COME PLAY WITH ME: EXPERIMENTAL EVIDENCE OF INFORMATION DIFFUSION ABOUT RIVAL GOODS

ABHIJIT BANERJEE[§], EMILY BREZA[†], ARUN G. CHANDRASEKHAR[‡], ESTHER DUFLO[♭], AND MATTHEW O. JACKSON[♮]

Abstract. We randomly invite households to come to a pre-specified, central location in 39 villages to participate in laboratory games. Because many households that were not directly invited turned up at our experiments, we study how the information about the opportunity to earn close to one day's wage diffuses through rural Indian villages. Furthermore, because some members of some of the villages had prior experience playing similar laboratory games, we ask how experience with a task affects information-spreading and -seeking behavior. Finally, we examine possible channels for strategic information diffusion. In our environment, participant slots for non-invited households are limited, making them rival goods. Additionally, participants could potentially receive larger payoffs from playing the laboratory games with their peers. Because of these two motivations, we examine how final participation patterns may reflect strategic behavior on the part of informed households.

JEL Classification Codes: C92, L14

Keywords: Social networks, Information Diffusion

1. Introduction

In developing countries, the village is an important unit for governance, co-investment, and resource distribution. Many of these types of goods may be rival: for example, the time of a health care or extension-service worker might be scarce; similary, the amount of grain or rice for distribution is generally fixed. In these

Date: August 2012 – Preliminary and Incomplete. Do not circulate.

Gowri Nagraj, Manaswini Rao and Dhruv Parthasarathy provided excellent research assistance. Special thanks to Horacio Larreguy, Cynthia Kinnan and Juan Pablo Xandri for sharing data from their previous work. We are grateful to the Russell Sage Foundation, the George and Obie Shultz Fund, and the MIT Center for International Studies for financial support. Breza and Chandrasekhar thank the National Science Foundation GRFP.

[§]Massachusetts Institute of Technology. Email: banerjee@mit.edu.

[†]Columbia Business School. Email: eb2840@columbia.edu.

[‡]Microsoft Research New England. Email: arc@microsoft.com.

[♭]Massachusetts Institute of Technology. Email: eduflo@mit.edu.

[♮]Stanford University, the Santa Fe Institute, and CIFAR. Email: jacksonm@stanford.edu.

cases, individuals again may strategically alert their friends to various opportunities, potentially not wanting too many individuals to learn about them. Other types of locally-distributed goods may be such that payoffs to households are complementary with the payoffs and actions of other participating households in the village. For example, an individual in a dairy cooperative earns more if the other producers in the village produce high quality milk; an individual might gain more utility from serving on a committee that has more like-minded members. As a result, when new opportunities arise in a village, individuals may have incentives to strategically inform certain friends and not others. In this paper, we seek to understand how information about the opportunity to obtain a rival good that may also have payoff complementarities diffuses through rural villages in India.

The development economics literature has made some headway toward understanding how non-rival information flows through villages. Foster and Rosenzweig (1995) find that indivduals earn higher profits from high yield variety seeds when their neighbors' experience with the seeds increases. Conley and Udry (2010) show that farmers of cash crops learn from their information neighbors about a new agricultural technology in Ghana and especially learn from those neighbors who have been recently successful. Similarly, Bandiera and Rasul (2006) study the adoption of a new cash crop in Mozambique and also find strong evidence of social learning. In the case of health, Kremer and Miguel (2007) show that individuals actually take up deworming medication less if direct friends or indirect second-order contacts randomly increase their own usage.

Researchers have also studied how information may diffuse through peer networks during financial crises. Iyer and Puri (2008) show that individuals are more likely to run on their bank if their neighbors are also running, while Kelly and Grada (2000) construct social networks for Irish immigrants in New York based on the Irish county of origin and model how financial panic is communicated across the network. They conclude that the social network is very important for explaining the spread of the panic.[1] In most of the existing literature, the social network is defined in a rather coarse fashion such as by grouping individuals based on an observable characteristic such as neighborhood. While the literature has been successful at identifying relationships between an individual's actions and the information set of

[1]In a relevant example from the developed world, Cohen et al. (2008) show that mutual fund managers overweight companies that have executives who were in their educational networks. This strategy substantially outperforms other comparable portfolio allocation. The authors conclude that the effect is due to superior information transmission along the social networks.

the peer group, we still do not have much concrete evidence regarding how more nuanced characteristics of the social network may impact the spread of information.

This paper attempts to contribute to the growing literature on using well-measured social networks to better understand diffusion processes. In their study of microfinance take-up, Banerjee et al. (2011) estimate a model of diffusion and conclude that individuals do indeed pass information about microfinance to other members of the network and that individuals who themselves took it up pass along the information with a higher probability. They also do not find any evidence that conditional on being informed, the actions of an individual's peer group do not impact the final take-up decision. Similarly, Banerjee et al. (2012) use an experiment to map out diffusion patterns. In this study, we use the same detailed networks data to learn about the case of information diffusion about the availability of scarce slots in a laboratory experiment.

In our study, we randomly go door-to-door inviting households to a set of laboratory games that are to be held two days later. We then measure which households learned about the opportunity and decided to attend the experimental sessions. We ask four questions: First, how does the network matter above and beyond simple peer group designations? Second, because some individuals had prior experience playing similar games, we explore how exposure to a product or experience impacts future take-up and or the propensity to spread information. Third, we explore the information diffusion process in the case of rival goods that may have payoff complementarities. Due to the nature of the games, we capped participation at 24 individuals per village. While some people were turned away from participating in the lab games, the firm cap implies that slots were rival.[2] Furthermore, during the laboratory games, individuals were more likely to earn higher payoffs if they played those games with their friends.[3] Finally we ask which village-level network characteristics[4] are correlated with high rates of information diffusion.

We find that overall, that the network does matter for diffusion in non-trivial ways. We show that the random invitations increase an individual's propensity to eventually play the lab experiment and that the random invitations of friends have a sizeable impact on an individual's take-up. Even informing an individual at

[2]Households were aware ex ante that there would be caps of the number of slots. We chose to randomly invite households because it is viewed as fair way of allocating slots by the village. The first-come-first-served rule for non-invited households was well-understood.
[3]See Breza et al. (2011) for a description of the results.
[4]especially those predicted by economic theory to matter

social distance 4 from another household causes a detectable increase in that other household's participation rate. We also show that there are significant experience effects for the household. Families that have previous experience are much more likely to attend the games. Also, individuals exhibit an increased likelihood if their friends have more experience with the games. We also present suggestive evidence of strategic behavior. In the structural models, we find that informed households with past experience are less likely to pass information to others than inexperienced households. We also show that individuals who have large fractions of friends who are in turn friends with invited people's friends are more likely to find out about the experiments. Our paper contributes to the literature by providing insight into the motivations behind the transmission of information. The rivalry of the good we offered makes it an especially interesting setting. Further, we are fortunate to be able to work with both high-quality networks data and random variation in which households have information.

The remainder of the paper is organized as follows. In section 2, we describe the experimental subjects, network and survey data sources, and the experimental design. In section 3, we present the specifications used in the reduced form analysis, while section 4 displays the results. In section 5, we propose, estimate and compare several structural models of information diffusion, while section 6 concludes.

2. Data and Experimental Design

2.1. **Setting.** We present experimental evidence on the diffusion of information about a village meeting and the opportunity to play a laboratory game for 39 villages located in Karnataka, India which range from a 1.5 to 3 hour's drive from Bangalore. We chose these villages as we had access to village census demographics as well as unique social network data set, previously collected in part by the authors. The data is described in detail in Banerjee et al. (2011) and Jackson et al. (2010).

The graph represents social connections between individuals in a village with twelve dimensions of possible links, including relatives, friends, creditors, debtors, advisors, and religious company. We work with an undirected and unweighted network, taking the union across these dimensions, following Banerjee et al. (2011) and Chandrasekhar et al. (2011a). As such, we have extremely detailed data on social linkages, not only between our experimental participants but also about the embedding of the individuals in the social fabric at large. We use the following notation: we have a collection of R villages, indexed by r and N_r individuals per village. Every village is associated with a social network $G_r = (V_r, E_r)$ consisting of a set V_r of vertices (households) and E_r of edges (denoting whether two vertices are linked or not in the village). To every network we associate an adjaceny matrix A_r which represents whether or not two vertices are linked. That is, $A_{r,ij} = 1$ if and only if $ij \in E$.

Moreover, the survey data includes information about caste, elite status and the GPS coordinates of respondent homes. In the local cultural context, a local leader or elite is someone who is a *gram panchayat* member, self-help group official, *anganwadi* teacher, doctor, school headmaster, or the owner of the main village shop.

2.2. **Experiment.** The experimental design was implemented in conjunction with the framed field experiments analyzed in Breza et al. (2011). Participation in the lab games of Breza et al. (2011) entailed attending one three-hour, and participants were compensated approximately Rs. 140 on average, or close to one day's wage for a low-skilled worker. Because the laboratory games required only 24 participants per village, and because the mean village size was significantly larger at ~192 households per village, we decided to recruit participants through random invitations.

Two days before each laboratory experiment in each study village, we randomly informed 18 households of the time and the place of the laboratory experiment. Invitees were told that they would have the opportunity to participate in laboratory games and earn, on average, more than Rs. 100 for approximately one morning of their time. They were informed that the invited individuals would receive a guaranteed slot in the experimental session, if they turned up at the pre-specified location. The surveyors made no reference to either the possibility of inviting others to the game or to what would happen if non-invited individuals reported to the experiment.

On the day of each village's experiment, our surveyors arrived at the pre-specified place and time and first logged in the names and other characteristics of all of the individuals who were waiting to participate. For the remainder of the paper, we designate y_i^s as an indicator for whether any member of household i showed up for the experiment at the pre-specified place and time. Approximately 12 individuals per village turned up before the experiments started, and the majority of these households were not directly invited to participate. While many individuals showed up for the experiment, the invitations did not generate sufficient attendance to satisfy the demands of the Breza et al. (2011) games. In the case that fewer than 24 individuals reported for the games, the surveyors went around house-to-house trying to recruit participants. In some cases, this secondary recruitment effort did encourage an over-supply of participants. Invited individuals were given first priority, and the remaining slots were filled on a first-come, first-served basis. Those individuals who ultimately participated in the games are captured in the indicator variable, y_i^p.

The laboratory games that were played among the 24 participants in each village were modified trust games. Thus, it is possible that participants could have received higher payoffs by recruiting their friends to also report to the games. The surveyors did not inform invited households that this would be the case.

Finally, it is important to note that in 32 of the 39 villages, our recruitment and laboratory experiments occurred in locations that had hosted laboratory games at some point in the previous two years. These games included the experiments described in Chandrasekhar et al. (2011a), Chandrasekhar et al. (2011b), and Chandrasekhar et al. (2012). A key feature of these experiments is that participants benefited from being able to share risk with fellow participants. Therefore, these experiments may have created beliefs that they might also benefit from collaboration with friends in the Breza et al. (2011) session.

2.3. **Descriptive Statistics.** Descriptive statistics at the individual level are presented in Table 1. We played games in villages with a total of 7502 households. 15% of the households in the sample had experience playing laboratory games over the prior two years. On average, 9.2% of households were invited to come play the new laboratory experiment. It should also be noted that 10.9% of households did eventually play the new games. The table also includes various network statistics at the individual level including measures of centrality and social distance.

We also show descriptive statistics aggregated at the village level in Table 2, because we are interested in cross-village comparisons and graph-level characteristics. In the 39 study villages, the average size is approximately 192 households. In addition to averaged individual statistics,[5] we show graph level statistics such as the first eigenvalue of the adjacency matrix, which has a mean value of 14.6 and a standard deviation of 2.9 across villages. We also display the second eigenvalue of the stochastic matrix, which has mean 0.81 and variance 0.07 across villages. (Please see Appendix A for a description of these statistics.)

[5]While the means are simply weighted versions of the means in the individual-level table, it is useful to be able to look at the cross-village variance in the average metrics.

3. Reduced Form Estimation Framework

We aim to characterize how information about the village experimental sessions spreads from the information plants. We first analyze information transmission using a reduced form, regression framework, exploiting the fact that informed individuals were randomly chosen.

Effects of Own and Peer Invitations. Because the invitees were randomly chosen, an immediate question of interest is how the invitation impacts both the individual's propensity to participate in the experiment and also the propensities of households in the invited household's social network to participate. The baseline regression we estimate is

$$y_{i,r} = \alpha_r + \beta invited_{i,r} + \gamma f\left(A_r, \mathbf{I}_r\right) + \delta f\left(A_r\right) + \epsilon_{i,r} \tag{3.1}$$

where i indexes the household, and v indexes the village. The dependent variable, $y_{i,r}$ is an indicator for either eventual participation in the lab games $(y^p_{i,r})$ or for showing up early to our experimental session $(y^s_{i,r})$, while $I_{i,r}$ is an indicator for whether or not individual i was randomly invited to participate. The third and fourth terms in the regression equation capture how the social network might diffuse information about the possibility to participate in our experiment. Namely, we let $\mathbf{I}_r := \{j \in V : \ I_{r,j} = 1\}$ – the collection of individuals in village v who received random invitations.

We consider several different definitions of $f\left(.\right)$. First, we ask if the number of invited friends influences a household's take-up of the game. In that case,

$$f_1\left(A_r\right) = \sum_{j \neq i}^{N_r} A_{r,ij}$$

which is simply the degree of individual i, and

$$f_1\left(A_r, \mathbf{I}_r\right) = \sum_{j \neq i} A_{ij} \cdot I_{j,r} \tag{3.2}$$

which is simply the number of invited friends. By including both terms in the regression, we control for the overall importance or popularity of the individual, so the "peer effect" coefficient, γ, only picks up the additional effect of randomly having an informed friend.

The second functional form that we try captures the idea that an individual might not only learn about the game through direct friends. Information may travel through longer paths to reach any individual. Let $D_{r,ij}$ represent the minimum distance between households i, j. If households i and j are not reachable, then $D_{r,ij} = \infty$. Let $\bar{D}_r$ denote the longest shortest-path between any two households in village r. Then, we define

$$f_2(A_r) = \sum_{j \neq i}^{N_r} A_{r,ij} \cdot \sum_{k=1}^{\bar{D}_r} \frac{1}{k} \mathbf{1}\{D_{r,ij} = k\} \text{ and} \tag{3.3}$$

$$f_2(A_r, \mathbf{I}_r) = \sum_{j \neq i}^{N_r} A_{r,ij} \cdot I_{j,r} \sum_{k=1}^{LP_v} \frac{1}{k} \mathbf{1}\{D_{r,ij} = k\}.$$

This function of the graph sums the inverse distance between household i and all other connected members of the village network. It makes an explicit assumption about how information transmission decays over longer paths.

Finally, we also ask if an individual's minimum distance from an invited member of the network affects that household's participation. Since every household shares an edge with at least one other household, $f_3(A_r) = 1$. Thus,

$$f_3(A_r, \mathbf{I}_r) = \min_{j \in V} \{D_{r,ij} : \; I_{j,r} = 1\}. \tag{3.4}$$

Information Decay. While the variants of Equation 3.1 capture some of the possible channels through which the network may matter, we can take one step further. A related question is how information transmission probabilities decay with social distance. Namely, we estimate

$$y_{i,r} = \alpha_r + \beta I_{i,r} + \sum_{k}^{\bar{D}_r} \gamma_k \sum_{j \neq i}^{N_r} A_{r,ij} \cdot \mathbf{1}\{D_{r,ij} = k\} \cdot I_{i,r} \tag{3.5}$$

$$+ \sum_{m}^{\bar{D}_r} \delta_m \sum_{j \neq i}^{N_r} A_{r,ij} \cdot \mathbf{1}\{D_{r,ij} = m\} + \epsilon_{i,r} \tag{3.6}$$

for $y_{i,r} \in \left\{ y_{i,r}^s, \; y_{i,r}^p \right\}$. We are also interested in testing

$$H_0 : \; \frac{\gamma_1}{\gamma_2} = \frac{\gamma_2}{\gamma_3} \tag{3.7}$$

In other words, does information flow in a multiplicative way? Suppose that conditional on being informed, I attend the session with probability $p_{participate}$. Suppose that I learn about information from one invited friend with probability q_1. Then, my participation likelihood is $p_{participate} q_1$. Similarly, suppose that conditional on one friend of social distance 2 being informed, I learn about the opportunity with probability q_2. Then we ask if it is the case that

$$p_{participate}q_2 = p_{participate}q_1^2$$

In a graph with unique paths between individuals, this would imply a constant transmission probability over time.

Effects of Past Experience. Because laboratory experiments of a similar structure with at least one common author had previously taken place in 32 of our 39 study villages either 1 or 2 years prior, we can ask if previous experience affects the decision to disseminate information about the time and place of the session or to participate, conditional on being informed.

The similarities with the experiments from Chandrasekhar et al. (2011a), Chandrasekhar et al. (2011b) and Chandrasekhar et al. (2012) are many. First, the experiments generally took place in large common areas of the village such as schools, temples, or dairy cooperative offices. Second, some of the survey staff was common between experiments. Third, once individuals reported to the experiments, registration procedures were very similar, with surveyors asking participants a basic set of questions that would allow each individual to be matched to the social networks data. Fourth, while the games had different economic content in each case, the broad goal of all of the experiments was to learn how social networks mediate the play of stylized games in the lab. As such, in many cases, payoffs were based on some aggregation of the play of several individuals. Fifth, individuals generally made decisions about the investment or allocation of resources in several rounds of play, and were randomly paid for one of the outcomes. Sixth, the overall levels of expected earnings were all around Rs120. Furthermore, all of the participants were part of the networks surveys of Banerjee et al. (2011) and were somewhat habituated to the idea of being surveyed.

It is also important to ask if prior experience with the game might impact a household's decision in a positive, negative, or unsigned fashioned. While we do not have any concrete evidence on player satisfaction, anecdotal evidence suggests that individuals by and large had a very positive experience playing the laboratory games. This is probably primarily driven by the fact that the experimental payoffs were the same order of magnitude as a day's wage, but only required a few hours of the respondent's time. During pilot rounds of the experiments in Breza et al. (2011), participants gave the unsolicited feedback that they hoped we would return to their villages in the future. It even seemed that the occasional individual who received low payoffs was not all that disappointed. We made it very clear during the experimental sessions that the computer would be randomly choosing which experimental round's results would comprise the final payouts. The survey staff did not receive any complaints by participants of this method being unfair.

Because of these similarities, past experience with the games might impact both who spreads information about the games, who looks for information about the games, and who is most likely to actually play the games when informed. We will defer the estimation of many of these types of parameters until Section 5. In the reduced form analysis we look for some baseline evidence of these effects. In the case of the spread of microfinance, Banerjee et al. (2011) find evidence that individuals who have taken microfinance are more likely to tell their friends about it. We look for a similar phenomenon.

Our main reduced form specification is

$$\begin{aligned} y_{i,r} &= \alpha_r + \beta I_{i,r} + \gamma Z_{i,r} + \delta_1 f_2\left(A_r\right) \\ &+\delta_2 f_2\left(A_v, \mathbf{I}_r\right) + \delta_3 f_2\left(A_r, \mathbf{Z}_r\right) \\ &+\delta_4 f_2\left(A_v, \mathbf{I}_r \circ \mathbf{Z}_r\right) + \epsilon_{i,v} \end{aligned} \tag{3.8}$$

where the function $f_2()$ is defined by Equation 3.3, and $Z_{i,r}$ in an indicator for whether household i played a laboratory game in the village over the preceding 2 years. If individuals enjoyed their experience in the past, then we would expect $\gamma > 0$. If we think that information about the previously played games easily diffuses to friends (conditional on that information being positive) then we would expect $\delta_3 > 0$. The δ_2 term captures learning and endorsement about the new opportunity from members of the network without any past experience who were randomly informed, while the δ_4 term captures any differential learning or endorsement from individuals who are both informed and who have past experience. We might expect δ_4 to be either positive or negative because of the rivalry of participation. Individuals with past experience may know how great the game is and might hesitate to broadcast information too loudly.

Village-Level Characteristics. While we only have data from 39 villages, we can still try to ask if there are any aggregate village characteristics that either help or hinder diffusion of information about the lab games. It may be the case that networks of specific shapes are better or worse for spreading information. Please see Appendix A for definitions of key network concepts.

First, we look for evidence of whether the average centrality (degree, eigenvector centrality, and betweenness centrality) of the randomly invited households impacts the overall fraction of households in the village who come to the participate in the experiment. Betweenness centrality is a notion about the fraction of shortest paths between all pairs of individuals on which a household sits. Thus giving information to more between individuals might imply better diffusion at the graph level.

Second, we also ask whether the average Shared Invited Neighbor Score in a village improves diffusion. If individuals do indeed prefer to pass information to clusters of friends, then it may be the case that if the village has a higher average Shared Invited Neighbor Score, then information about the experiments may spread to a higher fraction of individuals.

Finally, we also examine two separate network-level statistics that theorists have predicted to be important in the extent and the rate of diffusion. **?** show that the larger the first eigenvalue, the better information diffuses. Higher values of λ_1 imply more diffusion. Golub and Jackson (2009) suggest that for some types of networks, the second eigenvalue of the stochastized adjacency matrix provides a threshold for the rate at which information spreads. Smaller values implies faster rates.

Slots as Rival and Payoff Complementarities. One feature of the laboratory experiments (both past and current, from the perspective of the villagers) is that individuals might benefit from playing the games with a group of their friends. Most of the games played in the past had the feature that the payoff was determined by the joint play of pairs or triples of players. Individuals, therefore, may decide to strategically inform other residents of the village with the goal of recruiting the "right" group of friends. Specifically, it might be optimal to invite friends who are members of a clique who might in turn invite individuals who are closely linked to the original information source. Furthermore, the fixed number of participants used for the actual network experiments implies that slots are rationed. This is one hypothesis we investigate through our structural exercise.

4. Reduced Form Results

We now present the estimation of the specifications from section 3. In many of the regression tables, we consider two different outcome variables. In the household-level regressions, the first, "Early", is an indicator for whether a household turns up to the experiment in advance of the survey team. The second, "Late", is an indicator for whether a household eventually participates in the experimental sessions. We are more confident that individuals in the "Early" category became aware of the games through their village networks, while for "Late" participators, we sent our surveyors through the village to drum up extra interest. It should be noted, however, that invited households were guaranteed priority even if they did not turn up early.

Effects of Own and Peer Invitations. The basic reduced form, experimental results are presented in Table 3. Columns 1-3 use early turn-up as the dependent variable, while columns 4-6 use final participation. Further, the key network variable of interest is the number of invited friends in columns 1 and 4, the sum of the inverse distances to invited households in columns 2 and 5, and the minimum distance to an invited household in columns 3 and 6.

The first row of each specification shows the effect of being invited on either turning up early or late. Note that invited households are approximately 5 percentage points more likely to eventually participate in the experimental sessions in each specification. The coefficients are all quite precisely estimated. This is a large effect, considering overall turnout is on the order of 10% of households. However, note that columns 1-3 show that invited households are actually 2 percentage points less likely to arrive early to the experiments than the average, non-invited household. While perhaps surprising on face value, recall that invited households were guaranteed slots in the experiments regardless of when they actually arrived at the pre-specified location. Thus, it seems rational for invited households to wait to come to the games. Registering non-invited households and recruiting individuals to fill vacant slots did take up to one hour before the actual experiments began.

Each specification tells a consistent story vis a vis possible "peer effects" in information diffusion. Inviting one additional friend causes a household to increase its likelihood of coming early to the game by 0.9 percentage points and of eventually playing the game by 1.9 percentage points. Both effects are statistically significant. Similarly giving an additional invitation to an individual at social distance k increases an individual's likelihood of participating by $k * 2.9$ percentage points. Finally, going from having no connected, invited households to having a direct friend who is invited, increases the participation likelihood by 5.2 percentage points. All of the peer effect coefficients are significant at the standard levels. Furthermore, recall that in each specification, we control for the absolute number of friends, or the total sum of the inverse distances with all households in the network.

Information Decay. While the point estimates on the peer effects are large, Table 4 and the estimation of equation 3.5 gives us a better sense of the influence of invited peers as a function of social distance. Again, the model in column 1 uses early arrival as the dependent variable, while model 2 uses eventual participation. The first row of the table again shows the effect of an invitation on participation, and the coefficients look very similar to the previous specifications. Again, all of the functions of invited peer households are significant determinants of a given household's own participation decision. The effect of inviting one additional friend at minimum distance 1 on early arrival is 1.4 percentage points, while the effect sizes are 0.8, 0.4, and 0.4 percentage points giving an additional invitation to households at social distance 2, 3 or 4, respectively. It is quite striking that there is marginally significant effect for households at social distance 4. These regression results strongly imply that it is not simply enough to understand the incentives of the direct friends. The network is able to pass information between individuals at opposite sides of the graph.

Figure 1 displays the coefficients on the number of invited connections at social distance 1, 2, 3, or 4 variables from the "early" regression specification in a graphical format. Note that the propensity to turn up at the experiment (the y-axis) decreases as the distance between the household and each invited household (x-axis) increases. The relationship flattens between social distance 3 and social distance 4.

Given the coefficient estimates in column 1 of Table 4 and Figure 1, we perform tests to address the hypothesis in equation 3.7. First, we can conclude at standard significance levels that $\gamma_2 \neq \gamma_1$. However, the standard errors are too big to reject that $\gamma_2 = \gamma_3$. We find that the estimate $\widehat{\frac{\gamma_1}{\gamma_2}} = 1.707$ with standard error (0.550) while the estimate $\widehat{\frac{\gamma_2}{\gamma_3}} = 2.237$ with standard error (0.882). The test of $\frac{\gamma_1}{\gamma_2} = \frac{\gamma_2}{\gamma_3}$ cannot be rejected at standard levels using a non-linear Wald test. In fact, the test p-value is 0.5800, which is quite large. This suggests that if the network was organized with unique paths between individuals, information transmission probabilities would decay roughly exponentially in minimum social distance. However, because there are generally many paths between individuals, the pattern implies that the decay pattern is faster than exponential.

Effects of Past Experience. Table 5 presents evidence that past experience may have sizeable impacts on future participation. While past participation was not randomly assigned, individuals with past experience are approximately 13 percentage points more likely to turn up early to the experiment than individuals with no prior experience. This could be indicative that learning that the game is worthwhile has a causal impact on participation. It could also be the case that individuals who participated in the past are simply more central and thus are more likely to learn about the new opportunity. The coefficient in column 3 is even larger. Individuals with past experience are 22 percentage points more like to eventually participate in the new game. In this specification we also confirm that receiving an invitation increases participation by 2.9 percentage points. Furthermore, the coefficient on the interaction is quite large, 6 percentage points, but is not significant at the standard levels. While the coefficient on the interaction "Has Past Experience * Invited" is not significantly different from zero, a negative value might indicate that past participants understand that the registration process might last quite a while before the games actually begin.

In columns 2 and 4, we control for the sum of the inverse distances to all household in the village, which is one measure of centrality. It's only suggestive, but the first row coefficient on "Has Past Experience" barely changes from columns 1 and 3. The coefficient estimates (while many are only marginally significant) seem to imply that individuals are more likely to attend the games if they have more invited acquaintances or if they have more acquaintances with past experience. However, it does not seem to be the case that having invited friends who were themselves past participants has any additional effect on a household's own participation decision. If participation in the games is viewed as rival or if individuals prefer playing the games with their close friends, then perhaps, this non-result is not so surprising. Finally, it should also be noted that the results are robust to the alternate peer group specifications presented in Table 3.

Village-Level Characteristics. Finally, we analyze how village-level characteristics may influence the diffusion of information about our experiment. Table 6 describes the relationships between the average centrality of the invited households or the village's average Shared Invited Neighbor Score and village level turn-out. There is not much evidence that the average degree or the average eigenvector centrality of the invited households affects take-up. However, the average betweenness centrality of invited households is associated with increased take-up. Recall that we only have 39 observations in our regressions, so it is hard to make any definitive conclusions. The betweenness result is suggestive, though.

In column 4 of table 6, we find that villages with a higher Shared Invited Neighbor Score have substantially higher take-up. interestingly, those villages with higher baseline Shared Neighbor Scores have lower overall turn-out. In the strategic information diffusion setting, invited households may try to avoid passing information to other cliques with all uninvited households. As a result, information may diffuse thoroughly to the invited household's close friends, but may remain more or less local.

Finally we present correlations between other network characteristics and turn-out in Table 7. Columns 1-4 analyze the respective roles of the first eigenvalue of the adjacency matrix and the second eigenvalue of the stochastized adjacency matrix in information diffusion. In specifications 2 and 4, which have village size controls, neither measure seems to correlate with take-up. While network theory does predict a role for these measures, the theoretical results only cover threshold levels for the eigenvalues. We may not expect to see an effect if the values in all villages are far from those thresholds. In column 5, however, we do find that villages with higher levels of past experience do have substantially higher average turn-out levels.

5. Structural Model of Diffusion

While the reduced form results show that the network does matter for the spread of information in sometimes subtle ways, we propose and estimate a simple model to better understand the process by which information about the games spreads.

5.1. **Past Experience and Information Seeking.** First, we aim to better understand how informed households spread information about the opportunity both as a function of their own past experience as well as the experience of their friends. Furthermore, we look for evidence that individuals with past experience hold back as a result of the rival nature of slots in the experiment.

Model Time Line and Specification.

Full Model. Our model is a modified version of the information model estimated in Banerjee et al. (2011). However, in our case, we allow for both information seeking and differential propensities to spread information as a function of past experience. We propose a three period model with the following time line. On day 1, we invite a random set of households to a session of laboratory experiments 2 days later. The informed households then can pass information to their friends. On day 2, all of the currently informed individuals (either by random invite or by message from invited friends) can again pass information to their friends. Again, on the morning of day 3, all currently informed individuals can again pass information to their friends. Finally, after information has diffused, each household decides whether to participate in the session.

Working backwards, we assume that once informed, households with past experience participate in the experiments with probability $p_{Past,i}$ and that households without past experience participate with probability $p_{Not,i}$. Next, we specify the process by which informed households pass information to their friends. We estimate a set of transmission probabilities, $\left(q_{Not}^{not}, q_{Not}^{past}, q_{Past}\right)$, with q_{Past} representing the probability that an informed household with past experience tells any friend about the game. For individuals without past experience, we allow for differential transmission rates as a function of the experience level of the information recipient. The likelihood that an inexperienced individual passes information to an inexperienced friend is q_{Not}^{not} while the probability of her telling experienced friends is q_{Not}^{past}.[6] We assume that any household can only directly pass information to those households that share an edge in the graph.

Differential Seeking. We also specify and estimate a second model with parameters $(q^{not}, q^{past}, p_{Not}, p_{Past})$. In this setup, we allow individuals to transmit information to experienced and inexperienced households with different probabilities. We do not differentiate the experience levels of the speakers. An alternate way to think about this specification is that experienced individuals seek out information about new opportunities with different rates (perhaps more aggressively.)

Differential Speaking. Finally, we suggest a third model with parameters $(q_{Not}, q_{Past}, p_{Not}, p_{Past})$, where q_{Not} is the probability of the speaker transmitting information if the speaker is inexperienced. q_{Past} is the transmission probability if the speaker is experienced. We force individuals to be willing to speak with any other type of household with the same likelihood conditional on the individual's own experience.

Model Estimation. We discuss estimation for the full model, but the procedure is extremely similar for all models. The participation likelihoods (p_{No}, p_{Past}) are quite simple to estimate. Because for the subset of invited households, we know that each household was informed, and we know the participation outcome, we simply equate

[6]We limit the past participants to only one transmission probability, q_P for computational feasibility. We think that it is more likely for the experience of others to matter from the point of view of inexperienced households. We will try to relax this assumption in future drafts.

the participation probability with the average participation rate for experienced and inexperienced individuals.[7] Our large sample size and randomized invitation design makes this step quite simple. We denote estimates coming from this step as $(\hat{p}_{No}, \hat{p}_{Past})$.[8]

We are interested in understanding how information about the game diffuses, but we only observe the final participation decision of each household. Therefore, the outcome of interest (information penetration each period) is a latent variable. Given any guess for the full model parameters, $\left(q_{Not}^{not}, q_{Not}^{past}, q_{Past}, \hat{p}_{No}, \hat{p}_{Past}\right) = (q, \hat{p})$ we can then simulate the model for three periods. Let Z_r be the set of experienced households in village v, and let NE_v be the set of uninformed households in village v. Then at the end of the first period, the likelihood of a household i with past experience becoming informed is

$$\begin{aligned} \mathrm{P}\left(i \text{ no info}, t = 1 | i \in Z_r, q, \hat{p}\right) &= \prod_{j \in E_r, A_{r,ij}=1} \left(1 - \left(q_{Past}\mathbf{1}\left\{j \in Z_r\right\} + q_{Not}^{past}\mathbf{1}\left\{j \in NE_r\right\}\right) I_j\right) \\ &\quad \times \prod_{k \in NE_r, A_{r,ij}=1} \left(1 - \left(q_{Past}\mathbf{1}\left\{j \in Z_r\right\} + q_{Not}^{not}\mathbf{1}\left\{j \in NE_r\right\}\right) I_j\right) \end{aligned}$$

We can repeat this process for each period in the model. Then, predicted attendance at the experiment can be calculated in the following way:

$$\begin{aligned} \mathrm{P}\left(i \text{ participates} | i \in Z_r, q, \hat{p}\right) &= \hat{p}_{Past}\mathrm{P}\left(i \text{ informed}, t = 3 | i \in Z_r, q, \hat{p}\right) \\ \mathrm{P}\left(i \text{ participates} | i \in NE_r, q, \hat{p}\right) &= \hat{p}_{Not}\mathrm{P}\left(i \text{ informed}, t = 3 | i \in NE_r, q, \hat{p}\right) \end{aligned}$$

To estimate the information transmission parameters, we perform the method of simulated moments (MSM), calculating moments from the simulated models and comparing them to the empirical moments observed in the data. Because we are estimating 3 parameters in this way, we need to use at least three moments to provide identification. We base our moment selection on that of Banerjee et al. (2011) and also add variants of the moments that depend on past experience.

(1) Share of households with no neighbors taking up who participate.

[7] We use the early turn-out indicator as our key participation outcome for the structural exercise. However, recall that invited households are less likely to turn up early because of their prioritization. To solve this, when estimating the probabilities to come to the games, we look at $\max\{Early, Late\}$ for the invited households.

[8] Note that we implicitly assume that all individuals have the same participation likelihood conditional on experience and being informed regardless of being invited or not. It may be the case that the guaranteed participation would imply a higher participation rate for invited households. We plan to estimate the participation probabilities separately for invited and non-invited households in future versions.

(2) Share of experienced households with no neighbors taking up who participate.
(3) Share of households that are in the neighborhood of an invited household who participate.
(4) Share of experienced households that are in the neighborhood of an invited household who participate.
(5) Covariance of the fraction of households participating with the share of their neighbors who participate.
(6) Covariance of the fraction of experienced households participating with the share of their neighbors who participate.
(7) Covariance of the fraction of households participating with the share of second-degree neighbors who participates.
(8) Covariance of the fraction of experienced households participating with the share of second-degree neighbors who participates.

We adhere closely to the estimation and bootstrapping procedure of Banerjee et al. (2011), so we only provide a cursory explanation here. To estimate the information transmission parameters, we first create a grid of possible values for each of the three parameters. Then for each grid point, we simulate the model 75 times and calculate the average value of each of the 10 moments, $m_{sim,r}(q,\hat{p})$ for each village, r. The parameter estimates, $(\hat{q})$ are chosen as the minimizers of the GMM criterion function

$$\hat{q} = argmin_q \left(\mathbb{E}_R\left[m_{sim,r}(q) - m_{emp,r}\right]\right)\left(\mathbb{E}_R\left[m_{sim,r}(q) - m_{emp,r}\right]\right)'$$

Our bootstrap follows a Bayesian algorithm and allows for an arbitrary within-village correlation structure. Entire village blocks are sampled with replacement from the 39 total villages. Using the same set of average village moments, $m_{sim,r}(q)$, evaluated at each grid point, we can construct a new, criterion function and optimal parameter value for each bootstrap iteration. In the current version of the paper, we perform inferences based on 150 bootstrap iterations.

Estimation Results. Table 8 presents results from estimating the three information diffusion models. The first column shows the full model parameters and bootstrapped standard errors while columns 2 and 3 show results from the Information Seeking and Speaking models, respectively. Because the participation rates are estimated the same way for each model, the values are the same.

We find that individuals are substantially more likely to inform people who have played in the past. The difference is stark. In both columns 1 and 2 the transmission probability to individuals with past experience is close to 1 while the transmission probability to inexperienced households is close to 0. This may be caused by informed individuals choosing to inform those other individuals who they think would most benefit from the news. One alternative interpretation is that past players seek out information about future opportunities.

Column 3 shows the third specification, which looks for differential telling as a function of the status of the informed households. Interestingly, we find $q_N > q_P$, which implies that individuals with no past experience work harder to spread information about the experiments. The difference between the two parameter estimates is marginally significant. This pattern is again suggestive of individuals responding to the rivalry characteristic of the good. Finally, we can also compare q_P in the first model to the average $\bar{q_N} = 0.246$. Again, $\bar{q_N} > q_P$, but not statistically significantly. Again, this suggests that for rival goods, individuals who have a high gain from those goods are less likely to broadcast information about them.

5.2. **Which Friends to Tell?** We are currently investigating the strategic incentives to inform others. Given that there are a limited number of slots available to participate in the experiment and that there is likely to be greater value in participating in the experiment with friends, there may be differential incentives as to which sort of friend one would like to inform. We develop models underpinning these forces here.

5.2.1. *Myopic Model.* The first version of the model is purely myopic. The timing is as in the previous section. The primary difference is that individuals who are informed in turn inform their friends based on whether or not they share another friend in common. For example let i be the informed node and j and k are two friends of i. We say j is a supported neighbor of i if there exists an $l \in V$ such that $A_{il}A_{jl} = 1$. Assume j is a supported friend of i but k is not. We assume that i tells j, a supported friend, with probability q_s and i tells k, an unsupported friend, with probability q_u. In this model, we will estimate $(q_u, q_s, p_{Not}, p_{Past})$ and are interested in whether $q_s > q_u$. Specifically, if $q_s > q_u$, this provides evidence that individuals discriminate between which of their neighbors they provide information to and, moreover, are more likely to inform friends with whom they have other friends in common. This potentially allows them to limit the extent to which information leaks out of their local neighborhood.

5.2.2. *Boundedly Rational Model.* The second version of the model has strategic elements and a dynamic component. In this environment, ever period after an individual has become informed, she will choose whether or not to disseminate information to their supported neighbors and their unsupported neighbors. Specifically, in every period $t \leq 3$, an individual (if informed before t) picks $a_{i,t} \in \{0,1\}^2$. Here, $(0,0)$ means the individual tells neither unsupported nor supported neighbors, $(1,0)$ means that the individual tells only unsupported neighbors, $(0,1)$ means that the individual only tells supported neighbors, and $(1,1)$ means that both are told.

How is the decision made? Here we assume that an individual knows the other individuals who have been informed in previous periods as well. Then the individual assumes that each individual myopically passes information to each of their supported neighbors with probability q_s and unsupported neighbors with probability q_u in the present and subsequent periods. Using this heuristic, the individual computes the optimal choice of $a_{i,t}$ in the present period.

Payoffs are given by

$$V_{i,t}\left(a_{i,t}|\mathbf{I}^{t-1}\right) = \mathrm{E}\left[\left(\frac{24}{\#\text{Attending}}\right) \cdot \sum_{j \in V} \left\{\frac{1}{D_{ij}^{\gamma}} \cdot \text{Attending}_j\right\}\right]$$

where the expectation is taken with respect to the process above. Namely, given that all informed people in period $t-1$ know each others' identities, they compute the expected value of attending the game as a function of the probability that they will participate (24/#Attending) and the value of each other individual's attendance. Since these are games in which social distance has been shown to correlate highly with payoffs, we allow the payoffs to depend on D_{ij} as well as a decay parameter γ to be estimated in the model. As such, the model parameters are (q_u, q_s, γ) and we are interested in the distribution of actions $\left\{a_{i,t} : \ I_i^{t-1} = 1\right\}$. Our hypothesis, ceteris paribus, is that the proportion of times an individual chooses $(0,1)$ is higher than the proportion of times an individual chooses $(1,0)$.

6. Conclusion

The random nature of our invitations to play laboratory games offers a unique opportunity to better understand how information about rival goods diffuses through a network. We show through both the reduced form and structural estimation that individuals may sometimes hold back or strategically inform a subset of friends about beneficial village-level activities. In future work, we hope to be able to structurally model strategic information diffusion resulting from payoff complementarities.

References

Arora, S., S. Rao, and U. Vazirani (2004): "Expander flows, geometric embeddings and graph partitioning," in *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, ACM, 222–231.

Bandiera, O. and I. Rasul (2006): "Social Networks and Technology Adoption in Northern Mozambique*," *The Economic Journal*, 116, 869–902.

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. Jackson (2011): "The Diffusion of Microfinance," MIT working paper.

——— (2012): "Diffusion in Networks: Evidence from a Field Experiment," *Working Paper*.

Breza, E., A. G. Chandrasekhar, and H. Larreguy (2011): "Punishment on Networks: Evidence from a Framed Field Experiment," *Working Paper*.

Chandrasekhar, A. G., C. Kinnan, and H. Larreguy (2011a): "Informal Insurance, Social Networks, and Savings Access: Evidence from a lab experiment in the field," *MIT Working Paper*.

——— (2011b): "Information, Savings, and Informal Insurance: Evidence from a lab experiment in the field," *MIT Working Paper*.

Chandrasekhar, A. G., H. Larreguy, and J. Xandri (2012): "Testing Models of Social Learning on Networks: Evidence from a Framed Field Experiment," *Working Paper*.

Cohen, L., A. Frazzini, and C. Malloy (2008): "The Small World of Investing: Board Connections and Mutual Fund Returns," *Journal of Political Economy*, 116, 951–979.

Conley, T. and C. Udry (2010): "Learning about a new technology: Pineapple in Ghana," *The American Economic Review*, 100, 35–69.

Foster, A. D. and M. R. Rosenzweig (1995): "Learning by Doing and Learning from Others: Human Capital and Technical Change in Agriculture," *Journal of Political Economy*, 103, 1176–1209.

Golub, B. and M. Jackson (2009): "How homophily affects learning and diffusion in networks," *Working Papers*.

Hagen, L. and A. Kahng (1992): “New spectral methods for ratio cut partitioning and clustering,” *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 11, 1074–1085.

Iyer, R. and M. Puri (2008): “Understanding bank runs: the importance of depositor-bank relationships and networks,” *NBER Working Paper*.

Jackson, M., T. Barraquer, and X. Tan (2010): “Social Capital and Social Quilts: Network Patterns of Favor Exchange,” .

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

Kelly, M. and C. Grada (2000): “Market Contagion: Evidence from the Panics of 1854 and 1857,” *American Economic Review*, 1110–1124.

Kremer, M. and E. Miguel (2007): “The Illusion of Sustainability*,” *The Quarterly Journal of Economics*, 122, 1007–1065.

Figures

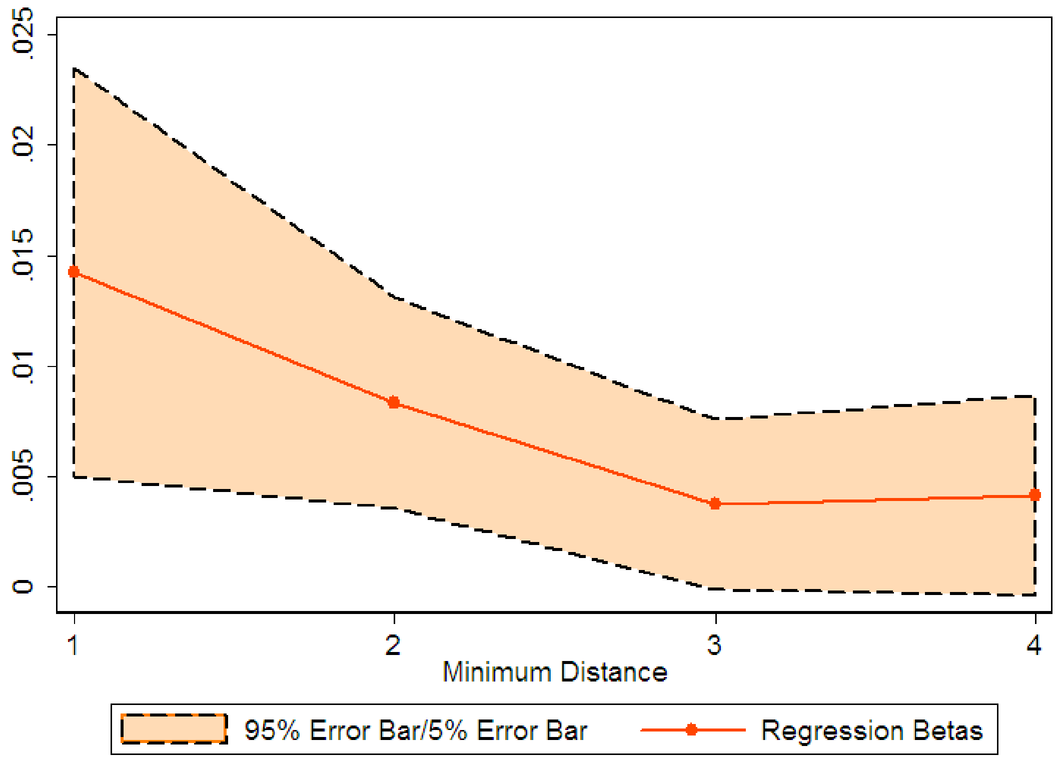

Figure 1. Impact of Randomly Invited Households on Participation as a Function of Social Distance.

Tables

Table 1. Individual-Level Summary Statistics

	Mean	Std. Dev.
Eigenvector Centrality	0.0530	(0.0489)
Betweenness Centrality	0.00819	(0.0135)
Number of Friends	8.978	(7.480)
Number of Min(Distance)=2 HHs	56.95	(34.80)
Number of Min(Distance)=3 HHs	87.72	(43.18)
Number of Min(Distance)=4 HHs	31.24	(34.55)
Sum of Inverse Distances to All Connected HHs	75.24	(29.63)
Shared Neighbor Score	2.002	(1.887)
Has Past Experience	0.153	(0.360)
Invited to Game	0.0936	(0.291)
Showed Up Early	0.0638	(0.245)
Eventually Participated	0.109	(0.312)
Total Number of Individuals	7502	

Table 2. Village-Level Summary Statistics

	Mean	Std. Dev.
Average Degree	8.852	(1.762)
Average Eigenvector Cent.	0.0562	(0.0120)
Average Betweenness Cent.	0.00887	(0.00249)
Avg. Degree of Inviteds	9.343	(2.033)
Avg. Eig. Cent. of Inviteds	0.0598	(0.0162)
Avg. Between. Cent. of Inviteds	0.00967	(0.00350)
Average Shared Neighbor Score	2.009	(0.767)
Avg. Shared Invited Neighbor Score	0.224	(0.133)
First Eigenvalue of Ajd. Mat.	14.58	(2.939)
Second Eigenvalue of Stoch. Mat.	0.812	(0.0700)
Past Experience Fraction	0.175	(0.117)
Showed Up Fraction	0.0699	(0.0466)
Village Size	192.4	(60.90)
Number of Villages	39	

Table 3. Basic Diffusion Regressions: Participation, Invitations and Invited Friends

	(1) Early	(2) Early	(3) Early	(4) Late	(5) Late	(6) Late
Invited to Game	-0.0226	-0.0167	-0.0228	0.0468	0.0565	0.0487
	(0.00817)	(0.00828)	(0.00833)	(0.0167)	(0.0168)	(0.0169)
Number of Invited Friends	0.00889			0.0185		
	(0.00435)			(0.00571)		
Number of Friends	0.00161			0.00345		
	(0.000623)			(0.000934)		
Sum 1/Dist. to Invited HHs		0.0172			0.0291	
		(0.00486)			(0.00572)	
Sum 1/Dist. to all HHs		-0.000705			-0.001000	
		(0.000413)			(0.000535)	
Min. Dist. to an Invited HH			-0.0250			-0.0520
			(0.00615)			(0.00762)
Constant	0.0437	0.00413	0.106	0.0578	-0.0135	0.186
	(0.00463)	(0.0102)	(0.00923)	(0.00614)	(0.0131)	(0.0114)
Observations	7502	7502	7100	7502	7502	7100
Adjusted R^2	0.031	0.031	0.030	0.026	0.023	0.017

Standard Errors are Clustered at the Village Level. Early is an indicator for showed-up early. Late indicates eventual participation. All specifications include village fixed effects.

Table 4. Participation as a Function of Invited HHs at Various Distances

	(1) Early	(2) Late
Invited to Game	-0.0182	0.0503
	(0.00803)	(0.0167)
Number of Friends	-0.000141	-0.00000271
	(0.000773)	(0.00116)
Number of Invited Friends	0.0142	0.0236
	(0.00471)	(0.00600)
Number of Min(Distance)=2 HHs	-0.000285	0.0000302
	(0.000236)	(0.000313)
Number of Min(Distance)=2 Invited HHs	0.00834	0.00959
	(0.00244)	(0.00286)
Number of Min(Distance)=3 HHs	-0.000289	-0.000234
	(0.000159)	(0.000218)
Number of Min(Distance)=3 Invited HHs	0.00373	0.00153
	(0.00196)	(0.00254)
Number of Min(Distance)=4 HHs	-0.000266	-0.000273
	(0.000206)	(0.000256)
Number of Min(Distance)=4 Invited HHs	0.00415	0.00317
	(0.00231)	(0.00275)
Constant	0.0227	0.0421
	(0.00940)	(0.0111)
Observations	7502	7502
Adjusted R^2	0.032	0.029

Standard Errors are Clustered at the Village Level. Early is an indicator for showed up to the experiments early. Late indicates eventual participation. All specifications include village fixed effects.

Table 5. Participation and Previous Experience

	(1) Early	(2) Early	(3) Late	(4) Late	(5) Early*	(6 La
Has Past Experience	0.137	0.130	0.223	0.209		
	(0.0184)	(0.0182)	(0.0172)	(0.0172)		
Invited to Game	-0.0207	-0.0171	0.0288	0.0324	-0.0149	0.1
	(0.00775)	(0.00751)	(0.0182)	(0.0180)	(0.0123)	(0.0
Has Past Experience * Invited	-0.0272	-0.0354	0.0638	0.0672		
	(0.0352)	(0.0378)	(0.0462)	(0.0464)		
Sum 1/Dist. to Connected HHs		-0.000851		-0.000739	-0.000314	-0.0
		(0.000460)		(0.000534)	(0.000939)	(0.00
Sum 1/Dist. to Conn., Inv. HHs		0.0125		0.0143	0.00871	0.0
		(0.00640)		(0.00818)	(0.0135)	(0.0
Sum 1/Dist. to Past Players		0.00474		0.00416		
		(0.00261)		(0.00305)		
Sum 1/Dist. to Past Play.*Invited		-0.0185		0.00780		
		(0.0129)		(0.0159)		
Constant	0.0454	-0.00124	0.0713	-0.0227	0.0223	0.0
	(0.00280)	(0.0114)	(0.00293)	(0.0153)	(0.0119)	(0.0
Observations	7502	7502	7502	7502	1621	16
Adjusted R^2	0.060	0.064	0.074	0.082	0.020	0.0

Standard Errors are Clustered at the Village Level. Early is an indicator for showed-up early.
Late indicates eventual participation. All specifications include village fixed effects
* Designates sample restriction to only those villages with no prior laboratory experience.

Table 6. Village Average Characteristics and Experiment Turn-Out

	(1)	(2)	(3)	(4)
Avg. Degree of Inviteds	0.00340			
	(0.00436)			
Average Degree	-0.00619			
	(0.00456)			
Avg. Eig. Cent. of Inviteds		0.775		
		(0.642)		
Average Eigenvector Cent.		0.0996		
		(1.294)		
Avg. Between. Cent. of Inviteds			4.453	
			(2.419)	
Average Betweenness Cent.			3.182	
			(7.191)	
Shared Invited Neighbor Score				0.218
				(0.0952)
Shared Neighbor Score				-0.0200
				(0.00980)
Constant	0.149	0.0465	0.00548	0.0640
	(0.0411)	(0.129)	(0.108)	(0.0441)
Observations	39	39	39	39
Adjusted R^2	0.134	0.142	0.193	0.217

Robust standard errors are reported.
The dependent var. in all cols. is the frac. of HHs that showed up early to the experiment.

Table 7. Village Network Characteristics and Experiment Turn-Out

	(1)	(2)	(3)	(4)	(5)
First Eigenvalue of Ajd. Mat.	-0.00429	-0.00211			
	(0.00193)	(0.00177)			
Village Size 100s		0.0134		0.00722	-0.00451
		(0.0220)		(0.0219)	(0.0199)
Inverse Village Size		14.08		14.22	-5.590
		(7.873)		(8.056)	(10.09)
Second Eigenvalue of Stoch. Mat.			-0.0987	0.0524	
			(0.129)	(0.0816)	
Village Had Prior Expts.					-0.00231
					(0.00111)
Past Experience Fraction					0.504
					(0.215)
Constant	0.132	-0.00619	0.150	-0.0684	0.0907
	(0.0315)	(0.0835)	(0.108)	(0.101)	(0.0886)
Observations	39	39	39	39	39
Adjusted R^2	0.048	0.199	-0.004	0.188	0.263

Robust standard errors are reported.

The dependent var. in all cols. is the frac. of HHs that showed up early to the experiment.

Table 8. Structural Estimation Results

	(1) Full Model	(2) Differential Seeking	(3) Differential Speaking
p_N	0.135	0.135	0.135
	(0.017)	(0.017)	(0.017)
p_P	0.521	0.521	0.521
	(0.060)	(0.060)	(0.060)
q_N^n	0.025		
	(0.021)		
q_N^p	0.971		
	(0.038)		
q^n		0.040	
		(0.014)	
q^p		0.983	
		(0.022)	
q_N			0.180
			(0.043)
q_P	0.167		0.100
	(0.056)		(0.039)

Appendix A. Glossary of Network Statistics

In this section we briefly discuss the network statistics used in the paper. Jackson (2008) contains an excellent and extensive discussion of these concepts which the reader may refer to for a more detailed reading.

Path Length and Social Proximity. The *path length* between nodes i and j is the length of the shortest walk between the two nodes. Denoted $\gamma(i,j)$, it is defined as $\gamma(i,j) := \min_{k\in\mathbb{N}\cup\infty}\left[A^k\right]_{ij} > 0$. If there is no such walk, notice that $\gamma(i,j) = \infty$. The *social proximity* between i and j is defined as $\gamma(i,j)^{-1}$ and defines a measure of how close the two nodes are with 0 meaning that there is no path between them and 1 meaning that they share an edge. In figure 2, $\gamma(i,j) = 2$ and $\gamma(i,k) = \infty$.

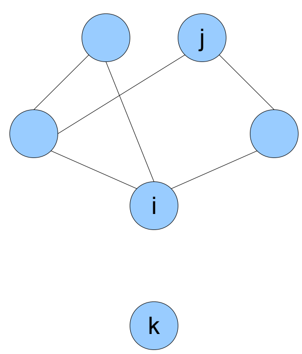

Figure 2. Path lengths i, j and i, k

Vertex characteristics. We discuss three basic notions of network importance from the graph theory literature: degree, betweenness centrality, and eigenvector centrality. The *degree* of node i is the number of links that the node has

$$d_i = \sum_{j=1}^{N} A_{ij}$$

. In figure 3(a), i has degree 6 while in (b) i has degree 2. While this is an intuitive notion of graphical importance, it misses a key feature that a node's ability to propagate information through a graph depends not only on the sheer number of connections it has, but also how important those connections are. Figure 3(b) illustrates an example where it is clear that i is still a very important node, though a simple count of its friends does not carry that information. Both betweenness centrality and eigenvector centrality address this problem.

The *betweenness centrality* of i is defined as the share of all shortest paths between all other nodes $j, k \neq i$ which pass through i. This is a normalized measure which is useful when thinking about a propagative process traveling from node j to k as taking the shortest available path.

The *eigenvector centrality* of i is a recursive measure of network importance. Formally, it is defined as the ith component of the eigenvector corresponding to the maximal eigenvalue of the adjacency matrix representing the graph.[9] The intuition for its construction is that one may be interested in defining the importance of a node as proportional to the sum of all its network neighbors' importances. By definition the vector of these importances must be an eigenvector of the adjacency matrix and restricting the importance measure to be positive means that the vector of importances must be the first eigenvector. Intuitively, this measure captures how well information flows through a particular node in a transmission process. Relative to betweenness centrality, a much lower premium is placed on a node being on the exact shortest path between two other nodes. We can see this by comparing figure 3(b), where i has a high eigenvector centrality and high betweenness, to (c), where i still has a rather high eigenvector centrality but now has a 0 betweenness centrality since no shortest path passes through i.

[9]The adjacency matrix A of an undirected, unweighted graph G is a symmetric matrix of 0s and 1s which represents whether nodes i and j have an edge.

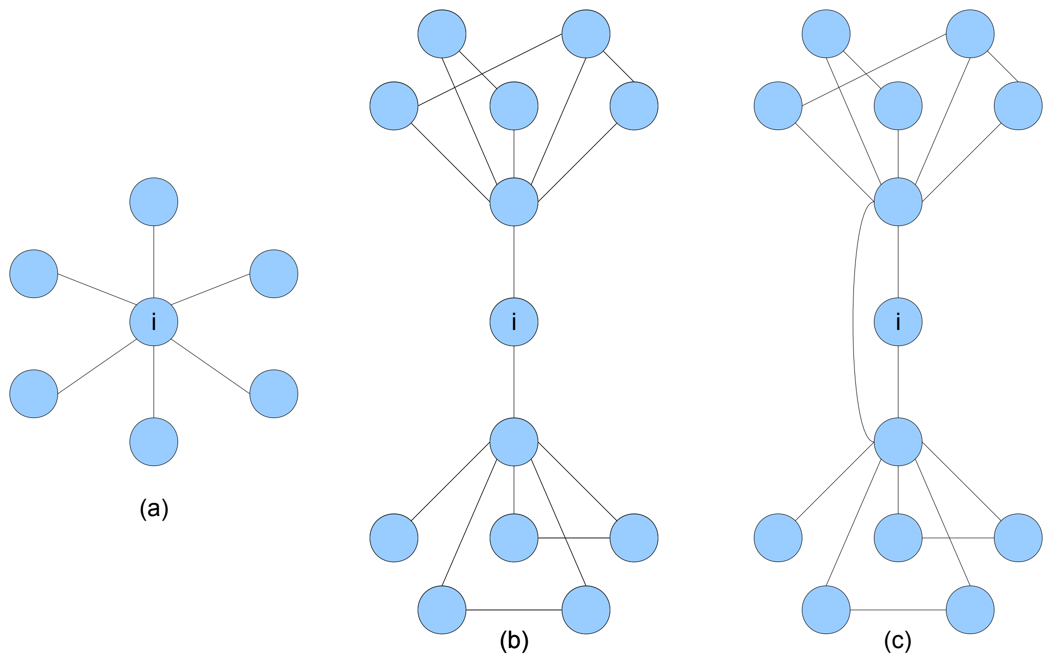

FIGURE 3. Centrality of node i

Spectral Partition. One exercise performed in graph theory is to partition the set of nodes into two groups such the information flow across the groups is low while the information flow within the groups is high. These partitions are of economic interest insofar we can think of information traveling from i to j not simply along the shortest path between the two nodes but through possibly many paths. The full flow process of information may carry important economic data. Network statistics which capture this feature, therefore, may be important to study.

There are numerous ways to partition the network including minimum cut, minimum-width bisections, and uniform sparsest cut. See Arora et al. (2004) for a recent discussion. The general result in this literature is that finding the cut is NP-hard. Consequently, approximation algorithms must be used.

We employ a simple approximation described as follows. Given a graph $G = (V, E)$, we are interestd in a partition of V into disjoint sets U and W such that $\frac{\sum_{i \in U} \sum_{j \in W} A(G)_{ij}}{|U||W|}$ is minimized. Following a simple approximation motivated by Hagen and Kahng (1992), we compute the "side" of node i based on the sign of ξ_i

where ξ_i is the eigenvector corresponding to the second smallest eigenvalue of the Laplacian of the graph G, defined as $L(G) = D - A$ where $D = diag\{d_1, ..., d_n\}$ a diagonal matrix of degrees and A is the adjacency matrix. Figure 4 illustrates the intuition of the partition. We say nodes i and j are on the same side of the spectral partition if $sign(\xi_i) = sign(\xi_j)$.

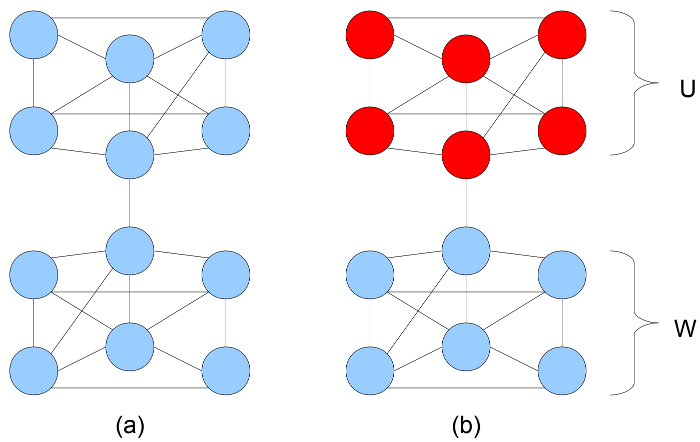

FIGURE 4. Spectral partition of V into U and W

First Eigenvalue. The first eigenvalue refers to the largest eigenvalue of the adjacency matrix. It gives a measure of how well information diffuses through a network. Higher values imply greater diffusion.

Fraction of Nodes in Giant Component. The giant component of a graph is the largest subset of nodes in which all pairs of nodes are reachable.

Second Eigenvalue of the Stochastized Adjacency Matrix. The stocastized adjacency matrix is a degree-scaled version of A_r. $\tilde{A}_{r,ij} = \frac{A_{r,ij}}{d_i}$. This matrix captures communication flows in the network. The second eigenvalue of $\tilde{A}_r$ is simply the second largest eigenvalue of the matrix. Golub and Jackson (2009) show that the second eigenvalue puts a bound on the rate of diffusion in some types of models. A larger value implies slower convergence of information.