UNIVERSITY OF CAMBRIDGE
Faculty of Economics

JANEWAY INSTITUTE CAMBRIDGE

CAMBRIDGE WORKING PAPERS IN ECONOMICS
JANEWAY INSTITUTE WORKING PAPERS

Experimental Evidence on Group Size Effects in Network Formation Games

Syngjoo Choi
Seoul National University

Sanjeev Goyal
University of Cambridge and NYUAD

Fulin Guo
University of Cambridge

Frédéric Moisan
Emlyon Business School, GATE UMR 5824

Abstract

This paper presents experimental evidence on games where individuals can unilaterally decide on their links with each other. Linking decisions give rise to directed graphs. We consider two classes of situations: one, benefits flow along the direction of the network paths (one-way flow), and two, when the benefits flow on network paths without regard to the direction of links (two-way flow). Our experiments reveal that in the one-way flow model subjects create sparse networks whose distance grows and efficiency falls as group size grows; by contrast, in the two-way flow model subjects create sparse and small world networks whose efficiency remains high in both small and large groups. We show that a bounded rational model that combines myopic best response with targeting a most connected individual provides a coherent account of our experimental data.

Reference Details

2417	Cambridge Working Papers in Economics
2412	Janeway Institute Working Paper Series
Published	26 March 2024
Websites	www.econ.cam.ac.uk/cwpe www.janeway.econ.cam.ac.uk/working-papers

Experimental Evidence on Group Size Effects in Network Formation Games

Syngjoo Choi* Sanjeev Goyal† Fulin Guo‡ Frédéric Moisan§

March 26, 2024

Abstract

This paper presents experimental evidence on games where individuals can unilaterally decide on their links with each other. Linking decisions give rise to directed graphs. We consider two classes of situations: one, benefits flow along the direction of the network paths (one-way flow), and two, when the benefits flow on network paths without regard to the direction of links (two-way flow). Our experiments reveal that in the one-way flow model subjects create sparse networks whose distance grows and efficiency falls as group size grows; by contrast, in the two-way flow model subjects create sparse and small world networks whose efficiency remains high in both small and large groups. We show that a bounded rational model that combines myopic best response with targeting a most connected individual provides a coherent account of our experimental data.

*Seoul National University. Email: syngjooc@snu.ac.kr

†University of Cambridge and New York University Abu Dhabi. Email: sg472@cam.ac.uk

‡University of Cambridge. Email: fg400@cam.ac.uk

§Emlyon Business School, GATE UMR 5824. Email: moisan@em-lyon.com

The paper was supported by the E.U.'s Future and Emerging Technology-Open Project IBSEN (contract no. 662725), the Cambridge-INET Institute and the Keynes Fund (University of Cambridge), Creative-Pioneering Researchers Program (Seoul National University), from the Center of Behavioral Institutional Design and Tamkeen under the NYU Abu Dhabi Research Institute Award CG005 and was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042, University of Lyon). We thank Stefano Caria, Marghirta Comola, Amit Dekel, Marcel Fafchamps, Francesco Feri, Moritz Janas, Miguel-Angel Melendez and Michael Xu for helpful comments.

1 Introduction

A large body of research studies the formation of networks and their welfare properties; leading examples include financial networks (Farboodi (2023)), development (Banerjee, Breza, Chandrasekhar, Duflo, Jackson, and Kinnan (2021)), production networks (Acemoglu, Carvalho, Ozdaglar, and A.Tahbaz-Salehi (2012), Acemoglu and Azar (2020)), infrastructure networks (Hendricks et al. (1995), Fajgelbaum and Schaal (2020), and information networks (Galeotti and Goyal (2010)). Underlying this large literature is the simple idea that individual entities compute the costs and benefits of linking and these trade-offs generate the networks that are observed in practice. As private actors and governments increasingly use network ideas it is important to understand the scope of this economic approach to network formation. Laboratory experiments offer the ideal context to test the essential mechanisms that shape linking. In this paper, we present experimental evidence on two models of network formation. A distinctive feature of our work is that the experiments cover both small and large groups – 10 and 50 subjects – and linking activity takes place in continuous time.[1]

We consider a game of pure linking taken from Bala and Goyal (2000).[2] In this game individuals can unilaterally decide to form links with others to access benefits – this could be information, for instance. A key element of the setting is the recursive flow of benefits – a link from A to B also gives A access to benefits that B has accessed through her links. Thus the returns to a link depend on the links that others create. This externality is central to the economic approach of network formation. There are two variants of the model: the *one-way* flow and the *two-way* flow of benefits.

In the one-way flow model, benefits flow along the directed paths of the network. The static theory is permissive – the empty network as well as a variety of connected networks arise in equilibrium. Our simulations of the best response dynamics suggest that so long as we start with a reasonable number of initial links, the limit network is the cycle network. The cycle maximizes aggregate payoffs across the relevant range of

[1]Our paper builds on the work of Berninghaus, Ehrhart, and Ott (2006), Friedman and Oprea (2012), Calford and Oprea (2017), and Goyal, Rosenkranz, Weitzel, and Buskens (2017) that draws attention to the role of continuous time experiments in testing theoretical models.

[2]There is a large literature on linking games, for early contributions, see e.g., Jackson and Wolinsky (1996), Feri (2007), Hojman and Szeidl (2008). Goyal (2023) provides a recent survey.

parameters. We conduct an experiment in which the focus is on group size – varying from 10 to 50. Given this prediction of a cycle network, the testable hypothesis is that, as we vary group size, efficiency remains high, distances grow, and linking and payoff inequality remain low. Our experimental findings for the one-way flow model are as follows: one, subjects create a sparse and connected network close to the cycle that attains over 90% efficiency for $N = 10$ and a sparse but unconnected network that attains less than 50% efficiency for $N = 50$. Two, distances and link inequality and payoff inequality all grow with group size.[3]

In the two-way flow model, links are created unilaterally but benefits flow both ways. The static theory is again permissive – the empty network as well as a variety of sparse connected networks arise in equilibrium. Our simulations of the best response dynamics suggest that so long as we start with a reasonable number of initial links, the limit network is the star network. The star maximizes aggregate payoffs across the relevant range of parameters. We conduct an experiment in which the focus is again on group size – varying from 10 to 50. Given the theoretical prediction of a star network, the testable hypothesis is that, as we vary group size, efficiency remains high, distances change very little, linking inequality grows, and payoff inequality is large but remains stable. Our experimental findings for the two-way flow model are as follows: one, subjects create connected and sparse networks that attain high efficiency close to 90% in the $N = 10$ treatment and over 80% in the $N = 50$ treatment. Two, distances remain low but increase in group size, and payoff inequality is large but remains stable, and linking inequality remains high in both group sizes.

Thus the move from small group to large group has a powerful effect on networks and welfare in the one-way flow model but relatively modest impact in the two-way flow model. We then turn to an examination of the reasons for this difference in impact of group size. We show that the sources of inefficiency observed in the experiment, especially in large groups, differs across the models: breakdown of connectedness in the one-way flow model, and overlinking and long distances in the two-way flow model.

We turn to an examination of the behavioral rules that give rise to these different

[3] A network is said to be connected if there is a directed path between any two nodes in it. Correspondingly, a network is unconnected if there is at least one pair of nodes which do not have a directed path between them.

group size effects in the two models. We do this by proposing and estimating a model that combines myopic best response with a behavioral rule of targeting a highest-degree individual (in this part we build on the work of Caria and Fafchamps (2020)). This analysis reveals that in all treatments except the one-way flow with N=50 case, deviations from theoretical predictions can be largely attributed to best responses combined with uniformly random noise. However, in the one-way flow model, individuals in large groups tend to also form links with the person with the most number of links. Linking with such a person is generally not optimal as what matters is access to individuals and this is a function of both direct and indirect connections. In the final part of the paper, we show through computer simulations that this simple behavioral model is appropriate in this environment as it can generate behavioral patterns that closely resembles the outcomes from the experiment.

The focus of the paper is on the group size effects within each of the two models: with this in mind, the interfaces and parameters used in the experiments are accordingly adjusted between the two models so as to make experimental tests clear. In particular, in line with the literature, we chose no decay in the flow of benefits in the one-way flow model (see e.g., Caria and Fafchamps (2020)) and decay in the two-way flow model (see Goeree et al. (2009)). This means that our experiments do not allow us to study the effects of flow models on network formation for a fixed group size.[4]

Our paper is a contribution to the experimental study of networks (for surveys of the literature refer to Choi, Gallo, and Kariv (2016) and Breza (2016)). A number of papers have experimentally studied the linking game proposed in Bala and Goyal (2000): see e.g., Berninghaus, Ehrhart, and Ott (2006), Callander and Plott (2005), Falk and Kosfeld (2012), Goeree, Riedl, and Ule (2009) and Caria and Fafchamps (2020). These papers consider small groups involving four and six subjects choosing links in discrete time; by contrast, we consider groups of 10 and 50 choosing links in continuous time.

Callander and Plott (2005), Falk and Kosfeld (2012) and Caria and Fafchamps (2020) study the one-way flow model and consider a parametric setting in which the cycle is an equilibrium and efficient network. Falk and Kosfeld (2012) show that

[4]Experiments on the effect of flow models on network formation would require one to consider an extended framework varying the flow patterns of benefits in a comprehensive manner (e.g., Olaizola and Valenciano (2014)). This is an interesting avenue for future research.

subjects in groups of 4 form the cycle network half of the time. Caria and Fafchamps (2020) show that subjects play myopic best response roughly half of the time and that they realize less than 70% of the efficiency gains. By contrast, we find that in small groups of $N = 10$ subjects create networks that are close to the cycle and attain over 90% of efficiency gains. This difference suggests that our experimental design with asynchronous linking in continuous time facilitates convergence to equilibrium (and efficient) networks.

Falk and Kosfeld (2012) and Goeree, Riedl, and Ule (2009) study the two-way flow model and consider a parametric setting in which the star is an equilibrium and efficient network. They find that subjects fail to form this network. By contrast, we find that in small groups, subjects form networks that are close to the star and attain close to 90% of possible efficiency gains. Moreover, to the best of our knowledge, our paper offers the first experimental evidence for network formation in large groups in both the one-way and the two-way flow model.

Our paper also contributes to the study of decision rules in network environments. This is a very active field of study: for learning, see Chandrasekhar, Larreguy, and Xandri (2020), Grimm and Mengel (2020), and Choi, Goyal, Moisan, and To (2023b); for network formation see Caria and Fafchamps (2020), Falk and Kosfeld (2012), Goeree, Riedl, and Ule (2009), Choi, Goyal, and Moisan (2023a), Horvath (2023), and van Leeuwen, Offerman, and Schram (2020); for games on networks see Charness, Feri, Meléndez-Jiménez, and Sutter (2014), Currarini, Feri, Hartig, and Meléndez-Jiménez (2023), Gallo and Yan (2023), and Choi, Goyal, Guo, and Moisan (2024)). Caria and Fafchamps (2020) show that in the one-way flow model environment with six individuals, some subjects depart from myopic best response and target a highest-degree individual. We advance their finding by clarifying its uses across group sizes and across one-way and two-way flow models.

The rest of the paper is organized as follows. In Section 2, we describe the one-way model and the experimental design and present the main experimental findings. In Section 3, we describe the two-way flow model and the experimental design and present the main experimental findings. In Section 4, we explain the data with a behavioral model. In Section 5, we carry out two sets of exercises: one, we show that our group size effects in one-way and two-way flow model carry over to a setting with

100 subjects, and two, we consider the dynamics of linking over the entire period of the experiment. In the main body of the paper we focus on the last 30 seconds of the experiment; here we ask if there are patterns of behaviour we have overlooked at the earlier stages of the experiment and if they differ across the one-way and the two-way flow model. Supplementary materials are presented in the Appendix.

2 One-way flow model

This section presents and tests the one-way flow model of networks taken from Bala and Goyal (2000): links are unilateral and costly, benefits accrue from individuals accessed through the directed paths to other individuals in the network. We start with the theoretical model. This is followed by a discussion of the experimental design and the equilibrium predictions and a statement of the hypothesis. We then present the experimental findings on network structure and on efficiency.

2.1 Theory

Let $\mathbf{N} = \{1, 2, \ldots, N\}$ with $N \geq 3$. Each player $i \in \mathbf{N}$ simultaneously and independently chooses a set of links g_i with others, $g_i = (g_{i1}, \ldots, g_{ii-1}, g_{ii+1}, \ldots, g_{iN})$, and $g_{ij} \in \{0, 1\}$ for any $j \in \mathbf{N}\backslash\{i\}$. Thus links are unilateral in this game. The set of strategies of player i is $s_i = G_i$, where $G_i = \{0, 1\}^{N-1}$. A strategy profile $s = (s_1, s_2, \ldots, s_N)$ specifies the links made by every player and induces a directed graph, g. Let $\eta_i(g) = |\{j \in \mathbf{N}\backslash\{i\} : g_{ij} = 1\}|$ be the number of links i has formed in g.

An individual A benefits from an individual B if and only if there exists a directed path from A to B in the network. For any pair of players i and j in g, the geodesic distance, denoted by $d(i, j; g)$, is the length of the shortest path between i and j in the directed network g. If no such path exists, the distance is set to infinity.

Given a strategy profile s, the payoffs of player i are:

$$\Pi_i(s) = V + \sum_{j \in \mathbf{N}\backslash\{i\}} \delta^{d(i,j;g)} V - \eta_i(g) k \tag{1}$$

where V is the value of benefit per connection, $\delta \in [0, 1]$ is the decay factor associated with indirect access to benefits, and k is the cost of linking with another player.

Define a network as efficient if it maximizes the sum of individual payoffs, across the set of all possible networks. The analysis of this model is summarized in the following result. In the experiment we will study a setting with zero decay, i.e., $\delta = 1$. Setting $\delta = 1$ has the analytical convenience that there exists a unique strict Nash equilibrium (for a discussion of the difficulties in characterizing Nash equilibrium that arise when $\delta < 1$, refer to Bala and Goyal (2000)). For this case the theoretical analysis is summarized as follows.

Proposition 2.1. *Suppose $\delta = 1$. For $k < V$ a Nash equilibrium network is connected and for $k > V$ a Nash equilibrium network is either connected or empty. Connected equilibrium architectures include flower networks (with a single hub and one or multiple cycles). For $k < V$ a strict Nash equilibrium network is a cycle and for $k > V$ a strict Nash equilibrium network is either the cycle or the empty network. If $k < (N-1)V$ the efficient network is a cycle and if $k > (N-1)V$ the efficient network is empty.*

Figure 1 presents examples of Nash equilibrium networks. Proposition 2.1 also tells us that equilibrium networks may depart from efficient networks: for $V < k < (N-1)V$ equilibrium includes empty and connected networks but the cycle is the unique efficient and strict Nash equilibrium network. To develop an understanding of the dynamic properties of different architectures we simulate myopic best response dynamics and obtain properties of limit networks.

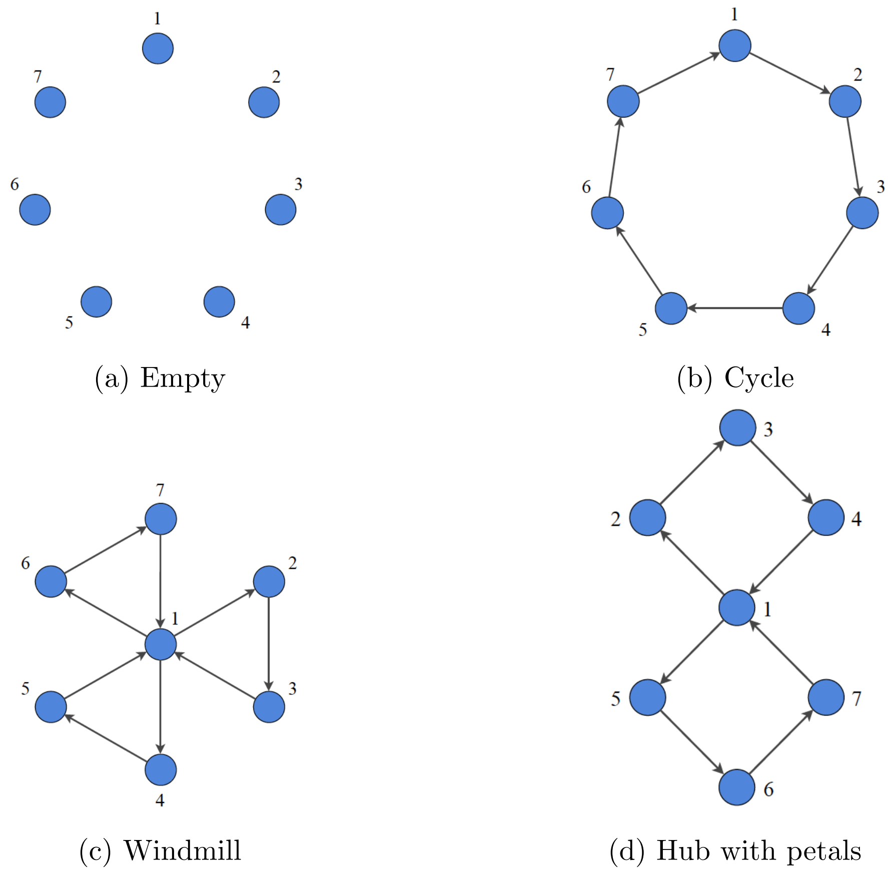


Figure 1: Nash networks in the one-way flow model

In the simulations of the dynamics we start from initial networks with low, medium, and high link density (the details are provided in Section A of the Appendix.) At each point a single individual (picked at random) makes a decision. This individual considers the potential payoffs that can be attained by adding, deleting, or switching a link with any other player, given the current network. Switching a link consists of a sequence of first adding (deleting) a link and then deleting (adding) another link. The individual chooses the option with the highest immediate payoff.

The parameters used in the simulations (and in the experiments) are as follows: $V = 10$ and $\delta = 1$ for $N = 10$ and $N = 50$. Under those parameter settings, the efficient network is a cycle and the strict Nash network is either the empty network or cycle (see Proposition 2.1). The cost of a link is adjusted across group size to keep incentives similar:[5] so $k = 20$ for $N = 10$, and $k = 100$ for $N = 50$. In over 99% of the simulations the process converged to the empty or cycle network. The remaining 1% corresponds to a network with two joined cycles that has similar characteristics to the cycle.

Table 1 also presents important properties of limit networks – connectedness, average degree, link inequality, and average distance. Recall that a component is a set of nodes with the property that every pair of nodes in it has a directed path running between them and there is no superset of such a set of nodes with this property. Connectedness of a network is the fraction of nodes that lies in its largest component. We define link inequality as the ratio of the difference between the largest degree and median degree divided by largest degree.[6] Finally, we define average distance as the mean distance among all pairs of nodes in the largest component of the network. Payoff inequality is the ratio of the difference between largest payoff and median payoff divided by the largest payoff.

	Empty		Cycle	
	$N = 10$	$N = 50$	$N = 10$	$N = 50$
One-way flow (120 obs.)	0%	4%	99%	96%
Connectedness	0	0	1	1
Indegree	0	0	1	1
Efficiency	0.15	0.03	1	1
Link inequality	0	0	0	0
Distance (largest component)	0	0	5	25
Payoff inequality	0	0	0	0

Table 1: Properties of limit networks in simulations under the one-way flow model

Notes: Every simulation iteration consists of randomly selecting an individual (with replacement) to add, delete, or switch a link as a myopic best response to the network resulting from the previous iteration. Different types of initial networks were considered (varying in link distribution and density). All simulations converged to a stable outcome. See section A in the Appendix for more details.

In the cycle network, every individual earns 80 for $N = 10$ and 400 for $N = 50$. Taken together, our theoretical results and the simulations suggest the following

[5] In the experiment, conversion rates were also adjusted to keep incentives similar across different group sizes. See Section E for more details.

[6] In the special case of the empty network with every node having degree 0, link inequality is set to 0.

hypotheses:

Hypothesis A *Under the one-way flow model, the limit network is the cycle. As group size grows from $N = 10$ to $N = 50$: (i) connectedness and efficiency remain high (equal to 1), (ii) distances grow (from 5 to 25), and (iii) link and payoff inequality remain low (equal to 0).*

2.2 Experimental design and procedures

The experiment consists of a continuous time game. The game is played over 6 minutes and is referred to as a round. The first minute is a trial period and the subsequent 5 minutes has payment consequences. The motivation for the trial period is to endogenize the selection of the initial network: we interpret the network created by subjects at the end of the trial period (minute 1) as the initial network for the subsequent payoff effective period (last 5 minutes). At any moment, during a round, each subject is informed about their network as follows: each subject is shown all the links in the entire network, and explicit visual information is provided to distinguish individuals they access (i.e., benefit from) from those they do not.[7] Figure 2 presents the decision screen observed by a subject. Subjects are shown their own payoff but not the payoffs of other subjects.

The top part of the screen depicts information about the timer indicating how much time has lapsed in the current round (the timer turns red when payoffs become effective, i.e., after more than 1 minute), and a comprehensive description of the subject's own payoff. Information about payoffs include gross earnings (from connections with others), the cost of linking (number of links formed multiplied by k), and the net earnings (costs subtracted from gross earnings).

The bottom part of the screen shows detailed information about the network (the subject's node is highlighted in yellow). The subject views the entire network but is explicitly shown the distinction between the nodes they access (in blue color) and those they do not (in grey color). The treatments with smaller group sizes use the

[7]We have tested a variant treatment that restricts network information to what is necessary to compute payoff i.e., only showing subjects the network structure of their own component). The results in this treatment were qualitatively similar but the failure of connectivity was greater in the larger group case. The data related to this extra treatment is available upon request.

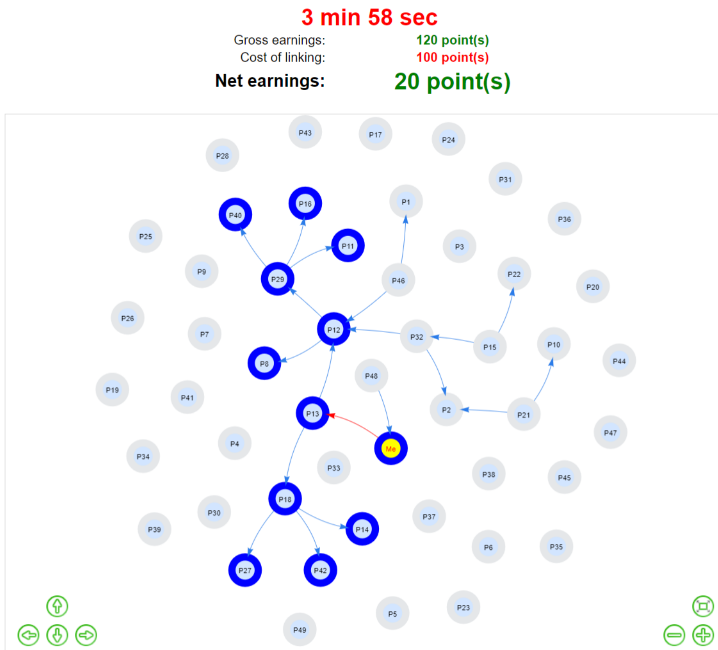


Figure 2: Example of decision screen in the one-way flow model

same interface (the scroll down feature is not available then because all players are directly visible on the screen).

Each game starts with the empty network. At any instant in the 6 minutes game, a subject can form/delete a link with any other subject by simply double-clicking on the corresponding node in the computer screen. At the end of each round, every subject is informed, using the same computer screen, of a time moment randomly chosen for payment. The subject is also provided detailed information on subjects' behavior at the chosen moment. Subjects' identification numbers were randomly reassigned at the beginning of every round, which makes it impossible to track the behaviors of others in previous rounds. This reduces potential repeated game effects and the possibility of reciprocation considerations playing a role in shaping behavior in later rounds.

Every group played 6 rounds. The first round was a trial round with no payoff relevance, the last 5 rounds were relevant for subjects' earnings. In analyzing the data, we will therefore focus on subjects' behavior and group outcomes from these last 5 rounds. Further details of the experimental procedures and network interface are described in Sections E and F of the Appendix.

There were 5 sessions in all: 1 session of 4 groups of 10 subjects for the $N = 10$ treatment, and 4 sessions of 50 subjects for the $N = 50$ treatment. In each experimental session, subjects were matched to form a group and interacted with the same subjects throughout the experiment. A total of 240 subjects participated in the experiment.

The experiment was conducted in the Laboratory for Research in Experimental and Behavioral Economics (LINEEX) at the University of Valencia. Sample instructions are presented in Section F of the Appendix. On average, subjects earned 12.3 euros (this includes a 5 euros show-up fee).

2.3 Experimental findings

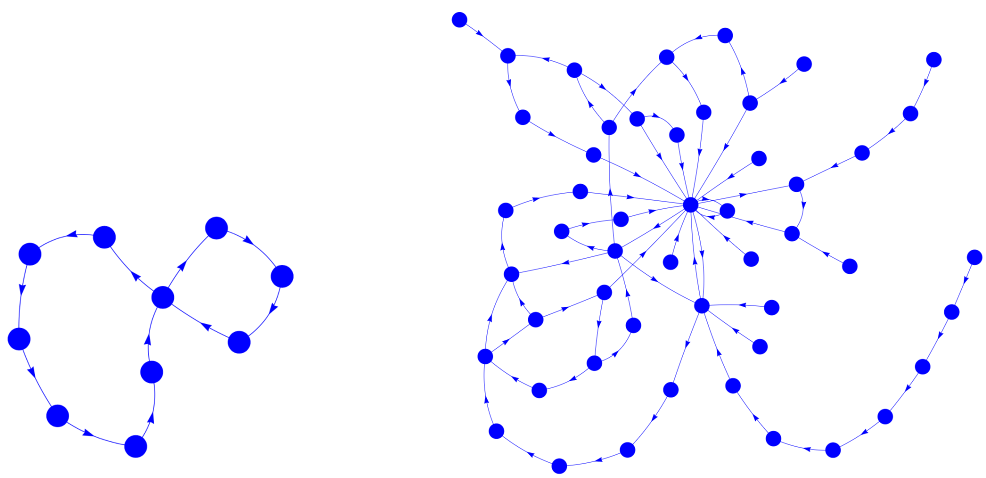
For simplicity, in the analyses that follow, the data used from every round of the game consists of observations (snapshots of every subject's choices) selected at intervals of one second. Unless stated otherwise, all statistical analyses consider data from the last 5 minutes (for the last five rounds of the experiment).

The dynamics of game outcomes are presented in section B of the Appendix . Because group outcomes become stable toward the end of the game, we compare the outcomes in the last 30 seconds of the game with theoretical benchmarks when presenting group size effects.

We start by examining the networks that have been created by subjects at the end of the trial period of each round (minute 1): as mentioned before, we interpret them as initial networks for the subsequent payoff-effective period (last 5 minutes). These initial networks have an average number of links per individual of 1.17 for $N = 10$ and 1.3 for $N = 50$ (detailed statistics are presented in Table 6 in section A.2 of the Appendix). It is worth noting that starting with these initial networks and simulating the best response dynamics as specified in Section 2.1, always leads to the cycle across both group sizes (these simulation dynamics are presented in Figure 17 in section A.2 of the Appendix). This is useful to bear in mind as we now turn to the experimental findings.

Figures 3a and 3b present the networks at minute 6 in groups of size $N = 10$

and $N = 50$. The networks in the figures are representative; the efficiency realized in these networks – 97.5% for N=10 and 44.9% for N=50 – is similar to the average efficiency observed in the last 30 seconds of the experiment. In particular, efficiency in the the experiments in the respective treatments is 90.7% for N=10 and 36.5% for N=50, respectively.[8] This group size effect is in contrast to Hypothesis A where efficiency is high across group sizes. We note that although the emerging network is sparse, link inequality due to the presence of a hub-like individual appears to be notable in the $N = 50$ group.



(a) $N = 10$: minute 6 (b) $N = 50$: minute 6

Figure 3: Snap shots of representative networks under the one-way flow model

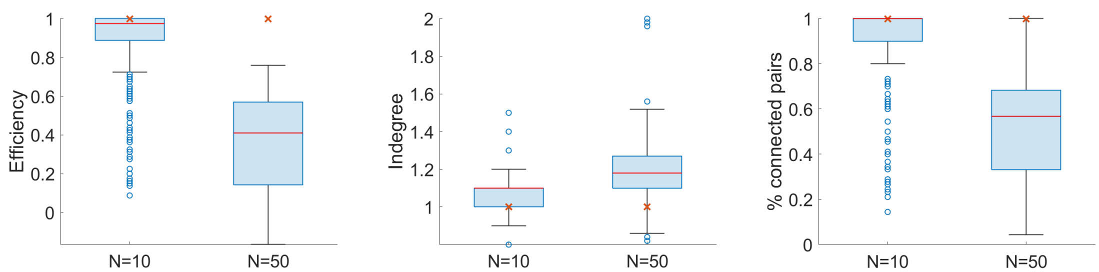
We now examine the patterns in the data systematically. Using the data of the last 30 seconds, we begin with the group size effect on efficiency.[9] Figure 4a shows that subjects attain over 90% efficiency in groups of size $N = 10$ – 50% of networks attain at least 97.5% efficiency, and 25% of networks maximize total welfare. By contrast, they attain under 40% efficiency in group size $N = 50$. This difference is large and statistically significant (two-sided Wilcoxon signed-rank (WSR) test: $p < 0.01$) and

[8]The animations of snap shots of the experiment are available at the following website: `https://moris.gate.cnrs.fr/software/animations_linking_game.html`.

[9]The dynamics of individual and group behavior over time is presented in Figure 20 in Section 5.

it is inconsistent with Hypotheses A.[10]

The inefficiency in the large group has two potential sources: over-linking and breakdown of connectedness. Figure 4b shows that networks are sparse: in the $N = 10$ group, average indegree is 1.07; in $N = 50$ group, the average indegree is 1.2. The difference across group sizes is modest but it is statistically significant (two-sided Wilcoxon signed-rank test with the group level average data: $p < 0.01$, number of observations: 20). On the other hand, Figure 4c shows that group size has powerful effects on connectedness: in group $N = 10$, connectedness is close to 100% (50% of networks are completely connected), whereas it reaches a little over 40% in the $N = 50$ case. Hence, we conclude that although over-linking is slightly more present in $N = 50$ group, connectivity breakdown is the main cause of the negative group size effect on efficiency in the one-way flow model. This negative group size effect on connectedness is inconsistent with Hypothesis A.



(a) Efficiency (b) Indegree (c) Connectedness

Figure 4: Effects of group size on efficiency under the one-way flow model

Notes: We use the data of last 30 seconds in every round. A red cross represents a theoretical benchmark. The red line in the box plot represents a median, the top and bottom edges of the box are the upper and lower quartiles, respectively, outliers (represented by circles) correspond to values more than 1.5 times the interquartile range away from the top or bottom of the box, and the whiskers above and below the box represent the nonoutlier maximum and minimum, respectively. Mean efficiency: 0.91 for N=10, 0.37 for N=50; mean indegree: 1.07 for N=10, 1.21 for N=50; mean connectedness: 0.93 for N=10, 0.52 for N=50.

To further understand the relation between group size and networks we study how far are the observed networks from the cycle. The *distance* of a network from the

[10]To address potential issues of learning across rounds, we replicate all the statistics reported in the paper with the data in the last two rounds. The results in the last two rounds are quite similar to those in the last 5 rounds. For example, in the last two rounds, the efficiency levels in the one-way flow model are 96% for $N = 10$ and 32% for $N = 50$, which are of similar magnitude as the results with the last five rounds in the paper.

cycle network is the number of link changes (adding or deleting a link) needed to move from that network to the cycle, normalized by the number of individuals. To get a sense of what this number means, note that starting from the empty network the distance to cycle is 1 in both group sizes, and starting from the complete network, the distance to cycle is N-2.

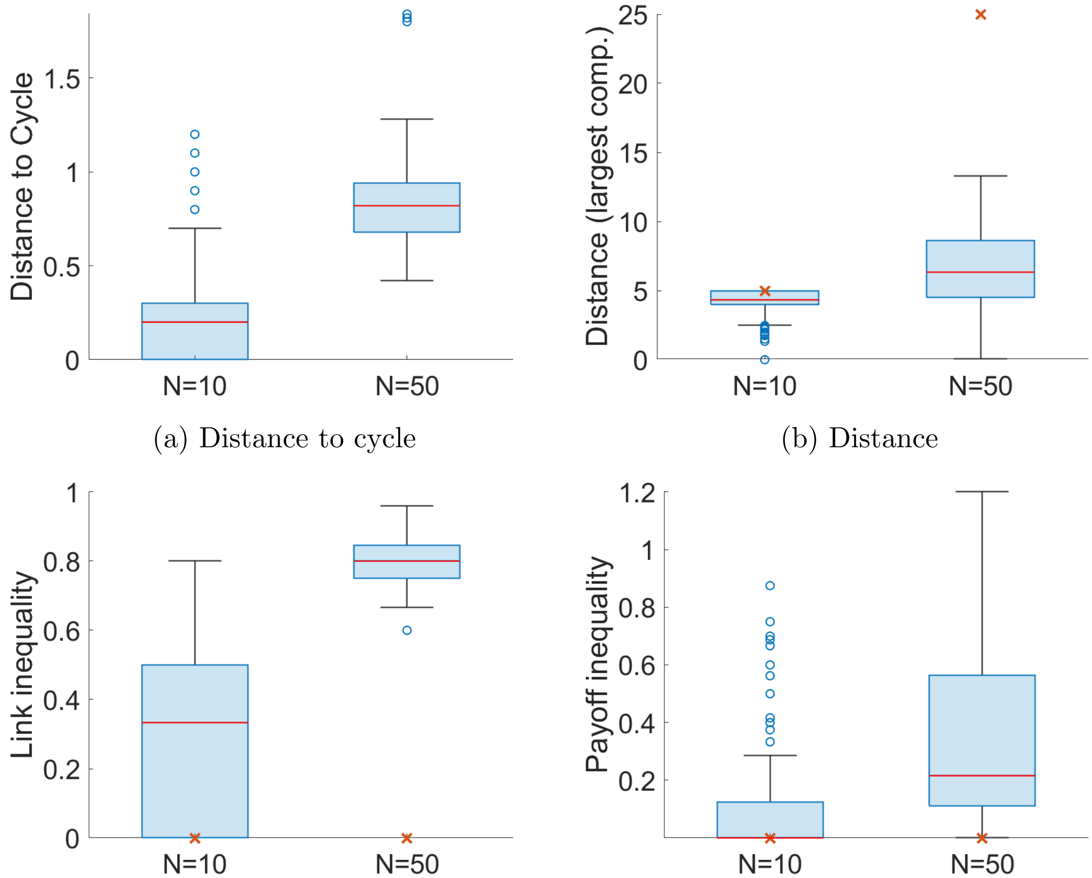
Figure 5a shows the distance to cycle in each group size. In group size $N = 10$, the cycle is reached perfectly for 25% of the observed networks (bottom edge of the box in the figure), the distance to cycle is below 0.2 for 50% of them (red horizontal line in the figure). This means that less than 2 link changes are needed to move from an experimentally observed network to the cycle network. By contrast, in $N = 50$, the distance to cycle is never lower than 0.42. Therefore, there is a powerful effect of group size on distance to cycle. This is inconsistent with Hypothesis A.

To deepen our understanding of network distances, we examine two network properties: distance and link inequality. Figure 5b shows that average distance in the largest component reaches 4.2 with little variance in $N = 10$ (as compared to 5 in a cycle), but is close to 7 with more heterogeneity in $N = 50$ (as compared to 25 in the cycle). The average distance under $N = 50$ is significantly larger than that for $N = 10$ (two-sided WSR test: $p < 0.01$). As group size increases, average distance grows but not to the degree to which the theory predicts in Hypothesis A.

Figure 5c shows that link inequality in $N = 10$ is around 0.26 but it grows with group size and is equal to 0.79 in group size $N = 50$. Recall that this ratio is equal to 0 in the cycle network. The inequality level is significantly different across the two group sizes (two-sided WSR test, $p < 0.01$). This is inconsistent with Hypothesis A.

Figure 5d shows that for group size $N = 10$ payoff inequality is close to 0 (50% of networks generate no inequality), in line with theoretical prediction, and for group size $N = 50$ it is on average around 0.5 (with 50% of networks below 0.22). The difference in payoff inequality is significant (two-sided WSR test, $p < 0.01$). This in inconsistent with Hypothesis A.

This payoff inequality can be seen to be a consequence of the breakdown of connectedness. Figure 19 in section B.1 of the Appendix shows that in $N = 10$, on average, 80% of subjects are connected to at least 90% of the group, while for $N = 50$, only 5% of subjects are connected to at least 90% of the group. Moreover, median



(a) Distance to cycle

(b) Distance

(c) Link inequality

(d) Payoff inequality

Figure 5: Effects of group size on networks and payoffs under the one-way flow model

Notes: We use the data of last 30 seconds in every round. A red cross represents a theoretical benchmark. The red line in the box plot represents a median, the top and bottom edges of the box are the upper and lower quartiles, respectively. Mean distance to cycle: 0.25 for N=10, 0.85 for N=50; mean distance: 4.22 for N=10, 6.58 for N=50; mean link inequality: 0.26 for N=10, 0.79 for N=50; mean payoff inequality: 0.07 for N=10, 0.50 for N=50.

number of others accessed differs significantly greatly across group sizes: for $N = 10$, the median node accesses 100% of others; for N=50, the median node only accesses 59% of others. These differences in median connectedness drive the group size effects on payoff inequality.

We summarize this discussion in our first finding.

Finding 1 In the one-way flow model experiment, the group size has a large impact on network structure, efficiency, and inequality. In particular,

(i) Networks distance to cycle increases from 0.25 in $N = 10$ to 0.85 in $N = 50$,

(ii) Efficiency falls from 0.91 in $N = 10$ to 0.37 in $N = 50$. This is mainly driven by the breakdown of connectedness,

(iii) Link inequality increases from 0.26 in $N = 10$ to 0.79 in $N = 50$. Payoff inequality grows from 0.07 in $N = 10$ to 0.5 in $N = 50$.

Thus an increase in group size leads to very large departure from the theoretical predictions. We develop a model of bounded rational individuals to explain this impact of group size in Section 4.

3 Two-way flow model

The two-way flow model assumes people mutually exchange benefits regardless of who creates the connection. Let us define the closure of g, an undirected network, denoted by $\bar{g}$ where $\bar{g}_{ij} = \max(g_{ij}, g_{ji})$ for every $i, j \in \mathbf{N}$. In other words, individuals benefit both from the link they form and the links they receive. For any pair of players i and j in g, the geodesic distance, denoted by $d(i, j; \bar{g})$, is the length of the shortest path between i and j in the undirected network $\bar{g}$. If no such a path exists, the distance is set to infinity.

Given a strategy profile s, the payoffs of player i are given by:

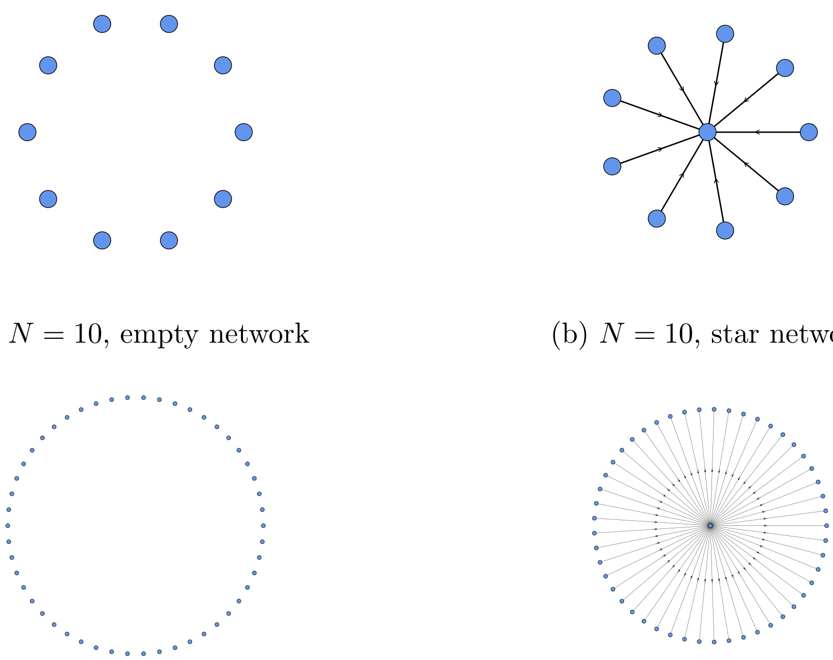
$$\Pi_i(s) = V + \sum_{j \in \mathbf{N} \setminus \{i\}} \delta^{d(i,j;\bar{g})} V - \eta_i(g)k \tag{2}$$

Our interest is in Nash equilibrium networks and efficient networks. In our study, we will focus on the case where decay is small, i.e., δ is close to 1 ($\delta = 0.9$ in the experiment). This is done to reduce the range of equilibrium: if $\delta = 1$, then any tree network is an equilibrium. Decay eliminates trees in which agents are far apart. In line with the restrictions in the one-way flow model, we will assume that the costs of a link are larger than return from an isolated agent but they are not too large so that linking with someone connected to everyone else is attractive: this means that $V\delta < k < V(\delta + (N-1)\delta^2)$. Bearing in mind these parametric restrictions, we state a result on equilibrium and efficiency networks in the two-way flow model (Goyal

(2023)).

Proposition 3.1. *A Nash equilibrium network is either connected or empty. If* $V\delta < k < V(\delta + (N-2)\delta^2)$ *then the empty network and the star network are Nash equilibrium networks. If* $2V[\delta - \delta^2] < k < 2V\delta + V(N-2)\delta^2$ *then the star is the unique efficient network.*

Figure 6 presents examples of equilibrium and efficient networks in the two-way flow model.



(a) $N = 10$, empty network

(b) $N = 10$, star network

(c) $N = 50$, empty network

(d) $N = 50$, star network

Figure 6: Nash and efficient networks: two-way flow

As there exist multiple equilibrium networks, in line with our strategy in the one-way flow model, we explore their robustness with respect to myopic best response dynamics. The rules of the dynamics are the same as described in the previous section on the one-way flow model. The parameters used in the simulations (and in the experiments) are as follows: $V = 10$ and $\delta = 0.9$ for $N = 10$ and $N = 50$. The cost of a link is adjusted across group size as in the one-way flow model: so $k = 20$

for $N = 10$, and $k = 100$ for $N = 50$. Under this choice of parameter values, the conditions from Proposition 3.1 are satisfied, and so the star is the unique efficient network. The empty and the star networks are both Nash networks. There is no general characterization of all equilibrium networks under these ranges of parameters. We conduct simulations of the best response dynamics, as in the one-way flow model. In our simulation runs we observe convergence to 2 types of network structures: the empty network and the star network (in which all links are paid for by the spokes).

Table 2 presents the properties of the network structures observed at the end of 80 simulations (further details of the procedure and statistics corresponding to each initial network can be found in Section A of the Appendix).

	Empty		Star	
	$N = 10$	$N = 50$	$N = 10$	$N = 50$
Two-way flow (120 obs.)	0%	1%	100%	99%
Connectedness	0	0	1	1
Indegree	0	0	0.9	0.98
Efficiency	0.15	0.03	1	1
Link inequality	0	0	0.89	0.98
Distance (largest component)	0	0	1.8	1.96
Payoff inequality	0	0	0.3	0.32

Table 2: Properties of limit networks in simulations under the two-way flow model

Notes: Every simulation iteration consists of randomly selecting an individual (with replacement) to add, delete, or switch a link as a myopic best response to the network resulting from the previous iteration. Different types of initial networks were considered (varying in link distribution and density). All simulations converged to a stable outcome. See section A of the Appendix for more details.

In the $N = 10$ case, the limit network is a star in all the simulations. In the $N = 50$ case, the limit network is empty in 1% of the cases and a star network in 99% of the cases; for all practical purposes, the dynamics converge to the star network for group size 10 and 50.

In the empty network individuals earn 10. In the star network, for $N = 10$ the hub and spokes earn 91 and 64, respectively; in the $N = 50$ case, they earn 451 and 308, respectively. The hub earns roughly 50% more than the spokes in both group sizes and payoff inequality changes little as group size increases.

Proposition 3.1 and the simulations motivate the following hypothesis in the two-way flow model.

Hypothesis B *Under the two-way flow model, the limit network is a star. As group size grows from $N = 10$ to $N = 50$ (i) connectedness and efficiency remain high (equal to 1), (ii) average distance changes little (from 1.8 to 1.96), (iii) linking inequality grows (from 0.89 to 0.98), and (iv) payoff inequality changes little (remains close to 0.3).*

3.1 Experimental design and parameters

The design is the same as in the one-way flow model except for one difference that pertains to the information on networks provided to subjects. In the two-way flow experiments, subjects see the nodes and the (undirected) links in their component. Recall that two nodes belong to the same component of the network, if there is a path connecting them in the undirected network as defined at the start of the section. For easy reference, we provide the network interface in the two-way flow model here.

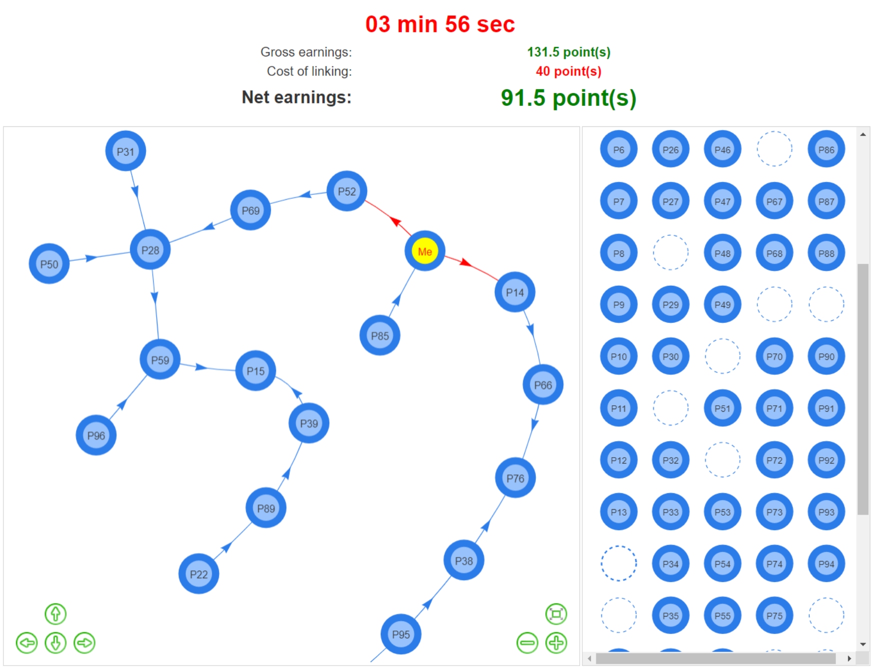


Figure 7: Decision screen in two-way flow model

The motivation of this design choice is to limit possible information overload for subjects by only providing the minimum amount of information required for them to compute their payoff at any moment of the game (payoffs are only affected by others from the same component). We note however that, in the experiment, subjects saw almost the entire network structure throughout the last 5 minutes of the game because of the very high level of connectedness (subjects' component includes almost all other subjects, as shown through Figure 9b).

In total there were 5 sessions in the two-way flow model experiment: 1 session of 4 groups of 10 subjects for the $N = 10$ treatment, 4 sessions of 50 subjects for the $N = 50$ treatment. A total of 240 subjects participated, and they earned on average 15.3 euros (this includes a 5 euros show-up fee). Further details of the experimental procedures and sample instructions are provided in Sections E and F of the Appendix.

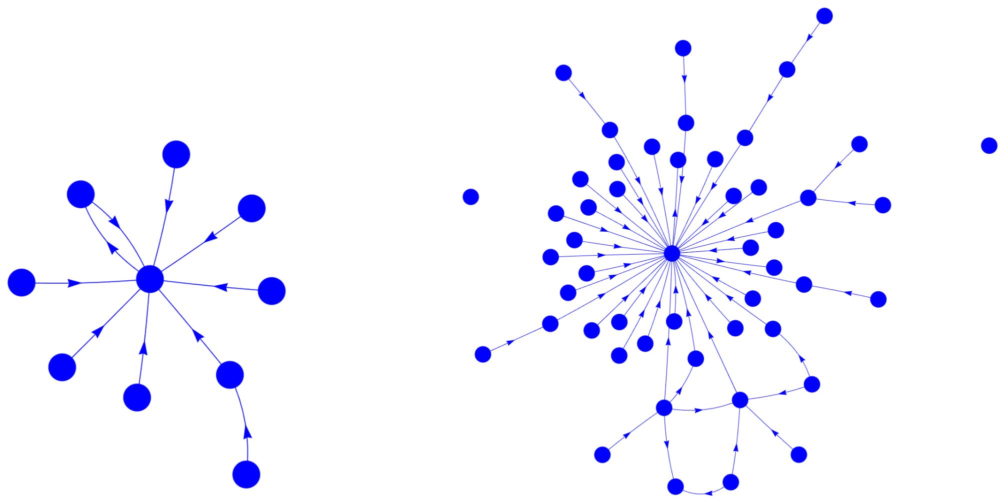
3.2 Experimental findings

We start by examining the initial networks selected by subjects at the end of the trial period of each round (minute 1). Those networks exhibit an average number of links per individual of 0.93 for $N = 10$ and 2.2 for $N = 50$ (detailed statistics are presented in Table 7 in section A.2 of the Appendix). Starting with these initial networks and simulating the best response dynamics leads to the star (these simulations are presented in Figure 18 in section A.2 of the Appendix).

Figures 8a and 8b present the networks observed at minute 6 in a group of each group size treatment. The networks in the figures are representative; the efficiency realized in these networks – 98.3% for N=10 and 82.5% for N=50 – is similar to the average efficiency observed in the last 30 seconds of the game (across the 20 rounds).

We first analyze the group size effects on efficiency. Figure 9a shows that efficiency falls with group size: in $N = 10$, subjects attain close to 0.9 (50% of networks attain above 0.94), in $N = 50$, efficiency is close to 0.8. This difference in efficiency is statistically significant (two-sided WSR test: $p < 0.01$). This group size effect on efficiency is inconsistent with the prediction stated in Hypothesis B.[11]

[11] We replicate all the statistics reported in the paper with the data in the last two rounds. The results in the last two rounds are quite similar to those in the last 5 rounds. For example, in the last two rounds, the efficiency levels in the two-way flow model are 89% for $N = 10$ and 83% for $N = 50$, which are of similar magnitude as the results with the last five rounds in the paper.



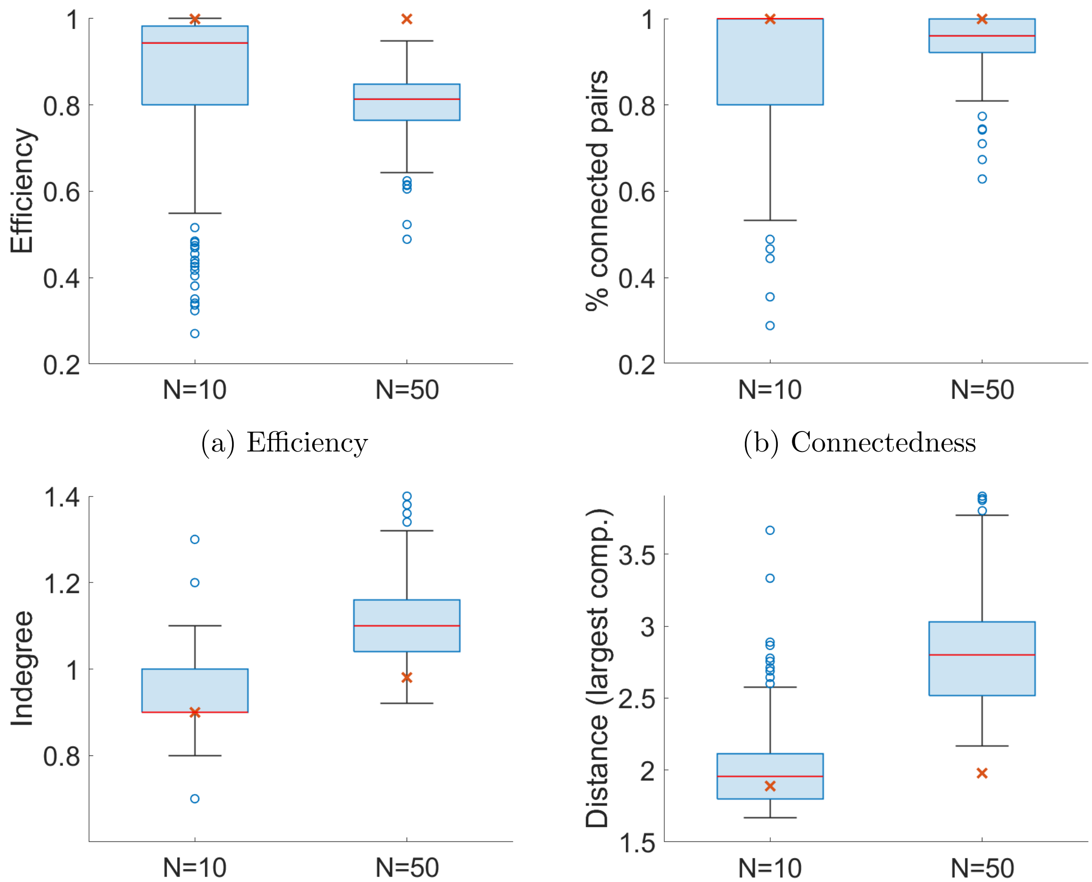
(a) $N = 10$: minute 6 (b) $N = 50$: minute 6

Figure 8: Snap shots of representative networks under the two-way flow model

In our context, there are three possible sources of inefficiency: breakdown of connectedness, over-linking, and large average distance. Figure 9b shows that the level of connectedness is quite high in both groups: 0.92 in $N = 10$ and 0.96 in $N = 50$. Figure 9c shows that the average indegree is 0.9 in $N = 10$ (with 25% of observations of the theoretical prediction) and goes up to 1.1 in $N = 50$. This difference in average indegree is modest but statistically significant (two sided WSR test with the group level average data: $p < 0.01$). Figure 9d shows that the average distance increases in group size: it is 2.01 in $N = 10$ (with 25% of observation round the theoretical prediction 1.9) and 2.8 in $N = 50$. This difference in distance is statistically significant (two sided WSR test: $p < 0.01$).

These differences in indegree and distance are deviations from the theoretical prediction summarized in Hypothesis B, and they move in the same direction as the group size effect on efficiency. We conclude that the negative group size effect on efficiency in the two-way flow model is mainly driven by over-linking (creating more links than necessary) and long distances in the large group.

We next turn to analyze the group size effects on network structure. We start with distance from star. Figure 10a shows that distance from star network in $N = 10$ is close to 0.4, while in $N = 50$ it is close to 0.9 (minimum distance observed is 0.36).



(a) Efficiency (b) Connectedness

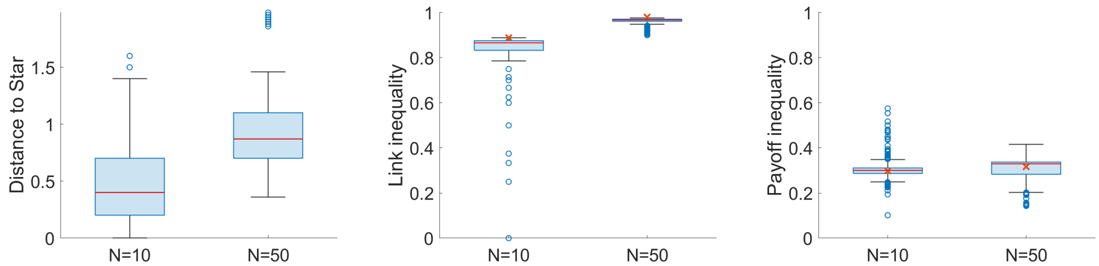
(c) Indegree (d) Distance

Figure 9: Effects of group size on efficiency under the two-way flow model

Notes: We use the data of last 30 seconds in every round. A red cross represents a theoretical benchmark. The red line in the box plot represents a median, the top and bottom edges of the box are the upper and lower quartiles, respectively. Mean efficiency: 0.88 for N=10, 0.81 for N=50; mean connectedness: 0.92 for N=10, 0.96 for N=50; mean indegree: 0.91 for N=10, 1.11 for N=50; Mean distance: 0.49 for N=10, 0.91 for N=50.

Thus there is a large impact of group size on distance to star. This is inconsistent with Hypothesis B.

Figure 10b shows that link inequality increases slightly in group size: the average ratio is 0.82 in $N = 10$, and 0.96 in $N = 50$ (two-sided WSR test, $p < 0.01$), with little variance in both treatments. Figure 10c shows that payoff inequality is close to theoretical prediction in both groups. The data patterns regarding inequalities of link and payoff are largely in line with the predictions stated in Hypothesis B.



(a) Distance to star (b) Link inequality (c) Payoff inequality

Figure 10: Effects of group size on networks and payoffs under the two-way flow model

Notes: We use the data of last 30 seconds in every round. A red cross represents a theoretical benchmark. The red line in the box plot represents a median, the top and bottom edges of the box are the upper and lower quartiles, respectively. Mean distance: 2.01 for N=10, 2.79 for N=50; mean link inequality: 0.82 for N=10, 0.96 for N=50; mean payoff inequality: 0.29 for N=10, 0.31 for N=50.

We summarize these observations in our second finding:

Finding 2: In the two-way flow model experiment, we observe modest group size effects on efficiency and network structure. As group size grows,

(i) Networks' average distance to star increases from 0.49 in $N = 10$ and 0.91 in $N = 50$,

(ii) Efficiency falls from 0.88 in $N = 10$ and 0.81 in $N = 50$. This is caused mainly by over-linking and long distance,

(iii) Link inequality remains high and increases from 0.82 in $N = 10$ and 0.96 in $N = 50$; payoff inequality remains low and changes little from 0.29 in $N = 10$ to 0.31 in $N = 50$

Overall, subjects form networks that are fairly similar to star networks in both small and large group sizes, characterized by their sparsity, high connectedness, small average distances, high link inequality.

4 Decision Rules

We found that an increase in group size has a powerful effect on networks and welfare and that these effects differ across the one-way and the two-way flow model. To develop an understanding of the sources of these group size effects, inspired by the work of Caria and Fafchamps (2020), we estimate a bounded rational model in which individuals make noisy choices that are either in line with myopic best response, targeting a highest-degree individual, or uniformly at random. The behavioral rule of targeting a highest-degree individual is relevant even in the one-way flow model due to the evidence of link inequality. We then simulate the estimated bounded rational model to check if it can replicate the group size effects found in the data.

To motivate the use of this bounded rational model, we first examine how close the observed behaviors of our subjects are to myopic best response. Figure 11 shows the average distribution of distance to best response in each group for each flow model. At any moment t, this is described as the minimum number of link changes (adding or deleting a link) for an individual to move from their realized action at t to the action (vector of links) that maximizes their payoff (given the choices made by others in the group at $t-1$).[12] Two observations can be made from this figure: first, distance to best response behavior is similar across group sizes in each flow model, both in terms of mean values and in terms of distribution. Perhaps more surprisingly, we observe that in the last 30 seconds, a large fraction of subjects make choices that are consistent with best response behavior. For the one-way flow model, the distance to best response is 0 in 86% of the cases for $N = 10$, and the distance is 0 in 80% of the cases for $N = 50$. For the two-way flow model, the distance to best response is 68% for $N = 10$, and 54% for $N = 50$. This suggests that the deviations from the equilibrium prediction of cycle in the one-way flow model and the star in the two-way flow model observed in Figures 5a and 10a, which are more significant in larger groups, is generated by a minority of individuals who deviate from best response.

In our decision model, individuals make decisions for 180 periods in total (i.e., 1 period is equivalent to 2 seconds in the experiment). We define one period as two seconds because switching links, which involves two consecutive choices, is considered a single action. It is assumed that each individual independently has a 0.1 probability of making an action in each period. This value is similar to the activity levels exhibited by subjects from our experiment. When they are picked, individuals make a decision in the following way: they make a best response to the network observed at the previous period with probability $1 - \epsilon_1 - \epsilon_2$, make a uniform random choice with

[12] For example, an individual needing to switch one link to reach their best response action would have a distance of 2 (deleting a link and adding another link)

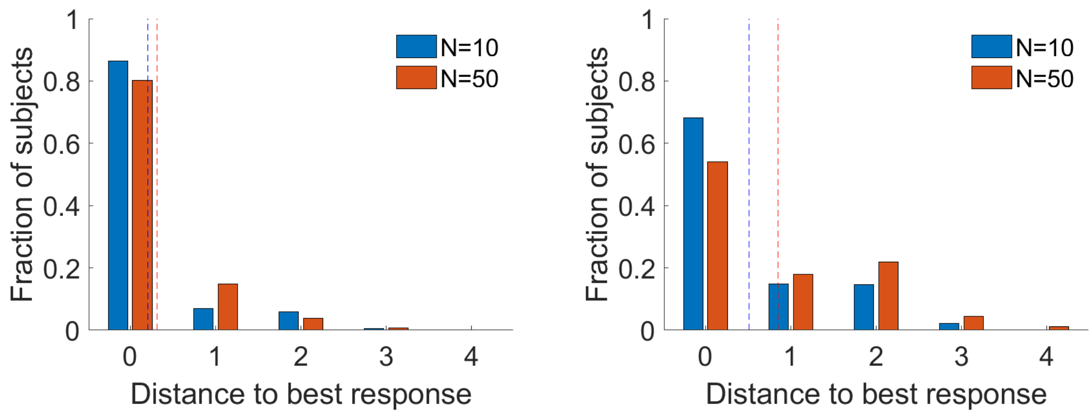


Figure 11: Average distribution of distance to best response

Notes: In these figures, each observation corresponds to a distribution computed at a moment of the last 30 seconds for a given group (snapshot), in a given round (600 observations overall); vertical dashed lines represent average distance to best response (for one-way flow model, 0.21 in $N = 10$ and 0.32 in $N = 50$; for two-way flow model, 0.51 in $N = 10$ and 0.85 in $N = 50$).

probability ϵ_1, and link to a highest-degree individual with probability ϵ_2. More specifically, the decision space in each period includes four types of choices: doing nothing, adding one link, deleting one link, and switching a link (deleting a link and adding a different link). Thus, for each individual i with out-degree m, the number of possible actions is $1 + (N - m - 1) + m + (N - m - 1)m = (m + 1)(N - m)$, where 1, $(N - m - 1)$, m, and $(N - m - 1)m$ represent the number of possible actions of doing nothing, adding a link, deleting a link, and switching a link, respectively.

When the trembling event (ϵ_1) occurs, each individual chooses uniformly at random over the decision space. When the event of targeting a highest degree node (ϵ_2) occurs, an individual will (1) add a link to a highest-degree node (if she has not formed that link and she is not the highest-degree node) and (2) delete one existing link or not delete any link (each with equal probability $\frac{1}{m+1}$, where m is outdegree). We conduct a generalized method of moments estimation of the two parameters $\{\epsilon_1, \epsilon_2\}$ in each treatment by matching three network statistics – indegree, connectedness, and link inequality – using the data of the last 30 seconds of the game. The details

Model	Group size	ϵ_1	ϵ_2	Obj. fcn
One-way	N=10	0.058***	0.008	1.381
		(0.009)	(0.008)	
	N=50	0.093**	0.197***	7.592
		(0.036)	(0.026)	
Two-way	N=10	0.217***	0.028	3.081
		(0.019)	(0.022)	
	N=50	0.163***	0.036	8.322
		(0.042)	(0.038)	

Notes: Simulation-based standard errors are reported in parentheses. ** denotes $p<0.05$. *** denotes $p<0.01$.

Table 3: Estimation results of the bounded rational model

of the GMM estimation method are given in section C of the Appendix.[13]

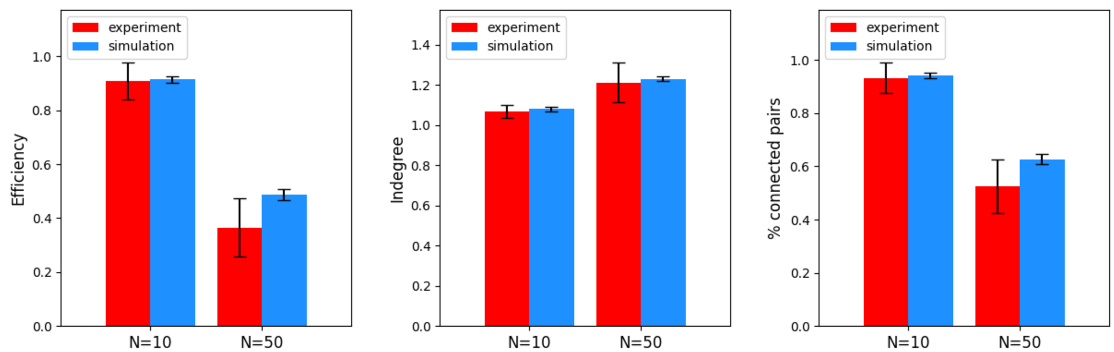
Table 3 reveals the treatment effects on the estimated parameters. First, the behaviour of a large proportion of subjects corresponds to best response and these probabilities range from 71% to 93% across group sizes in both models. Second, in the one-way flow model with $N = 50$, the estimated probability of targeting the highest degree individual is significantly positive (approximately 0.2), while this number is close to zero in all other three treatments. Also, the sum of the two types of errors is the smallest (≈ 0.07) in the one-way model with $N = 10$, while the sum is between 0.2 and 0.3 in the other three treatments.[14]

We use the estimated parameters to simulate the network formation dynamics. We start with the networks at the end of the first minute in the experiments and simulate the remaining five minutes of the game. The network patterns in the last 30 second are robust to variation in the initial networks. 200 simulations are conducted for each treatment. Figures 12 and 13 present the results of the last 30 seconds of simulated dynamics from the estimated behavioral model, along with the results from

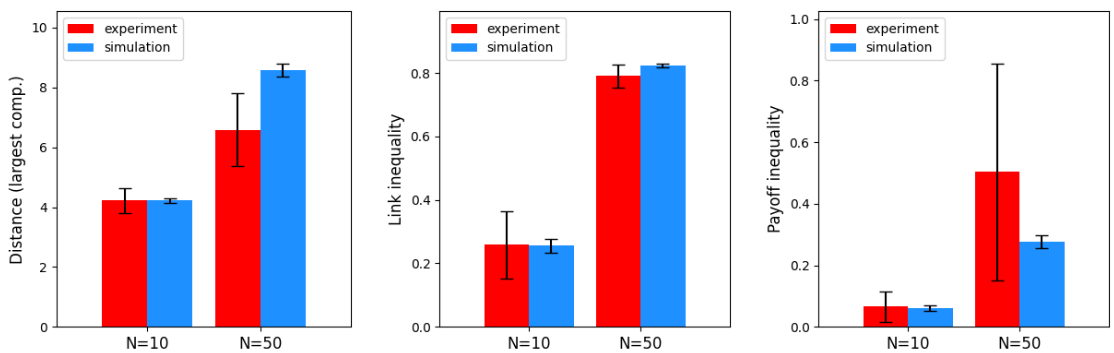
[13]For simplicity, this paper estimates common parameter values across all players. A variation could involve a finite mixture model that accounts for heterogeneity in individual behavioral parameters, estimating both these parameters and the proportion of each player type (See, e.g., Costa-Gomes et al. (2001); Sutter et al. (2013)).

[14]One might think that it is natural for individuals to link to the highest-degree individual in the case of two-way flow. Indeed, connecting to the highest-degree individual is often a best response in the two-way flow model, which is captured by the best response probability (i.e., $1 - \epsilon_1 - \epsilon_2$). The parameters can be identified because best response and targeting the highest degree individual are not identical.

the data, for each of the six statistics – efficiency, indegree, connectedness, distance, link inequality, and payoff inequality. The simulation results show that the behavioral model captures the treatment effect of group sizes and fits the data quite well. In the one-way flow model, as shown in Figure 12, the behavioral model replicates the high (low) efficiency, high (low) connectedness, and low (high) link inequality under the case of $N = 10$ (resp. $N = 50$) found in the experiments. In the two-way flow model, the difference in simulation results between $N = 10$ and $N = 50$ is modest, consistent with the experimental outcomes.



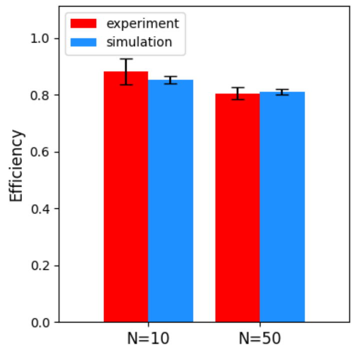
(a) Efficiency (b) Indegree (c) Connectedness



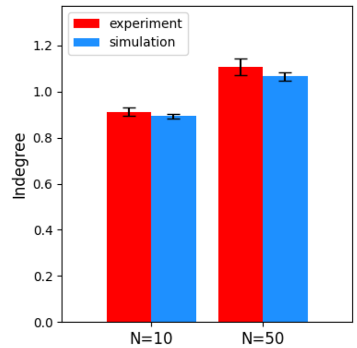
(d) Distance (e) Link inequality (f) Payoff inequality

Figure 12: Comparing the model fits with the data under the one-way flow

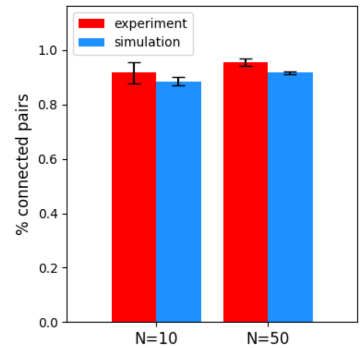
Notes: Error bars display standard 95% confidence interval around the mean.



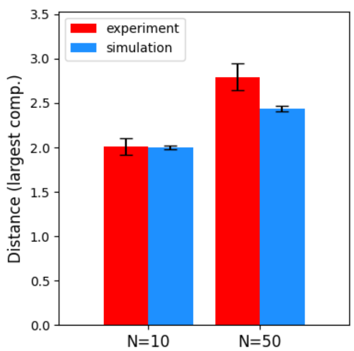
(a) Efficiency



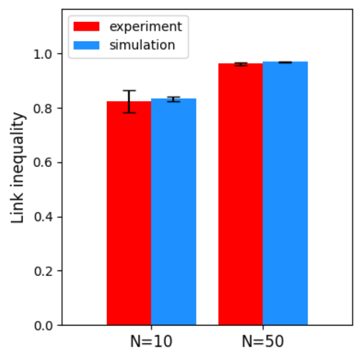
(b) Indegree



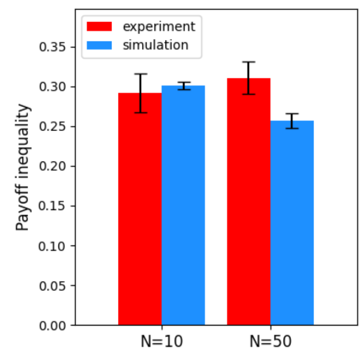
(c) Connectedness



(d) Distance



(e) Link inequality



(f) Payoff inequality

Figure 13: Comparing the model fits with the data under the two-way flow

Notes: Error bars display standard 95% confidence interval around the mean.

5 Discussion

This section further explores two themes of our work. One, we ask if the group size effects we have reported are robust to larger group sizes; we study group sizes of 100. Two, we consider the dynamics of linking over the entire period of the experiment: note that in the paper, so far, we have focused on the last 30 seconds of the experiment. We would like to examine the overall dynamics to see if there are patterns of behaviour we have overlooked at the earlier stages of the experiment and if they differ across the one-way and the two-way flow model.

5.1 Extrapolation with group size

Our principal finding is that an increase in group size has powerful and distinct effects in the two models. To illustrate the robustness of this finding, we ask if similar effects obtain as we move from groups of size 50 to groups of size 100. We do this by simulating the link formation dynamics with three group sizes $N = 10, 50, 100$ by adjusting the cost of a link and using the behavioral model estimated in Section 4. The details are shown in section D of the Appendix.

Figures 22 and 23 (in the section D of the appendix) present the simulation outcomes of the last 30 seconds for $N = 100$, alongside those for $N = 10$ and $N = 50$. They show that in the one-way flow model, as group size increases further to $N = 100$, the deviation from equilibrium increases, as evidenced by decreasing efficiency and connectedness and increasing link and payoff inequality. In contrast, in the two-way flow model, the impact of group sizes on network structures is modest, as evidenced by high efficiency, connectedness, and link inequality, along with modest payoff inequality across all group sizes.

The extrapolation exercise reinforces the primary finding in the paper that transitioning from small to large groups significantly affects networks and welfare in the one-way flow model, while having a relatively modest impact in the two-way flow model. A possible interpretation of our results is that subjects find it more difficult to coordinate in large groups, especially under the one-way flow model. Following up on the work by Weber (2006), efficient coordination in large groups could be achieved by starting with a small group size playing the game and gradually increasing the number of players. We leave this as an important avenue for future work large networks as networks often start off with small groups and then the group size grows over time.

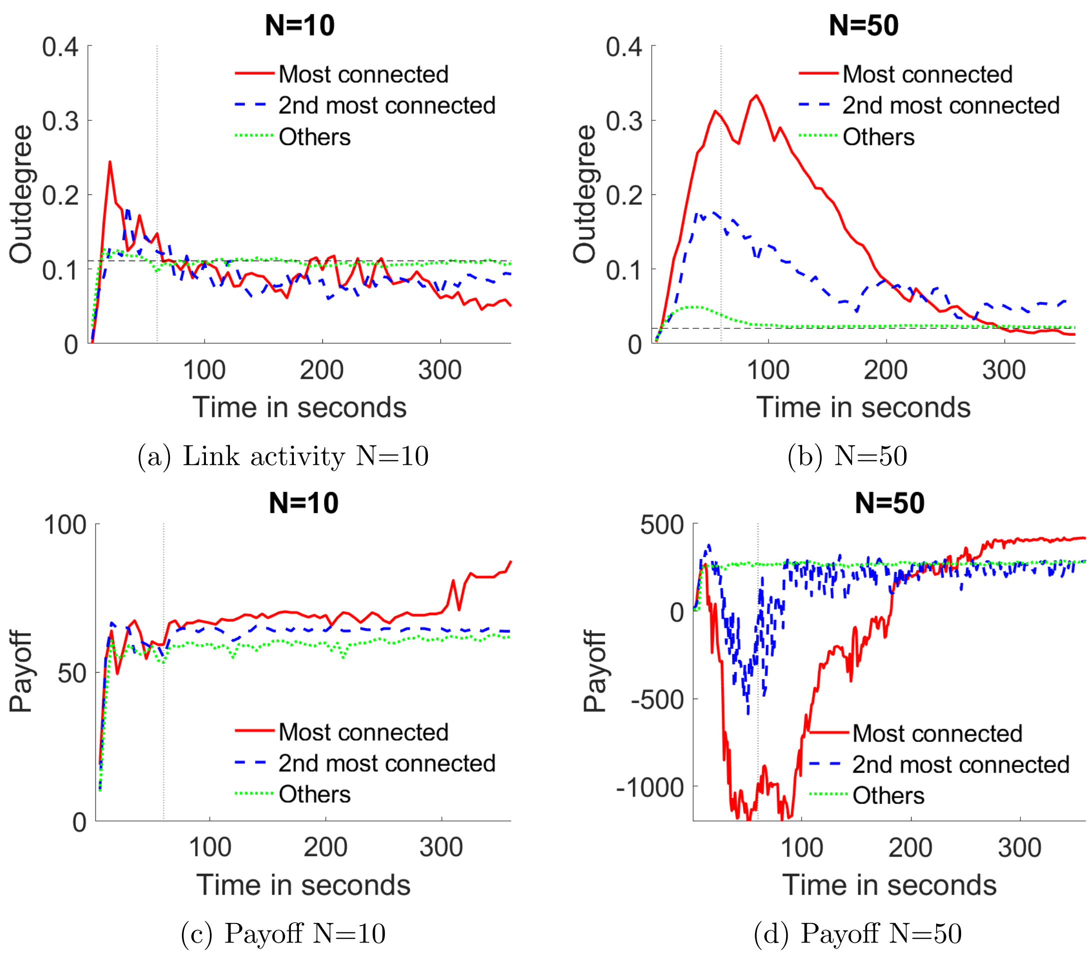
5.2 Dynamics of Behavior

Our analyses have focused on the limit behavior observed in the last 30 seconds of the game. We now briefly discuss the dynamics of behavior through the last entire payoff relevant period of the game – the last 5 minutes. Looking at the temporal data in the one-way flow model does not reveal new insights since behavior is fairly stable over time (see Figure 20 in the Appendix). Matters are however more interesting in the two-way flow model and we turn to this subject now.

We note that aggregate behavior becomes fairly stable over time (see Figure 21 in the Appendix) but the hubs can exhibit different dynamics. Figure 14 presents the time series of the average number of links created (outdegree) and median payoffs obtained by three types of subjects across group sizes in the two-way flow model: the most connected individual (with highest degree), the second most connected individual (second highest degree), and the others. We observe that in the small groups, the hub (most connected individual) forms few links but in the large group the hub forms a very large number of links initially in a bid to become central – the idea being that these initial links encourage others to link with them. Over time, however, the hub gradually deletes its own links to economise on costs (see Figure 14b). Thus, in large groups, the hub invests at the early stages (forming links is costly) and hopes to recover these costs and make profits by becoming the hub (without forming any link eventually). Interestingly, such forward looking behavior is exhibited only by a few individuals who compete with each other to become the hub. While we observe higher payoffs earned by the hub in the latter stage of the game (see Figure 14d), as expected, it turns out that they do not compensate for the very large losses made in the initial investment phase. In aggregate, the average payoffs for the most connected individual are negative in the large groups (they are around -60). Since forming zero link yields a payoff of $V = 10$, this strategy is clearly dominated. So why do subjects choose to make large investments and earn negative payoffs?

We note that there is no such over linking in the small group and that the hub then earns larger payoffs than the spokes. Thus the problem of negative payoffs arises as we raise the group size. We speculate that it may be information overload caused by the complexity of the evolving network in the large group that may be giving rise to the problem of over-linking early in the game. In other words, the subjects seeking to become the hub do not appreciate the full payoff consequences of their actions and therefore mistime their linking activity. It is also worth noting that this behavior by the most connected individuals largely explains the fall in efficiency observed in the early part of the game for large groups.

This analysis of dynamic behavior in the experiment allows us to make two general observations. First, while a few individuals seem to follow some sophisticated forward looking reasoning, the behavior of the vast majority of subjects over time is largely consistent with a myopic best response decision rule combined with targeting most connected individual and some random noise, as discussed in the previous section. Second, although the extreme behavior of a few individuals can influence who will be selected as the hub in the network, it is not necessary to reach the limit networks observed in the experiment. Indeed, our simulation exercise shows that the dynamics of a myopic best response with noise leads to a star network and therefore (roughly) reproduces the observed patterns.



(a) Link activity N=10

(b) N=50

(c) Payoff N=10

(d) Payoff N=50

Figure 14: Dynamics of subjects' behavior and payoffs under the two-way flow

References

Acemoglu, D. and P. D. Azar (2020): "Endogenous production networks," *Econometrica*, 88, 33–82.

Acemoglu, D., V. Carvalho, A. Ozdaglar, and A.Tahbaz-Salehi (2012): "The Network Origins of Aggregate Fluctuations," *Econometrica*, 80, 1977–2016.

Bala, V. and S. Goyal (2000): "A Non-Cooperative Model of Network Formation," *Econometrica*, 68, 1181–1231.

Banerjee, A., E. Breza, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, and C. Kinnan (2021): "Changes in social network structure in response to exposure to formal credit markets," Tech. rep., National Bureau of Economic Research.

Berninghaus, S. K., K.-M. Ehrhart, and M. Ott (2006): "A network experiment in continuous time: The influence of link costs," *Experimental Economics*, 9, 237–251.

Breza, E. (2016): "Field Experiments, Social Networks and Development," in *Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoulle, A. Galeotti, and B. Rogers, Oxford University Press.

Calford, E. and R. Oprea (2017): "Continuity, Inertia, and Strategic Uncertainty: A Test of the Theory of Continuous Time Games," *Econometrica*, 85, 915–935.

Callander, S. and C. R. Plott (2005): "Principles of network development and evolution: an experimental study," *Journal of Public Economics*, 89, 1469–1495.

Caria, A. S. and M. Fafchamps (2020): "Can people form links to efficiently access information?" *The Economic Journal*, 130, 1966–1994.

Chandrasekhar, A. G., H. Larreguy, and J. P. Xandri (2020): "Testing models of social learning on networks: Evidence from two experiments," *Econometrica*, 88, 1–32.

Charness, G., F. Feri, M. A. Meléndez-Jiménez, and M. Sutter (2014): "Experimental games on networks: Underpinnings of behavior and equilibrium selection," *Econometrica*, 82, 1615–1670.

Choi, S., E. Gallo, and S. Kariv (2016): "Networks in the Laboratory," in *Oxford Handbook of the Economics of Networks*, ed. by Y. Bramoulle, A. Galeotti, and B. Rogers, Oxford University Press.

Choi, S., S. Goyal, F. Guo, and F. Moisan (2024): "Experimental Evidence on the Relation Between Network Centrality and Individual Choice," *Jane Institute Working Paper University of Cambridge*.

Choi, S., S. Goyal, and F. Moisan (2023a): "Brokerage rents and intermediation networks," *Proceedings of the National Academy of Sciences*, 120, e2301929120.

Choi, S., S. Goyal, F. Moisan, and Y. Y. T. To (2023b): "Learning in Networks: An Experiment on Large Networks with Real-World Features," *Management Science*, 69, 2778–2787.

Costa-Gomes, M., V. P. Crawford, and B. Broseta (2001): "Cognition and behavior in normal-form games: An experimental study," *Econometrica*, 69, 1193–1235.

Currarini, S., F. Feri, B. Hartig, and M. A. Meléndez-Jiménez (2023): "Transmission and use of information in network games," *European Economic Review*, 155, 104443.

Erdős, P. and A. Rényi (1960): "On the evolution of random graphs," *Publ. Math. Inst. Hung. Acad. Sci*, 5, 17–60.

Fajgelbaum, P. D. and E. Schaal (2020): "Optimal transport networks in spatial equilibrium," *Econometrica*, 88, 1411–1452.

Falk, A. and M. Kosfeld (2012): "Its all about Connections: Evidence on Network Formation," *Review of Network Economics*, 11.

Farboodi, M. (2023): "Intermediation and voluntary exposure to counterparty risk," *Journal of Political Economy*, 131, 3267–3309.

Feri, F. (2007): "Stochastic stability in networks with decay," *Journal of Economic Theory*, 135, 442–457.

Friedman, D. and R. Oprea (2012): "A Continuous Dilemma," *American Economic Review*, 102, 337–63.

Galeotti, A. and S. Goyal (2010): "The law of the few," *The American Economic Review*, 100, 1468–1492.

Gallo, E. and C. Yan (2023): "Efficiency and equilibrium in network games: An experiment," *Review of Economics and Statistics*, 105, 1515–1529.

Goeree, J. K., A. Riedl, and A. Ule (2009): "In search of stars: Network formation among heterogeneous agents," *Games and Economic Behavior*, 67, 445–466.

Goyal, S. (2023): *Networks: An Economics Approach*, MIT Press.

Goyal, S., S. Rosenkranz, U. Weitzel, and V. Buskens (2017): "Information Acquisition and Exchange in Social Networks," *Economic Journal*, 127, 2302–2331.

Grimm, V. and F. Mengel (2020): "Experiments on belief formation in networks," *Journal of the European Economic Association*, 18, 49–82.

Hendricks, K., M. Piccione, and G. Tan (1995): "The economics of hubs: The case of monopoly," *The Review of Economic Studies*, 62, 83–99.

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

Horvath, G. (2023): "Network formation and efficiency in linear-quadratic games: An experimental study," *Working Paper*.

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

Olaizola, N. and F. Valenciano (2014): "Asymmetric flow networks," *European Journal of Operational Research*, 237, 566–579.

Price, D. d. S. (1976): "A general theory of bibliometric and other cumulative advantage processes," *Journal of the American society for Information science*, 27, 292–306.

Price, D. J. d. S. (1965): "The scientific foundations of science policy," *Nature*, 206, 233–238.

Sutter, M., S. Czermak, and F. Feri (2013): "Strategic sophistication of individuals and teams. Experimental evidence," *European Economic Review*, 64, 395–410.

van Leeuwen, B., T. Offerman, and A. Schram (2020): "Competition for status creates superstars: An experiment on public good provision and network formation," *Journal of the European Economic Association*, 18, 666–707.

Weber, R. A. (2006): "Managing growth to achieve efficient coordination in large groups," *American Economic Review*, 96, 114–126.

ONLINE APPENDICES

A Simulation Details

A.1 General predictions

Given the partial characterization of the equilibrium networks across both models, we perform computer simulations to explore which network structure(s) are most likely to emerge from a myopic best response dynamics. More specifically, we consider simulations in which each iteration consists of randomly selecting an individual to make a choice (with replacement). In this case, when selected to make a choice, an individual considers the potential payoffs that could immediately be reached (according to either model) by adding, deleting, or switching a link with any other person from the group, given the most recent network structure (at the end of the previous iteration). The action with the highest immediate payoff is then updated accordingly. Under this simulation setting, we vary the type of initial network structure to identify any potential effect on the dynamics and the converging networks.

To generate these initial networks, we consider two well-known network models: the Erdos-Renyi model that generates random networks with equally distributed connections (Erdős and Rényi, 1960), and the Price model that generates scale free networks with significant inequality in the distribution of connections (Price, 1965, 1976). For each model, we also consider three levels of connection density: we choose an average number of links per individuals of 1 (low density), 1.5 (medium density) and 2 (high density). We run 20 independent simulations for each of the 6 types of initial networks (initial networks generated according to the above stochastic models differ across all simulations). Every simulation converged to a stable outcome.

Table 4 presents the frequency of limit networks and the dynamics of best response dynamics. Table 4 tells us that in all simulations with average outdegree larger than 1, the dynamics converge to the cycle. In the case of outdegree 1, dynamics converge to the cycle in 85% of the cases with Erdos-Renyi initial networks and they converge to the cycle in 95% of the cases with Price initial networks. In only one case do they converge to the network that is neither a cycle nor the empty network across all these simulations. This occurrence is found for the case $N = 10$, when starting from an egalitarian network (Erdos-Renyi model) with low connectivity (mean outdegree of 1). In this case, the network consists of two joined cycles, and therefore share similar statistics as the cycle network (therefore this case is not excluded from the time series analysis provided in Figure 15).

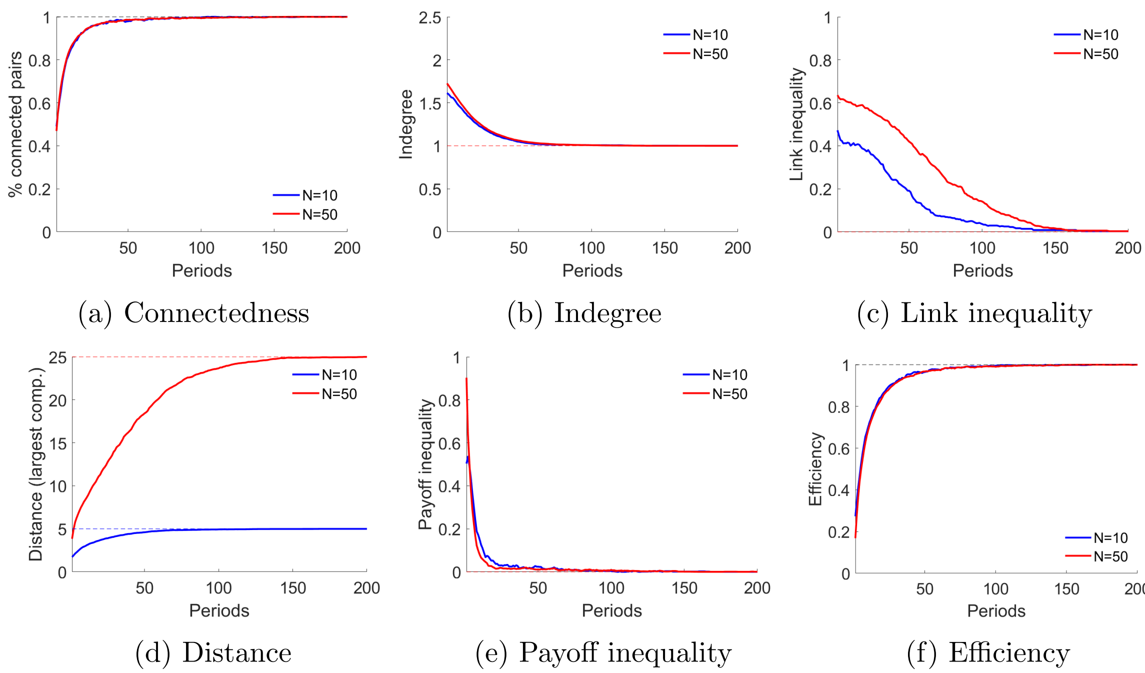
Figure 15 presents the simulation dynamics: we see that convergence is rapid – by period 50 the network is connected and attains full efficiency. There is a small group size effect on rates of convergence with larger groups being slightly slower to converge, especially on the dimensions of distance.

Initial network			Empty		Cycle	
Structure	Outdegree	#	$N = 10$	$N = 50$	$N = 10$	$N = 50$
	1	20	0%	15%	95%	85%
Erdos-Renyi	1.5	20	0%	0%	100%	100%
	2	20	0%	0%	100%	100%
	1	20	0%	5%	100%	95%
Price	1.5	20	0%	0%	100%	100%
	2	20	0%	0%	100%	100%

Table 4: Frequency of limit networks in computer simulations under the one-way flow model (myopic best response dynamics)

Table 5 presents the frequency of limit networks observed for each type of initial network fed into the simulations. It shows that if outdegree is larger than 1 then dynamics converge to the star network is all cases. If outdegree is equal to 1, the dynamics, starting from Erdos-Renyi initial network, converge to the star network in all cases for $N = 10$ and in 95% of the cases for $N = 50$. If outdegree is equal to 1, the dynamics, starting from Price initial network, converge to the star network in all cases for $N = 10$ and for $N = 50$.

Figure 16 describes the dynamics followed by the simulations that converge to the star network in the two-way flow model. We see that convergence is rapid – by period 50 the network is connected and attains close to full efficiency. There is difference in the dynamics with respect to distance but otherwise the group size effects are modest.

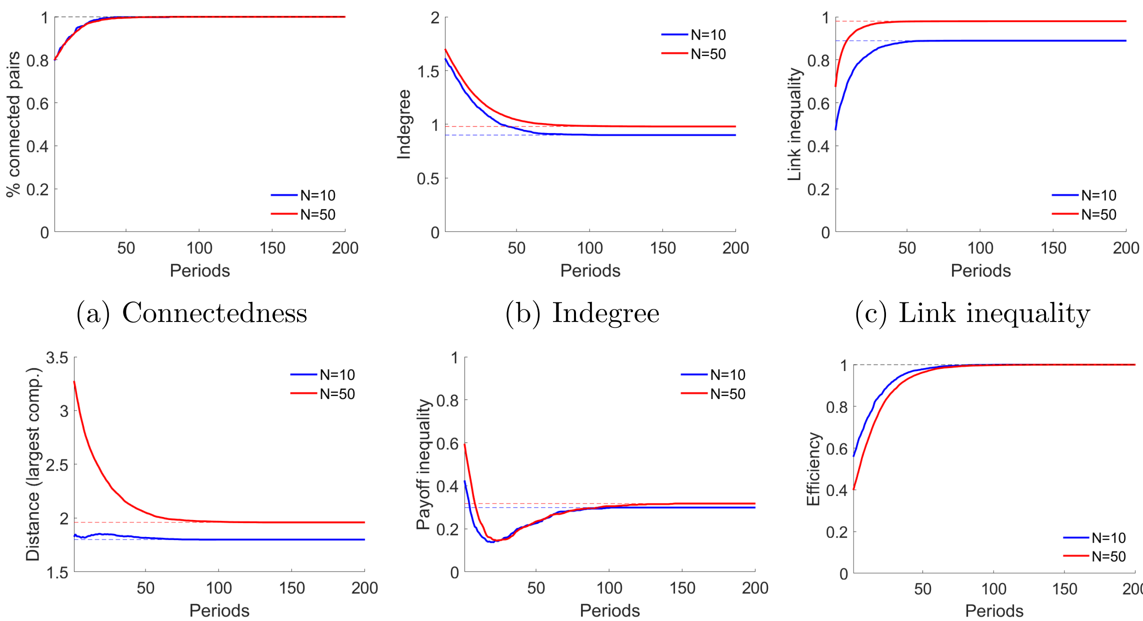


(a) Connectedness (b) Indegree (c) Link inequality

(d) Distance (e) Payoff inequality (f) Efficiency

Figure 15: Time series of network statistics across simulations of myopic best response dynamics converging to the cycle network in the one-way flow model (119 obs. for $N = 10$, 116 obs. for $N = 50$)

Notes: for group size comparisons, each period corresponds to one simulation iteration for N=10, and to 5 subsequent simulation iterations for N=50 (average values across the 5 iterations are then reported). So each individual has 10% chance of making a choice at any given period, in both treatments.



(a) Connectedness (b) Indegree (c) Link inequality

(d) Distance (e) Payoff inequality (f) Efficiency

Figure 16: Time series of network statistics across simulations of myopic best response dynamics converging the star network in the two-way flow model (120 obs. for $N = 10$, 119 obs. for $N = 50$)

Notes: for group size comparisons, each period corresponds to one simulation iteration for N=10, and to 5 subsequent simulation iterations for N=50 (average values across the 5 iterations are then reported). So each individual has 10% chance of making a choice at any given period, in both treatments.

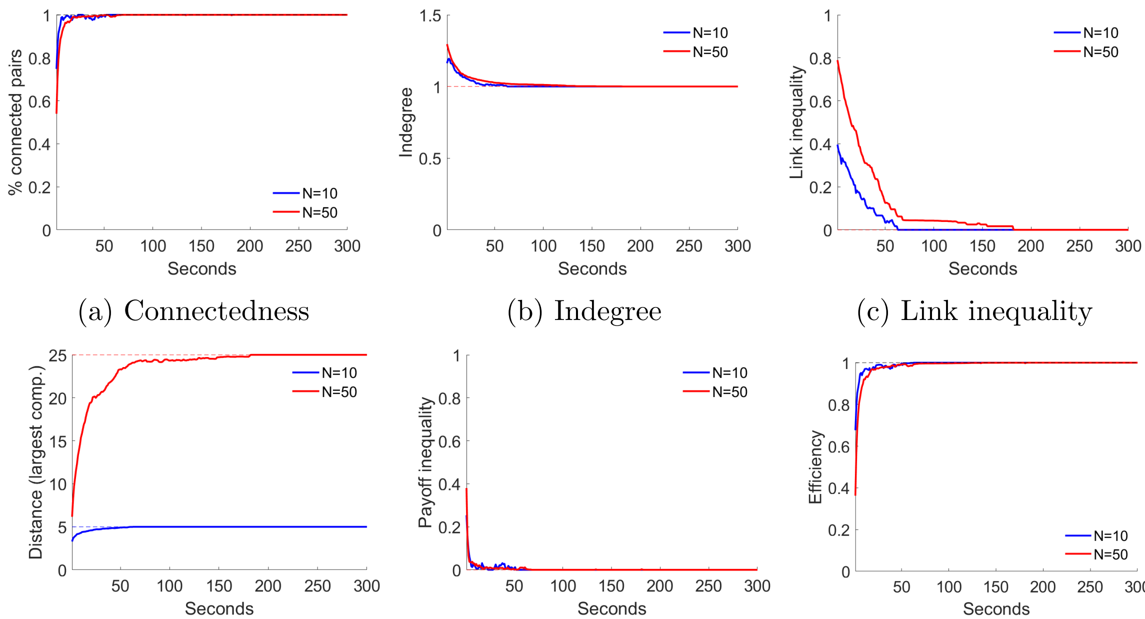
Initial network			Empty		Star	
Structure	Outdegree	#	$N = 10$	$N = 50$	$N = 10$	$N = 50$
Erdos-Renyi	1	20	0%	5%	100%	95%
	2	20	0%	0%	100%	100%
	3	20	0%	0%	100%	100%
Price	1	20	0%	0%	100%	100%
	2	20	0%	0%	100%	100%
	3	20	0%	0%	100%	100%

Table 5: Frequency of limit networks in computer simulations under the two-way flow model (myopic best response dynamics)

A.2 Predictions in the experiment

Simulations from the previous sections made general assumptions on the initial networks, which can affect the selection of the limit network through best response dynamics. We now consider the specific initial networks that have been selected by subjects in our experiment at the end of the trial (non-payoff effective) period of each round (second 60).

Descriptive statistics about those initial networks are presented in Tables 6 and 7 for the one-way and two-way flow models, respectively. Figures 17 and 18 present the time series of simulation dynamics based on best response dynamics when starting from those initial networks across both models. We observe that those simulations always converge to the cycle under one-way flow model and the star under two-way flow model, respectively. Moreover, the simulations show that convergence occurs within 100 seconds in this case (assuming each player makes on average one move every 5 second, as observed in the experiment). This suggests that the actual duration in our experiment (300 second payoff-effective period in each round) was sufficient to converge to the limit networks.

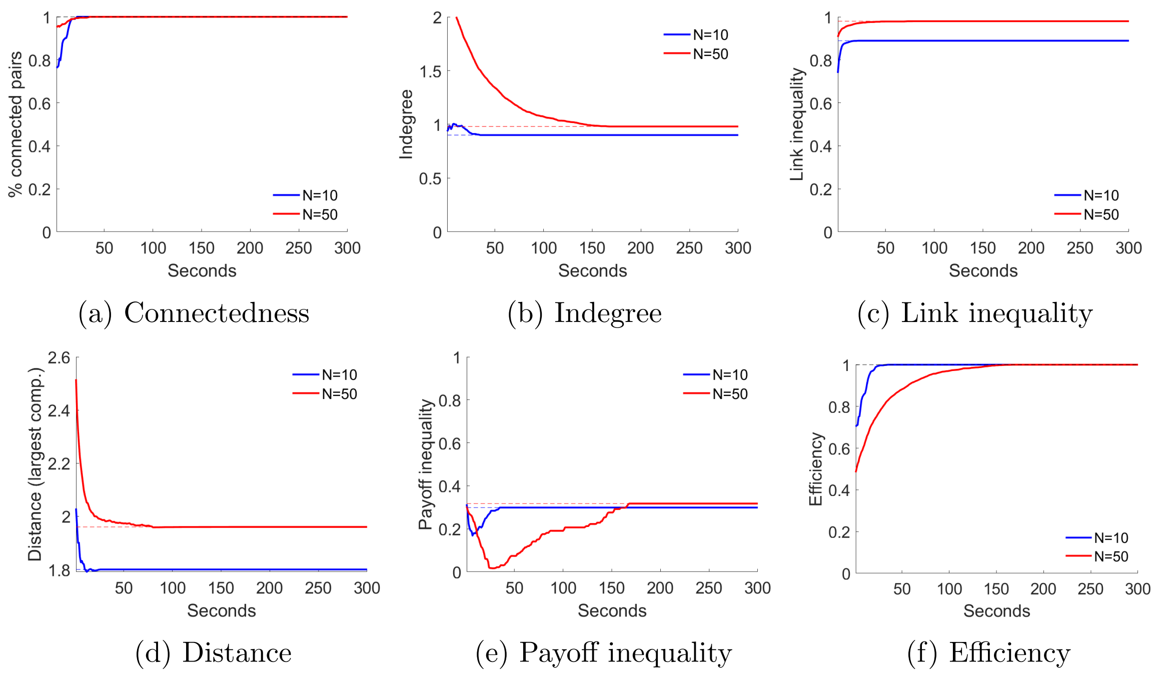


(a) Connectedness (b) Indegree (c) Link inequality

(d) Distance (e) Payoff inequality (f) Efficiency

Figure 17: Simulations of myopic best response dynamics in the one-way flow model, starting from experimental initial networks (20 obs. for $N = 10$ and $N = 50$)

Notes: Initial networks correspond to observations at second 60 of each round of the experiment. For consistency with the experimental design, each simulation runs for 300 seconds, and every agent has a 5% probability of making a choice at any given second across both treatments (as observed in the experiment).



(a) Connectedness (b) Indegree (c) Link inequality

(d) Distance (e) Payoff inequality (f) Efficiency

Figure 18: Simulations of myopic best response dynamics in the two-way flow model, starting from experimental initial networks (20 obs. for $N = 10$ and $N = 50$)

Notes: Initial networks correspond to observations at second 60 of each round of the experiment. For consistency with the experimental design, each simulation runs for 300 seconds, and every agent has a 5% probability of making a choice at any given second across both treatments (as observed in the experiment).

	Initial Networks	
	$N = 10$	$N = 50$
Connectedness	0.72	0.41
Average indegree	1.17	1.3
Efficiency	0.64	0.2
Link inequality	0.41	0.8
Distance (largest component)	3.18	4.26
Payoff inequality	0.26	0.72
Distance to cycle	0.58	1.03

Table 6: Properties of initial networks observed in the experiment under the one-way flow model (end of trial period; 20 obs.)

	Initial Networks	
	$N = 10$	$N = 50$
Connectedness	0.24	0.44
Average indegree	0.93	2.2
Efficiency	0.16	0.02
Link inequality	0.72	0.9
Distance (largest component)	1.71	2.67
Payoff inequality	0.57	0.8
Distance to cycle	1.09	2.21

Table 7: Properties of initial networks observed in the experiment under the two-way flow model (end of trial period; 20 obs.)

B Additional Results

B.1 One-way flow model

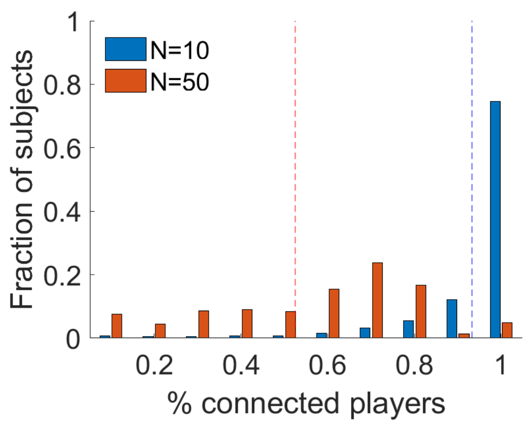
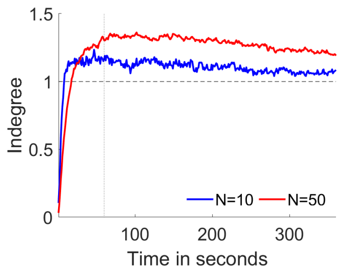
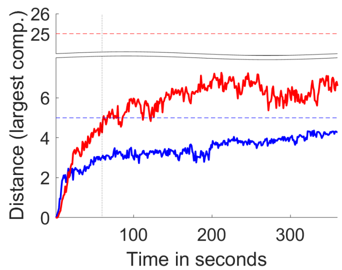


Figure 19: Average distribution of connectedness

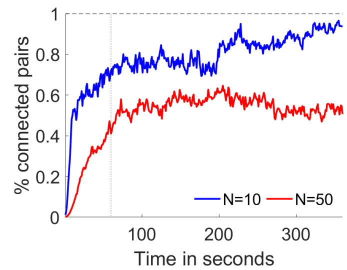
Notes: In this figure, each observation corresponds to a distribution computed at a moment of the last 30 seconds for a given group (snapshot), in a given round (600 observations overall); vertical dashed lines represent average connectedness (0.93 in $N = 10$ and 0.52 in $N = 50$).



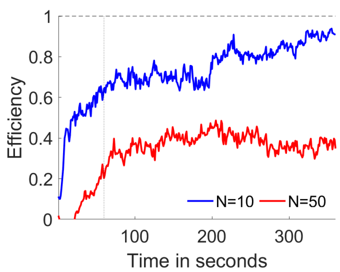
(a) Indegree



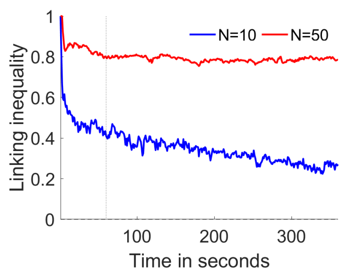
(b) Distance



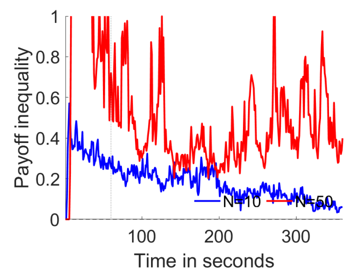
(c) Connectedness



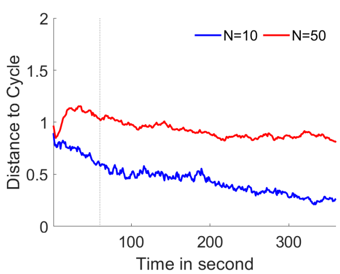
(d) Efficiency



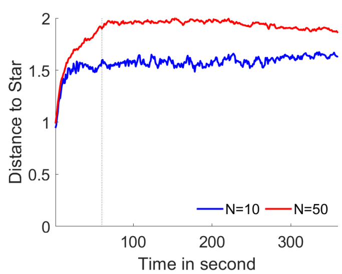
(e) Link inequality



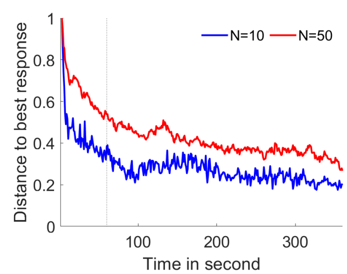
(f) Payoff inequality



(g) Distance to Cycle



(h) Distance to Star

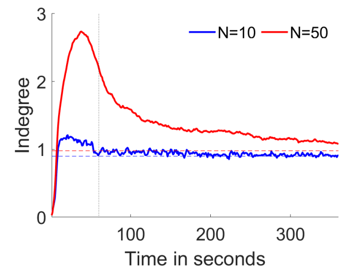


(i) Distance to best response

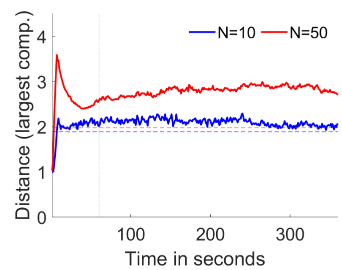
Figure 20: Treatment effects and dynamic behavior in one-way flow model

Notes: Horizontal dashed lines represent equilibrium predictions. Vertical lines represent the end of the trial period and the beginning of the payoff effective period.

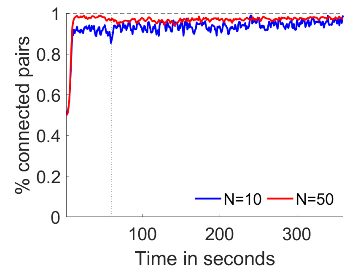
B.2 Two-way flow model



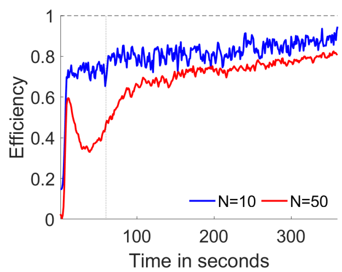
(a) Indegree



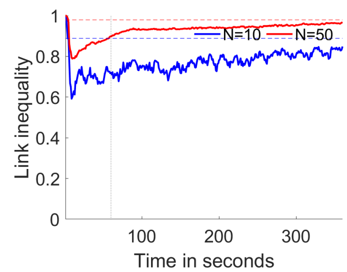
(b) Distance



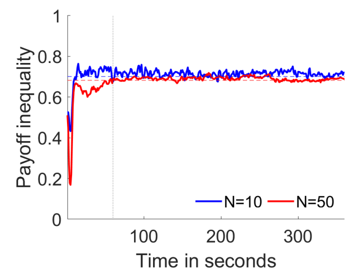
(c) Connectedness



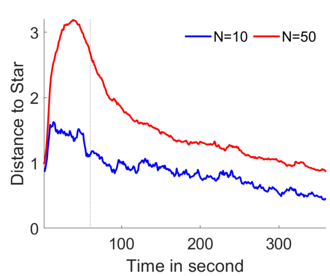
(d) Efficiency



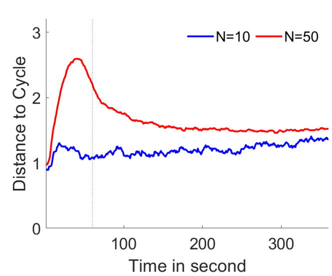
(e) Link inequality



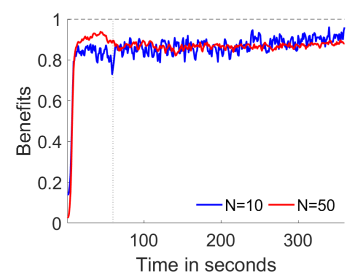
(f) Payoff inequality



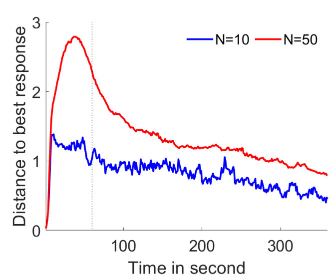
(g) Distance to Star



(h) Distance to Cycle



(i) Benefits



(j) Distance to best response

Figure 21: Treatment effects and dynamic behavior in two-way flow model

Notes: Horizontal dashed lines represent equilibrium. Vertical lines separate trial period from payoff effective period.

C Estimation of the Behavioral Parameters

The network dynamics $\tilde{G}_t(\epsilon)$ depends stochastically on the two noisy parameters ϵ_1 (the probability of making a uniformly random choice) and ϵ_2 (the probability of targeting a highest-degree individual). To estimate these parameters, we minimize the discrepancy between the simulated network statistics and the actual network statistics observed in the last 30 seconds. We use the generalized method of moments (GMM) where three statistics are considered: average indegree, connectedness, and degree inequality. The moment condition is

$$m(\epsilon_1, \epsilon_2) = E[\overline{X}(G) - \overline{X}(\tilde{G}(\epsilon_1, \epsilon_2))] = 0$$

where $\overline{X}(\tilde{G}(\epsilon_1, \epsilon_2)$ denotes the statistics of network $\tilde{G}$ under parameters ϵ_1 and ϵ_2, and $\overline{X}(G)$ denotes the actual statistics. In the estimation, $\overline{X}(\tilde{G}(\epsilon_1, \epsilon_2)$ is computed as the average statistics in the last 15 periods (equivalent to the last 30 seconds) over 128 simulations, while $\overline{X}(G)$ is the average statistics in the last 30 seconds of the experiment encompassing all groups and all rounds except the first round. The objective function is as follows:

$$(\epsilon_1, \epsilon_2) = argmin\ \hat{m}(\epsilon_1, \epsilon_2)^T W \hat{m}(\epsilon_1, \epsilon_2) \tag{3}$$

We use a two-step estimation strategy where in the first step we use an identity matrix as the weight matrix W and update it according to $W_{(1)} = \left[\hat{m}(\epsilon_1, \epsilon_2)\hat{m}(\epsilon_1, \epsilon_2)^T\right]^{-1}$. In the second stage, we use the updated $W_{(1)}$ to estimate the parameters in (3).

We use the Python library Constrained Optimization By Linear Approximation (COBYLA) optimizer, which is suitable for the case where the objective function is derivative-free and there are constraints in the parameters.

To compute the standard error of the estimates, we replicate the aforementioned estimation procedure 10 times. This stochasticity of the estimates arise from the randomness in simulations. The mean of the 10 outcomes is used as the estimate, and the standard deviation of them serves as the standard error.

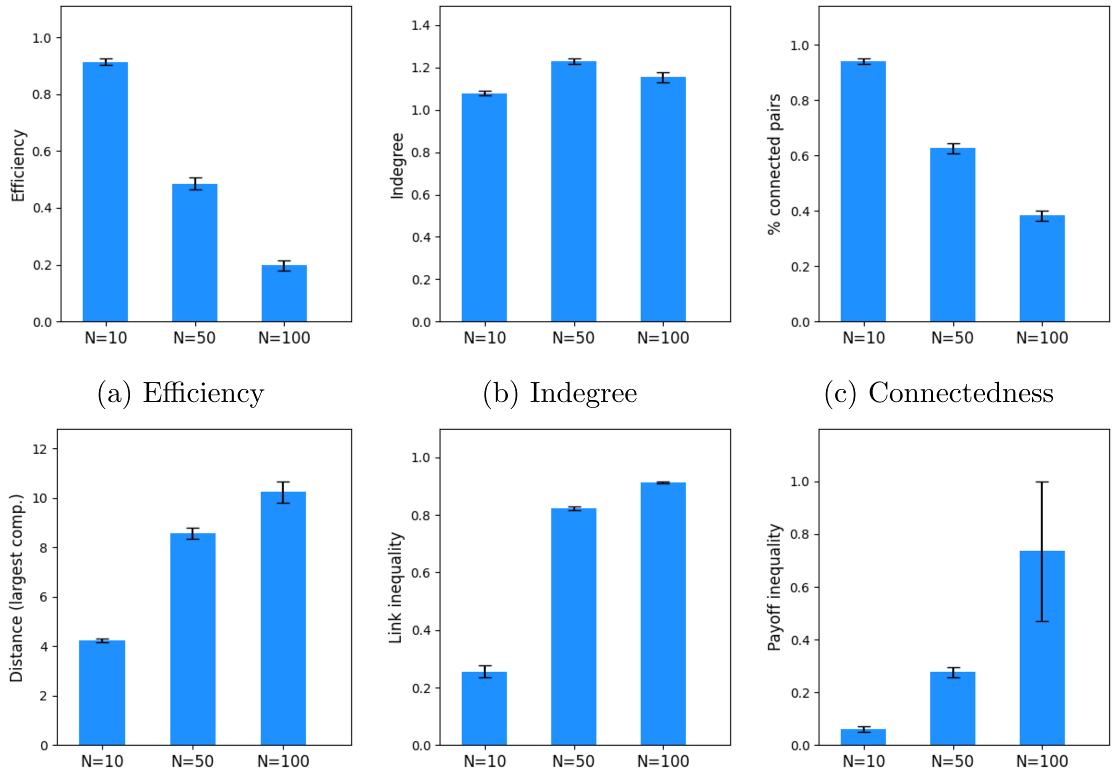
D Extrapolations

We simulate the dynamics of link formation with $N = 100$ players using the behavioral model described in Section 4 to investigate whether the main results of the paper can be generalized to larger group sizes in both the one-way and two-way flow models. The parameters for the game under $N = 100$ follow the main experiments. That is, the cost of linking is $k = 2N$, and $\delta = 1$ for the one-way model and $\delta = 0.9$ for the two-way model.

The simulation procedures for $N = 100$ follow those described in Section 4, with the exception that we assume players make uniformly random choices in the first minute to begin with a relatively dense network.[15] That is, when a player is picked to make a decision, she makes a choice uniformly at random from the whole choice set (which includes all applicable actions involving adding, deleting, or switching a link). As in Section 4, each player has a probability of 0.1 to make a decision in any given second. We use the estimated parameter values for $N = 50$ to simulate the network dynamics for $N = 100$. The overall network patterns for $N = 100$ are robust to reasonable variations in the values of the behavioral parameters. Two hundred simulations are conducted for each model.

Figures 22 and 23 present the results of the last 30 seconds of simulated dynamics for $N = 100$, alongside the results for $N = 10$ and $N = 50$ as presented in Section 4, for each of the six statistics: efficiency, indegree, connectedness, distance, link inequality, and payoff inequality. The simulation results indicate that in the one-way flow model, the deviation from equilibrium increases with group sizes, as evidenced by decreasing efficiency and connectedness and increasing link inequality. In contrast, in the two-way flow model, group sizes have modest impacts on network structures, as demonstrated by, for example, high efficiency and high connectedness.

[15] Recall that in the experiment, the first minute is payoff-irrelevant. The network patterns in the last 30 seconds of the simulated game are robust to whether we start with uniformly random choices in the first minute or begin with the actual network from the experiments for $N = 10, 50$. Those patterns for the case of $N = 100$ are robust to reasonable variations of players' active rate in the first minute (i.e., how frequently players make a change in the first minute).



(a) Efficiency (b) Indegree (c) Connectedness

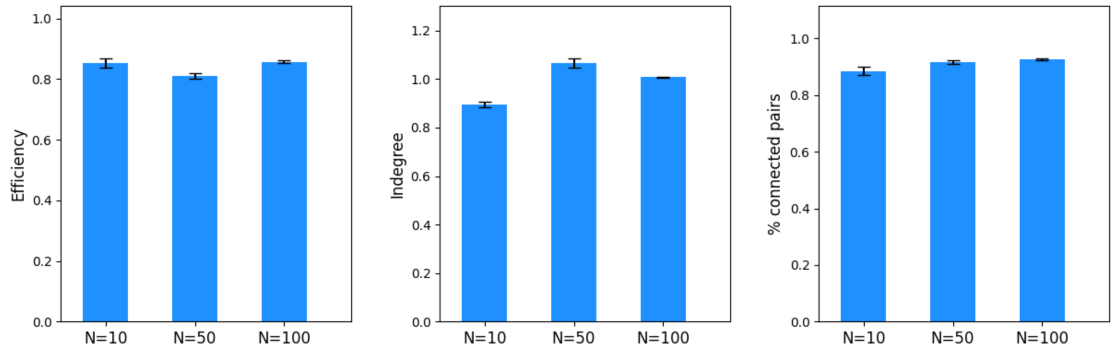
(d) Distance (e) Link inequality (f) Payoff inequality

Figure 22: Simulation results for the one-way flow model

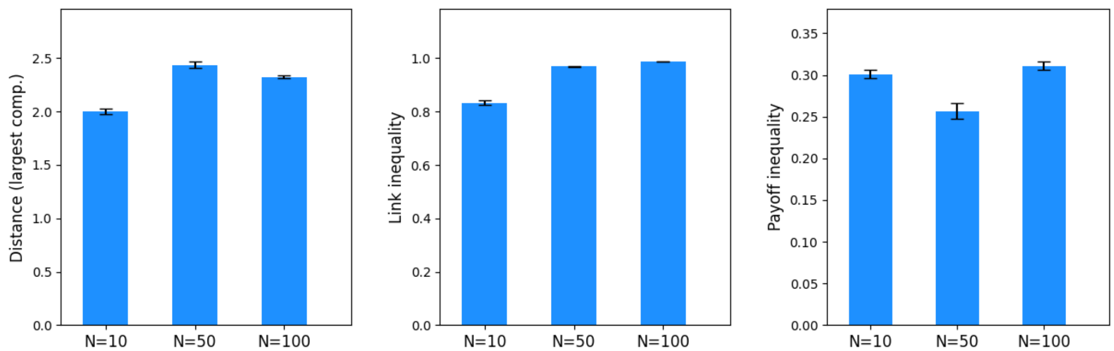
Notes: Error bars display standard 95% confidence interval around the mean.

E Additional Details of Experimental Procedures

A subject participates in only one of the experimental sessions. Subjects read the instructions, which were also read aloud by an experimenter to guarantee that they all received the same information. While reading the instructions, the subjects were provided with a step by step interactive tutorial which allowed them to get familiarized with the experimental software and the game. Subjects interacted through computer terminals and the experimental software was programmed using HTML, PHP, Javascript, and SQL. This procedure is common to all the treatments. Sample instructions and interactive tutorials for all the experiments are shown in section F of the Appendix .



(a) Efficiency (b) Indegree (c) Connectedness



(d) Distance (e) Link inequality (f) Payoff inequality

Figure 23: Simulation results for the one-way flow model

Notes: Error bars display standard 95% confidence interval around the mean.

At the beginning of each experiment, each subject was endowed with an initial balance of 50 points in the $N = 10$ treatment, and 250 points in the $N = 50$ treatment. Subjects' total earnings in the experiment were equal to the sum of earnings across the last 5 rounds and the initial endowment.[16] Earnings were calculated in terms of experimental points and then exchanged into euros at the rates of 40 points being equal to 1 euro for the $N = 10$ treatments, and 200 points being equal to 1 euro for the $N = 50$ treatments.[17] On average, a session lasted 90 minutes.

[16]In case of negative total earnings, the corresponding subject would simply earn 0 point from the game.

[17]The different conversion rates and initial endowments are justified by the different linking costs across different treatments, as an attempt to maintain similar earnings.

F Experimental Instructions and Tutorials

[In the following instructions, N is to be replaced with a value from $\{10, 50\}$, R with a value from $\{40, 200\}$, C with a value from $\{20, 100\}$, and E with a value from $\{50, 250\}$, depending on the treatment]

[All experiments]

Please read the following instructions carefully. **These instructions are the same for all the participants**. The instructions state everything you need to know in order to participate in the experiment. If you have any questions, please raise your hand. An experimenter will answer your question.

You can earn money by earning points during the experiment. The number of points that you earn depends on your own choices and the choices of other participants. At the end of the experiment, the total number of points that you have earned will be exchanged at the following rate:

$$R \textbf{ points} = \textbf{1 Euro}$$

The money you earn will be paid out in cash at the end of the experiment. The other participants will not see how much you earned.

Details of the experiment

The experiment consists of 6 (six) independent rounds of the same form. The first round is for practice and does not count for your payment. The next 5 rounds will be counted for your payment. At the beginning of each round, you will be grouped with $N-1$ other participants. This group will remain fixed throughout the 6 rounds. Each of the other participants will be randomly assigned an identification number of the form "Px" where x is a number between 1 and $N-1$. Those numbers will be randomly changed across every round of the experiment. The actual identity of the participants will not be revealed to you during or after the experiment. The participants will always be represented as blue circles on the decision screen. You are always represented as a yellow circle identified as "ME".

Each round will last for **6 (six) mins: the first minute will be a trial period, only the subsequent 5 minutes will be relevant for the earnings**. Your earnings in a given round will be based on everyone's choice **at a randomly selected moment in the last 5 minutes of the round**. In other words, any decision made before or after that randomly chosen moment will not be used to determine points that you earn in the round. This precise moment will be announced to everyone only at the end of the round, along with your choice and that of others connected with you in the network at that moment.

At the beginning of the experiment, you are given **an initial balance of E points**. Your final earnings at the end of the experiment will consist of the sum of points you earn across the 5 last rounds plus this initial balance (the first round will be used to familiarize yourself with the game and will have no influence on your earnings). Note that if your final earnings (i.e., the sum of your earnings across the 5 last rounds plus the initial balance) go below 0, your final earnings will be simply treated as 0.

In each round, every participant will be allowed to form links with other participants or delete links that were previously created by him- or herself at any moment during the 6 minutes. You are linked with another person if you form a link with that person or that person forms a link with you (or both). Each link you form costs you **C points**. You do not pay any cost for links formed by others. In order to form or delete a link with a participant, you will simply need to double-click the corresponding node on the computer screen. A network resulted from your choice and choices of other participants at any moment will be updated in your computer screen in real time.

The participants that you are linked with (regardless of whether you or they form the links) are called your neighbours. You are said to be **connected** with another participant when there exists a sequence of links connecting you with that person in the network.

[Two-way flow model experiment only]

The computer screen will be split into two parts:

- **<u>The left side of the screen presents you and participants that you are connected with</u>**.
- **<u>The left side of the screen presents you and participants that you are</u>**

connected with.

[One-way flow model experiment only]

The participants that you form a link with are called your neighbours. You are said to **access another participant either if you are linked to this person or if there is a path – a sequence of directed links – leading from you to this person in the network**.

The computer screen presents all the participants and links created by them. You and every participant that you can access are coloured in blue**. All the other players are coloured in** grey.

[All experiments]

Each node is described by their identification number "Px". Identification numbers "Px" are randomly assigned in every round. Therefore, every player is likely to have a different ID in different rounds.

At the very beginning of the round when no link is formed, you will be on the left side of the screen and all the other participants will be shown on the right side of the screen.

You may revise your choices at any moment before the round ends. During a round, you will also be informed about every other participant's most recent decision (formed links), which will be updated every 2 seconds or whenever you change your own choice.

Earnings

Your earnings at any moment of the round are determined by **the benefits that you obtain** *minus* **the costs that you incur** from the network at that moment.

The costs that you incur from the network are equal to C points times the number of links created by you.

The benefits that you obtain from the network are equal to the sum of benefits you receive from each of the other participants to whom you are connected, plus 10 points.

[Two-way flow model experiment only]

The benefit you receive from each participant depends on the distance between you and that participant. This distance is defined by the smallest number of links that connect you with the participant in the network. For example, the distance between you and each of your neighbours is 1. The distance between you and each neighbour of your neighbours (who is not your neighbour) is 2.

You receive a benefit of 10×0.9^d points from a participant who is connected with you from a distance d. Given this form of benefit, for example, you receive a benefit of $10 \times 0.9^1 = 9$ points for each of your neighbours, and $0 \times 0.9^2 = 8.1$ points for each neighbour of your neighbours that is not your neighbour. You receive 0 point from each participant whom you are not connected with.

[One-way flow model experiment only]

By yourself you receive **10 points**. In addition, you receive a benefit of **10 points** from every participant you can access. You receive 0 point from each participant whom you cannot access. The total benefits you obtain in a network therefore correspond to 10 times the number of participants that you can access, plus 10 points (that correspond to the value you receive on your own). Note that the benefit you receive from a participant does not depend on the distance between you and that participant in the network.

[All experiments]

One moment in the last 5 minutes of the round will be randomly chosen to determine every participant's real earnings in the round.

Tutorial

[Two-way flow model experiment only]

Please follow this simple tutorial simulating a simple virtual scenario on the computer screen. In this tutorial you are interacting with 9 other players, and every link you form costs you 20 points. **Note** that this setting is only illustrative and slightly differs from the real game described above (you will then interact with $N-1$ other players and every link will cost you C points). In the initial state, you are not linked with anyone: you start with 10 points.

- Initially, the nodes on the right side of the screen represent all other players (in this simulation, those players are not real people). You may choose to form a link with any player by simply double clicking on the corresponding node. For example, forming a link with P4 reveals that each of P2 and P3 forms one link with P4. Forming a link with P4 costs you 20 points (in red on the screen), but it also generates a benefit of 35.2 points (10 + 9 (= 10×0.9^1 from P4) + 8.1 (= 10×0.9^2 from P2) + 8.1 (= 10×0.9^2 from P3). Your resulting earnings are **15.2 points (= 35.2 points - 1 link × 20 points)**.

- After forming a link with P4, you observe that some nodes remain not connected with you (P1, P5, P6, P7, P8, and P9 on the right side). However, forming an additional link with P9 (by double clicking on the corresponding node) reveals that all those nodes were connected with one another and that you are now connected with every participant. You were not allowed to observe them before because they were not linked with any node you were connected to. You can now observe them because there exists a sequence of links connecting you from any of them (for example, P5 is connected to you via P9). Remember that you can only see players that are connected with you. Your resulting earnings become **44.7 points (= 84.7 points - 2 links × 20 points)**.
- Alternatively, you may choose to remove a link that you previously formed by double clicking on the corresponding node. For example, after forming links with P4 and P9, removing the link with P4 makes players P2, P3, and P4 move to the right side of the screen, as they are not connected with you anymore.
- You may also shape the visual structure of the network by dragging nodes as it pleases you.

[One-way flow model experiment only]

Please follow this tutorial that simulates a scenario on the computer screen. In this tutorial you are interacting with 9 other players, and every link you form costs you 20 points. **Note** that this setting is only illustrative and slightly differs from the real game described above (you will then interact with 49 other players and every link will cost you 100 points). In the initial state, you are not linked with anyone, and there are 3 players that can access you (P4 is linked with you, and both P2 and P3 are linked with P4): you start with 10 points. In this simulation, players are not real people.

- You may choose to form a link with any player by simply double clicking on the corresponding node. For example, by forming a link with P9, you gain access to P1, P6, P7, and P9. Forming a link with P9 costs you 20 points (in red on the screen), but it also yields a total benefit of 40 points (10 from P1 + 10 from P6, 10 from P7 + 10 from P9). Your resulting earnings are **30 points (= 10(own) + 10(P1) + 10(P6) + 10(P7) + 10(P9) - 1 link × 20 points)**.

- After forming a link with P9, forming an additional link with P5 or P8 (by double clicking on the corresponding node) you update your total earnings to **20 points (= 10(own) + 10(P1) + 10(P6) + 10(P7) + 10(P9) + 10(P5) - 2 links × 20 points)**.

- You may also choose to remove a link that you previously formed by double clicking on the corresponding node. For example, after forming links with P9 and P5, removing the link with P9 does not change the benefits you earn from others since your link with P5 already allows you to indirectly access P9 and P9's connections. In other words, your previous link with P9 was redundant, and removing it increases your total earnings to **40 points (= 10(own) + 10(P1) + 10(P6) + 10(P7) + 10(P9) + 10(P5) - 1 link × 20 points)**.

- Note that you do not benefit from players who can access you but who you cannot access. This is the case of players P2, P3, and P4 here. Forming a link with P2 will make you access both P2 and P4, whose node will then become blue. In this case, your total earnings will however remain the same (**40 points**) as you will gain 20 extra points (=10(P2) + 10(P4)) but pay 20 points for the additional link.

- You may also shape the visual structure of the network by dragging nodes as it pleases you.

[All experiments]

Summary

Here is a brief description of information available on the decision screen:

1. The timer indicates elapsed time since the beginning of the round. Any round lasts **6 minutes**. A moment will be randomly selected **in the last 5 minutes**

to determine everyone's payoff. The time displayed will turn red when entering this interval.

2. **Only decisions made at the randomly selected moment in the round** matter to directly determine the earnings. The payoff may be negative at the end of a round. However, starting from a balance of E points, any negative total of points at the end of the 5 rounds will be equivalent to 0 point.

   [Two-way flow model experiment only]

3. A participant is connected with you if there exists a sequence of links connecting you to that person in the network.

   [One-way flow model experiment only]

4. You can access a participant if there exists a sequence of directed links that go from you to that person in the network. The corresponding node is coloured in **blue**. Similarly, a participant can access you if there is a sequence of directed links pointing out from this participant to you. The corresponding node is coloured in **grey**.

   [All experiments]

5. For every participant you are connected with, you receive 10×0.9^d points where d represents the smallest number of links that connect you with that person in the network.

6. However, you receive 0 points from every participant you are not connected with. For every link you form, you pay C points.

7. You are represented as the yellow node, and your ID is "ME".

8. Every other node's ID is represented as "Px" (inside the node) where x is a number. Every node has a unique ID, which is randomly reassigned in every round.