Theory and Experiments on a Local Public Goods Game: Inequality Aversion and Welfare Preference

Yang Zhang[**], Longfei He[*,†]

[*]College of Management and Economics, Tianjin University, Tianjin, 300072

[**] Brunel Business School, Brunel University London

In this paper, we investigate social preferences in a network game, where the network structure determines whose action affects the payoff of which player. We develop alternative theories that incorporate inequality aversion and welfare preference into the context of dominant-strategy network games, and test their implications in laboratory experiments. When the economic return is relatively low, we observe that subjects contribute more than the amount that would maximize their monetary profit; moreover, subjects at the central network positions contribute more than those at the periphery. These anomalies suggest that subject behavior is mainly driven by the welfare preference and not as much by either inequality aversion or self-interest, regardless of the network structures considered. In a supplemental experiment with an increased economic return, we find that the advantage of social preferences over self-interest in driving individual activities varies with the underlying network architecture. We also estimate the behavioral parameters and interpret the results in relation to the network topology.

Keywords: Social and economic networks, Game theory, Social preferences, Laboratory experiment, Network games

JEL classifications: L14, C72, D91, C92

[†] Corresponding author: helf@tju.edu.cn (L. He)

1. Introduction and Literature Background

The concept of externalities is central to the study of the economics of networks. In the presence of externalities, one derives a payoff not only from the action taken by oneself, but also from the actions of others. In recent years, scholars have begun to focus on "local externalities," where one's payoff is affected by the actions of a specific subset of individuals (one's neighbors) in the population. For example, if one person adopts a new technology to reduce pests on his farm, then nearby farms (and only the farms nearby) will also benefit. If one country takes effort to reduce pollution, the environment of adjacent countries also improves. In pharmaceutical or high-tech industries, research discoveries by one firm may spillover to its trade partners (neighbors). The theories involving local externalities offer insights into how the outcomes of games and social welfare are shaped by the interaction (network) structure. We refer interested readers to the literature reviews by Jackson and Zenou (2015) and Jackson et al. (2017).

However, most theoretical works on network games assume that individuals pursue their self-interests, which would seem inapplicable in many relevant social settings. For example, in a friendship or kinship network, agents may consider the well-being of their friends or relatives (neighbors) when playing the neighborhood game. In the workplace, employees may derive their utilities by comparing their earnings and performances with those of their colleagues (neighbors). Collaborators on a project—or coauthors on a paper—may place the goal of their team prior to their individual gains. These give rise to the study of other-regarding preferences. Notably, concerning this matter, see the seminal discussions in Kahneman et al. (1986a) and Kahneman et al. (1986b) , as well as models and analyses in Fehr and Schmidt (1999), Bolton and Ockenfels (2000), Charness and Rabin (2002) and the references therein. To the best of our knowledge, our paper is among the first to bridge the gap between research on network games and that on social preferences.

Through both theoretical modeling and laboratory experiments, we demonstrate the significance of incorporating social preferences into the analysis of games on networks. We develop alternative models that incorporate aversion to inequality and pursuit of welfare into the neighborhood games. These models produce qualitatively different insights, which we test using data pertinent to various network configurations.

In our model of inequality aversion [resp. welfare preference], players care not only about their pecuniary payoffs, but also about the fairness of the payoff distribution within their neighborhood [resp. welfare within their neighborhood]. In that sense, our models extend the seminal works of Fehr and Schmidt (1999) and Charness and Rabin (2002) into a *network* context. Further model comparisons based on empirical data reveal that the welfare model is more likely to capture subject behavior than inequality aversion and self interest when the economic return is relatively low, regardless of the network structures considered. When the economic return rises, the advantage of social preferences over self-interest in driving individual activities appears to differ by the underlying network topology. Specifically, in the star network, the welfare preference continues to outweigh inequality aversion and self-interest; whereas in the circle network, self-interest grows to an extent almost equally influential compared to social preferences.

This paper contributes to the theories of network games. Bramoullé et al. (2014), Bramoullé and Kranton (2007) and Ballester et al. (2006) all investigate games with linear best replies. The games they study exhibit either strategic substitute or strategic complement. Another school of research studies network games under incomplete network information. Examples include Sundararajan (2007) on strategic complementary games as well as Galeotti et al. (2010) on such games and those with strategic substitutes. In addition, there is remotely related literature that concerns games and experiments with strategic substitutes on endogenously formed networks, such as Hojman and Szeidl (2008), Cho (2009), Galeotti and Goyal (2010) and van Leeuwen et al. (2020).[2] Moreover, other studies (Zhang and Chen, 2019; Fainmesser and Galeotti, 2016; Bloch and Quérou, 2013; Candogan et al., 2012) explore the pricing issues based on network games. Elliott and Golub (2013) examine the design of networks that implement the Pareto-efficient outcomes of a public goods game. For more comprehensive reviews on the literature of network games, see the recent surveys of Jackson and Zenou (2015) and Jackson et al. (2017). In all the works cited so far, a common assumption is that individuals care only about their own payoffs. Conversely, the theories we explore here consider prosocial motives and assume that the agent's utility is also affected by the payoffs of others in their network neighborhood. In the literature of network games, Bourlès et al. (2017) and Ghiglino and Goyal (2010) are in a similar spirit to our work. Ghiglino and Goyal (2010) assume that a player derives utility from comparing her own consumption with that of her neighbors, whereas in our model such comparison occurs between players' payoffs. We thus extract qualitatively different insights from that in Ghiglino and Goyal

[2] The literature devoted to network formation, which dates back to the seminal works of Jackson and Wolinsky (1996) and Bala and Goyal (2000), constitutes a separate branch of research on social and economic networks. We are aware of that literature but will not include it in the review, given the focus of our research. Interested readers can find more relevant discussions in the survey papers cited in the text, as well as in Jackson (2008) and Goyal (2009).

(2010).[3] Bourlès et al. (2017) examines a kind of player utility that resembles our networked welfare preference. However, they study a game of linear transfers while our game features nonlinear public goods. We also study inequality aversion in network games, which is not present in Bourlès et al. (2017). In addition to the differences noted above, our work is distinguished from both Bourlès et al. (2017) and Ghiglino and Goyal (2010) by our adoption of an experimental approach.

Directly related to our paper is a handful of laboratory studies on network games. Earlier works such as Keser et al. (1998), Berninghaus et al. (2002), Cassar (2007) and Corbae and Duffy (2008) only consider bilateral (coordination or cooperation) games played between each pair of connected players. This stream of research is reviewed in Kosfeld (2004). More recent experimental works, e.g., Judd et al. (2010), Rosenkranz and Weitzel (2012), Berninghaus et al. (2013), Charness et al. (2014), Boosey and Issac (2016) and Boosey (2017) examine games that one plays multilaterally with all neighbors. More specifically, the games explored in Judd et al. (2010) involve coordination, and those in Rosenkranz and Weitzel (2012), Berninghaus et al. (2013), Boosey and Issac (2016) and Boosey (2017) all feature public goods.[4] Charness et al. (2014) investigate the games under both complete and incomplete information.[5] Our work falls into the latter (multilateral game) category, but it differs from the existing literature in two important ways. First, we develop formal behavioral models—in the context of network games—of both inequality aversion and welfare preference. Further, notably, these models apply to more general network structures than the ones explored in the prior experimental literature. Second, we tested competing hypotheses on network behavior, thus estimate the behavioral parameters, and interpret the estimation results in relation to the network topology.

3 For example, the player's utility is negatively affected by the consumption of her neighbors in Ghiglino and Goyal (2010), but it is not necessarily so in our model.

4 The public goods setting in the study of Rosenkranz and Weitzel (2012) borrows from that of Bramoullé and Kranton (2007), and exhibits nonlinearity. Berninghaus et al. (2013), Boosey and Issac (2016) and Boosey (2017) investigate linear public goods, which is a classic environment for experimental research. In Berninghaus et al. (2013), players on a square network choose both the size and the location of a contribution, which generates a geographically decayed benefit for neighbors. In Boosey and Issac (2016), the public goods contribution occurs on a complete network, whereas monitoring and punishment are restricted by some specific network layouts (i.e., complete network, circle, or a particular asymmetric architecture). Boosey (2017) contrasts the behavior of subjects on a circle network who observe their neighbors' average payoff (or average contribution) with that of subjects who are informed of the average payoff (or average contribution) regarding all players.

5 Charness et al. (2014) show that subjects under complete network information and strategic complements maximize aggregate payoffs to a considerable extent (Results 2, 6 of Charness et al. (2014)). This could be viewed as an analogy to our finding regarding the prevalence of welfare preference. Nevertheless, our setup allows for a systematic test of welfare preference against inequality aversion based on formal analytical models, which would not be available in the setting of Charness et al. (2014).

This paper is also akin to the research on social comparison, which examines how an agent's behavior changes in response to the influence of her peers. Most of the extant literature focuses on the case where individuals compare their attributes with the aggregate of the entire group—for instance, the average or median score across the population (see, among others, Frank, 1985; Hopkins and Kornienko, 2004; Chen et al., 2010; Card et al., 2012). Nevertheless, there are some recent studies (e.g., Roels and Su, 2014) that inspect social comparisons when the agents are exposed to full group information. Such exposure means, in our terminology, that the reference network is *complete* (i.e., everyone's activity is visible to everyone else). Furthermore, Immorlica et al. (2017) investigate social comparison in the presence of an explicit reference network. In their paper, social status is defined locally (i.e., players compare only with their neighbors), and the comparison is in one direction (i.e., players incur disutility only when their status is inferior to that of another player). Our work is distinguished from the literature on social comparison primarily by our consideration of payoff externalities—we consider the case where neighbors can not only observe each other's earnings, but also affect each other's earnings by their own actions. As such, the network in our study does *not only* function as a reference structure ("who receives information on whom" or "who is visible to whom") *but also* defines a payoff structure ("whose decision affects whose profit"). That explains why our findings substantially deviate from those of works in the social comparison vein. For instance, we observe scenarios in which the subjects are more motivated to maximize the welfare of their neighborhoods than to enforce the fairness of payoff distribution.

The network externality in our paper can be interpreted as a kind of public goods shared by neighbors. Most experimental research on public goods employs a linear setting for the benefit and cost of providing the good (Rege and Telle, 2004; Masclet and Villeval, 2008; Fischbacher and Gächter, 2010; Weimann et al., 2012 – see also the reviews by Chaudhuri, 2011 and Ledyard et al., 1995). However, we adopt a nonlinear setup that features a strictly positive contribution as the player's dominant strategy. Much as in Keser (1996), Isaac and Walker (1998) and Cason and Gangadharan (2015), the presence of an interior optimum enables us to test the deviation from the standard prediction toward both directions. More importantly, the nonlinear setting facilitates the estimation of behavioral parameters (which enter the equilibrium prediction).[6]

As for the rest of the paper, Section 2 introduces the models of social preferences in network games. In Section 3, we present the experimental design and protocols. Section 4 records the

[6] In the linear setting, some behavioral parameters only show up in the conditions that determine which equilibrium takes place but do not enter the equilibrium prediction, thus creating difficulty for the estimation.

experimental results, for which we conduct both aggregate-level and individual-level analyses. Subsequently, Section 5 documents a second experiment aimed at studying the interaction between network architecture and economic stake. Finally, Section 6 presents the conclusion. As a note on the writing style, we will randomly use "his" and "her" when referring to an anonymous third-person throughout this paper.

2. Model

2.1. Base model

The network consists of a set of players and the connections between them. Player i and j are [not] connected if $g_{ij} = 1$ [0]. We consider undirected links—$g_{ij} = g_{ji}$—and normalize $g_{ii} = 0$. Connected players are *neighbor*s. The set of player i's neighbors constitutes her *neighborhood*. We use $d_i \coloneqq \Sigma_j g_{ij}$ to denote the number of neighbors player i has or the *degree* of player i.

Each player i makes a (nonnegative) contribution, x_i. Player i's payoff, $\pi_i(\boldsymbol{x})$, is determined by her own contribution x_i (scalar) and the contributions of others, $\boldsymbol{x}_{-i}$ (vector); and the combined vector $\boldsymbol{x} \coloneqq (x_i; \boldsymbol{x}_{-i})$ represents the *action profile* for all players.

$$\pi_i(\boldsymbol{x}) = \delta(x_i + \textstyle\sum_{j\neq i} g_{ij} x_j) - \frac{1}{2} x_i^2. \qquad (1)$$

This payoff setting reflects local externality. A player's payoff is affected not only by her own action but also by the actions of those connected to her in the network. The payoff function also resembles a public goods setting, as for each individual player, everyone in her neighborhood (including herself) indiscriminately benefits from her contribution, while she alone bears the cost of contribution. Moreover, the setting features nonlinearity and an interior dominant-strategy equilibrium (in a similar spirit to that in Keser (1996)). If every player maximizes her payoff as in (1), there is a dominant strategy to contribute δ regardless of the player's network position. This yields the first theoretical conjecture on the equilibrium play (Remark 1).

REMARK 1 (Self-interest equilibrium). *The players contribute δ units regardless of their network positions.*

Note that the presence of a dominant strategy in our experiment rules out the possibility of risk attitudes to affect behavior (because there is no risk in decision-making). At the same time, it largely reduces the cognitive burden associated with maximizing one's own profit. As a result, the remaining bias in decision-making should be attributed to the concern for others' profits—the *social preference*. We then develop two competing models that embed social preferences in network games. In Section 2.2 we examine a model of welfare preference, and Section 2.3 features a study of a model of inequality aversion. Both models assume that a player's utility increases with her own payoff, while the models differ by the treatment of other's payoffs in one's utility function.

2.2. The Model of Welfare

In the welfare model, an individual's utility amounts to a weighted sum of her own payoff and those of her neighbors. This model is based on the framework of Charness and Rabin (2002), but developed here into a network context. Formally, we have

$$\Pi_i(\boldsymbol{x}) = \left(1 - \sum_j \lambda_{ij} g_{ij}\right)\pi_i(\boldsymbol{x}) + \sum_j \lambda_{ij} g_{ij} \pi_j(\boldsymbol{x}), \quad (2)$$

where $\pi_i(\boldsymbol{x})$ is player i's pecuniary payoff defined in (1), and the coefficient λ_{ij} captures the relative importance of player j's payoff in i's objective. The term $1 - \sum_j \lambda_{ij} g_{ij}$ is the weight that player i reserves for her own monetary payoff, which we assume is nonnegative.[7] If the total weight assigned to others' payoffs, $\sum_j \lambda_{ij} g_{ij}$, is equal to 1, player i is purely altruistic or a disinterested individual who seeks to maximize the surplus of all other players. However, if $\sum_j \lambda_{ij} g_{ij} = 0$, player i becomes *homo economicus* as assumed in standard game theory. When the network is complete ($g_{ij} \equiv 1$), our welfare model becomes a multi-person variant of the model described in the appendix of Charness and Rabin (2002).[8] We can also rewrite the welfare utility function as $\Pi_i(\boldsymbol{x}) = \left(1 - \sum_j L_{ij}\right)\pi_i(\boldsymbol{x}) + \sum_j L_{ij} \pi_j(\boldsymbol{x})$, where $L_{ij} \coloneqq \lambda_{ij} g_{ij}$. This expression can be interpreted as stating that the individual assigns different weights to different persons, where a nonzero weight may exist for a person not connected to the focal individual. Most of our results for the welfare model can be easily extended for this more general specification; however, here we write $\lambda_{ij} g_{ij}$ to separate the effect of the network structure from that of altruism.

THEOREM 1 (Welfare equilibrium). *The welfare equilibrium profile* $\boldsymbol{x}^*$ *on network* G *is such that, for each player* i, $x_i^* = \frac{\delta}{1 - \sum_j \lambda_{ij} g_{ij}}$. *If also* $\lambda_{ij} \equiv \lambda_i$ *for all* j, *then*

---

[7] Similar in spirit to the welfare preference in our paper, Neilson and Wichmann (2014) develop a model on the homophily among network neighbors in their valuations of an exogenous public good. Unlike Neilson and Wichmann (2014), we treat the public good as endogenous. Our paper also differs from theirs in that (i) we develop an additional model of inequality aversion, and (ii) the theoretical predictions of our models are contrasted and tested by laboratory experiment.

[8] We simplify certain aspects of the CR model (e.g., by removing the weight on the worst-off payoff) and also normalize the weights to put forth the central scheme of weighted welfare maximization and embed it into the network structure.

$$x_i^* = \frac{\delta}{1-\lambda_i d_i} \tag{3}$$

All the proofs are provided in Appendix B. Under the welfare preference, a player cares about the collective wealth within her neighborhood. Hence, Theorem 1 states that a player's contribution increases with $\sum_j \lambda_{ij} g_{ij}$, or her connectivity weighted by the strength of her welfare concern. If one weights neighbors' payoffs indiscriminately, then her equilibrium contribution level is increasing in the number of her neighbors (d_i). In other words, a welfarist will raise his contribution when it benefits more people. As suggested in Theorem 1, the equilibrium play under welfare preference exhibits the following pattern.

REMARK 2 (Welfare equilibrium). *Players contribute more than δ independent of their network positions. Moreover, in a given network, players with more neighbors make higher contributions.*

2.3. The model of inequality aversion

In the seminal work of Fehr and Schmidt (1999), a player suffers disutility if her payoff is either less or greater than that of any of her peers. In this section, we shall develop their model into a network context. To be specific, player i maximizes her utility as follows:

$$\Pi_i(\boldsymbol{x}) = \pi_i(\boldsymbol{x}) - \sum_j \alpha_{ij} g_{ij} \left(\pi_j(\boldsymbol{x}) - \pi_i(\boldsymbol{x})\right)^+ - \sum_j \beta_{ij} g_{ij} \left(\pi_i(\boldsymbol{x}) - \pi_j(\boldsymbol{x})\right)^+, \tag{4}$$

where $\pi_i(\boldsymbol{x})$ is one's (pecuniary) payoff defined as in equation (1), $(y)^+ \coloneqq \max\{y, 0\}$, and the coefficients α_{ij} [β_{ij}] respectively represent the marginal reduction of player i's utility for each unit by which her monetary payoff is lower [higher] than that of her neighbor j. We assume that $\sum_j \beta_{ij} g_{ij} < 1\ \forall i$.[9] In our model, players are concerned about the fairness of payoff distributions within their social neighborhoods but not about the fairness over the entire society. This could be the case where individuals either (i) do not observe the payoffs of those who are not their friends or colleagues, or (ii) only use their friends or colleagues as their reference groups.[10] In our terminology, Fehr and Schmidt (1999) assume a complete network with everyone connected to everyone else. In contrast, our model generalizes Fehr and Schmidt (1999) by allowing an agent to

---

[9] This assumption is compatible with existing evidence that many subjects are far less averse to advantageous inequality than to disadvantageous inequality. See Loewenstein et al. (1989), Ferrer-i-Carbonell (2005), Ho and Su (2009), Immorlica et al. (2017) and the references therein.

[10] The existence of localized social comparison has garnered some empirical support, e.g. Luttmer (2005), Neumark and Postlewaite (1998).

compare his or her payoff with a specific subset of the population; and it is the network structure that determines "who compares the profit with whom."

THEOREM 2 (equilibrium of Inequality Aversion)**.** *There exists a pure strategy equilibrium in the game with inequality aversion, with any given network structure.*

Theorem 2 establishes the existence of equilibrium in the game of inequality aversion. However, note that multiple equilibria may exist under Theorem 2, which will be made clearer when we introduce the equilibria on specific networks (see the Corollaries in Appendix A). Next, we will explore some regularities of the IA equilibrium, which will be useful for subsequent empirical analysis.

REMARK 3 (equilibrium of Inequality Aversion). *In a given network, players who earn more [less] than all of their neighbors contribute more [less] than the self-interest level.*

To understand Remark 3, notice that when one's contribution increases, her neighbors' payoffs increase faster than her own payoff (which may even decrease).[11] Thus, a fairness-minded player earning higher [lower] than all neighbors will have the incentive to increase [decrease] her contribution relative to the self-interest level (δ) to reduce the payoff gap. The formal proof of Remark 3 directly follows the player's best response (Proposition A-1 in Appendix A) and is thus omitted.

3. Experiment

3.1. Experimental design

In this section, we present a simple experiment to test the competing theories outlined in Section 2. By manipulating the *network structure*, our experiment consists of three *treatments*, as recorded in Table 1 below.

Table 1. Summary of experimental design

Network Structure	Star	Circle	Complete
Treatment	*S*	*C*	*Com*

Note: between-subjects design, $\delta = 10$, network size = 4

[11] This happens because one's contribution benefits oneself and everyone connected to her indiscriminately, while one alone bears the cost of contribution. See (1).

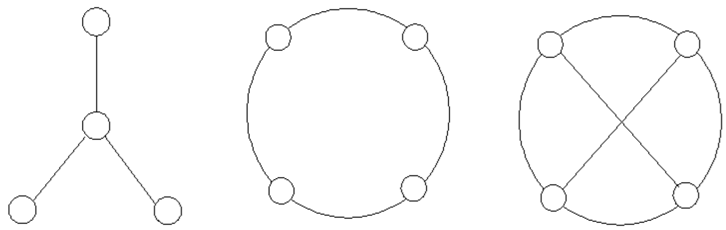


Figure 1. Experimental network structures

Our experiment employs the three network structures depicted in Figure 1: star, circle, and complete. Generally speaking, by a *star* we refer to a network with a single node at the center (called the *core*) and multiple nodes that are connected to the core but not to each other (the *periphery*). Let the star's core node be denoted by c, and denote by $1,2 \ldots N_p$ the N_p nodes at the periphery. The *circle* is a N-vertex network in which each node is connected only to the two adjacent nodes. In the *complete* network, there exists a connection between any two nodes; and we denote the number of players by N. For the graphs illustrated in Figure 1, $N_p = 3, N = 4$, so that all networks have the same size (4 players). We choose these networks for their simplicity and because they serve as miniatures of real-world social and organizational systems. The star network represents a monarchy, with the core assuming the most power or prestige; the circle is like a democratic society, with players distributed at egalitarian positions. The complete network resembles the conventional environment for game theory research (with global interactions). In all treatments, the subjects are connected according to the network structures shown in Figure 1 (depending on their treatments).

3.2. Experimental protocol

The experiment consists of 20 rounds. In each round, the subjects play the game described in Section 2.1. In the star networks, a subject is located—throughout the experiment—either at the core or at the periphery.[12] With the design of repeated play and fixed *role*s (either core or periphery in star networks), we can obtain more experienced subject behavior.[13] Meanwhile, in order to maintain the one-shot nature of the game, we hold a rematch of the subjects every round according to the network structure. Specifically, we divide each treatment into five cohorts, each of which

[12] In other words, no subject can be placed at the core in one round and at the periphery in another round. If a subject is assigned at the core [the periphery] in one round, s/he will remain at the core [the periphery] in all rounds of the experiment.

[13] This design is adopted because alternating over different roles (core or periphery) in the experiment will likely disrupt the subject's learning.

consists of three [two; two] star [circle; complete] networks in treatment *S* [*C*; *Com*]; and we restrict the rematches within each cohort. Notably, a rematch preserves the subject's role as core or periphery in the star treatments. Thus, we guarantee that the subjects' roles are fixed, whereas the partnerships constantly change; therefore, every round resembles a one-shot game. After each round, subjects were informed of their own payoffs and those of their neighbors, while the anonymity was maintained for all players.[14]

We collected data from 140 subjects in September 2020. The subjects were mainly registered students at a major public university in China. All subject participation was monetarily incentivized, and the average earning per subject was 37.06 Chinese Yuan. The software we used was programmed in z-Tree (Fischbacher, 2007).

4. Results

4.1. Aggregate patterns

OBSERVATION 1. *In all treatments, players on average contribute more than predicted by the standard theory.*

Figure 2 plots the average subject contribution in star, circle, and complete network treatments. In the figure (as well as other figures in this paper), the vertical bar represents the 95% confidence interval for the overall mean of the plotted variable; and the dashed line (when imposed) denotes the standard prediction. Figure 2 shows that the actual play deviates significantly from the self-interest benchmark. The average contribution falls below δ in none of the periods in the star treatment and complete network treatment. In the circle treatment, the average contribution by period exceeds the self-interest amount with merely a few exceptions toward the end of the experiment. The summary statistics of subject contribution is presented in Table 2. For Observation 1, one sample *t*-test clustered by cohorts shows a *p*-value of 0.001 [0.040] for core [peripheral] players in treatment *S*, $p = 0.018$ for subjects in treatment *C*, and $p = 0.002$ for those in treatment

[14] To avoid negative payoffs and associated payment issues, we set an upper bound on the contribution that can be made by a subject in the experiment. The bound is sufficiently high so that the theoretical equilibria continue to hold as they fall within the bound. Nevertheless, our models can also be formally extended to accommodate the upper bound. Such an extension could be conducted in a manner similar to our handling of the nonnegativity constraint on the amount of contribution (see, for example, the proof of Proposition A-1).

Com.[15, 16] The observed pattern of behavior is consistent with Remark 2, which is derived under the welfare preference. Note that the experimental result on the complete network replicates the existing finding regarding over-contribution in the public goods experiment with an interior dominant strategy (Keser, 1996). We also find that subject contribution slightly declines throughout the game in all treatments. This trend is in line with the common laboratory finding that contribution to the public good tends to fall over time (see, for example, Fischbacher and Gächter, 2010).

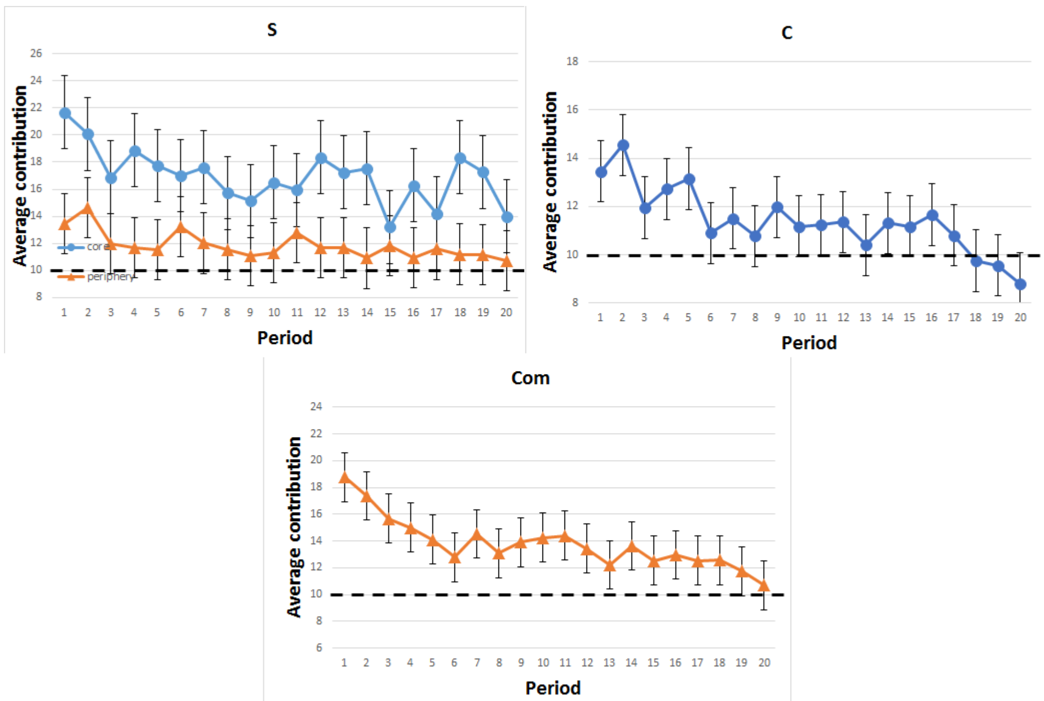


Figure 2. Average contributions in all treatments

Table 2. Summary of contributions

S		*C*	*Com*
Core	Periphery		
16.97 [0.97]	11.86 [0.80]	11.41 [0.46]	13.81 [0.66]

Mean [standard error (clustered by cohorts)]

OBSERVATION 2. *In star networks, core nodes, on average, earn more, and contribute more than do peripheral nodes.*

[15] The *p*-values are one-sided, with the patterns in Observation 1 formulated as the alternative hypotheses.

[16] The *t*-test clustered by cohorts employed in our paper takes into account the dependence across observations from the same cohort. The test is conducted in a regression framework so that the *t*-test problem is transformed to a test of the regression coefficient. Thus, we could apply the standard regression approach of clustering the standard errors on the cohort level.

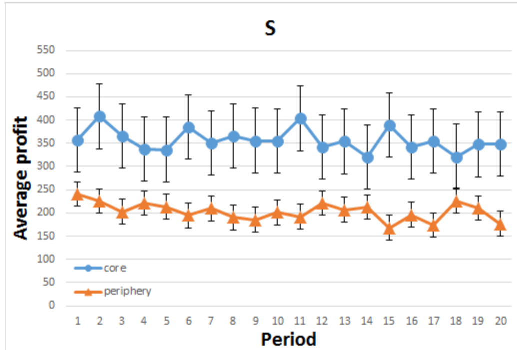


Figure 3. Payoffs in the star treatment

Table 3. Payoffs in the star treatment

S	
Core	Periphery
357.01	203.51
[24.97]	[9.54]

Mean [standard error (clustered by cohorts)]

In the star networks, we find that core subjects make a higher profit than peripheral ones (Figure 3). Table 3 shows that the profit of the core dominates that of the periphery (p-value = 0.004 from a two-sample t-test clustered by cohorts). If peripheral subjects were to pursue fairness, they should contribute less than δ (by Remark 3). However, we find their contribution to be higher than the self-interest level (Observation 1), which runs counter to inequity aversion. Moreover, as Figure 2 suggests, the core players contribute more than do peripheral ones (p-value = 0.006 from a two-sample t-test clustered by cohorts) in the star networks,[17] which supports Remark 2.

OBSERVATION 3. *In circle and complete networks, subjects who earn less than all of their neighbors on average contribute less, but still more than* δ, *in the subsequent period.*

In treatments *C* and *Com*, we find the average contribution tends to fall for the subjects who earned less than all their neighbors in the previous round. This case is plotted in Figure 4 as *lag profit < neighbors'* and *lowest profit in the previous round* in *C* and *Com* respectively. Nevertheless, even in this case, the reduced contribution remains higher than δ, which likely stems from the welfare preference. These are visualized in Figure 4, where the change in the amount of contribution over consecutive periods appears more often negative than positive in the circle treatment, and uniformly negative in the complete network treatment, for those having earned below all neighbors. Meanwhile for those subjects, the average contribution levels in most periods

[17] For the tests for Observation 2: All the p-values are one-sided, with the patterns in Observation 2 formulated as the alternative hypotheses.

stay higher than δ in both treatments (Figure 4).[18] For Observation 3, one-sample *t*-tests clustered by cohorts show a *p*-value of 0.003 for treatment *C* and 0.002 for treatment *Com* (*contribution reduction*), and $p = 0.010$ for *C* and 0.008 for *Com* (*contribution higher than* δ).[19]

When having earned less than all neighbors, a fairness-minded subject would tend to decrease his or her contribution to reduce the payoff gap (Lemma A-1). This is consistent with Observation 3. However, for a fairness equilibrium to sustain, the contribution has to fall below δ in the case of earning less than all of one's neighbors (Remark 3), which is clearly rejected by the experimental result. Overall, although Observation 3 reflects a mix of welfare concern and inequality aversion, the evidence for the former seems stronger than that for the latter. As we will see in Section 4.2, the preference for welfare indeed outweighs the aversion to inequality in the circle and the complete networks at the individual level.

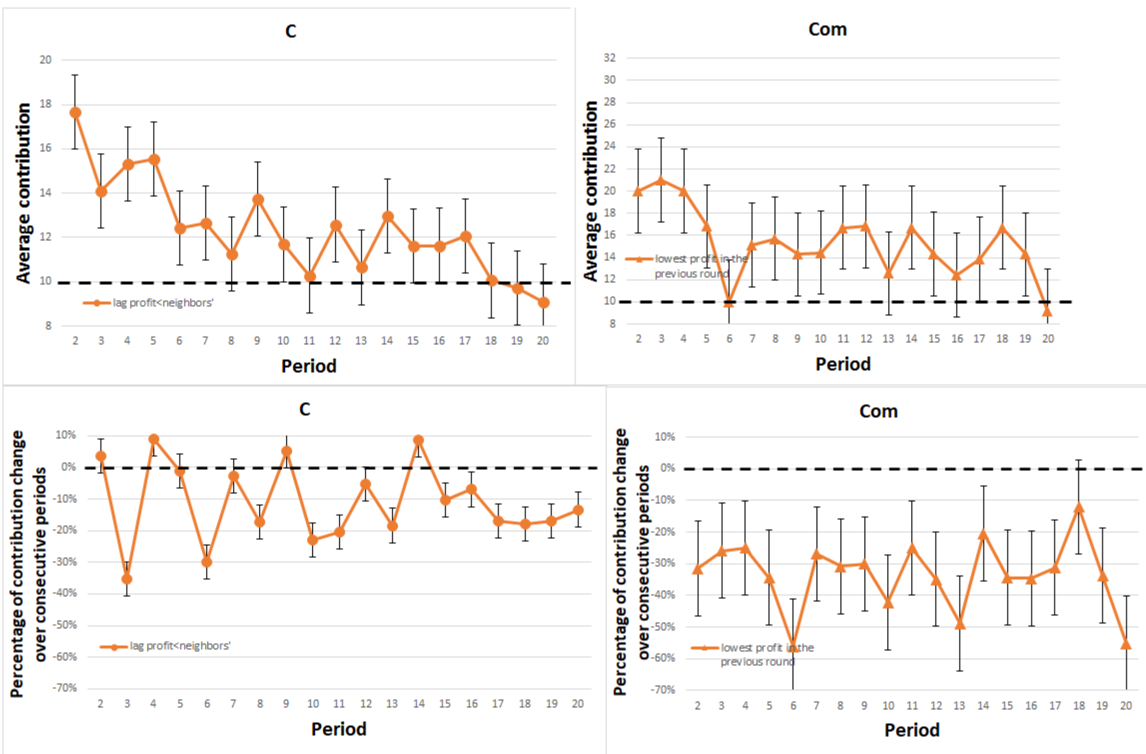


Figure 4. Behavior in treatments *C* and *Com* conditional on the focal player earning less than all neighbors in the previous round

[18] In the circle and the complete network treatments, our analysis accounts for the potential lag effect in subjects' decision-making, as they cannot see and respond to their neighbors' profits until the next decision round. As for the analysis of the star treatments, considering the lag effect would make little difference, because the average profit of the core is uniformly higher than that of the periphery in every period.

[19] The *p*-values are one-sided, with the patterns in Observation 3 formulated as the alternative hypotheses.

In sum, these experimental results suggest that the theory based on self interest is largely inconsistent with the actual behavior in network games. In general, subjects contribute more than what would maximize their own profit. When they are at the central position in the star network, the subjects earn more, and make more contribution, than those at the outer positions. Furthermore, subjects in the circle and the complete networks reduce their contribution while still contributing more than the self-interest amount when they earned less than all their neighbors. These behavioral patterns indicate that subjects tend to exhibit more preference for welfare than aversion to inequality. In the next section, we will see that the same conclusion also holds at the individual level. Our experimental design also allows us to explore the network effects from an angle that is not captured by equilibrium analysis. Accordingly, see Appendix C.2 for an investigation of the global network effect.

4.2. Individual-level analysis

Having established that the welfare preference dominates both self-interest and inequality aversion at the aggregate level, we then commence an individual-level investigation and estimation of behavioral parameters. We fit the behavioral parameters $\alpha_i, \beta_i, \lambda_i$ for each individual i.[20] As such, the estimation is based on the welfare equilibrium characterized in (3), as well as the equilibria of inequity aversion laid out in explicit forms in Corollaries A-1, A-2, and A-3 (in Appendix A). For the maximum likelihood estimation (MLE), we attach a zero-mean and normally distributed noise to the theoretically predicted action of each player i, and also estimate its standard deviation (σ_i).

Table 4 records the log-likelihood (LL) and Akaike information criterion (AIC) values of the models in all treatments. In all the treatments, the welfare model achieves a higher likelihood than that of inequality aversion, and does so with fewer parameters. Consequently, the welfare model has unambiguously lower AIC values, which favors it over inequity aversion in the model selection. The welfare model is also superior to the self-interest model in terms of the AIC comparison. For expositional convenience, we shall respectively abbreviate the welfare, inequality aversion, and self-interest models with W, IA, and SI.

[20] We assume that an individual i treats her neighbors indifferently in assessing her utility. In other words, we set $\alpha_{ij} \equiv \alpha_i, \beta_{ij} \equiv \beta_i$ in (4) and $\lambda_{ij} \equiv \lambda_i$ in (2) for all j in the fairness and the welfare model.

Table 4. Model comparison*

Treatment	*S*	*C*	*Com*
LL	-2465.80 [-2994.73;-3092.29]	-1506.06 [-1577.72;-1582.37]	-1915.26 [-2055.44; -2070.95]
AIC (df) [21]	5171.59 (120) [6349.45 (180); 6304.57 (60)]	3172.13 (80) [3395.43 (120); 3244.73 (40)]	3990.53 (80) [4350.88 (120); 4221.90 (40)]

* Model W [Model IA; Model SI]

Why does the pursuit of welfare outweigh inequality aversion in subjects' behavior in our experiment? We argue that the social connections in our experiment may impart a sense of *group* to the connected agents, so that they feel more affiliated when making decisions. This conjecture is supported by the study on social identity, which shows the group saliency can lead individuals to pursue the group's goal (Akerlof and Kranton, 2000; Akerlof and Kranton, 2005; Kranton, 2016). In particular, Chen and Li (2009) demonstrate that subjects exhibit significantly less envy and more charity with partners from the same group and that the in-group subject behavior can be well interpreted through the Charness and Rabin (2002) model. This echoes the prevalence of welfare preference observed in our setting, which resembles a *networked* version of the Charness and Rabin (2002) model. Further, it has also been shown in simple distribution experiments that the efficiency preference dominates inequality aversion (see Engelmann and Strobel, 2004; Engelmann and Strobel, 2006 and references therein). In our experiment, we find that the same pattern of model selection also prevails in games in networks.

Table 5 reports detailed parameter estimates of the behavioral models, where the notations with subscript c and p respectively denote the parameter estimates for the core and peripheral subjects in the star networks. In addition, σ_i^{IA} $[\sigma_i^{W}]$ denotes the individual σ-estimates under the inequity aversion [welfare] model. As an exemplary interpretation of results from the welfare model, note that λ_c averages 0.128 in treatment *S*. That means, for an average subject at the core, the profit of a single neighbor of hers constitutes 12.8% of the focal subject's utility. As the subject has three such neighbors, she ends up with 38.4% of interest attributed to neighbor profits, and 61.6% to her own.

[21] Recall that each treatment includes 5 cohorts, each star [circle; complete network] cohort contains 3 [2; 2] networks, and each network comprises 4 subjects; hence, there are 60 [40; 40] subjects in each star [circle; complete network] treatment. For each individual, the IA [welfare] model has 3[2] parameters, while the baseline model involves only the σ_i-term. Altogether, these give the degrees of freedom (df) specified in Table 5.

Table 5. MLE results

treatment	Model IA					Model W		
	α_c	α_p	β_c	β_p	σ_i^{IA}	λ_c	λ_p	σ_i^{W}
S	0.066*	0.048	0.108	0.147	5.246	0.128	0.142	4.168
	(0.040)	(0.022)	(0.010)	(0.041)	(0.571)	(0.011)	(0.036)	(0.341)
	α_i		β_i		σ_i^{IA}	λ_i		σ_i^{W}
Com	1.78×10^{-7}*		0.042		6.477	0.081		5.167
	(8.68×10^{-8})		(0.011)		(0.831)	(0.011)		(0.699)
C	0.001*		0.025		3.875	0.055		3.319
	(0.001)		(0.009)		(0.797)	(0.012)		(0.623)

Mean (standard error clustered by cohort)
All parameters significant at 0.05 level except for those marked with * (per *t*-test clustered by cohort)

Some remarks are in order concerning the estimation of the inequality aversion model. As the star network naturally leads to asymmetric payoffs, both core and peripheral players will expect that the earning is higher for the core and lower for the periphery. In this case, whether a fairness concern will emerge depends on to what degree this inequality is "within tolerance." In treatment *S*, the core subjects earn, on average, approximately 1.75 times as much as the peripherals (see Table 3). As such, a peripheral player with $\alpha_p = 0.048$ (as estimated in Table 5) incurs a fairness disutility that amounts to about $0.048 \times (1.75 - 1) = 3.6\%$ of the monetary profit s/he earns. For a core player with $\beta_c = 0.108$ (as estimated in Table 5), s/he suffers from a disutility that accounts for $3\beta_c\left(1 - \frac{1}{1.75}\right) = 13.9\%$ of his/her financial payoff. Thus, the loss from inequality can be nontrivial (especially for the core players) if the core earns more than the periphery. Furthermore, if the actual outcome of the game goes against the players' expectations—if it turns out that the core earned lower than the periphery in some period—the players will experience even stronger fairness disutility. A core subject would be more upset if she ends up receiving less payoff than the periphery than would a peripheral subject when earning less than the core ($\alpha_c > \alpha_p$). Meanwhile, a player at the periphery would feel sorrier for a core earning less than her, than would a core player when earning more than the periphery ($\beta_c < \beta_p$). Both $\alpha_c > \alpha_p$ and $\beta_c < \beta_p$ reflect how the player's network position shapes her perception of fairness; notably, this is indeed what we observed with average parameter estimates in Table 5 (i.e., 0.066 > 0.048 and 0.108 < 0.147, respectively). In addition, we observe that $\alpha_p < \beta_p$ in treatment *S* and $\alpha_i < \beta_i$ in treatment *C* and *Com*.[22] Although the fact $\alpha < \beta$ is inconsistent with the common understanding that people are more averse to disadvantageous inequality than to advantageous inequality (see e.g. Ho and Su, 2009; Immorlica et al., 2017), this inconsistency in our case reflects that the welfare model explains the data better than that of inequality aversion. To see this, note that welfare preference and inequality aversion differ in their treatments of neighbors' payoffs that are higher than one's own. In such cases, a fairness-minded player wants to *reduce* her neighbor's payoff, whereas a player endowed with welfare preference wants to *increase* her neighbor's payoff. Hence, a forced estimation of the inequity aversion model (where α is constrained to be nonnegative), when the underlying "true" model is that of welfare preference, should yield values of α that are close to zero (because the "true" α-values should be less than zero). Subsequently, this leads to $\alpha < \beta$.

Our results in this section can be summarized as follows. We show that the welfare preference outperforms both inequality aversion and self-interest in driving the individual behavior in the network games we examine. We also find that one's perception of fairness is leveraged by her network position, which explains the observed patterns of estimated parameters, α and β.

5. Effects of Economic Return

In this section, we conduct an additional experiment to examine the effect of increasing economic return and its interplay with the network structure. We denote by *HS* and *HC* the new treatments involving star (*S*) and circle (*C*) networks under a higher (*H*) economic return ($\delta = 30$). Consistent with the foregoing experimental protocol, the treatment *HS* [*HC*] consists of five cohorts, each of which entails three [two] four-player star [circle] networks as shown in Figure 1. In total, we recruited 60[40] subjects for *HS* [*HC*] from the same subject pool used for the main experiment in Section 3. Each subject plays 20 rounds of the game described in Section 2.1 with non-repetitive partners who are rotated within cohorts so that every round resembles a one-shot game. In the star networks, the role of subjects is fixed (either core or periphery) throughout the experiment. After each round, the subjects are informed of the payoffs of their own and those of their neighbors, while the anonymity is maintained for all players. Other procedures parallel those applied to the main experiment in Section 3. Appendix C provides supplemental materials for the experiment, including the organization and the matching protocol within each cohort (Appendix C.1), sample instruction (Appendix C.3), and sample software screenshots (Appendix C.4).

---

[22] The statement $\alpha_p < \beta_p$ in *S* is weakly significant with $p = 0.058$ from a one-sided *t*-test clustered by cohorts. The same test produces $p = 0.035$ [0.011] for the statement $\alpha_i < \beta_i$ in *C* [*Com*].

OBSERVATION 4. *In the star network, the welfare preference remains dominant over inequality aversion and self-interest with an increased economic return. However, in the circle network, the advantage of social preferences over self-interest is weakened with an increased economic return.*

In *HS*, players at both the core and periphery contribute more than δ at the aggregate level.[23] As presented in Figure 5, the average contribution of the core stays above δ in every period—so does that of the periphery, with minor exceptions in 2 rounds out of 20. Conversely, the contribution in *HC* starts off high but later slides down to a lower level (Figure 5). Inferred from Table 6, the average contribution in *HC* departs from the profit maximizing level only by 2.23% (i.e. $\frac{|29.33-30|}{30} = 2.23\%$). In fact, a one-sample *t*-test clustered by cohorts fails to reject the null hypothesis that the contribution in *HC* never deviates from δ (with a two-sided *p*-value = 0.634). On the contrary, recall that the average contribution in treatment *C* differs from the self-interest level by 14.1% (see Table 2). Thus, the self-interest seems to be enhanced in the circle networks when the economic return rises higher.

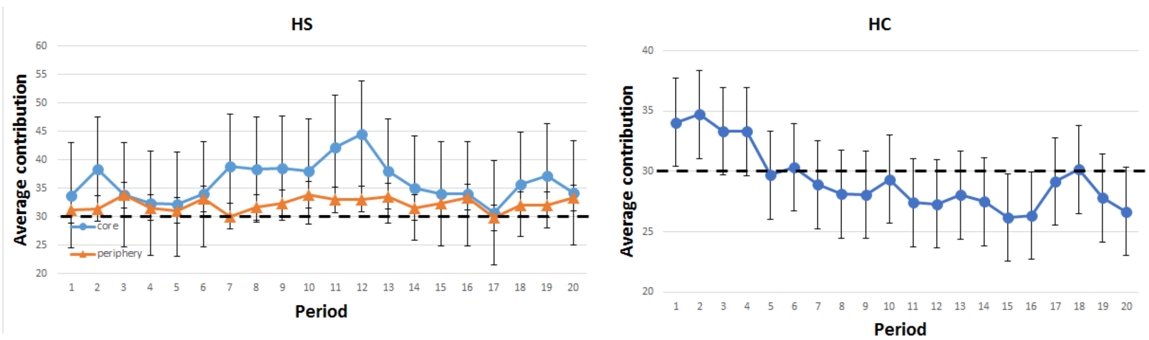


Figure 5. Average contributions in treatments with increased economic return

Table 6. Summary of contributions in treatments with increased economic return

HS		*HC*
Core	Periphery	
36.16 [3.31]	32.21 [0.81]	29.33 [1.31]

Mean [standard error (clustered by cohorts)]

As displayed in Figure 6 and Table 7, core nodes earn more than peripheral nodes in treatment *HS*.[24] In this case, Remark 3 indicates that a peripheral player concerned with fairness should contribute below δ. However, we have observed the peripheral contribution being greater than δ

[23] For this statement, *p*-value = 0.027 for peripheral players by the one-sample *t*-test clustered by cohorts. The same test produces weak significance (*p*=0.068) for the core players.

[24] One-sided *p*-value = 0.000 from two-sample *t*-test clustered by cohorts

(see the preceding paragraph), thereby yielding a contradiction to the prediction of inequality aversion. In treatment *HC*, we find the contributions are most likely to shrink for subjects who earned less than both of their neighbors in the previous round.[25] Specifically, see Figure 7 for an illustration by periods, where most contribution changes (in percentages) are negative. As explained in Section 4.1, this is consistent with the conjecture of inequality aversion, as lowering one's contribution in this case can help reduce the gap between one's profit and her neighbors' (Lemma A-1).

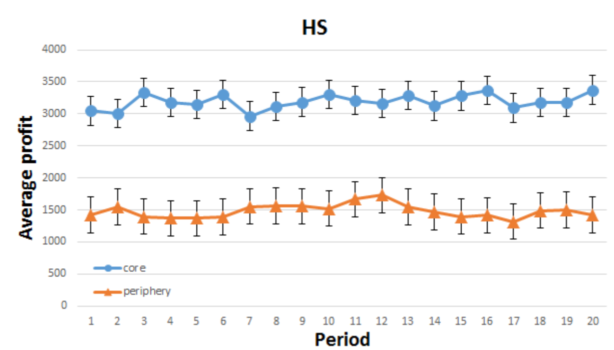


Figure 6. Payoffs in *HS*

Table 7. Payoffs in *HS*

HS	
Core	Periphery
3191.05	1481.91
[79.91]	[100.18]

Mean [standard error (clustered by cohorts)]

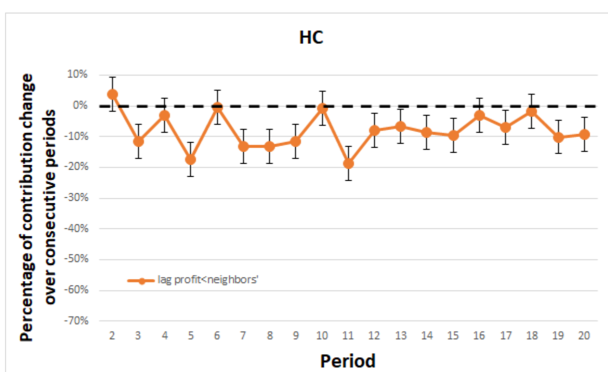


Figure 7. Change of contribution in *HC*
when the focal player earned less than both neighbors in the previous round

We also estimate the behavioral models based on Corollaries A-1, A-2 and (3) in Theorem 1 by following the same approach outlined in Section 4.2. As suggested by Table 8, the welfare model

[25] One-sided p-value = 0.010 from one-sample t-test clustered by cohorts

has the unambiguously lowest AIC value in treatment *HS*, which favors it over inequity aversion and self-interest in the model selection. Nevertheless, the performance gaps between models become substantially smaller in treatment *HC*, with the welfare model winning by a narrow margin—excelling in AIC by only 0.7% over the self-interest model. The parameter estimation in Table 9 also confirms the patterns $\alpha_c > \alpha_p$, $\beta_c < \beta_p$, and $\alpha_p < \beta_p$ in treatments *HS*.[26] These echo the results in the main experiment (Section 4) and reflect the way one's network location affects her fairness concern and the fact that the welfare preference outperforms inequality aversion in capturing the behavior in the star networks (see Section 4.2 for analogous arguments). In addition, compared to the parameter estimates in the main experiment (Table 5), the estimated λs for core, periphery, and circle subjects are all lower in treatments with an increased economic return.[27] This observation suggests that if the stake becomes higher, the individual cares more about oneself.

Table 8. Model comparison*

Treatment	*HS*	*HC*
LL	-3389.39 [-3705.55;-3777.66]	-2277.14 [-2305.51; -2332.84]
AIC (df) [28]	7018.78 (120) [7771.10 (180); 7675.32 (60)]	4714.27 (80) [4851.02 (120); 4745.67 (40)]

* Model W [Model IA; Model SI]

[26] For $\alpha_c > \alpha_p$, a two-sample t-test clustered by cohorts yields a one-sided p-value of 0.001. The same test indicates weak significance for $\beta_c < \beta_p$ with $p = 0.089$. The statement $\alpha_p < \beta_p$ is also statistically significant with $p = 0.046$ by a one-sided t-test clustered by cohorts.

[27] For this statement, two-sample t-tests clustered by cohorts yield a one-sided p-value of 0.048 for λ_p and 0.001 for λ_c in star treatments and 0.047 for λ_i in circle treatments.

[28] Recall that each treatment includes 5 cohorts, each star [circle] cohort contains 3 [2] networks, and each network comprises 4 subjects; hence, there are 60 [40] subjects in each star [circle] treatment. For each individual, the IA [welfare] model has 3 [2] parameters in the star treatment and 3 [2] parameters in the circle treatment, while the baseline model involves only the σ_i-term. Altogether, these give the degrees of freedom (df) as specified in Table 5.

Table 9. MLE results

treatment	Model IA					Model W		
	α_c	α_p	β_c	β_p	σ_i^{IA}	λ_c	λ_p	σ_i^{W}
HS	0.330	0.023	0.055	0.209	9.425	0.060	0.072	8.202
	(0.052)	(0.007)	(0.014)	(0.088)	(0.842)	(0.014)	(0.017)	(0.799)
	α_i		β_i		σ_i^{IA}	λ_i		σ_i^{W}
HC	0.044*		0.013		9.115	0.029		8.730
	(0.030)		(0.004)		(2.394)	(0.009)		(2.494)

Mean (standard error clustered by cohort)
All parameters significant at 0.05 level except for those marked with * (per *t*-test clustered by cohort)

In summary, we find that when the economic return rises, the welfare preference remains more prevalent than inequality aversion and self-interest in the star networks. In contrast, self-interest gains more ground in the circle networks under a higher stake. Our estimation confirms the patterns of parameters observed in the foregoing main experiment and illustrates how one's network location and the level of economic return influence the extent of social preferences.

6. Conclusion and Discussion

We investigate social preferences in network games where each player engages with an exogenous subset of the player population. The network is configured as a set of relationships that establish "whose action affects whose payoff." We develop theoretical models that incorporate inequality aversion (Fehr and Schmidt, 1999) and welfare preference (Charness and Rabin, 2002) into a baseline dominant-strategy network game, and test our theories with experimental data. When the economic return is relatively low, subjects are observed to make higher contributions than the self-interest amount. We also find that subjects at the core positions contribute more than those at the periphery. These behavioral regularities collectively suggest welfare preference—rather than inequality aversion or self-interest—as the driving force in the game play in all the networks considered. When the economic return is increased, however, the advantage of social preferences over self-interest in shaping subject behavior tends to vary with the network structure. Our estimation of the behavioral parameters reveals how one's perceptions of fairness and welfare are affected by the network topology.

In their seminal work, Fehr and Schmidt (1999) (p.851) remark that the effect of network topology on inequality aversion is a vital subject for future research: "*Another set of questions concerns the choice of the reference group. ... There may well be interactive structures in which some agents have a salient position that makes them natural reference agents.*" This idea is materialized and analyzed in our paper. In a similar vein, we also extend the classic model of welfare preference (Charness and Rabin, 2002) into the network environment. This study offers a starting point for testing social preferences on more realistic social or economic networks, and developing more general models on social preferences in network games.

Our current experiments were not conducted on real social networks. Instead, we arbitrarily assigned individuals to our experimental networks. The tradeoff we faced in this aspect (control vs. realism) is typical in the laboratory experiments. Nonetheless, it is reasonable to suppose that using real social networks would actually enhance the welfare preference because true friends presumably care more about each other's profit than do random partners. Also note that our present experimental setup does not involve communication between neighbors. However, as communication can enhance cooperation in public goods provision (see Chaudhuri, 2011; Ledyard et al., 1995), neighbors shall be more aware of the group's goal when the communication is in place. Consequently, the welfare preference may further firm up in the presence of communication.

A promising direction for future research would concern the learning and dynamics in the behavior of network games. For example, we observe that there is usually a declining trend in player contributions over time. In the context of public goods provision, Fischbacher and Gächter (2010) has explained this downward contribution as a consequence of imperfect conditional cooperation – that is, the players only partly match the contributions of others at every iteration. Whether a similar mechanism is responsible for the intertemporal pattern found in our setting, and how such mechanics interact with the network layout, are subjects that merit further exploration.

Our present experiment features a dominant strategy in the base game, as a means to control for nuisance factors (e.g., risk attitudes, cognitive complexity) other than social preferences. Nevertheless, a future study beyond this design could incorporate complementarity or substitutability among neighbors' contributions. Whether adding such "synergy" in neighbor actions could alter the preference of players remains an interesting question.

Owing to the page limit, the appendices can be separated from the manuscript and posted online. The notations in the appendices inherit from those in the main text unless otherwise clarified.

Appendix A. The Equilibrium of Inequality Aversion

A.1. Strategy and best response

Given the individual payoff function in the main text as (1), a marginal increase of x_i will cause the payoff of *every* neighbor of player i to increase by δ. Thus, one's own action x_i will *not* unilaterally change the order of her neighbor's payoffs. Given the action profile $\boldsymbol{x}$, denote by $\pi_{(1)_i}(\boldsymbol{x})$, $\pi_{(2)_i}(\boldsymbol{x}) \dots \pi_{(d_i)_i}(\boldsymbol{x})$ the 1st, 2nd, …d_ith *least* pecuniary payoff of player i's neighbor. Given $\boldsymbol{x}_{-i}$, let $X_{(k)_i}(\boldsymbol{x}_{-i})$ represents the threshold contribution level for player i, with which i's payoff equals $\pi_{(k)_i}(\boldsymbol{x})$, that is, $\pi_i\left(X_{(k)_i}(\boldsymbol{x}_{-i}); \boldsymbol{x}_{-i}\right) = \pi_{(k)_i}(\boldsymbol{x})$. Then the following lemma holds:

LEMMA A-1. $X_{(d_i)_i}(\boldsymbol{x}_{-i}) < X_{(d_i-1)_i}(\boldsymbol{x}_{-i}) \dots < X_{(2)_i}(\boldsymbol{x}_{-i}) < X_{(1)_i}(\boldsymbol{x}_{-i})$.

The proof of Lemma A-1 is straightforward. For player i, $\frac{\partial \pi_i(x)}{\partial x_i} = \delta - x_i$, while for i's neighbor j, $\frac{\partial \pi_j(x)}{\partial x_i} = \delta$. $\frac{\partial \pi_i(x)}{\partial x_i} < \frac{\partial \pi_j(x)}{\partial x_i}$. Thus, when x_i increases, player i's neighbors' payoffs increase faster than player i's payoff (which may even decrease). Subsequently, Lemma A-1 follows. Without loss of generality, normalize $X_{(0)_i}(\boldsymbol{x}_{-i}) = +\infty$ and $X_{(d_i+1)_i}(\boldsymbol{x}_{-i}) = -\infty$, $\forall i$. We refer to the interval $\left(X_{(k)_i}(\boldsymbol{x}_{-i}), X_{(k-1)_i}(\boldsymbol{x}_{-i})\right)$ as *range-k*. Let $\underline{d_i} := \max\left\{t \middle| X_{(t)_i}(\boldsymbol{x}_{-i}) \geq 0, t \leq d_i\right\}$. In other words, $\underline{d_i}$ is the highest rank of i's payoff in her neighborhood that player i can ever attain by reducing her contribution level, given $\boldsymbol{x}_{-i}$. By definition, this rank must be upper bounded by i's number of neighbors—$\underline{d_i} \leq d_i$. Let ϕ be the best response function to the contribution profile $\boldsymbol{x}$, $\phi \equiv (\phi_i)_{\text{all } i}$. Elementwise, ϕ_i is the best response function for an individual player i to her neighbor action $\boldsymbol{x}_{-i}$. Now we are ready to present a foundational result for the model of inequality aversion.

PROPOSITION A-1. a) *Player i's inequality-averse utility $\Pi_i(x_i; \boldsymbol{x}_{-i})$ is piecewise concave with respect to her own contribution, x_i.* b) *There exists r such that the unique best response*

$$\phi_i(\boldsymbol{x}_{-i}) = \begin{cases} \frac{\delta}{1+\sum_{j=(r)_i}^{(d_i)_i}\alpha_{ij}-\sum_{j=(1)_i}^{(r-1)_i}\beta_{ij}}, & \text{if } \frac{\delta}{1+\sum_{j=(r)_i}^{(d_i)_i}\alpha_{ij}-\sum_{j=(1)_i}^{(r-1)_i}\beta_{ij}} \in \left(X_{(r)_i}(\boldsymbol{x}_{-i}), X_{(r-1)_i}(\boldsymbol{x}_{-i})\right) \\ X_{(r)_i}(\boldsymbol{x}_{-i}), & \text{if } X_{(r)_i}(\boldsymbol{x}_{-i}) \in \left(\frac{\delta}{1+\sum_{j=(r)_i}^{(d_i)_i}\alpha_{ij}-\sum_{j=(1)_i}^{(r-1)_i}\beta_{ij}}, \frac{\delta}{1+\sum_{j=(r+1)_i}^{(d_i)_i}\alpha_{ij}-\sum_{j=(1)_i}^{(r)_i}\beta_{ij}}\right) \end{cases}$$

c) *A contribution profile* $\boldsymbol{x}$ *is an equilibrium of inequality aversion, if and only if* $\boldsymbol{x} = \phi(\boldsymbol{x})$**.**

One can show that $\Pi_i(x_i; \boldsymbol{x}_{-i})$ is concave with respect to x_i within any range-k, and is continuous at any point $X_{(k)_i}(\boldsymbol{x}_{-i})$. Hence, the best response of player i to her neighbor action $\boldsymbol{x}_{-i}$ is obtained at either an interior point (which falls in the range that implements the corresponding payoff order) or a boundary point (which equalizes the player's payoff with that of a neighbor of hers). An analogy to our equilibrium characterization in the network game literature is Proposition 1 from Bramoullé et al. (2014), where their equilibrium conditions (i) and (ii) correspond to the interior and corner best responses in our equilibrium of inequality aversion (i.e., Proposition A-1, part b). It is important to note that Proposition A-1 holds for *arbitrary network topology*. Proposition A-1 lays the foundation for the derivation of equilibria under inequality aversion in the star network (Corollary A-1), the circle network (Corollary A-2), and the complete network (Corollary A-3). The continuity and piecewise concavity of utility function, as revealed in Proposition A-1, is rudimentary to the existence of pure strategy equilibrium under inequality aversion (Theorem 2).

A.2. Equilibrium on the experimental networks

Theorem 2 establishes the existence of pure strategy equilibrium with inequality aversion. However, for the experimental networks we used, we have to obtain the explicit form of equilibria for parameter estimation. To proceed, more notations are required. For the star network, given an action profile $\boldsymbol{x}$, order the peripheral nodes 1,2… N_p such that their payoffs increase with their index—$\pi_1(\boldsymbol{x}) < \pi_2(\boldsymbol{x}) < \cdots < \pi_{N_p}(\boldsymbol{x})$. We consider all possible scenarios as follows: 1) *Type-1 payoff separating* (PS1), where the core's payoff is higher than all peripheral nodes' payoffs: $\pi_1(\boldsymbol{x}) < \pi_2(\boldsymbol{x}) < \cdots < \pi_{N_p}(\boldsymbol{x}) < \pi_c(\boldsymbol{x})$; 2) *Type-2 payoff separating* (PS2), where the core's payoff is lower than all peripheral nodes' payoffs: $\pi_c(\boldsymbol{x}) < \pi_1(\boldsymbol{x}) < \pi_2(\boldsymbol{x}) < \cdots < \pi_{N_p}(\boldsymbol{x})$; and 3) *Payoff embedding* (PE), where the core's payoff lies in the midst of peripheral nodes' payoffs: there exists r such that $\pi_1(\boldsymbol{x}) < \cdots < \pi_{r-1}(\boldsymbol{x}) < \pi_c(\boldsymbol{x}) < \pi_r(\boldsymbol{x}) < \cdots < \pi_{N_p}(\boldsymbol{x})$. Denote by

x_c^* and x_p^* the equilibrium action of core and periphery, respectively, and denote the equilibrium profile by $\boldsymbol{x}^* := \left(x_c^*, x_p^*\big|_{p=1,2\ldots N_p}\right)^{\mathrm{T}}$.

COROLLARY A-1 (equilibrium of inequality aversion in star). *There are three possible equilibria of inequality aversion in a star.*

Case PS1: $x_p^* = \frac{\delta}{1+\alpha_p}$, *for* $p \in \{1,2 \ldots N_p\}$, *and* $x_c^* = \frac{\delta}{1-\beta_c N_p}$ *if* $\pi_1(\boldsymbol{x}^*) < \cdots < \pi_{N_p}(\boldsymbol{x}^*) < \pi_c(\boldsymbol{x}^*)$.

Case PS2: $x_p^* = \frac{\delta}{1-\beta_p}$, *for* $p \in \{1,2 \ldots N_p\}$,*and* $x_c^* = \frac{\delta}{1+\alpha_c N_p}$ *if* $\pi_c(\boldsymbol{x}^*) < \pi_1(\boldsymbol{x}^*) < \cdots < \pi_{N_p}(\boldsymbol{x}^*)$.

Case PE: There exists $r \in (1, N_p]$,*such that* $x_p^* = \frac{\delta}{1+\alpha_p}$, *for all* $p \in \{1,2, \ldots r-1\}$; $x_p^* = \frac{\delta}{1-\beta_p}$ *for all* $p \in \{r, r+1, \ldots N_p\}$; *and* $x_c^* = \frac{\delta}{1+\alpha_c(N_p-r+1)-\beta_c(r-1)}$ *if* $\pi_1(\boldsymbol{x}^*) < \cdots < \pi_{r-1}(\boldsymbol{x}^*) < \pi_c(\boldsymbol{x}^*) < \pi_r(\boldsymbol{x}^*) < \cdots < \pi_{N_p}(\boldsymbol{x}^*)$.

The equilibria in Corollary A-1—and Corollaries A-2 and A-3 to follow—can be worked out by self-reinforcement as follows. Start with an order of payoffs imposed by a candidate profile $\boldsymbol{x}^*$. If the best response (given in Proposition A-1) under this order of payoffs results in $\boldsymbol{x}^*$ itself, then $\boldsymbol{x}^*$ becomes an equilibrium. The formal proofs of Corollary A-1 (and of Corollaries A-2 and A-3 to follow) are omitted as they immediately follow from Proposition A-1.

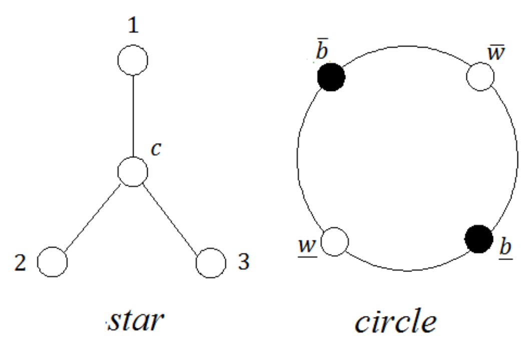


Figure A-1. The networks illustrated with notations

We next study the game of inequality aversion on circle networks. For the four-player circle used in the experiment, we can divide the nodes into two groups—each of which contains nodes that share a common neighborhood. Without loss of generality, the two groups are labeled *b* and *w* (respectively colored black [*b*] and white [*w*] in Figure A-1). We use an upper [lower] bar to represent the node with higher [lower] payoff in each group. Hence, by definition, we have $\pi_{\bar{w}}(\boldsymbol{x}) > \pi_{\underline{w}}(\boldsymbol{x})$ and $\pi_{\bar{b}}(\boldsymbol{x}) > \pi_{\underline{b}}(\boldsymbol{x})$ for given $\boldsymbol{x}$. Subsequently, there is the consideration of all possible scenarios in equilibrium: 1) *payoff separating* (PS), where $\pi_{\bar{w}}(\boldsymbol{x}) > \pi_{\underline{w}}(\boldsymbol{x}) > \pi_{\bar{b}}(\boldsymbol{x}) >$

$\pi_{\underline{b}}(\boldsymbol{x})$; 2) *payoff overlapping* (PO), where $\pi_{\bar{w}}(\boldsymbol{x}) > \pi_{\bar{b}}(\boldsymbol{x}) > \pi_{\underline{w}}(\boldsymbol{x}) > \pi_{\underline{b}}(\boldsymbol{x})$; and 3) *payoff embedding* (PE), where $\pi_{\bar{w}}(\boldsymbol{x}) > \pi_{\bar{b}}(\boldsymbol{x}) > \pi_{\underline{b}}(\boldsymbol{x}) > \pi_{\underline{w}}(\boldsymbol{x})$.[29]

COROLLARY A-2 (equilibrium of inequality aversion in circle). *There are three possible equilibria of inequality aversion in a four-player circle network.*

Case PS. $x_b^* = \frac{\delta}{1+2\alpha_b}$ *for all* $b \in \{\bar{b}, \underline{b}\}$, *and* $x_w^* = \frac{\delta}{1-2\beta_w}$ *for all* $w \in \{\bar{w}, \underline{w}\}$ *if* $\pi_{\bar{w}}(\boldsymbol{x}^*) > \pi_{\underline{w}}(\boldsymbol{x}^*) > \pi_{\bar{b}}(\boldsymbol{x}^*) > \pi_{\underline{b}}(\boldsymbol{x}^*)$.

Case PO. $x_{\bar{b}}^* = \frac{\delta}{1+\alpha_{\bar{b}} - \beta_{\bar{b}}}$, $x_{\underline{b}}^* = \frac{\delta}{1+2\alpha_{\underline{b}}}$. $x_{\bar{w}}^* = \frac{\delta}{1-2\beta_{\bar{w}}}$, $x_{\underline{w}}^* = \frac{\delta}{1+\alpha_{\underline{w}}-\beta_{\underline{w}}}$ *if* $\pi_{\bar{w}}(\boldsymbol{x}^*) > \pi_{\bar{b}}(\boldsymbol{x}^*) > \pi_{\underline{w}}(\boldsymbol{x}^*) > \pi_{\underline{b}}(\boldsymbol{x}^*)$.

Case PE. $x_b^* = \frac{\delta}{1+\alpha_b - \beta_b}$ *for all* $b \in \{\bar{b}, \underline{b}\}$, $x_{\bar{w}}^* = \frac{\delta}{1-2\beta_{\bar{w}}}$, $x_{\underline{w}}^* = \frac{\delta}{1+2\alpha_{\underline{w}}}$ *if* $\pi_{\bar{w}}(\boldsymbol{x}^*) > \pi_{\bar{b}}(\boldsymbol{x}^*) > \pi_{\underline{b}}(\boldsymbol{x}^*) > \pi_{\underline{w}}(\boldsymbol{x}^*)$.

One implication of Corollary A-2 is that, if a player's payoff is higher [lower] than both her neighbors in the circle network, then she unambiguously contributes more [less] than the self-interest amount, driven by a sense of sympathy [envy]. This is consistent with Remark 3. When the player's payoff falls between those of her neighbors, whether she contributes more than or less than the self-interest level will depend on the relative strength of her sympathy and envy. If, specifically, a player's feeling of envy from earning less than one of her neighbors dominates her compassion for the other neighbor who earns even less, then the focal player's contribution should be pulled below δ.

Finally, let us turn to the equilibrium of inequality aversion on the complete network. Denote by $r_i(\boldsymbol{x}) \equiv r_i$ the payoff rank of player i under profile $\boldsymbol{x}$, and by $\pi_{(r)}(\boldsymbol{x})$ the player payoff ranked at r in the network for given $\boldsymbol{x}$. The equilibrium is then characterized by the following corollary:

COROLLARY A-3 (equilibrium of inequality aversion in complete network). *There is a possible equilibrium of inequality aversion in a complete network of* N *players, such that* $x_i^* = \frac{\delta}{1+(N-r_i)\alpha_i-(r_i-1)\beta_i}$ *and* $\pi_{(1)}(\boldsymbol{x}^*) < \cdots < \pi_{(r_i)}(\boldsymbol{x}^*) < \pi_{(r_i+1)}(\boldsymbol{x}^*) < \cdots < \pi_{(N)}(\boldsymbol{x}^*), \forall i$.

---

[29] Note that payoff scenarios other than PS, PO, and PE can be generated by exchanging labels (i.e., b vs. w, upper bar vs. lower bar).

Corollary A-3 works out in the same flavor as Corollaries A-1 and A-2, except that everyone falls in the same neighborhood in the complete network (where one can introduce a common payoff ranking over all players). As implied by Corollary A-3, the player with the lowest [highest] earning in the complete network will contribute lower [higher] than δ, which is consistent with Remark 3. In estimating the inequality-aversion model (Corollaries A-1, A-2, and A-3), we use the actual profit observed in each period to determine the candidate equilibrium for that period. For example in a star network, if in some period the payoff of the core is greater than those of all in the periphery, then the subjects are supposed to play equilibrium PS1 in that period (where the contribution of any peripheral player p is estimated by $\frac{\delta}{1+\alpha_p}$ and that of the core player c estimated by $\frac{\delta}{1-\beta_c N_p}$).[30]

Appendix B. Proofs

PROOF OF THEOREM 1

Observe $\frac{\partial \pi_i(x)}{\partial x_i} = \delta - x_i$, $\frac{\partial \pi_j(x)}{\partial x_i} = \delta g_{ji}$ for $j \neq i$, and $\Pi_i(\boldsymbol{x}) = \left(1 - \sum_j \lambda_{ij} g_{ij}\right)\pi_i(\boldsymbol{x}) + \sum_j \lambda_{ij} g_{ij} \pi_j(\boldsymbol{x})$. Thus

$$\frac{\partial}{\partial x_i}\Pi_i(\boldsymbol{x}) = \left(1 - \sum_j \lambda_{ij} g_{ij}\right)\frac{\partial \pi_i(\boldsymbol{x})}{\partial x_i} + \sum_j \lambda_{ij} g_{ij} \frac{\partial \pi_j(\boldsymbol{x})}{\partial x_i}$$

$$= \left(1 - \sum_j \lambda_{ij} g_{ij}\right)(\delta - x_i) + \delta \sum_j \lambda_{ij} g_{ij} g_{ji}$$

Given $1 - \sum_j \lambda_{ij} g_{ij} > 0$, $\frac{\partial}{\partial x_i}\Pi_i(\boldsymbol{x})$ decreases in x_i. Therefore, x_i^* is identified by the first-order condition $\frac{\partial}{\partial x_i}\Pi_i(\boldsymbol{x}) = 0$.

$$x_i^* = \delta + \delta \frac{\sum_j \lambda_{ij} g_{ij} g_{ji}}{1 - \sum_j \lambda_{ij} g_{ij}}$$

[30] Note that the prerequisites of Corollaries A-1, A-2, and A-3 are stated in terms of estimated parameters (e.g., $\pi_1(\boldsymbol{x}^*) < \cdots < \pi_{N_p}(\boldsymbol{x}^*) < \pi_c(\boldsymbol{x}^*)$); therefore, it is an approximate to qualify these prerequisites using actual profit data. Such approximation becomes necessary given the rotation of partnership and the individualized parameter estimation.

When $g_{ij} = g_{ji} \in \{0,1\}$, we simplify the above expression as $x_i^* = \frac{\delta}{1-\sum_j \lambda_{ij} g_{ij}}$. If additionally we have $\lambda_{ij} \equiv \lambda_i$ for all j, then it follows that $x_i^* = \frac{\delta}{1-\lambda_i d_i}$. ■

PROOF OF THEOREM 2

By Proposition A-1, Player i's inequality-averse utility $\Pi_i(x_i; \boldsymbol{x}_{-i})$ is piecewise concave with respect to her own contribution, x_i. It is also continuous everywhere, including at each boundary point that equalizes the focal player's payoff with that of one of her neighbors. Given the action space is convex, the existence of pure strategy equilibrium follows from standard arguments (e.g., Theorem 1, Debreu, 1952). ■

PROOF OF PROPOSITION A-1

In the range-$(k+1)$, we have the following:

$$\Pi_i(\boldsymbol{x}) = \pi_i(\boldsymbol{x}) - \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} \left(\pi_j(\boldsymbol{x}) - \pi_i(\boldsymbol{x})\right) - \sum_{j=(1)_i}^{(k)_i} \beta_{ij} \left(\pi_i(\boldsymbol{x}) - \pi_j(\boldsymbol{x})\right)$$

Rearranging,

$$\Pi_i(\boldsymbol{x}) = \pi_i(\boldsymbol{x}) \left(1 + \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij}\right) - \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} \pi_j(\boldsymbol{x}) + \sum_{j=(1)_i}^{(k)_i} \beta_{ij} \pi_j(\boldsymbol{x})$$

Thus the derivative of $\Pi_i(\boldsymbol{x})$ within range-$(k+1)$ is:

$$\frac{\partial \Pi_i(\boldsymbol{x})}{\partial x_i} = \left(1 + \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij}\right)(\delta - x_i) - \delta \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} + \delta \sum_{j=(1)_i}^{(k)_i} \beta_{ij}$$

The concavity of $\Pi_i(\boldsymbol{x})$ within range-$(k+1)$ is ensured by the fact that $1 + \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij} > 0$, which can be implied from the assumption $1 - \sum_j \beta_{ij} g_{ij} > 0$.

Given this, player i's best response within range-$(k+1)$, hereafter denoted by $\phi_i^{k+1}(\boldsymbol{x}_{-i})$, is as follows.

$\phi_i^{k+1}(\boldsymbol{x}_{-i}) =$

$$\begin{cases} \delta + \delta \frac{\sum_{j=(1)_i}^{(k)_i} \beta_{ij} - \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij}}{1+\sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij}}, & if \ \delta + \delta \frac{\sum_{j=(1)_i}^{(k)_i} \beta_{ij} - \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij}}{1+\sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij}} \in \left(X_{(k+1)_i}(\boldsymbol{x}_{-i}), X_{(k)_i}(\boldsymbol{x}_{-i})\right) \\ X_{(k)_i}(\boldsymbol{x}_{-i}), & if \ \delta + \delta \frac{\sum_{j=(1)_i}^{(k)_i} \beta_{ij} - \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij}}{1+\sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij}} > X_{(k)_i}(\boldsymbol{x}_{-i}) \\ X_{(k+1)_i}(\boldsymbol{x}_{-i}), & if \ \delta + \delta \frac{\sum_{j=(1)_i}^{(k)_i} \beta_{ij} - \sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij}}{1+\sum_{j=(k+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k)_i} \beta_{ij}} < X_{(k+1)_i}(\boldsymbol{x}_{-i}) \end{cases}.$$

When player i's contribution x_i increases, her payoff rank within neighborhood (k) is lowered per Lemma A-1. Thus, $\delta + \delta \frac{\sum_{j=(1)_i}^{(k-1)_i} \beta_{ij} - \sum_{j=(k)_i}^{(d_i)_i} \alpha_{ij}}{1+\sum_{j=(k)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(k-1)_i} \beta_{ij}}$ decreases.

Notice that $\Pi_i(\boldsymbol{x})$ is continuous at every $X_{(k)_i}(\boldsymbol{x}_{-i})$. Thus, $\Pi_i(\boldsymbol{x})$ is globally concave with respect to x_i. The resulting unique maximizing solution $\phi_i(\boldsymbol{x}_{-i})$ should be obtained at either an interior or a boundary point of some range-r. In the former case (interior), the optimal solution is the point in range-r where the first derivative of $\Pi_i(\boldsymbol{x})$ equals 0:

$$\phi_i(\boldsymbol{x}_{-i}) = \delta + \delta \frac{\sum_{j=(1)_i}^{(r-1)_i} \beta_{ij} - \sum_{j=(r)_i}^{(d_i)_i} \alpha_{ij}}{1 + \sum_{j=(r)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(r-1)_i} \beta_{ij}} = \frac{\delta}{1 + \sum_{j=(r)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(r-1)_i} \beta_{ij}},$$

if $\frac{\delta}{1+\sum_{j=(r)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(r-1)_i} \beta_{ij}} \in \left(X_{(r)_i}(\boldsymbol{x}_{-i}), X_{(r-1)_i}(\boldsymbol{x}_{-i})\right)$.

In the latter case (boundary), the first derivative of $\Pi_i(\boldsymbol{x})$ turns from positive to negative at the optimum $X_{(r)_i}(\boldsymbol{x}_{-i})$. That is,

$$\phi_i(\boldsymbol{x}_{-i}) = X_{(r)_i}(\boldsymbol{x}_{-i}),$$

if $X_{(r)_i}(\boldsymbol{x}_{-i}) \in \left(\frac{\delta}{1+\sum_{j=(r)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(r-1)_i} \beta_{ij}}, \frac{\delta}{1+\sum_{j=(r+1)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{(r)_i} \beta_{ij}} \right)$.

Notice the above cases are valid as $\frac{\delta}{1+\sum_{j=\left(\underline{d}_i+1\right)_i}^{(d_i)_i} \alpha_{ij} - \sum_{j=(1)_i}^{\left(\underline{d}_i\right)_i} \beta_{ij}} \geq 0$, which guarantees the optimal contribution, $\phi_i(\boldsymbol{x}_{-i})$, is nonnegative. Finally, a contribution profile $\boldsymbol{x}$ is an equilibrium if and only if the player contributions are best responses to each other—$\boldsymbol{x}_i = \phi_i(\boldsymbol{x}_{-i})$ for each i, or in a vector form, $\boldsymbol{x} = \phi(\boldsymbol{x})$. That concludes the proof. ■

Appendix C. Experiment

C.1. Experimental protocol

In connection with Sections 3 and 5 in the main text, we further illustrate below the structuring of the experiment and the matching protocol in Figures C-1 and C-2 respectively.

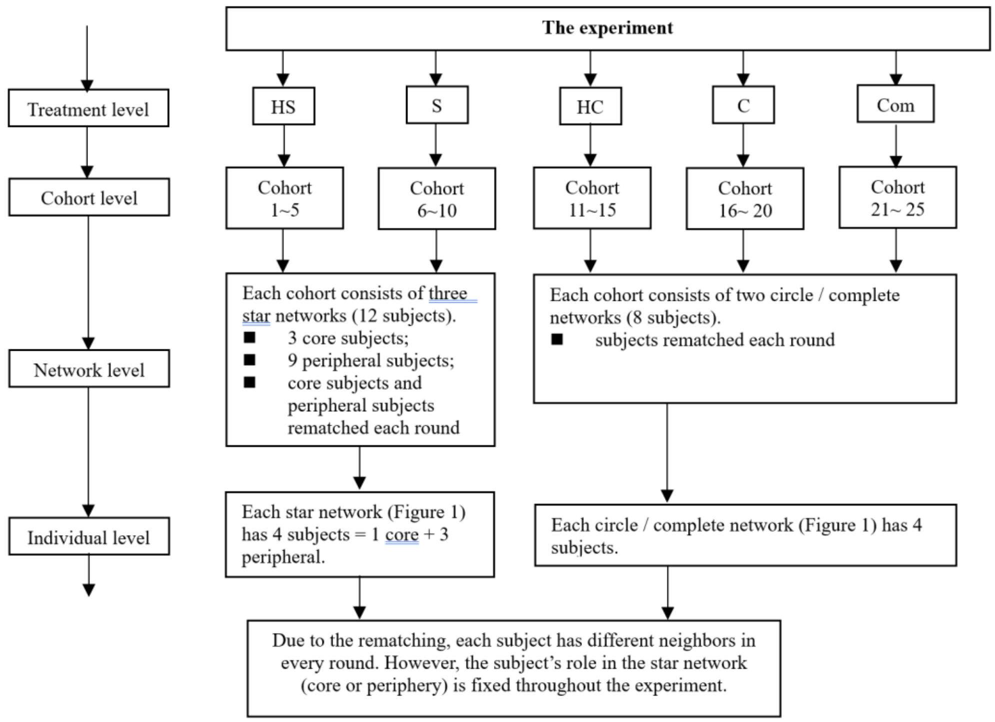
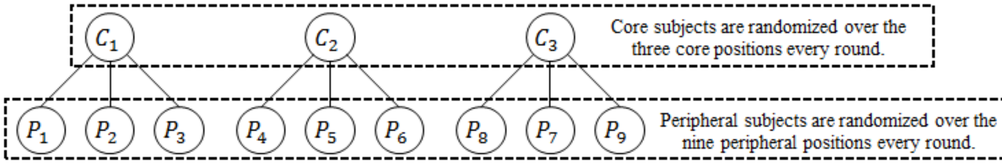
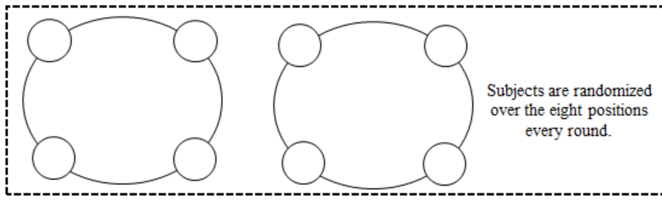


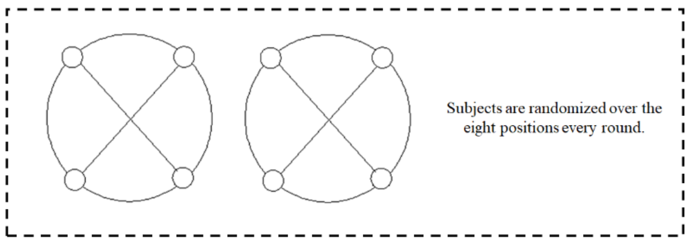
Figure C-1. The organization of the experiment



(a) Star cohort



(b) Circle cohort



(c) Complete network cohort

Figure C-2. The matching protocol within a cohort

C.2. Global network effect

Note that the network effects observed in our main experiment in Section 4 are *local*, in the sense that we contrast subject behaviors in the same network with different numbers of neighbors (e.g. core *vs.* periphery). Now, we will examine *global* network effects by comparing the behavior across networks while controlling for one's number of neighbors. Specifically, we will compare behavior in *Com* with that of the core players in treatment *S*. Thus, we control for one's number of neighbors (3) while varying the outer network to examine the play.

Figure C-3 shows that the subjects contribute less in *Com* than those at the core position in *S* (the latter denoted as *S core* in the figure) with one-sided *p*-value = 0.009 by a two-sample *t*-test clustered by cohorts. We view this as the effect of reference structure. Although having the same degree, the core of a star and a member of a complete network face drastically different networks outside their neighborhoods. The hierarchical configuration of the star network exposes its core to a "social responsibility" to raise its contribution, whereas such exposure does not take place in a complete network owing to the structural balance.

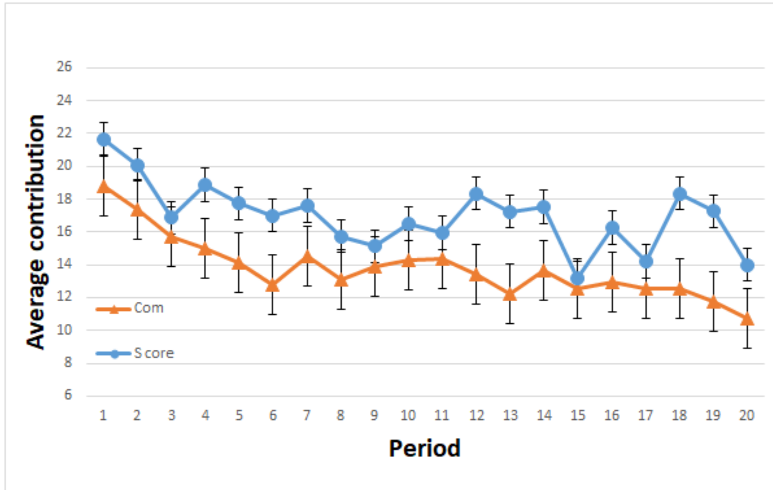


Figure C-3. Global network effect on contribution

In this section, we find that, while having the same number of connections, subjects in the complete network contribute significantly less than those at the core position in the star network. Thus, we showcase how the global network structure could affect individual contributions by manipulating the saliency of one's network position.

C.3. Experimental instruction

We present the experimental instruction for treatment *HS* in this appendix. The instruction can be straightforwardly adapted to other treatments with the change of payoff or network structures. The original instruction delivered to the subjects was written in Chinese, and we present here the version translated to English.

****************************** Instruction: *HS* **********************************

Welcome and thank you for participating in this experiment. In this experiment you will earn money. From now on until the end of the experiment, please do not communicate with other participants. If you have any question, please raise your hand. An experimenter will come to your place and answer your question privately. At the end of the experiment, the balance of your account will be converted into Chinese Yuan according to the conversion rate: 1 unit of experimental profit = 0.001 Chinese Yuan, and paid to you right away.

The experiment lasts for 20 rounds. In each round, participants will be organized in a network shown in Figure 1 below. In this network, people connected to you are your neighbors. Every round your neighbors will be different people. If you ever take a core position (C) in any round, you will remain at the core position throughout the experiment. Similarly, if you ever take a peripheral position (P1, P2, or P3) in any round, you will remain at the peripheral position throughout the experiment. Your position in the network will be marked blue, shown on the computer screen.

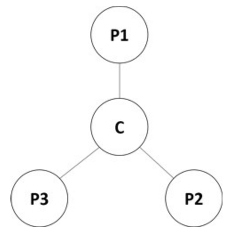


Figure 1. The network

In each round of the experiment, you will determine the amount of a "contribution". On the one hand, every unit of your contribution will return you a benefit of 30 units. On the other hand, making the contribution incurs you a cost that equals 0.5×(the amount of your contribution)^2. Furthermore, your neighbors also benefit from your contribution – Every unit of your contribution will deliver a benefit of 30 units to each of your neighbors. Likewise, every unit contributed by each of your neighbors will also bring you 30 units of benefit. Every round, you can contribute up to 75 units.

Examples.

Suppose you are player P1 in the network above. Your neighbor is therefore player C. If player C contributes 30 and you (player P1) contribute 20, your profit will be:

$$30 \times 20 + 30 \times 30 - \frac{1}{2} \times 20^2 = 1300$$

Similarly, if you are player C in the network, your neighbors are player P1, P2, P3. If player P1, P2, P3 contribute 20, 30, 40 respectively, and you (player C) contribute 30, your profit will be:

$$30 \times 30 + 30 \times (20 + 30 + 40) - \frac{1}{2} \times 30^2 = 3150$$

After each round, you will get to see your profit, your contribution, and your neighbor(s)'s profit this round.

Before the game starts, you are required to complete a quiz, which covers the important knowledge about the game. The quiz will take place on your computer. You will not start to play the game unless you answer all the questions in the quiz correctly.

****

C.4. Software interface

The experimental software was programmed with zTree (Fischbacher, 2007). Subjects began with a quiz testing their understanding of the game, with no earning accumulated to the game. The quiz screens are shown in Figure C-4. The actual decision-making interfaces are found in Figure C-5. The original text on the experimental interface was in Chinese, and we present here the version translated to English. For conciseness, we only include the interfaces for treatment *C*. The other cases not covered in the screenshots were presented to the subjects in an analogous manner.

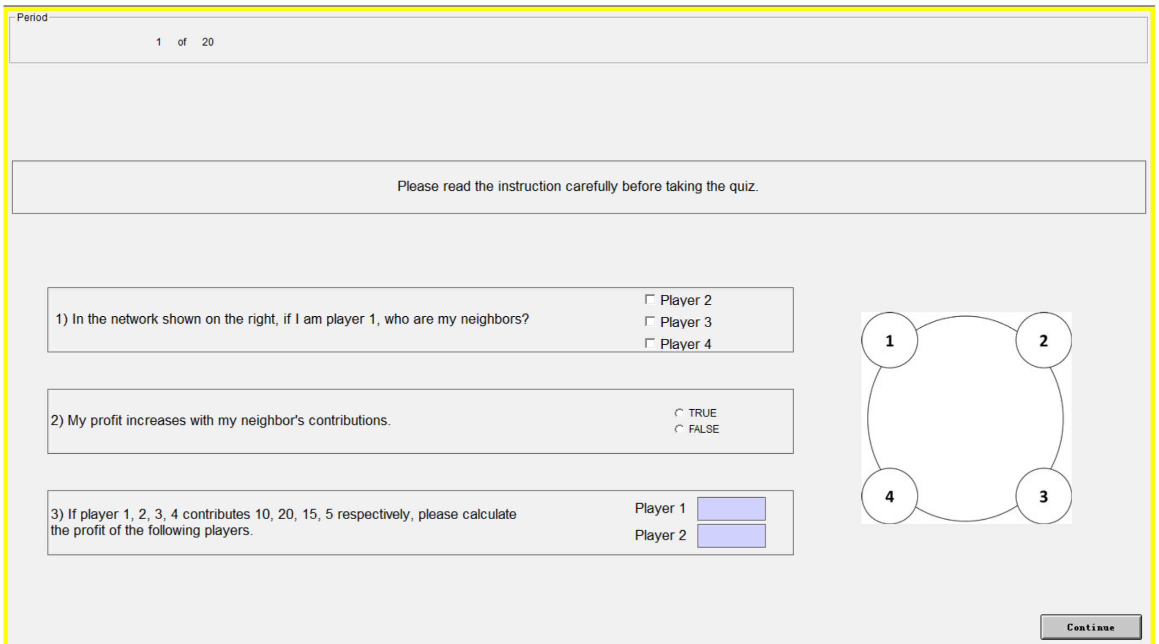


Figure C-4. Quiz

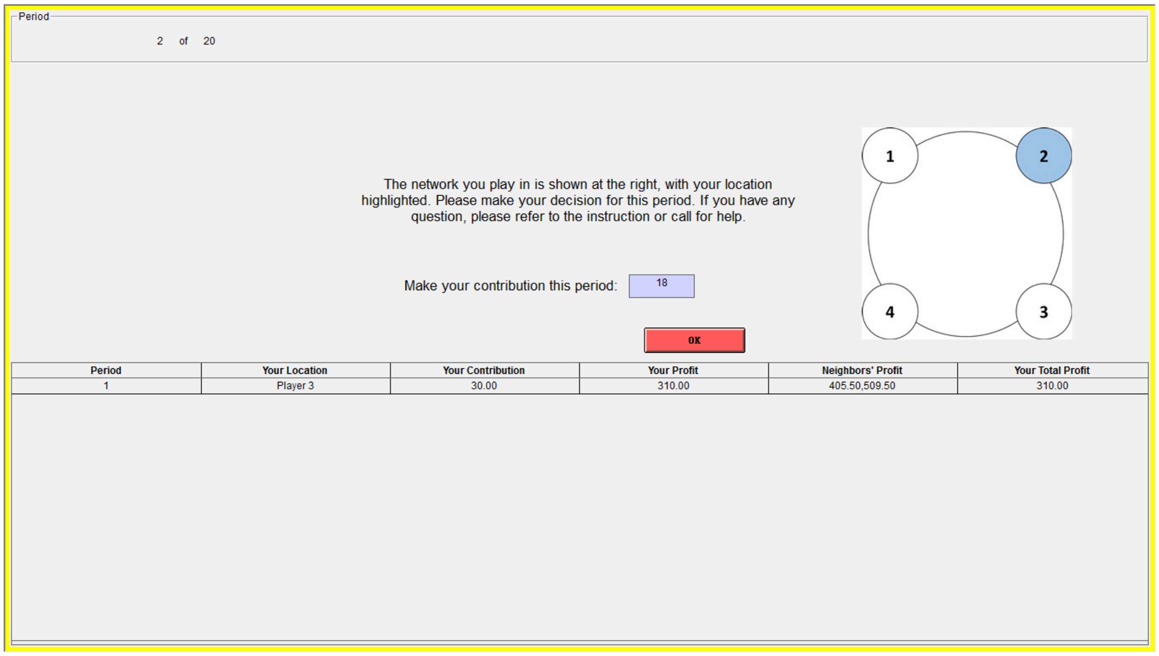


Figure C-5. Decision making

References

Akerlof, G., & Kranton, R. (2000). Economics and Identity. *Quarterly Journal of Economics*, *115*(03), 715–753.

Akerlof, G., & Kranton, R. (2005). Identity and the Economics of Organizations. *Journal of Economic Perspectives*, *19*(1), 9–32.

Bala, V., & Goyal, S. (2000). A Noncooperative Model of Network Formation. *Econometrica*, *68*(5), 1181–1229.

Ballester, C., Calvo-Armengol, A., & Zenou, Y. (2006). Who’s who in networks. wanted: the key player. *Econometrica*, *74*(5), 1403–1417.

Berninghaus, S., G¨uth, W., & Schosser, S. (2013). Reciprocity in Locating Contributions: Experiments on the Neighborhood Public Good Game. *Games*, *4*(02), 144–162.

Berninghaus, S. K., Ehrhart, K.-M., & Keser, C. (2002). Conventions and local interaction structures: experimental evidence. *Games and Economic Behavior*, 177–205.

Bloch, F., & Quérou, N. (2013). Pricing in social networks. *Games and Economic Behavior*, *80*, 243–261.

Bolton, G. E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *The American Economic Review*, 166–193.

Boosey, L. A. (2017). Conditional cooperation in network public goods experiments. *Journal of Behavioral and Experimental Economics*, *69*, 108–116.

Boosey, L., & Issac, M. R. (2016). Asymmetric network monitoring and punishment in public goods experiments. *Journal of Economic Behavior & Organization*, *132*, 26–41.

Bourlès, R., Bramoullé, Y., & Perez-Richet, E. (2017). Altruism in Networks. *Econometrica*, *85*(2), 675–689.

Bramoullé, Y., & Kranton, R. (2007). Public goods in networks. *Journal of Economic Theory*, *135*(1), 478–494.

Bramoullé, Y., Kranton, R., & D'Amours, M. (2014). Strategic Interaction and Networks. *The American Economic Review*, *104*(3), 898–930.
Candogan, O., Bimpikis, K., & Ozdaglar, A. (2012). Optimal pricing in networks with externalities. *Operations Research*, *60*(04), 883–905.
Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at Work: The Effect of Peer Salaries on Job Satisfaction. *American Economic Review*, *102*(6), 2981–3003.
Cason, T., & Gangadharan, L. (2015). Promoting Cooperation in Nonlinear Social Dilemmas through Peer Punishment. *Experimental Economics*, *18*, 66–88.
Cassar, A. (2007). Coordination and cooperation in local, random and small world networks: experimental evidence. *Games and Economic Behavior*, *58*(02), 209–230.
Charness, G., Feri, F., Melendez, M., & Sutter, M. (2014). Experimental games on networks: underpinnings of behavior and equilibrium selection. *Econometrica*, 1615–1670.
Charness, G., & Rabin, M. (2002). Understanding Social Preferences with Simple Tests. *Quarterly Journal of Economics*, *117*, 817–869.
Chaudhuri, A. (2011). Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics*, *14*(1), 47–83.
Chen, Y., Harper, M. F., Konstan, J., & Li, S. X. (2010). Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens. *The American Economic Review*, *100*(4), 1358–1398.
Chen, Y., & Li, S. X. (2009). Group Identity and Social Preferences. *The American Economic Review*, *99*(01), 431–457.
Cho, M. (2010). Endogenous formation of networks for local public goods. *International Journal of Game Theory*, *39*(04), 529–562.
Corbae, D., & Duffy, J. (2008). Experiments with network formation. *Games and Economic Behavior*, *64*(01), 81–120.
Debreu, G. (1952). A social equilibrium existence theorem. *Proceedings of the National Academy of Sciences*, *38*(10), 886–893.
Elliott, M., & Golub, B. (2013). A network approach to public goods. *Proceedings of the Fourteenth ACM Conference on Electronic Commerce* , 377–378. New York.
Engelmann, D., & Strobel, M. (2004). Inequality Aversion, Efficiency, and Maximin Preferences in Simple Distribution Experiments. *The American Economic Review*, *94*(04), 857–869.
Engelmann, D., & Strobel, M. (2006). Inequality aversion, efficiency, and maximum preferences in simple distribution experiments:Reply. *American Economic Review*, *96*(05), 1918–1923.
Fainmesser, I. P., & Galeotti, A. (2016). Pricing network effects. *The Review of Economic Studies*, *83*(1), 165–198.
Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(03), 817–868.
Ferrer-i-Carbonell, A. (2005). Income and well-being: an empirical analysis of the comparison income effect. *Journal of Public Economics*, *89*(5–6), 997–1019.
Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 171–178.
Fischbacher, U., & Gächter, S. (2010). Social Preferences, Beliefs, and the Dynamics of Free Riding in Public Goods Experiments. *The American Economic Review*, *100*(01), 541–556.
Frank, R. H. (1985). The Demand for Unobservable and Other Nonpositional Goods. *The American Economic Review*, *75*(1), 101–116.
Galeotti, A., & Goyal, S. (2010). The Law of the Few. *The American Economic Review*, *100*(04), 1468–1492.
Galeotti, A., Goyal, S., Jackson, M. O., Vega-Redondo, F., & Yariv, L. (2010). Network Games. *The Review of Economic Studies*, *77*(01), 218–244.

Ghiglino, C., & Goyal, S. (2010). KEEPING UP WITH THE NEIGHBORS SOCIAL INTERACTION IN A MARKET ECONOMY. *Journal of the European Economic Association*, *8*(1), 90–119.

Goyal, S. (2009). *Connections: An Introduction to the Economics of Networks.* Princeton, NJ: Princeton University Press.

Ho, T. H., & Su, X. (2009). Peer-induced fairness in games. *The American Economic Review*, *99*(05), 2022–2049.

Hojman, D., & Szeidl, A. (2008). Core and periphery in networks. *Journal of Economic Theory*, *131*(01), 295–309.

Hopkins, E., & Kornienko, T. (2004). Running to Keep in the Same Place Consumer Choice as a Game of Status. *The American Economic Review*, *94*(4), 1085–1107.

Immorlica, N., Kranton, R., Manea, M., & Stoddard, G. (2017). Social Status in Networks. *American Economic Journal: Microeconomics*, *9*(1), 1–30.

Isaac, R. M., & Walker, J. (1998). Nash as an Organizing Principle in the Voluntary Provision of Public Goods: Experimental Evidence. *Experimental Economics*, *1*, 191–206.

Jackson, M. O. (2008). *Social and Economic Networks*. Princeton, NJ: Princeton University Press.

Jackson, M. O., Rogers, B. W., & Zenou, Y. (2017). The Economic Consequences of Social-Network Structure. *Journal of Economic Literature*, *55*(1), 49–95.

Jackson, M. O., & Wolinsky, A. (1996). A Strategic Model of Social and Economic Networks. *Journal of Economic Theory*, *71*, 44–74.

Jackson, M. O., & Zenou, Y. (2015). Games on networks. In P. Young & S. Zamir (Eds.), *Handbook of Game Theory* (Vol. 4, pp. 34–61). Amsterdam: Elsevier Publisher.

Judd, S., Kearns, M., & Vorobeychik, Y. (2011). Behavioral Dynamics and Influence in Networked Coloring and Consensus. *Proceedings of the National Academy of Sciences*, *108*(16), 6689–6689.

Kahneman, D., Knetsch, J., & Thaler, R. (1986a). Fairness and the Assumptions of Economics. *Journal of Business*, *59*(4), 285–300.

Kahneman, D., Knetsch, J., & Thaler, R. (1986b). Fairness as a Constraint on Profit Seeking: Entitlements in the Market. *The American Economic Review*, *76*(4), 728–741.

Keser, C. (1996). Voluntary contributions to a public good when partial contribution is a dominant strategy. *Economics Letters*, *50*(03), 359–366.

Keser, C., Ehrhart, K.-M., & Berninghaus, S. K. (1998). Coordination and local interaction: experimental evidence. *Economics Letters*, *58*, 269–275.

Kosfeld, M. (2004). Economic networks in the laboratory: A survey. *Review of Network Economics*, *3*(1), 20–41.

Kranton, R. E. (2016). Identity Economics 2016: Where Do Social Distinctions and Norms Come From? *American Economic Review*, *106*(5), 405–409.

Ledyard, J., Kagel, J., & Roth, A. E. (1995). Public Goods: A Survey of Experimental Research. In *The Handbook of Experimental Economics* (pp. 111–194). Princeton, NJ: Princeton University Press.

Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social Utility and Decision Making in Interpersonal Contexts. *Journal of Personality and Social Psychology*, *57*(3), 426–441.

Luttmer, E. (2005). Neighbors as Negatives Relative Earnings and Well-Being. *The Quarterly Journal of Economics*, *120*(3), 963–1002.

Masclet, D., & Villeval, M.-C. (2008). Punishment, inequality, and welfare: a public good experiment. *Social Choice and Welfare*, *31*(3), 475–502.

Neilson, W., & Wichmann, B. (2014). Social networks and non-market valuations. *Journal of Environmental Economics and Management*, *67*(02), 155–170.

Neumark, D., & Postlewaite, A. (1998). Relative income concerns and the rise in married women's employment. *Journal of Public Economics*, *70*(1), 157–183.

Rege, M., & Telle, K. (2004). The impact of social approval and framing on cooperation in public good situations. *Journal of Public Economics*, *88*(7–8), 1625–1644.

Roels, G., & Su, X. (2014). Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points. *Management Science*, *60*(3), 606–627.

Rosenkranz, S., & Weitzel, U. (2012). Network structures and strategic investments: an experimental analysis. *Games and Economic Behavior*, *75*(2), 898–920.

Sundararajan, A. (2008). Local network effects and complex network structure. *The B.E. Journal of Theoretical Economics*, *7*(01), pp.000010220219351704l319.

van Leeuwen, B., Offerman, T., & Schram, A. (2020). Competition for Status Creates Superstars: an Experiment on Public Good Provision and Network Formation. *Journal of the European Economic Association*, *18*(2), 666–707.

Weimann, J., Brosig-Koch, J., Hennig-Schmidt, H., Keser, C., & Stahr, C. (2012). *"Public goods contribution in large groups"FEMM Working Papers 120009, Otto-von-Guericke University Magdeburg, Faculty of Economics and Management.*

Zhang, Y., & Chen, Y.-J. (2020). Optimal nonlinear pricing in social networks under asymmetric network information. *Operations Research*, *68*(03), 818–833.