $t'_{lm} = t_{lm}$ for all other $\{l, m\}$. Then g' and T' satisfies both conditions of Definition 9 and g is not strongly stable.

Case 2: : All players with types equivalent to i hasx B_r^* links: Then there exists j and k such that $jk \in g$, $ij \notin g$ and $ik \notin g$. Without loss of generality, take $t_{kj} \geq 0$. Consider the following network $g' = g + ik - kj$ and transfer scheme T' where $t'_{ik} = t'_{ki} = c$ and $t'_{lm} = t_{lm}$ for all other $\{l, m\}$. g' and T' satisfies both conditions of Definition 9 and g is not strongly stable, which is a contradiction, which completes the proof. □

This finishes the first part of the proof where transfers are allowed. If transfers are not allowed, the following lemma derives an upper bound on the number of players who do not attain their cooperation bound.

Lemma 11. *For any $k < \overline{B}$ there can be at most $k + 1$ players with k links who do not attain their cooperation bound.*

Proof. Let g denote a strongly stable network. For a contradiction, assume for some $k < \overline{B}$, there are $k' > k$ players with k number of links who do not attain their cooperation bound. Then at least two of those players are not linked, let i and j denote those players. Then $g' = g \cup \{ij\}$ satisfies the conditions of Definition 7 and g is not stable. □

Lemma 12. *If two players attain their cooperation bound and are linked with a player who have fewer links than them, then they must be linked.*

Proof. Let i and j denote the two players who are linked with a player who have fewer links than them and who attain their cooperation bound. Let k_i and k_j denote these players ($k_i = k_j$ is possible). Then $g' = g \cup \{ij\} \setminus \{ik_i, jk_j\}$ satisfies the conditions of Definition 7 and g is not stable. □

Note that if all players have types from $\mathcal{V}$, for any society with N members, there can be at most $\sum_{b=1}^{\overline{B}}(b+1)$ players who do not attain their cooperation bound, which does not depend N. Hence as $N \to \infty$, the fraction of players who attain their cooperation bounds (and fraction of players who have links with others who have fewer links than them) goes to 1.

A.9. Proof of Proposition 7

I first derive $\tilde{c}_r$. Rewrite equation 17 in the following way to derive the cut-off cost that can support $B^*(c_p)+1$ relationships.

$$\tilde{c}_r = \max_{c<c_p}\left\{c : c \leq \frac{\delta\alpha v(1-p)^{B^*(c_p)}p}{1-\delta+\frac{\alpha}{B^*(c_p)+1}\delta[1-(1-p)^{B^*(c_p)+1}]}\right\} \tag{29}$$

If $B^*(c_r) = B^*(c_p)$, strong stability of the network where all players have B^* links follows from following the same steps in the proof of Theorem 2. If $B^*(c_r) > B^*(c_p)$, The rest of the proposition follows from Proposition A.8 assuming there are two types with (v, c_r, δ) and (v, c_p, δ)

A.10. Proof of Proposition 8

First, note that Lemma 4 does not depend on M, hence applies in this setting too.

Lemma 13. *Let $ij \in g$, where i is a rich player and j is a poor player. Then g is not strongly stable.*

Proof. For a contradiction, assume g satisfies the conditions of the lemma and is strongly stable. There are two cases, either there exists another rich player who has fewer than B_r^* links or all rich players have B_r^* links.

Case 1: There exists a rich player who has fewer than B_r^* links: Let k denote this player. Consider the following network $g' = g + ik$ and transfer scheme T' where $t'_{ik} = t'_{ki} = c$. Then g' and T' satisfies both conditions of Definition 9 and g is not strongly stable.

Case 2: : All rich players have B_r^* links:

Then there exists j and k such that $jk \in g$, $ij \notin g$ and $ik \notin g$. Without loss of generality, take $t_{kj} \geq 0$. Consider the following network $g' = g + ik - kj$ and transfer scheme T' where $t'_{ik} = t'_{ki} = c$. g' and T' satisfies both conditions of Definition 9 and g is not strongly stable, which is a contradiction, which completes the proof. □

A.11. Proof of Lemma 1

To prove stability, note that all the community links can be supported by grim trigger strategies where all community members removes their links with a deviator. All non-community links can be supported by Theorem 1.

To see strong stability, first note that no player wants to remove a link unilaterally. Next, no sustainable deviation makes any community member better off, since adding a non-community links in a sustainable way requires reduction in cooperation. Finally, if ij both belong to small communities, then from Theorem 2 we know that their coalition cannot violate strong stability, which finishes the proof.

A.12. Proof of Proposition 9

Now, we assume g is strongly stable.

Lemma 14. *If i has a bilateral link, then i has at most B^* links.*

Proof. If there is a player that has a bilateral link and has more than B^* links, then the player who has the most links within the players who have a bilateral link will not provide a favor to any of its bilateral relationships. □

This lemma also proves the third statement in the Proposition.

Corollary 5. *If i has a bilateral link with j, then j has at most B^* links.*

Lemma 15. *If $\phi(i) \in \overline{\Phi}$ and $\phi(i)$ has full community cooperation, then i does not have any non-community links.*

Proof. It is clear that any non-community (i.e. bilateral) link will be deviated by i. □

This lemma proves the second statement in the Proposition. Moreover, from the corollary and last lemma it is clear that for any member of a large community, full community cooperation is strictly better than having a bilateral link, proving the first statement in the proposition.

A.13. Proof of Proposition 10

Note that

$$U(n+1,\kappa) = \frac{\delta}{1-\delta}\Big[\alpha v\Big(1-(1-p)^n\Big) - \alpha c\Big(1-(1-p)^n\Big)\Big] - n\kappa \equiv \beta(n+1,\kappa) - n\kappa \quad (30)$$

Observe that $\beta(n,\kappa)$ is strictly concave in n, thus so does $U(n,\kappa)$. Next, $\lim_{n\to\infty}\beta(n,\kappa) = \frac{\delta\alpha}{1-\delta}(v-c)$ and $\lim_{n\to\infty}\beta(n,\kappa) = -\infty$. Hence, $U(n,\kappa)$ has a unique interior optimum for $n \in \mathbb{R}$. Then U is maximized at either the floor or ceiling of the interior optimum, which proves the first part. Second part is immediate from $\frac{U(n,\kappa)}{\partial n \partial \kappa} < 0$.

To compare community enforcement with pure bilateral enforcement, note that when κ is small, cooperation with any number of players can be supported at a very low cost, therefore community enforcement is optimal. Second, when κ is large, clearly the cost of maintaining the community links dominates the benefit of cooperation, therefore bilateral enforcement is optimal. To finish the proof, we show that if community enforcement is optimal for a κ, then community enforcement is optimal under any $\kappa' < \kappa$. Let $B(\kappa)$ denote the optimal community size under κ. Note that

$$U(B(\kappa'),\kappa') \geq U(B(\kappa),\kappa') > U(B(\kappa),\kappa) \tag{31}$$

where the first inequality holds from the optimality of $B(\kappa')$ under κ' and second from the fact that U is decreasing in κ. Thus community cooperation under κ' is better than bilateral cooperation. Finally, if community cooperation is optimal under κ, then $B(\kappa) > B^*$ since cooperation of B^* players can be sustained without any cost.

A.14. Proof of Proposition 11

The utility under bilateral enforcement with courts is given by

$$U(N,\gamma) = \alpha(v-c)\left(1-(1-p)^{B^*(\gamma)}\right) - \frac{C(\gamma)}{N} \tag{32}$$

Since $B^*(\gamma)$ is increasing in γ and is defined on integers and $C(\gamma)$ is continuous, $U(\cdot,\gamma)$ only has jump discontinuities, thus is discontinuous at most countably many points.

Lemma 16. *$U(\cdot,\gamma)$ is upper-semi continuous in γ.*

Proof. From Corollary 4, $B^*(\gamma)$ is increasing in γ. Since $U(\cdot,\gamma)$ is increasing in $B^*(\gamma)$, at each point $U(\cdot,\gamma)$ is discontinuous, it jumps up. Since $U(\cdot,\gamma)$ is continuous at any other point, is is usc. □

Since $\lim_{\gamma\to c} C(\gamma) = \infty$, and the first term of $U(N,\gamma)$ is bounded, there exists $\epsilon > 0$, such that $U(N,\gamma) < U(N, c-\epsilon)$ for all $\gamma > c-\epsilon$. Then the existence of an optimal $\gamma^*(N)$ follows from the compactness of $[0, c-\epsilon]$ and upper semi-continuity of $U(\cdot,\gamma)$.

To prove the second part, for a contradiction, assume that $N > N'$ but $\gamma^*(N) < \gamma^*(N')$, where $\gamma^*(N)$ and $\gamma^*(N')$ are maximizers of $U(N,\gamma)$ and $U(N',\gamma)$. From optimality of $\gamma^*(N')$, we have $U(N', \gamma^*(N')) - U(N', \gamma^*(N)) \geq 0$. Moreover

$$(U(N, \gamma^*(N')) - U(N, \gamma^*(N))) - (U(N', \gamma^*(N')) - U(N', \gamma^*(N)))$$
$$= - \left(\frac{C(\gamma^*(N'))}{N} - \frac{C(\gamma^*(N))}{N} \right) + \left(\frac{C(\gamma^*(N'))}{N'} - \frac{C(\gamma^*(N))}{N'} \right) > 0 \tag{33}$$
$$> 0$$

where the inequality follows from $N' < N$ and $C(\gamma^*(N')) - C(\gamma^*(N)) > 0$ and implies $(U(N, \gamma^*(N')) - U(N, \gamma^*(N))) > 0$, contradicting the optimality of $\gamma^*(N)$.

To compare the two bilateral enforcement mechanisms, observe that $\max_\gamma U(\gamma, N) = U(\gamma^*(N), N)$ is increasing in N since $U(\gamma^*(N), N) \geq U(\gamma^*(N'), N) > U(\gamma^*(N'), N')$ when $N > N'$. The result then follows from observing the optimality of bilateral enforcement with courts when $N \to \infty$ and bilateral enforcement when $N \leq B^*(0)$.

To compare the efficiency of bilateral enforcement with courts and community enforcement We first show that as $N \to \infty$, bilateral enforcement with courts can approximate the first best payoff of $\alpha(v - c)$ arbitrarily.

Lemma 17. $\lim_{N\to\infty} \max_\gamma U(N, \gamma) = \alpha(v - c)$

Proof. Let ϵ be given. First, note that there exists $\hat{\gamma}$ such that $\alpha(v - c)(1 - (1 - p)^{\hat{\gamma}}) > \alpha(v - c) - \epsilon/2$. Next, observe that there exists $\hat{N}(\hat{\gamma})$ such that for all $N \geq \hat{N}$, $C(\hat{\gamma})/N < \epsilon/2$, which proves the result. □

Let $U(B(\kappa), \kappa)$ denote payoff at the optimal community size under κ. Clearly, $U(B(\kappa), \kappa) < \alpha(v - c)$. The result than follows from the previous lemma and the fact that $\max_\gamma U(\gamma, N) = U(\gamma^*(N), N)$ is increasing in N.

B. Prisoners Dilemma as the Stage Game

B.1. Model and Results

In some settings, the interaction between players might be mutually beneficial. For example, players may represent workers in a firm, and their interaction might be cooperating in a joint project.[37] Prisoner's dilemma is a more fitting model to such interaction than the favor exchange game studied in previous sections. The payoff matrix of prisoners dilemma is denoted by G_p.

[37]For example, both players can either exert costly effort or slack off, but regardless of their effort, both obtain the payoff if the project is successful, which has higher probability in the case where they both exert effort.

	C	D
C	(v, v)	$(-l, g)$
D	$(g, -l)$	$(0, 0)$

where v, l and g satisfies standard prisoner's dilemma assumptions.[38] The players expected utility for a period with network g is given by

$$\tilde{u}_i(g) = \alpha v \left(1 - (1-p)^{d_i(g)}\right) + \alpha v \sum_{j \in N_i(g)} \left(\frac{1 - (1-p)^{d_j(g)}}{d_j(g)} \right) \tag{34}$$

Both the favor exchange game and prisoner's dilemma have been used to analyze social cooperation.[39]. As argued in previous sections, players cannot substitute one relationship for the other, the probability that they interact at any given period, as well as the value of each relationship does not depend on the network structure. Therefore, like the favor exchange game, in prisoner's dilemma, whether two players can cooperate or not is determined by a comparing the benefit of a defection today and the future value of continued cooperation. Indeed, one can formally show that under monopolistic cooperation, the set of stable networks for both games are equivalent.

Proposition 12. *Under monopolistic cooperation, for any G_f with (v, c), there is a G_p with $(v, g^*(v, c), l)$ such that the set of stable networks are same under both games where*

$$g^*(v, c) = v + 2c\Big(\frac{1 + \hat{p}\delta - \delta}{1 - \delta}\Big) \tag{35}$$

However, this equivalence breaks down if favors are substitutable. The main difference is that, unlike favor exchange, interaction is mutually beneficial for the players of Prisoner's Dilemma and would like to be called to play as frequently as possible. Thus, when favors are substitutable, players' preference over the degrees of their neighbors is reversed, any player prefers their neighbors to have as few links as possible.[40]

The following definition for strong stability under G_p accounts for the change in the gain from a deviation by replacing c with $g - v$.

Definition 10. *A relationship ij is sustainable at g if for all $x \in \{i, j\}$,*

$$\frac{\delta}{1-\delta}\tilde{u}_i(g') - \frac{\delta}{1-\delta}\tilde{u}_i(g' - xy) \geq g - v \tag{36}$$

[38] That is, $v, g, l > 0$ and $2v > g - l$.

[39] For their results, Jackson et al. (2012) note that "the same results apply to the play of a prisoners dilemma between players, or other forms of trust and cooperation games with free-rider or short-term deviation challenges."

[40] Formally, if $ij \in g$ and $jk \notin g$, then $\tilde{u}_i(g + jk) < \tilde{u}_i(g)$.

Definition 11. *A stable network g is strongly stable if for every $\{i, j\} \subset N$, and g' that is obtainable from g via deviations by $\{i, j\}$, one of the two holds either (i) $ij \in g'$ and is not sustainable or (ii) $\tilde{u}_x(g) \geq \tilde{u}_x(g')$ for all $x \in \{i, j\}$.*

The difference in the preferences over the degrees of the neighbors between these two games is important for the structure of strongly stable networks. Unlike favor exchange, where some players are not desirable neighbors due to their low degree, players prefer having links with low degree players. Thus, in any strongly stable network, all players would attain the upper bound.

Proposition 13. *Let $M = M_p$. For any $(N, \alpha, p, G_p, \delta)$, there is a $B_p(\alpha, p, G_p, \delta) = B^*$ such that a network is strongly stable if and only if all players have B^* links.*

B.2. Proofs

B.2.1. Proof of Proposition 12

Note that in the favor exchange game, the stability of the networks is determined by comparing the cost of providing the favor today to the change in the continuation value for the rest of the game

$$\left(-c + \frac{\delta}{1-\delta}\hat{p}(v-c)\right) \tag{37}$$

Rearranging, the complete network is stable if and only if

$$c\left(1 + \frac{\delta \hat{p}}{1-\delta}\right) \leq \frac{\delta}{1-\delta}\hat{p}v \tag{38}$$

In the prisoner's dilemma case, the same deviation corresponds to defecting in the stage game, thus equation 37 corresponds to

$$g \leq v + 2\frac{\delta}{1-\delta}\hat{p}v \tag{39}$$

Rearranging, we obtain

$$\frac{g-v}{2} \leq \frac{\delta}{1-\delta}\hat{p}v \tag{40}$$

equating LHS of equations 37 and 40, we obtain the value stated in the proposition. The proof then follows from the exactly same steps of the proof of Proposition 1.

B.2.2. Proof of Proposition 13

Let

$$l(m,n) = v - g + \frac{\delta\alpha v}{1-\delta}\left((1-p)^{m-1} - (1-p)^m + \frac{1-(1-p)^n}{n}\right) \tag{41}$$

$l(m,n)$ denotes the change in the expected payoff of a player with m links from defecting a player with n links under a bilateral equilibrium where they have m and n links, respectively. The following lemma shows that this determines he stability of a network.

Lemma 18. *A network g is stable if and only if for all ordered pairs $ij \in g$, $l(d_i(g), d_j(g)) \geq 0$.*

Proof. Let g denote a network that satisfies the condition. We will show that the following strategy-belief pairs constitute a bilateral equilibrium:

$$\sigma_i(h_{t,i}, j) = \begin{cases} \{K, C\} & \text{if } ij \in N_i(g_t) \text{ and all favors are performed between } i \text{ and } j \text{ at } h_{t,i} \\ \{R, D\} & \text{otherwise} \end{cases} \tag{42}$$

and μ_i^* puts probability 1 to the history where all previous favors have been performed and links have been kept apart from the ones i have already observed otherwise.

Let σ_i denote another strategy that deviates from σ_i^* at a history $h_{t,i}$. In what follows, we will show that, for each $j \neq i$, replacing $\sigma_i(h_{t,i}, j)$ with $\sigma_i^*(h_{t,i}, j)$ weakly increases the payoff of i, thus σ_i is not a profitable deviation. Let $i \neq j$ denote another player.

Case 1: There was a previous deviation in $h_{t,ij}$: Exactly same as in the proof of Proposition 2.

Case 2: There was not a previous deviation in $h_{t,ij}$:

Let $\mathcal{D}(h_{t,i})$ denote the set of all k such that there was a previous deviation in $h'_{t,ik}$, where h' is a subhistory of h. Let $\tilde{g}$ denote the network obtained by removing all links ik for $k \in \mathcal{D}(h_{t,i})$. Note that $\mu_i^*(h_{t,i})$ puts probability 1 on $\tilde{g}$. Let ω_R denote the probability that R is played under $\sigma_i(h_{t,i}, j)$ and ω_D denote the probability that $\{K, D\}$ is played under $\sigma_i(h_{t,i}, j)$. Let $d_i(g) = m$ and $d_j(g) = n$. The difference in the expected continuation payoff between σ^* and σ is given by $G_D(m, n, \omega_D) + G_R(m, n, \omega_R)$ where

$$G_D(m,n,\omega_D) = \omega_D\alpha\left(v + v\frac{\delta}{1-\delta}\left(\alpha(1-(1-p)^m) + \sum_{k\in d_i(g)}\alpha\frac{(1-(1-p)^{d_k(g)})}{d_k(g)}\right)\right)$$
$$-\omega_D\alpha\left(g + v\frac{\delta}{1-\delta}\left(\alpha(1-(1-p)^{m-1}) + \sum_{k\in d_i(g),k\neq j}\alpha\frac{(1-(1-p)^{d_k(g)})}{d_k(g)}\right)\right) \tag{43}$$

$$G_R(m,n,\omega_R) = \alpha v + \omega_R\left(v\frac{\delta}{1-\delta}\left(\alpha(1-(1-p)^m) + \sum_{k\in d_i(g)}\alpha\frac{(1-(1-p)^{d_k(g)})}{d_k(g)}\right)\right)$$
$$-\omega_R\alpha\left(v\frac{\delta}{1-\delta}\left(\alpha(1-(1-p)^{m-1}) + \sum_{k\in d_i(g),k\neq j}\alpha\frac{(1-(1-p)^{d_k(g)})}{d_k(g)}\right)\right) \tag{44}$$

Note that if $G_D(m,n,\omega_D) \geq 0$ implies $G_R(m,n,\omega_R) \geq 0$. Moreover, rearranging equation 43, we obtain

$$G_D(m,n,\omega_D) = v - g + \frac{\delta}{1-\delta}\alpha v\left((1-p)^{m-1} - (1-p)^m + \frac{(1-(1-p)^n)}{n}\right) \tag{45}$$

which is equal to $l(m,n)$. Thus, whenever $l(m,n) \geq 0$, σ is not a profitable deviation. Next, let $d_i(g) = m$ and $d_j(g) = n$, assume $l(m,n) < 0$ and g is a stable network. This means that, for all $h_{t,i}$ without a previous deviation, $\sigma_i(h_{t,j}) = \{K, C\}$. However, the strategy σ' that has $\sigma_i(h_{t,j}) = \{K, D\}$ and follows σ^* in every other node increases the expected payoff of i by $-G_D(m,n,\omega_D) = -l(m,n) > 0$, and is a profitable deviation. Thus g is not a stable network.

□

Now, note that $l(n,n)$ is decreasing in n and $\lim_{n\to\infty} l(n,n) = v - g < 0$. Thus, there exists B^* such that for all $n \leq B^*$, $l(n,n) \geq 0$ and for all $n > B^*$, $l(n,n) < 0$. Thus by Lemma 18, the network where all players have B^* links is stable.

Lemma 19. *Let g denote a network where $d_i(g) > B^*$. If g is strongly stable, then there cannot be a j with $d_j(g) < B^*$.*

Proof. Since $d_i(g) > B^*$ and $d_j(g) < B^*$, there exists at least one player that is linked with i but not linked with j, let k denote this player. Consider the network $g' = g - ik + jk$. It is easy to check that both conditions of Definition 11 is satisfied under g' where $S = \{j, k\}$. ☐

Lemma 20. *Let g denote a network where $d_i(g) > B^*$. If g is strongly stable, then there cannot be a j with $d_j(g) > B^*$.*

Proof. First, note that $ij \notin g$ since $d_i(g) > B^*$ and $d_j(g) > B^*$ implies $l(d_i(g), d_j(g)) < 0$. Thus, all neighbors of i and j have at most B^* links. Thus, there exists k, k' such that $ik \in g$ and $jk' \in g$, but $kk' \notin g$. Consider the network $g' = g - ik + jk$. It is easy to check that both conditions of Definition 11 is satisfied under g' where $S = \{j, k\}$. ☐

Lemma 21. *Let g denote a strongly stable network. Then there cannot be an i with $d_i(g) > B^*$.*

Proof. Note that $d_i(g) > B^*$, and i from Lemma 20, i is the only such player. Thus, there exists j and k such that $jk \notin g$, but $ij \in g$ and $ik \in g$. Consider the network $g' = g - ik + jk$. It is easy to check that both conditions of Definition 11 is satisfied under g' where $S = \{j, k\}$. ☐

Lemma 22. *Let g denote a strongly stable network. Then there cannot be an i with $d_i(g) < B^*$.*

Proof. Assume for a contradiction $d_i(g) < B^*$. Then there exists j such that $ij \notin g$. If $d_j(g) < B^*$, then Consider the network $g' = g + ij$. It is easy to check that both conditions of Definition 11 is satisfied under g' where $S = \{i, j\}$. If $d_j(g) = B^*$, then let k denote a player with $jk \in g$. Consider the network $g' = g + ij - jk$. It is easy to check that both conditions of Definition 11 is satisfied under g' where $S = \{i, j\}$. ☐

Taken together, the lemmas above prove the result.