Brokerage: an experimental study[*]

Syngjoo Choi[†]   Sanjeev Goyal[‡]   Frédéric Moisan[§]

December 30, 2022

Abstract

We study the determinants of intermediation networks and brokerage rents through an examination of two pricing rules: criticality and betweenness. Under criticality, stable networks involve one or more interconnected cycles; under betweenness, stable networks contain one or more hubs with a disproportionate share of links. Under criticality, distances grow but degree and payoff inequality remain modest as group size grows; under betweenness, distances remain unchanged but degree and payoff inequality explode as group size grows. Our experiment shows that subjects create networks that are in line with the predictions of the theory.

JEL: C92, D83, D85, Z13.

[†]Department of Economics, Seoul National University. Email: syngjooc@snu.ac.kr

[‡]Faculty of Economics and Christ's College, University of Cambridge and NYUAD. Email: sg472@cam.ac.uk

[§]Emlyon Business School & GATE. Email: moisan@em-lyon.com

1 Introduction

Intermediation is a salient feature of the modern economy. In some contexts, as in online retail, there exist dominant intermediaries with direct links to vast numbers of buyers and sellers, while in other cases, as in tea and coffee supply chains and financial instruments, there exist long paths of intermediation (Spulber [1996],BelleFlamme and Peitz [2015],Goyal [2023]). Networks offer us a language to reason about the process of intermediation. Trades between two actors can be realized if they have a direct link or if they are indirectly linked through a chain of intermediaries.[1] As links are costly to maintain, actors would like to economize on the number of links they maintain. Suppose there are n actors arranged in a hub-spoke network, with $n-1$ links. Most of the exchange takes place among spokes and involve the same single intermediary – the hub. On the other hand, consider a cycle network containing all actors: it has n links and involves potentially long chains of intermediation. Moreover, as everyone is symmetric, every actor will act as an equal intermediary and everyone earns an equal payoff.[2] The goal of the paper is to examine the economic circumstances that give rise to hub-spoke networks (that are very unequal) as against networks dominated by cycles (that are much more equal).

Our focus will be on the role of pricing rules that allocate the surplus from bilateral trades between connected pairs of subjects in the network. The first pricing rule – *Criticality pricing* – assumes that a trader earns rents for a trade between two other traders if it is critical for the trade: it lies on all paths of trade between these traders. This reflects situations where network path lengths do not matter. The second pricing rule – *betweenness pricing* – considers a situation in which a trader earns rents from a trade between two other traders only if it lies on the shortest path between them. This leads to payoffs for an actor that are proportional to its betweenness centrality. We borrow the criticality pricing rule from Choi et al. [2017] and Goyal and Vega-Redondo [2007] and the betweenness pricing rule from Kleinberg et al. [2008] and Galeotti and Goyal [2014]. We suppose that individuals play a best response and the dynamics help us identify limit networks. With criticality pricing, the limit networks involve one or more interconnected cycles or a hybrid star-cycle structure; with betweenness pricing, the networks contain one or more hubs (with a disproportionate share of links) Under criticality pricing, distances grow with

---

[1] The links may embody trust relations, reflect communication channels, or simply reflect physical infrastructure (such as train, road or shipping services).

[2] We recall that $n-1$ is the minimum number of links needed to ensure that there is a path between every pair of actors.

1

size but degree and payoff inequality remain modest across group sizes; under betweenness pricing distances remain unchanged but degree and payoff inequality explode as group size grows.

The different limit networks arise due to subtle differences in local incentives for linking under the two pricing rules. In a real world setting, groups are sometime very large and individuals choose linking at different points in time. The individual decision problem is complicated because the attractiveness of links depends on the overall structure of links. As group size grows, these informational and computational requirements become progressively more challenging. So it is quite unclear if subjects will abide by the local incentive pressures that drive the dynamics in our model. This motivates an experimental investigation of the power of local incentives in selecting networks.

The work of Friedman and Aprea [2012], Goyal et al. [2017], and Choi, Goyal, and Moisan [2022] suggests that continuous time experiments offer subjects more opportunities for choice and for learning and that they may offer better prospects for convergence to equilibrium than discrete time experiments. Our experimental platform builds on this insight. Relative to the earlier paper by Choi et al. [2022], the innovation in the present paper is that we allow for two-sided linking protocol that the intermediation models use. This is made possible by a new and transparent visual representation on the status of the linking relationship among subjects at any point in time. The paper considers a design with criticality and betweenness pricing rules, with three group sizes varying from 10 to 50 to 100. This allows us to systematically examine how local incentives matter and how they interact with group size.

Our first major finding is that pricing rules have strong effects on the macroscopic features of the network. In line with theory, under the criticality rule, a network with multiple cycles with long path lengths is observed, and it generates modest payoff inequality across all group sizes. Also, in line with theory, under betweenness pricing, networks with multiple hubs emerge and this leads to extreme payoff inequality in the large groups. Our second finding concerns efficiency: subjects create networks that attain high efficiency under criticality but attain less efficiency under betweenness pricing. This fall in efficiency under betweenness pricing is due to greatly increased link proposals by a few hub subjects: the hubs hope to earn large rents and the other subjects agree to link proposals so as to economize on rents in long intermediation chains. By contrast, under criticality pricing, subjects have no incentive to form links to shorten intermediation paths (once a cycle has been created.)

2

Our paper is a contribution to the study of intermediation. The economics literature studies pricing by intermediaries, their ability to reduce frictions, and to extract surpluses (Goyal [2023]. For an early model, see [Rubinstein and Wolinsky, 1987]; for more recent work on intermediaries in networks see [Condorelli et al., 2016],Choi, Galeotti, and Goyal [2017], and Manea [2018]. Condorelli and Galeotti [2016] provide a survey of this work. For experiments on trading in networks and on intermediation, see Gale and Kariv [2009], Kariv et al. [2018], Charness et al. [2007] and Choi, Galeotti, and Goyal [2017]. A related strand of work examines how market power emerges through the deliberate creation of links in a network formation setting: in addition to Goyal and Vega-Redondo [2007], Kleinberg et al. [2008] and Buskens and Van de Rijt [2008], recent contributions include Farboodi [2014] and Kotowski and Leister [2018]. To the best of our knowledge, the present paper offers the first experimental exploration of how intermediation pricing rules shape incentives to form links and determine the network and the distribution of payoffs.

Inequity aversion is an important theme in the experimental literature([Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000] and [Kosfeld et al., 2009]). In the context of networks, Falk and Kosfeld [2012] and Goeree et al. [2009] argue that inequity aversion explains why experimental subjects do not create the hub-spoke structure predicted by the theory. How can we reconcile our findings with these earlier findings? Recall that in the standard model [Fehr and Schmidt, 1999, Bolton and Ockenfels, 2000], a typical individual is sensitive to an average across payoff differences between themselves and every other person. As a result, while an inequity averse individual may disapprove of the large payoff difference between themselves and the wealthiest person, they would be more tolerant of the average inequality. In other words, aversion to isolated inequalities is diluted in large groups and has limited effects on individual incentives.

The rest of the paper is organized as follows. In Section 2, we describe the pricing models and formulate our hypotheses. Section 3 describes the design of the experiment. Section 4 presents the main experimental findings.

2 Theory

We consider a game with $N = \{1, 2, ..., n\}$ individuals, where $n \geq 3$. Relationships between nodes are conceptualized in terms of binary variables, so that a relationship either exists or does not exist. Denote by $g_{ij} \in \{0, 1\}$ a relationship between two nodes i and j. The variable g_{ij} takes on a value of 1 if there exists a link between i and j and 0, otherwise.

Links are undirected, i.e., $g_{ij} = g_{ji}$. The set of nodes taken along with the links between them defines the network; this network is denoted by g and the collection of all possible networks on n nodes is denoted by $\mathcal{G}$. Given a network g, in case $g_{ij} = 0$, $g + g_{ij}$ adds the link $g_{ij} = 1$, while if $g_{ij} = 1$ in g, then $g + g_{ij} = g$. Similarly, if $g_{ij} = 1$ in g, $g - g_{ij}$ deletes the link g_{ij}. Let $N_i(g) = \{j | g_{ij} = 1\}$ denote the nodes with whom node i has a link; this set will be referred to as the *neighbors* of i. Let $\eta_i(g) = |N_i(g)|$ denote the number of connections/neighbors of node i in network g. Furthermore, let $d(i, j; g)$ denote the geodesic distance between players i and j in network g.

Individuals propose links with others. The strategy of a player i is a vector of link proposals $s_i = [s_{ij}]_{j \in N \setminus \{i\}}$, with $s_{ij} \in \{0, 1\}$ for any $j \in N \setminus \{i\}$. The strategy set of player i is denoted by S_i. A link between agents i and j is formed if both propose a link to each other, i.e., $g_{ij} = s_{ij} s_{ji}$. A strategy profile $s = (s_1, s_2, \ldots, s_n)$ induces an undirected network $g(s)$.[3] The network $g(s) = \{g_{ij}\}_{i,j \in N}$ is a formal description of the pairwise links that exist between the players. There exists a path between i and j in a network g if either $g_{ij} = 1$, or if there is a distinct set of players $i_1, \ldots, i_n$ such that $g_{ii_1} = g_{i_1 i_2} = g_{i_2 i_3} = \ldots = g_{i_n j} = 1$. All players with whom i has a path defines the component of i in g, which is denoted by $C_i(g)$.

Suppose that players are traders who can exchange goods and that this exchange creates a surplus of V. This exchange can be carried out only if these traders have a link or if there is a path between them. There is a fixed (marginal) cost c per individual for every link that is established. On the other hand, any proposal that is not reciprocated carries no cost.

The central issue here is how are potential surpluses allocated between the different parties to the trade. In the case where two traders have a link, it is natural that they split the surplus equally, each earning $\frac{V}{2}$. If they are linked indirectly, then the allocation of the surplus depends on the nature of competition between the intermediary agents. One simple idea is to view these paths as being perfect substitutes. Another possibility is that the paths offer differentiated trading mechanisms. We will develop models of intermediation that build on these ideas.

We start with the notion of paths between traders being perfect substitutes. Formally, a trader i is said to be critical for trader j and k if i lies on every path between j and k in the network. Denote by $T(j, k; g)$ the set of players who are critical for j and k in

---

[3]With a slight abuse of notation, for simplicity, we will write g instead of $g(s)$.

network g and let $t(j, k; g) = |T(j, k; g)|$ denote the number of critical players between j and k. Following Choi, Galeotti, and Goyal [2017] and Goyal and Vega-Redondo [2007], for every strategy profile $s = (s_1, s_2, \ldots, s_n)$ the net payoffs to player i are given by:

$$\Pi_i^{crit}(s) = \underbrace{\sum_{j \in C_i(g)} \frac{V}{t(i, j; g) + 2}}_{\text{Access Benefits}} + \underbrace{\sum_{j,k \in N \setminus \{i\}} V \frac{I_{i \in T(j,k;g)}}{t(j, k; g) + 2}}_{\text{Brokerage Rents}} - \eta_i(g)k \tag{1}$$

where $I_{i \in T(j,k)} \in \{0, 1\}$ stands for the indicator function specifying whether i is critical for j and k. We shall refer to it as the model of criticality-based pricing, which relies on two distinct sources of benefits: access benefits and brokerage rents.

One feature of this payoff model is slightly implausible: trade can take place along arbitrarily long paths even when shorter paths are available, and that too without any costs. However, if costs are similar, it is easy to see that a small friction arising out of distance would make a large difference in the flow of trade. To take this factor into account, we consider a simple alternative model of intermediation that emphasizes the role of shortest paths between traders.

Let $n_{jk} = (d(j, k; g) - 1)$ denote the number of intermediaries on a shortest path between j and k in network g. Trade surplus between j and k is equally distributed among the source and destination j and k, and among the intermediaries on the shortest path. In the case of multiple shortest paths, one of them is randomly chosen. Therefore, the (ex-ante) expected return for any trader i is in proportion to the shortest paths between j and k that i lies on. Formally, we write $b_{jk}^i(g) \in [0, 1]$ to denote *betweenness* of player i between j and k.[4] Given a strategy profile $s = (s_1, s_2, \ldots, s_n)$, the net payoffs to player i are given by:

$$\Pi_i^{btwn}(s) = \underbrace{\sum_{j \in C_i(g)} \frac{V}{n_{ij} + 2}}_{\text{Access Benefits}} + \underbrace{\sum_{j,k \in N \setminus \{i\}} V \frac{b_{jk}^i}{n_{jk} + 2}}_{\text{Brokerage Rents}} - \eta_i(g)k \tag{2}$$

This is the model with betweenness pricing; we borrow this pricing rule from Kleinberg, Suri, Tardos, and Wexler [2008].

Following Jackson and Wolinsky [1996], a network is said to be pairwise stable if no one can benefit by removing any existing link, and no pair can mutually benefit by adding

---

[4]Formally, $b_{jk}^i(g) = \frac{\text{\# shortest paths between j and k on which i lies}}{\text{\# shortest paths between j and k}}$.

a non existing link with each other. Online Appendix A provides a formal definition of this concept together with a description of pairwise stable networks under both pricing rules (criticality and betweenness).

Figure 1 presents examples of pairwise stable networks. The efficiency of an outcome is measured as the ratio between the sum of individual payoffs and the maximum sum of individual payoffs that can be achieved:

$$E(s) = \frac{\sum_i \Pi_i(s)}{\max_{s'} \sum_i \Pi_i(s')} \tag{3}$$

It follows that $E(s) \leq 1$. A network is said to be socially efficient if it maximizes social welfare, i.e., $E(s) = 1$.

In our model, the intermediation rents cancel out; there is therefore no benefit in having a cycle in a network, as that merely adds to the costs, without any gain in additional trades being realized. So, every component in an efficient network must be minimally connected or a singleton. From standard considerations we know that a (non-empty) network with multiple components cannot be efficient (Goyal and Vega-Redondo [2007]). The total payoffs in a minimally connected network are $\frac{Vn(n-1)}{2} - 2(n-1)c$ and they are equal to 0 in the case of an empty network. So it follows that an efficient network is minimally connected if $c < \frac{Vn}{4}$, and empty otherwise. A prominent example of minimally connected network is the star network (see Figure 1(b)); as it contains the same number of links $(n-1)$, a minimal connected network with multiple hubs is also efficient. We note that a cycle network contains n links and is connected, so it is close to being efficient for large n, i.e., $E(cycle)$ is close to 1 for large n.

We define payoff inequality as the ratio of the highest payoff to the median payoff.

$$I(s) = \frac{\max_i(\Pi_i(s))}{med_i(\Pi_i(s))} \tag{4}$$

It follows that $I(s) \geq 1$. The outcome is equal if $I(s) = 1$. Observe that the outcome is equal in the empty and the cycle network. By contrast, in the star network, under both the criticality and betweenness pricing, the hub and spoke earn, respectively:

$$V(n-1)\left[\frac{1}{2} + \frac{n-2}{6}\right] - (n-1)c \qquad V\left[\frac{1}{2} + \frac{n-2}{3}\right] - c \tag{5}$$

The ratio of hub payoffs to the median payoff too is unbounded (the median payoff

<div align="center">(a) Cycle   (b) Star</div>

<div align="center">(c) Cycle-star   (d) Multiple hubs</div>
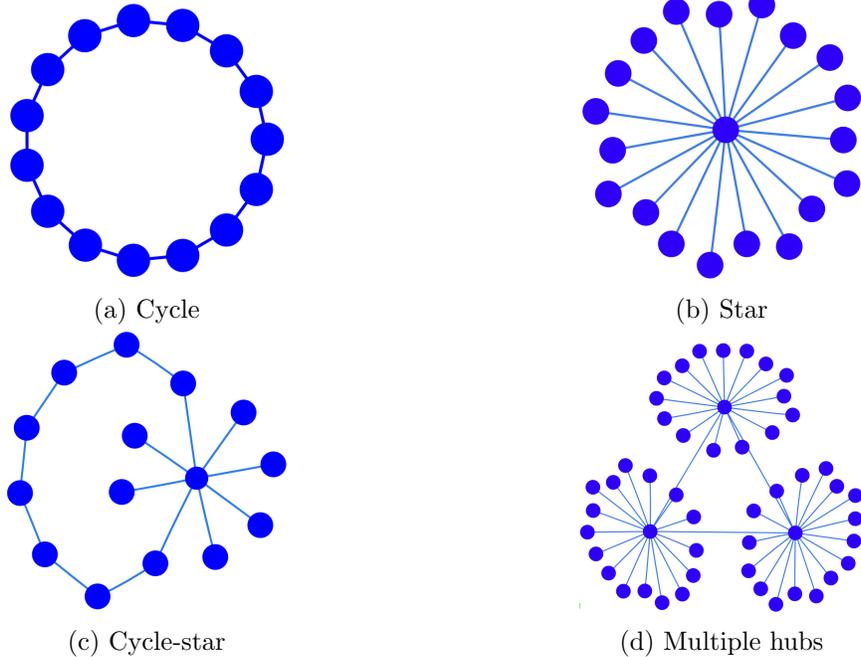
<div align="center">Figure 1: Examples of pairwise stable networks</div>

corresponds to the spoke's payoff).

Our theoretical results shows that a wide range of networks – in terms of architecture and payoffs – are pairwise stable. We next examine their robustness.

2.1 Dynamics of Linking and Hypotheses

To develop an understanding of robustness, we simulate myopic best response dynamics and obtain properties of limit networks under the two pricing rules. By definition every pairwise stable network will be stable under a best response dynamic. So the goal of the exercise is to understand the respective size of the basin of attraction of the different pairwise stable networks and to understand if these basins are very different under the two pricing protocols.

The dynamic works as follows: suppose that at each point a single individual (picked at random) makes a decision. This individual considers the potential payoffs that can be attained (according to either pricing rule) by adding or deleting a single link with any other player, given the current network. The individual chooses the option with the highest

<div align="center">7</div>

immediate payoff. In an iteration of the simulation, every individual gets 100 opportunities to make a decision.

The basic parameters used in the simulations are as follows. The value of trade between any two traders is $V = 10$. There are three groups sizes, 10, 50 and 100. The cost of a link is adjusted across scale to keep incentives as similar as possible: so $k = 8$ for $n = 10$, $k = 40$ for $n = 50$, and $k = 80$ for $n = 100$.

Examples of pairwise stable limit networks are presented in Figures 1(c) and 1(d): Cycle-star networks are commonly observed under criticality pricing for all group sizes and networks with multiple hubs are commonly observed under betweenness pricing (the number of hubs increases with group size).

Table 1 presents the average statistics in the resulting network structures at the end of the 80 simulations run for three group sizes under each pricing protocol. Further details of the procedure and statistics corresponding to each initial network can be found in Appendix B.

	Criticality			Betweenness		
	$N = 10$	$N = 50$	$N = 100$	$N = 10$	$N = 50$	$N = 100$
Connectedness	1	1	1	1	1	1
Average Degree	2	2.01	2.01	1.88	2.02	2.07
Max degree / Median degree	1.58	3.83	2.8	5.83	23	37.96
Distance (largest component)	2.65	11.04	23.02	2.09	2.53	2.68
Max payoff / Median payoff	1.77	5.84	7.19	3.63	14.51	25.63
Efficiency	0.95	0.99	0.99	0.98	0.99	0.98
Obs.	80	80	80	80	80	80

Table 1: Network statistics from computer simulations (after 100 iterations of myopic best response dynamics)

We note that networks are (almost) connected. Under criticality pricing, the networks involve one or more interconnected cycles or a hybrid star-cycle structure. By contrast, under betweenness pricing, the networks contain one or more hubs (with a disproportionate share of links). As a result, distances grow under criticality but remain small under betweenness across group size. Under criticality pricing, degree and payoff inequality grow but remain modest across group sizes, but under betweenness pricing they explode as group size grows. Finally, as networks are connected and have very low average degree, they attain close to full efficiency in all cases. These observations suggest the following

hypotheses:

Hypothesis A *Under criticality pricing networks are connected. However, as group size grows (i) distances grow (ii) degree and payoff inequality remain modest.*

Hypothesis B *Under betweenness pricing networks are connected. However, as group size grows (i) distances remain unchanged (ii) degree and payoff inequality become very large.*

Hypothesis C *Networks are close to fully efficient for all group sizes under both pricing protocols.*

3 Experimental design

The experiment uses the same parameters as the simulations. For ease of reference, we recall that there are three group sizes – 10, 50 and 100. The value of trade between any two traders is $V = 10$. The cost of a link is adjusted across scale to keep incentives as similar as possible: so $k = 8$ for $n = 10$, $k = 40$ for $n = 50$, and $k = 80$ for $n = 100$. Thus the experiment consists of 6 treatments in all: 3 group sizes $\times$ 2 pricing protocols.

In each experimental session, a continuous time game is played over 6 minutes and is referred to as a round. The first minute as a trial period and the subsequent 5 minutes as the game with payment consequences. Every group played 6 rounds.

In a round, at any moment, the subject is shown the entire network of reciprocated links. In addition, every subject is shown all outstanding link proposals – made and received – that involve him or her. Every subject is also provided full information on the payoffs of everyone. Figure 13 in Online Appendix E presents the decision screen observed by subjects. At any instant in the 6 minutes game, a subject can make or remove a proposal to another subject by simply double-clicking on the corresponding node in the computer screen. Any reciprocated proposal leads to the formation of a link. Non-reciprocated links were represented through different node shapes (see Online Appendix C for details).

At the end of each round, subjects' earnings are determined based on the network structure observed at a randomly selected moment. The first round was a trial round with no payoff relevance and only the last 5 rounds were relevant for subjects' earnings. In analyzing the data, we will focus on subjects' behavior and group outcomes from these last 5 rounds. Further details about the procedure are provided in Online Appendix D.

4 Experimental Results

For simplicity, in all the empirical analyses, the data is organized on a second to second basis. So, every round yields us 360 observations – snapshots of every subject's choices in the group. Although some information about choice dynamics between two time intervals may be lost, we believe that the second by second record is adequate for our purposes. Moreover, unless otherwise stated, all analyses are focused on data from the last 5 payoff relevant minutes of each round of the game. Using this data set, we run panel regressions for the treatment effects with time fixed effects and standard errors clustered at the group level and report regression results in Table 8 in Online Appendix F.5.

Figures 2(a) and 2(b) show the snap shots of the criticality and betweenness treatments respectively for a large group ($N = 100$) at the end of a representative round of the experiment.[5] These snap shots draw attention to three points: first, under both pricing protocols, subjects create sparse and connected networks; two, the pricing protocol leads to the emergence of equal and dispersed networks under criticality and to unequal and small distance networks under betweenness pricing; third, we can infer that there is little payoff inequality in the criticality treatment while there is great inequality under betweenness pricing. We now present a systematic analysis of the experimental data.

We begin with a summary of the findings on connectedness: connectivity is very high and similar across treatments: on average, 98.6% for $N = 10$, 99% for $N = 50$, 98.7% for $N = 100$ for criticality pricing, and 97.5% for $N = 10$, 99.4% for $N = 50$, 98.1% for $N = 100$ for betweenness pricing. This confirms the prediction of connectedness in both Hypotheses A and B.

We next turn to the treatment effects on distance and degree and payoff inequality. Our principal findings are summarized in Figure 2. Distance means average geodesic distance in the largest component.[6] Degree inequality is defined by the ratio of highest degree divided by the median degree and payoff inequality is defined as in equation (4).

Figures 2(c) and 2(f) show that as group size grows, distances grow under criticality pricing but change only slightly under betweenness pricing. Under criticality pricing, aver-

---

[5]Behavioral dynamics of the groups that generated the network structures presented in Figures 2(a) and 2(b) can be viewed through our interactive tool at the following websites: `https://networks.econ.cam.ac.uk/net_formation/animation_brokerage_critical.php` (Figure 2(a)), and `https://networks.econ.cam.ac.uk/net_formation/animation_brokerage_betweenness.php` (Figure 2(b)).

[6]The average size of the largest component under criticality pricing is 9.9 in the $N = 10$ group, 49.7 for the $N = 50$ group, and 99.3 for the $N = 100$. Under betweenness pricing, it is 9.9 in the $N = 10$ group, 49.9 for the $N = 50$ group, and 99.1 for the $N = 100$

(a) Snapshot: Criticality $N = 100$

(b) Snapshot: Betweenness $N = 100$

(c) Distance: Criticality

(d) Degree inequality: Criticality

(e) Payoff inequality: Criticality

(f) Distance: Betweenness

(g) Degree inequality: Betweenness

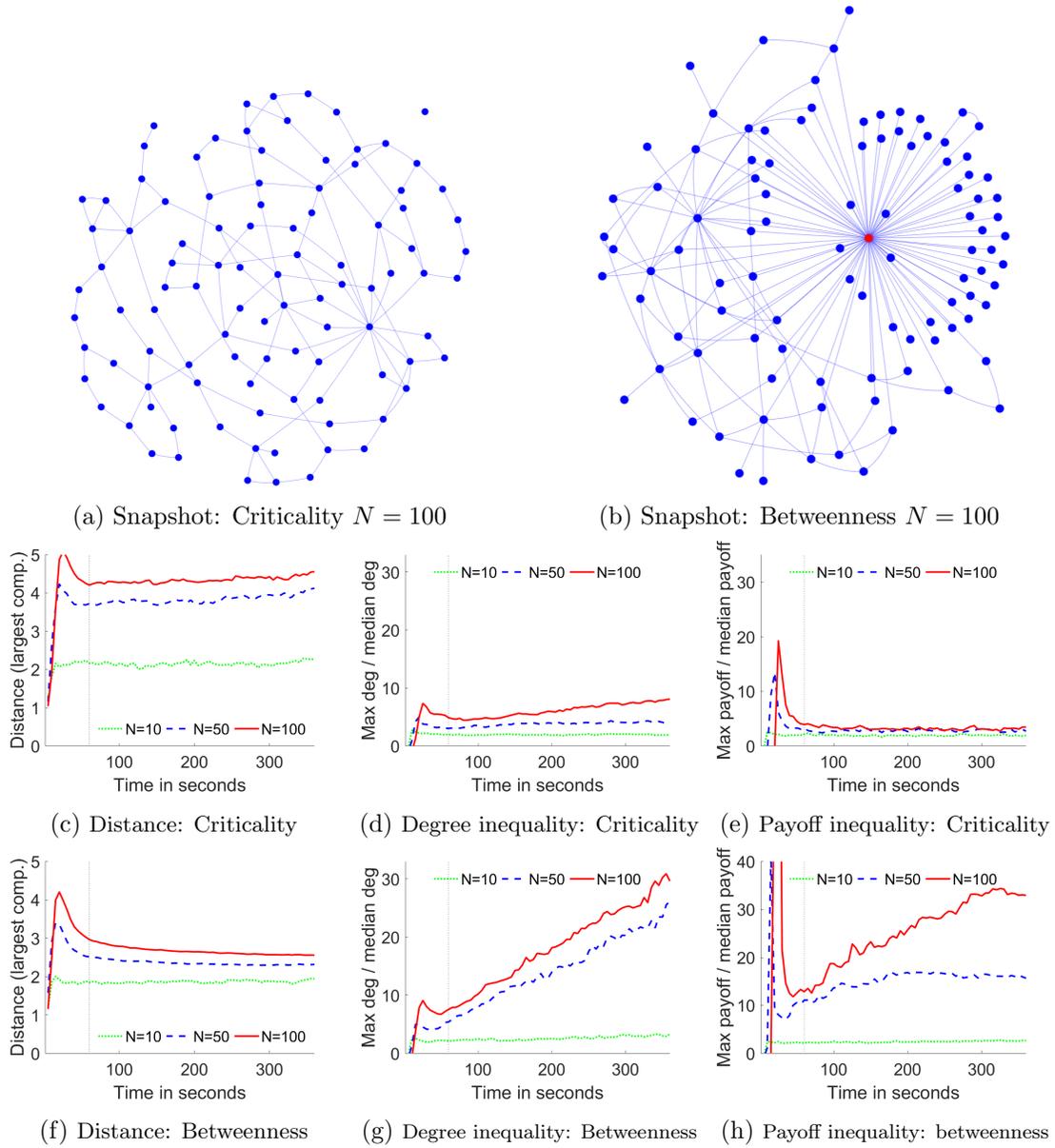(h) Payoff inequality: betweenness

Figure 2: Network structure and dynamics

age distance is around 2 in the small group, around 4 in $N = 50$ and above 4 in $N = 100$. On the other hand, under betweenness pricing, average distance is somewhere between 2 and 3 in both small and large groups. The regression analysis in the first column of Table

11

8 in Online Appendix F.5 confirms the treatment effects on distance at the 1% significance level.

Figures 2(d), 2(e), 2(g), and 2(h) show that as group size increases, degree and payoff inequality remain modest under criticality pricing but become very large under betweenness pricing: under critical pricing, payoff inequality is modest – between 2 and 3 – across group sizes. By contrast, under betweenness pricing, payoff inequality grows massively with group size – from around 2 to 16 to 35. The regression analysis in the second and third columns of Table 8 in Online Appendix F.5 confirms the treatment effects on degree and payoff inequality at the 1% significance level. These findings on networks and payoff inequality are consistent with Hypotheses A and B.

We summarize our first main finding as follows:

Result 1: Subjects create networks that are in line with Hypotheses A and B. The networks are close to connected on all cases. Under criticality pricing, average distance grows,while degree and payoff inequality remain modest across group sizes. By contrast, under betweenness pricing, average distance remains stable but degree and payoff inequality grow massively with increase in group size.

We next turn to efficiency. Figure 3(a) shows that efficiency is high for all group sizes under criticality pricing. However, Figure 3(b) shows that the level of efficiency is lower under betweenness pricing and falls sharply as group size grows. There are only minor effects of group size under criticality pricing , but rather large group size effects under betweenness pricing. The regression analysis in the fourth column of Table 8 in Online Appendix F.5 confirms the treatment effects on efficiency at the 1% significance level. This finding on efficiency goes against Hypothesis C.

Efficiency of a network is a function of the connectedness of the network and the number of links in the network – connectedness enhances welfare, additional links lowers welfare. We have noted that subjects create connected networks in all cases. So the differences in efficiency must be due to differences in number of links. Figures 3(c) and 3(d) present average degree across treatments. We see that under criticality pricing, average degree lies between 2 and 3 and changes only slightly as group size increase. On the other hand, under betweenness pricing average linking grows substantially as group size grows. The regression analysis in the fifth column of Table 8 in Online Appendix F.5 confirms the treatment effects on average degree at the 1% significance level. Thus pricing protocol and group size interact to push up linking activity.
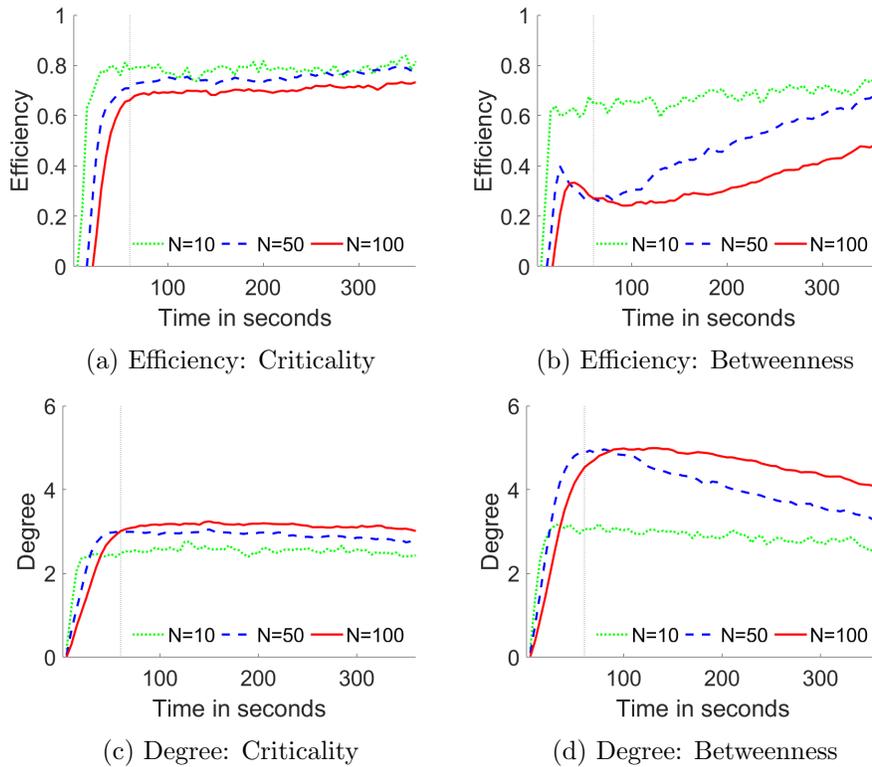
(a) Efficiency: Criticality   (b) Efficiency: Betweenness

(c) Degree: Criticality   (d) Degree: Betweenness

Figure 3: Efficiency and Degree

These observations are summarized as follows:

Result 2 (Efficiency) Efficiency remains high under criticality pricing but falls sharply with group size under betweenness pricing. This difference is due to over-linking under betweenness pricing.

Subjects create networks that are consistent with Hypotheses A and B regarding distance and degree and payoff inequality but these networks violate Hypothesis C on efficiency. The main reason for the lower efficiency under betweenness pricing is excessive linking. We briefly examine individual behaviour next.

4.1 Individual Behavior

We define three types of subjects in terms of the number of links they receive– one, the most popular individual, two, the 2nd most popular individual, and three, all the other

13

individuals. Figure 4 plots the time series of the average fraction of the number of link proposals made by each type to the total number of link proposals. In the large groups, there are major differences in the link proposals made by the two most popular individuals. Under criticality pricing, most subjects form two links and no one forms a very large number of links, thus keeping the average degree close to 2. This is confirmed by Table 2 showing that forming exactly 2 links is dominant among subjects across all group sizes, and such behavior generates the highest payoffs on average (note that forming a large number of links is never profitable in this case). This is explained by the fact that, once a cycle is created, there are no brokerage rents to be earned under criticality, which dampens the incentive to propose links. By contrast, under betweenness pricing, there are brokerage rents to be earned even once the network is connected and contains a cycle. These rents are especially large in the big groups. Table 2 indeed shows that the highest payoffs are earned by a small fraction of subjects forming at least 5 links for $N = 10$, 15 links for $N = 50$, and 20 links for $N = 100$ (in this case, payoffs increase with the number of links). On the other hand, among others with a lower number of links (the vast majority), the most profitable behavior is to form exactly one link, which is commonly exhibited by many subjects. As a response to such incentives, the number of link proposals by the most popular individual is large and grows over time (see Figures 4(e) and 4(f)). In addition, the second most popular subject also proposes a large number of links and this number remains high across time.[7] At the start of the experiment, the proposals are reciprocated by other subjects embedded in long paths, who seek to reduce intermediation rent payments. This leads to significant over-linking in large groups under betweenness pricing until at least one major hub has been identified to connect with.

---

[7]In order to formally capture subjects' response to such incentives across treatments, we consider a simple model of myopic best response with noise in Online Appendix F.6. By estimating the level of noise in each treatment, we find that it is very similar for both pricing protocols in the small group ($N = 10$), but significantly different in the large groups ($N = 50$ and $N = 100$): subjects exhibit less noise under betweenness pricing, suggesting more consistency with a myopic best response behavior. Moreover, given a fixed pricing protocol, this analysis suggests significantly less random behavior as the group size increases.
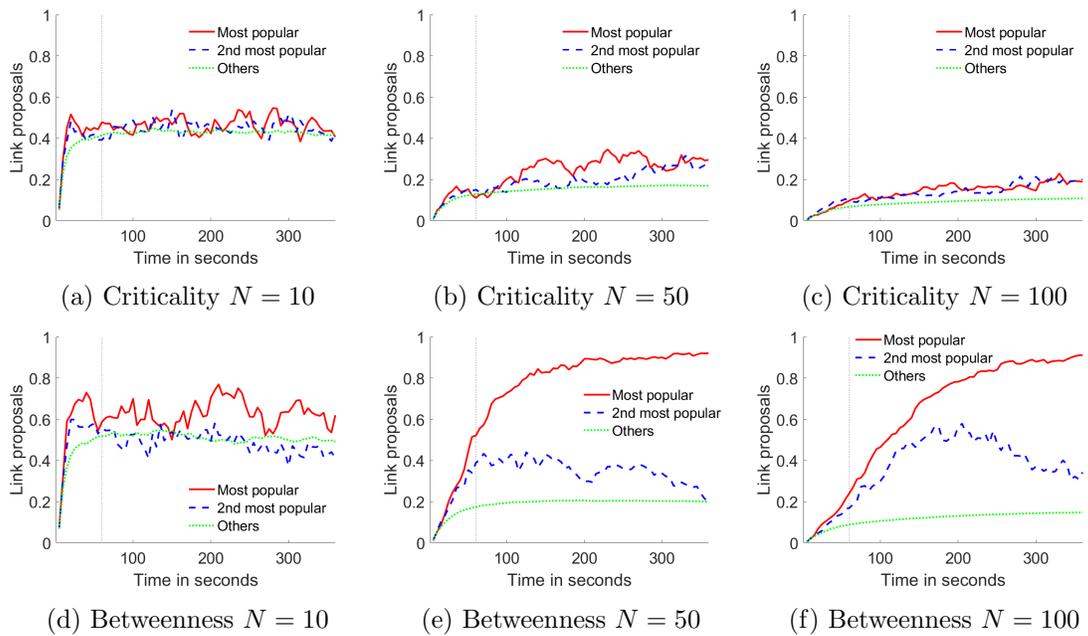
(a) Criticality $N = 10$ (b) Criticality $N = 50$ (c) Criticality $N = 100$

(d) Betweenness $N = 10$ (e) Betweenness $N = 50$ (f) Betweenness $N = 100$

Figure 4: Fraction of link proposals (number of link proposals divided by $N - 1$)

		Criticality			Betweenness		
Links	Stats	$N=10$	$N=50$	$N=100$	$N=10$	$N=50$	$N=100$
1	%	11.3	8.4	8.2	21.5	30.8	25
	Payoff	20.6	109.2	220.4	20.1	100.9	176.3
2	%	**45.1**	**42.7**	**42.2**	25.9	19.2	21.2
	Payoff	**26.3**	**154.9**	**314.3**	17.6	76.3	127.2
3	%	26.6	24.5	22.5	21	13.6	14.6
	Payoff	24	133.8	270.3	17.2	53.5	81.9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
5-10	%	5.5	11.8	13.2	**16.3**	19.7	18
	Payoff	25.2	58.6	92.4	**40.1**	-0.9	-27.7
10-15	%	.	0.6	1.7	.	3.9	5.9
	Payoff	.	-38.9	0	.	2.1	-112
15-20	%	.	0.1	0.5	.	1.2	2.2
	Payoff	.	-75.3	92.4	.	114.8	-38.4
20-25	%	.	0	0.1	.	**0.7**	**1**
	Payoff	.	-387.4	-247.1	.	**451.7**	**183.7**

Table 2: Average fraction of subjects with specific numbers of links (from 1 to 25) and associated mean payoffs (values in bold highlight the most profitable level of connectivity for each treatment)

References

J. Andreoni and J. Miller. Giving according to garp: An experimental test of the consistency of preferences for altruism. *Econometrica*, 70(2):737–753, 2002.

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

J. Barnes and P. Hut. A hierarchical o(n log n) force-calculation algorithm. *Nature*, 324: 446–449, 1986.

P. BelleFlamme and M. Peitz. *Industrial Organization: Markets and Strategies*. Cambridge University Press, 2015.

G. E. Bolton and A. Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American economic review*, 90(1):166–193, 2000.

M. Bostock, V. Ogievetsky, and J. Heer. D^3 data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.

V. Buskens and A. Van de Rijt. Dynamics of networks if everyone strives for structural holes. *American Journal of Sociology*, 114(2):371–407, 2008.

G. Charness, M. Corominas-Bosch, and G. R. Frechette. Bargaining and network structure: An experiment. *Journal of Economic Theory*, 136(1):28–65, 2007.

S. Choi, A. Galeotti, and S. Goyal. Trading in networks: theory and experiments. *Journal of the European Economic Association*, 15(4):784–817, 2017.

S. Choi, S. Goyal, and F. Moisan. Connectors and influencers. 2022.

D. Condorelli and A. Galeotti. Intermediation in networks. In Y. Bramoulle, A. Galeotti, and B. Rogers, editors, *Oxford Handbook of the Economics of Networks*. Oxford University Press, 2016.

D. Condorelli, A. Galeotti, and L. Renou. Bilateral trading in networks. *The Review of Economic Studies*, 84(1):82–105, 2016.

P. Eades. A heuristic for graph drawing. *Congressus numerantium*, 42:149–160, 1984.

A. Falk and M. Kosfeld. Its all about connections: Evidence on network formation. *Review of Network Economics*, 11:Article 2, 2012.

M. Farboodi. Intermediation and voluntary exposure to counterparty risk. *Available at SSRN 2535900*, 2014.

E. Fehr and K. M. Schmidt. A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, 114(3):817–868, 1999.

D. Friedman and R. Aprea. A continuous dilemma. *American Economic Review*, 102(1): 337–63, 2012.

T. M. Fruchterman and E. M. Reingold. Graph drawing by force-directed placement. *Software: Practice and experience*, 21(11):1129–1164, 1991.

D. M. Gale and S. Kariv. Trading in networks: A normal form game experiment. *American Economic Journal: Microeconomics*, 1(2):114–32, 2009.

A. Galeotti and S. Goyal. Competing chains. Work in progress, 2014.

J. K. Goeree, A. Riedl, and A. Ule. In search of stars: Network formation among heterogeneous agents. *Games and Economic Behavior*, 67(2):445–466, 2009.

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

S. Goyal and F. Vega-Redondo. Structural holes in social networks. *Journal of Economic Theory*, 137(1):460–492, 2007.

S. Goyal, S. Rosenkranz, U. Weitzel, and V. Buskens. Information acquisition and exchange in social networks. *Economic Journal*, 127:2302–2331, 2017.

C. A. Holt and S. K. Laury. Risk aversion and incentive effects. *American Economic Review*, 92(5):1644–1655, 2002.

Y. Hu. Efficient, high-quality force-directed graph drawing. *Mathematica Journal*, 10(1): 37–71, 2005.

M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. *Journal of economic theory*, 71(1):44–74, 1996.

M. Jacomy, T. Venturini, S. Heymann, and M. Bastian. Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. *PloS one*, 9(6), 2014.

S. Kariv, M. H. Kotowski, and C. M. Leister. Liquidity risk in sequential trading networks. *Games and Economic Behavior*, 109:565–581, 2018.

J. Kleinberg, S. Suri, É. Tardos, and T. Wexler. Strategic network formation with structural holes. In *Proceedings of the 9th ACM Conference on Electronic Commerce*, pages 284–293. ACM, 2008.

M. Kosfeld, A. Okada, and A. Riedl. Institution formation in public goods games. *American Economic Review*, 99(4):1335–55, 2009.

M. H. Kotowski and C. M. Leister. Trading networks and equilibrium intermediation. *HKS Working Paper No. RWP18-001*, 2018.

M. Manea. Intermediation and resale in networks. *Journal of Political Economy*, 126(3): 1250–1301, 2018.

B. Rammstedt and O. P. John. Measuring personality in one minute or less: A 10-item short version of the big five inventory in english and german. *Journal of research in Personality*, 41(1):203–212, 2007.

A. Rubinstein and A. Wolinsky. Middlemen. *The Quarterly Journal of Economics*, 102(3): 581–593, 1987.

D. F. Spulber. Market microstructure and intermediation. *Journal of Economic perspectives*, 10(3):135–152, 1996.

APPENDIX

A   Theory

Following Jackson and Wolinsky [1996], a network is said to be pairwise stable if no one can benefit by removing any existing link, and no pair can mutually benefit by adding a non existing link with each other.

Definition 1. *A strategy profile s^* leads to a pairwise stable network $g(s^*)$ if the following conditions hold:*

- *For any $i \in N$ and any $s_i \in S_i$, $\Pi_i(s^*) \geq \Pi_i(s_i, s^*_{-i})$;*

- *For any $i, j \in N$ and any $s \in S$ such that $g_{ij}(s^*) = 0$ and $g(s) = g(s^*) + g_{ij}$, if $\Pi_i(s) > \Pi_i(s^*)$, then $\Pi_j(s) < \Pi_j(s^*)$.*

Here is a description of pairwise stable networks under criticality pricing rule.

Proposition 1. *Suppose payoffs are given by (1). A pairwise stable network always exists.*

- *The complete network is not pairwise stable for $n \geq 4$*

- *If $k < V \sum_{i=1}^{n-2} \frac{i}{2(2+i)}$ then a cycle network is pairwise stable. In addition, a network with multiple cycles and spokes is pairwise stable for particular subsets of parameters.*

- *If $V/6 < k < Vn/3 - V/6$ then a star network is pairwise stable. In addition, a tree with two or more hubs can be pairwise stable for particular subsets of parameters.*

- *If $k > V/2$ then the empty network is pairwise stable and if $k > Vn/3 - V/6$ then it is the unique pairwise stable network.*

Proof. Pairwise stability of the Empty network is straightforward.

In the Star network, the hub does not benefit from removing a link if $c < \frac{Vn}{3} - \frac{V}{6}$. Under the same condition, no spoke can benefit from removing a link. No pair of spokes can benefit by adding a link with each other if $c > \frac{V}{6}$. The Star network is therefore pairwise stable if $\frac{V}{6} < c < \frac{Vn}{3} - \frac{V}{6}$.

In the Cycle network, no pair of players can benefit by adding a link because $c > 0$ and it cannot change the access benefit $(= \frac{(n-1)V}{2})$ or the intermediation rent $(= 0)$. Given that

that the access benefit obtained by a player removing a link is $\sum_{i=1}^{n-1} \frac{V}{i+1}$, such an action is not profitable as long as $c < \frac{V(n-1)}{2} - \sum_{i=1}^{n-1} \frac{V}{i+1} = \sum_{i=1}^{n-2} \frac{Vi}{2(2+i)}$. The Cycle network is therefore pairwise stable if $c < \sum_{i=1}^{n-2} \frac{Vi}{2(2+i)}$.

In the Complete network, if $n = 3$ and $c < \frac{V}{6}$, no one can benefit by removing a link. If $n > 3$ and $c > 0$, then any player can benefit by removing a link because it cannot reduce either access benefit $(= \frac{(n-1)V}{2})$ or intermediation rent $(= 0)$. Therefore. the Complete network is not pairwise stable for a sufficiently large n.

Finally, it is easy to see that, for any values of $c > 0$ and $n \geq 3$, there is at least one pairwise stable network among the Empty, Star, and Cycle networks. $\qquad\square$

We next provide a description of pairwise stable networks under betweenness pricing rule.

Proposition 2. *Suppose payoffs are given by (2). A pairwise stable networks always exists.*

- *If $k < V/6$ then the unique pairwise stable network is complete.*

- *For fixed values of c and V: a cycle is not pairwise stable for large enough n.*

- *If $V/6 < k < Vn/3 - V/6$ then a star network is pairwise stable. In addition, a tree with two or more hubs can be pairwise stable for particular subsets of parameters.*

- *If $k > V/2$ then the empty network is pairwise stable and if $k > Vn/3 - V/6$ then it is the unique pairwise stable network.*

Proof. Pairwise stability of the empty network is straightforward.

Pairwise stability of the Star network follows the same arguments as in the proof of Proposition 1.

In the Complete network, no player can benefit by removing a link as long as $c < \frac{V}{6}$.

In the Cycle network, it is easy to see that, the gain in benefits (access benefits and brokerage rents) for adding a link between two players sitting at an extreme distance from one another increases with n. As a result, if n is sufficiently large, then such a move becomes profitable for both players. This argument naturally extends to any k-cycle-s-Star network with fixed values of k and s, and a sufficiently large n.

Finally, it directly follows that, for any values of $c > 0$ and $n \geq 3$, there is at least one pairwise stable network among the Empty, Star, and Complete networks. $\qquad\square$

B Computer Simulations

Given the large variety of pairwise stable networks under both pricing protocols, 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 of 100 iterations, each of which consisting of all individuals making a choice in a random sequence (so that each individual makes the same number of choices, i.e., 100, in any given simulation). In this case, when selected to make a choice, an individual considers the potential payoffs that could immediately be reached (according to either pricing protocol described above) by adding or deleting a single link with any other person from the group, given the most recent network structure. The connection corresponding to the option with the highest immediate payoff is then updated accordingly: if a proposal was (not) previously formed with that person, then this proposal is removed (added). Under this simulation setting, we vary the type of initial network structure to identify any potential effect on the dynamics and the converging networks. More specifically, we consider two well-known network models: the Erdos-Renyi model that generates random networks with equally distributed connections, and the Barabasi-Albert model that generates scale free networks with significant inequality in the distribution of connections (Barabási and Albert [1999]). For each model, we also consider two levels of connection density: in the Erdos-Renyi model, we choose an average degree of 2 (low density)[8] and 5 (high density); in the Barabasi-Albert model, we start from an initial network of 3 individuals forming a line (2 individuals connected with the same third person), and we choose the number of links to be added by any newly introduced individual to be 1 (low density) and 3 (high density). We run 20 independent simulations for each of the 4 types of initial networks (initial networks generated according to the above stochastic models differ across all simulations).

To describe the dynamics followed by the simulations across the different treatments, Figures 5, 6, 7, 8, 9, and 10 show the time series corresponding to the the different statistics of interest (degree, connectedness, efficiency, distance, degree inequality, payoff inequality).

Tables 3, 4, 5, and 6 further describe the various statistics at the end of each simulations depending on the different types of initial network (Erdos Renyi with low/high density, Barabasi-Albert with low/high density). Those results show overall consistency of the simulation outcomes being reached, regardless of the initial network.

---

[8]We do not consider initial networks with lower density as they too often lead the simulations to converge to the empty network, which is of less interest to our study.
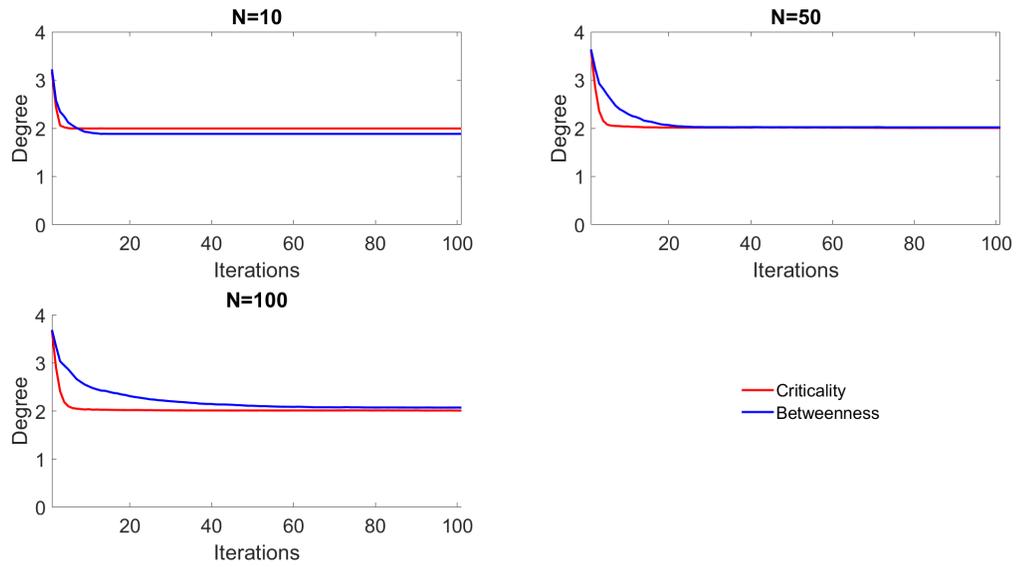
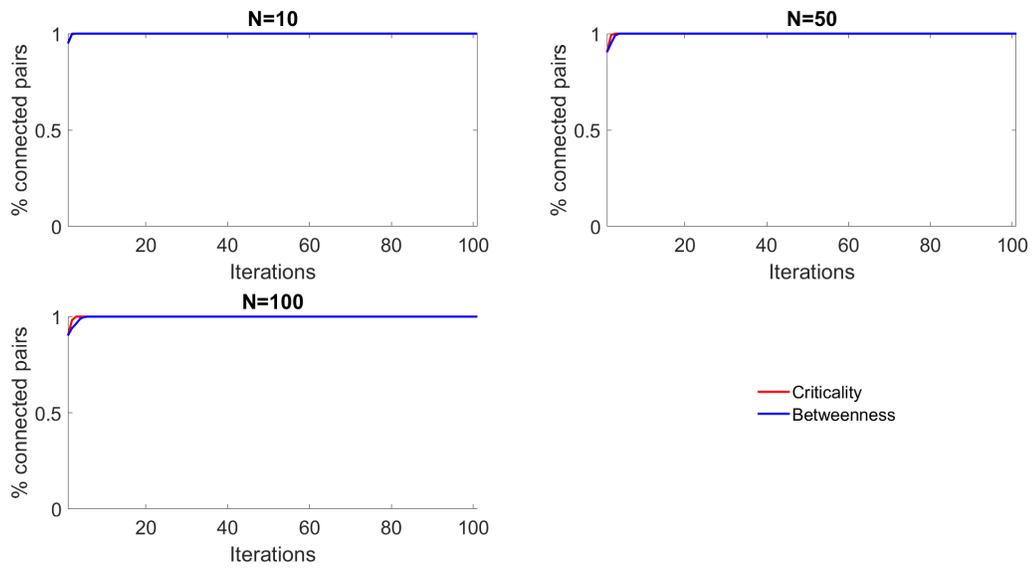Figure 5: Time series of the mean degree across simulations (N=80)

Figure 6: Time series of the connectedness ratio across simulations (Nb of obs.=80)
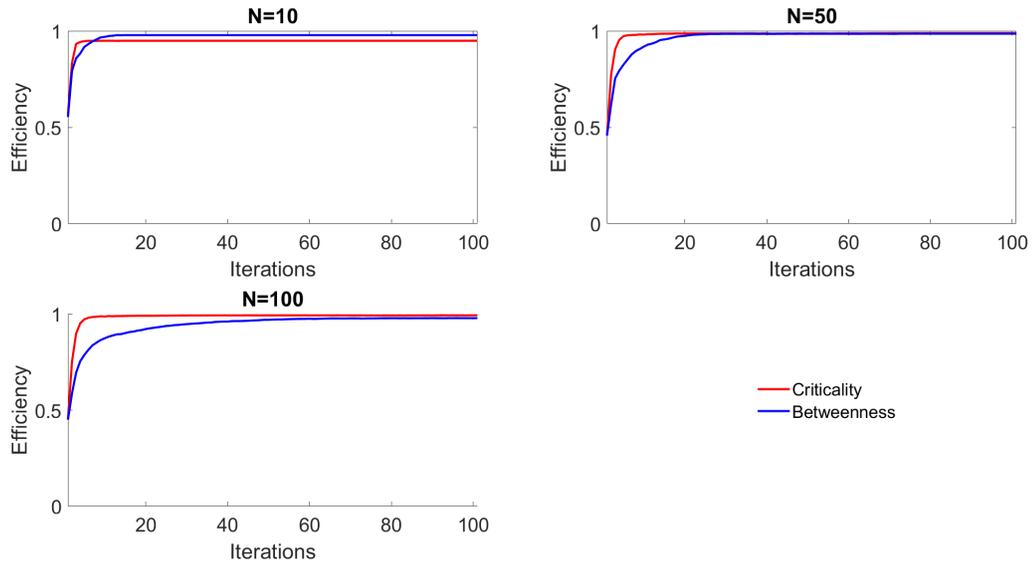
23

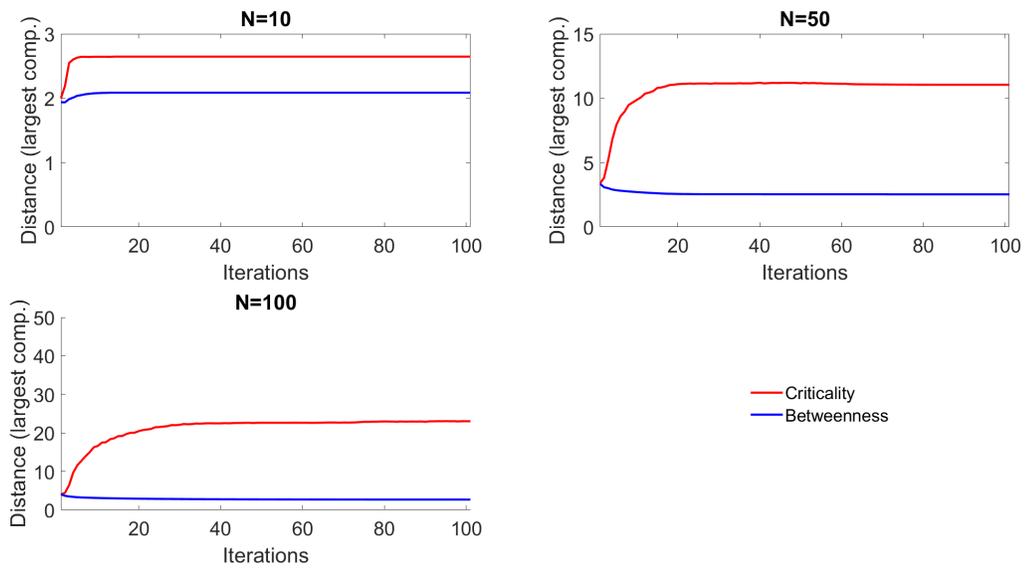Figure 7: Time series of the mean efficiency across simulations (Nb of obs.=80)

Figure 8: Time series of the mean distance in the largest component across simulations (Nb of obs.=80)
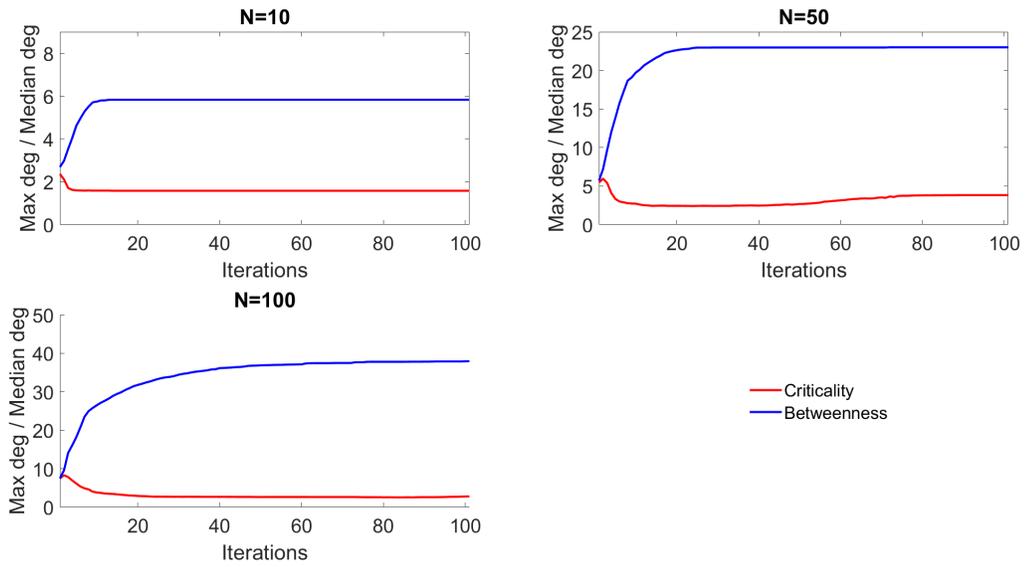
Figure 9: Time series of the mean degree inequality across simulations (Nb of obs.=80)
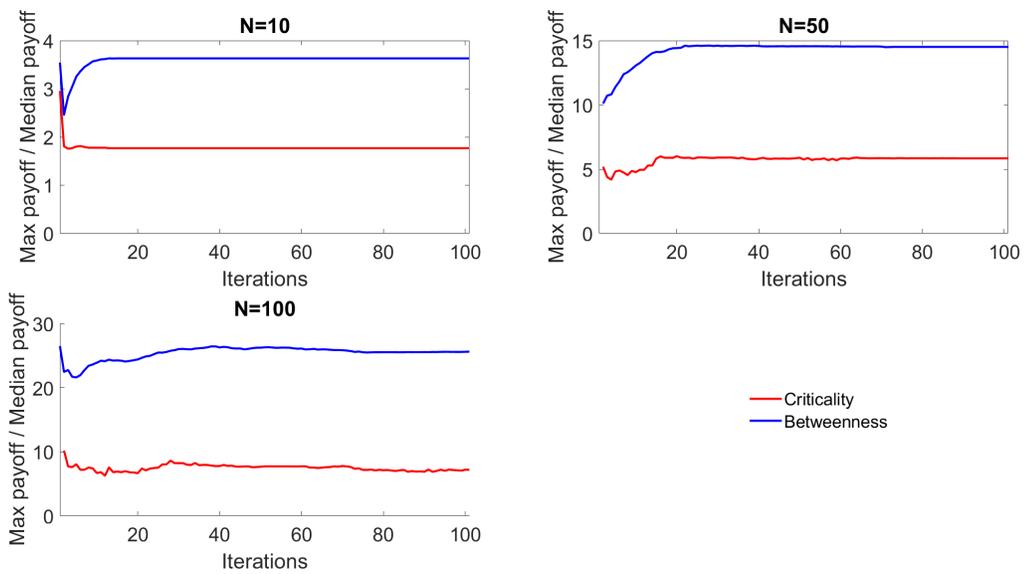

Figure 10: Time series of the mean payoff inequality across simulations (Nb of obs.=80)

25

	Criticality			Betweenness		
	$N = 10$	$N = 50$	$N = 100$	$N = 10$	$N = 50$	$N = 100$
Connectedness	1	1	1	1	1	1
Average degree	2	2.01	2.01	1.84	2.01	2.05
Max degree / Median degree	1.25	2.6	2.2	5.85	23.25	38.1
Distance (largest component)	2.71	11.55	23.8	2.1	2.52	2.67
Max payoff / Median payoff	1.37	3.91	3.86	3.67	14.61	25.62
Efficiency	0.95	0.99	0.99	0.99	0.99	0.98
Obs.	20	20	20	20	20	20

Table 3: Network statistics from computer simulations when starting from Erdos-Renyin networks (mean degree = 2), after 100 iterations of myopic best response dynamics

	Criticality			Betweenness		
	$N = 10$	$N = 50$	$N = 100$	$N = 10$	$N = 50$	$N = 100$
Connectedness	1	1	1	1	1	1
Average degree	2	2	2	1.92	2.02	2.09
Max degree / Median degree	1.28	2.33	2.18	5.9	22.95	34.25
Distance (largest component)	2.71	12.09	24.35	2.04	2.53	2.71
Max payoff / Median payoff	1.44	3.09	2.8	3.54	14.39	23.43
Efficiency	0.95	0.99	1	0.97	0.99	0.97
Obs.	20	20	20	20	20	20

Table 4: Network statistics from computer simulations when starting from Erdos-Renyin networks (mean degree = 5), after 100 iterations of myopic best response dynamics

	Criticality			Betweenness		
	$N = 10$	$N = 50$	$N = 100$	$N = 10$	$N = 50$	$N = 100$
Connectedness	1	1	1	1	1	1
Average degree	1.98	2.01	2.01	1.8	2.03	2.08
Max degree / Median degree	2.33	5.6	3.48	6.2	22.45	40.7
Distance (largest component)	2.51	10.41	22.27	2.13	2.54	2.66
Max payoff / Median payoff	2.42	7.37	9.33	3.8	14.15	26.84
Efficiency	0.95	0.99	0.99	1	0.98	0.98
Obs.	20	20	20	20	20	20

Table 5: Network statistics from computer simulations when starting from Barabasi-Albert networks (1 link per every newly introduced individuals), after 100 iterations of myopic best response dynamics

	Criticality			Betweenness		
	$N = 10$	$N = 50$	$N = 100$	$N = 10$	$N = 50$	$N = 100$
Connectedness	1	1	1	1	1	1
Average degree	2	2.01	2.01	1.97	2.01	2.06
Max degree / Median degree	1.48	4.8	3.35	5.38	23.35	38.8
Distance (largest component)	2.65	10.13	21.68	2.07	2.54	2.69
Max payoff / Median payoff	1.84	9	12.77	3.51	14.89	26.63
Efficiency	0.95	0.99	0.99	0.96	0.99	0.98
Obs.	20	20	20	20	20	20

Table 6: Network statistics from computer simulations when starting from Barabasi-Albert networks (3 links per every newly introduced individuals), after 100 iterations of myopic best response dynamics

C   Experiment: Challenges and methodology

Experiments in this setting pose at least several challenges to human subjects because of (i) the large decision space, (ii) the complexity of payoff functions, and (iii) the observability of link proposals and network structure. We discuss these challenges and explain how our experimental software and design address each of them. Some of the discussion is taken from our companion paper, Choi, Goyal, and Moisan [2022].

Learning and dynamics. We wish to allow subjects ample opportunities to learn about the environment of decision making, other subjects' behaviors, and how to respond optimally to them. Because of the complexity of payoff functions and decision making, the issue of learning and behavioral convergence is particularly important. To address these issues, we run the experiment in continuous time with near real time updating on the evolution of network structure and link proposals made by subjects.[9] Continuous time experiments can also offer better prospects for convergence than discrete time experiments (see e.g., Friedman and Aprea [2012]).

In our experiment, the game is played in continuous time for 6 minutes during which every subject was free to asynchronously adjust their link proposals. Because subjects face a complex problem of decision making and need some time to figure out the game and coordinate their actions, a trial time of one minute is provided (during which subjects start choosing their actions with no monetary consequence). After the trial period is over, the subsequent 5 minutes are payoff relevant and one second is randomly chosen to determine subjects' earnings in the game. This information is publicly known to subjects.

Two-sided linking protocol. The intermediation models we consider use the two-sided linking protocol with which a link between two individuals is formed if and only if both individuals consent to form it. This protocol distinguishes between link proposals (i.e., choices made by subjects) and realized links (i.e., link proposals that are reciprocated). Compared to the one-sided linking protocol, the two-sided linking protocol introduces an extra layer of the relationship between any two individuals: the pair is linked, or unlinked with none of them making a link proposal, or unlinked with only one of them making a link proposal to the other. Keeping track of such information in large groups can be a challenge for human subjects.

---

[9]More precisely, the network depicted on any subject's screen is updated every 2 seconds or whenever the subject makes a new decision.

In order to make it easy for the decision maker to grasp the relation with any individual, we use the visual representation on the status of the linking relationship between the decision maker, denoted by Me, and an individual as shown in Table 7: An individual who neither made a proposal to nor received a proposal from the decision maker is represented with a circle shape; If an individual sent a link proposal to the decision maker who did not reciprocate it, that individual is depicted with a square shape; If an individual receives a link proposal from the decision maker but did not reciprocate it, the individual is represented with a triangle shape; If both an individual and the decision maker make link proposals to each other, the link between them is visualized with the individual being shaped with a circle.
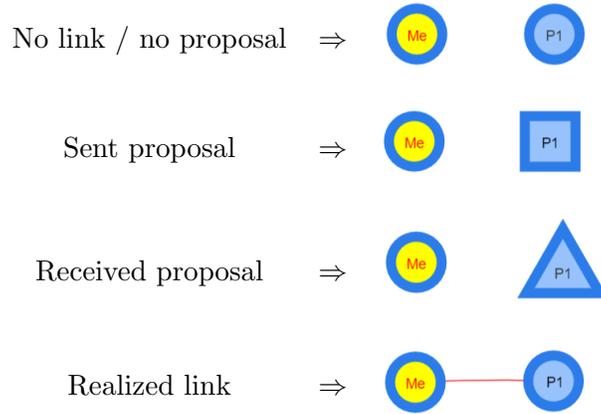
Table 7: Visualization of proposals and link

On the other hand, from the decision maker's perspective, information about the relation between any other two individuals is provided in the binary form: the corresponding pair is shown to be either linked or not linked. Therefore, no information about the detail of unlinked individuals is provided to the decision maker.

Figure 11 illustrates our method of distinguishing these different cases. In the initial network depicted on the left side of Figure 11, the decision maker who is represented with a yellow node identified as "me" does not make any link proposal, but receives link proposals from players P2, P3, and P4; these individuals are triangle shaped. From the network on the left, if the decision maker makes link proposals to P2, P3, P4, and P6, the network changes to the right side of Figure 11. The decision maker then have three realized links with P2, P3, and P4, and one pending link proposal to P6. On the other hand, the
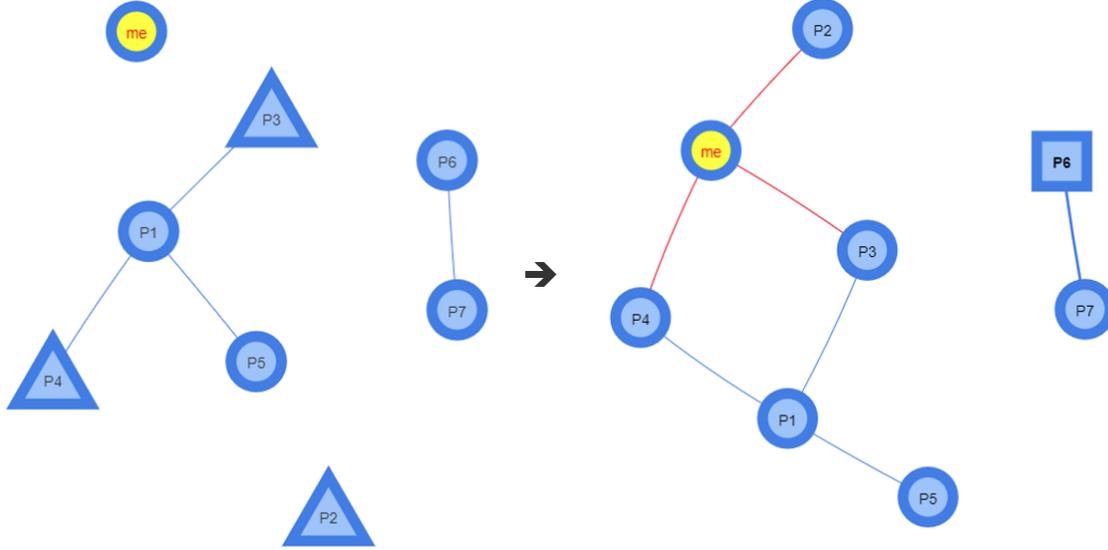
Figure 11: Example of network updates after new choices

decision maker can only see the realized links between any other players (e.g., between P1 and P5); no information is provided about unlinked pairs (e.g., the pair of P5 and P7 may be unlinked because either P5, P7, or both P5 and P7 do not make a link proposal).

C.1 Network visualization

Existing studies of network formation in economics have considered small group sizes such as 4 or 8 people and visualized evolving networks with fixed positions of nodes. When the group size increases, such a representation of networks with fixed positions of nodes makes it very difficult for subjects to perceive network features.

For example, consider a group of 20 people with fixed positions of nodes in a circle as depicted in Figure 12(a); the exact network is barely perceptible by observing this figure. The same network structure can be represented in a transparent manner in Figure 12(b).

For subjects to learn their optimal choices, they must have a good idea of evolving networks. An appropriate tool for visualizing networks and their changes in real time is thus critical in running the experiment in continuous time. This leads us to develop an experimental software including an interactive network visualization tool that allows the network to automatically reshape itself in response to decisions made by subjects. We

(a) Fixed visualization　　　　　　(b) Adaptive visualization
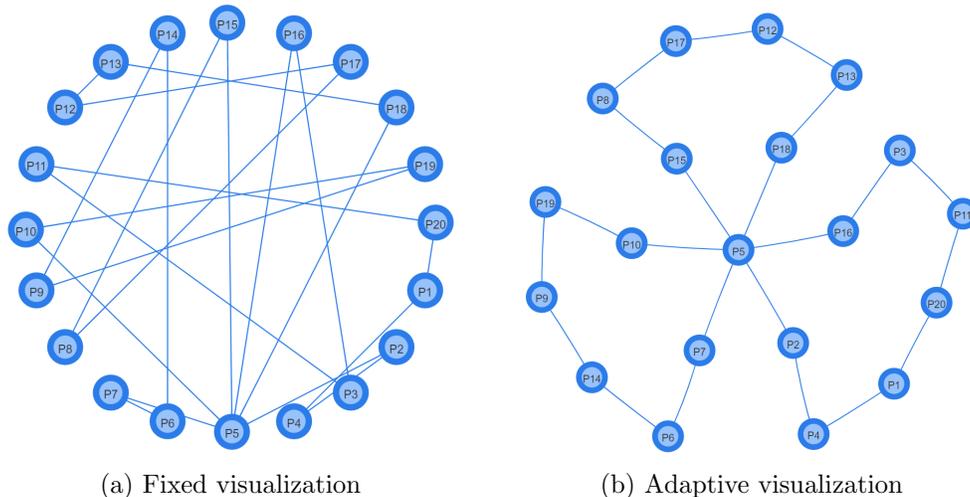
Figure 12: Examples of network visualization

use *force-directed algorithms* to visualize networks in real time (see, e.g., Eades [1984], Fruchterman and Reingold [1991], Hu [2005], Bostock et al. [2011], Jacomy et al. [2014]).

Clearly, different network environments will offer differ levels of transparency and information on network architecture. Our strategy is to start with a visualization approach that is efficient and that allows us to systematically explore the effects of different variables – such as scale and variations in information on networks and payoffs (taken up below). Of course, the experimental platform is flexible enough to incorporate other ways of representing networks and can be used to explore the effects of network visualization itself on human behavior and network formation. Thus, the three experiments reported in this paper could be interpreted as benchmark findings with efficient network visualization.

The algorithms used in the experimental software can be understood with three building components as follows. An illustration of technical details is provided in Online Appendix.

Barnes-Hut approximation algorithm. Due to the heavy load of computing the repulsion forces in large-scale networks, we adopt the approximation algorithm proposed by Barnes and Hut [1986]. It first constructs a quad-tree by recursively dividing the 2-dimensional space into same size quadrants such that every node can eventually be associated with exactly only one region based on its position in space (or leaf of the tree). In this way, network nodes are placed in the two-dimensional Cartesian coordinate system.

The approximation algorithm then aggregates groups of nodes located in the same

31

region to determine a unique force that approximates the sum of individual forces in that group as if the group of nodes were a single node. Specifically, starting from the largest region of the quad-tree, the algorithm assesses the Euclidean distance $dist(o, c)$ between a given node o and the center of mass of that region c. If this distance is sufficiently large, then the group of nodes in the corresponding region is considered as a single node. Otherwise, the process is iterated by considering subregions from the tree. The condition with distance trades off the speed of the simulation against its accuracy.

Modelling forces. The force-directed algorithms use attraction and repulsion forces between nodes in the network and gravity force toward the center of the screen, in order to readjust their positions in two-dimensional space and improve the overall visibility on the subjects' screen.

Any two nodes o and o' in the network repulse each other with a *repulsion force* $F_r(o, o')$ in order to avoid overlaps and allow a sparse visualization of the network. It is modelled as a decreasing function of the Euclidean distance between two nodes $dist(o, o')$, implying that close nodes repulse more than distant nodes. Two connected nodes o and o' in the network apply an *attractive force* $F_a(o, o')$ towards each other to allow for visual proximity. A classical approach of modelling attraction force is a linear and positive relation with the distance, implying that close nodes attract less than distant nodes. Finally, every node o applies a *gravity force* $F_g(o)$ to the center of the spatialization space O to pull the entire network towards the center of the screen. In particular, such a force allows disconnected components to be within reasonable distance from each other, and therefore more easily visualized on the screen.

In summary, nodes are attracted by gravity and other nodes they are linked with, and repulsed by other nodes they are not linked with.

The net force vector applied to any node o resulting from the above three forces is then given by the following form of weighted sum (where F_x and F_y represent corresponding force vectors applied to the x and y axes of the Euclidean space respectively):

$$F_x(o) = \frac{x_O - x_o}{dist(o, O)} F_g(o) + \sum_{o' \in N \setminus \{o\}} \frac{x_{o'} - x_o}{dist(o, o')} F_a(o, o') + \sum_{o'' \in N \setminus \{o\}} \frac{x_{o''} - x_o}{dist(o, o'')} F_r(o, o'') \quad (6)$$

$$F_y(o) = \frac{y_O - y_o}{dist(o, O)} F_g(o) + \sum_{o' \in N \setminus \{o\}} \frac{y_{o'} - y_o}{dist(o, o')} F_a(o, o') + \sum_{o'' \in N \setminus \{o\}} \frac{y_{o''} - y_o}{dist(o, o'')} F_r(o, o'') \quad (7)$$

Note that the computation of the repulsion force for every node can be a complex task,

especially in the context of large networks. In order to address this issue, the experimental software approximates this computation using the well-known algorithm introduced by Barnes and Hut [1986]. More concretely, it finds groupings of nodes in proximity and determines a repulsion force $F_r(o, c)$ between node o and the group of nodes with a center of mass c, in replacement of the brute force method of computing repulsion forces between all pairs of nodes.

We turn back to Figure 12 to derive some intuition of how the net force equations aggregate forces for every node and the network is visualized in the two-dimensional space. The adaptive visualization in Figure 12(b) is obtained by using the force-directed algorithm. The network has a petal-like structure with three independent sub-components connected through a common player, P5. The visualization algorithm makes P5 to be located at the center of the screen because the neighbors of P5 repluse each other and surround P5, while each pair of P5's neighbors belonging to the same sub-component are in close proximity and positioned side by side. The three forces then operate to make the rest of players located to draw non-overlapping petal-like structures.

Dynamic adjustment. The above equations (6) and (7) describe the net forces that are applied for the visualization of the network, given the positions of all nodes and the links between nodes. When the network changes, the algorithm updates dynamically the network visualization by computing the corresponding velocity of nodes on both coordinate axes.

In our large-scale experiment, this visualization tool improves graphical clarity of evolving networks and helps subjects distinguish between those who are more connected and those who are less connected. It is wothwhile to note that this tool allows interaction between the subject and the network: while the nodes are subject to the above attraction and repulsion forces, they can also be freely manipulated by the participant through the usual drag-select functionality. The creation and removal of links is also interactive through double-clicking on corresponding nodes. This network visualization tool is built on the open source Javascript library *vis.js*.

D   Experimental design and procedures

The experiment consists of 6 treatments in all: 3 group sizes $\times$ 2 pricing protocols. There were in total 18 sessions: 1 session with 4 groups of 10 subjects, 4 sessions with 50 subjects,

and 4 sessions with 100 subjects for each of the criticality and betweenness treatments. In each experimental session, subjects were matched to form a group and interacted with the same subjects throughout the experiment. A total of 1280 subjects participated in the experiment.

A subject participated in only one of the experimental sessions. After subjects read the instructions, the instructions were 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.

In each experimental session, a continuous time game is played over 6 minutes and is referred to as a round. The first minute as a trial period and the subsequent 5 minutes as the game with payment consequences. Every group played 6 rounds.

In a round, at any moment, the subject is shown the entire network of reciprocated links. In addition, every subject is shown all outstanding link proposals – made and received – that involve him or her. Every subject is also provided full information on the payoffs of everyone (this is done by mentioning the numeric value of the payoffs for every subject next to their player ID). However, subjects are *not* shown unreciprocated links among other pairs. The principal motivation for this design choice was to keep the information options available to a subject manageable. Figure 13 presents the screen observed by subjects.

At any instant in the 6 minutes game, a subject can make or remove a proposal to another subject by simply double-clicking on the corresponding node in the computer screen. Any reciprocated proposal leads to the formation of a link. Non-reciprocated links were dealt with the protocol described in section C. At any moment, every subject is shown the amount of access benefits, brokerage rents, overall cost of linking, and net payoffs. Finally, the subjects are also provided with information about the net payoffs of every other player (given within the corresponding node of the network).

At the end of each round, every subject is informed of a time moment randomly chosen for payment. The subject is also provided detailed information on subjects' behavior at the chosen moment, through the corresponding network structure. While the groups were fixed in a session, subjects' identification numbers were randomly reassigned at the beginning of every round in order to reduce potential repeated game effects. The first round was a trial round with no payoff relevance and only the last 5 rounds were relevant for subjects' earnings.

At the beginning of the experiment, every subject was endowed with an initial balance

of 80 points for the $N = 10$ treatments, 400 points for the $N = 50$ treatments, and 800 points for the $N = 100$ treatments. Subjects' total earnings in the experiment were given by the sum of earnings across the last 5 rounds and the initial endowment. Earnings were calculated in terms of experimental points and then exchanged into euros at the rate of 20 points being equal to 1 euro in the $N = 10$ treatments, 110 points being equal to 1 euro in the $N = 50$ treatments, and 220 points being equal to 1 euro in the $N = 100$ treatments.[10] Each session lasted on average 90 minutes, and subjects earned on average about 16.4 euros (including a 5 euros show-up fee).

All experimental sessions were conducted in the Laboratory for Research in Experimental and Behavioral Economics (LINEEX) at the University of Valencia. The experimental sessions of the $N = 100$ treatment were conducted through the internet connection between LINEEX and the Laboratory for Experimental Economics (LEE) at the University Jaume I of Castellón. In this case the number of subjects was then evenly distributed across the two locations. Subjects in the experiment were recruited from online recruitment systems of the LINEEX and LEE.

At the end of the experiment, subjects took incentivized tasks to elicit social preferences and risk preferences. They are a modified version of Andreoni and Miller [2002] and Holt and Laury [2002], respectively. In addition, subjects answered a brief version of the Big Five personality inventory test adapted from Rammstedt and John [2007], a comprehension test related to the experimental game, and a debriefing questionnaire including demographic information. More details about them can be found in Online Appendix **??**.

E   Network game interface

The decision making interface used in the experiment is similar across all treatments. More specifically, Figure 13 illustrates a (fictitious) example of a subject's computer screen in Treatment **Criticality**. 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 1 minute has passed), and a comprehensive description of the subject's own payoff. Information about payoffs include own benefits from own connections, brokerage rents, the cost of linking, and the net earnings. The bottom part of the screen shows detailed information about the network (the subject's node is highlighted in yellow).

---

[10]Different conversion rates and initial endowments are used to maintain similar earnings across treatments.
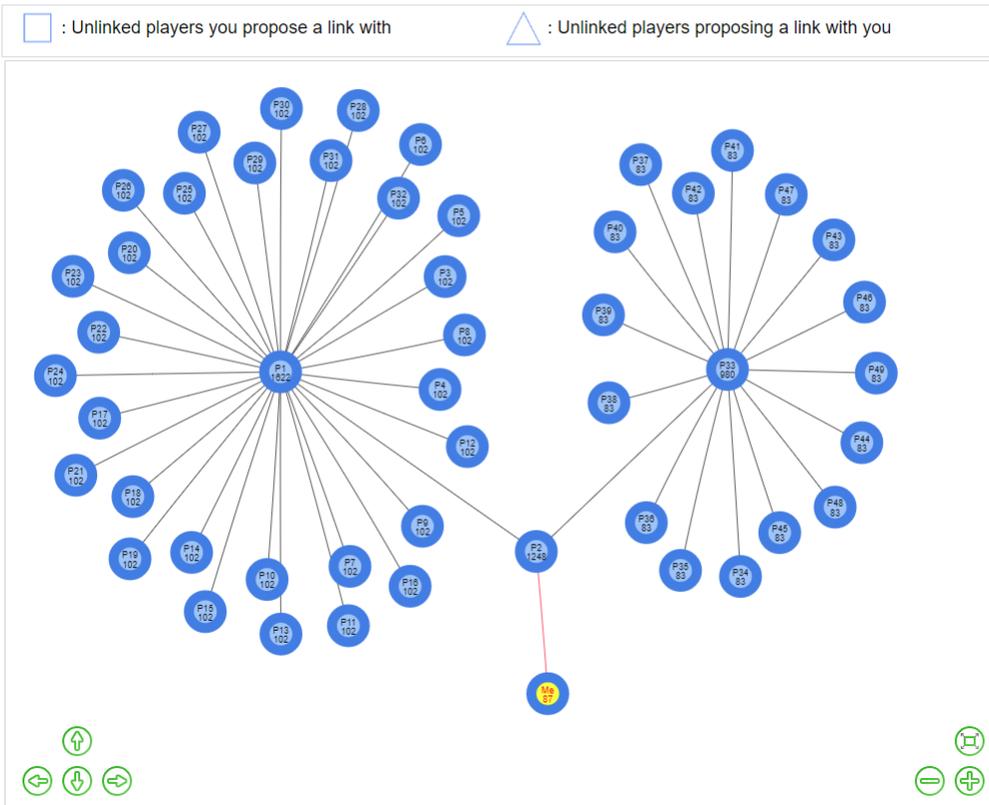
Figure 13: Example of decision screen for Criticality Treatment

F   Additional Tables and Figures

F.1   Comparison of pricing protocols

Figures 14(a) and 14(b) show the snap shots of the criticality treatment at minute 3 and minute 6, respectively. Figures 14(c) and 14(d) show that the dynamics in the betweenness treatment.

Figures 15, 16, 17, 18, and 19 show the time series of various group statistics (average degree, distance in the largest component, degree, inequality, payoff inequality, and efficiency) in comparison with the theoretical predictions from the computer simulations described in Section B.

F.2   Network Centrality Measures

We consider two standard measures of network centrality: closeness and degree centrality.

Closeness centrality of a player i in a given network g captures how close i is from all other players. Formally, it is calculated as $C_c(i;g) = \sum_{j \neq i} \frac{n-1}{d(i,j;g)}$. Closeness centrality of a network g with n players, as used in Figure 20, corresponds to $C_c(g) = \frac{\sum_{i=1}^{n}[C_c^{max}(g) - C_c(i;g)]}{(n-2)(n-1)/(2n-3)}$ where $C_c^{max}(g) = \max_i C_c(i;g)$. Average closeness centrality of networks realized across different treatments in the experiment is depicted in Figure 20.

Degree centrality of a player i in a given network g captures the fraction of links realized by i. Formally, it corresponds to $C_d(i;g) = \sum_{j \neq i} \frac{\eta_i(g)}{n-1}$. Degree centrality of a network g with n players, as used in Figure 21, corresponds to $C_d(g) = \frac{\sum_{i=1}^{n}[C_d^{max}(g) - C_d(i;g)]}{n-2}$ where $C_d^{max}(g) = \max_i C_d(i;g)$. Average closeness centrality of networks realized across different treatments in the experiment is depicted in Figure 21.

F.3   Alternative Inequality Measure

Figure 22 depicts an alternative to the inequality measure presented in the main text, based on the well known Lorenz curve and Gini coefficient.

F.4   Explanations of Macroscopic Patterns

Figures 23 and 24 show the distribution of the two types of benefits that can be obtained in the experiment: access benefit represents the gains from trading with others in the network (through being connected to them), and brokerage rent represents the gains from acting as

(a) Criticality $N = 100$ at minute 3

(b) Criticality $N = 100$ at minute 6

(c) Betweenness $N = 100$ at minute 3
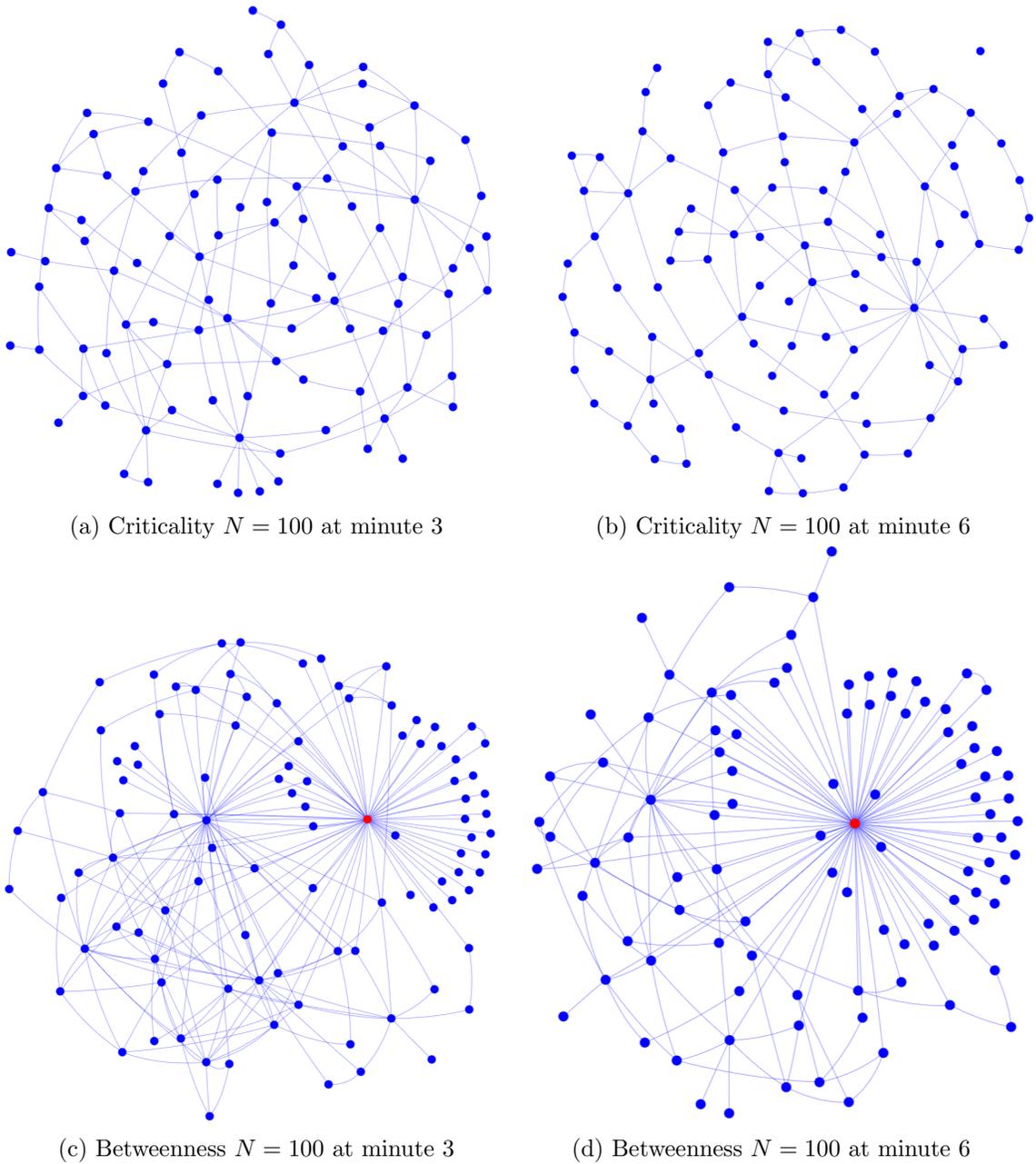
(d) Betweenness $N = 100$ at minute 6

Figure 14: Snap shots of the Experiment

an intermediary between any pair of traders in the network. Figure 23 shows that access benefits dominate to generate earnings in the groups. Moreover, we see a significantly
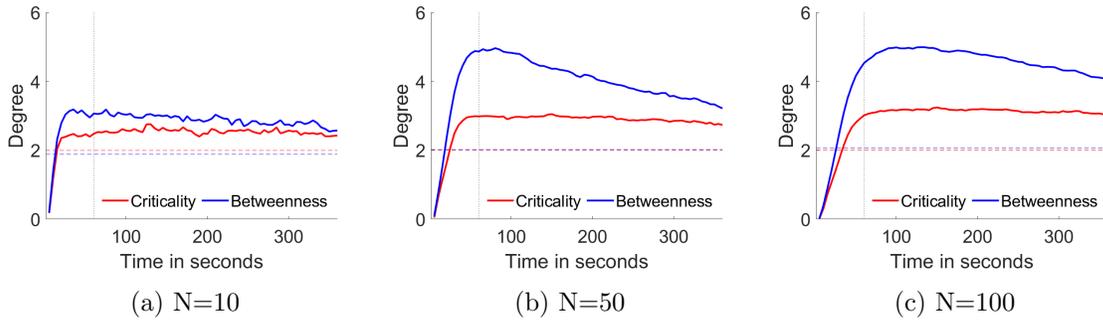
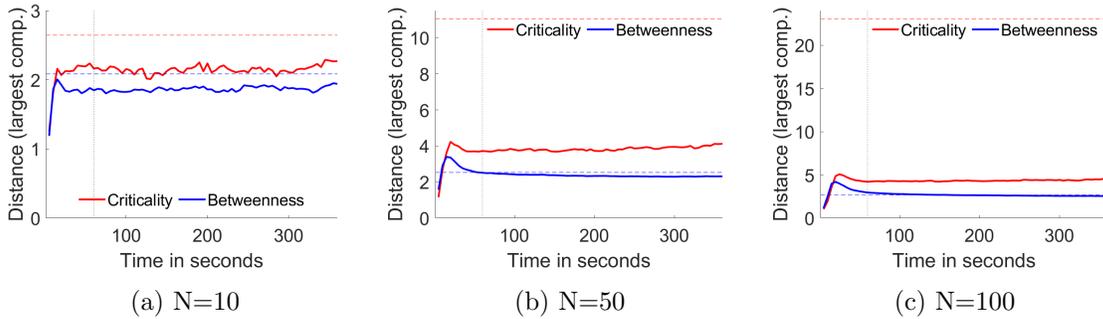Figure 15: Mean degree (dotted lines represent the expected values predicted by the computer simulations)

Figure 16: Mean distance (dotted lines represent the expected values predicted by the computer simulations)
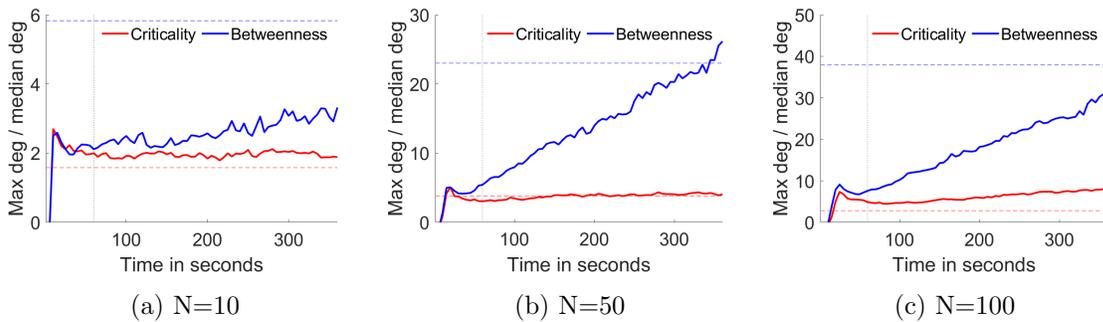
Figure 17: Mean degree inequality (dotted lines represent the expected values predicted by the computer simulations)
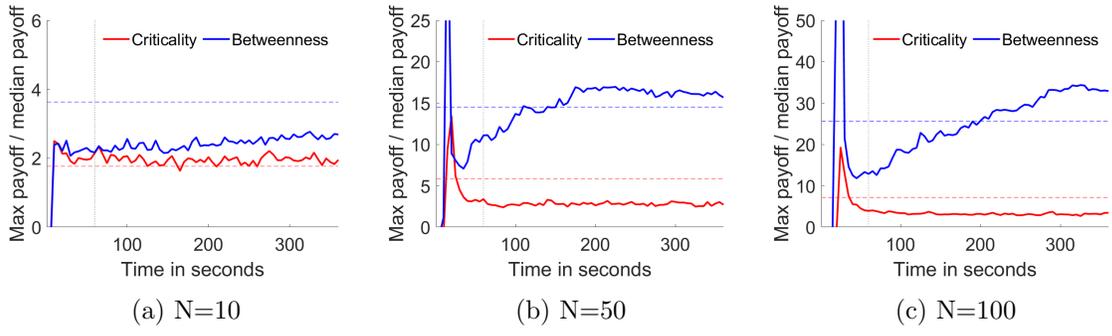
Figure 18: Mean payoff inequality (dotted lines represent the expected values predicted by the computer simulations)
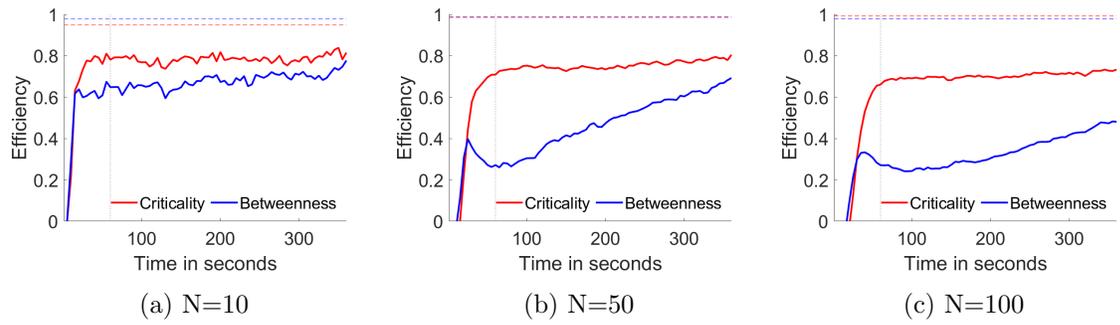
Figure 19: Mean efficiency (dotted lines represent the expected values predicted by the computer simulations)
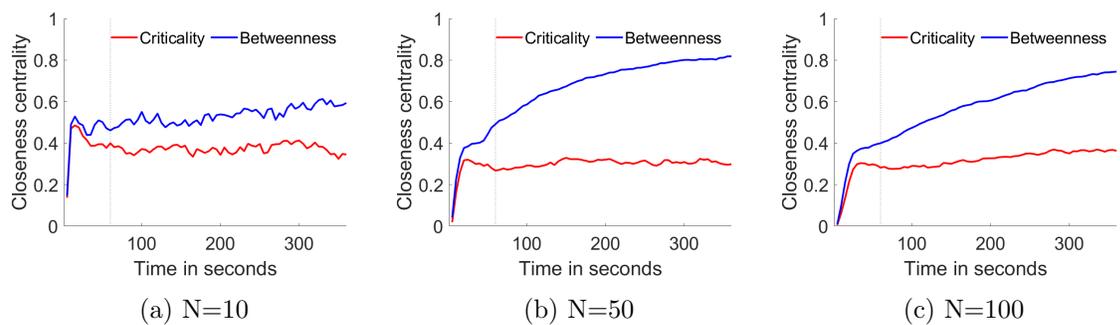
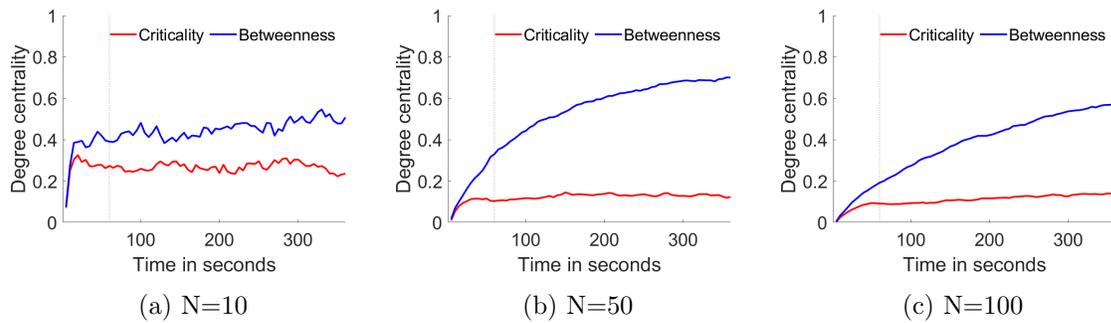Figure 20: Closeness Centrality in the Experiment
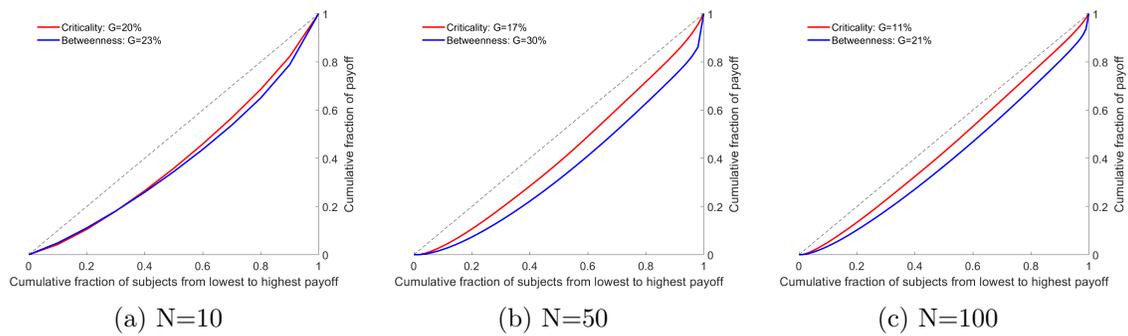
40

Figure 21: Degree Centrality in the Experiment

Figure 22: Lorenz curve and Gini Coefficient

larger fraction of brokerage rent under the betweenness pricing. Figure 23 further shows different distributions for the most connected individuals, especially in the large groups where individual earnings are mostly obtained through brokerage rents under betweenness pricing. This indicates that brokerage rent is the source of payoff inequality observed in those treatments.
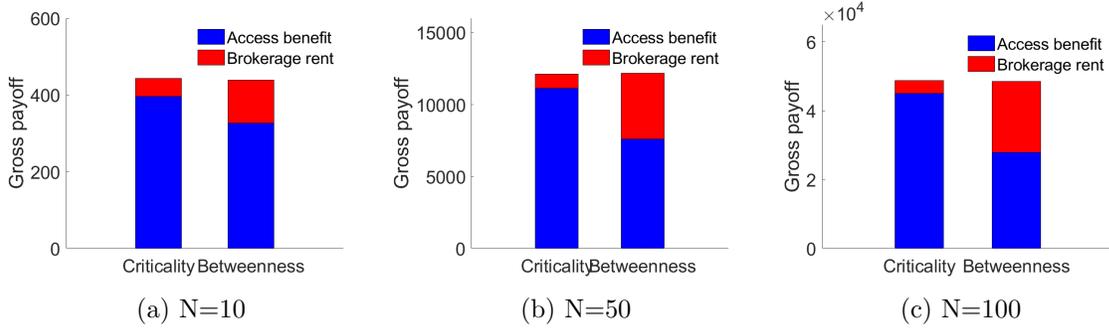

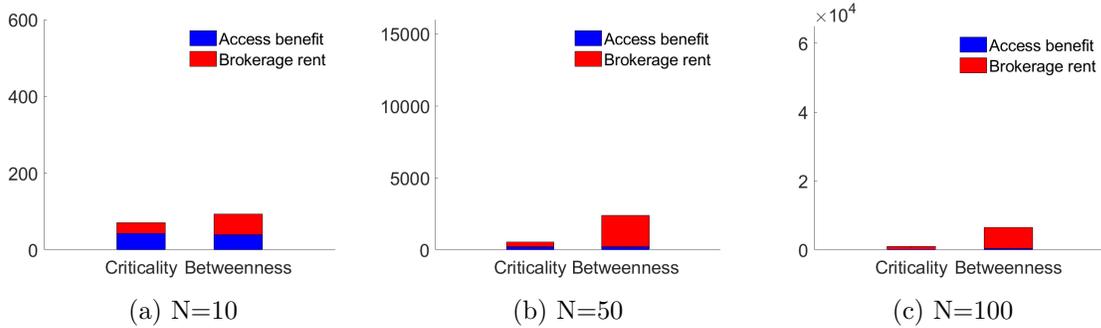Figure 23: Distribution of gross payoffs: Sum of individual benefits

Figure 24: Distribution of gross payoffs: Maximum individual benefits

F.5   Regression table

F.6   Individual Decision Rule

In order to understand how subjects update their link proposals across the different treatments of the experiment, we consider a simple model of myopic best response with noise. The idea of noisy best responses is that the probability that player i selects a link proposal vector s_i^t at time t is proportional to the material payoffs associated with s_i^t, conditional

	Distance	Payoff inequality	Degree inequality	Efficiency	Degree
$N = 50$	1.689***	0.850***	1.876***	-0.030***	0.372***
	(0.037)	(0.037)	(0.037)	(0.037)	(0.037)
$N = 100$	2.193***	1.284***	4.226***	-0.081***	0.597***
	(0.044)	(0.044)	(0.223)	(0.016)	(0.063)
$(N = 10) \times$Between	-0.282***	0.499***	0.677***	-0.103***	0.332***
	(0.029)	(0.137)	(0.319)	(0.023)	(0.065)
$(N = 50) \times$Between	-1.486***	12.558***	11.239***	-0.269***	1.150***
	(0.037)	(0.581)	(0.993)	(0.006)	(0.030)
$(N = 100) \times$Between	-1.681***	22.816***	12.689***	-0.364***	1.506***
	(0.063)	(2.450)	(1.222)	(0.020)	(0.085)
Obs.	36,000	35,998	36,000	35,974	36,000
R-squared	0.917	0.763	0.617	0.672	0.751

Notes: Standard errors clustered at the group level (24 clusters in total) are shown in parentheses. *** represents significance level 1%. All regressions include a list of dummies for time seconds (time fixed effects).

Table 8: Panel Regression Results of Group Outcomes

on the link proposal matrix by the others at time $t - 1$ denoted by s_{-i}^{t-1}.

Let s_i^t denote a vector of link proposals that can be made by subject i at time $t > 1$ such that $s_{ij}^t \neq s_{ij}^{t-1}$ for some $j \neq i$ and $s_{ik}^t = s_{ik}^{t-1}$ for all $k \neq j$. Let S_i^t be the set of all such vectors s_i^t for subject i at time $t-1$. At any time $t > 1$ when subject i makes a choice, we assume i selects a link proposal vector $s_i^t \in S_i^t$ with the following choice probability, conditional on a link proposal matrix s_{-i}^{t-1} chosen by all other subjects at time $t - 1$:

$$\Pr(s_i^t | s_{-i}^{t-1}) = \frac{\exp\left(\lambda \Pi_i\left(s_i^t, s_{-i}^{t-1}\right)\right)}{\sum_{h_i^t \in S_i^t} \exp\left(\lambda \Pi_i\left(h_i^t, s_{-i}^{t-1}\right)\right)}$$

where $\Pi_i(s) = \Pi^{crit}(s)$ or $\Pi_i(s) = \Pi^{btwn}(s)$ denote the normalized payoff function,[11] and $\lambda \geq 0$ represents the degree of noise: when $\lambda = 0$, subjects' choices are purely random; when $\lambda \to \infty$, subjects choose optimally with no decision error. Note that the above

---

[11] As the scale of individual payoffs grows significantly with group sizes, we normlize them through the conversion rates that were used in the experiment to obtain similar average earnings across treatments. More precisely, this normalization consists of dividing individual payoffs by 20 for $N = 10$, 110 for $N = 50$, and 220 for $N = 100$

decision rule rules out the choice of "doing nothing", as our focus is on understanding how subjects update their connections in the network.

Let the experimental data be organized with a finite number of group outcomes at the round level r, $X_{rt} \equiv \left(x_{rt}^1, x_{rt}^2, ..., x_{rt}^m \right)$, over time $t = 1, ..., 360$. We consider three main outcomes at the group level (i.e., $m = 3$):

- Connectedness defined by ratio of pairs connected with each other;

- Average degree per individual;

- Degree inequalilty defined by the ratio of maximum degree to median degree.

We aim at explaining the data with the simple model of myopic best response with noise described above. For any link proposal matrix s^{t-1} observed at time $t-1$ in the experiment (for a given group and a given round), we predict the next link proposal matrix at time t (i.e., s^t) by exclusively applying this decision rule to all subjects of the group who actually made a choice at time t in the experiment (i.e., the link proposals associated with inactive subjects at time t remain unchanged). These predictions allow us to generate the stochastic time series of group outcomes, denoted by $\{ Y_t (\lambda) \}_{t=1}^{360} \equiv \left\{ \left(y_t^1 (\lambda), y_t^2 (\lambda), ..., y_t^m (\lambda) \right) \right\}_{t=1}^{360}$.

The method of estimating parameter λ is based on the generalized method of moments (GMM) that minimizes the average distance between X_{rt} and the expectation of $Y_t (\lambda)$, i.e. $E [Y_t (\lambda)]$, over different rounds r at every time t. That is, we consider the following moment condition with a vector-valued $m \times 1$ function at every time t:

$$D_t (\lambda) \equiv E \left[X_{rt} - E \left[Y_t (\lambda) \right] \right] = 0$$

We then take the sample average of this moment condition over different rounds $r = 1, ..., R$, at every time t

$$\widehat{D}_t (\lambda) = \frac{1}{R} \sum_{r=1}^R \left[X_{rt} - E \widetilde{[Y_t (\lambda)]} \right]$$

where $E \widetilde{[Y_t (\lambda)]}$ denotes the average of group outcomes generated from the model with parameter λ with a sufficiently large number of simulations (in practice, we consider 1000 such simulations for each estimation exercise). Note that the gap between $E [Y_t (\lambda)]$ and $E \widetilde{[Y_t (\lambda)]}$ is assumed to vanish as the number of simulations increases.

Using the theory of generalized method of moments, we can find λ that minimizes the

following objective function

$$\sum_{t=61}^{360} \left(\frac{1}{R} \sum_{r=1}^{R} \left[X_{rt} - E\widetilde{[Y_t(\lambda)]} \right] \right)' W \left(\frac{1}{R} \sum_{r=1}^{R} \left[X_{rt} - E\widetilde{[Y_t(\lambda)]} \right] \right)$$

where W is a positive-definite $m \times m$ weighing matrix.

Practically, we use the two-step estimation approach as follows:

- In the first step, we take W as the identity matrix and compute $\widehat{\lambda}_{(1)}$ that minimizes the objective function.

- In the second step, we compute an estimated $\widehat{W}_{(1)}$ based on the first-step estimation $\widehat{\lambda}_{(1)}$:

$$\widehat{W}_{(1)} = \sum_{t=61}^{360} \left(\frac{1}{R} \sum_{r=1}^{R} \left[X_{rt} - E\left[Y_t\widetilde{\left(\widehat{\lambda}_{(1)} \right)} \right] \right] \right) \left(\frac{1}{R} \sum_{r=1}^{R} \left[X_{rt} - E\left[Y_t\widetilde{\left(\widehat{\lambda}_{(1)} \right)} \right] \right] \right)'$$

We then plug $\widehat{W}_{(1)}$ in the objective function

$$\sum_{t=61}^{360} \left(\frac{1}{R} \sum_{r=1}^{R} \left[X_{rt} - E\widetilde{[Y_t(\lambda)]} \right] \right)' \widehat{W}_{(1)} \left(\frac{1}{R} \sum_{r=1}^{R} \left[X_{rt} - E\widetilde{[Y_t(\lambda)]} \right] \right)$$

and compute $\widehat{\lambda}$ that minimizes the above objective function.

Note that our focus is on estimating the noise parameter to predict behavior in the payoff effective period (last 300 seconds) of every round from the experiment.

Given the noisy best response model generates stochastic time series of group outcomes, we approximated the model predictions by averaging the group statistics across 1000 simulations (at any given second, for each parameter estimation). In order to assess the robustness of such estimations, we repeated the procedure for 10 iterations for each treatment.

The resulting parameter estimations across all treatments are reported in Table 9.

In order to compare the performance of the noisy best response model through the estimated parameter λ, we simulated the corresponding dynamics when applying the average values for λ from Table 9, and starting from the link proposal matrices observed at the end of the trial period across our experimental sessions. In such simulations, the set of active

	Criticality			Betweenness		
	$N = 10$	$N = 50$	$N = 100$	$N = 10$	$N = 50$	$N = 100$
λ	3.51 (0.1)	5.35 (0.06)	6.38 (0.03)	2.78 (0.32)	10.04 (0.08)	11.02 (0.27)
Predict. error	62866260806	141194948	4	40944894522	198130157	8
Nb obs./est.	6000	6000	6000	6000	6000	6000
Nb est.	10	10	10	10	10	10

Notes: Standard errors are shown in parentheses. The reported prediction errors (outputs of objective function) are per observation ($\times$1e-18).

Table 9: Estimated noise parameter λ across treatments (average values based on 10 estimations per treatment, 6000 observations per estimation, and 1000 simulations per observation)

players (updating at least one link proposal) at every second in a simulation was perfectly matched with the set of active subjects in a round of the experiment (each simulation was therefore matched with one round and one group of the experiment). Figures 25, 26, 27, 28, 29, and 30 show the resulting dynamics (solid lines) in comparison with the experimental data (dashed lines), across various statistics (link inequality, efficiency, degree, payoff inequality, connectedness, distance).
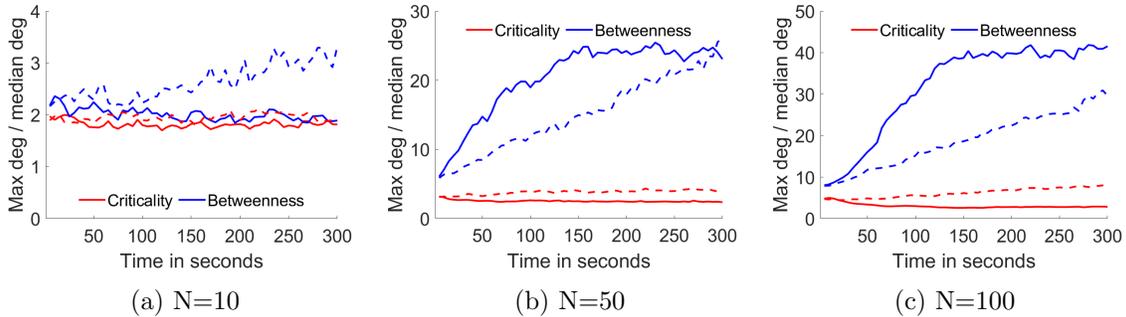

(a) N=10 (b) N=50 (c) N=100

Figure 25: Link inequality (solid lines represent the simulations based on the myopic best response model using the estimated noise parameter *lambda*; dashed lines represent the experimental data)

We observe that the degree of noise (captured by the value of λ) is very similar for both pricing protocols in the small group (t-test: p=0.04 for $N = 10$), but is significant different in the large groups (t-tests: p<0.001 for $N = 50$): subjects exhibit less noise under betweenness pricing, suggesting more consistency with a myopic best response be-
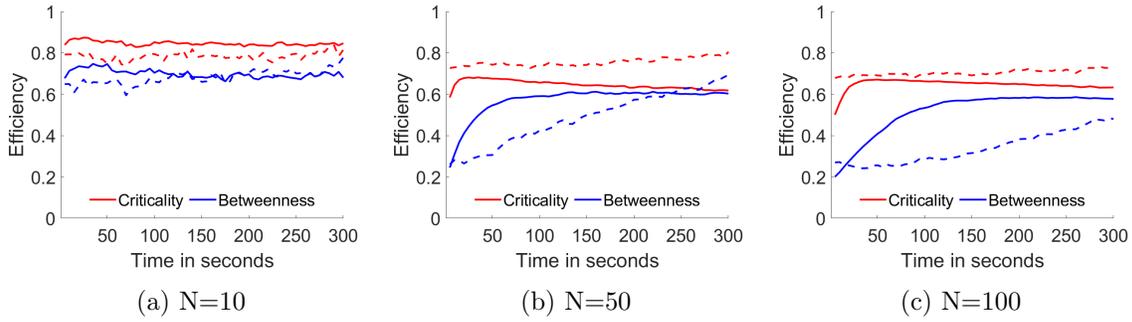
Figure 26: Efficiency (solid lines represent the simulations based on the myopic best response model using the estimated noise parameter *lambda*; dashed lines represent the experimental data)
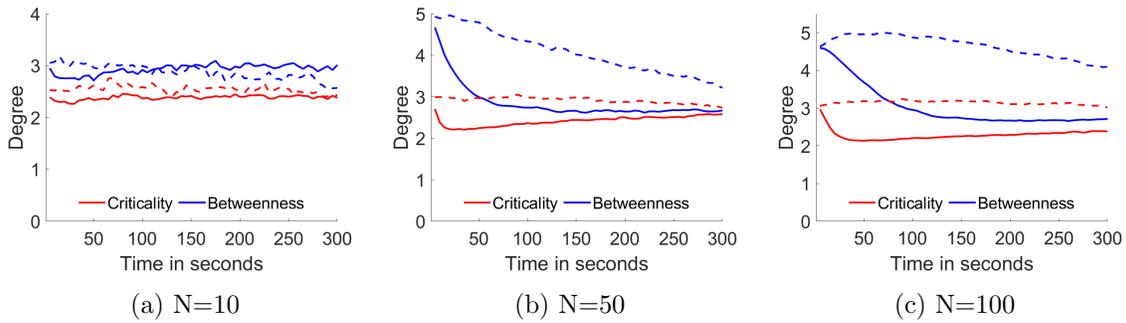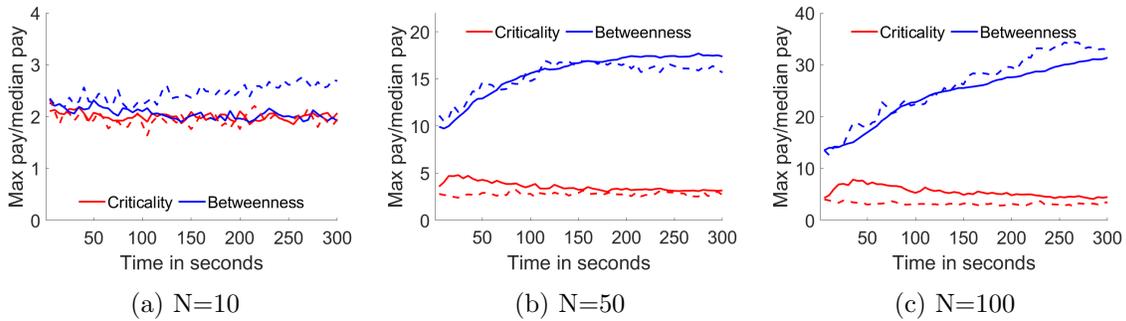

Figure 27: Mean degree (solid lines represent the simulations based on the myopic best response model using the estimated noise parameter *lambda*; dashed lines represent the experimental data)

Figure 28: Payoff inequality (solid lines represent the simulations based on the myopic best response model using the estimated noise parameter *lambda*; dashed lines represent the experimental data)
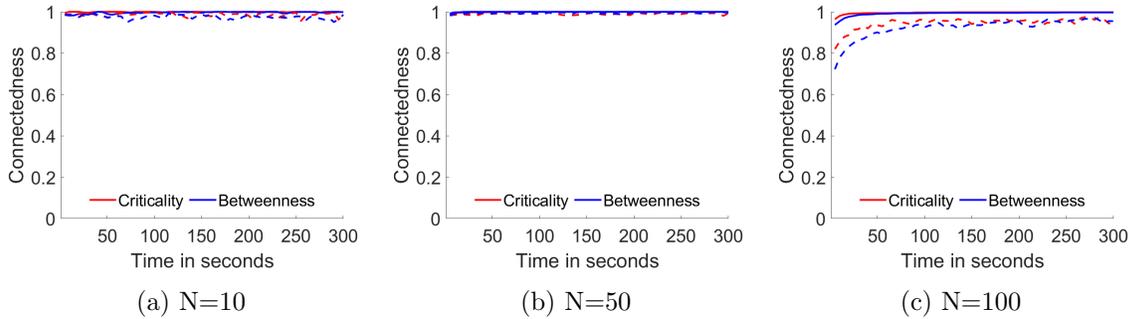
Figure 29: Connectedness (solid lines represent the simulations based on the myopic best response model using the estimated noise parameter *lambda*; dashed lines represent the experimental data)
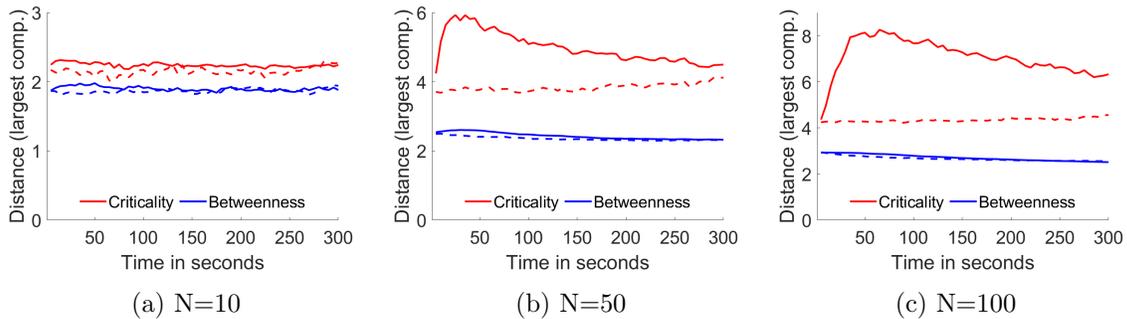

Figure 30: Distance in the largest component (solid lines represent the simulations based on the myopic best response model using the estimated noise parameter *lambda*; dashed lines represent the experimental data)

havior. This results can be explained by the salience of the best response action in this case, which mostly consists in forming a unique link with the most connected individual. By comparing the estimated noise parameter across group sizes for a fixed pricing protocol, this analysis also suggests significantly less random behavior in the larger groups (t-tests comparing any pairs of group sizes: $p<0.001$). This observation is somewhat counterintuitive as more complexity is found in the larger groups (in terms of strategy space, payoff computation, network visualization), making it more difficult for subjects to identify the optimal choices. A possible explanation for this result is that the information of the network structure together with everyone's payoff information, which are provided in real time in the experiment, facilitates the subjects' learning of the payoff consequences to different types of actions (i.e., by looking at others' behavior and payoffs, they can easily identify which actions are more beneficial than others).