Connectors and Influencers

Syngjoo Choi* Sanjeev Goyal† Frédéric Moisan‡

November 12, 2019

Abstract

Empirical networks exhibit inequality in connections. It is important to understand the economic mechanisms that account for this inequality. This paper develops a new experimental platform to study network formation. The platform integrates a network visualization tool using the algorithm of Barnes and Hut [1986] with an interactive tool of asynchronous choices in continuous time. It allows for large groups in excess of 100 subjects.

The platform is used to test the model of linking and efforts in Galeotti and Goyal [2010]. The theory predicts a 'star' network in which the spokes pay for links with a single hub. There are two equilibrium effort configurations: the center makes all the effort (the influencer outcome) and the hub makes zero effort (the pure connector outcome).

The experiments provide robust evidence for specialization in linking and efforts; this is specially striking in large groups. Scale and availability of others' payoff information interact in powerful ways: in a baseline treatment where subjects observe only own payoffs, increasing scale leads to the influencer outcome. When subjects see everyone's payoffs, with increasing scale the outcome is the pure connector outcome. A learning rule that combines myopic best response and connection seeking provides an explanation for these findings.

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. Email: sg472@cam.ac.uk

‡Faculty of Economics, University of Cambridge. Email: fm442@cam.ac.uk

The paper was supported by the FET-Open Project IBSEN (contract no. 662725), the Keynes Fund and by Creative-Pioneering Researchers Program (Seoul National University). We thank Francis Bloch, Olivier Compte, Vince Crawford, Matt Elliott, Philippe Jehiel, Michael Kosfeld, Boris van Leeuwen, Rosemarie Nagel and Stephanie Rosenkranz for helpful comments.

1 Introduction

Large scale social networks are a defining feature of contemporary economy and society. Empirical research suggests that such networks exhibit a *law of the few*: the distribution of links is very unequal.[1] Given the prominence of these networks, it is important to understand the principles underlying their formation.

The economic approach takes the view that individuals compare the costs and benefits of linking. Beginning with the early work of Bala and Goyal [2000] and Jackson and Wolinsky [1996], this idea has been explored in a number of papers on network formation. A general take away from this literature is that purposeful linking activity leads to the 'law of the few'.[2] This result has been the subject of extended experimental investigation: in a test of the Bala and Goyal [2000] model, Falk and Kosfeld [2012] and Goeree, Riedl, and Ule [2009] show that experimental subjects do not create such networks; in a recent paper, van Leeuwen, Offerman, and Schram [2019] report that the specialization in linking and efforts predicted by the Galeotti and Goyal [2010] model is not observed in the laboratory. These experimental findings raise a question mark about the validity of an economic approach to understand networks.

A common element of existing network formation experiments is that the number of subjects is small (typically ranging between 4 and 8). Moreover, practically all the experiments require subjects to make simultaneous choices in discrete time. In a real world setting, groups are very large and individuals typically choose effort and linking at different points in time. The individual decision problem is complicated because the attractiveness of links depends on the efforts of individuals *and* also on the efforts by the neighbours of these individuals. As group size grows, these informational requirements become more demanding. So it is quite unclear if we can extend the findings from the small group experiments to more realistic settings. The work of Berninghaus, Ehrhart, and Ott [2006], Friedman and Oprea [2012] and Goyal et al. [2017] 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 paper builds on this insight.

A large-scale continuous-time experiment on network formation generates a great deal of information that is in principle relevant for decision making. This may be cognitively too

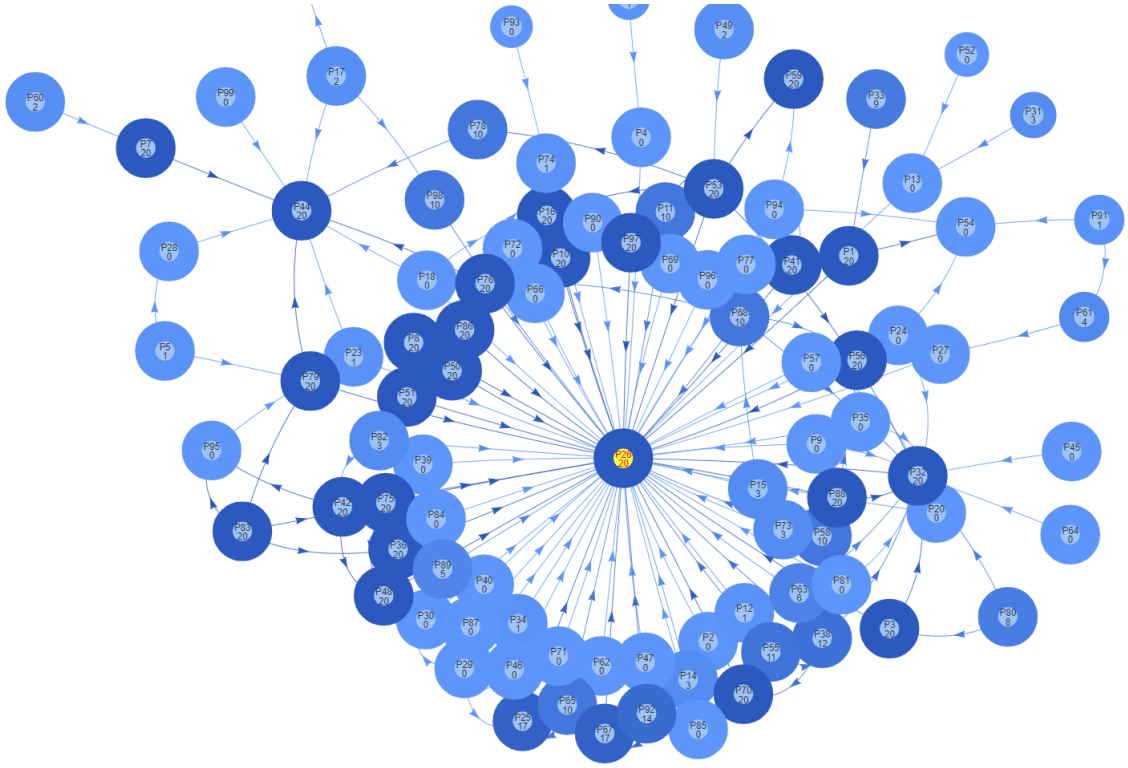
[1] See Barabsi and Albert [1999], Goyal, Moraga, and van der Leij [2006], and Jackson and Rogers [2007].
[2] See e.g., Hojman and Szeidl [2008]; Bramoulle, Galeotti, and Rogers [2016] survey the literature.

demanding for individuals and can undermine the rationale for controlled experimentation. In order to handle such concerns, we develop a new experimental platform. Three aspects of the platform are worth noting. Firstly, it includes a network visualization tool that uses the Barnes-Hut approximation algorithm (Barnes and Hut [1986]). This algorithm allocates nodes and edges in a two-dimensional space to improve visual clarity of network presentation. Secondly, we integrate this tool for network visualization with the interactive tool of dynamic choices. This feature allows individuals to form and remove links and change effort levels instantly. The integration enables us to update rapidly evolving networks in real time on the computer screen. Finally, the platform is flexible in information provision both with regard to what subjects know about the network and what they know about the actions and payoffs of different subjects.

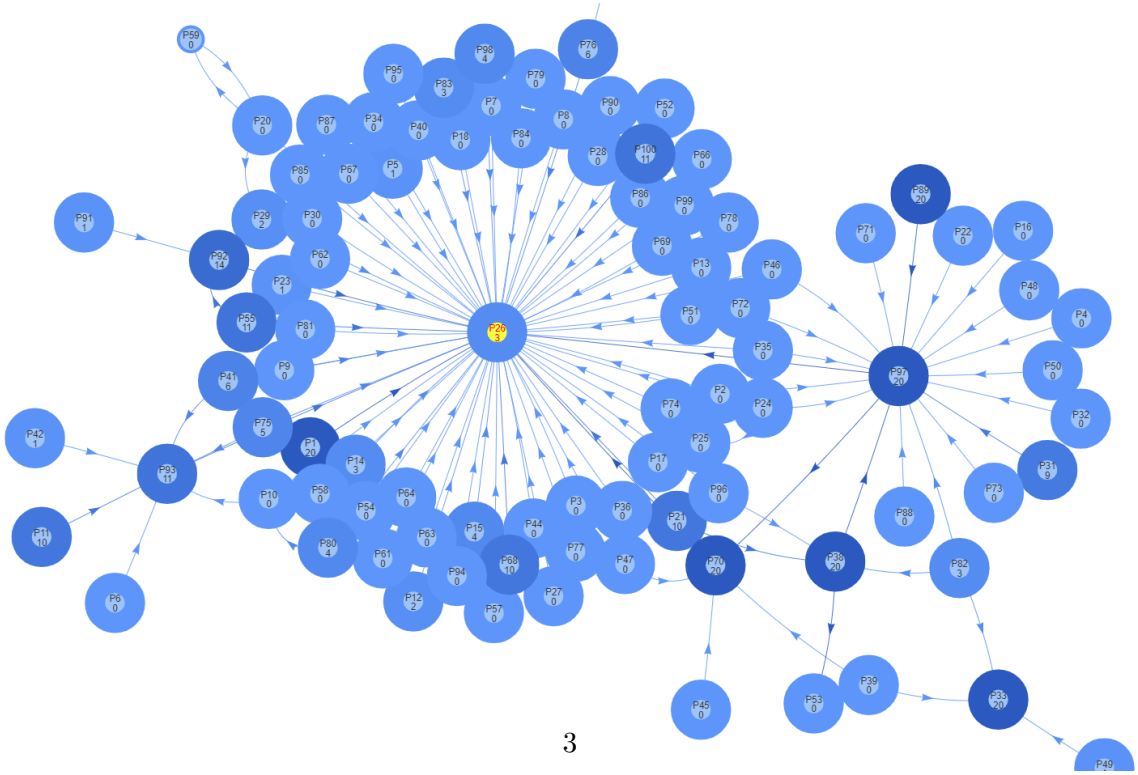
The design of the experiment builds on a model of linking and efforts taken from Galeotti and Goyal [2010]. The theory predicts that every (strict Nash) equilibrium of this game is a 'star' network in which the spokes pay for links with a single hub. There are two equilibrium effort configurations: the center makes all the effort (the pure influencer outcome) and the hub makes zero effort (the pure connector outcome). The goal of the paper is to test these predictions. There are four group sizes $4, 8, 50, 100$ and each of these groups plays the linking and effort game over 6 minutes. There are two information treatments: in the baseline treatment, subjects observe only their own payoffs, while in the payoff information treatment a subject observes the payoffs of everyone. Taken together, we therefore have a 4×2 design. This design enables us to vary the strategic uncertainty and cognitive complexity facing subjects and therefore offers a general environment to test the theory.

We start with the baseline information treatment. Figures 1 and 2 present snap shots taken from the experiment with a hundred subjects. Initially, at minute 1, subject P26 emerges as a hub with the maximum effort 20. There are other subjects who make maximal effort (such as P97). At minute 3, P26 continues to be a hub but has substantially lowered her effort. Due to this shading of effort, she starts to lose some of her links to subject P97, who has kept her effort at 20. The transition becomes clearer in Figure 2a at the 5 minute mark, when the initial hub subject P26 has lost most of her links to the emerging hub P97. Figure 2b confirms that this transition is stable until the end of the game.

Our *first* finding is that, in all four group sizes, there is specialization in linking and efforts. This manifests itself in the clearest form in the large groups (as in the 100 subject experiment reported above). Our *second* finding concerns individual behavior. In all four

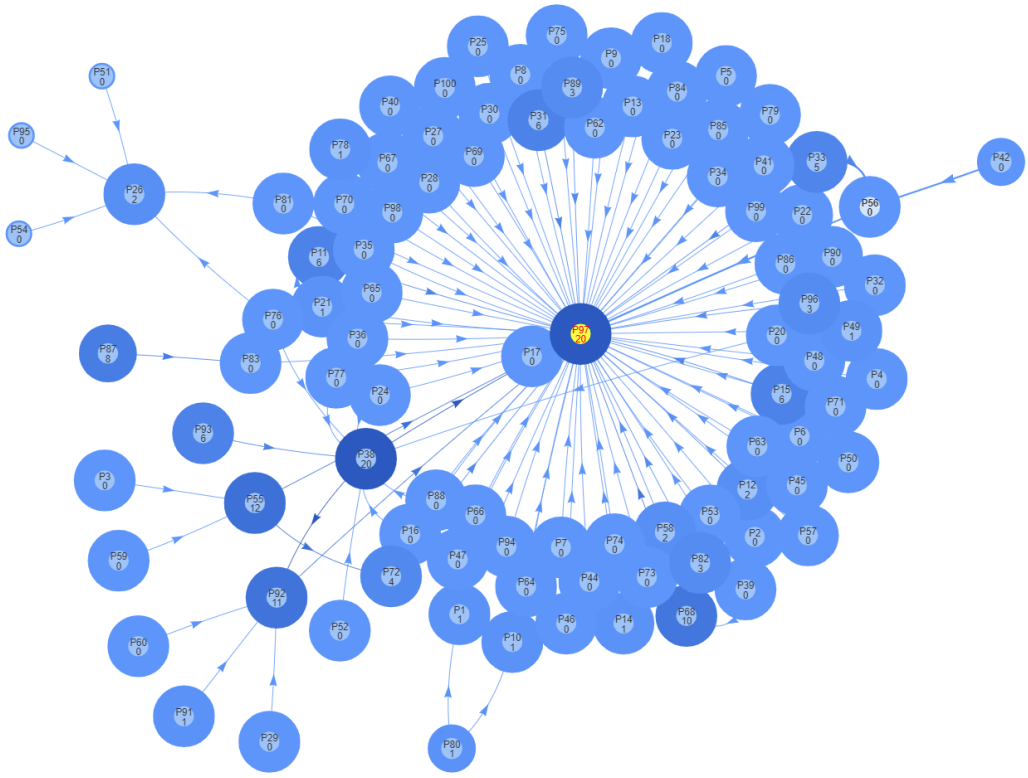


(a) At minute 1

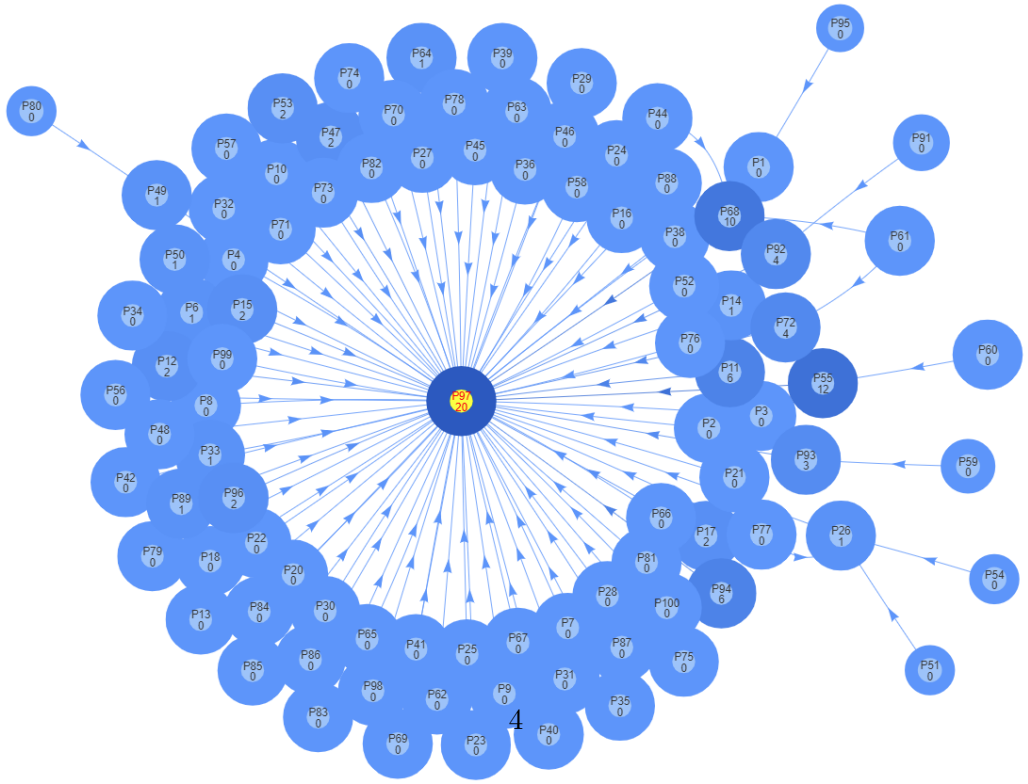


(b) at minute 3

Figure 1: Snap shots of a dynamic game



(a) At minute 5



(b) At minute 6

Figure 2: Snap shots of a dynamic game (cont.)

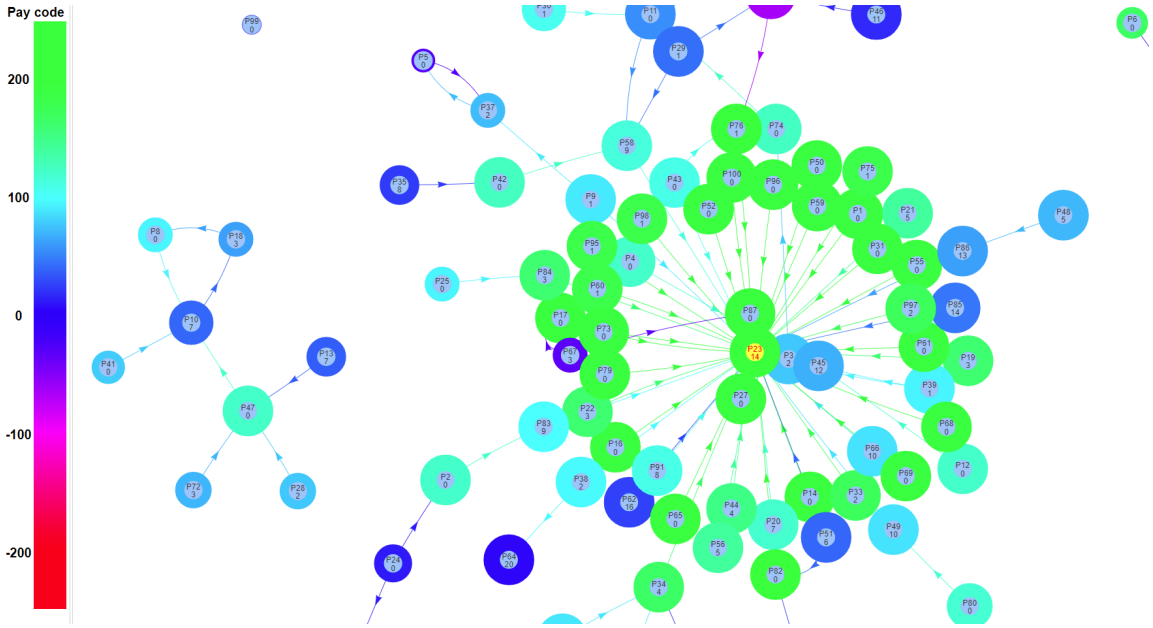
group sizes, highly connected individuals exert large efforts. In particular, in small groups the efforts of the most connected individual are close to the equilibrium prediction of the static model. By contrast, in the large groups of fifty and hundred subjects the most connected subject chose efforts that are far higher than the equilibrium prediction. As a consequence, in the large groups, the hubs earns less than the peripheral nodes.

One possible explanation for the high efforts is that individuals enjoy non-monetary benefits from being a hub, and this incentive is reinforced in larger groups. An alternative hypothesis is that in larger groups, the complexity of the dynamics overwhelms individuals and they are led into large efforts, in spite of the lower payoffs. To examine these explanations we design a treatment in which subjects are shown the payoffs of everyone. The provision of payoff information on everyone facilitates comparison of payoff performances with others. This may make it easier for subjects to understand the payoff implications of their own choices. Availability of such information may also alter individual behavior due to imitation possibilities (Schlag [1998], Huck et al. [1999], and Camerer [2003]).

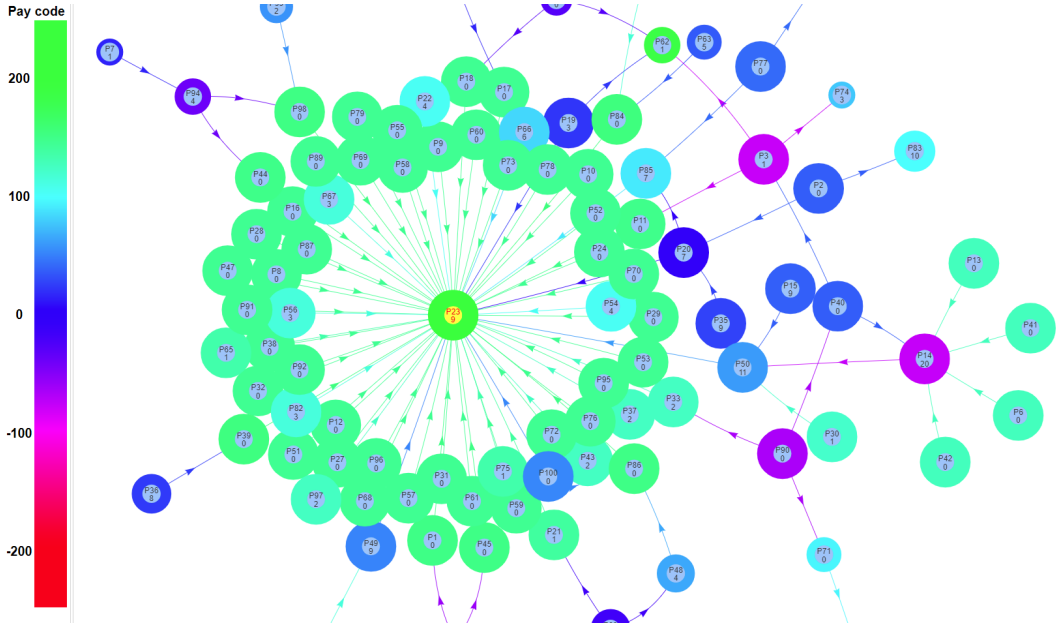
Figures 3 and 4 present snap shots taken in the payoff information treatment with a hundred subjects. Observe that the specialization in linking continues to hold in this setting. However, there is a major change in the behavior of individuals seeking to become a hub: the most connected individual (P23) starts at a high effort 14, but then shades her efforts. The key difference with the baseline is that the outcome is closer to a "pure connector outcome" in some groups and, in most groups, the most connected individual earns much more than the peripheral individuals.

Our *third* finding is that, in all four group sizes, there is specialization in linking and in efforts, and this specialization is more transparent in the large groups. Our *fourth* finding is that, in the payoff information treatment, in small groups there is a strong positive correlation while in large groups there is a weak correlation between connectedness and effort. Indeed, in some large groups the most connected individual puts in 0 effort leading to the pure connector outcome (as in Figures 3 and 4). The pure connector outcome is in sharp contrast to the pure influencer outcome observed in the baseline treatment.

These powerful treatment effects motivate an examination of individual decision rules. We study the behavior of three types of subjects—most connected, 2nd most connected, and the others. The effort dynamics bring out two broad patterns: one, they show large effects of group size and payoff information on the behavior of the two most connected individuals. Two, these Figures also show that other – poorly connected – subjects behave similarly across the group sizes and information treatments: they make low effort that is

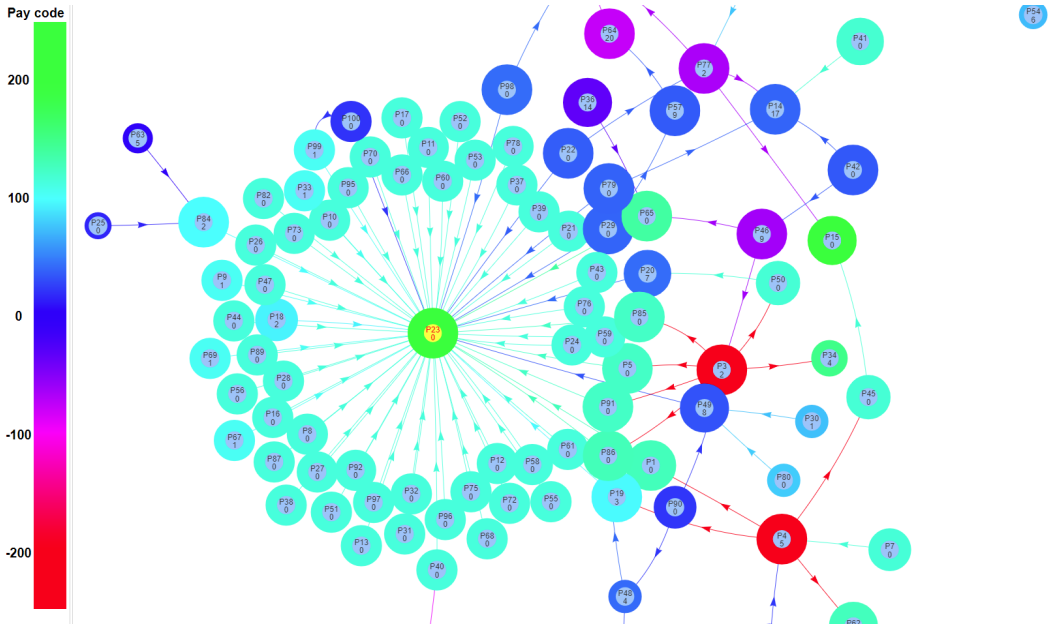


(a) At minute 1

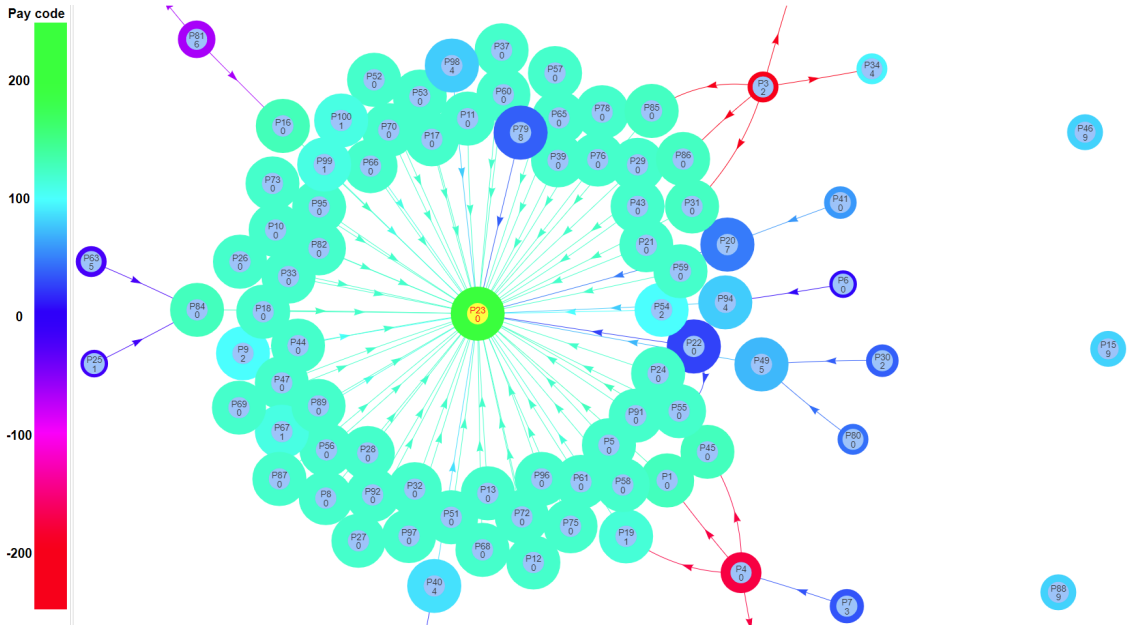


(b) at minute 3

Figure 3: Snap shots with payoff information



(a) At minute 5



(b) At minute 6

Figure 4: Snap shots with payoff information (cont.)

declining over time. We show that the effort dynamics of less connected subjects closely match myopic best response efforts. And we show that the differential effort choices of the two most connected subjects across treatments are well accounted for by a decision rules that combines myopic best response and seeking high degree.

We proceed by classifying individuals into two categories: a group that includes those who choose low efforts and a group that includes those who choose high effort. In order to accommodate the different patterns of effort dynamics observed across the treatments, we propose a learning rule that combines myopic best response and competition for hub status. We use this learning rule to explain the effort behavior of the two most connected individuals in each treatment. It is because the effects of group size and payoff information are mostly on the behavior of the highly connected individuals and the other individuals behave similarly across treatments. We find that the learning rule we propose provides a coherent account of the effects of group size and payoff information on effort dynamics, whereas the myopic best response rule alone performs poorly in fitting effort dynamics. Therefore it suggests that given the heterogeneity in subjects' desire of competing for connection, scale and payoff information change the behavior of highly connected individuals and lead to different patterns of specialization in linking and efforts.

The paper makes two contributions to the economics of networks: one, it presents a new platform for large scale network formation experiments in continuous time; this platform will soon be made public and it should help in advancing experimental study of network questions. Two, we run an experiment on the model of Galeotti and Goyal [2010] on this platform and present robust experimental evidence in support of the 'law of the few' property of networks. Furthermore, we show how information and scale jointly shape the selection of the pure influencer and the pure connector equilibrium, respectively. Our findings mark a major departure from the literature: they offer the first robust evidence in support for a standard economic model of network formation; earlier experimental papers reject this theory, see Falk and Kosfeld [2012], Goeree, Riedl, and Ule [2009] and van Leeuwen, Offerman, and Schram [2019]. The present paper is the first paper to report evidence of the pure connector outcome. Finally, our work reveals that scale has a bearing on the decision rules that individuals use. Taken together, these points make a powerful argument for the use of large scale continuous time experiments in the study of network problems.[3]

[3]There is a also a literature on experiments on games in networks and on games in which players choose partners and then play a coordination game, see e.g., Charness and Sutter [2014], Riedl, Rohde, and Strobel

Our paper also contributes to the experimental literature of continuous time games. Existing studies are built on a development of an experimental software, called ConG (Pettit, Friedman, Kephart, and Oprea [2014]) and have focused on small group interaction (see e.g., Friedman and Oprea [2012]; Calford and Oprea [2017]). The novelty of our paper is that we develop an experimental software that enables us to study large group interaction. In order to overcome information overload of evolving networks and relax subjects' cognitive bounds in information processing, our software integrates the network visualization tool with the interactive tool of asynchronous choices in real time. This is achieved by adopting an enhanced communication protocol between the server and subjects' computers. It allows us to run both network visualization and asynchronous dynamic choices in real time without communication congestion and lagged responses, even when participants are interacting remotely from different physical locations.

2 Theory

We present a model of linking and efforts taken from Galeotti and Goyal [2010].

Let $N = \{1, 2, \ldots, n\}$ with $n \geq 3$. Each player $i \in N$ simultaneously and independently chooses a level of effort $x_i \in \mathbf{R}$ and a set of links g_i with others to access their efforts such that $g_i = (g_{i1}, \ldots, g_{ii-1}, g_{ii+1}, \ldots, g_{in})$, and $g_{ij} \in \{0, 1\}$ for any $j \in N \backslash \{i\}$. Let $G_i = \{0, 1\}^{n-1}$. We define the set of strategies of player i as $S_i = \mathbf{R} \times G_i$, and the set of strategies for all players as $S = S_1 \times \ldots \times S_n$. A strategy profile $s = (x, g)$ specifies efforts and the links made by every player. Observe that g is a directed graph; the closure of g is an undirected network denoted by $\bar{g}$ where $\bar{g_{ij}} = \max(g_{ij}, g_{ji})$ for every $i, j \in N$. The undirected link between two players reflects exchange of benefits from efforts. Let $\eta_i(g) = |\{j \in N : g_{ij} = 1\}|$ be the number of links i has formed. 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 $\bar{g}$. If no such path exists, the distance is set to infinity. Define $N_i^l(\bar{g}) = \{j \in N : d(i, j; \bar{g}) = l\}$ to be set of players at distance l from i in $\bar{g}$.

[2016] and Kearns, Judd, and Vorobeychik [2012]. The interest there is on how networks effect behavior and on how allowing for endogenous networks affects this behavior. The focus of the current paper is quite different.

Given a strategy profile $s = (x, g)$, the payoffs of player i are:

$$\Pi_i(x, g) = f(x_i + \sum_{l=1}^{n-1} a_l(\sum_{j \in N_i^l(\bar{g})} x_j)) - cx_i - \eta_i(g)k \tag{1}$$

where c denotes the constant marginal cost of efforts, k the cost of linking with another player, and a_l reflects the spillover across players who are at distance l. So if $j \in N_i^l(\bar{g})$, then the value of agent j's effort to i is given by $a_l x_j$. Throughout, it is assumed that $a_1 = 1$, $a_2 \in (0, 1)$, and $a_l = 0$, for all $l \geq 3$. The benefit function $f(y)$ is twice continuously differentiable, increasing, and strictly concave in y. For simplicity, also assume that $f(0) = 0$, $f'(0) > c$, and $\lim_{y \to \infty} f'(y) = m < c$. Under these assumptions there exists a number $\hat{y} \in X$ such that $f'(\hat{y}) = c$.

There are no general equilibrium characterization results available for this model[4] The following result characterizes equilibrium when linking costs are relatively large.

Proposition 1. *Suppose payoffs are given by (1), and that $a_1 = 1$, and $a_2 \in (0, 1)$. Then there exists a $\hat{k}$, such that for $k \in (\hat{k}, c\hat{y})$ the following is true. The equilibrium network is a periphery sponsored star. There exist two possible effort equilibrium configurations:*

- *the pure influencer outcome: the hub invests $\hat{y}$ and everyone else invests 0.*
- *the pure connector outcome: the hub invests 0 and everyone else invests $\hat{y}/(1 + (n - 2)a_2)$.*

Proof. The first step is to observe that in equilibrium every individual must access at least $\hat{y}$. This is true because if someone is accessing less than $\hat{y}$, then due to the concavity of the $f(.)$ function, she can simply increase her utility by raising effort so that the total access equals $\hat{y}$.

The second step is to show that players will form one link or zero link, for sufficiently large linking costs. Observe that an isolated individual will choose $\hat{y}$. So it follows that in a network with connections, no one will ever choose more than $\hat{y}$. Note that if link costs are close to $c\hat{y}$ then it is not profitable to form links with two individuals who each choose $\hat{y}$. So the only situation in which an individual, A, may choose two or more links arises if an individual accesses significantly more than $\hat{y}$ through each link. Consider a link between

[4]The analysis of Galeotti and Goyal [2010] focuses on polar cases in which $a_1 = 1$ and $a_l = 0$, for all $l \geq 2$ and the case where $a_l = 1$, for all l. Our formulation allows for indirect flow of benefits with decay; this appears to be a natural case.

A and B. Iterating on optimal effort, it is true that if B chooses $\hat{y}$ then every neighbor of B must choose 0. So A accesses more than $\hat{y}$ only if B chooses strictly less than $\hat{y}$. If a neighbour of B chooses a positive effort then it must be the case then this person must meet the first order condition on optimal efforts: her total efforts invested and accessed must equal $\hat{y}$. As this person is a neighbour of B, it follows that A cannot access more than $\hat{y}$ via the link with B. So, A will form at most one link in equilibrium.

The third step considers effort configurations. Take the situation in which some individual (say) A chooses $\hat{y}$. It is optimal for everyone else to choose effort 0 and form a link with this person. And it is clearly optimal for A to choose $\hat{y}$ faced with zero efforts by everyone else.

To conclude the proof, we need to show that the pure connector outcome is the only possible equilibrium in a situation where no player chooses $\hat{y}$. Observe first that the pure connector outcome is an equilibrium so long as $k < c\hat{y}(n-2)a_2/(1+(n-2)a_2)$. Observe that $c\hat{y}(n-2)a_2/(1+(n-2)a_2)$ converges to $c\hat{y}$, as n gets large.

The next step is to rule out any other possible equilibrium. The key observation here is that any equilibrium network must have diameter less than or equal to 2. Suppose the diameter of a component is 3 or more. We know from step 2 that the component must be acyclic. So consider two furthest apart leaf nodes. A variant of the 'switching' argument, developed in Bala and Goyal [2000], shows that one of the two leaf players have a strict incentive to deviate. So every component must have diameter 2. Given that the network is acyclic, this implies it must be a star. It is now possible to apply standard agglomeration arguments to deduce that multiple components cannot be sustained in equilibrium.

Finally, the hub player must choose zero. Suppose not. By hypothesis the hub chooses less than $\hat{y}$. Given that a_1 and $a_2 < 1$, both the hub and the spokes cannot be accessing exactly $\hat{y}$. A contradiction that implies that the hub must choose zero effort.

□

In the pure influencer equilibrium, we witness an extreme version of the 'law of the few': a single person receives all the links formed in society and also carries out all the efforts. The pure connector equilibrium retains the specialization in links: a single person receives all links, but the efforts are evenly spread out. Interestingly, in both equilibria the creation of links is basically egalitarian – $n-1$ players each form one link. For large k values, the payoff distribution is only slightly unequal in the pure influencer equilibrium. However, the payoff inequality can be very large in the pure connector equilibrium (especially if k

is large and a_2 is small). We note that the pure connector equilibrium holds only for a sufficiently large group size n, i.e., $n \geq 2 + k/(a_2(c\hat{y} - k))$.

We now specify the parameters used in the experiment. The function $f(.)$ is taken from Goyal et al. [2017].

$$f(y) = \begin{cases} y(29 - y) & \text{if } y \leq 14 \\ 196 + y & \text{else} \end{cases} \tag{2}$$

For simplicity, the efforts are assumed to take on integer values only and there is an upper bound, $\overline{x} = 20$. So the efforts set is given by $X = [0, 20]$. The cost of effort $c = 11$ and the cost of a link $k = 95$; finally, the decay parameter $a_2 = 1/2$. Given these parameters, it can be checked that $\hat{y} = 9$.

There exists a pure influencer equilibrium in which a single individual chooses 9, all other individuals choose 0 and form a link with the positive effort player. In principle, there exists a pure connector equilibrium in which the periphery players each choose $18/n$, for any $n \geq 50$.[5] Given the integer constraints, this equilibrium is no longer feasible (for $n \geq 50$, $0 < 18/n < 1$ is not an integer). In the treatments with 50 and 100 subjects, the periphery sponsored star where 18 peripheral individuals choose 1 and the rest of the subjects choose 0 constitutes an 'approximate' equilibrium (for details see Online Appendix A).[6] Figure 5 illustrates the pure influencer equilibrium and the pure connector (approximate-)equilibrium.

To summarize, in the pure influencer equilibrium, the hub chooses effort 9, while the spokes choose 0. The hub earns 81, while the spokes each earn 85. In the pure connector equilibrium, the hub chooses effort 0, eighteen spokes choose 1 each, while the other spokes choose 0. The hub earns 198, the active spokes 74, and the inactive spokes 85.

[5] The pure connector equilibrium does not hold in the experimental setting for any $n < 50$.

[6] The periphery player who chooses effort 1 and forms a link with the hub earns 79.25. This person could earn 81 by deleting the link and instead choosing effort level 9.

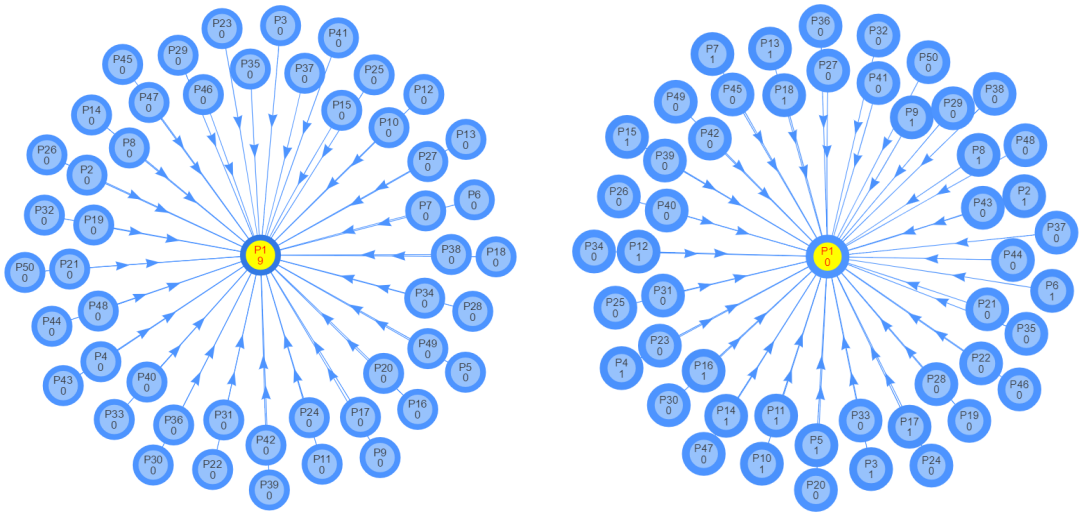


Figure 5: Pure influencer and pure connector equilibrium, $n = 50$

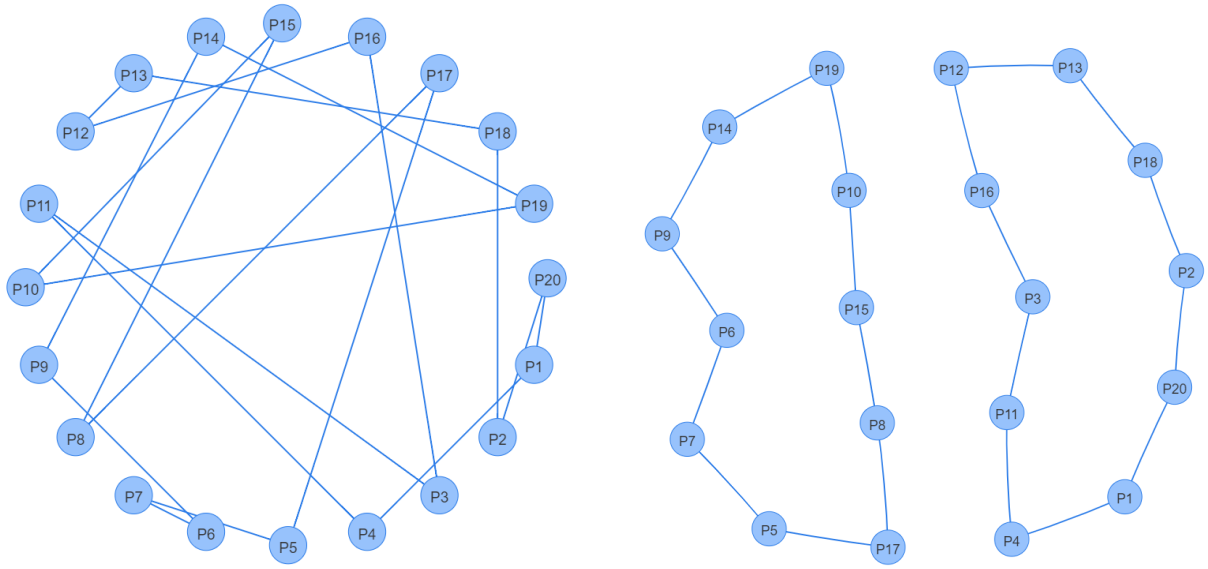
3 Experiment

3.1 Challenges and methodology

As the complexity of subjects' decision making increases in scale, large-scale experiments on network formation pose several major challenges. This section discusses these challenges and explains how our experimental software and design address each of them.

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 (e.g., Goyal et al. [2017]; van Leeuwen et al. [2019]). When the group size increases, such a representation of networks with fixed positions of nodes make 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 6a; the exact network is barely perceptible by observing this figure. The same network structure can be represented in a transparent manner in Figure 6b.

For subjects to learn their optimal choices, they must have a good idea of the evolving networks. An appropriate tool for visualizing networks 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



(a) Fixed visualization

(b) Adaptive visualization

Figure 6: Examples of network visualization

itself based on the evolving structure. We use the Barnes-Hut approximation algorithm [Barnes and Hut, 1986] for grouping nodes in a network that are sufficiently nearby and adjust their relative positions on the subject's computer screen. This algorithm enables us to apply repulsion forces between nodes so that they are sufficiently separated from one another, attractive forces to nodes that are directly linked with each other, and gravity to all the nodes with respect to a central origin on the screen such that nodes not linked with each other remain within reasonable distance from each other.

The network visualization in Figure 6b was made using this algorithm. 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. More details regarding the specifics of this visualization tool (including model parameters characterizing attraction and repulsion forces) can be found in Online Appendix B. It is important to emphasize 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*.

Learning and dynamics. It is important to offer subjects adequate opportunities to learn about the environment of decision making, other subjects' behaviors, and how to respond optimally to them. In view of the strategic complexity alluded to above, the issues of learning and behavioral convergence are particularly complicated. To address them, we build on the work of Berninghaus et al. [2006], Friedman and Oprea [2012] and Goyal et al. [2017], and run the experiment in continuous time with near real time updating—of all actions and linking by everyone.[7]

In our experiment, the game is played in continuous time for 6 minutes during which every subject is free to asynchronously adjust their actions of efforts and linking. 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. The structure of the experiment is publicly known to subjects.

Running the continuous time experiments in large groups poses a number of technical challenges. First, every action made by a subject on her computer must be updated instantly on the computer screens of all other participants through the server computer. Network visualization must be also correspondingly updated in real time. As the group size increases, the information flows across the computer network increases dramatically. This can cause communication congestion and lagged responses. Another challenge with a large scale experiment is that it is constrained by the limited capacity of existing laboratories. Large groups that cannot fit into a single lab therefore require remote interactions between subjects in different geographical locations (that is, across different labs). In order to handle both of these technical challenges, we use a Websocket protocol with enhanced two-way communication between the server and subjects' computers. It fits naturally into the environment of asychronous choices in real time and the updates are made only when necessary. Our Websocket technology relies on the Javascript run-time environment *Node.js*. [8]

[7] Although the experimental software allows for real time updating of actions, we voluntarily introduce some latency in our experiment to avoid any possible confusion caused by some overload of activity on the subjects' screen. More precisely, the network depicted on any subject's screen is updated every 5 seconds or whenever the subject makes a decision.

[8] Since it only requires an internet connection and is compatible with most existing web browsers (e.g.,

Network information. In addition to the issue of network visualization, there is the issue of network information available to individual subjects. To get a sense of the range of possibilities, consider two extreme scenarios: one, subjects only observe their own neighbors in the current network, and two, subjects get to see the entire network. The information and cognitive load implied by the latter scenario grows rapidly in size of the group. In view of this potential trade-off between transparency of network change and information and cognitive overload, we choose to inform each subject of a local structure of the network within a (geodesic) distance 3.

So given a fixed network, for every subject, we can partition the entire group of subjects into two mutually exclusive subgroups: those who are located within distance 3 from the subject, and those who are located outside this set. Figure 7 provides an illustration of network visualization and information in the experiment with 50 subjects. The left side of Figure 7 shows the group of subjects within distance 3 (and all their links with other subjects within distance 3). The right side of Figure 7 collects the subjects who lie at a distance greater than 3. Observe that in addition to local network information, subjects are informed about every subject's effort—presented as a number within the corresponding node along with that subject's ID. A node's total access to public goods is captured by the size of that node.

Information on Payoffs. We now turn to information on payoffs: clearly subjects need to be able to see their own payoffs in order to learn the profitability of different linking and effort combinations.[9] What about the information on others' payoffs?

The literature of learning in games provides some guidance on this question, see Camerer [2003] for a survey. In adaptive models such as reinforcement learning and experience-weighted attraction learning (Camerer and Ho [1999]), players ignore information on payoffs of other individuals. In models of imitation learning (Schlag [1998]) and sophisticated learning (Camerer et al. [2002]), players would behave differently if the payoffs of others are known. In the recent body of network experiments (e.g., Goeree et al. [2009] and Falk and Kosfeld [2012]), researchers have tended not to show subjects the payoffs of others. However, when information on others' payoffs is available in particular in large groups where it is difficult to infer such information, subjects may follow a different behavioral rule. In

Google Chrome, Mozilla Firefox, Internet Explorer), this technology makes no specific restriction on the physical location of every participant.

[9]Details about the costs and benefits are provided to the subjects to facilitate their comprehension of their own payoff, as illustrated in Online Appendix D

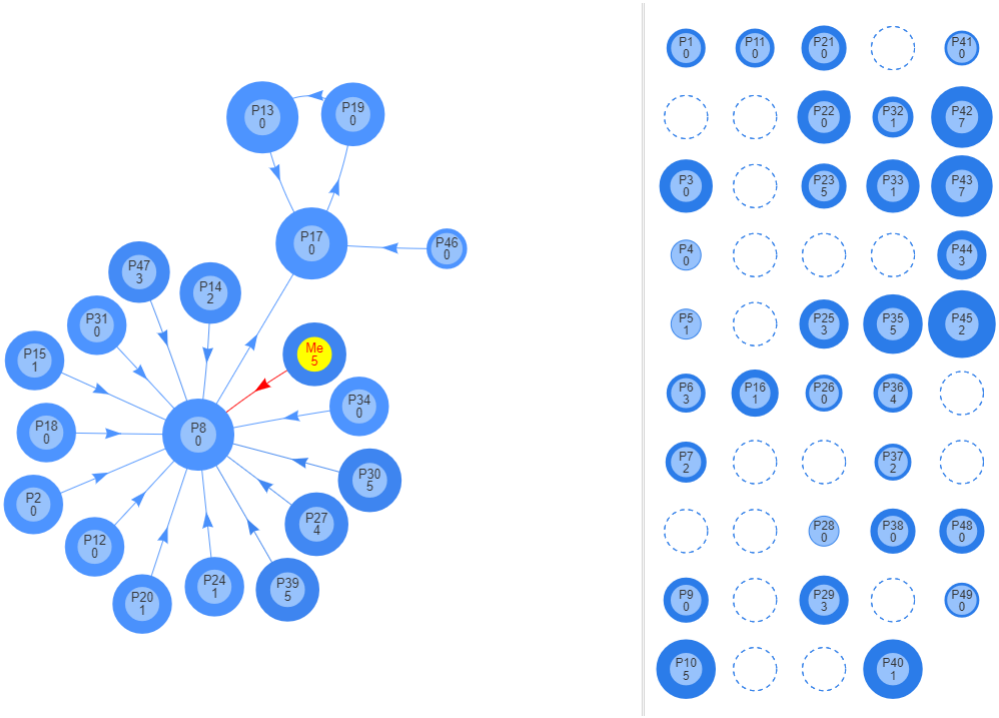


Figure 7: Network information

fact, the experimental literature documents that human subjects may behave differently when information on the payoffs of other individuals is available (e.g., Huck et al. [1999]).

Building on these strands of research, it is possible to argue that in games with small groups of subjects, showing the payoffs of others may not be a first order issue, as subjects can compute these payoffs themselves in a fairly straightforward manner. However, in a dynamic game with a hundred subjects—and with the network and efforts configuration constantly evolving—an individual may find it much harder to compute the payoffs of other subjects. The knowledge of others' payoffs may be an important factor in experimental design. The first reason is learning dynamics: observing the others' payoffs could assist subjects in better appreciating the trade-offs associated with different courses of action. The second reason is fairness considerations: the two equilibria described in Proposition 1 exhibit very different level of payoff inequality across players. The pure-influencer equilibrium exhibits a minor payoff difference between the hub player and the spoke players, whereas the pure-connector equilibrium yields a much larger payoff difference between the hub player and the spokes players. These considerations motivate treatments in which we vary the level of information on others' payoff.

In the baseline treatments subjects are shown their own payoffs but *not* others' payoffs. A subject is also shown the efforts and public good access for all other subjects, as shown in Figure 7. In principle, therefore, a subject can infer the gross payoffs of any subject.. But we believe that such inference would be challenging for subjects during a large scale continuous-time game, where the network and effort levels are evolving rapidly. To facilitate learning, we add information about every player's payoff through a set of color codes as illustrated by Figure 8. Specifically, the border of every node is coloured: the colour varies from green (high positive payoff) to red (high negative payoff). The scale of the colour code is presented at all times on the left hand side, as in Figure 8.

3.2 Treatments and design details

We vary the group size $N \in \{4, 8, 50, 100\}$ and the visibility of others' payoff. Table 1 summarizes the 4×2 structure of our experiment.

All the treatments are based on the payoff function (1). Recall that the marginal cost of effort is set to $c = 11$. This implies an optimal effort of $\hat{y} = 9$. The cost of linking $k = 95$. We restrict effort x as any positive integer value not exceeding 20, i.e., $\overline{x} = 20$. Finally, we set $a_1 = 1$, $a_2 = 0.5$, and $a_l = 0$, for all $l \geq 3$.

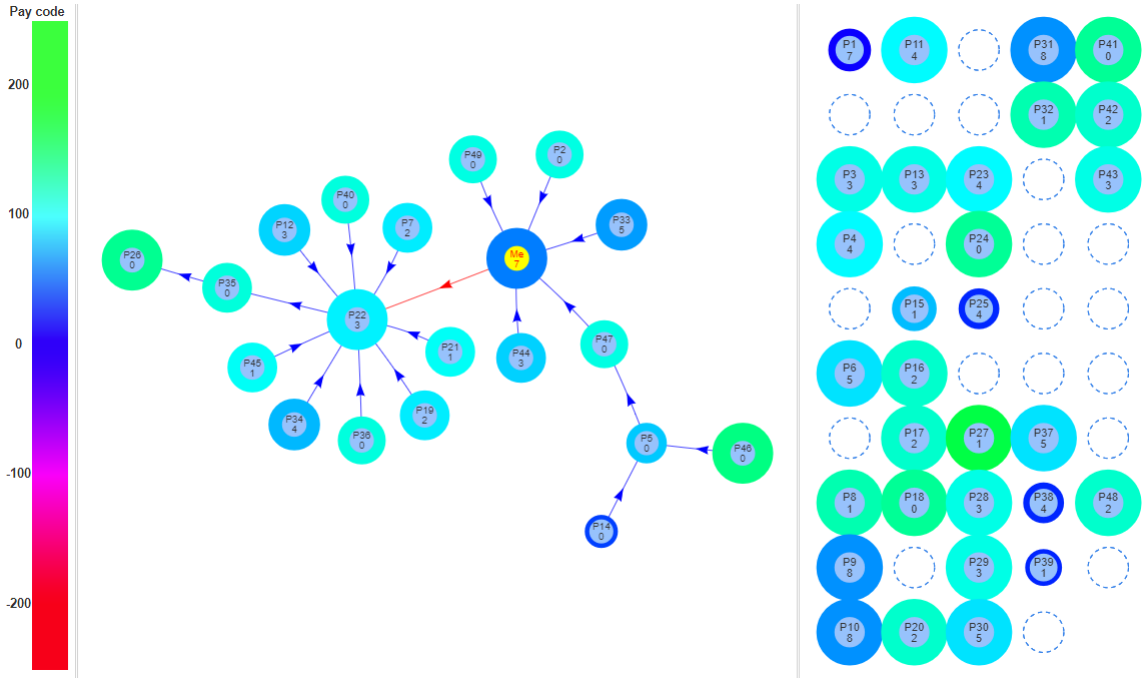


Figure 8: Screen shot of the Payoff Information Treatment

		Group size			
		$N = 4$	$N = 8$	$N = 50$	$N = 100$
Others' payoff information	**NO**	**Baseline4**	**Baseline8**	**Baseline50**	**Baseline100**
	YES	**PayInfo4**	**PayInfo8**	**PayInfo50**	**PayInfo100**

Table 1: Experimental Treatments

At any instant in the 6 minutes game, a subject can form or delete a link with any other subject by simply double-clicking on the corresponding node in the computer screen. If the subject forms a link with another subject on the right side of the screen (i.e., someone who is in more than 3 geodesic distance away), that subject along with his neighbors and neighbors' neighbors would be transferred to the left side of the computer screen. In a case where the subject removes a link with another subject on the left side of the screen, that subject would be transferred to the right side of the computer screen if they go more than 3 links apart and would remain in the left side of the screen otherwise.

During the experiment, each subject can also choose any level of effort by moving a slider varying from 0 to 20 by increments of 1. This slider is provided on top of the decision screen along with other payoff-relevant information including the subject's gross earnings (i.e., the benefit $f(x)$ where x is the total amount of information the subject has access to), cost of effort, cost of linking, and resulting earnings (i.e., payoff $\Pi_i(x_g)$). Further information on the screen is provided in Online Appendix D.

3.3 Experimental procedures

The experiment was conducted at the Laboratory for Research in Experimental and Behavioral Economics (LINEEX) located in University of Valencia and at the Laboratory for Experimental Economics (LEE) that is located at the University Jaume I of Castelln . All the treatments except for $N = 100$ treatments were conducted at the LINEEX. The experimental sessions with $N = 100$ subjects were conducted through an Internet connection between LINEEX and LEE (the number of subjects was then evenly distributed across the two locations). Subjects in the experiment were recruited from online recruitment systems of the two laboratories. 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. Subjects interacted through computer terminals and the experimental software was programmed using HTML, PHP, Javascript, and SQL. Sample instructions and interactive tutorials are available in Online Appendix C.

There were in total 18 sessions: 1 session of 16 subjects for each of the Baseline4 and PayInfo4 treatments, 1 session of 32 subjects for each of the Baseline8 and PayInfo8 treatments, 4 sessions of 50 subjects for each of the Baseline50 and PayInfo50 treatments, and 3 sessions of 100 subjects for each of the Baseline100 and PayInfo100 treatments. In each experimental session, subjects were (randomly) matched into a fixed group (if there were more than one group in a session) and interacted with the same subjects throughout the experiment. Therefore, there are 4 independent groups for each of the $N = 4$, $N = 8$, and $N = 50$ treatments and 3 independent groups for each of the $N = 100$ treatments. A total of 1096 subjects participated in the experiment.

The experiment consists of 6 rounds of the continuous-time game, each of which lasted for 6 minutes with the first minute as a trial period and the subsequent 5 minutes as the game with payment consequence. At the end of each round every subject was informed, using the same computer screen, of a time moment randomly chosen for payment, detailed information on subjects' behavior at the chosen moment including a network structure and all subjects' efforts, and the resulting earning of the subject. While the membership of a group was fixed within a session, subjects' identification numbers were randomly reassigned at the beginning of every round in order to reduce potential reputation effects. The first round was a trial round with no payoff relevance and the subsequent 5 rounds were effective for subjects' earnings. In analyzing the data, we will focus on subjects' behavior and group outcomes from the last 5 rounds. At the beginning of the experiment, each subject was endowed with an initial balance of 500 points and added positive earnings to or subtracted negative earnings from that initial balance. Subjects' total earnings in the experiment amounted to 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 100 points being equal to 1 euro. Each session lasted on average 90 minutes, and subjects earned on average about 18 euros (this included a 5 euros show-up fee).

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 these facts can be found in Online Appendix E.

3.4 Connecting theory and experiment

The static model focuses on the trade-off between personal efforts and linking with others. The analysis reveals that individual incentives and strategic interaction lead to a fairly clear cut resolution of these trade-offs: the network has a very specific structure and there are only two possible configurations of efforts possible. This sets a clear line for the experiment. Our interest is in understanding network formation in large groups. To facilitate individual experimentation and learning we consider a design in continuous time with asynchronous choice: this offers ample scope for experimentation and learning. However, this dynamic game opens the possibility of signalling, cheap talk, and reputation building, forces that go far beyond the original static game. The mapping from the static theory to the experiment is not straightforward.

Our philosophy is that if the arguments in the static model are robust, then subjects should abide by the predictions of the theory in an experimental setting that incorporates realistic elements – such as dynamic linking and effort choice – more accurately. Keeping this in mind, for the purposes of the experiment, we take the following 'high level' view of the predictions of the theory:

1. *Law of the Few*: a small fraction of individuals receive most of the links and carry out most of the efforts. An increase in group size leads to greater specialization in linking and efforts.

2. *Strategic uncertainty*: there exist multiple equilibria; these equilibria differ in actions, linking and payoffs across individuals.

4 Results: Baseline Treatments

We highlighted three key points from Figures 1 and 2: (i) extreme specialization in linking and efforts; (ii) very large efforts and intense competition among a few subjects to become

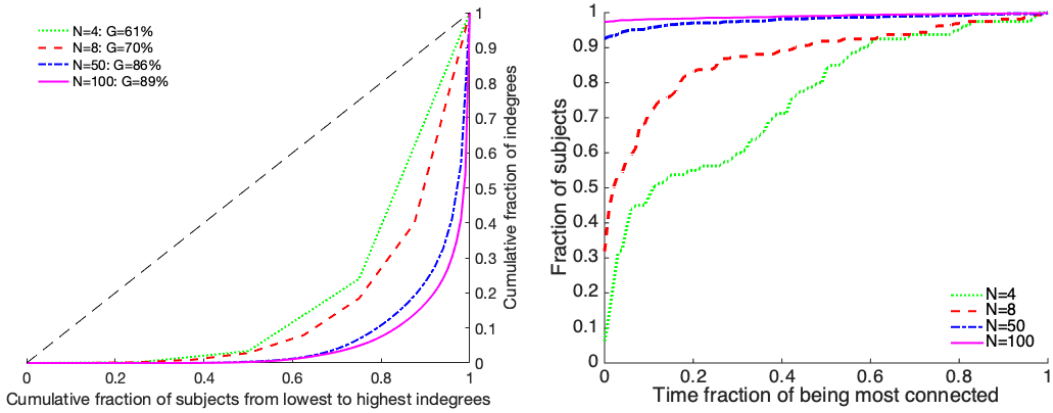
the hub; and (*iii*) the emergence of the pure influencer outcome. In this section we examine the experimental data more systematically.

For simplicity in all the data analyses that follow, the data used from every round of the game consists of 360 observations (snapshots of every subject's choices in the group) selected at regular time intervals of one second. Although some information about choice dynamics between two time intervals may be lost, we consider the possible impact of such a simplification as negligible to our analyses. Moreover, unless stated otherwise, all analyses are focused on data from the last 5 minutes of each round of the game.

4.1 Macroscopic Patterns

The first statistic we consider is the number of connections. For any individual, their *indegree* is the number of incoming links from other individuals. The interest is in the specialization/inequality in the indegree. We present two different ways of looking at this issue. The Lorenz curve plots the cumulative fraction of subjects, ranked from least connected to most connected, against the cumulative fraction of total indegrees. The (instantaneous) Lorenz curves are then averaged across seconds of the last five minutes, across rounds, and across groups in each treatment. Figure 9a presents these (averaged) Lorenz curves and the corresponding Gini coefficients of indegree across different group sizes. They reveal that specialization is present in every group size, but that it becomes especially striking as the group size increases. This is well reflected in the Gini coefficient measure: it is 0.61 for Baseline4, 0.70 for Baseline8, 0.86 for Baseline50, and 0.89 for Baseline100. By organizing the group-level average data, we conduct t-test for the null hypothesis on the equality of Gini coefficients between a small group ($N = 4$ or $N = 8$) and a large group ($N = 50$ and $N = 100$). We reject it with 5% significance level.

Consider next the number of individuals who become hubs. Specifically, we consider the time fraction (number of seconds out of 5 minutes) for which the individual is most connected. Figure 9b shows the cumulative distributions of time fraction of being most connected and mean indegree ratio. The fraction of subjects who *never* become the most connected player are very high for the large group treatments—0.97 for Baseline100 and 0.93 for Baseline50; this fraction is significantly lower for the smaller groups—0.31 for Baseline8 and 0.06 for Baseline4. It suggests that only a few subjects had any chance of being most connected in the large group treatments, whereas most of the subjects in the small group treatments experienced moments when they were most connected. In this



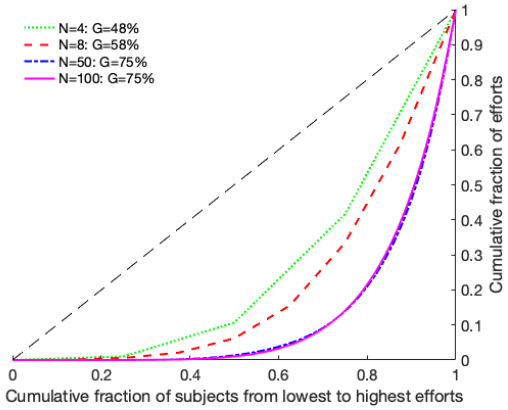
(a) Lorenz Curves and Gini Coefficient of Linking (b) Time fraction

Figure 9: Linking in the baseline treatments

sense, specialization grows with scale.

We turn to efforts. Again, the main indicator is the Lorenz curve and Gini-coefficients. Figure 10a presents (averaged) Lorenz curves and Gini coefficients, across different group sizes. Specialization in efforts is present in every group size and it is especially striking in larger groups. This is well reflected in the Gini coefficient of efforts: 0.48 for Baseline4, 0.58 for Baseline8, 0.75 for Baseline50, and 0.75 for Baseline100. The difference between Gini coefficient in the small group treatment ($N = 4$ and $N = 8$) and that in the large group treatment is statistically significant (p-values < 0.01 from t-test with the group-level data).

In order to look into the details of specialization next consider a variable of mean effort ratio. An individual's effort ratio at every second is defined as her effort divided by the sum of efforts across individuals at that second. We compute the mean of effort ratios across the five minutes for each individual. With this variable, we compute its cumulative distribution in any round and consider the average across rounds and groups. Specialization in efforts becomes substantially more pronounced in large groups. The fraction of subjects whose mean effort ratio is low increases significantly in group size. For instance, relative frequencies of subjects with mean effort ratio being less than or equal to 0.05 are 0.19 for Baseline4, 0.42 for Baseline8, 0.91 for Baseline50, and 0.99 for Baseline100. The distribution of mean effort ratio for a small group treatment first order dominates that for a large group treatment at the usual significance level (p-value < 0.01



(a) Lorenz curves and Gini coefficients of Efforts

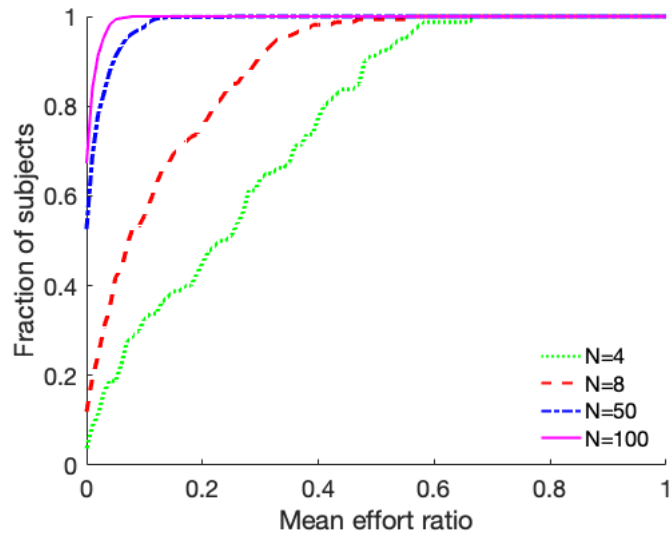


Figure 10: Distribution of Efforts

from the Kolmogorov-Smirnov test).

Result 1 *Specialization in linking and efforts is present in all group sizes and becomes significantly higher as group size increases.*

We consider the relation between indegrees, efforts and payoffs. Recall from Proposition 1, that there are two equilibria, corresponding to the pure influencer and the pure connector outcomes. In the former there is a positive correlation between efforts and indegrees and a (weak) negative relation between indegrees and payoffs. By contrast, in the latter equilibrium, there is a negative correlation between indegrees and efforts, and a positive correlation between indegrees and payoffs. We run linear regression analysis of mean indegree ratio on efforts interacted with the dummies for small group ($N = 4$ and $N = 8$) and large group ($N = 50$ and $N = 100$) and a median regression of (median) payoffs on mean indegree ratio interacted with the dummies for small group and large group. Table 2 reports the regression results with and without controlling demographic information, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality. All regressions include a constant, a dummy for large group, and dummies for rounds. Robust standard errors, clustered by individual subject, are reported in parenthesis. We use the median regression analysis to minimize the impact of outliers

Table 2: Regression analysis in the baseline treatments

	Indegree ratio (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	4.47***	4.48***		
	(0.43)	(0.43)		
Effort × Large group	0.62***	0.62***		
	(0.07)	(0.07)		
Indegree ratio (%) × Small group			-0.05	-0.03
			(0.14)	(0.16)
Indegree ratio (%) × Large group			-2.79***	-2.75***
			(0.61)	(0.51)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.580	0.581	0.109	0.157

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

in payoffs.[10]

This regression brings out the positive correlation between efforts and indegree: this relation is statistically significant and positive in the baseline treatments of small groups and large groups. The regression coefficients for efforts is smaller in the large group treatments than in the small group treatments. This is partly because the range of effort is wider in the large group treatments while the range of the indegree ratio is similar across the treatments. Next, we note that the association between indegree and payoffs is weak and insignificant in the small group baseline treatments. There is, however, a strong negative and significant correlation between linking and payoff in the large group baseline treatments. A one percent increase in mean indegree ratio is associated with 2.75 decrease in median payoff for the large group baseline treatments. These associations are robust to the inclusion of additional controls. We summarize the discussion on the relation among effort, linking and payoff as follows.

[10]In Online Appendix F.1 we report the same regression analysis by replacing mean indegree ratio with time fraction of being most connected. The regression results with both variables are quite similar.

Result 2 *There is a positive correlation between effort and indegrees in all group sizes. The correlation between indegrees and payoffs is insignificant in the small groups and significantly negative in the large groups.*

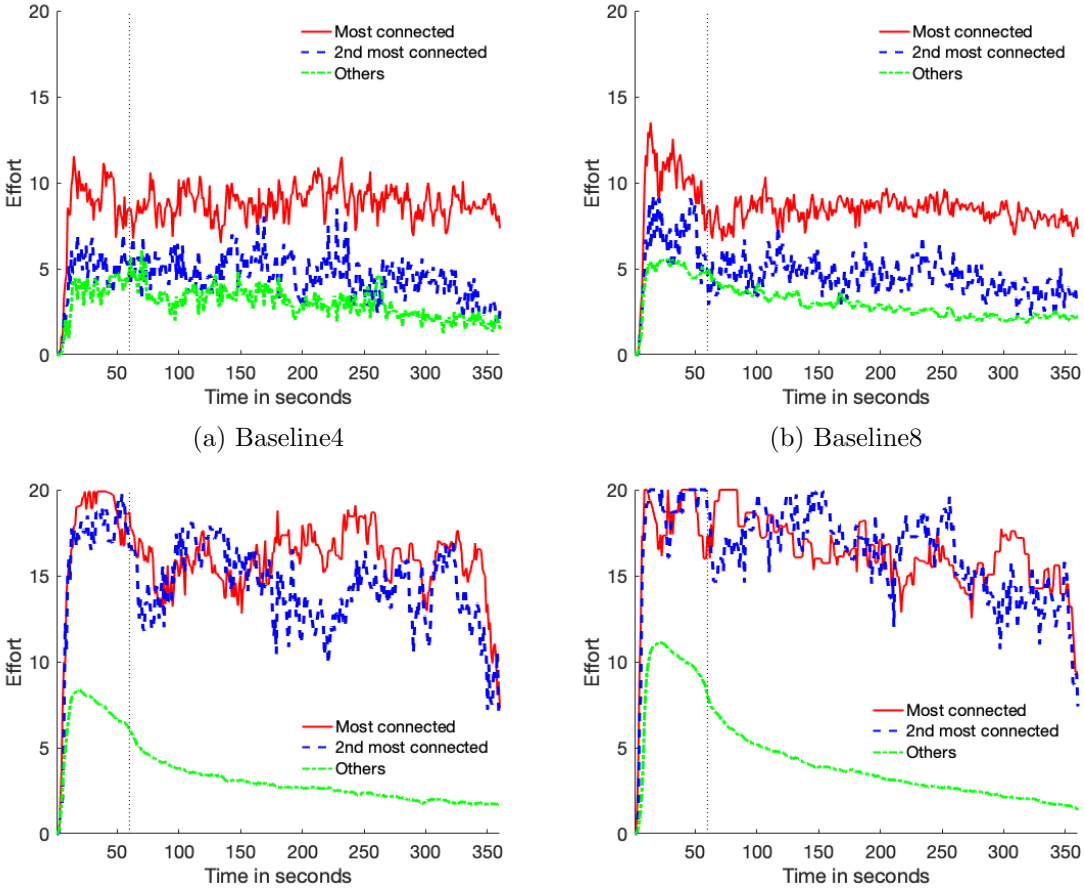
4.2 Individual Behavior and Competition Dynamics

This section examines the effects of group size more closely through a study of individual level behavior. The snap shots in Figures 1 and 2 suggest that there are different types of subjects with distinct dynamics of efforts during the game—the two most connected subjects who are competing with each other and the rest of the subjects.

We start with an examination of the dynamics of efforts made by the three different types of subjects identified at every second of the experiment—most connected, 2nd most connected, and the others. Figure 11 presents the average time series of effort for each of the group sizes. The end of the trial minute is represented by the vertical dotted line. We observe very sharp increase in effort by the most connected individual as we move from group size 8 to 50. The other interesting feature of the data is the relative levels of effort between the top two connected individuals: in the small groups there is a persistent gap between their efforts; in the large groups there is a very small gap in effort levels between the top two connected individuals. On the other hand, the average level of effort made by the others is low in all group sizes and steadily decreases over time. These time series patterns suggest that an increase in group size leads to significantly greater competition to become a hub.

We next look at the dynamics of median payoffs obtained by the three different types of subjects in Figure 12. The two most connected subjects do not perform better than the other subjects in the large groups. In particular, the 2nd most connected subjects in both Baseline50 and Baseline100 obtain persistently lower payoffs over time than those of the other subjects. This is a consequence of the high efforts. The most connected subjects in the Baseline50 also get persistently lower payoffs than the other subjects except for the last 10 seconds. In the Baseline100, they earn as much as the others for brief periods but the average payoffs are lower than others' payoffs. By contrast, in small groups, the payoffs earned are stable and very similar among the three different types of subjects.

In order to make a statistical assessment on the treatment effects on subjects' behavior, we conduct linear regression analyses of mean efforts and oudegree (the number of links) made by each type of subjects—most connected, 2nd most connected, and the others—on



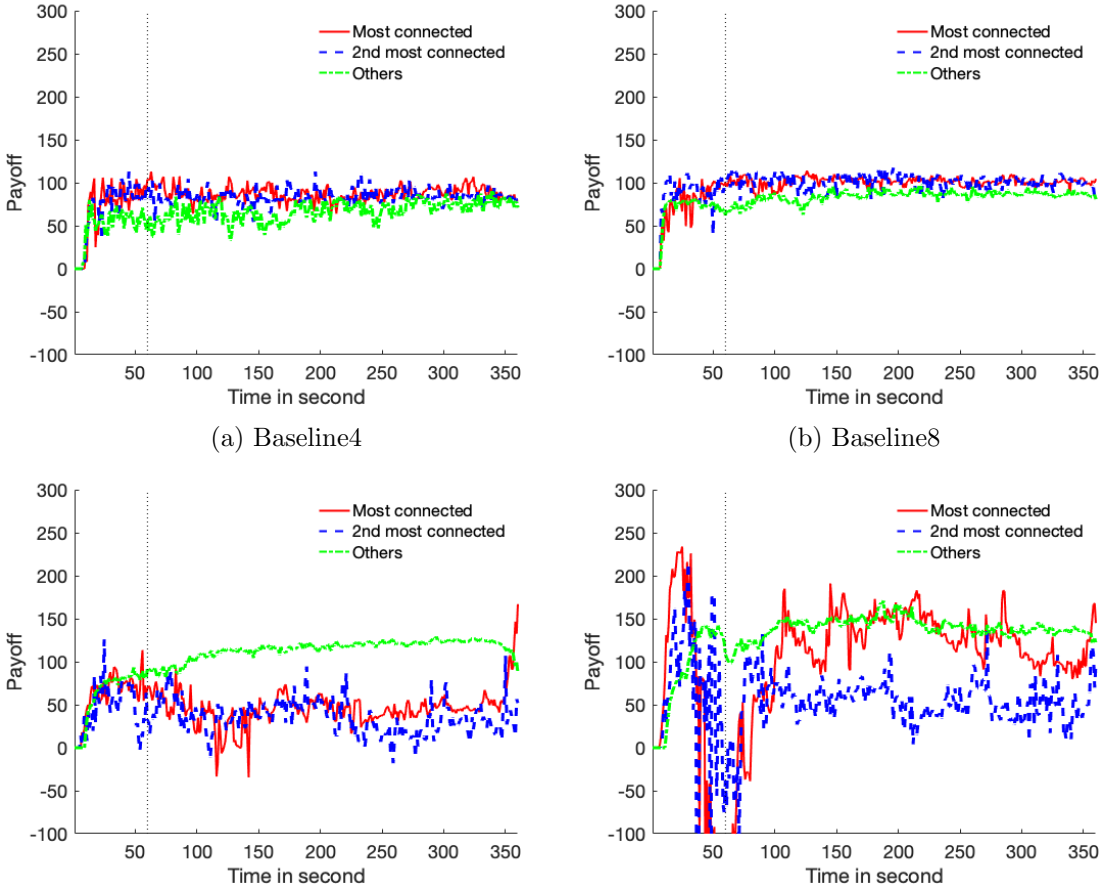
(a) Baseline4

(b) Baseline8

(c) Baseline50

(d) Baseline100

Figure 11: Time series of efforts for the three different types of subjects



(a) Baseline4

(b) Baseline8

(c) Baseline50

(d) Baseline100

Figure 12: Time series of median payoffs for the three different types of subjects

the dummy of large groups ($N = 50$ or 100). In this analysis, we define the types of subjects with the ranking of the fraction of time (across the five minutes) in which a subject is most connected in a round.[11] The most connected individual is the subject who receives the most links for the largest fraction of time. The 2nd most connected individual is similarly defined. We refer to the rest of subjects as the 'others'.

Table 3 reports the regression results after controlling for round dummies, demographic information, comprehension test score, experimental measures of risk aversion and altruism, and personality. Robust standard errors clustered by individual subject are reported. Average efforts and outdegrees for each type of subjects in the small groups ($N = 4$ and 8) are also reported for comparison.

We observe statistically significant and large treatment effect on efforts and outdegree. The most connected subject chose 68% more effort and about one link more in the large groups as compared to the small groups. The 2nd most connected subject made 173% more effort in the large groups than in the small groups. These patterns confirm that competition is more intense in the large groups.[12]

These effects of group size on subjects' behavior in turn have large effects on their payoffs. Table 4 reports median regression results on the effects of scale on individual median payoffs.[13] As expected, the two most connected subjects earned substantially less in the large groups than in the small groups: 27% less for the most connected subject, albeit less strongly significant, and 55% less for the 2nd most connected subject. [14] And thanks to the intense competition of the two most connected subjects, the other subjects earned 44% more in the large groups than in the small groups.

The discussion on individual behavior, the competition dynamics, and payoffs is summarized as follows.

Result 3 *An increase in group size intensifies the competition between the two most*

[11] Figure 21 in Online Appendix F.2 presents histograms showing the time fraction of different efforts over 5 minutes for the three different types of subjects across group sizes in the baseline treatment. The two most connected subjects in the large groups chose the maximum effort level, 20, for the majority of time, whereas they in the small groups chose significantly less with the mode of the most connected subject's effort being around the equilibrium effort level, 9

[12] Tables 9 and 10 in Online Appendix F.1 report the replications of Table 3 by splitting the two large groups. The results remain similar with each of the large groups.

[13] Due to outliers of payoffs, we conduct median regression analysis with median payoffs.

[14] Tables 11 and 12 in Online Appendix F.1 report the replications of Table 4 by splitting the two large groups. The negative effects of large group on median payoffs for the two most connected subjects are stronger in $N = 50$ than in $N = 100$.

Table 3: Scale effects on effort and outdegree in the baseline treatments

	Mean effort			Mean outdegree		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Large group	6.00***	9.04***	0.62*	1.03***	0.75*	0.24***
	(1.05)	(1.10)	(0.32)	(0.35)	(0.40)	(0.05)
Average in small group	8.77	5.24	2.65	0.20	0.62	0.90
Number of observations	75	75	2590	75	75	2590
R-squared	0.59	0.64	0.04	0.38	0.41	0.03

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 4: Scale effects on payoffs in the baseline treatments

	Median payoff		
	most connected	2nd most connected	others
Large group	-23.75*	-44.94**	37.12***
	(13.25)	(18.13)	(2.90)
Median in small group	86.50	81.00	85.00
Number of observations	75	75	2590
R-squared	0.19	0.23	0.08

Notes: Robust standard errors are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

connected subjects. It leads to a significant increase in their efforts and outdegree. A consequence of this is a decline in their payoffs, relative to the other subjects.

5 Payoff Information

We found that, as the group size grows, individuals compete fiercely to become hubs. This leads them to invest very large amounts and, as a result, their earnings suffer. Indeed, in some cases the hubs actually make negative earnings![15] This is a striking and somewhat unexpected outcome. An obvious concern is that the game becomes very complex with the increased number of players and individuals may not understand the payoff implications of different choices. To examine the scope of this idea we consider a design in which subjects are shown the payoffs of everyone. This section presents the findings for this treatment.

5.1 Macroscopic Patterns

Figure 13 begins by presenting the average of Lorenz curves and Gini coefficients of indegree across seconds of the last five minutes of the game, across rounds, and across groups in the payoff information treatment. We again observe the aggregate effect of scale on indegree distribution in the payoff information treatment, albeit to a lesser degree compared to the baseline treatments. The Gini coefficients are larger in the large group sizes than in the small group sizes: 61% for PayInfo4, 77% for PayInfo8, 84% for PayInfo50, and 81% for PayInfo100. Comparing these statistics from the baseline treatments, we observe an increase of Gini coefficient in PayInfo8 relative to Baseline8 and a decrease of this in PayInfo100 relative to Baseline100. In spite of this, we observe a statistical difference in this measure of specialization of linking between PayInfo4 and each of PayInfo50 and PayInfo100 (p-value < 0.01 from t-test with the group-level average data). These scale effects are also seen in the cumulative distributions of the time fraction of being most connected and mean indegree ratio (see Online Appendix F.2).

Turning to efforts, Figure 14 presents the (averaged) Lorenz curves and Gini coefficients across groups. The scale effects we see are similar to what we observed for the baseline treatment. The Gini coefficient increases in group size: 38% for PayInfo4, 57% for PayInfo8, 74% for PayInfo50, and 74% for PayInfo100. We observe a statistical difference in this

[15] We observe that 25% (13%) of the 1st most connected subject's sample in the Baseline50 (Baseline100) earned negative earnings. There is no incidence of negative earnings for the most connected subject in the small group treatments. Negative earnings are often made by excessive efforts and multiple links.

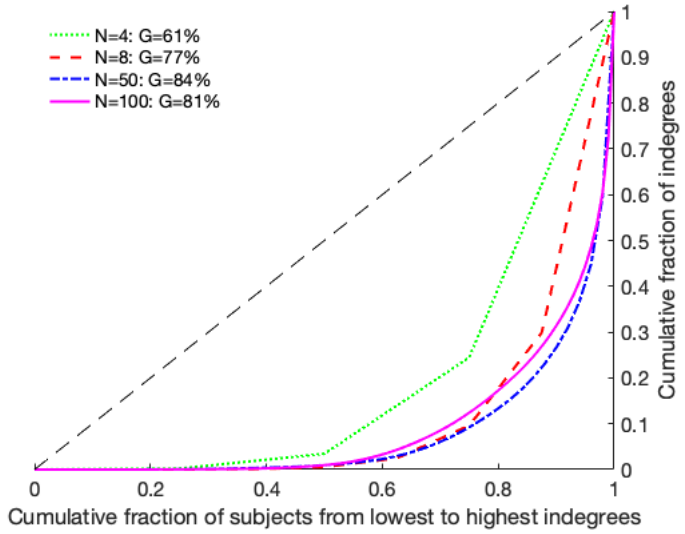


Figure 13: Lorenz curves and Gini coefficients of indegrees: Payoff Information treatments

measure of specialization of efforts between PayInfo4 and each of PayInfo50 and PayInfo100 (p-value < 0.01 from t-test with the group-level average data) and between PayInfo8 and PayInfo50 (p-value < 0.04). We also observe similar scale effects with the cumulative distributions of mean effort ratio (see Online Appendix F.2).

We next consider the relation between efforts and linking and the relation between linking and payoff in the payoff information treatments. As in the baseline treatments in Section 4.1, we run linear regression analysis of mean indegree ratio on efforts interacted with the dummies for small groups and large groups and a median regression of (median) payoffs on mean indegree ratio interacted with the same dummies for small groups and large groups. Table 5 reports the regression results with and without controlling the set of additional variables. Robust standard errors, clustered by individual subject, are reported in parenthesis.

Firstly, starting with the regression results about the large group payoff information treatments, we find that the relation between efforts and linking is significantly positive but relatively weak. On the other hand, the relation between linking and payoff in the large group treatments is strongly positive. Overall we interpret that showing information on others' payoff makes subjects choose efforts cautiously and leads to the relation between

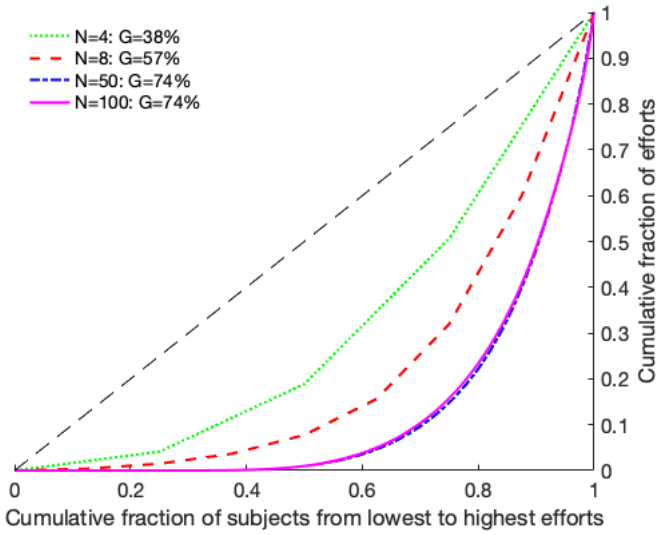


Figure 14: Lorenz curves and Gini coefficients of Efforts: Payoff Information treatment

Table 5: Regression analysis when information on others' payoff is observable

	Indegree ratio (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	5.42***	5.37***		
	(0.55)	(0.55)		
Effort × Large group	0.32***	0.32***		
	(0.06)	(0.06)		
Indegree ratio (%) × Small group			0.26**	0.28***
			(0.12)	(0.11)
Indegree ratio (%) × Large group			1.81***	1.75***
			(0.16)	(0.12)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.521	0.523	0.002	0.015

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

linking and payoff which is predicted by the pure-connector equilibrium.

We next look at the regression results about the small group payoff information treatments. As in the baseline treatments, we find a strong and positive association between efforts and linking in the small group payoff information treatment. This pattern is in line with the corresponding prediction of the pure-influencer equilibrium. When we analyze the relation between linking and payoff, we observe a relatively weak and positive relation between linking and payoff.

These observations are summarized as follows.

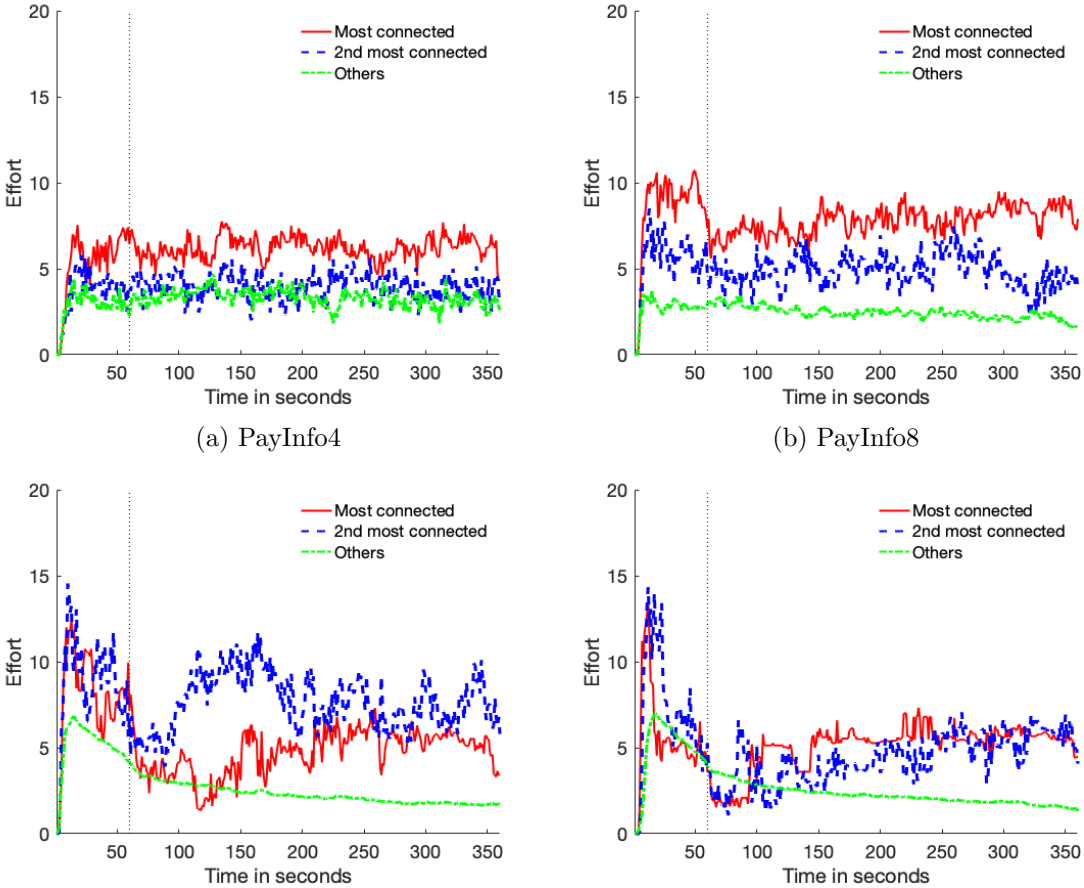
Result 4 *(i) Specialization in linking and efforts continues to hold in the payoff information treatments. (ii) In the small groups, the correlation between linking and effort is strongly positive, while the correlation between linking and payoff is weak. (iii) In the large groups there is a strongly positive correlation between linking and payoffs, while the correlation between efforts and linking is weak.*

A comparison of Result 4 with Result 2 reveals that growing group size increases specialization in linking and efforts under both the baseline and the payoff treatment. The second finding is that payoff information interacts powerfully with scale: the negative correlation between degrees and payoffs is reversed, subjects make more cautious effort choices and move away from a pure influencer outcome towards a pure connector outcome.

We now examine the decision rules that give rise to these scale and payoff information effects.

5.2 Individual Behavior and Competition Dynamics

We start with Figure 15 that presents the time series of average efforts (the end of the trial minute is represented by the vertical dotted line) for the three different types of subjects identified at every second in each of the group sizes. Compared to Figure 11 in the baseline treatments, we observe that competition dynamics in the large group payoff information treatments is quite different: the time series of efforts made by two most connected subjects are substantially lower when information on others' payoffs is visible. By contrast, in the small groups, the dynamics of efforts is similar across the payoff information treatment and the baseline. The behavior of 'other' subjects is similar across the two information treatment and across different group sizes.



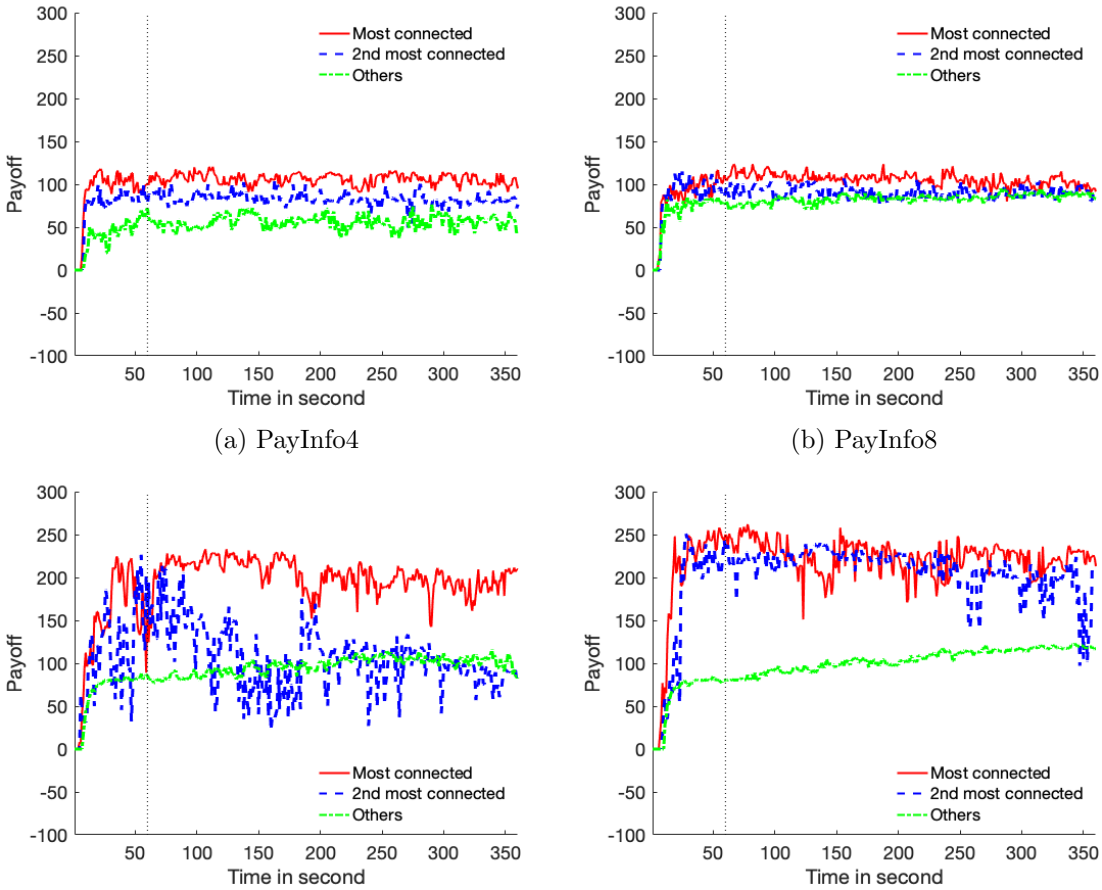
(a) PayInfo4

(b) PayInfo8

(c) PayInfo50

(d) PayInfo100

Figure 15: Time series of efforts in the payoff information treatment



(a) PayInfo4

(b) PayInfo8

(c) PayInfo50

(d) PayInfo100

Figure 16: Time series of median payoffs in the payoff information treatment

Figure 16 presents time series of median payoffs for the three types of subjects in the payoff information treatment. We would like to compare this figure with Figure 12 that summarizes the outcomes in the baseline treatments: this comparison reveals that there is a sharp increase in the payoffs of the most connected subjects in the large groups. In the small groups, the payoffs are similar across the two information treatments. Putting together the observations from Figure 15 and Figure 16, we are led to conclude that the availability of information on others' payoffs leads to lower efforts by the two most connected subjects and this in turn creates large payoff gains for them.

We next conduct linear regression analyses of mean efforts and outdegree made by each

type of subjects on the dummies of payoff information and large group ($N = 50$ or 100) and their interaction dummy. As was done in Table 3, we define each type using the time fraction of being most connected in the 5 minutes.[16]

Table 6 report the regression results with controlling for round dummies, demographic information, comprehension test score, experimental measures of risk aversion and altruism, and personality. Average efforts and outdegrees for each type of subjects in the large groups ($N = 50$ or 100) baseline treatment are also reported for comparison. The coefficient of the interaction dummy captures the difference-in-differences effect for the treatments of large group and payoff information. On the other hand, the coefficient of the payoff information describes the payoff information effect in the small groups.

In large groups, we observe a significantly negative effect of payoff information on efforts. Compared to the corresponding subject type in the large group baseline treatments, the most connected subject makes 61% less effort and the 2nd most connected subject makes 68% less effort. The rest of subjects also made 28% less effort in the large group payoff information treatments.[17] Hence, all the subjects in the large groups lowered their efforts when information on others' payoffs is available. In contrast, in the small groups, subjects' efforts did not respond to the availability of information on others' payoffs.

We then proceed to check the effect of payoff information on subjects' payoffs. Table 7 reports the median regression results of payoffs on the dummies of payoff information and large group as well as their interaction with the same set of control variables as in Table 6. We observe substantial impacts of payoff information on subjects' earnings in the large groups. The median payoffs increase by 202% and 260%, respectively, for the most connected subject and the 2nd most connected subject. In contrast, we observe a 11% drop of the payoffs for the other subjects.[18]

Result 5 *(i). In the small groups, subjects do not change efforts and linking behavior*

[16]Figure 22 in Online Appendix F presents histograms showing the time fraction of different efforts over 5 minutes for the three different types of subjects across group sizes in the payoff information treatment. We observe a drastic change of efforts in the large groups: in the payoff information treatments, the two most connected subjects substantially lower their level of efforts and the unique mode of the distribution is at zero effort. On the other hand, we observe little change of efforts in the small group sizes by showing information on others' payoffs.

[17]Tables 14 and 15 in Online Appendix F.1 replicate Table 6 by considering $N = 50$ and $N = 100$ separately. The negative effect of payoff information remain similar in each case.

[18]Tables 16 and 17 in Online Appendix F.1 report the replication of Table 7 by considering the cases of $N = 50$ and $N = 100$ separately. The results remain overall similar although the negative effect of payoff information for the most connected subject in $N = 100$ is only about 61% increase and imprecised estimated.

Table 6: Treatment effects on effort and outdegree

	Mean effort			Mean outdegree		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Payoff info	-0.75	0.52	0.00	-0.05	0.24	0.05
	(0.77)	(0.70)	(0.36)	(0.84)	(0.18)	(0.07)
Large group	6.30***	8.41***	0.62**	1.14	0.94***	0.25***
	(1.04)	(1.19)	(0.30)	(0.93)	(0.33)	(0.05)
Payoff info × Large group	-9.24***	-9.00***	-0.91**	2.34	-0.80**	-0.01
	(1.41)	(1.63)	(0.39)	(2.42)	(0.39)	(0.07)
Average in large group baseline	15.12	13.22	3.22	1.90	1.32	1.12
Number of observations	150	150	5180	150	150	5180
R-squared	0.53	0.51	0.05	0.43	0.34	0.04

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 7: Treatment effects on payoffs

	Median payoff		
	most connected	2nd most connected	others
Payoff info	6.71	-12.75***	-10.56***
	(11.54)	(4.53)	(1.97)
Large group	-30.33*	-42.76**	36.20***
	(17.08)	(17.34)	(1.90)
Payoff info × Large group	119.24***	120.76***	-14.07***
	(29.18)	(29.02)	(2.30)
Median in large group baseline	59.00	47.00	126.50
Number of observations	150	150	5180
R-squared	0.09	0.17	0.09

Notes: Robust standard errors are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

significantly in response to the availability of information on others' subjects. (ii). In the large groups, the two most connected individuals make substantially lower efforts in the payoff information treatment as compared to the baseline treatment. This results in large payoff gains for them.

We are led to the view that, when groups are large, it is difficult for a subject to keep track of the payoff implications of different choices. When information on payoffs of others is not directly shown, some subjects are willing to make large efforts in order to attract links from others. In doing so, these subjects appear not to understand that these large efforts lead to much lower payoffs. However, when information on others' payoffs is made directly observable, individuals become more cautious in their effort decisions. This is in contrast with what we observe in small groups: there subjects are able to keep track of the payoff implications of their effort more accurately and the impact of showing information on everyone's payoffs has relatively small impact on subjects' effort behavior.

6 Matching Effort Dynamics with Learning Rules

The effort dynamics presented in Figures 11 and 15 bring out two general points: one, they show large effects of group size and payoff information on the behavior of the two most connected individuals. Two, these Figures show that other – poorly connected – subjects behave similarly across the group sizes and information treatments: they make low effort that is declining over time. Figures 27 and 28 in the Online Appendix F.2 present the effort dynamics of less connected subjects and compare them to their myopic best response efforts. Myopic best response is a good match to the observed behavior. Consequently, in what follows, we focus on the behavior of the two most connected individuals. We will argue that their behavior fits well with a decision rule that combines myopic best response and high degree seeking.

Let $\{(x_{it}, D_{it})\}_{t=1}^{360}$ denote the sample of individual i's effort and indegree over periods, $t = 1, ..., 360$. We use the sample of two most connected individuals i and j who compete to be a hub. Because the computer screen was updated every 5 seconds or whenever the individual made a decision, we allow 3 seconds time lag in defining effort levels predicted by the learning rule.[19] The learning rule we consider has two parameters (θ and $\overline{x}$) to be estimated from the data, with the following features:

[19]Results presented in this section are robust to different values of time lag around 5 seconds.

$$x_{it}(\theta_i, \overline{x}_i) = \begin{cases} \overline{x}_i & \text{if} \quad |D_{i,t-3} - D_{j,t-3}| \leq \theta_i, \ x_{it-3} > \ x_{jt-3} \\ x_{jt-3} & \text{if} \quad |D_{i,t-3} - D_{j,t-3}| \leq \theta_i, \ x_{it-3} \leq \ x_{jt-3} \\ x_{mbr,t} & \text{if} \quad |D_{i,t-3} - D_{j,t-3}| > \theta_i \end{cases}$$

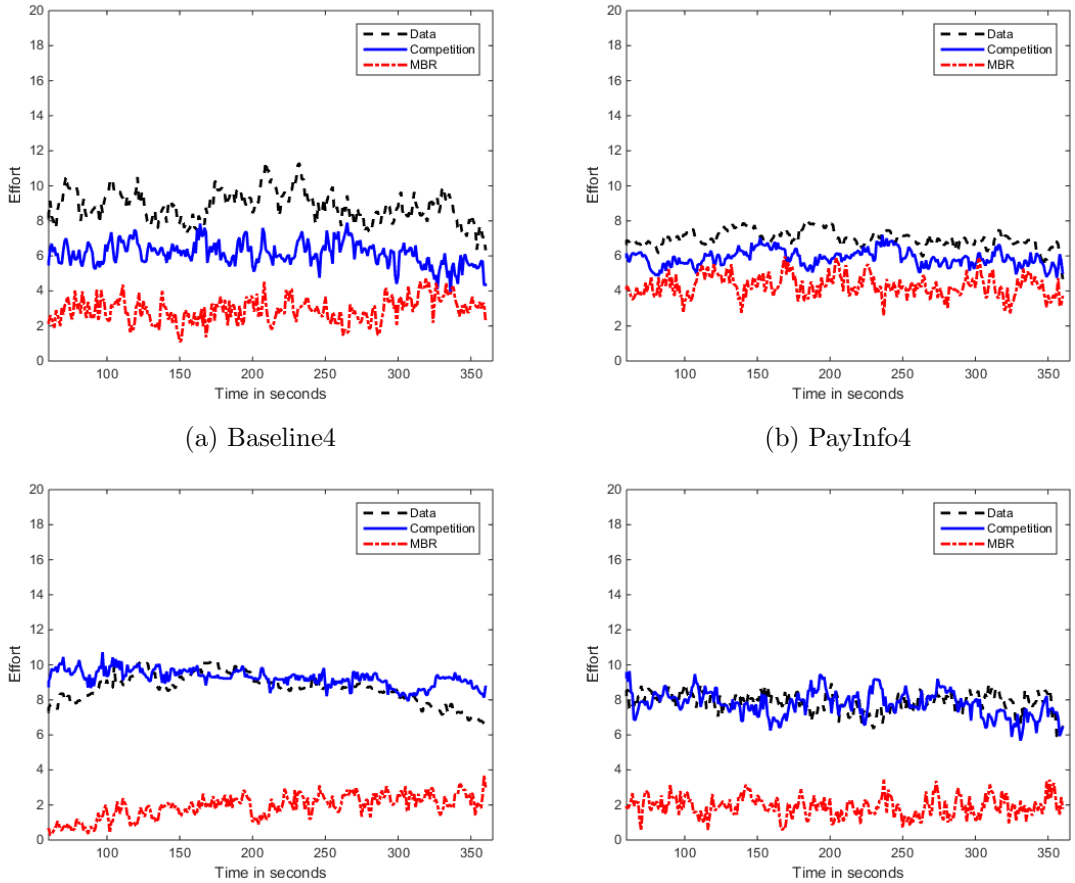
This decision rule has two features that are worth noting. First, competition for hub status is captured by the difference between indegrees obtained by the two individuals. Specifically, if the gap between their indegrees is larger than θ, the competition is not longer active, either because an individual is sufficiently ahead of the other or has fallen too far behind. In this situation, the individual chooses an effort predicted by myopic best response. The second part of the rule applies when the two players are in competition to become a hub: in this situation, the individual chooses either $\overline{x}$ if he chose a higher effort than the other or, otherwise, imitates the other's (higher) effort.

We estimate the two parameters, θ_i and $\overline{x}_i$, for each individual i by minimizing the sum of the distance between observed efforts and those predicted by the learning rule. In order to assess a goodness of fit of the learning rule, we compare the time series of observed efforts with those predicted by estimated learning rules.

Figure 17 and Figure 18 present the fitting of our learning rule to observed effort dynamics of the most connected individual during the last 5 minutes across treatments of group size and payoff information. The black-color dashed line represents effort dynamics observed in the experimental data. The blue-color solid line describes the dynamics of efforts predicted by the estimated learning rules. For the purpose of comparison, we also draw the time series of efforts predicted by myopic best response (the red-color dotted line). These figures show that our decision rule provides a good fit for the effort dynamics in the experiment; this is specially so if we compare it with the myopic best response rule.

In the large group baseline treatments, the two individuals compete strongly by choosing effort close to the maximum level of effort at the start (of the payoff relevant period) and decrease their efforts over time. The learning rule we propose captures such patterns of effort dynamics closely. On the other hand, in the large payoff information treatments, the two individuals tend to start with low efforts and lower their efforts slightly over time. The learning rule captures this pattern of efforts. The close fit is obtained through different estimates on the two parameters of the decision rule – θ and $\overline{x}_i$ – as we vary scale and payoff information.

Figure 25 and Figure 26 in Online Appendix F.2 report the same kind of figures for the second most connected individual. The figures show that the learning rule also performs



(a) Baseline4

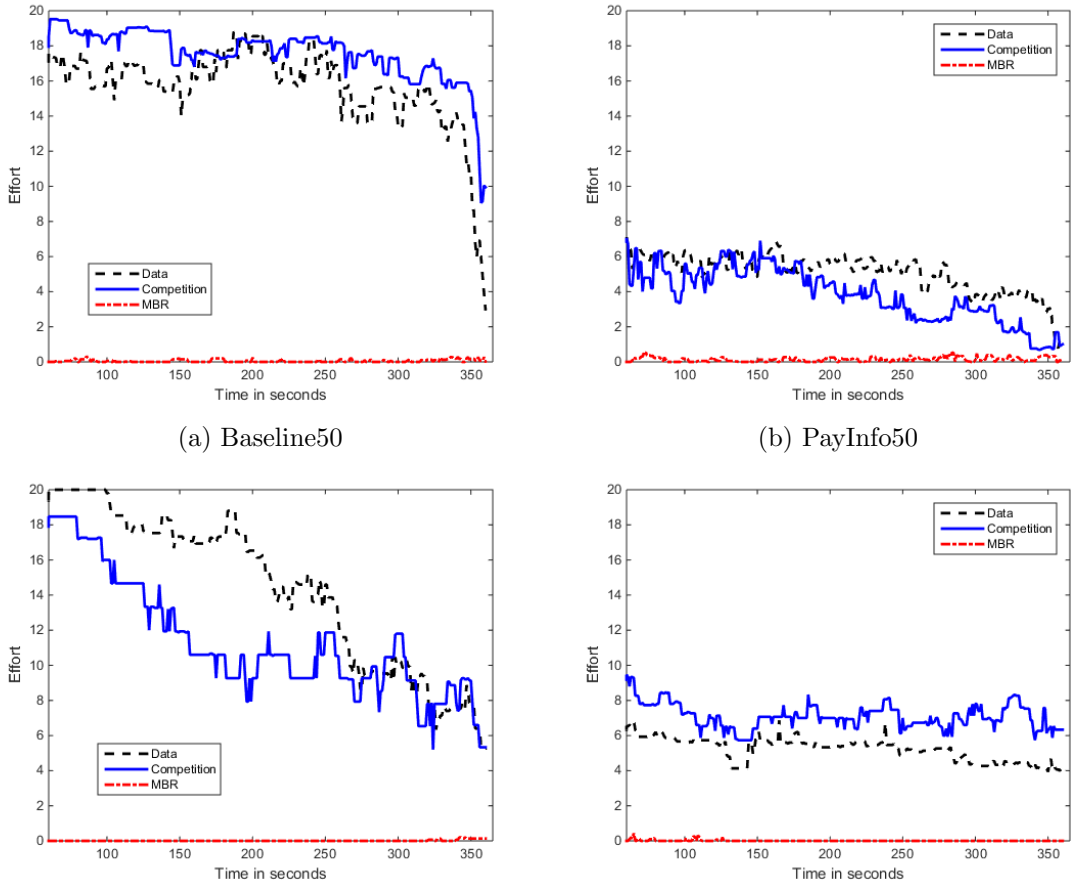
(b) PayInfo4

(c) Baseline8

(d) PayInfo8

Figure 17: Fitting effort dynamics with learning rules: most connected individual

well in fitting their effort dynamics.



(a) Baseline50

(b) PayInfo50

(c) Baseline100

(d) PayInfo100

Figure 18: Fitting effort dynamics with learning rules: most connected individual (cont.)

7 Conclusion

There is a large body of research that describes the structure of large empirical networks. A recurring theme in this work is that networks exhibit great inequality in connections. The economic theory of network formation shows that the trade-off between the costs of linking and the benefits of direct and indirect links is resolved in strategic models in favor of unequal networks. However, experiments on these models shows that subjects do not form such networks. This mismatch between the theory and the experimental evidence is the context for our paper.

We develop a new platform for the study of network formation. The platform allows for continuous time linking and effort choice and it allows for large scale experiments (up to 100 subjects). The paper presents an experiment on this platform; we test the predictions on specialization on linking and efforts in the model of Galeotti and Goyal [2010].

Our experiments provide strong support for the specialization prediction. Moreover, and in line with the theory, the specialization is more pronounced in larger groups. The second finding is that scale interacts powerfully with provision of information on others's payoffs. In the treatment where subjects see only their own payoffs as group size grows, the most connected individuals compete fiercely—they exert large efforts and have small earnings. By contrast, when a subject sees everyone's payoffs, as groups size grows, there is limited competition among highly connected subjects—they exert little effort and have large earnings. In the former setting, subjects always pick the pure influencer outcome, while in the latter case they often pick the pure connector outcome. We show that the treatment effects on effort dynamics can be reconciled with the individual learning rule that combines myopic best response and a desire to have many connections.

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.

V. Bala and S. Goyal. A non-cooperative model of network formation. *Econometrica*, 68 (5):1181–1231, 2000.

A.-L. Barabsi 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.

S. K. Berninghaus, K.-M. Ehrhart, and M. Ott. A network experiment in continuous time: The influence of link costs. *Experimental Economics*, 9(3):237251, 2006.

Y. Bramoulle, A. Galeotti, and B. Rogers. *Oxford Handbook on Economics of Networks*. Oxford University Press, 2016.

E. Calford and R. Oprea. Continuity, inertia, and strategic uncertainty: A test of the theory of continuous time games. *Econometrica*, 85:915–935, 2017.

C. Camerer. *Behavioral Game Theory: Experiments in Strategic Interaction*. Princeton University Press, 2003.

C. Camerer and T. H. Ho. Experience-weighted attraction learning in normal form games. *Econometrica*, 67:827–874, 1999.

C. Camerer, T. H. Ho, and J.-K. Chong. Sophisticated experience-weighted attraction learning and strategic teaching in repeated games. *Journal of Economic Theory*, 104: 137–188, 2002.

F. F. M. M.-J. Charness, G. and M. Sutter. Experimental games on networks: Underpinnings of behavior and equilibrium selection. *Econometrica*, 82:1615–1670, 2014.

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

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

A. Galeotti and S. Goyal. The law of the few. *The American Economic Review*, 100(4): 1468–1492, 2010.

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, J. Moraga, and M. van der Leij. Economics: an emerging small world? *Journal of Political Economy*, 114(2):403–412, 2006.

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

D. Hojman and A. Szeidl. Core and periphery in networks. *Journal of Economic Theory*, 139:295–309, 2008.

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

S. Huck, H.-T. Normann, and J. Oechssler. Learning in cournot oligopoly-an experiment. *Economic Journal*, 109:C80–C95, 1999.

M. Jackson and A. Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71:44–74, 1996.

M. O. Jackson and B. W. Rogers. Meeting strangers and friends of friends: How random are social networks? *American Economic Review*, 97(3):890–915, 2007.

M. Kearns, S. Judd, and Y. Vorobeychik. Behavioral experiments on a network formation game. *ACM EC 2012*, 2012.

J. Pettit, D. Friedman, C. Kephart, and R. Oprea. Software for continous game experiments. *Experimental Economics*, 17:631–648, 2014.

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. Riedl, I. M. Rohde, and M. Strobel. Efficient coordination in weakest-link games. *The Review of Economic Studies*, 83(2):737–767, 2016.

K. Schlag. Why imitate, nd if so, how? a bounded rational approach to multi-armed bandits. *Journal of Economic Theory*, 78:130–156, 1998.

B. van Leeuwen, T. Offerman, and A. Schram. Competition for Status Creates Superstars: an Experiment on Public Good Provision and Network Formation. *Journal of the European Economic Association*, 2019.

ONLINE APPENDICES

A Theory

The following proposition shows that, under discrete values of personal effort, a sufficiently high cost of linking implies a pure influencer equilibrium (for any group size n) and an approximate pure connector equilibrium (for a sufficiently large group size).

Proposition 2. *Suppose payoffs are given by (1), $a_1 = 1$, $a_2 \in (0,1)$, $\hat{y} \in X = \{0, 1, 2, \ldots, \overline{x}\}$, and $c(\hat{y} + a_2 - 1) < k < c\hat{y}$.*

(a) Every Nash equilibrium $s^* = (x^*, g^*)$ is such that g^* is a periphery sponsored star structure where the hub is a pure influencer investing $\hat{y}$ and every spoke invests 0.

(b) If $n \geq 2 + \frac{\hat{y}-1}{a_2}$, there exist pure connector ϵ-equilibria where the hub invests 0 while m spokes invest 1 and others invest 0 (with m s.t. $(m-1)a_2 < \hat{y} \leq 1 + (m-1)a_2$).

Proof. It follows from Proposition 1 that the pure influencer equilibrium always hold, regardless of n. Moreover, the pure connector equilibrium holds only if $n \geq 2 + \frac{k}{a_2(c\hat{y}-k)}$, in which case it requires every spoke to personally invest $\frac{\hat{y}}{1+(n-2)a_2}$. Since $c(\hat{y} + a_2 - 1) < k$ implies $\hat{y} < \frac{k}{c} + 1$, we have that $n \geq 2 + \frac{k}{a_2(c\hat{y}-k)} > 2 + \frac{\hat{y}-1}{a_2}$, and consequently $\frac{\hat{y}}{1+(n-2)a_2} < 1$ for any $n \geq 2 + \frac{\hat{y}-1}{a_2}$. However, since the lowest positive effort that can be made in the game is 1, no more than $m = \frac{\hat{y}-1}{a_2} + 2$ players (with $m < n$) can benefit from making such minimum positive effort. In this case, each of those positive investors accesses $(m-1)a_2$ from forming a link and therefore earns $f(1 + (m-1)a_2) - c - k$, which is strictly less than if they unilaterally deviate by forming no link and investing $\hat{y}$ since $c(\hat{y} + a_2 - 1) < k$ can be rewritten as $(m-1)a_2 < \frac{k}{c}$. As a result, the pure connector outcome is an ϵ-equilibrium whenever $\epsilon > f(\hat{y}) - f(1 + (m-1)a_2) - c(\hat{y} - 1)$ where m is the number of investing spokes such that $(m-1)a_2 < \hat{y} \leq 1 + (m-1)a_2$. □

B Network visualization

The experimental software uses the well known Barnes-Hut approximation algorithm as introduced by Barnes and Hut [1986], which provides a low complexity simulation technique to compute the forces applied to any node as influenced by every other node in a network. Such forces are computed through three distinct sources:

- All other nodes from the network: all nodes apply a repulsion force F_r to each other to avoid overlaps and allow a sparse visualization of the network. This force is the only used by Barnes and Hut [1986].
- Connected nodes (with direct links only) in the network: nodes that are linked with each other apply attractive forces F_s towards each other to allow for visual proximity of connected nodes.
- Point of origin O: nodes are applied a gravitational force F_{cg} to a center of origin 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 directly linked with.

The Barnes-Hut algorithm consists in first constructing a quad-tree by recursively dividing the visual space into same size groups such that every node can eventually be associated with exactly one group based on its visual position (leaf of the tree). Any such group may however associate several nodes such that the aggregated forces applied from those nodes can be approximated through a unique force (as if the group of nodes were a single node). More precisely, starting from the largest group of the Barnes-Hut quad-tree (the root), the algorithm assesses the distance between a given o and the center of mass of that group of nodes: if the distance is sufficiently large (according to a given exogenous threshold), then the group of nodes is considered as a single node, else the process is iterated by considering subgroups from the tree (nodes sufficiently close to o will therefore be considered independently). Such approximation is known to considerably reduce computational complexity for computing forces applied to every node. The root of the Barnes-Hut quad-tree represents the whole visual space.

More formally, we define the distance between a node o and a node m (or group of nodes represented as a single node m) as $dist(o, m)$. The repulsion force applied to node o by m is determined as

$$F_r(o, m) = \frac{K_g.M_m}{dist(o, m)^3} \tag{3}$$

Where K_g captures the gravitational constant such that $K_g < 0$ to obtain the repulsion effect. In Equations (3), it is assumed that the mass of every node is 1. However, the mass M_m may be larger when representing a group of nodes (M_m represents the number of nodes in that group, as described by the Barnes-Hut algorithm). Similarly, the attraction force applied to node o by m corresponds to

$$F_s(o,m) = \frac{K_s}{dist(o,m)} \cdot \begin{cases} 0 & \text{if } o \text{ and } m \text{ are not linked} \\ (L - dist(o,m)) & \text{if } o \text{ forms a link with } m \\ (dist(o,m) - L) & \text{if } m \text{ forms a link with } o \end{cases} \tag{4}$$

Where L defines the resting length of an edge, and K_s the spring gravity constant such that $K_s > 0$ to obtain the attraction effect. Note from Equation (4) that the force applied on two linked nodes is symmetric, i.e., both nodes are equally attracted by each other. Finally, we define the central gravity force applied to node o as

$$F_{cg}(o) = \frac{K_{cg}}{dist(o,O)} \tag{5}$$

Where O represents the position of the point of origin, and K_{cg} the central gravity constant such that $K_{cg} > 0$ to obtain the attraction effect.

The net force vector applied to any node o resulting from the above three forces is then:

$$F_x(o) = d_x(o,O).F_{cg}(o) + \sum_{m \in N \setminus \{o\}} d_x(o,m).(F_r(o,m) + F_s(o,m)) \tag{6}$$

$$F_y(o) = d_y(o,O).F_{cg}(o) + \sum_{m \in N \setminus \{o\}} d_y(o,m).(F_r(o,m) + F_s(o,m)) \tag{7}$$

The above static properties describe the net forces that are applied in the network, given the positions of all nodes and the links between nodes. The resulting dynamic update of the network is achieved by computing the corresponding velocity of nodes on both coordinate axes. More precisely, the velocity applied to a node o at a time t on both coordinate axes

(x and y) is determined as follows:

$$V_x(o,t) = max(V_{max}, (F_x(o) - D.V_x(o,t-1)).T + V_x(o,t-1)) \tag{8}$$

$$V_y(o,t) = max(V_{max}, (F_y(o) - D.V_y(o,t-1)).T + V_y(o,t-1)) \tag{9}$$

Where D represents the damping factor determining how much of the velocity from the previous physics simulation iteration carries over to the next iteration, T the time step for the discrete simulation, and V_{max} the maximum velocity of nodes (used to increase time to stabilization). We assume no initial velocity, i.e., $V_x(o,0) = V_y(o,0) = 0$. Given such velocity, the position update of a node o at any time t directly follows:

$$x(o,t) = x(o,t-1) + V_x(o,t).T \tag{10}$$

$$y(o,t) = y(o,t-1) + V_y(o,t).T \tag{11}$$

The discrete simulation terminates and node o stabilizes whenever the associated velocity (on both coordinate axes) becomes sufficiently low with respect to some given threshold (V_{min}). More precisely, the convergence rules are:

$$V_x(o,t) < V_{min} \tag{12}$$

$$V_y(o,t) < V_{min} \tag{13}$$

Model parameter setting used in the experiment:

- $K_g = -2000$
- $K_s = 0.04$
- $K_{cg} = 0.3$
- $L = 95$
- $D = 0.09$
- $T = 0.5$
- $V_{min} = 0.3$
- $V_{max} = 10$

C Experimental instructions

[In the following instructions, N is to be replaced with any value from $\{3, 7, 49, 99\}$ *depending on the treatment]*

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. One of the experimenters 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 exchange rate:

100 points = 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 other participants. This group will remain fixed throughout the 6 rounds. Each of the participants will be randomly assigned an identification number of the form "Px" where x is a number between 1 and N. 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 **6 (six) mins: the first minute will be a trial period, only the latter 5 minutes will be relevant for the earnings**. Your earnings in a given round will be based on everyones choice **at a randomly selected moment in the last 5 mins of the round**. In other words, any decision made before or after that randomly

chosen moment will not be used to determine your points. This precise moment will be announced to everyone only at the end of the round, along with the corresponding behavior and earnings.

At the beginning of the experiment, you are given an initial balance of 500 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 capital (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 endowment) go below 0, your final earnings will be simply treated as 0.

In each round, every participant will have choose two types of actions:

- **How many any units to buy/invest**: You may buy at most 20 units. Each unit costs you **11 points**.
- **Add/delete links with other participants**: You are linked with another person if you form a link with that person or that person forms a link with you (or both). You do not pay any fee for links formed by others. The people that you are linked with (regardless of whether you or they form the links) are called your neighbours. You automatically have access to **all units bought by your neighbours** as well as **half of the units bought by your neighbours' neighbours** (see below for an example). Each link you form costs you **95 points**.

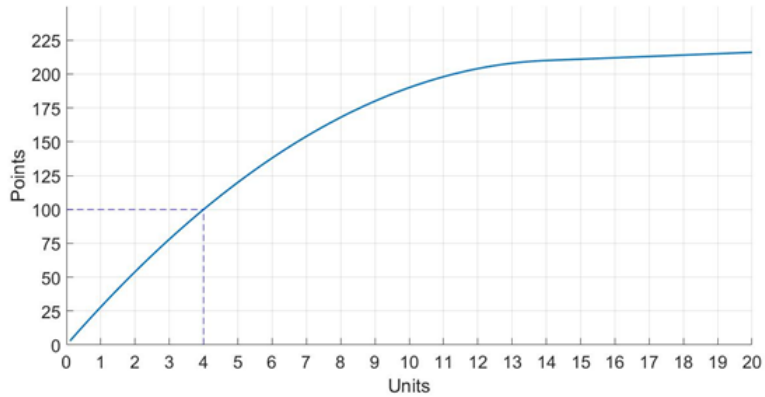
You may revise your choices at any moment before the round ends. During a round, you will also be informed about every other participants most recent decision (units bought and formed links), which will be updated every 5 seconds or whenever you change your own choice.

At any moment, the total number of units you have access to (i.e., units you bought + units bought by your neighbours + units bought by your neighbours neighbours) generates points for you according to the following figure (for example, accessing 4 units generates 100 points, as shown by the dotted lines):

Moreover, having access to 20+m units generates 216+m points.

The computer screen will be split into two parts:

- **The middle side of the screen presents you and your local neighbourhood**. More precisely, you will see your neighbours, the neighbours of your neighbours, and



the neighbours of neighbours neighbours. In other words, you will see the participants that are up to 3 links away from you.

- **The right side of the screen presents participants outside of your local neighbourhood**.
- **The left side of the screen presents the code for the players' net earnings in the network**. *[Payoff information treatment only]* The inner circle of each node from the middle or right part side of the screen is characterized by some color, which varies from **green** (high positive net payoff) to **red** (high negative net payoff) depending on the players corresponding net earnings.

Each node is described by their identification number "Px" and the number of units that they buy. Identification numbers "Px" are randomly assigned in every round. Therefore, every player is likely to have a different ID in different rounds. In the initial state of the network, nobody buys any unit and no link is formed.

Tutorial

Please follow this simple tutorial simulating a simple virtual scenario on the computer screen. In this tutorial you are interacting with 9 other players. In the initial state, you are not linked with anyone and you do not buy any units: you start at 0 points.

1. The slider allows you to choose how many units you wish to buy yourself. For example, buying 4 units costs you 44 points (= 4 units × 11 points, in red on the

screen) and generates 100 points (according to the figure from the previous page, in green on the screen).

2. Initially, the nodes on the right side of the screen represent all other players (in this simulation, those players are not real people). The size of node reflects the total number of units bought by that node and the units accessed via the network. For example, P1-P4 are the largest nodes because these players have access to the most units.

3. 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 P1, P2, and P3 each form a link with P4, and P9 forms a link with P1. Forming a link with P4 costs you 95 points (in red on the screen), but it also gives you access to 8.5 units (7 from P4 + 0.5 $\times$ 1 from P1 + 0.5 $\times$ 1 from P2 + 0.5 $\times$ 1 from P3), which generates 174 points (according to the above figure, describing the benefit function in green on the screen). If you do not buy any additional unit yourself, your resulting net payoff is **79 points (= 174 points − 1 link $\times$ 95 points)**.

4. After forming a link with P4, you observe that some nodes remain unobserved (P5, P6, P7, and P8 on the right side). However, forming an additional link with P9 (by double clicking on the corresponding node) reveals that those nodes all form a link with P9. You were not allowed to observe them before because they were 4 nodes away from you (for example, P5 were connected to you via P4, P1, and P9). You can now observe them because they are only 2 nodes away from you (for example, P5 is connected to you via P9 only). Remember that you can only see players that are at most 3 nodes away. Assuming you still do not buy any unit yourself, your resulting net payoff is **16 points (= 206 points from accessing 12.5 units − 2 links $\times$ 95 points)**.

5. 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 leads to players P2 and P3 becoming unobserved again, as they are now more than 3 nodes away from you.

6. Note that varying the amount of units you buy directly affects the sizes of the nodes you are linked with as well as their neighbours. Indeed, the amount of units they

each have access to includes the units you buy (the larger this amount, the larger the node).

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

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 mins**. A moment will be randomly selected **in the last 5 mins** to determine everyones 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 500 pts, any negative total of points at the end of the 5 rounds will be equivalent to 0 point.

3. The amount of units you have access is equal to the sum of **(1)** the units bought by you, **(2)** the units bought by your neighbours, and **(3)** half of the units bought by your neighbours neighbours.

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

5. Every other nodes 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.

6. The size of each node determines **how many units that node has access to** (units bought personally plus units accessed from others, directly and indirectly).

7. The amount of units **bought personally by** a player is mentioned inside the corresponding node.

8. *[Payoff information treatment only]* The color of each node determines **that nodes net earnings** according to the code depicted on the left side of the screen.

D Network game interface

The decision making interface used in the experiment is similar across all treatments. More specifically, Figure 19 illustrates a (fictitious) example of a subject's computer screen in Treatment **Baseline100**. 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), the subject's own effort, which can be modified via the slider, and a comprehensive description of the subject's own payoff. Information about payoffs include gross earnings (output of function $f(.)$), the cost of effort (own effort multiplied by c), the cost of linking (number of links multiplied by k), and the net earnings (costs substracted 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's local network is represented on the left, other players outside of the subject's local network are found on the right. Note that a scrolldown feature is available for the subject to explore every player outside of his/her local network. Baseline treatments with smaller group sizes use the very same interface (the scrolldown feature is not available then because all players are then directly visible on the screen).

Similarly, Figure 20 illustrates a (fictitious) example of a subject's computer screen in Treatment **PayInfo100**. The only difference with the decision screen from Figure 19 is about the wider range of colors used to represent the border of each node depicted in the network. Any given node's color is directly associated with that node's corresponding payoff, according to the scale presented on the left part of the screen. payoff-information treatments with smaller group sizes use the very same interface.

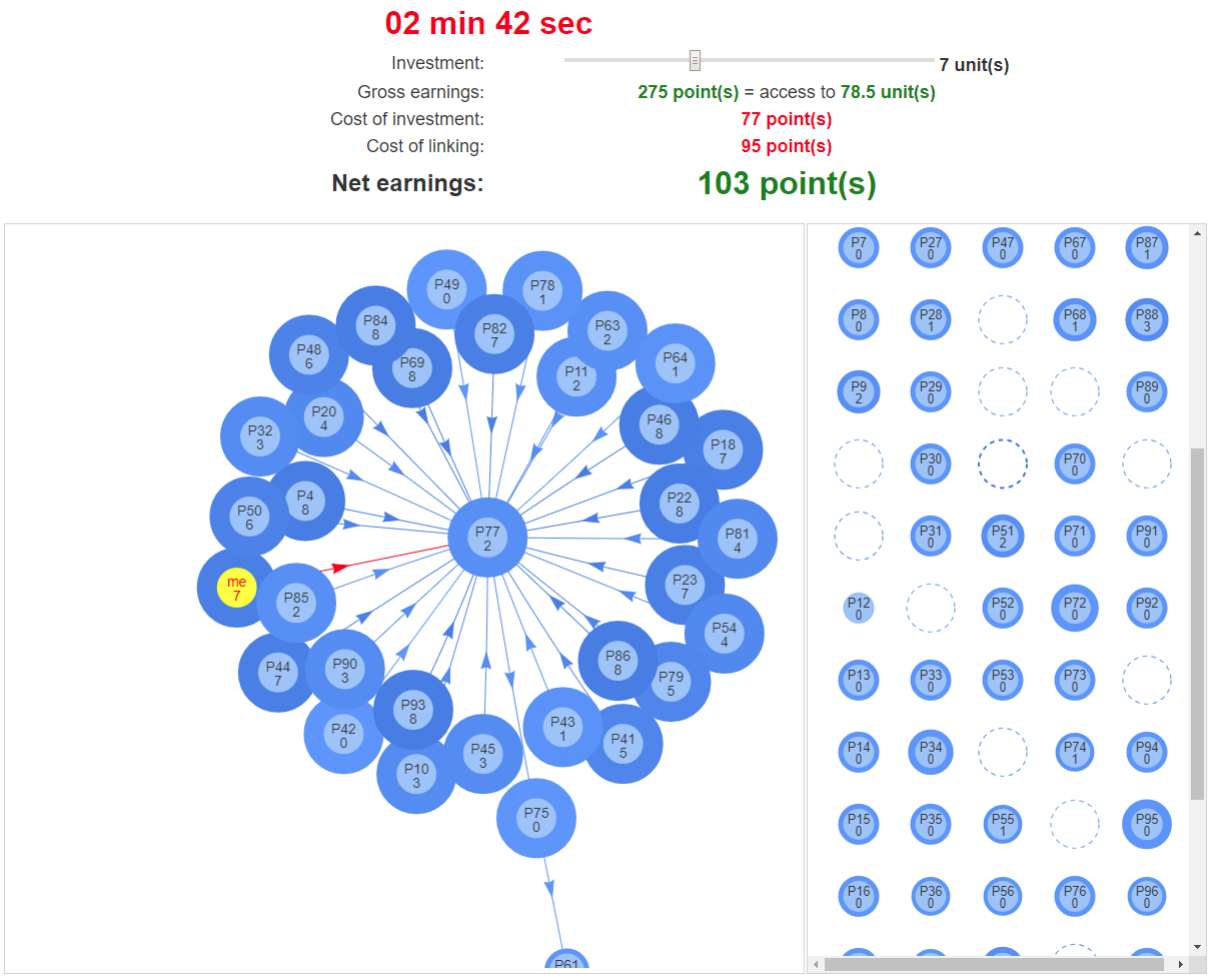


Figure 19: Example of decision screen for Treatment **Baseline100**

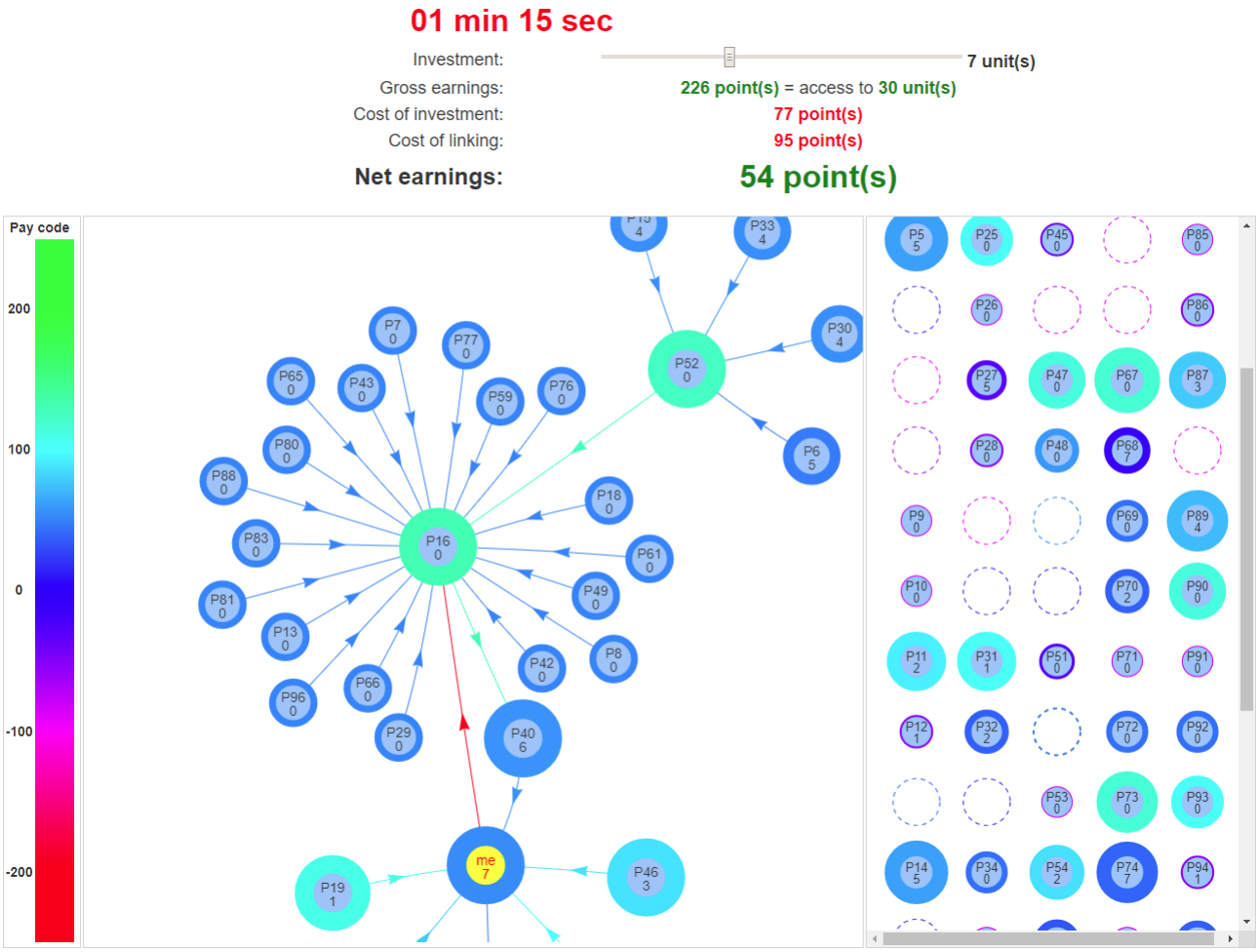


Figure 20: Example of decision screen for Treatment **PayInfo100**

E Questionnaires

At the end of the experiment, subjects answered a set of surveys aiming at measuring various types of individual differences. More precisely, incentivized measures of comprehension in network game, social preferences, and risk preferences were used. Finally non incentivized personality measures were used before which subjects filled up a debriefing questionnaire that includes demographics information.

E.1 Comprehension check

In order to assess the subjects' comprehension of the network game played during the experiment, we provided 5 questions, each of which with a unique correct answer. Each correct answer was rewarded with 0.1 euro for the subject.

The following first 2 questions were used across all treatments (correct answers are "11 pts" to question 1, and "95 pts" to question 2).

Question 1: In the previous game, how many points did investing one unit cost you?

- 1 pts
- 11 pts
- 21 pts
- 31 pts
- 41 pts
- 51 pts
- more than 51 pts

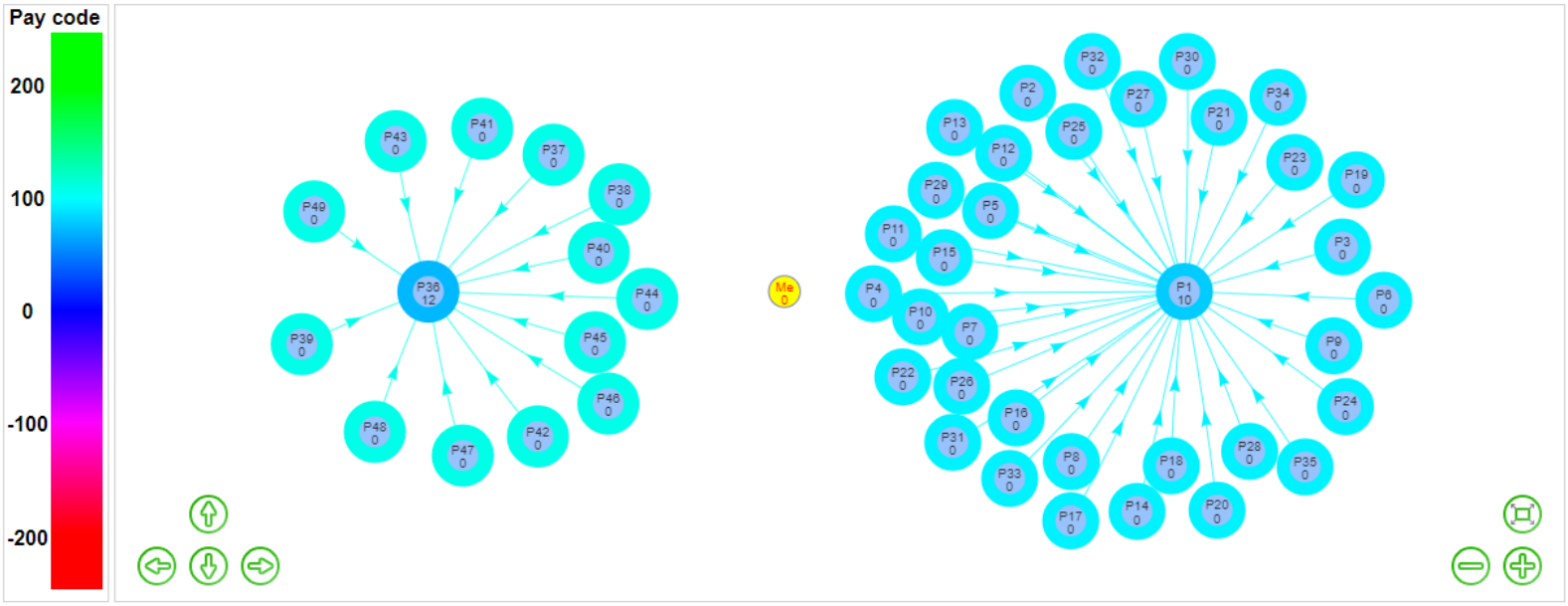
Question 2: In the previous game, how many points did forming a link cost you?

- 0 pt
- 25 pts
- 45 pts
- 65 pts
- 95 pts
- 115 pts
- more than 115 pts

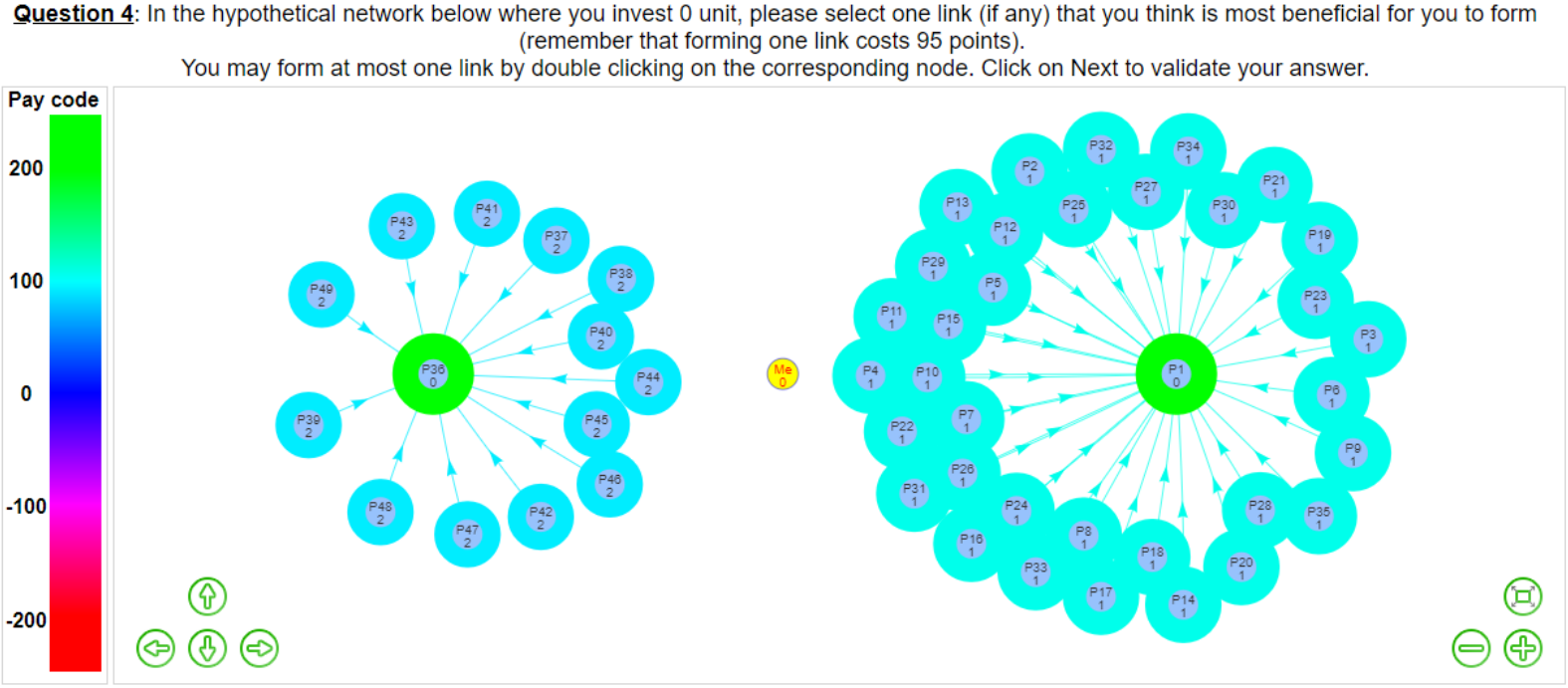
The third question depicted below was used in the payoff information treatment with $n = 50$ (the correct answer is "P36"). This question was adapted in all other treatments by matching the number of nodes to the group size in the experiment, and by removing the colors in the baseline treatments.

The following questions 4 and 5 below were also used in the payoff information treatment with $n = 50$ (correct answers are "P1" for both questions 4 and 5). These questions were

Question 3: In the hypothetical network below where you invest 0 unit, please select one link (if any) that you think is most beneficial for you to form (remember that forming one link costs 95 points).
You may form at most one link by double clicking on the corresponding node. Click on Next to validate your answer.

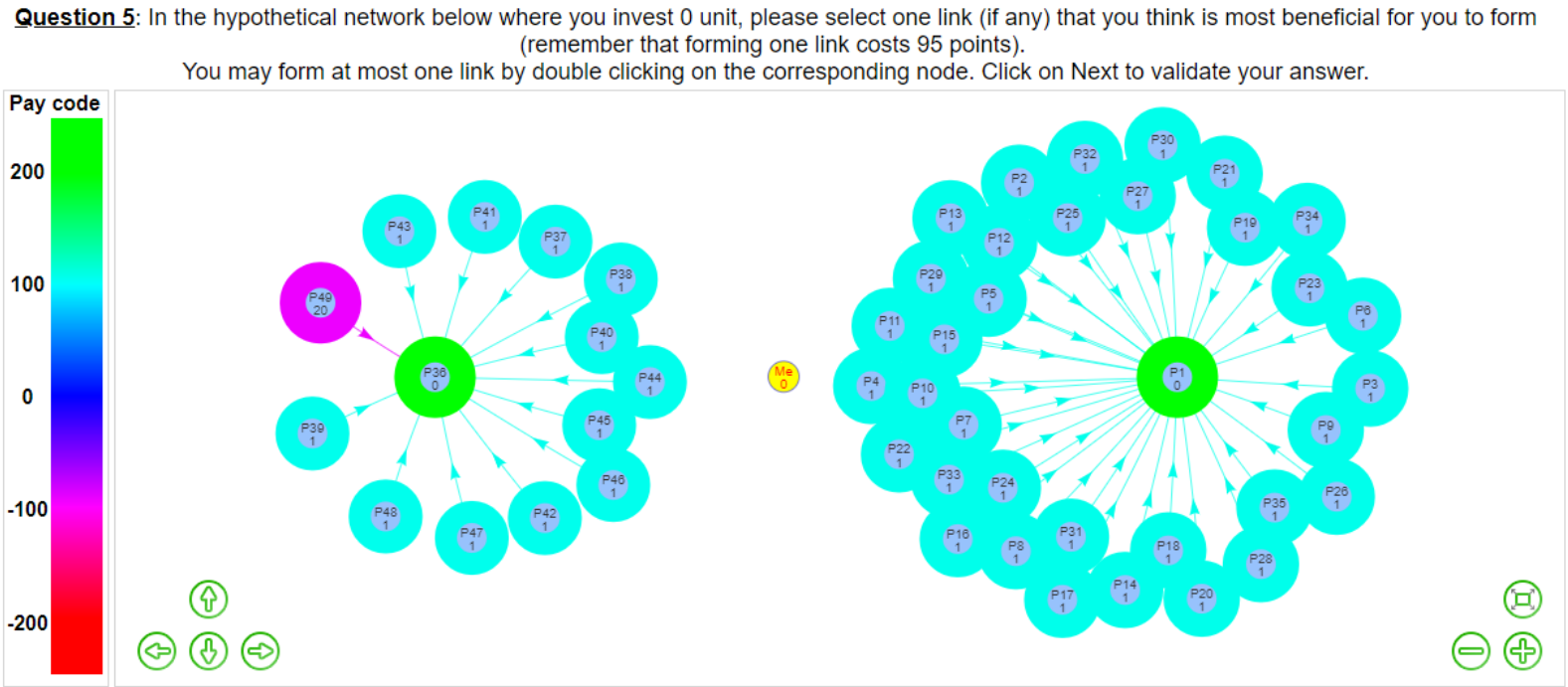


however adapted only in other treatments where $n > 4$ by again matching the number of nodes to the group size in the experiment. The reason for filtering the small group treatments (with $n = 4$) is that the limited number of nodes did not allow representing the corresponding scenarios. As before, these questions were also adapted to the baseline treatments by simply removing the colors to match the design of the actual game that subjects played.



E.2 Social preferences

The social preferences measure was adapted from Andreoni and Miller [2002] and involved a series of five money allocation tasks between the decision maker and some anonymous external participants of another experiment at the LINEEX lab (corresponding payments were therefore made to these external passive participants). The five tasks used in our experiment were represented through sliders as shown in the following figure:

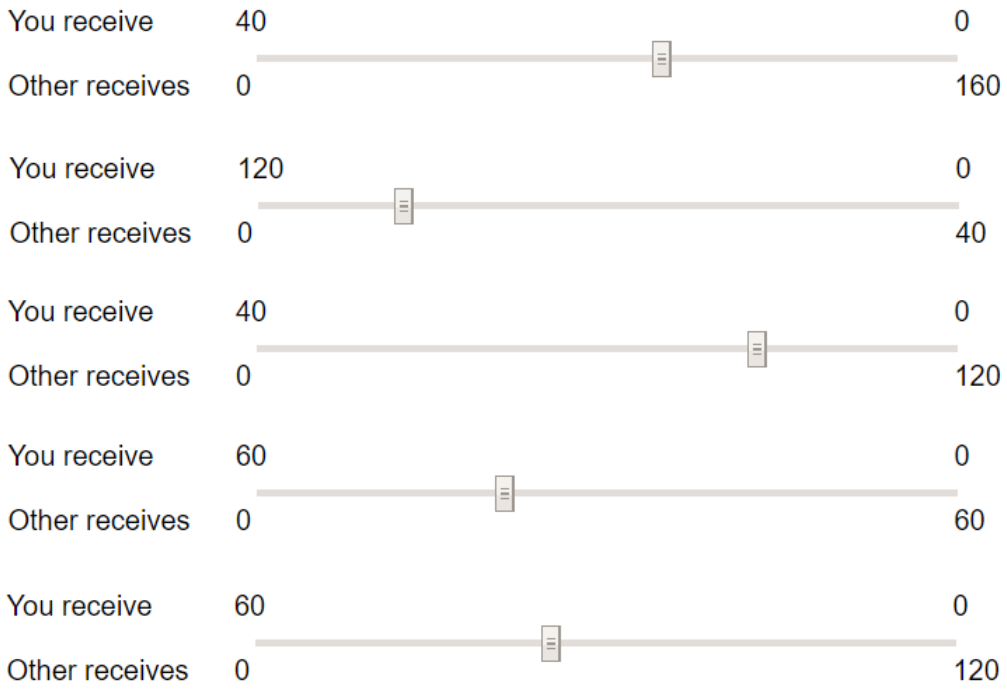


Note however that each question was presented in a different screen, and the order of presentation was randomized for every subject. Furthermore, 50 points were worth 1 euro both the subject, and the other anonymous external participant. Detailed instructions provided to the subjects, as well as a screenshot highlighting one of the above five questions are described below.

Instructions: You are asked to answer a series of 5 questions, each of which consists of selecting an allocation of points that you most prefer between yourself and an anonymous randomly selected person who is participating to a different experiment in this lab. At the end of the study, we will randomly select your allocation for 1 of the 5 questions to determine the payments for both you and the other person in this part. Your decisions will remain unknown to the other persons you are matched with.

E.3 Risk preferences

The risk preference measure was adapted from Holt and Laury [2002] and consisted of a series of five binary choices between lotteries, presented as in the figure below.



E.4 Personality test

Non incentivized measures were used through a simplified version of the Big Five personality inventory test adapted from Rammstedt and John [2007], as shown below.

Question 1

Please select your preferred allocation on the slider below
(values are in points, with 50 points = 1 euro):

You receive	17
Other receives	93

You receive	40	0
Other receives	0	160

Next

You are now asked to make 5 independent choices between two lotteries. According to **Lottery A**, you can win 2.00€ with a certain probability **p**, and 1.60€ otherwise. According to **Lottery B**, you can instead win 3.85€ with the same probability **p**, and 0.10€ otherwise. For each of the following 5 choices, which only differ in the value of the probability **p**, please select the lottery that you prefer. At the end of the study, we will randomly select one of your 5 preferred lotteries to determine your payment in this question.

	Lottery A			**Lottery B**
Choice 1:	2.00€ with probability 20/100, 1.60€ with probability 80/100	○	○	3.85€ with probability 20/100, 0.10€ with probability 80/100
Choice 2:	2.00€ with probability 35/100, 1.60€ with probability 65/100	○	○	3.85€ with probability 35/100, 0.10€ with probability 65/100
Choice 3:	2.00€ with probability 50/100, 1.60€ with probability 50/100	○	○	3.85€ with probability 50/100, 0.10€ with probability 50/100
Choice 4:	2.00€ with probability 65/100, 1.60€ with probability 35/100	○	○	3.85€ with probability 65/100, 0.10€ with probability 35/100
Choice 5:	2.00€ with probability 80/100, 1.60€ with probability 20/100	○	○	3.85€ with probability 80/100, 0.10€ with probability 20/100

Next

How well do the following statements describe your personality?

I see myself as someone who...	Disagree strongly	Disagree a little	Neither agree nor disagree	Agree a little	Agree strongly
1. ... is reserved	○	○	○	○	○
2. ... is generally trusting	○	○	○	○	○
3. ... tends to be lazy	○	○	○	○	○
4. ... is relaxed, handles stress well	○	○	○	○	○
5. ... has few artistic interests	○	○	○	○	○
6. ... is outgoing, sociable	○	○	○	○	○
7. ... tends to find fault with others	○	○	○	○	○
8. ... does a thorough job	○	○	○	○	○
9. ... gets nervous easily	○	○	○	○	○
10. ... has an active imagination	○	○	○	○	○

Next

F Appendix tables and figures

F.1 Regression tables

F.2 Appendix figures

Table 8: Regression analysis in the baseline treatments: time fraction

	Time fraction of being most connected (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	5.30***	5.29***		
	(0.50)	(0.50)		
Effort × Large group	0.84***	0.83***		
	(0.13)	(0.13)		
Indegree ratio (%) × Small group			-0.02	0.04
			(0.13)	(0.15)
Indegree ratio (%) × Large group			-1.10***	-1.15***
			(0.13)	(0.15)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.407	0.409	0.086	0.136

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 9: Scale effects on effort and outdegree in the baseline treatments

	Mean effort			Mean outdegree		
	most connected	2nd most connected	others	most connected	2nd most connected	others
$N = 50$	6.61***	7.27***	0.32	0.99***	0.74	0.15***
	(1.08)	(1.41)	(0.32)	(0.28)	(0.52)	(0.06)
Average in small group	8.77	5.24	2.65	0.20	0.62	0.90
Number of observations	60	60	1120	60	60	1120
R-squared	0.61	0.59	0.04	0.49	0.47	0.04

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 10: Scale effects on effort and outdegree in the baseline treatments

	Mean effort			Mean outdegree		
	most connected	2nd most connected	others	most connected	2nd most connected	others
$N = 100$	6.64***	11.06***	0.88***	2.03**	0.72*	0.30***
	(1.54)	(1.10)	(0.32)	(0.77)	(0.38)	(0.05)
Average in small group	8.77	5.24	2.65	0.20	0.62	0.90
Number of observations	55	55	1630	55	55	1630
R-squared	0.62	0.83	0.04	0.53	0.46	0.06

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 11: Scale effects on payoffs in the baseline treatments

	Median payoff		
	most connected	2nd most connected	others
$N = 50$	-40.81***	-51.09**	28.82***
	(10.20)	(23.61)	(1.73)
Median in small group	86.50	81.00	85.00
Number of observations	60	60	1120
R-squared	0.39	0.23	0.11

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 12: Scale effects on payoffs in the baseline treatments

	Median payoff		
	most connected	2nd most connected	others
$N = 100$	16.54	-25.41*	53.20***
	(29.95)	(14.54)	(2.77)
Median in small group	86.50	81.00	85.00
Number of observations	55	55	1630
R-squared	0.20	0.38	0.14

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 13: Regression analysis when information on others' payoff is observable: time fraction

	Time fraction of being most connected (%)		Median payoff	
	(1)	(2)	(1)	(2)
Effort × Small group	6.20***	6.13***		
	(0.71)	(0.71)		
Effort × Large group	0.43***	0.43***		
	(0.15)	(0.15)		
Time fraction (%) × Small group			0.24**	0.28***
			(0.11)	(0.07)
Time fraction (%) × Large group			1.15***	1.13***
			(0.09)	(0.14)
Additional controls	No	Yes	No	Yes
Number of observations	2740	2740	2740	2740
R-squared	0.302	0.305	0.010	0.004

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, a dummy for large group, and dummies for rounds. Additional controls include age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 14: Treatment effects on effort and outdegree

	Mean effort			Mean outdegree		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Payoff info	-1.35*	0.21	-0.04	0.74*	0.29*	0.05
	(0.76)	(0.65)	(0.37)	(0.45)	(0.17)	(0.07)
$N = 50$	6.53***	6.48***	0.30	1.48***	0.95**	0.17***
	(1.21)	(1.37)	(0.33)	(0.36)	(0.44)	(0.06)
Payoff info $\times$ $N = 50$	-8.90***	-5.72***	-0.39	-0.29	0.00	-0.10
	(1.83)	(1.93)	(0.44)	(0.57)	(0.54)	(0.08)
Average in Baseline50	15.70	11.33	2.84	1.18	1.28	1.04
Number of observations	120	120	2240	120	120	2240
R-squared	0.54	0.38	0.04	0.43	0.41	0.07

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 15: Treatment effects on effort and outdegree

	Mean effort			Mean outdegree		
	most connected	2nd most connected	others	most connected	2nd most connected	others
Payoff info	-0.55	0.20	0.06	-0.57	0.02	0.03
	(0.79)	(0.70)	(0.36)	(0.89)	(0.12)	(0.07)
$N = 100$	6.53***	10.26***	0.87***	0.75	0.86**	0.30***
	(1.33)	(1.48)	(0.32)	(1.94)	(0.34)	(0.06)
Payoff info $\times$ $N = 100$	-9.82***	-12.48***	-1.28***	2.30	-0.89**	-0.01
	(1.65)	(2.01)	(0.41)	(2.47)	(0.37)	(0.08)
Average in Baseline100	14.35	15.73	3.48	2.86	1.40	1.17
Number of observations	110	110	3260	110	110	3260
R-squared	0.53	0.68	0.07	0.54	0.32	0.04

Notes: Robust standard errors, clustered by individual subject, are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

Table 16: Treatment effects on payoffs

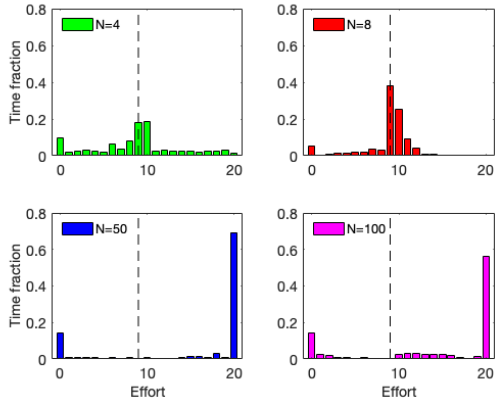
	Median payoff		
	most connected	2nd most connected	others
Payoff info	0.76	-18.73***	-6.07**
	(8.94)	(4.04)	(2.38)
$N = 50$	-57.53***	-48.79**	29.99***
	(11.82)	(19.15)	(2.46)
Payoff info $\times$ $N = 50$	142.27***	100.25***	-14.96***
	(18.28)	(29.02)	(2.26)
Median in Baseline50	48.50	51.00	118.00
Number of observations	120	120	2240
R-squared	0.21	0.14	0.11

Notes: Robust standard errors are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.

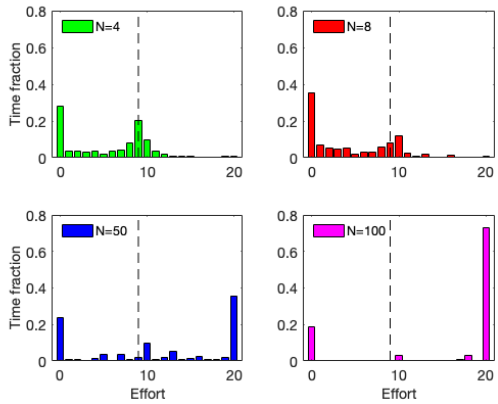
Table 17: Treatment effects on payoffs

	Median payoff		
	most connected	2nd most connected	others
Payoff info	-0.64	-15.63***	-11.36***
	(12.59)	(5.31)	(2.05)
$N = 100$	40.61**	-43.22	51.44***
	(17.34)	(27.99)	(1.85)
Payoff info $\times$ $N = 100$	92.09	160.98***	-20.70***
	(150.87)	(29.02)	(2.67)
Median in Baseline100	153.00	42.50	140.50
Number of observations	110	110	3260
R-squared	0.07	0.26	0.13

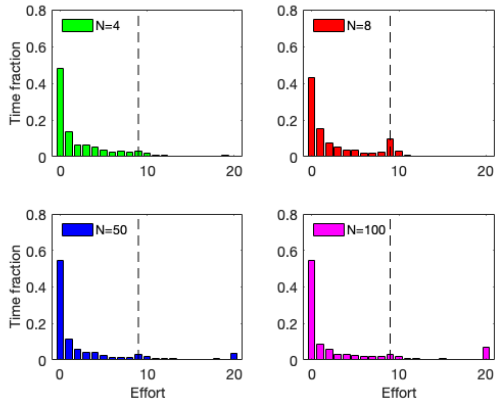
Notes: Robust standard errors are report in parenthesis. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. All regressions include a constant, round dummies, age, female, education, comprehension test score, experimental measures of risk aversion and altruism, and Big 5 personality.



(a) the 1st most connected

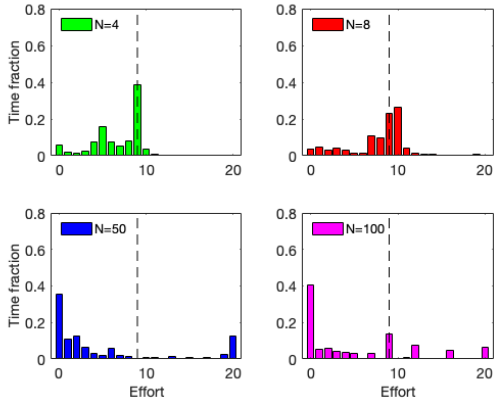


(b) the 2nd most connected

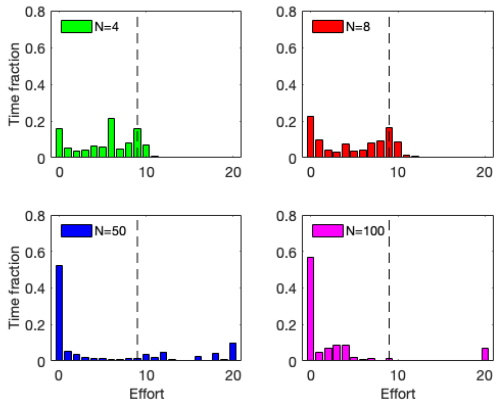


(c) the others

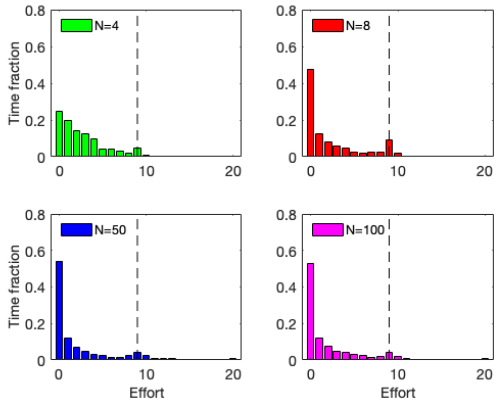
Figure 21: Distribution of efforts in the baseline treatment



(a) the 1st most connected

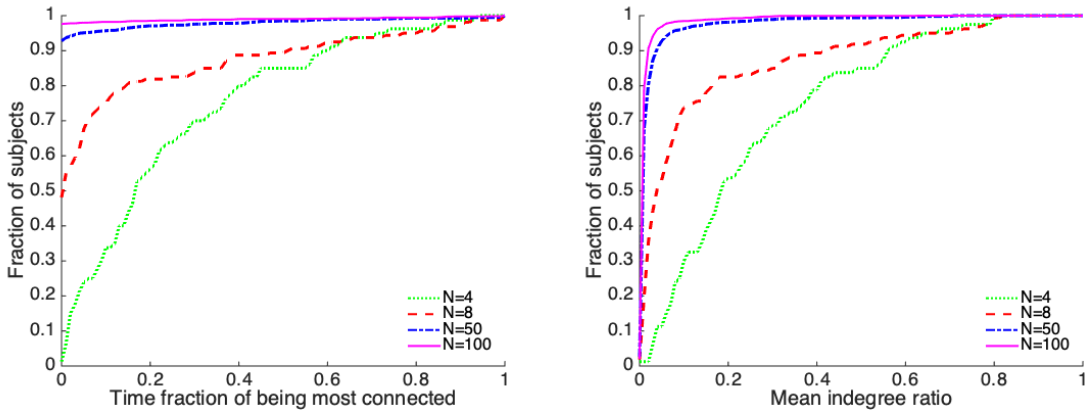


(b) the 2nd most connected



(c) the others

Figure 22: Distribution of efforts in the payoff information treatment



(a) Time fraction

(b) Mean indegree ratio

Figure 23: Distribution of linking: information on others' payoff

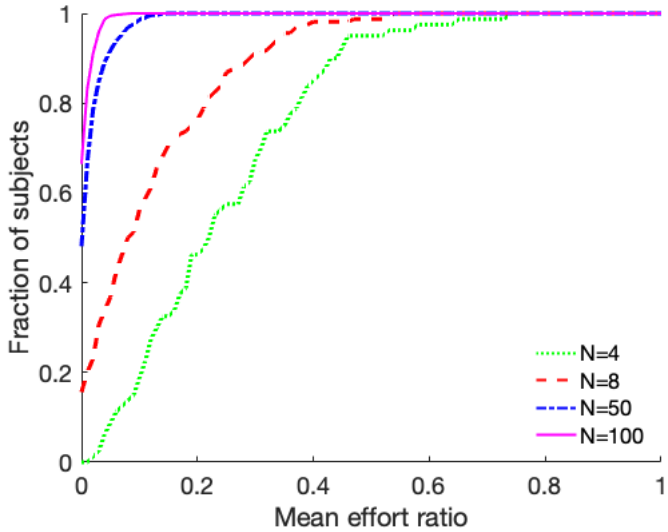
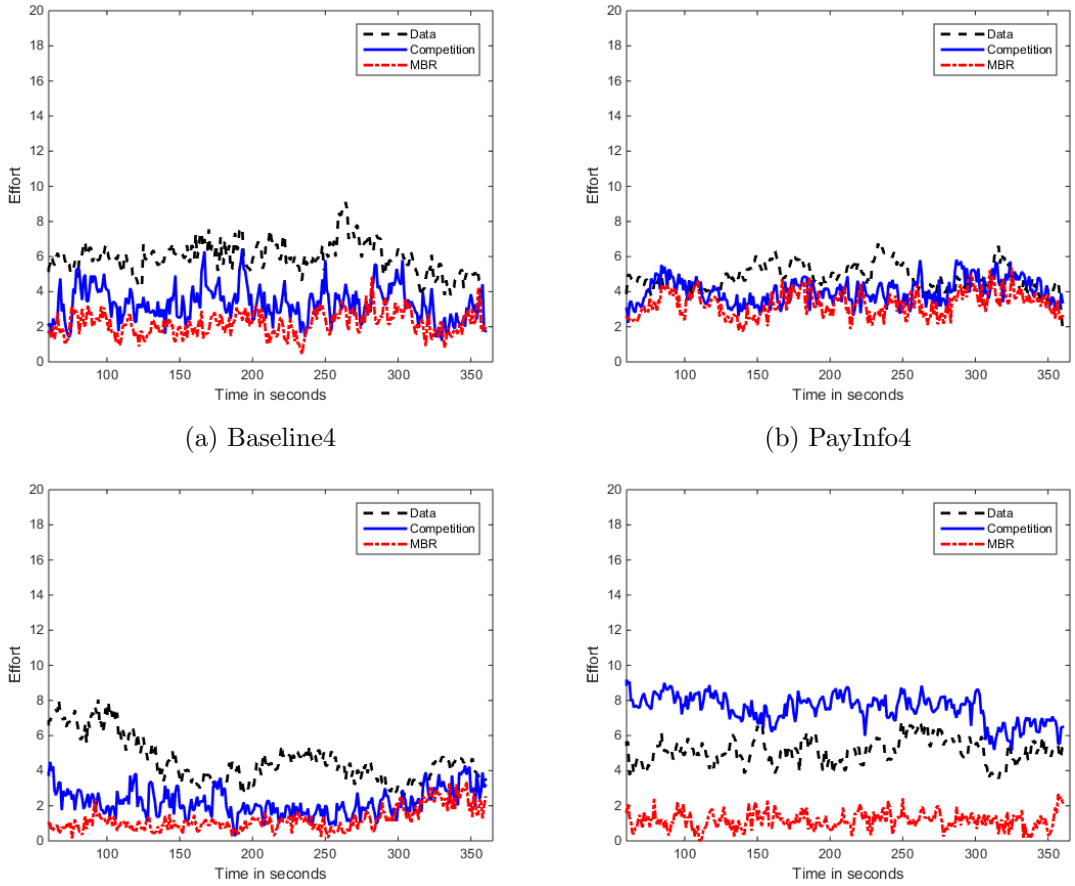


Figure 24: Distribution of Efforts in the payoff information treatments



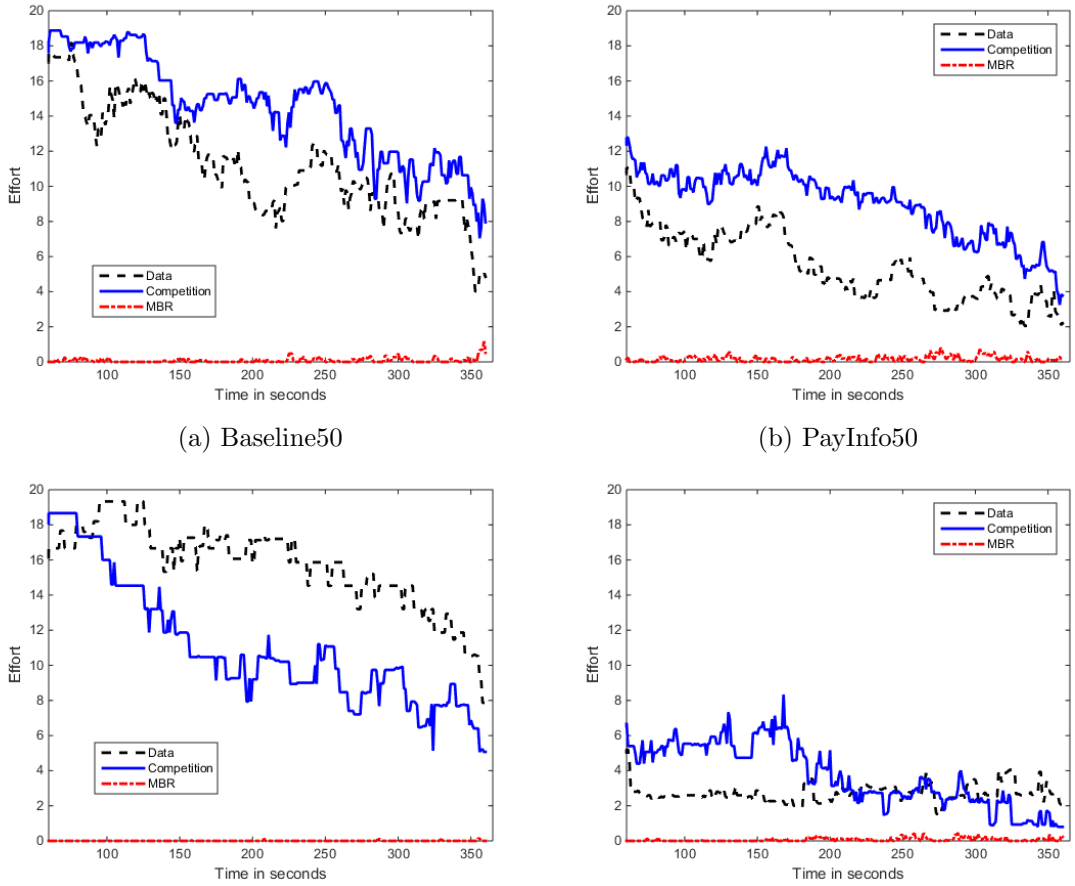
(a) Baseline4

(b) PayInfo4

(c) Baseline8

(d) PayInfo8

Figure 25: Fitting effort dynamics with learning rules: 2nd most connected



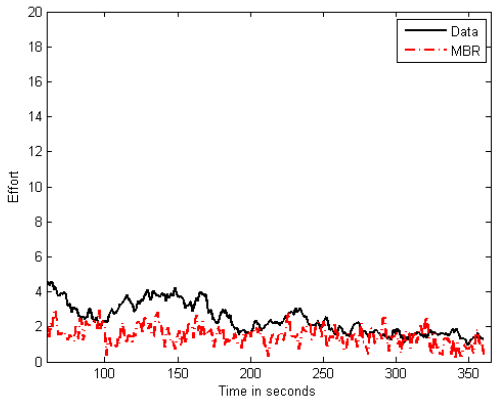
(a) Baseline50

(b) PayInfo50

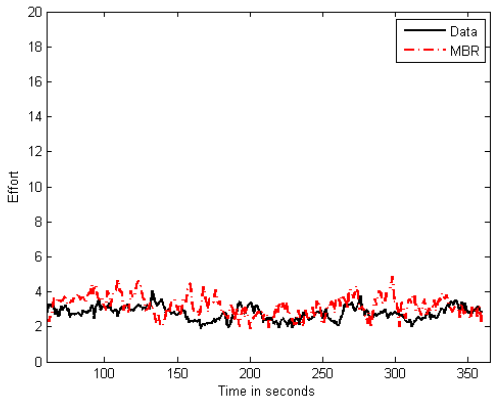
(c) Baseline100

(d) PayInfo100

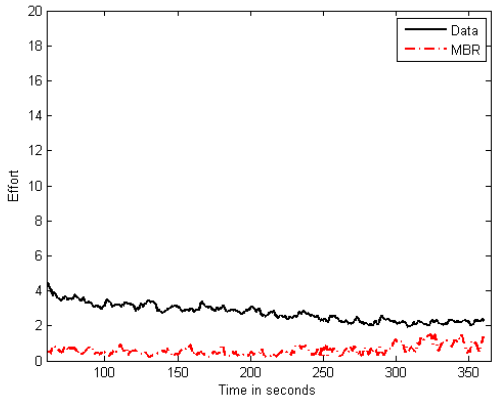
Figure 26: Fitting effort dynamics with learning rules: 2nd most connected (cont.)



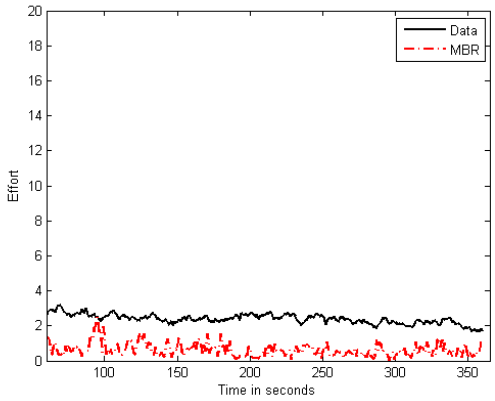
(a) Baseline4



(b) PayInfo4

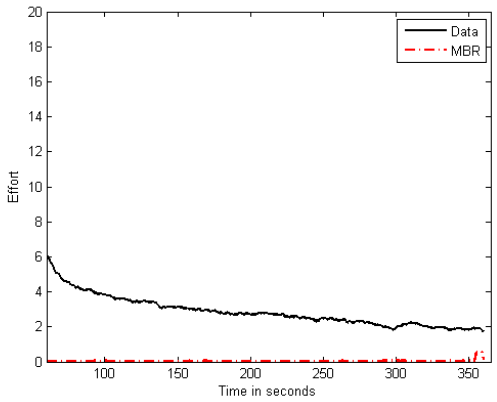


(c) Baseline8

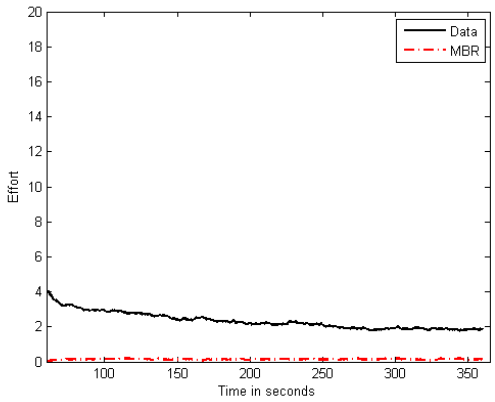


(d) PayInfo8

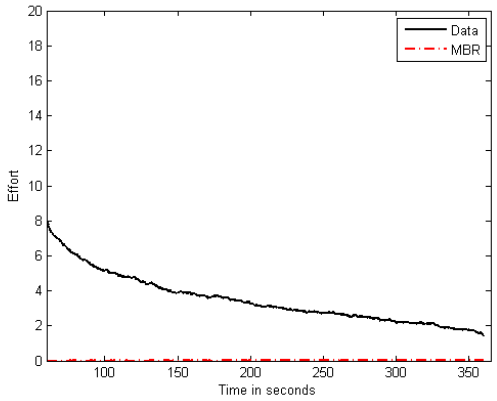
Figure 27: Fitting effort dynamics with learning rules: others



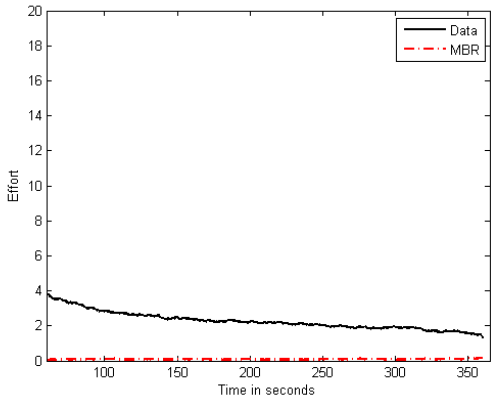
(a) Baseline50



(b) PayInfo50



(c) Baseline100



(d) PayInfo100

Figure 28: Fitting effort dynamics with learning rules: others (cont.)