Emergent Star Networks with Ex Ante Homogeneous Agents

Rong Rong and Daniel Houser

August 2012

Discussion Paper
Emergent Star Networks with Ex Ante Homogeneous Agents

Abstract: The acquisition and dispersion of information, a critical aspect of economic decisions, can occur through a network of agents (Jackson, 2009). Empirical and theoretical findings suggest that an efficient information dispersion network takes the form of a star: small numbers of agents gather information and distribute it to a large group. Despite these findings, controlled tests of this theory failed to find evidence of the emergence of such networks with an exception of Goeree et al (2008), which reports frequent star network formation with ex ante heterogeneous agents and perfect information. Empirical evidence suggests that these conditions may not always be feasible in natural environments (Feick and Price, 1987; Conley and Udry, 2010). Complement to earlier experimental research, we study institutional conditions under which star networks emerge in the presence of ex ante homogeneous agents. We find that investment limits and the “right-of-first-refusal,” both of which regularly coexist with star networks in natural environments, have a surprisingly strong ability to promote the formation of star networks with ex ante homogeneous agents. Using a cluster analysis, we trace the large positive effects of these institutions to the impact they have on individuals’ behavioral rules. In particular, we find that these institutions encourage individual rationality as well as positive habits, which then lead to the frequent and stable emergence of star networks. Our results may have important implications for environments characterized by ex ante homogeneous agents, e.g., those found in technology, marketing and agricultural sectors.

JEL classification: D85, D02, C92
Keywords: Social networks, star network formation, cluster analysis, experiments
I. Introduction

How information is initially acquired and subsequently dispersed among people is widely studied in economics (Rogers, 1995). In many relevant contexts, it occurs through networks of agents (Jackson, 2009). Empirical and theoretical findings suggest that efficient information networks take the form of a star: small numbers of agents gather information and then distribute it to a larger group (Weimann, 1994; Bala and Goyal, 2000; Galeotti and Goyal, 2010). Despite the theoretical advances, one persistent challenge has been discovering the conditions under which star networks emerge within controlled laboratory environments. One study that successfully generated star networks imposed the requirements of ex ante agent heterogeneity and perfect information (Goeree, et al., 2008). Unfortunately, however, these conditions may not characterize some natural environments (Feick and Price, 1987; Conley and Udry, 2010); therefore, the results leave open the question of which naturally-occurring conditions work to promote star network emergence in presence of ex ante homogeneous agents.

The earliest works on star networks date back to the 1950s. In their pioneering paper, Katz and Lazarsfeld (1955) coined the term “opinion leaders” to describe a small subset of highly connected people\(^1\). Half a century later, studies continue to provide empirical support for the existence of opinion leaders (Weimann, 1994; Katz and Lazarsfeld, 2006). Opinion leaders clearly make a difference. For instance, empirical evidence has shown that words from opinion leaders boost sales of consumer products (Godes and Mayzlin, 2009), contribute to the prevention of AIDS (Kelly et al., 1992), and transmit political thought and ideas (Roch, 2007). Given opinion leaders’ importance in disseminating information, people in both the private and public sectors are eager to discover how to locate and influence them (Iyengar et al., 2008). A deeper understanding of the emergence and characteristics of star networks, a stylized model of opinion leader, could help to facilitate such efforts.

Research on star networks connect to the empirical literature of so-called scale-free networks, referring to networks where most nodes are connected to a very small number of nodes, and a few nodes are connected to a very large number of other nodes. Examples of scale-free

\[^1\] The concept of “influentials” can largely be used interchangeably (Merton, 1968; Gladwell, 2000).
networks include scientific citation networks, internet Pagerank networks, airport networks, etc\textsuperscript{2}. Star networks are scale-free networks, and thus understanding how star networks emerge can inform more generally how empirically relevant scale-free networks emerge in natural environments.

Theoretical studies of star networks have shown that under certain conditions, star networks emerge endogenously and include efficient and stable equilibria (Bala and Goyal, 2000; Galeotti and Goyal, 2010\textsuperscript{3}). Certain crucial conditions underlie the formation of equilibrium star networks, namely that: (1) information can be shared (meaning it is non-rival); and (2) agents are able to form links unilaterally\textsuperscript{4}. While strong, these conditions are easily implementable in laboratory tests and are attractive in that many environments where information dispersion is important are characterized by these conditions\textsuperscript{5}.

To investigate star-network formation, economists have begun to collect laboratory data from participants in various network environments (Callander and Plott, 2005; Falk and Kosfeld, 2003; Goeree et al, 2008). Loosely speaking, in a typical network formation experiment, players decide how to form “links” with other players in light of the benefits those links confer. Unfortunately, these studies have almost always failed to find the emergence of star networks\textsuperscript{6}. To our knowledge, the single exception is Goeree et al (2008), which found star networks to emerge only after imposing substantial agent heterogeneity among perfectly-informed participants. The authors argued that this heterogeneity promoted star networks by simplifying the network coordination problem.

Despite their success, the conditions of ex ante heterogeneity used by Goeree et al (2008) may not always be satisfied outside the laboratory (Feick and Price, 1987; Conley and Udry, 2010). Complementing the findings of Goeree et al (2008), we investigate network formation in the presence of ex ante homogeneous agents. In particular, we study whether certain naturally-

\textsuperscript{2} Detailed examples of scale-free networks could be found at (Barabasi, 2004).
\textsuperscript{3} Some non-game-theoretical models of star network formation build upon preferential attachment and study the behavior of large networks (Barabasi and Albert, 1999; Jackson and Rogers, 2007). A direct test of those models needs validation of its behavioral assumptions, which is difficult to achieve with a lab experiment.
\textsuperscript{4} Jackson and Wolinsky (1996) discussed two cases in which those two conditions are lacking. They found that network efficiency and stability is hard to achieve under regular payoff functions.
\textsuperscript{5} Knowledge generated in academia is one example. Open source software is another example. This excludes cases where information is protected by IPR and dispersion of that information requires bilateral agreements such as those found in some enforceable contracts.
\textsuperscript{6} Falk and Kosfeld (2010) found equilibrium “wheel” networks to emerge, but were not able to observe the formation of equilibrium star networks.
occurring institutional features may promote star-network formation. We collect data from laboratory experiments to examine whether efficient star network formation can be promoted by: (1) making network investment sequential; (2) limiting investment in acquiring information; and (3) enforcing the “right of first refusal,” which ensures that an initial investor is able, should they desire to do so, to continue his/her investment.

In selecting treatment conditions, we choose those that might bear relevance to the way networks emerge in natural settings. To begin, it is very common to create networks through a process of sequential decisions. For example, online networking sites, such as Facebook or Twitter, regularly ask current users to invite their friends to join the site, and then those friends ask their friends. We incorporate sequential decisions into our experiment design.

Investment limits are a regular feature of natural environments, as (1) technologies are too expensive to allow repeated R&D investment (Dimasi et al, 2003; Dimasi and Grabowski, 2007) and (2) government policies exist to prevent wasteful repeated investments (Tran, 2009). Moreover, when personal relationships are at stake, information investment may face natural constraints with regard to time or distance (Marsden and Campbell, 1984). In this paper, investment limits will be modeled as a type of individual budget constraints.

The “right of first refusal” (which we denote by RFR) is a business contract where the privilege of investment is granted to an investor who has invested in the immediate past. Aside from being a common business practice, RFR emerges regularly whenever economic outcomes favor persistent investments on one agent rather than the spreading of resources across multiple smaller investments. Take, for instance, the situation in which low income families in developing countries decide which child to send to school; firms choosing which employees to give additional training; or funding agencies selecting which scholars’ projects to fund. In each of these cases, the predictability of the investor’s identity serves as the coordination device that may facilitate star network formation. Our experiment design allows us to vary the existence of RFR and measure its impact on star network formation.

The key finding of our paper is that star networks can reliably emerge with ex ante homogeneous agents. As noted above, this complements previous findings by Goeree et al (2008) and helps to explain the emergence of star networks in natural environment where agent heterogeneity may play a smaller role. In particular, we find that combining investment limits with RFR generates robust star networks, and that these networks are “stable” in the sense that
they tend rapidly to reemerge after falling into disequilibrium. The effect persists in either simultaneous or sequential decision environments. Moreover, investment limits alone generate about half of the equilibrium outcomes compared to when they are combined with the RFR. Surprisingly, sequential decision alone does not seem to promote equilibrium networks.

We also examine individual decision rules using cluster analysis and find that the behavior of our lab subjects separates into clearly defined behavioral clusters. We then investigate how different institutions may shift subjects into different types of behavioral rules. Our analysis demonstrates that investment limits in simultaneous decision environments promote the formation of positive habits, which translate into more frequent coordination in star networks. Investment limits in sequential environments generate frequent star networks, largely due to the fact that agents are more likely to follow the rule of rationality predicted by the theory.

Our investigation provides both methodological and substantive contributions. Methodologically, we use cluster analysis to distinguish different behavioral rules used by subjects. These differences help to explain why different institutions lead to different levels of star network formation. To our knowledge, we are the first to implement cluster analysis to shed light on the behavioral rules used in social networks, and the first to explain institutional impacts in social network formation environments using these behavioral rules.

Substantively, the findings of our paper have direct implications for information dispersion, which is especially relevant to, e.g., the technology and agricultural sector. Our results suggest that: (1) difficulties with coordination may lead to undesirable network outcomes; and (2) institutional features that promote sustained investment by a single individual facilitate the formation of efficient networks.

The remainder of the paper is organized as follows: The next section briefly reviews the theoretical and experimental literature on network formation. Section 3 lays out the theoretical background of the study. Section 4 presents the experimental design and procedure, and sets up the hypothesis. Section 5 reports experimental results. Section 6 discusses cluster analysis. Section 7 concludes.
II. Literature Review

II.1 Theoretical work on star network formation

Many theoretical studies have attempted to shed light on the process of network formation in general (Jackson, 2003), and recently specific theoretical progress has been made on understanding the conditions under which star networks can form (Bala and Goyal, 2000; Bramoulle et al., 2004; Galeotti and Goyal, 2010). For all the cases that we study in this paper, equilibrium star networks are also efficient. Star networks feature asymmetry in equilibrium actions by participants, because it pays to send links when others invest and vice versa. Note that this environment is characterized by “strategic substitutes”, and includes in general both anti-coordination games and games related to public goods provision.

An early paper by Bala and Goyal (2000) studied an environment with non-rival network goods and the possibility of forming links unilaterally. They found that star networks emerge in equilibrium only when the benefit of information flows between two agents regardless of who sends the link. Their study was followed by Bramoulle et al. (2004), who examined network formation in an anti-coordination game. They found that the shape of the equilibrium network need not be a star; with the exact network shape depending on the cost of link formation. More recently, a study by Galeotti and Goyal (2010) extended the model of Bala and Goyal (2000) by endogenizing the choice to invest. Their study showed that star networks emerge in equilibrium as well. These advances of course leave open the question of whether the conditions required by theory are sufficient to generate star networks reliably in a controlled laboratory environment.

We now turn to a detailed review of the empirical studies of network formation in the laboratory.

II.2 Experiments on star network formation

Despite the abundance of empirical evidence related to star networks, we are aware of only three experimental studies on star network formation (Callander and Plott, 2005; Falk and

---

7 Bramoulle and Kranton (2007) discussed public goods provision in network settings extensively. However, networks in their study are exogenous.
8 In their model, if only the link sender receives information from the link receiver, the equilibrium network is a wheel.
9 Galeotti and Goyal (2010) predicts peripheral-sponsored stars, which are different from the center-sponsored stars found in Bala and Goyal (2000) and Bramoulle et al. (2004).
Kosfeld, 2003; Goeree et al, 200811). Falk and Kosfeld (2003) tested the theory of Bala and Goyal (2000). In particular, they studied whether and how equilibrium networks can form under “one-way” and “two-way” information flows. In contrast with theoretical predictions, they found that when information flows two ways the network fails to converge to a star. They concluded that the need for asymmetric strategies combined with inequality aversion might contribute to the difficulty in realizing star networks.

Callander and Plott (2005) also tested Bala and Goyal (2000) in the lab. They considered various conditions that differed in terms of the linking cost, as well as the value of information. They also examined the impact of having network agents with heterogeneous payoff structure, an issue unaddressed by the model. Their main finding was that star networks did not consistently emerge under theoretical conditions, and that even introducing payoff heterogeneity did not lead to systematic formation of star networks. Consequently, they report that “significant and persistent inefficiency” is a feature of all of their network environments.

In light of the complications with generating star networks, and following Callander and Plott (2005), Goeree et al (2008) explored whether common knowledge of agent heterogeneity combined with two-way information flows might promote star networks. They reported that: (1) compare to homogeneous agent treatment, significantly more stars are observed when agents’ payoff are heterogeneous12; and (2) perfect information about the nature of heterogeneity plays an important role in facilitating the coordination on star networks.

Like the above studies, we explore what conditions may facilitate the emergence of star networks. But in contrast, our study emphasizes the importance of homogeneous agent assumption and explores how institutional characteristics may impact network formation in presence of agent homogeneity.

While Callander and Plott (2005) and Goeree et al (2008) demonstrated the importance of individual heterogeneity in network environments, there may be some environments where individual differences play a smaller role. For instance, information about heterogeneity may not always be easily available in natural environments, due to the fact that it goes unobserved. Indeed, substantial empirical research on market mavens has found no differences between the

---

11 Another study by Deck and Johnson (2004) tests the model of Droste et al (2000) where the predicted network is a lattice.
12 According to their experimental data, the heterogeneity on cost of linking does not seem to significantly promote star network formation.
observable characteristics of agents who play different roles in the network (Feick and Price, 1987; Geissler and Edison, 2005; Wiedman et al, 2001; Williams and Slama, 1995). Others have pointed out that obtaining information about the costs and benefits of other network agents in agricultural environment may be difficult, given that people have an incentive to conceal their private information (Conley and Udry, 2010). Moreover, game theoretical models that predict star networks have always assumed ex ante agent homogeneity\textsuperscript{13}.

In view of the fact that individual-level information is costly and sometimes infeasible to obtain, and to avoid introducing artificial focal points, we design our experiments to include ex ante homogeneity. We investigate conditions under which ex ante identical agents will take asymmetric equilibrium actions to establish efficient and stable star networks.

III. Theoretical Background

Our study is based on the model of network game in Galeotti and Goyal (2010). In their model, a group of identical rational agents face the choice of either investing in information or obtaining it less expensively by linking to another who currently invested in information. The level of investment by agent \(i\) is discrete \(x_i \in \{0,1\}\). The set of links sent by agent \(i\) is denoted by a vector 
\[
g_i = (g_{i1}, \ldots, g_{i,n-1}, g_{i,n})
\]
where \(g_{ij} = 1\) if player \(i\) sent a link to player \(j\). Linking choices are then combined to determine the directed network structure 
\[
g = (g_1, g_2, \ldots, g_n)
\]
The key assumptions of the model are that information is non-rival and flows both ways across network links.\textsuperscript{15}

The non-directed version of the network is denoted by \(\bar{g}\), where \(\bar{g}_{ij} = \max\{g_{ij}, g_{ji}\}\) for each agent \(i\) and \(j\). Define \(N(i; g) = \{j : g_{ij} = 1\}\) as the set of agents to whom \(i\) has sent a link and \(N(i; \bar{g}) = \{j : \bar{g}_{ij} = 1\}\) as a set of agents with whom \(i\) has been connected. The payoff to agent \(i\) is
\[
\pi_i(x_i, g_i) = f(x_i + \sum_{j \in N(i; \bar{g})} x_j) - cx_i - |N(i; g)|k
\]
\textsuperscript{13} Jackson and Lopez-Pintado (2011), Larrosa and Tahme (2011), and Vandenbossche and Demuynck (2010) developed models with heterogeneous agents. However, none of these relate to incentives associated with information acquisition or diffusion; therefore, the predictions generated from those models are not star-shaped networks.

\textsuperscript{14} A directed graph is a graph where the edges have a direction associated with them.

\textsuperscript{15} These assumptions are important due to the fact that they closely characterize certain situations of information dispersion in natural environments. For example, knowledge about agricultural technology is mostly non-rival, and could be shared between personal connections of farmers regardless of the linking direction.
where \( c > 0 \) reflects the cost of investing, \( k > 0 \) is the cost of sending one link and \( |N(i; g)| \) refers to the cardinality of the set \( N(i; g) \).

Different specifications for \( f \) define different types of games. In this paper, we follow Galeotti and Goyal (2009) and assume \( f \) is a step function

\[
\begin{cases}
  f(y_i) = 1 & \text{if } y_i \geq 1 \\
  f(y_i) = 0 & \text{if } y_i < 1
\end{cases}
\]

Where \( y_i = x_i + \sum_{j \in N(i; g)} x_j \). The above return function \( f(y_i) \) resembles the payoff structure of best shot game in the widely studied public good games literature. The advantage to using a step function is that it provides sharp equilibrium predictions that can be more easily tested in the laboratory\(^{16}\).

According to theory, every equilibrium of the network best shot game is a star network when \( k < c \).\(^{17}\) The intuition is as follows: if in equilibrium the sole investor deviates and does not invest, then the group obtains no information, implying a lower payoff for everyone including the investor. Similarly, if a person who has linked to the investor deviates by not linking, choosing to link to another (who in equilibrium cannot have the information), or becoming an investor oneself, then in all cases such deviations clearly lead to lower payoffs. Therefore, the star network is a Nash equilibrium. Note also that all star network equilibria in the best shot game are efficient (in the sense that equilibria are not Pareto ranked). This feature of the network best shot game, as well as its clean equilibrium predictions, leaves it ideal for laboratory testing. In the following section we detail our design, which follows the network best shot game closely.

**IV. Experiment design and hypothesis**

Our experiment is designed to examine how naturally-occurring institutions affect star network formation with ex ante homogenous agents. Institutional characteristics such as sequential decisions, investment limits and the RFR often coexist with star networks. We conjecture that these institutional characteristics may be important conditions for the formation of star networks in naturally occurring environments. Our laboratory study brings these

\(^{16}\) Instead of star network, the general model prediction a so-called core-peripheral network, where a few interconnected agents invest in information while the rest of agents connects to them. A star network is a special core-peripheral network that includes a single agent in the core.

\(^{17}\) When \( k > c \), the unique equilibrium is an empty network.
institutional features into a controlled laboratory setting and examines the effect of each on star network formation.

IV.1 Experiment design

IV.1.1. General environment

Our experiment design is based on Galeotti and Goyal (2010), but modified as described in Galeotti and Goyal (2007)\(^{18}\). This modification leads to the sharp prediction that star networks are the unique Nash equilibrium, and are also efficient\(^{19}\). To the best of our knowledge, our study is the first to examine the network formation process where agents make simultaneous linking and information investing decisions.

Each experimental session includes 16 subjects randomly divided into four groups. All subjects participate in three stage games. Each stage game consists of a random number of rounds\(^{20}\). Groups are fixed during each stage game, and each group member holds a unique ID: J,K,L or M. We avoid using “A” as an ID because it may be focal.

In each round, decision-makers decide to whom to link among their other three group members and also whether to purchase information. Table 1 details the costs and benefits associated with each action a player can take. If a participant purchases information, she pays a cost of E$0.9 and earns the value of information, E$3, with certainty. On the other hand, if a player decides to send a link to another player, she pays a cost of E$0.5 per link. When one subject links to another subject who has purchased information, the subject who chooses to link also earns E$3. Subjects who link to other subjects that have not purchased the information pay a cost of E$0.5, but earn nothing. Costs and payoffs remain fixed throughout all three stage games and all treatments.\(^{21}\).

---

\(^{18}\) This environment reflect in some ways a job contact network where people turn to other people to obtain information on job opening.

\(^{19}\) In this earlier version of the paper, Galeotti and Goyal (2007) described a best-shot game where the investment decision is binary. An agent could choose either to invest in one unit or no units of the information. The optimal level of investment for a group is also set to be one, so that any additional investment is inefficient.

\(^{20}\) There are always at least 4 rounds in a stage. After round 4, the game has a random stopping probability of 0.04 at any given round. To keep control over the length of the real experiment, we use the predetermined length 16, 44 and 24 for experimental stages I, II and III respectively. Those numbers are generated using a random number device. The practice stage always lasts for 8 rounds.

\(^{21}\) We choose these parameters to ensure that the predicted equilibrium and efficient networks are star-shaped.
Table 1. Costs and benefits associated with player’s actions

<table>
<thead>
<tr>
<th>Parameter value</th>
<th>Cost of sending link</th>
<th>E$0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost of investing in information</td>
<td>E$0.9</td>
<td></td>
</tr>
<tr>
<td>Value of information</td>
<td>E$3</td>
<td></td>
</tr>
<tr>
<td>Number of player in a group</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

Subjects submit their decisions using the decision screen (see Appendix A, Fig. 1). Then, a display screen informs all players of the current network outcome and each group member’s payoff (see Appendix A, Fig. 2).

Within each of the three stage games, the payoff is determined by the accumulated earnings over all rounds. Players are informed about their own stage payoff at the end of each stage. They are also reminded that they will be re-matched with players with whom they have not played previously, and that their stage payoff will not be carried over to the new stage. Each subject’s earnings for the experiment are determined by one randomly-determined stage game (each with equal probability).

IV.1.2 Treatment Design

Within the general experimental environment described above, we study the effects of three institutional characteristics of network formation. We examine sequential decisions and investment limits on network formation, both individually and jointly, using a two-by-two treatment design. A fifth treatment then studies the effect of the “right of first refusal.”

In a two-by-two design, we vary the sequence of decisions in one dimension to be either sequential or simultaneous. In simultaneous treatments, subjects from the same group make their decisions at the same time, not knowing what other subjects would choose. In sequential treatments, only one subject makes a decision per round. Players make decisions according to the alphabetical order of their ID (first J, then K, L and finally M) with full knowledge of the choices made by earlier decision makers. Further, players earn money even on rounds for which they do not make a decision, with their payoff determined by their most recent previous choice in combination with the choices of others.

The second dimension of our design varies the existence of investment limits. Absent investment limits, players can invest in information and links at will, independently of other players’ decisions. On the other hand, in treatments with investment limits, the following three conditions hold: (i) in each round, each player can either send a link or invest in information, but
cannot do both; (ii) each player can send at most one link; and (iii) at most one player can invest in information at any given time. We refer to the treatment without limits as the “baseline” in both the sequential and simultaneous environments, and denote treatments with investment limits as “limits.”

Notice that Seq_L and Seq_B differ in two ways: while investment is limited in Seq_L, it also implies the RFR, by which we mean that a person who currently invested in information has the right to continue his/her investment. The reason is that in Seq_L, a subject who has invested in information will continue to hold it until their next decision, and nobody else will be able to invest in additional information. Consequently, the only way they can lose the information is if they give up the information. It follows that comparing Seq_L to Seq_B measures the total effect of the investment limits combined with the RFR.

While these two effects cannot be separated in our sequential environment, it is possible to achieve separation in a simultaneous setting. To do this, we construct a fifth treatment that builds on Sim_L but eliminates the RFR. In any given round, agents who choose to invest in information will have an equal chance to obtain the information, regardless of whether he/she invested the information in the previous round. This treatment is denoted as simultaneous-limits with no RFR (Sim_L_NoRFR).

In summary, we investigate network formation in five treatments that differ in terms of the sequence of moves, whether investment is limited, and the existence or nonexistence of an RFR. We list the properties of these five treatments in Table 2.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Decision sequence</th>
<th>Investment limits?</th>
<th>RFR?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_B</td>
<td>Sequential</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Seq_L</td>
<td>Sequential</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Sim_B</td>
<td>Simultaneous</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Sim_L</td>
<td>Simultaneous</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Sim_L_NoRFR</td>
<td>Simultaneous</td>
<td>Y</td>
<td>N</td>
</tr>
</tbody>
</table>

### IV.2. Explanation of Treatment Design

We select our treatment conditions on the basis that each might bear relevance to the way star networks emerge in natural settings.
• Why sequential decisions?

It is very common to create networks through a process of sequential decisions. For example, online networking sites, such as Facebook or Twitter, regularly ask current users to invite their friends to join the site, and then those friends ask their friends. Similarly, given that it is infeasible to achieve face-to-face interaction with large numbers of people at the same time and location, social connections that exist “off-line” also generally occur through a sequential process. For example, when farmers consult others before adopting certain new seeds, it is likely that they observe what other farmers have done and turn to the ones that have experimented and gathered experience with the seeds (Foster and Rozensweig, 1995; Conley and Udry, 2010). The emergence of a “focal player,” in this case the center of the star network, is endogenously determined.

• Why investment limits?

Investment limits are a regular feature of natural environments. They can take two forms. First, repeated investment in information may not be feasible. One reason is that technologies are expensive to invent (Dimasi et al, 1991; Dimasi and Grabowski, 2007), often leaving it extremely costly for multiple groups to obtain the capital needed to invest in developing the same idea. Further, policies may exist to prevent inefficient investment in information. For example, to reduce inefficiency caused by excessive rent seeking, many governments choose to monopolize “insider” information through procurement auctions, which work to limit investment (Tran, 2009).

Second, sending multiple links may not be feasible. Non-monetary constraints in terms of time or social distance become important in this regard. This can become especially significant in the context of personal relationships, where time spent together is an important factor and often a binding constraint (Marsden and Campbell, 1984).

In natural environments, we often find that efficient results are achieved in contexts where resources are limited. It is plausible that efficient star networks might emerge naturally in the presence of investment limits as a response to limited resources. We pursue this possibility in our experiments by designing treatments with investment limits.
The specific limits we consider reduce the set of actions each player can choose, which consequently reduces the number of possible network outcomes in our experiment\textsuperscript{22}. Note that these limits are respected in any equilibrium, though they are not sufficient for equilibrium. Indeed, after imposing these restrictions, there remain 521 unique network outcomes, of which only four are equilibrium outcomes.

- **Why the “right of first refusal”?**

  While the RFR regularly appears in business contracts, it also shares a broader implication for any situation where long-term investment emerges naturally. An example is investment in personal education. Low-income families in developing countries may only be able to support formal education for one sibling (Song, Appleton and Knight, 2006). The RFR determines that the child who first begins education has the right to persist in his/her schooling. Other siblings would then expect to receive the return on the education by connecting with their educated sibling, i.e., the family forms a star network. Coordinating on educational outcomes in this way is efficient, and similar advantages may be at play when deciding which person or group is to receive advantaged access to information and knowledge, e.g., the choice of which employees will receive special training or which scholars will receive research grants. Our experiment design allows us to investigate the effect of RFR on improving coordination in networks in a controlled environment. Note that in sequential play environments, RFR cannot be separated from investment limits. An advantage of simultaneous play environments is that the impact of RFR can be distinguished from that of investment limits.

**IV.3 Hypotheses**

Our above discussion suggests, in general, that network environments that include sequential decisions, investment limits, and the “right of first refusal,” are expected to facilitate star network formation. The effects can be demonstrated in multiple ways. We discuss three possible measures that capture the effects of these institutional characteristics: equilibrium frequency, network stability, and individual rationality.

\textsuperscript{22} The conditions we impose reduce the number of players’ possible actions from 16 to 5, which then subsequently reduce the number of possible network outcomes from 65536 to 512.
**Frequency of Equilibrium**

We count a network graph as a star if and only if there is one member who chooses to invest in information and the other three agents send exactly one link to the sole investor. For each stage of the game, equilibrium frequency is found by dividing the total number of star networks by the total number of rounds in that stage game. Then, the mean frequency of star networks is found by averaging over the frequencies in the 24 stage games of each treatment.

**Hypothesis 1.** Star networks emerge more frequently in environments characterized by sequential decisions, investment limits, and the RFR. In light of the above discussion, we expect the frequency of realized star networks to follow the order below (where A>B denotes that the star networks are expected to occur more frequently in A than B):

1) Effect of sequential decisions:
   - Seq_B>Sim_B; Seq_L>Sim_L
2) Combined effect of investment limits and the RFR:
   - Sim_L>Sim_B; Seq_L>Seq_B
3) Effect of RFR alone (only in simultaneous environments):
   - Sim_L>Sim_L_NoRFR
4) Effect of the investment limits alone (only in simultaneous environments):
   - Sim_L_NoRFR>Sim_B

**Network Stability**

Network configuration may change over time. The stability of a network is key to many real world applications, e.g., R&D in business firms (Dodd et al, 2003). We use two measures to investigate the stability of networks among treatments: duration of equilibrium and duration of disequilibrium. For each stage, the duration of (dis-)equilibrium is calculated as the average number of consecutive (non-)star outcomes divided by the number of rounds in the stage. We then average this measure over all 24 stage games for each treatment.

**Hypothesis 2.** This hypothesis again follows from our argument that star networks are promoted by sequential decisions, investment limits, and the RFR. This implies that environments with these features should exhibit a longer mean duration of continuous star networks and, after falling out of equilibrium, a shorter mean duration of disequilibrium.
Specifically, we hypothesize the following regarding the mean duration of continuous star networks:

1) Effect of sequential decisions:
   Seq_B>Sim_B; Seq_L>Sim_L

2) Combined effect of investment limits and the RFR:
   Sim_L>Sim_B; Seq_L>Seq_B

3) Effect of the RFR alone (only in simultaneous environments):
   Sim_L>Sim_L_NoRFR

4) Effect of the investment limits alone (only in simultaneous environments):
   Sim_L_NoRFR>Sim_B

   Similarly, we form hypotheses regarding the mean duration of disequilibrium after falling out of equilibrium. Note the direction of the effect is reversed from the first set of hypotheses, since the less time it takes to restore a star network, the better an institution is at promoting stability.

5) Effect of sequential decisions:
   Seq_B<Sim_B; Seq_L<Sim_L

6) Combined effect of investment limits and the RFR:
   Sim_L<Sim_B; Seq_L<Seq_B

7) Effect of the RFR alone (only in simultaneous environments):
   Sim_L<Sim_L_NoRFR

8) Effect of the investment limits alone (only in simultaneous environments):
   Sim_L_NoRFR<Sim_B

- **Individual Rationality**

  Theory predict that for a rational player at any given time, if no player in the group invested in information, then the rational choice is to invest. While if at least one player in the group invested in information, then the rational choice is to send a link to the investor. Our individual rationality measure takes value 1 if a subject makes a choice consistent with rationality, and 0 otherwise.

  It is possible, in principle, to have a high mean individual rationality, while also having a low frequency of equilibrium (e.g., if three group members make systematically “good” decisions, but one member makes systematically “bad” decisions”). In general, however, we
would expect high individual rationality in network formation games to imply faster convergence to star networks, as well as more stable network outcomes. Just as above, we expect sequential decision-making, investment limits and the “right of first refusal” to each promote individual rationality.

**Hypothesis 3.** Individual rationality is greater in environments characterized by sequential decisions, investment limits and the RFR.

1) Effect of sequential decisions:
   Seq_B>Sim_B; Seq_L>Sim_L

2) Combined effect of investment limits and the RFR:
   Sim_L>Sim_B; Seq_L>Seq_B

3) Effect of the RFR alone (only in simultaneous environments):
   Sim_L>Sim_L_NoRFR

4) Effect of the investment limits alone (only in simultaneous environments):
   Sim_L_NoRFR>Sim_B

**IV.4. Experimental Procedure**

The experiment sessions were conducted between December 2010 and March 2011 in the ICES laboratory at George Mason University. Subjects were recruited via email from registered students at George Mason University. Each subject participated in only one session and none had previously participated in a similar experiment.

In total, 160 subjects participated in the computerized experiment programmed with z-Tree (Fischbacher 2007). Each experimental session lasted between 120 and 150 minutes. Subjects’ total earnings were determined by the Experimental Dollars (E$) earned at the end of the experiment, which were then converted at a rate of E$3 per US dollar. The average earnings were $25.28, ranging from a maximum of $53 to a minimum of $8 across all sessions.

In all treatments, before a session started, subjects were seated in separate cubicles to ensure anonymity. They were informed of the rules of conduct and provided with detailed instructions (see appendix B for the instruction of Sim_B as an example). The instructions were read aloud. In order to guard against confusion, after subjects finished reading the instructions, they were asked to complete a quiz. An experimenter checked their answers. Then the
experiment worked through the quiz questions on a white board in front of the laboratory. The experiment began after all subjects confirmed they had no further questions.

We ran 2 sessions for each treatment condition. Thus, in the end, we obtained 672 network graphs for each treatment (excluding the practice stage). Most of our analysis assumes 24 observations (eight groups each of which plays three stage-games with perfect strangers) for each treatment.

V. Results

Section IV.1 presents results from treatment comparisons at the aggregate level. Section IV.2 uses a cluster analysis approach to shed light on decisions at the individual level.23

V.1 Aggregate Analysis

We present results in the order of hypotheses listed in Section III.3. First, we discuss results concerning the frequency of star networks. Then we investigate two network stability measures: the duration of equilibrium and disequilibrium. Finally, we discuss how the proportion of individually rational moves varies across treatments and over time.

V.1.1 Frequency of Star Networks

The mean frequency of star networks in each of our treatments is shown in Figure 1. It is clear from this figure that star networks emerge at different rates, with baseline treatments displaying the lowest frequency of star networks. More formally, our findings are as follows:

---

23 All of the below excludes the non-incentivized practice stage. Our results remain qualitatively identical if practice stage data are included.
Result 1. (Test of hypothesis 1.1) Sequential decisions do not increase the frequency of star networks.

We found star networks to emerge with frequency 12.6% and 12.7% in Sim_B and Seq_B, respectively (p=0.667\(^{24}\)). On the other hand, when investment limits and the RFR are both present, 53.5% of networks formed in Seq_L are star shaped in comparison to 65.3% in Sim_L; this is also insignificant at standard levels (p=0.054).

Result 2. (Test of hypothesis 1.2) More star networks emerge when investment limits combined with the RFR are present.

Sim_L generated 65.3% of star networks, while only 12.6% of networks in Sim_B are star shaped. This difference is significant (p< 0.001). Agents in Seq_L form star networks 53.5% of time. When compared with the 12.6% in Seq_B, the difference is again significant (p< 0.001). Thus, our data provide clear evidence supporting the positive impact of investment limits and the RFR on star network formation.

\(^{24}\)Unless otherwise indicated, all p-value refers to two-tailed Mann-Whitney test.
Result 3. (Test of hypothesis 1.3) The RFR promotes star network formation in simultaneous decision environments.

The right most two bars in Figure 1 correspond to Sim_L and Sim_L_NoRFR. The only difference between these two treatments is that the RFR is present in the former but absent in the latter. In Sim_L star networks emerge at a rate of 65.3%\(^2\). The frequency in Sim_L_NoRFR treatment (32.4%) is significantly lower than this (p = 0.0016). This is evidence that the RFR promotes star networks in simultaneous environment.

Result 4. (Test of hypothesis 1.4) In simultaneous environments star networks emerge more frequently with than without investment limits.

Sim_L_NoRFR generates star networks at a rate of 32.4%, while the frequency in Sim_B is 12.6% (p=0.0034). Thus, investment limits promote star networks under simultaneous decision-making.

---

\(^2\) The frequency of star network we found here, 65.3%, is much higher than the best case in Goeree et al(2008) with agent heterogeneity condition.
Result 5. (Time Trend) More star networks are formed in later rounds of the experiment in Sim_L, Seq_L and Sim_L_NoRFR.

Figure 2 depicts the mean frequency of star networks over time. In all treatments the frequency of equilibrium is low in early rounds and then increases. Despite this common increase, when considering only the final five rounds, we can reject the hypothesis that mean frequencies of star networks are the same in all treatments (Kruskal-Wallis test, p=0.0006).

To analyze differences in the rate at which equilibrium changes over time and between treatments, we use an OLS regression with fixed effects at the group level. The results are described in Table 3. The key finding is that Seq_L displays a statistically significantly higher rate of increase in star networks over time than any other treatment (see Table 4). In contrast, Sim_B and Seq_B have the lowest rates among all treatments (Chow test, maximum p-value less than 0.0028, See table 4).

<table>
<thead>
<tr>
<th>Table 3. Linear Regression on the evolution of star network frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>round</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>stage</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>constant</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>fe(group)</td>
</tr>
<tr>
<td>R²</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4. Chow test for the same slope of regression (p-value)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Seq L</td>
</tr>
<tr>
<td>Sim_B</td>
</tr>
<tr>
<td>Sim_L</td>
</tr>
<tr>
<td>Sim_L_NoRFR</td>
</tr>
</tbody>
</table>

V.1.2 Network Centrality

Networks are said to have high centrality when a small number of people have a large number of connections. So, for example, network centrality is highest in our environment when one person receives three links, and no others receive any links, as occurs in a star network. It is
lowest, on the other hand, when everyone sends links to each other person. Here we take a similar approach to Georee et al (2008) and report the “degree centrality” per group. The following equation gives the formal definition of this measure:

\[
\text{degree-centrality} = \frac{1}{(n-1)(n-2)} \sum_{i,j \in N} \left[ \max_{j \in N} \text{degree}_j - \text{degree}_i \right]
\]  

(1)

where \( N = \{1, 2, \ldots n\} \) is the set of agents in the network, while \( i \) and \( j \) are two typical members of the set.

The above equation normalizes the centrality measure of a star network to be one. Intuitively, centrality measures closer to one suggest the network is more similar to a star. Our goal is to determine whether agents form networks that are more star-like over time and whether this tendency differs between treatments.

**Result 6. Centrality is higher in treatments with investment limits and the RFR.**

Figure 3 shows the mean degree centrality under each treatment condition. After Bonferroni correction for multiple comparisons, the difference between Sim_B and Sim_L, Seq_B and Seq_L remains statistically significant (p=0.000 and 0.078 respectively). Thus, under investment limits and the RFR, more star-like networks form in both sequential and simultaneous environments.

**Figure 3. Mean degree centrality measure across treatments**
IV.2 How stable are the star networks?

We consider two measures of stability: (1) once a star network formed, how many rounds does it last? (2) after falling out of equilibrium, how many rounds does it take to restore a star network? Figure 4 demonstrates the average length of continuous star networks (in percentage term) across treatments.

Result 7. (Test of hypothesis 2.1) Sequential decisions do not reduce the duration of star networks.

In Sim_B, once a star network forms, it will last on average for only 2.1 rounds (0.075% of the total number of rounds in a stage game). In Seq_B, however, star networks sustain themselves for an average of 2.24 rounds. The difference is not statistically significant (Mann_Whitney two tailed, p=0.6201). Similarly, when comparing results in Sim_L (0.298%) and Seq_L (0.445%), we cannot reject the null hypothesis that their equilibrium duration is identical (p=0.2742).

Figure 4 The length of continuous equilibrium networks (%) across treatments

Result 8. (Tests of hypotheses 2.2, 2.3 and 2.4) Star network duration is longer in treatments with investment limits or RFR.

Observed duration for continuous star networks parallels the results found regarding the frequency of equilibrium.
In simultaneous environments, when investment limits are introduced (Sim_L_NoRFR), the average length of continuous star network increases to 5.8 rounds, a significant increase from the result of 2.1 rounds in Sim_B (p=0.0098). Adding the RFR in Sim_L, the length of the equilibrium network increases to 12.46 rounds, a significant difference compared both to Sim_L_NoRFR and Sim_B (p=0.0032 and p<0.001 respectively).

In the sequential environment, the same institutions of investment limits and the RFR are introduced. In Seq_L, star networks last for about 8.34 rounds in a 28-round stage game. Agents in Seq_L do not sustain the same star as long as they did in Sim_L (p<0.001).

**Result 9. (Test of hypothesis 2.5) Sequential decisions do not shorten the duration of disequilibrium in the Baseline condition, but do under investment limits.**

Figure 5 shows that the duration of disequilibrium before reestablishing a star network does not differ between Sim_B and Seq_B (p=0.7142), while it does between Sim_L and Seq_L (p=0.0371).

**Figure 5. The length of Continuous Disequilibrium networks across treatments**

---

26 The longest-lasting star network occurs in Seq_L treatment, wherein a group of subjects kept playing a single star network for 43 rounds in their 60-round stage game without disruption.
Result 10. (Test of hypothesis 2.6, 2.7, 2.8) Disequilibrium durations are shorter under investment limits or the RFR.

Our data support hypotheses 2.6 and 2.7. Comparing Sim_L_NoRFR to Sim_L, the disequilibrium length falls from 10.1 rounds to 5.5 rounds, a significant decrease (p=0.014). In Sim_B and Sim_L it takes on average 17.2 and 5.5 rounds to return to equilibrium after a disruption. In Seq_B and Seq_L it takes 18.8 and 7 rounds on average to return to equilibrium play. Both comparisons are statistically significant (p<0.001).

V.3 Individual Rationality

We define a decision to be rational if it satisfies both of the following conditions: (1) if any other agent has invested in information, one should link to her; (2) if no other agent has invested in information, one should invest. The percentage of decisions that satisfy these conditions are plotted in Figure 6 for each treatment condition.

Figure 6. Mean percentage of rational choices made under each treatment condition

Result 11. (Test of hypothesis 3.1) Sequential decisions do not improve individual rationality.

We find no statistically significant differences in individual rationality between Sim_B and Seq_B (p=0.0512) or between Sim_L and Seq_L (p=0.2879).
Result 12. (Test of hypothesis 3.2) The combination of investment limits and the RFR promotes rational choice at individual level.

In simultaneous environments, about 40.8% of choices made are rational in the baseline condition, while in Sim_L, the number increases significantly to 79.3% (p<0.001). In sequential environments, the effect is smaller (an increase from 52.2% to 85.5%) but remains highly significant (p<0.001).

Result 13. (Test of hypothesis 3.3) The RFR alone increases individual rationality in the simultaneous environment.

The last two bars in Figure 6 show that the existence of the RFR improves individual rationality. About 51.9% of choices made in Sim_L_NoRFR are individually rational, a significant decrease in comparison to Sim_L (p=0.0023).

Result 14. (Test of hypothesis 3.4) Investment limits promote rational choice at the individual level in simultaneous environments.

The percentages of individual rational moves for Sim_L_NoRFR and Sim_B were 59.1% and 40.8% respectively (p=0.0010).

VI. Behavioral Rules

The purpose of this section is to draw inferences about the behavioral rules of individuals in our various treatments. Our maintained assumption is that behavioral rules in all treatments are formed using elements from a menu of information that are finite and identical, but that different treatments lead to rules that differ at the level of usage on the information. Without ex ante knowledge of what kind of rules may exist, we use cluster analysis to detect them.27 Compared to regressions, cluster analysis can better explore patterns within a complex environment where the classification structure may not be well defined. It allows us to explore behaviors among individuals without the need to pre-define the nature or number of possible rules (see also Houser et al, 2004).

27 Cluster analysis, as a numerical method for classification, serves the function of organizing a large and complicated data set into a small number of groups of objects so that the original data set could be easily understood based on patterns of similarity. It is widely used in fields such as astronomy, biology and marketing, and increasingly in economics (Fisher, 1963; Hirschberg et al, 1991; Houser, et al, 2004).
In this section, we first discuss our method for cluster analysis (the k-means algorithm) and then describe how we implemented cluster analysis with our data.

VI.1. The k-means clustering algorithm

With cluster analysis, one develops indices and criteria to know in a mathematically precise way how “close” or “far apart” objects are to each other. A variety of distance measures have been proposed. The one we use here is Euclidean distance:

\[ d_{ij} = \left( \sum_{k=1}^{p} (x_{ik} - x_{jk})^2 \right)^{1/2} \]  

where \( x_{ik} \) and \( x_{jk} \) are, respectively, the \( k^{th} \) variable value of a \( p \)-dimensional observations for individual \( i \) and \( j \).

An informative clustering includes groups such that the distance between objects in the same group is small, while the distance between groups is large. Based on this simple intuition, a variety of so-called “dissimilarity indices” (formed by combining distance measures across agents) have been suggested.

The particular index described below comprises the foundation for the k-means clustering algorithm, which we use to perform our analysis. Let \( d_{ql,kv} \) be the Euclidean distance between the \( l^{th} \) object in the \( q^{th} \) group and the \( v^{th} \) object in the \( k^{th} \) group. Then the k-means dissimilarity index takes the following form:

\[ h_l(m) = \frac{1}{2n_m} \sum_{i=1}^{n_m} \sum_{v=1}^{n_m} d_{ml, mv}^2 \]  

This index measures the within sum of square, i.e., the sum of squared distances between an object in cluster group \( m \) and the mean of all objects that belong to group \( m \)\(^{28}\).

Ideally, one would be able to go through all combinations of objects to determine the one that yields the lowest dissimilarity index within each group. However, when the number of objects is not small, it becomes extremely computationally burdensome to do so\(^{29}\). This has led

\(^{28}\) Indices that measure the separation between groups are also used in many other methods. We refer interested readers to Everitt et al (2011).

\(^{29}\) According to Liu (1968), the number of possible partitions one must consider in order to partition 100 network agents into five groups is \( 6.6 \times (10^{67}) \).
scholars to develop numerical algorithms to approximate clustering solutions; among these, k-means is used most widely (Steinley, 2006).

Our analysis is based on a canonical k-means algorithm (Hartigan and Wong, 1979). It involves iteratively updating partitions by relocating objects into the group whose mean is closest and then recalculating group means.

- **Determining the number of clusters**

A cluster analysis requires one to determine the number of clusters a dataset contains. A large variety of methods have been proposed for this purpose. In a Monte Carlo analysis, Milligan and Cooper (1985) report that among 30 methods they compared, the top performer was Calinski and Harabasz (1974) (which we denote by C-H). We use C-H to determine the number of clusters in our data.

C-H suggested that the number of clusters, $g^*$, should be chosen to maximize $C(g)$, where $C(g)$ is given by:

$$C(g) = \frac{\text{trace}(B)}{g-1} \left/ \frac{\text{trace}(W)}{n-g} \right.$$  \hspace{1cm} (4)

where $B$ is the usual between (clusters) sum of squared deviations from (overall) mean, and $W$ the within (clusters) sum of squared deviations from (cluster) mean.

**VI.2. Results from k-means analysis**

Our analysis proceeds in two steps. First, we estimate for each individual the parameters that characterize the way they make decisions given information. Then, we use cluster analysis to group similar individuals into behavioral rules. In particular, we run a linear regression for each individual with the decision to invest (or not) as a binary dependent variable, on a constant, a dummy for whether investing is rational and an index characterizing the subjects investing behavior in the previous two rounds (see also Kurzban and Houser, 2005). Then, we use the k-means algorithm to cluster these estimates into groups of behavioral rules. We repeat the above analysis for the linking decision.

---

30 Another successful technique developed by Duda and Hart (1973) works with hierarchical cluster methods. The network data do not fit these types of cluster analysis.

31 We use the sum of past two choices as a proxy for past choice in this analysis since the immediate past choices are highly likely to be collinear with our dependent variables in all treatments with investment limits.
VI.2.1 Behavioral rules

The independent variables we include in our regressions are meant to capture a person’s: (i) base rate willingness to invest or link to others (captured by the regression’s constant); (ii) consistency with individual rationality (captured by the dummy variable that takes value one if it is optimal to invest (or link)); and (iii) propensity to form a “habit” of choice in the sense that they do what they did before (captured by the variable indicating the lagged decisions for the past 2 rounds). Equations 13 and 14 specify our regression equations for investing and linking respectively:

\[
\text{invest}_{i,t} = \beta_1 \cdot \text{rational}_{i,t}^p + \beta_2 \cdot \sum_{s=1}^{2} \text{invest}_{i,t-s} + \beta_3 + \epsilon_{i,t} \quad (5)
\]

\[
\text{linksending}_{i,t} = \gamma_1 \cdot \text{rational}_{i,t}^l + \gamma_2 \cdot \sum_{s=1}^{2} \text{linksending}_{i,t-s} + \gamma_3 + \delta_{i,t} \quad (6)
\]

where

\[
\text{rational}_{i,t}^p = \begin{cases} 
1, & \text{if subject } i \text{ should have purchased information at round } t \\
0, & \text{otherwise}
\end{cases} \quad \text{according to individual rationality}
\]

\[
\text{rational}_{i,t}^l = \begin{cases} 
1, & \text{if subject } i \text{ should have sent link at round } t \\
0, & \text{otherwise}
\end{cases} \quad \text{according to individual rationality}
\]

\[
\text{invest}_{i,t-s} = \begin{cases} 
1, & \text{if subject } i \text{ invested in information in round } t-s \\
0, & \text{otherwise}
\end{cases}
\]

\[
\text{linksending}_{i,t-s} = \text{number of links subject } i \text{ sent in round } t-s
\]

The above regressions are repeated for each individual. We end up with 142 and 152 subjects in our sample for the investing and linking regressions, respectively. Each individual’s estimates can be represented by a point in 3-space (See Appendix C, panels a and b).

1.2.2 K-means clustering

We implement our k-means cluster analysis, as well as cluster number selection, using R. Based on the C-H index, we find three clusters in both investing and linking decisions (See Appendix D, panel a and b, for the 3-space plot).

---

32 We drop 18 subjects for regressions on investing decisions, as there is zero variation in dependent variables. For the same reason, we drop 8 subjects for regression on linking decisions.
a) Investing decisions

The three panels of Figure 9 are the three 2-space projections of the estimates \( \{\beta_1, \beta_2, \beta_3\} \) from Figure 8, Panel a. Each point represents an individual’s estimates from his/her investing decisions regression. Points with the same marker belong to the same cluster.

Figure 9. Projections of Estimates from Investing Decision

33 We have found substantial differences in K-means clustering results produced by the standard packages in Stata, R and Matlab. We traced it to differences in the specific numerical algorithms used by each package (Steinley, 2008) and decided to adopt clustering package from R, as it is most robust to the choice of initial points. In particular, we use kmeansruns from package cluster and initialize the same clustering analysis with 1000 different starting points. 34 Decision rules for investing and linking differs in terms of interpretation and the range of measurement. Hence, we discuss them separately.
It is clear from visual inspection that our clusters are well-separated. To provide statistical evidence on the strength of this separation, we analyze the separation along each independent variable’s axis. Mann-Whitney tests find significant differences between all pairs of clusters in each axis (p<0.001), with the exception of the constants in the triangle and round clusters (the lack of separation in this single case can be clearly seen from panels (a) or (b)).

Not only are the clusters clearly separated, the location of the clusters also carries meaningful interpretation in our sample. Table 5 provides the mean estimate for each
independent variable and for each cluster, and also reports whether that mean is significantly different from zero.

**Table 5. The Mean of Estimates from Regression on Investing Decision**

<table>
<thead>
<tr>
<th></th>
<th>Square cluster</th>
<th>Triangle cluster</th>
<th>Round cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rational to invest</td>
<td>0.8190 (0.0000)</td>
<td>0.3411 (0.0000)</td>
<td>-0.0978 (0.0054)</td>
</tr>
<tr>
<td>Lagged choice</td>
<td>-0.0408 (0.1480)</td>
<td>0.1745 (0.0000)</td>
<td>0.0782 (0.0120)</td>
</tr>
<tr>
<td>Base rate(constant)</td>
<td>0.0175 (0.2589)</td>
<td>0.0137 (0.7066)</td>
<td>0.4279 (0.0000)</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>57</td>
<td>46</td>
<td>39</td>
</tr>
</tbody>
</table>

Note: p-value from Wilcoxon signed-rank test in parentheses

Based on the results from Table 5, we summarize the characteristics of the three clusters that define the three behavioral rules used by our subjects.

(1) We define the cluster indicated with circular markers as the “Rational” type. People that belong to this cluster are guided by the rationality of the current opportunity to invest. They focus relatively less on their past choices, and their base rate of investing is near zero. Consequently, if it is neither rational to invest nor had they invested before, the subjects in this group would be relatively unlikely to invest.

(2) We define the cluster indicated by triangle markers as the “Habit” type. Subjects in this cluster are guided by rationality, but relatively less so than the Rational type. Instead, their current decisions follow closely their past decisions. As with the “Rational” type, the base rate of investing of the Habit type is near zero.

(3) We define the cluster indicated by square markers as the “Dogmatic” type. When making their investing decisions, subjects of this type place relatively little weight on whether it is rational to invest, and also are less guided by whether they have invested before. We find that the Dogmatic subjects have the highest base rate of investing among all three types.

We now investigate how the institutional characteristics in our various treatments affect the type of behavioral rules subjects use. Table 6 reports the frequency of types by treatment.
Table 6. Number of Individuals in each Treatment and each Type

According to Investing Decisions

<table>
<thead>
<tr>
<th></th>
<th>Seq_B 12.6% of star</th>
<th>Seq_L 53.5% of star</th>
<th>Sim_B 12.7% of star</th>
<th>Sim_L 65.3% of star</th>
<th>Sim_L_NoRFR 32.4% of star</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Rational” Round</td>
<td>9 (31.03)</td>
<td>23 (92.00)</td>
<td>0 (0.00)</td>
<td>3 (11.11)</td>
<td>22 (70.97)</td>
</tr>
<tr>
<td>“Habit” Triangle</td>
<td>8 (27.59)</td>
<td>2 (8.00)</td>
<td>3 (10.00)</td>
<td>24 (88.89)</td>
<td>9 (29.03)</td>
</tr>
<tr>
<td>“Dogmatic” Square</td>
<td>12 (41.38)</td>
<td>0 (0.00)</td>
<td>27 (90.00)</td>
<td>0 (0.00)</td>
<td>0 (0.00)</td>
</tr>
<tr>
<td>Total</td>
<td>29 (100.00)</td>
<td>25 (100.00)</td>
<td>30 (100.00)</td>
<td>27 (100.00)</td>
<td>31 (100.00)</td>
</tr>
</tbody>
</table>

Note: percentage in parenthesis

As noted above, star networks emerge in fewer than 13% of our two baseline treatments (Seq_B and Sim_B). This low level of star network formation coincides with a concentration of Dogmatic type subjects (41.38% and 90% respectively). That is to say, having a concentration of players using the Dogmatic investing rule is not conducive to star network formation.

On the contrary, for the Seq_L treatment, which generates a relatively high percentage of star networks, the large majority of subjects (92%) choose to behave rationally. The other highly effective treatment, Sim_L, generates 65.3% of star networks. Its success at generating star network coincides with a high level of Habit typed subjects (88.89%), a few Rational subjects (11.11%) and no Dogmatic subjects.

The Sim_L_NoRFR treatment generates a medium level of star networks (32.4%). No subject in this treatment belongs to the Dogmatic type. In particular, most of them (90.97%) follow rational behavioral rules.

b) Linking decisions

Similar to the above analysis, the three panels of Figure 10 project each 3-vector estimate \(\{\gamma_1, \gamma_2, \gamma_3\}\) from regression on linking decisions (Equation 14) into corresponding 2-space.

Again, we find clear visual separation between our clusters from estimates on linking behavior, and Mann-Whitney tests support significant differences between all pairs of clusters on all three axes (p<0.001), again with the exception of the estimates of the constants between the
round and triangle clusters (p=0.5279) (this is apparent from visual inspection of Panels a and b in Figure 10).

Interestingly, the location of the clusters closely resembles those found for investing decisions. Consequently, we assign the same labels, Rational, Habit and Dogmatic, for each of these clusters as well. Table 7 reports the mean of each estimate for each cluster and the Wilcoxon signed-rank p-value for the test of whether the cluster’s mean is significantly different from zero.

Figure 10. Projections of Estimates from Linking Decision

(a)

(b)
Table 7. The Mean of Estimates from Regression Analysis of Linking Behavior

<table>
<thead>
<tr>
<th></th>
<th>Round cluster</th>
<th>Triangle cluster</th>
<th>Square cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rationality</td>
<td>0.7746 (0.0000)</td>
<td>0.2693 (0.0000)</td>
<td>-0.0461 (0.2112)</td>
</tr>
<tr>
<td>Lagged choice</td>
<td>-0.0187 (0.4091)</td>
<td>0.2179 (0.0000)</td>
<td>0.0745 (0.0041)</td>
</tr>
<tr>
<td>Base rate(constant)</td>
<td>0.1735 (0.0000)</td>
<td>0.1331 (0.0000)</td>
<td>0.6980 (0.0000)</td>
</tr>
<tr>
<td>Number of subjects</td>
<td>37</td>
<td>75</td>
<td>40</td>
</tr>
</tbody>
</table>

Note: p-value from Mann-Whitney test in parenthesis

Based on the characteristics of the three clusters described in Table 7, we define three behavioral rule types as follows:

(1) We define the round cluster to be a “Rational” type. People who belong to this cluster make decisions that are guided largely by the rationality of their current choice. They have a significantly positive (but small) base rate of linking, but they do not take account of their past choices when making their decisions.

(2) We define the triangle cluster as a “Habit” type. People in this cluster make choices that resemble their previous choices. It turns out that rationality is also statistically significant, but the coefficient is smaller than that of the Rational type.
(3) We define the square cluster as a “Dogmatic” type. Subjects in this group sent links to others at a high base rate; rationality is not a statistically significant predictor of choices for this type and while previous choices have a statistically significant coefficient that is small in magnitude. The base rate of linking is 69.8% (statistically significantly higher than either of the other types \((p<0.001)\)).

To see how institutions interact with types, we report types by treatment in Table 8.

### Table 8. Number of Individuals in Different Treatments and Types According to Linking Decisions

<table>
<thead>
<tr>
<th></th>
<th>Seq_B</th>
<th>Seq_L</th>
<th>Sim_B</th>
<th>Sim_L</th>
<th>Sim_L NoRFR</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Rational” round</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Habit” Triangle</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Dogmatic” square</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>13 (40.63)</td>
<td>20 (71.43)</td>
<td>0 (0.00)</td>
<td>2 (6.67)</td>
<td>2 (6.25)</td>
</tr>
<tr>
<td></td>
<td>6 (18.75)</td>
<td>4 (14.29)</td>
<td>12 (40.00)</td>
<td>24 (80.00)</td>
<td>29 (90.63)</td>
</tr>
<tr>
<td></td>
<td>13 (40.63)</td>
<td>4 (14.29)</td>
<td>18 (60.00)</td>
<td>4 (13.33)</td>
<td>1 (3.13)</td>
</tr>
<tr>
<td></td>
<td>32 (100.00)</td>
<td>28 (100.00)</td>
<td>30 (100.00)</td>
<td>30 (100.00)</td>
<td>32 (100.00)</td>
</tr>
</tbody>
</table>

Note: percentage in parenthesis

We were surprised that the clusters found in the linking analysis resemble so closely the clusters found in our analysis of investing behaviors. In both cases, the two baseline treatments with the lowest frequency of star networks also have the highest percentage of subjects belonging to Dogmatic type (40.63% and 60% for Seq_B and Sim_B respectively).

Result from Seq_L shows that the majority of subjects are the Rational type. And while Sim_L has the most frequent star network formation, it also has a high percentage of Habit type subjects.

To investigate the relationship between the behavioral rules players use when making linking or investing decisions, Figure 11 plots the percentage of players belonging to each type in linking decisions conditional on each type in investing decisions\(^{35}\). 78% of subjects that are the

---

\(^{35}\) Note that by construction, players in the investment-limits treatments who make rational investing decisions also necessarily make rational linking decisions. This is true only for this treatment and the players who behave
Habit type in investing decisions are also Habit type in linking decisions. Similarly, there is substantial overlap among participants classified as Rational and Dogmatic between linking and investing decisions. Indeed, a Pearson Chi-square test rejects that type classifications are independent between investing and linking decisions (p<0.001).

**Figure 11. Frequency of Each Type in Investing Decision Conditional on Types in Linking Decision**

**VI.3. Discussion**

Our results strongly suggest that efficient star network formation could form but only under the “right” institutions. By clustering individuals into decision rules, we further understand why certain institution works better than others.

In a complex network formation environment such as ours, it can be difficult to determine ex ante the types of rules participants are likely to follow. Our cluster analysis approach allows one to draw inferences about number and nature of rules at use in the population without the need to make strong ex ante assumptions.

We analyzed investing and linking decisions separately. Both analyses yielded clearly separated and easily interpretable clusters. In both cases, subjects fall into one of the following rationally. Our design does not otherwise imply any correlation between the behavioral rules followed by players when making linking or investing decisions.
decision types: Rational, which makes choice according to individual rationality; Habit, which is more heavily influenced by their previous choices; and Dogmatic, which displays a high base rate of investing or linking regardless of rationality or previous choice. We found it interesting that when interacting behavioral rules with treatments, we again obtained similar results: institutions more effective at promoting star networks seem to encourage coordination through either rationality or dependence on previous choices. Finally, the fact that individual classifications are correlated between environments adds additional comfort that the rules we have discovered are in fact a valid behavioral characterization.

These results suggest that institutions impact network formation by changing the frequency of different behavioral rules people follow. Institutions work best at promoting efficient network when they promote active decision-making: we found no cases where networks reliably emerged when there was a high frequency of dogmatic types. Further, the behavioral rules that are consistent with efficient network formation include two kinds: rationality rule and habit forming rules. We conjecture that successful institutions seem to promote “thinking” in the sense that they provide people the relevant information needed to achieve better coordination.

VII. Conclusion
Star networks emerge naturally in many social environments, and theory indicates star network equilibria are efficient. Based on a theory suggested by Galeotti and Goyal (2007, 2010), we study star network formation using laboratory experiments. Previous experimental studies of network formation suggest that persistent star networks could emerge in the lab, but only under ex ante agent heterogeneity (Goeree et al, 2008). This contrasts with natural environments, where star networks frequently emerge even when agents are ex ante homogeneous (Feick and Price, 1987; Conley and Udry, 2010). We conjectured that certain naturally-occurring institutional characteristics, such as sequential decisions, investment limits, and the “right of first refusal,” may help to explain why star networks form in these natural (un-designed) environments.

Our main finding is that investment limits and the “right of first refusal” promote star network formation. In comparison to baseline treatments, we find that environments with those features realize increased star-network frequency, improved network stability and higher levels of individually rational choices. Using a cluster analysis to recover similarities in individual decision rules, we further find that different institutions encourage different behavioral rules, in
the sense that some institutions promote rationality and “good” habits better than others. In particular, sequential environments with investment limits and the RFR tended to promote rationality, and simultaneous environments with investment limits and the RFR tended to promote habits that were conducive to the formation of stable star networks.

To our knowledge, our paper reports the first experimental investigation of the impact of institutions on star network formation. Results from this study suggest how one might design mechanisms to promote the efficient flow of information through social networks relevant to political campaigns, franchise training or agricultural innovation\(^36\).

Finally, this paper offers a direct test of network formation theory (Galeotti and Goyal, 2010). Our findings suggest that the institutional environment can substantially promote the formation of star networks by altering the behavioral rules used among agents. It would be profitable for future theoretical research to analyze the specific role of investment limits and the right of first refusal within an environment characterized by agents with multiple behavioral types.

---

\(^{36}\) For example, our study suggests that when an important agricultural innovation occurs, designing a network formation process that incorporates investment limits or the RFR may lead to more knowledge transmission, therefore higher adoption rate, with lower cost.
References


Jackson, M. an D. Lopez-Pintado (2011): Diffusion in Networks with Heterogeneous Agents and Homophily', working paper


Köhn HF, D. Steinley and M. Brusco (2010): The p-median model as a tool for clustering


Song, L., S. Appleton and J. Knight (2006): Why do girls in rural China have lower school enrollment? World Development. 34(9), 1639-1653


Appendix A. Z-tree experimental interface

Fig. A An example of the decision screen
Appendix A, continue

Fig B. An example of the display screen

<table>
<thead>
<tr>
<th></th>
<th>Player J</th>
<th>Player K</th>
<th>Player M</th>
<th>Yourself</th>
</tr>
</thead>
<tbody>
<tr>
<td>Links sent to...</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-</td>
</tr>
<tr>
<td>connections formed</td>
<td>Y</td>
<td>N</td>
<td>N</td>
<td>-</td>
</tr>
<tr>
<td>with...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of items</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>obtained</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Earnings (E$) for this round</td>
<td>3.1</td>
<td>3.5</td>
<td>1.0</td>
<td>3.5</td>
</tr>
<tr>
<td>Accumulated earnings (E$) for this stage</td>
<td>9.3</td>
<td>8.0</td>
<td>3.0</td>
<td>5.5</td>
</tr>
</tbody>
</table>
Appendix B. Multiple Comparison Correction:

We find most of the results in section IV.1 hold when we correct the p-value for multiple comparisons. Table A-D report pair-wise comparisons between treatments with a Bonferroni correction for equilibrium frequency, equilibrium duration, disequilibrium duration and individual rationality. Table entries show row mean-column mean (p-value in parenthesis).

Table A. Differences in the mean frequency of star networks by treatments

<table>
<thead>
<tr>
<th></th>
<th>Seq_B</th>
<th>Seq_L</th>
<th>Sim_B</th>
<th>Sim_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_L</td>
<td>0.4078 (0.000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_B</td>
<td>-0.0011 (1.000)</td>
<td>-0.4089 (0.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_L</td>
<td>0.5260 (0.000)</td>
<td>0.1182 (1.000)</td>
<td>0.5271 (0.000)</td>
<td></td>
</tr>
<tr>
<td>Sim_L_noRFR</td>
<td>0.1963 (0.080)</td>
<td>-0.2115 (0.044)</td>
<td>0.1974 (0.076)</td>
<td>-0.3297 (0.000)</td>
</tr>
</tbody>
</table>

Table B. Difference of the mean equilibrium duration by treatments

<table>
<thead>
<tr>
<th></th>
<th>Seq_B</th>
<th>Seq_L</th>
<th>Sim_B</th>
<th>Sim_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_L</td>
<td>0.223 (0.022)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_B</td>
<td>0.052 (1.000)</td>
<td>-0.217 (0.028)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_L</td>
<td>0.369 (0.000)</td>
<td>0.147 (0.409)</td>
<td>0.364 (0.000)</td>
<td></td>
</tr>
<tr>
<td>Sim_L_NoRFR</td>
<td>0.132 (0.664)</td>
<td>-0.091 (1.000)</td>
<td>0.126 (0.778)</td>
<td>-0.238 (0.011)</td>
</tr>
</tbody>
</table>

Table C. Difference of the mean disequilibrium duration by treatments

<table>
<thead>
<tr>
<th></th>
<th>Seq_B</th>
<th>Seq_L</th>
<th>Sim_B</th>
<th>Sim_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_L</td>
<td>-0.361 (0.001)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_B</td>
<td>0.058 (1.000)</td>
<td>0.419 (0.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_L</td>
<td>-0.416 (0.000)</td>
<td>-0.055 (1.000)</td>
<td>-0.474 (0.000)</td>
<td></td>
</tr>
<tr>
<td>Sim_L_NoRFR</td>
<td>-0.253 (0.048)</td>
<td>0.108 (1.000)</td>
<td>-0.310 (0.006)</td>
<td>0.164 (0.650)</td>
</tr>
</tbody>
</table>

Table D. Difference of the mean disequilibrium duration by treatments

<table>
<thead>
<tr>
<th></th>
<th>Seq_B</th>
<th>Seq_L</th>
<th>Sim_B</th>
<th>Sim_L</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq_L</td>
<td>0.333 (0.000)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_B</td>
<td>-0.114 (0.314)</td>
<td>-0.447 (0.000)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sim_L</td>
<td>0.270 (0.000)</td>
<td>-0.062 (1.000)</td>
<td>0.385 (0.000)</td>
<td></td>
</tr>
<tr>
<td>Sim_L_NoRFR</td>
<td>0.069 (1.000)</td>
<td>-0.264 (0.000)</td>
<td>0.183 (0.007)</td>
<td>-0.202 (0.002)</td>
</tr>
</tbody>
</table>
Appendix C. 3-space Plot for Individuals’ Estimates by Treatment

(a) Investing decisions

(b) Linking decisions

Note: different markers represent different treatments

- Seq_B ; ▲ Seq_L; - Sim_B; ♦ Sim_L; ○ Sim_L_NoRFR
Appendix D. 3-space Plot for Individuals’ Estimates by Cluster

(a) Investing decisions

(b) Linking decisions