Examining the Effect of Social Network on Prediction Markets through a Controlled Experiment*

Liangfei Qiu[a※], Huaxia Rui[b], Andrew Whinston[c]

[a] *Department of Economics, University of Texas at Austin, Austin, TX 78712, USA*

[b] *Simon School of Business, University of Rochester, Rochester, NY 14627, USA*

[c] *McCombs School of Business, University of Texas at Austin, Austin, TX 78712, USA*

※Corresponding author. E-mail address: liangfeiqiu@utexas.edu (L. Qiu).

Abstract

This paper examines the effect of a social network on prediction markets using a controlled laboratory experiment that allows us to identify causal relationships between a social network and the performance of an individual participant, as well as the performance of the prediction market as a whole. Through a randomized experiment, we first confirm the theoretical predictions that participants with more social connections are less likely to invest in information acquisition from outside information sources and perform significantly better than other participants in prediction markets through free-riding. We further show that when the cost of information acquisition is low, a social-network-embedded prediction market outperforms a non-networked prediction market. We also find strong support for peer effects in prediction accuracy among participants.

Keywords: information exchange, social networks, prediction markets

* We thank Eric K. Clemons at Wharton, Oliver Hinz at Technische Universität Darmstadt, Robert J. Kauffman at Singapore Management University, Sarah Rice at University of Connecticut, Dale O. Stahl at UT Austin, and Thomas A. Weber at EPFL for their helpful comments. We also thank the seminar participants at the 2012 International Conference on Information System (ICIS) and the 46th Hawaii International Conference on System Sciences (HICSS).

1. Introduction

Prediction markets have long been regarded as an effective way to tap into the wisdom of crowds by aggregating dispersed information within a social system [4, 6, 15, 16, 20, 35]. Several empirical studies have demonstrated the power of prediction markets in areas such as political science [4], supply chain management [6, 22], marketing [10, 16, 33], and finance [5]. In most of the previous literature, researchers have assumed that the participants in the prediction markets are isolated: They receive small bits and pieces of independent information and cannot affect the decisions of other participants. However, in reality, people often mobilize their social networks to collect information and opinions on a variety of issues. CNBC recently reported an effective information exchange network through which tweeting with fellow farmers has become a way for participants in a far-flung and isolating business to compare notes on everything from weather conditions to new fertilizers.[1] These tweets are dramatically accelerating the flow of information that may give investors an edge in the commodities market. With the advance of information technologies and the rise of social media, information exchange is ubiquitous these days. Indeed, people can use their smartphones or computers to share information with their social network neighbors at almost any place, at any time. The ubiquity of information exchange on social networks and the lack of understanding about their effects on prediction markets motivate us to explore the following research question: How does information exchange among the participants of a prediction market affect the behavior and performance of the network's participants?

Only a few attempts have been made in the previous literature to address this research

[1] The information is from CNBC News, March 8, 2011. The CNBC reporter called the phenomenon "Trading on Twitter." Grisafi, known as @IndianaGrainCo on Twitter, said he tweets with at least 15 farmers on a regular basis to check on crop conditions.

question or other similar questions. For example, in a different context, Coval and Moskowitz [13] asked a similar question and found that social networks help fund managers earn above-normal returns in nearby investments: The average fund manager generates an additional 2.67% return per year from local investments, relative to nonlocal holdings. The closest research to the present paper is a recent work that used game theory to study the Bayes-Nash equilibrium of an incomplete information game among participants in a social-network-embedded prediction market [30]. They found a symmetric equilibrium by which participants with few social connections typically exert effort to acquire information, whereas participants with many social connections typically free-ride others' information. However, in their stylized model, they made several simplifying assumptions: 1) people can always observe information from their direct neighbors; and 2) people are fully rational and have infinite computation capacity to integrate information in an optimal way.

Apparently, these assumptions might not hold in some real-world contexts. However, relaxing these assumptions in an analytical model could easily yield intractability of the results. To further our understanding of the research question without confining ourselves to these assumptions, we take a different approach in this paper by carrying out an experimental study. In particular, to address the research question, we test a series of hypotheses through our experiments. First, we test how participants' degree (the number of social connections) in the social network influences their decisions regarding whether to invest in information acquisition and affects their performance in the form of earnings. Following [30], information acquisition in our paper specifically refers to information gathering from outside sources and does not include asking network neighbors for information. Unlike an experimental approach, the traditional econometric methods are often subject to identification difficulty because the network structure is usually endogenously determined [29], as a result, empirically disentangling the unobserved individual characteristics (e.g., the predictive ability) from the actual effects of network degree on an individual's information acquisition and prediction performance is difficult. In our controlled experiment, participants are randomly assigned to different network positions, which allows us to identify the causal relationship between network structure and the individual's information acquisition and prediction performance. The experimental results are consistent with the theoretical prediction: participants with higher degrees in the social network are less likely to invest in information acquisition, compared with participants that have lower degrees, and they actually earn more by free-riding neighbors' information.

Second, the wisdom of crowd effect has been extensively studied in the literature [28]: The average of many individuals' estimates can cancel out errors and be surprisingly close to the truth. However, this approach requires independent estimates, which are rare in a social networking world. Lorentz et al. [28] demonstrate that sharing information corrupts the wisdom of the crowds. Contrary to previous work, our study shows that information sharing in a social network need not undermine the wisdom of crowd effect. The experimental results suggest that when the cost of information acquisition is low, a social-network-embedded prediction market outperforms a prediction market without a social network in terms of prediction accuracy. On the other hand, when the cost of information acquisition is high, we do not find any significant difference between the performances of these two types of prediction markets. In addition to these two major hypotheses, we also test whether the structure of the underlying social network has any effect on the performance of prediction and the experimental results suggest that network structure does matter.

The rest of the paper is organized as follows. Next, we review the related literature. In the third section, we outline a simple analytical model that motivates our hypotheses tested in the experiment. We describe the experiment and the analysis of the experimental results in the fourth section. In the fifth section, we present some simulation results on the prediction performance that complement our experimental results. In the sixth section, we conclude the paper.

2. Literature Review

A large body of literature explores the role of social networks in student alcohol use [19], product adoption [1], financial markets [12], the use of technology [34, 35], and health plan choice [32]. The standard empirical approach is a regression of an individual's behavior on his or her social connectedness or his or her peers' behaviors. The growing literature on the identification of the effect of network structure and social influence has recognized an econometric challenge: The network structure is endogenously determined [18, 29]. In our context, the network structure can be the result of past prediction performance. The confounding factors, such as participants' unobserved characteristics, make it difficult to identify the causal effect of network structure on an individual's behavior. For example, the positive correlation between social connectedness and individuals' prediction performance can be driven either by the actual social effect or the unobserved individual characteristics. In the first case, individuals gain from their social ties. In the second case, individuals self-select their friends and tend to associate with the participants having high predictive ability. Both of the two cases are theoretically plausible and need to be empirically distinguished. Failure to account for the second case might lead to an overestimation of the effect of social connectedness.

Researchers in the existing empirical literature have addressed this econometric challenge using different strategies. One approach was the use of natural experiment [38]. Sacerdote [31] studied peer effects among college roommates in a natural experiment: Freshmen entering Dartmouth College were randomly assigned to dorms and to roommates. A second approach relied on the panel nature of the data to control the unobserved charateristics. Sorensen [32] examined the effect of social learning on University of California employees' choices of health plans using a rich panel data set. After controlling for the department-specific unobservables, the estimated social effects were smaller but remained significant. A third approach was the use of exogenous instrument variables. Gaviria and Raphael [19] corrected the spurious estimates of school-based peer effects by instrumenting for peer behavior using the average behavior of the peers' parents. Our method belongs to a fourth approach: a randomized laboratory-controlled experiment. In our present experiment, participants are randomly assigned to different network positions in prediction markets.

The present study is also closely related to the literature on prediction markets. Researchers in previous studies have focused on how to elicit dispersed private information, for example, by using some variation of scoring rules. Scoring rules do not suffer from the irrational participation or thin market problems that plague standard prediction markets. They instead suffer from a thick market problem, namely how to produce a single consensus estimate when different people give differing estimates. Hanson [23] suggested a new mechanism for prediction markets, the market scoring rule, which combines the advantages of markets and scoring rules. The market scoring rule avoids the problems by being automated market makers in the thick market case and simple scoring rules in the thin market case. Fang et al. [15] proposed a proper scoring rule that elicits agents' private information, as well as the precision of the information. In their work, the agents' private signals are independent. In our present social-network-embedded prediction market, the information that participants have is correlated with that of their friends.

The present work is also related to the work on network games by Galeotti et al. [18], who provided a framework to analyze strategic interactions in an incomplete information network game. Golub and Jackson [21] discussed how network structure influences the spread of information and the wisdom of the crowds.

A handful of research has examined the mechanisms of prediction markets using laboratory experiments. Healy et al. [24] found that the performance of the prediction market mechanisms is significantly affected by the complexity of the environment. Jian and Sami [27] compared two commonly used mechanisms of prediction markets: the probability-report mechanism and the security-trading mechanism. A great deal of attention has also been paid to the experimental work that considers the effect of exogenously specified network structures on outcomes [9]. Hinz and Spann [25] examined the effects of different network structures on bidding behaviors in name-your-own-price auctions. Bapna et al. [3] studied the effect of the strength of social ties on Facebook using a field experiment. To the best of our knowledge, our paper is the first to study the effect of network structure on individual behavior and on forecasting performance in prediction markets using a laboratory controlled experiment, thus enriching the literature by identifying the causal effect of the social network on prediction markets.

3. A Simple Model of a Social-Network-Embedded Prediction Market

3.1 Model Setup

In this section, we set up a simple model of a social-network-embedded prediction market, which both captures the key features of the experiment and serves as the benchmark for the hypotheses we test in the experiment. Table 1 summarizes the notations used for our model.

Table 1. Summary of Notations

V	The random variable that the principal wants to forecast
n	The number of participants in the prediction market
V_0	The prior mean of V
ρ_V	The prior precision of V
k_i	The number of Participant i's friends
S_i	Participant i's private signal
ε_i	The signal's error
ρ_ε	The precision of participants' signal error
x_i	The prediction reported by Participant i
m_i	Whether Participant i acquires information
σ_i	Participant i's mixed strategy of information acquisition
c	The cost of information acquisition

A principal wants to forecast the realization of a random variable V. In reality, V could be movie box office revenue, future demand for electricity, or election outcomes. The principal resorts to n participants to obtain an accurate prediction. For ease of exposition, we refer to the principal as "he" and each participant as "she." Before receiving any private information, the principal and the participants share a common prior on the distribution of V, given by:

$$V \sim N(V_0, 1/\rho_V), \tag{1}$$

where V_0 is the mean of the normal distribution, and ρ_V is the precision of the prior.

Participants in the prediction market are linked to each other according to a social network, and information is transmitted over the network. The social network $\Gamma = (N, L)$ is given by a finite set of nodes $N = \{1,2,\ldots,n\}$ and a set of links $L \subseteq N \times N$. Each node represents a participant in a prediction market. The social connections between the participants are described by an $n \times n$ dimensional matrix denoted by $g \in \{0,1\}^{n\times n}$ such that:

$$g_{ij} = \begin{cases} 1, & \text{if } (i,j) \in L \\ 0, & \text{otherwise} \end{cases}.$$

Let $N_i(g) = \{j \in N: g_{ij} = 1\}$ represent the set of friends of Participant i. The degree of Participant i is the number of Participant i's friends: $k_i(g) = \#N_i(g)$. The principal does not know the social network graph. For simplicity, we assume the network is undirected, but the results also hold for directed networks.

Each participant is risk neutral and can access a private independent information source at a cost c. m_i is a binary variable indicating whether Participant i acquires information. Participants exchange information over the social network: For simplicity, we assume that they can observe their direct friends' information, but not their second-order friends' (friend's friend) information. More precisely, if Participant i acquires information from her private source ($m_i = 1$), she observes a conditionally independent private signal and passes it to her friends:

$$S_i = V + \varepsilon_i, \varepsilon_i \sim N(0,1/\rho_\varepsilon), \tag{2}$$

where ρ_ε is the precision of Participant i's information source for $i = 1,2,\dots,n.$ The signals' errors $\varepsilon_1,\dots,\varepsilon_n$ are independent across participants and are also independent of V. We assume that the precision of all participants' information sources is equal, which implies that no one is especially well informed, and that the valuable information is not concentrated in a very few hands.

The principal designs a quadratic loss function to elicit the private information of participants. A participant's payoff function is given by:

$$w(m_i, x_i, V) = a - b(x_i - V)^2 - m_i c, \tag{3}$$

where x_i is the prediction reported by Participant i, and $b(x_i - V)^2$ is a quadratic penalty term for mistakes in the forecast. Notice that the optimal report for Participant i is $x_i^* = E[V|I_i]$, where I_i is the information set of Participant i, which includes both the information she acquires and the information passed to her from the social network Γ. We can also use other strictly proper scoring rules (see [15]). The qualitative results remain unchanged.

A participant follows a two-step decision procedure. In the first stage, all of the participants decide whether to acquire information simultaneously. In the second stage, a participant makes use of her signal, as well as of the signals of her friends, to report her best prediction.

We first focus on the optimization problem in the second stage. In the second stage, Participant i's best prediction x_i^* depends on whether Participant i and her friends acquire information; thus, x_i^* is a function of m_i and $m_{N_i(g)}$, where $m_{N_i(g)} \in \{0,1\}^{k_i}$ is the action profile of Participant i's friends, and it represents whether Participant i's friends acquire information.

If Participant i acquires information ($m_i = 1$), she forms her private belief from the private signal S_i, as well as from information she obtains from her neighbors, and her payoff is:

$$a - b\left[x_i^*\left(m_i = 1, m_{N_i(g)}\right) - V\right]^2 - c.$$

If the participant has decided not to acquire information ($m_i = 0$), she forms the belief only from her neighbors' signals, and her payoff is:

$$a - b\left[x_i^*\left(m_i = 0, m_{N_i(g)}\right) - V\right]^2.$$

3.2 Equilibrium Results

Given the action profile of her friends, Participant i's utility is given by

$$u\left(m_i, m_{N_i(g)}\right) = E_V\left[a - b\left[x_i^*\left(m_i, m_{N_i(g)}\right) - V\right]^2 - m_i c\right], \tag{4}$$

where E_V is the expectation with respect to V. The utility $u\left(m_i, m_{N_i(g)}\right)$ depends on whether Participant i and her neighbors acquire information.

Following [17], we assume that each participant observes her own degree k_i, which defines her type, but does not observe the degree or connections of any other participant in the network. For example, people who graduated from the same MBA program might have a good sense of their classmates after graduation, but they do not know who the friends of these classmates are. Another example is that people only pay attention to a subset of their friends in the Facebook and Twitter network, given their limited cognitive resources. They don't know to whom their friends pay attention.

Each participant's belief about the degree of her friends is given by:

$$\Pi(\cdot\,|k_i) \in \Delta\{1, \ldots, k_{max}\}^{k_i},$$

where k_{max} is the maximal possible degree, and $\Delta\{1, \ldots, k_{max}\}^{k_i}$ is the set of probability distribution on $\{1, \ldots, k_{max}\}^{k_i}$. For simplicity, we make an assumption that neighbors' degrees are all stochastically independent, which means that Participant i's degree is independent from the degree of one of her randomly selected friends. This assumption is true for many random networks, such as the Erdös-Rényi random graph [14].

A strategy of Participant i is a measurable function $\sigma_i\colon\{1\ldots, k_{max}\} \to \Delta\{0,1\}$, where $\Delta\{0,1\}$ is the set of probability distributions on $\{0,1\}$. This strategy simply says that a participant observes her degree k_i, and on the basis of this information she decides whether to acquire information. Notice that $\Delta\{0,1\}$ means that the participant adopts a mixed strategy: She randomizes her actions with some probabilities in $m_i = 1$ and in $m_i = 0$. The strategy profile of Participant i's friends is denoted by $\sigma_{N_i(g)}$.

We focus on symmetric Bayes-Nash equilibria, where all participants follow the same strategy σ. A Bayes-Nash equilibrium is a strategy profile, such that each participant with degree k_i chooses a best response to the strategy profile of her friends. Let $\phi\big(m_{N_i(g)}, \sigma, k_i\big)$ be the probability distribution over $m_{N_i(g)}$ induced by $\Pi(\cdot\,|k_i)$. The expected payoff of Participant i with degree k_i and action m_i is equal to:

$$U\big(m_i, \sigma_{N_i(g)}; k_i\big) = E_{m_{N_i(g)}} u\big(m_i, m_{N_i(g)}\big) = \sum\nolimits_{m_{N_i(g)}} \phi\big(m_{N_i(g)}, \sigma, k_i\big) u\big(m_i, m_{N_i(g)}\big), \tag{5}$$

where $E_{m_{N_i(g)}}$ is the expectation with respect to $m_{N_i(g)}$. We say that Participant i's strategy σ_i is non-increasing if $\sigma_i(k_i)$ dominates $\sigma(k_i')$ in the sense of first-order stochastic dominance (FOSD) for each $k_i' > k_i$. In other words, if the strategy σ_i is non-increasing, high-degree participants randomize their actions with less probability in $m_i = 1$ and thus are less likely to acquire information.

Proposition 1 gives us the basic result of the Bayes-Nash equilibrium.

Proposition 1. *There exists a symmetric Bayes-Nash equilibrium that is non-increasing in degree. There exists some threshold* $k^* \in \{0,1,2,\ldots\}$, *such that the probability of choosing to acquire information satisfies:*

$$\sigma(m_i = 1|k_i) = \begin{cases} 1, & for\ k_i < k^* \\ 0, & for\ k_i > k^*. \\ (0,1], & for\ k_i = k^* \end{cases}$$

Furthermore, the expected payoffs are non-decreasing in degree.

The proof is included in the appendix. Proposition 1 has very clear implications. The participant's equilibrium action is weakly decreasing in her degree. In other words, the more friends she has, the less willing she is to acquire information. Participants can free ride on the actions of their friends. If Participant i has more friends, she is more likely to benefit from the signals passed around by her friends. It should also be emphasized that participants who have more friends earn higher payoffs under the appropriate monotone equilibrium because of the positive externalities. Here, higher degree participants exert lower efforts but earn a higher payoff than do their less connected peers. The non-increasing property of equilibrium actions implies that social connections create personal advantage. In the network game with positive externalities, well-connected participants earn more than poorly connected participants. Note that the threshold degree k^* is a function of parameters such as c, ρ_ε, and ρ_V.

After making the decision on information acquisition, each participant reports the best point estimation. The purpose of prediction markets is to generate fairly accurate predictions of future events by aggregating the private information of a large population. How does the principal aggregate these small bits and pieces of relevant information that exist in the opinions and intuitions of diverse individuals? We assume that the principal adopts a simple averaging rule, and his prediction is $\frac{1}{n}\sum_{i=1}^{n} x_i^*$. Note that the simple averaging rule is optimal only when all the participants' forecasts are independent and equally accurate; however, it is a good operational rule for limited information (e.g., [2]). In our networked prediction markets, the principal has limited information: He does not know the social network graph. In this case, the principal cannot propose a weighted averaging rule and simply follows the operational rule of thumb: "Use equal weights unless you have strong evidence to support unequal weighting of forecasts" ([2], p. 422).

4. An Experimental Analysis on Network Structure and Forecasting Performance

In this section, we compare the performance of non-networked prediction markets (NNPM) with the performance of social-network-embedded prediction markets (SEPM) using controlled laboratory experiments. We also take into account the network structure in which the participants are embedded. Our experiment demonstrates that network structure has a significant effect on the individual’s behavior of information acquisition and the prediction market performance.[2] Eighty undergraduate students were recruited as subjects from a large university,

[2] An important problem in network games is the existence of multiple equilibria. One way to reduce the equilibrium multiplicity is to introduce incomplete information about network structure [17]. Another way is to use an experimental examination. In this section, we show that the monotonic equilibrium described in Proposition 1 is consistent with our experimental result.

and they had no previous experience in prediction market experiments. There were four experimental sessions, each consisting of five groups. We restricted our attention to the case of four-person networks, so each group consisted of four randomly assigned participants. The average earnings were $8.50 per person, including a $1.50 show-up fee, for a 40-minute session.

4.1 Experimental Design

Similar to the setup of the theoretical model, participants were asked to predict a random variable V during the experiment. The common prior is given by equation (1) in our model setup, and in the experiment we set $V_0 = 10$ and $\rho_V = 0.5$. Each participant could receive a private signal S_i at a cost c. The signal S_i is given by equation (2), and $\rho_\varepsilon = 1$.

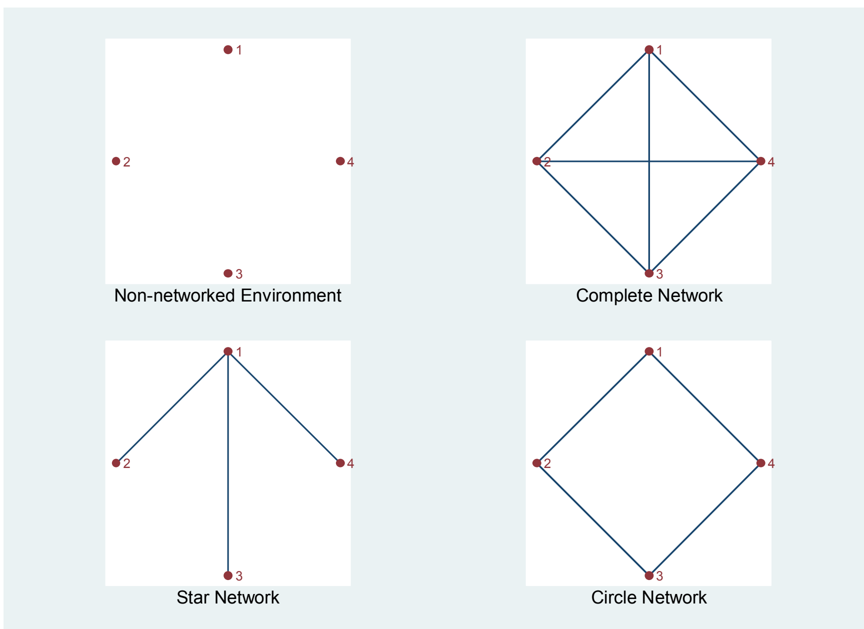


Figure 1. Network Structures

The experiment had a 4×2 design: four different treatments of network structures $\times$ two levels of information acquisition cost. The four treatments of network structures include: 1. the baseline treatment, non-networked environment; 2. complete network; 3. star network; and 4. circle network. They are illustrated in Figure 1. In all of the treatments, subjects participated in the experiment via the computer system we developed. Throughout the experiment, the subjects were not allowed to communicate in person and could not see others' screens. The only communication channel available to them was to chat via designated Gmail accounts. In the baseline treatment, $N_i(g) = \emptyset$, and each participant was isolated. In a complete network, Participant i was connected to three other participants (we call them Participant i's "friends"). As Figure 2 shows, the participant "crecaustin02" could chat with "crecaustin", "crecaustin03", and "crecaustin04" through Gmail. Similarly, the communication networks are given by a star network and a circle network in Treatments 3 and 4, respectively.

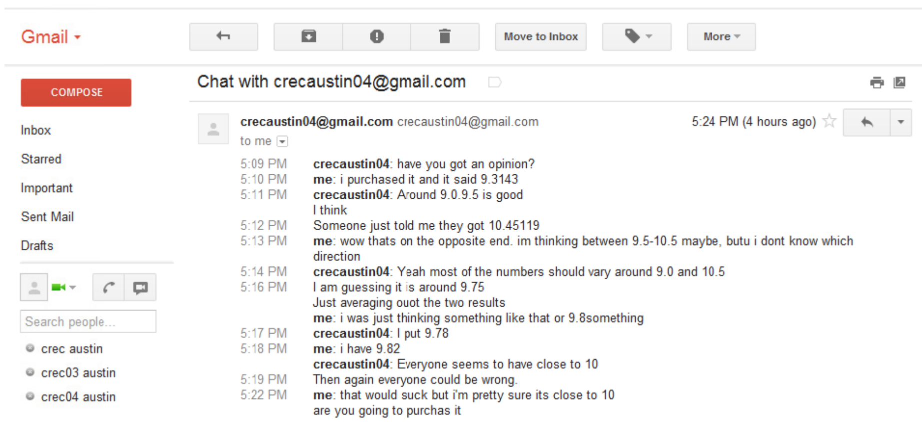


Figure 2. A Screenshot of the Communication over a Complete Network

Each treatment was conducted in an experimental session with two independent decision-rounds. In Round 1, the cost of information acquisition was \$0.50. In Round 2, the cost of information acquisition was \$1. The order of the acquisition cost was always low and then high. To minimize the effect of reputation, each round started with the randomly formed groups (four-person networks). Participants in each round followed a two-stage decision process: information acquisition and prediction. Figure 3 depicts a flow chart of experiment round t, $t = 1$, 2 (the only difference between the two rounds is the cost of information acquisition). In the stage of information acquisition, participants made their decisions about whether to purchase a signal from an outside expert. If they paid the cost of information acquisition *c*, they would receive a private signal. Once all the decisions of information acquisition were made, participants could communicate over the given network under each treatment (After the experiment, we checked the participants' chat history and found that no one misreported the private signal to others). In this stage, every participant chatted with every neighbor at the same time. It means that focal participants might receive information from their second-order friends.

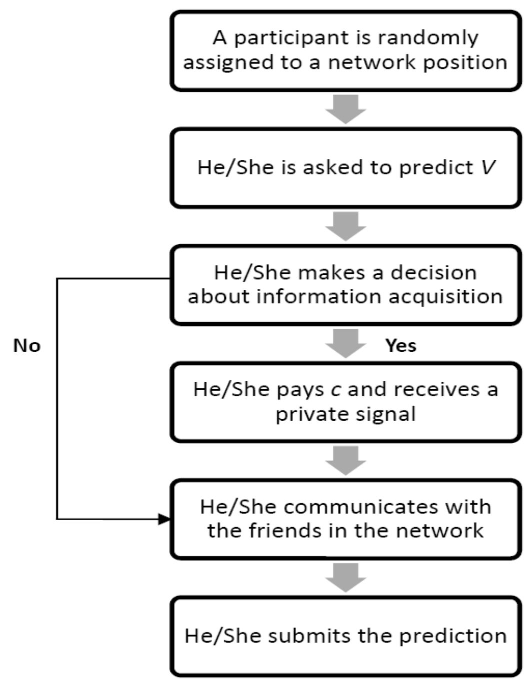


Figure 3. The Flow Chart of the Experiment Round *t*, *t* = 1, 2

After checking the chat history, we found that participants received a substantial amount of information from their first-order friends but less information from their second-order friends. On average, a participant received information from 61.88% of her first-order friends and 4.38% of her second-order friends. This result implies that most participants were willing to exchange their private signals with others but were less willing to tell others the information they got from someone else. Specifically, we find that central participants in a star network only exchanged their own signals with peripheral participants. This information diffusion pattern is consistent with the exchange theory that explains the reciprocity based on the idea of socially embedded behavior [26]. Peripheral participants had no other information channels, except for their own private signals and the information from the hub. Thus, central participants exchanged only their private signals, excluding information from others with peripheral nodes according to reciprocity and norms of fairness [11].

After the experiment, the computer system calculated the total payoff of each participant according to the payoff function (3). We set $a = 5$ and $b = 1$. Therefore, the maximum payoff for each round was \$5. We are interested in testing the following four hypotheses. Hypotheses 1 and 2 are motivated by Proposition 1 in the analytical model.

Hypothesis 1. *Each individual's information acquisition is non-increasing in the participant's degree.*

Hypothesis 2. *The participant's earnings are non-decreasing in the degree.*

Hypothesis 3 is motivated by the following arguments: Even if participants are isolated in a non-networked environment, individual estimates are no longer independent because of the common prior (public information). The existence of a social network facilitates the dissemination of private information among participants, which effectively puts more weights on private information when participants' predictions are aggregated in the prediction market. Such adjustment is beneficial to the forecasting accuracy because it corrects to a certain extent a possible bias toward the common prior. Social networks need not undermine the wisdom of crowd effect, especially when people share a common prior. On the other hand, when the cost of information acquisition is very high, the existence of a social network can impede information acquisition by the community as a whole because of possible free-riding opportunities, thus lowering the forecasting accuracy of the prediction market.

Hypothesis 3. *An SEPM outperforms an NNPM when the cost of information acquisition is low.*

Hypothesis 4 is motivated by a large body of literature on the identification and estimation of peer effects [1]. Peer effects are economically important because they are present in many decision domains, such as students' academic performance [31], mutual fund managers' portfolio choices [12], and health plan choices [32]. In our experiment, we tested whether the prediction performance of a participant is influenced by the members of the group to which they belong.

Hypothesis 4. *Peer effects exist in the prediction accuracy among participants.*

In a star network, the central participant has an above-average influence. Hypothesis 5 is motivated by the following arguments: If the hub has a relatively wrong estimate, the above-average influence exacerbates the problem and hurts the prediction market performance significantly.

Hypothesis 5. *In a star network, the prediction market performance is positively correlated with the performance of the central participant.*

4.2 Summary Statistics

Table 1 summarizes the statistics of participants' predictions under different network structures. We perform two variance-comparison tests: the Variance ratio test (F test) and the Brown–Forsythe test. Note that the F test relies on the assumption that the samples come from normal distributions, and the Brown–Forsythe test [7] provides robustness against many types of non-normal data while retaining good power. By calculating the standard deviations of predictions under different treatments, we find that the standard deviation under a complete network is significantly lower than the standard deviation under a non-networked environment (1.237 vs. 2.186, F test: $p < 0.01$; Brown–Forsythe test: $p < 0.05$), which suggests that communications lead to greater consensus about the true value.

Table 2. Descriptive Statistics of the Participants' Predictions

	Mean	The Std. Dev.	Obs
Non-Networked Environment	9.853364	2.186319	40
Complete Network	9.893607	1.237169	40
Star Network	9.916362	2.025045	40
Circle Network	9.616113	1.532874	40

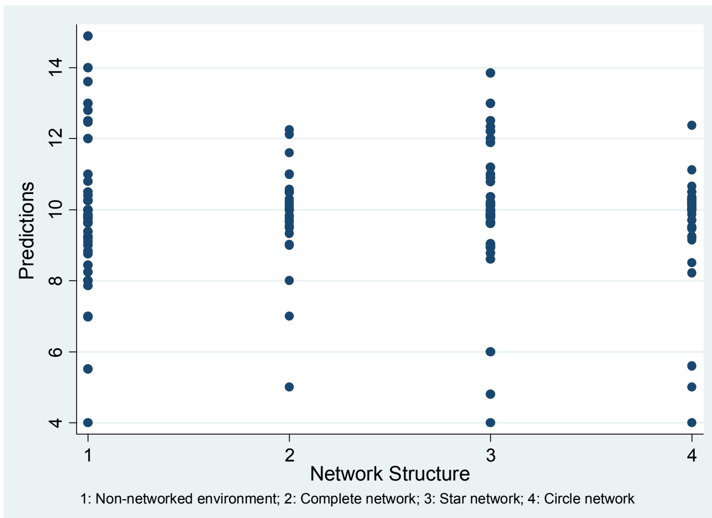


Figure 4. Predictions under Different Network Structures

As Table 2 and Figure 4 show, the variation of the prediction also depends on the network structure. The standard deviation of the predictions under a star network is significantly higher than the standard deviation under a circle network (2.025 vs. 1.533, *F* test: $p < 0.05$; Brown–Forsythe test: $p < 0.05$), and the standard deviation of the predictions under a circle network is significantly higher than the standard deviation under a complete network (1.533 vs. 1.234, *F* test, $p < 0.10$; Brown–Forsythe test: $p < 0.10$). The denser the network is, the lower the standard deviation of the predictions (the density of the network: complete > circle > star > non-networked). To address the problem that the underlying observation may not be independent, we also compute the average prediction in each four-person group (essentially removing the within group correlation) and then test the standard deviation under different network structures. The result is robust. The intuition is that participants communicate with each other more effectively in a denser network. Thus, information exchange reduces the variance of the predictions.

4.3 Experimental Results: Testing of H1

Do participants play an equilibrium strategy of information acquisition in social networks? To test this hypothesis, we first compute the mean of information acquisition (if Participant i acquires information, $m_i = 1$; otherwise $m_i = 0$) when participants' degree varies. Figure 5 shows that the equilibrium strategy of information acquisition is decreasing in the number of connections.

We then run a logistic regression of participants' information acquisition decision on their degree and the cost of information acquisition:

$$\text{logit}\, E(acquisition_i | degree_i, cost_i, sdummy_i) = \beta_0 + \beta_1 degree_i + \beta_2 cost_i + \beta_3 sdummy_i, \quad (6)$$

where, *sdummy*, a dummy variable included for a robustness check, indicates whether the participant having three connections is in a star network (because such participants can also be in a complete network). We find that participants' information acquisition behavior is indeed consistent with the equilibrium strategy predicted by the analytical model: A larger number of connections leads to a lower probability of information acquisition. The result is shown in Column 1 of Table 3. We find that the probability of participants' acquiring information decreases with the degree and the cost of information acquisition. Roughly speaking, the logit estimates should be divided by four to compare them with the linear probability model estimates [36]. For example, Column 1 of Table 2 shows that adding a degree can reduce the probability of information acquisition by 7.6%.

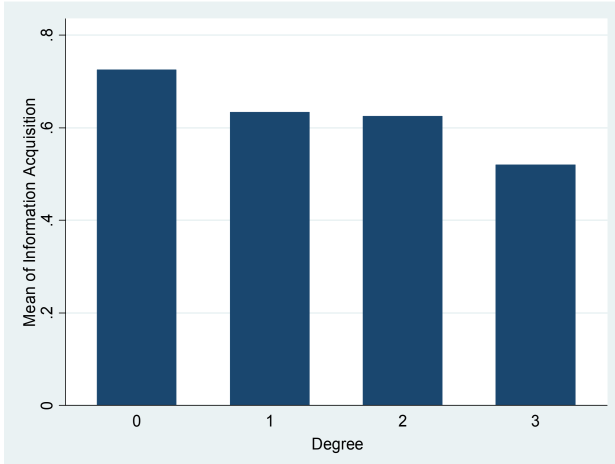


Figure 5. Information Acquisition and Participants' Degrees

Column 2 suggests that the result is also robust to the inclusion of *sdummy* (Note that the P value for the coefficient of *degree* is 0.064, which is very close to a 5% significance level). Small sample size is a common problem for the experimental method. The validity of z-statistics depends on the asymptotic distribution of large samples. When the sample size is insufficient for straightforward statistical inference, bootstrapping is useful for estimating the distribution of a statistic without using asymptotic theory. In Column 3, we use bootstrapping to compute the standard errors and find that the result is robust (we draw a sample of 160 observations with replacement, and repeat this process 10,000 times to compute the bootstrapped standard errors). To account for the possible unobserved heterogeneity of participants, we control for the subjects' latent characteristics using a random effects model in Column 4, and the result is robust.

Table 3. Logistic Regression Analysis of Information Acquisition Using Model (6)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Logit	Logit	Bootstrapping	Random effects	Cluster effects	Cluster Bootstrap	Robust Std. Err.
degree	-0.304**	-0.293*	-0.304**	-0.296**	-0.293***	-0.293***	-0.293**
	[-2.010]	[-1.855]	[-2.017]	[-2.201]	[-6.692]	[-3.361]	[-2.092]
cost	-1.302***	-1.302***	-1.302***	-1.317***	-1.302**	-1.302***	-1.302**
	[-3.706]	[-3.706]	[-3.567]	[-3.352]	[-2.492]	[-4.162]	[-3.712]
sdummy		-0.154		-0.156	-0.154	-0.154	-0.154
		[-0.213]		[-0.211]	[-0.313]	[-0.890]	[-0.213]

Constant	1.693***	1.686***	1.693***	1.704***	1.686***	1.686***	1.686***
	[4.403]	[4.374]	[4.393]	[3.882]	[4.191]	[6.333]	[4.560]
Observations	160	160	160	160	160	160	160
R-squared	0.186	0.187	0.186	0.111	0.187	0.187	0.187

z or t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

As shown in Table 2, the standard deviations are not the same under different network structures. Potential problems arise with statistical inference in the presence of clustering effects. Default standard errors that ignore clustering can greatly understate true standard errors [8]. Wooldridge [37] provided an econometric approach to analyzing cluster sample. Following his approach, we compute the variance matrices that are robust to arbitrary cluster correlation and unknown heteroskedasticity.[3] In our context, the observations are clustered into different network topologies. Standard errors are adjusted for clusters in Column 5, and the result is similar. A practical limitation of inference with cluster-robust standard errors is the assumption that the number of clusters is large. Cameron, Gelbach, and Miller [8] show that cluster bootstraps can lead to considerable improved inference when there are few clusters. Column 6 shows that the results of the cluster bootstrap are robust. Because of the strong suspicion of heteroskedasticity, we also compute the heteroskedasticity-robust *t* statistics using the Huber-White sandwich estimators in Column 6 to check the robustness of our results. The robust *t* statistics can deal with the concerns about the failure to meet standard regression assumptions, such as heteroskedasticity [36]. Our results are robust to the case when the modeling errors depend on the explanatory variables, such as *degree* and *sdummy*. Note that different network topologies can be linearly predicted from the variables *degree* and *sdummy* (dummy variables indicating network structures are redundant when we have the two explanatory variables, *degree* and *sdummy*, so adding additional dummy variables indicating network structures causes the

[3] Wooldridge's example [36] is to estimate the salary-benefits tradeoff for elementary school teachers in Michigan. Clusters are school districts. Units are schools within a district.

problem of multicollinearity). Thus, the results in Column 7 are also robust to the case when the modeling errors depend on different network topologies.

4.4 Experimental Results: Testing of H2

We also examine the effect of social connections on individuals' earnings. Figure 6 shows that the mean of earnings for each round is increasing in the number of connections.

Next, we run an ordinary least squares (OLS) regression of earnings on the degree and the cost of information acquisition:

$$earnings_i = \beta_0 + \beta_1 degree_i + \beta_2 cost_i + \beta_3 sdummy_i + \beta_4 acquisition_i + \varepsilon_i. \quad (7)$$

Table 4 shows that the participants' earnings increase with the degree and decrease with the cost of information acquisition. Being excluded from these connections is thus a handicap for a participant. The basic result remains unchanged when we add a star network dummy variable, *sdummy*, or a dummy variable, *acquisition*, indicating whether a participant acquires information. Column 4 shows that the result is robust when we use the method of bootstrapping. The result of a random effects model is similar and reported in Column 5. In Column 6, we account for clustering in data. Column 7 reports the results of the cluster bootstrap. We also run a robust regression and compute the robust *t* statistics in Column 8. Our estimators are shown to be robust to various kinds of misspecification. The experimental results in Tables 3 and 4 thus support Hypotheses 1 and 2. Because of the randomization of the network position assignments, our experimental results do not suffer from the identification problem related to the endogenous network structure and reveal causality rather than mere correlation.

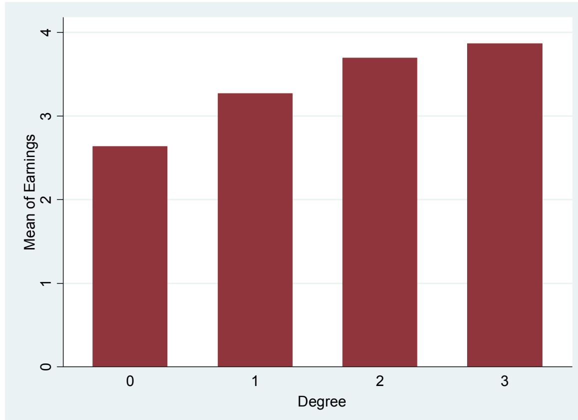


Figure 6. Earnings and Participants' Degrees

Table 4. OLS Regression Analysis of the Participants' Earnings Using Model (7)

VARIABLES	(1) OLS	(2) OLS	(3) OLS	(4) Bootstrapping	(5) Random effects	(6) Cluster effects	(7) Cluster Bootstrap	(8) Robust Std. Err.
degree	0.214**	0.228**	0.179**	0.214**	0.228**	0.228**	0.228**	0.228**
	[2.119]	[2.142]	[2.060]	[2.131]	[2.012]	[2.691]	[2.101]	[2.230]
cost	-0.774***	-0.774***	-1.001***	-0.774***	-0.786***	-0.774**	-0.774***	-0.774***
	[-3.286]	[-3.277]	[-4.161]	[-3.303]	[-3.482]	[-2.232]	[-4.460]	[-3.282]
sdummy		-0.217	-0.246		-0.217	-0.217	-0.217	-0.217
		[-0.424]	[-0.494]		[-0.404]	[-1.224]	[-0.792]	[-0.353]
acquisition			-0.789***					
			[-3.142]					
Constant	3.560***	3.552***	4.234***	3.560***	3.558***	3.552***	3.552***	3.552***
	[15.22]	[15.09]	[13.42]	[16.31]	[14.52]	[47.10]	[19.44]	[16.14]
Observations	160	160	160	160	160	160	160	160
R-squared	0.287	0.287	0.343	0.287	0.187	0.287	0.287	0.287

z or t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

4.5 Experimental Results: Testing of H3

Hypothesis 3 predicts that when the cost of information acquisition is low, an SEPM outperforms an NNPM. In our experiment, each group is a prediction market, so we have 20 prediction markets in total. The performance of a prediction market *g* is measured by forecast accuracy:

$$MAccuracy_g = 1 - Absolute\ Percentage\ Error_g = 1 - \frac{|F_g - V_g|}{V_g},$$

where F_g is the forecast of prediction market *g*, calculated as the average of all four participants' predictions (the principal's prediction) in that market, and V_g is the realization of the random variable in market *g*. To test Hypothesis 3, we perform t-tests and Monte Carlo permutation tests with 10,000 permutations. A t-test relies heavily on the asymptotic distributional assumption and may not perform well when the sample size is small. A Monte Carlo permutation test gives a non-parametric way to compute the sampling distribution because no assumption on the sampling distribution is required (for another example, see Jian and Sami [27] who also compared the performance of different prediction markets using a permutation test).

We find that when the cost of information acquisition is low ($0.50), a complete networked prediction market significantly outperforms an NNPM (*t* statistics: $p = 0.04$; permutation test: $p = 0.02$). When the cost of information acquisition is high ($1), the performance difference is not significant (*t* statistics: $p = 0.42$; permutation test: $p = 0.47$). Therefore, the superior forecasting performance of a networked prediction market decreases with the cost of information acquisition. The relative performance of a networked prediction market to a non-networked prediction market depends on the cost of information acquisition.

4.6 Experimental Results: Testing of H4 and H5

Hypothesis 4 states that participants' prediction accuracy can be affected by the accuracy of other participants in their network. There are several challenges in identifying the peer effects [29]. First, network formation could be endogenous: Individuals self-select their friends. For example, many social networks exhibit homophily: People are more prone to make friends with those who are similar to themselves. This makes it difficult to disentangle the selection effect and the real peer effects. This challenge is similar to the identification problem in estimating the effects of network structure. In our experiment, the “friends” of a participant were randomly assigned. Random assignment implies that a participant’s background characteristics, such as predictive ability, are uncorrelated with their friends’ background characteristics. This approach allows us to take care of the first challenge.

Second, Participants *i* and *j* can affect each other simultaneously. This reflection problem [29] causes a difficulty in identifying the actual causal effect if we adopt a linear-in-means specification: Participant *i*’s prediction performance is a linear function of the average performance level of his or her friends.

The reflection problem can be overcome by introducing nonlinearities into social interactions [26]. The prediction accuracy of Participant *i* is influenced by the maximal accuracy of her friends:

$$Accuracy_i = \beta_0 + \beta_1 max_{j \in N_i} Accuracy_j + \beta_2 \Omega_\mathrm{i} + \varepsilon_i. \tag{8}$$

$N_i(g) = \{j \in N: g_{ij} = 1\}$ is the set of friends of Participant i, Ω_i represents the control variables, and the prediction accuracy of Participant *i* is given by:

$$Accuracy_i = 1 - Absolute\ Percentage\ Error_i = 1 - \frac{|x_i - V|}{V},$$

where x_i is the prediction of Participant *i*, and V is the realization of the random variable in the corresponding prediction market. In our experiment, this specification is reasonable because participants with high predictive ability share their “forecasting formula.” The performance of a participant directly depends on whether she has a clever friend. For example, as shown in Figure 7, a clever participant proposed a useful average rule. As a result, a participant’s prediction is influenced by her friends with the best forecasting performance.

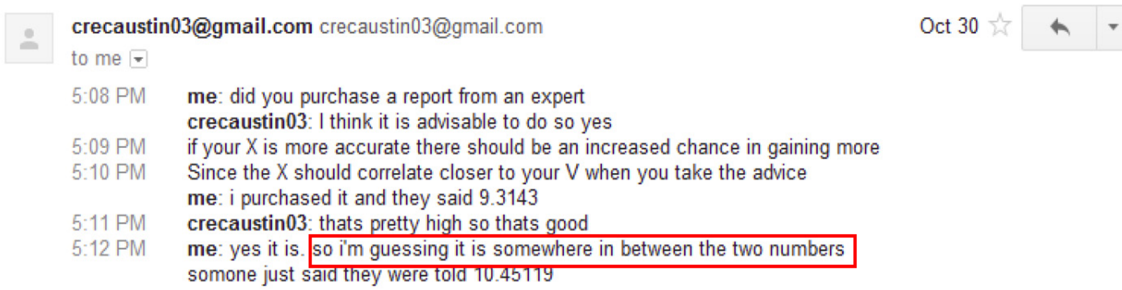


Figure 7. A Screenshot of Chats between Two Participants in the Experiment

Table 5 presents the regression results. Again, we control for participants' degree, the cost of information acquisition, and the dummy variable, *acquisition*. The variable, *social influence*, represents the maximal accuracy of Participant *i*'s friends. Our interest is the coefficient on social influence, β_1, and we find that the coefficient is significantly positive in Column 1. This result is also robust to a different model specification in Column 2 and the use of bootstrapping. The coefficient implies that a 1% increase in the maximal accuracy of the friends of a focal player is associated with a roughly 0.5% increase in the focal player's prediction accuracy. This coefficient is moderate in size and seems plausible.

Table 5. Estimation of Peer Effects using the Regression Model in (8)

VARIABLES	(1) OLS	(2) OLS	(3) Bootstrapping
social influence	0.460***	0.492***	0.460***
	[6.299]	[6.391]	[2.606]
acquisition	-0.0147	-0.0262	-0.0147
	[-0.362]	[-0.632]	[-0.374]
cost		-0.0387	
		[-1.474]	
degree		-0.2461	
		[-0.601]	
Constant	0.483***	0.483***	0.483***
	[6.874]	[6.874]	[2.853]
Observations	120	120	120
R-squared	0.255	0.271	0.255

z or t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

Hypothesis 5 states that peripheral nodes are influenced by the central participant in a star network, so the prediction market performance is positively associated with the prediction performance of the central participant. To test this hypothesis, we run an OLS regression of the prediction market accuracy of a star network in market *g* ($MAccuracy_g$) on the prediction accuracy of the central participant ($Accuracy_g$), the cost of information acquisition, and the number of participants acquiring information in market *g* (*signal*):

$$MAccuracy_g = \beta_0 + \beta_1 Accuracy_g + \beta_2 cost + \beta_3 signal + \varepsilon_i. \tag{9}$$

Table 6 shows the regression results. In Column 1, we find that a 1% decrease in the prediction accuracy of the central node is associated with a 0.534% decrease in the prediction market accuracy. When a hub has a relatively wrong estimate, it will cause a serious problem in a star networked prediction market. Column 2 shows that the result is robust after we control for the information acquisition in the market. To address the small sample concern, we do bootstrapping in Column 3, and the positive correlation is still significant.

Table 6. Estimation of the Hub Effects using the Regression Model in (9)

VARIABLES	(1) OLS	(2) OLS	(3) Bootstrapping
Accuracy	0.534***	0.520***	0.534**
	[7.380]	[6.721]	[2.061]
cost	-0.0828	-0.0775	-0.0828
	[-1.621]	[-1.450]	[-1.282]
signal		-0.0316	
		[-0.733]	
Constant	0.478***	0.559***	0.478*
	[6.882]	[4.242]	[1.893]
Observations	10	10	10
R-squared	0.896	0.904	0.896

z or t-statistics in brackets, *** p<0.01, ** p<0.05, * p<0.1

5. Extension

5.1 What Happens with a Complex Social Network?

One shortcoming of the controlled experiment approach is that the network structure is relatively simplistic. A natural question is whether our hypotheses are supported when the underlying social network is more complicated. In particular, from a manager's perspective, seeing how well Hypothesis 3 is supported with a more complex social network is important. This is because if an SEPM always (weakly) dominates an NNPM, then the manager of a prediction market should always promote the use of social networks among the participants. In this section, we conduct numerical simulations based on the analytical model to further analyze the effects of social networks on the forecast accuracy of prediction markets when the social network is more complex. The simulation results complement our findings from the experiment by demonstrating that a social network is actually a double-edged sword in a prediction market: When the cost of information acquisition is low, a social network can promote forecast efficiency, as suggested by our experimental results, but if the cost of information acquisition is high, it could decrease the prediction performance.

Using our analytical model, we conduct a variety of agent-based simulations in the social-network-embedded prediction markets. In every simulation round, a random social network that includes 100 participants is generated, using a 100×100 dimensional matrix. Following the Erdős–Rényi random graph model, we assume that the link between two participants is formed with independent probability p in our simulation. We set the parameter values for the common prior $V \sim \mathrm{N}(V_0, 1/\rho_V) = \mathrm{N}(10, 2)$, and the noise of the signal $\varepsilon_i \sim \mathrm{N}(0, 1/\rho_\varepsilon) = \mathrm{N}(0, 1)$. The results are robust for other parameter values. On the basis of Proposition 1, we can compute the fixed point, the threshold degree k^*, and then further compute the prediction by each participant, which enables us to compute the forecasting accuracy of the prediction market.

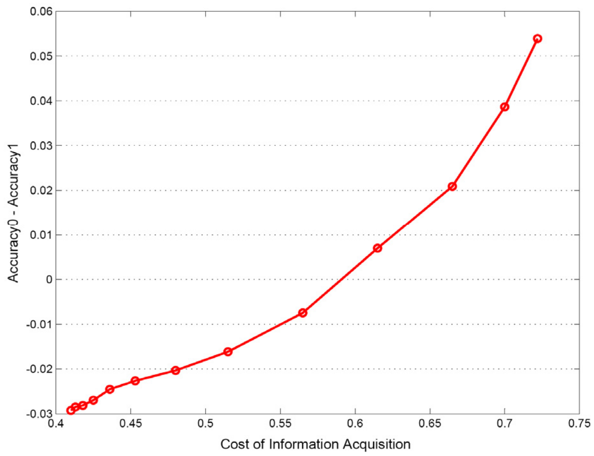
In the simulation, we use two measures of prediction market performance: the forecast accuracy and the mean squared errors (MSE) of the prediction market. Recall that the forecast of prediction market g, F_g, is the simple average of all 100 participants' predictions (the principal's prediction) in that market.

For each cost level of information acquisition, we run 1,000 simulations for both the SEPM and the NNPM, and then we compute the estimated forecast accuracy and the MSE. Figure 6 illustrates the effect of the cost of information acquisition on prediction market performances of the SEPM and the NNPM. The figure is drawn for parameter values n = 100, p = 0.3, V_0 = 10, ρ_V = 0.5, ρ_ϵ = 1, and b = 1. Accuracy0 represents the forecast accuracy computed in the NNPM, and Accuracy1 represents the forecast accuracy in the SEPM. The forecast accuracy is defined as:

$$Accuracy_g = 1 - Mean\ Absolute\ Percentage\ Error_g = 1 - \frac{1}{1000}\sum_{j=1}^{1000}\frac{|F_{gj} - V|}{V}, g = 0,1.$$

Figure 8(a) shows that when the cost of information acquisition is low, the SEPM outperforms the NNPM in terms of forecast accuracy, and when the cost is high, the NNPM outperforms the SEPM. In Figure 8(b), this result is robust to a different measure of prediction market performance: MSE. MSE0 represents the MSE computed in the NNPM, and MSE1 represents the MSE in the SEPM. When c is small, $\mathrm{MSE0} - \mathrm{MSE1} > 0$, which means that the SEPM outperforms the NNPM. As c increases, $\mathrm{MSE0} - \mathrm{MSE1}$ decreases, and when c is large enough, the NNPM performs better than the SEPM.

(a) Forecast Accuracy



(b) MSE

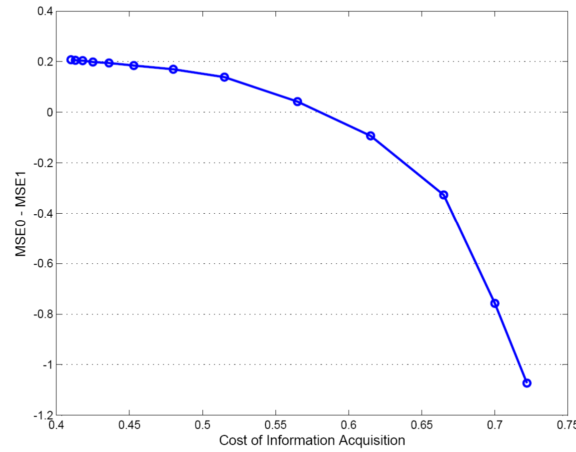


Figure 8. A Comparison between the Performances of the SEPM and the NNPM (Erdős–Rényi random graph)

The Erdős–Rényi random graph may be inappropriate for modeling some real-life phenomena. Typical real-world social networks possess additional structure that is absent in the Erdős–Rényi random graph. For example, the Erdős–Rényi random graph does not exhibit power laws. Using the similar simulation approach, we can also study the prediction performance under more realistic social networks, such as the Preferential Attachment graph [26]. In the Preferential Attachment graph, two participants are more likely to be socially connected if they have a common acquaintance. Note that the Preferential Attachment graph has two parameters: the number of participants n and the total number of edges in the graph e. To compare the Preferential Attachment graph with the Erdős–Rényi random graph we already discussed, we calculate the expected number of edges in the Erdős–Rényi random graph (n = 100, p = 0.3): $[n(n\text{-}1)p]/2 = 1485$ (corresponding to the mean degree 29.7). Thus, we do a robustness check on the Preferential Attachment graph for parameter values n = 100, e = 1485, $V_0 = 10$, $\rho_V = 0.5$, $\rho_\epsilon = 1$, and b = 1. Figure 9 shows that the results are robust.

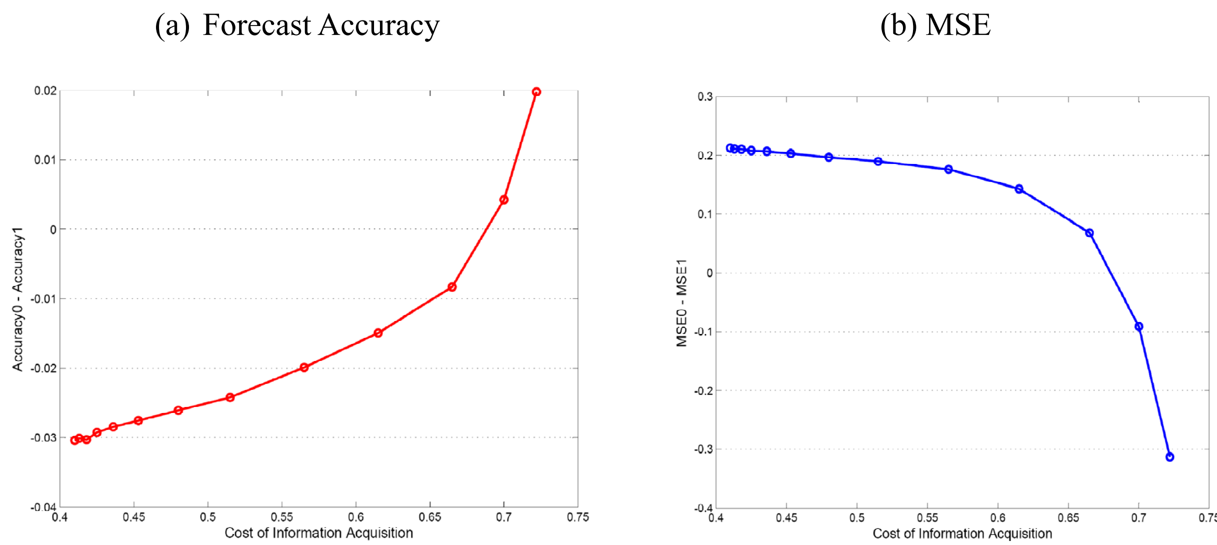


Figure 9. A Comparison between the Performances of the SEPM and the NNPM (Preferential Attachment graph)

This simulation analysis suggests the following results.

Simulation Result: *The performance of an SEPM increases compared to an NNPM with decreasing information acquisition costs.*

There are two implications of this result: First, when the cost of information acquisition is low, a social network can enhance forecast accuracy in prediction markets. Second, a social network also has a negative effect on the forecast accuracy of a prediction market when the cost of information acquisition is high. The second implication is driven by the fact that in our analytical model, social networks could reduce people's incentive to acquire information and could then be detrimental to the forecast accuracy of the prediction market as a whole. Our result depends crucially on the cost of information acquisition. Coval and Moskowitz [13] show that investors prefer to hold local firms rather than distant ones, because the cost of acquiring information about companies located near investor is lower.[4] Similarly, if a prediction market is

[4] The "Home bias puzzle" has been widely studied in the finance literature [13]. Investment managers exhibit a strong preference for domestic equities.

created for forecasting the performance of a firm, participants have easier access to private information and have the lower travel, time, and research costs associated with obtaining private information.[5] If participants in the United States are trying to predict the performance of a Chinese company, the cost of acquiring private information is extremely high.[6]

These implications are critical to understanding how to use social networks to improve the performance of prediction markets. Our present results suggest the following guidance for the business practice of prediction markets: When the predicted event is simple, which is interpreted as a low information acquisition cost, we recommend a social-network-based prediction market. When the predicted event involves complicated issues, which can be interpreted as a high cost of information acquisition, the traditional non-networked prediction market is preferred. For example, it is rather difficult for people to know some information about the event, "Hugo Chavez to no longer be the President of Venezuela before midnight ET 31 Dec 2012" (Intrade Prediction Market). However, it is relatively easy to have some ideas about the *Twilight* movie box office (Iowa Electronic Markets). Whether to use social networks in prediction markets depends on the cost of information acquisition.

5.2 What Happens When the Signals Are Misleading?

In the previous analysis, we assume that the private signals in the market are informative. However, under some circumstances, the signals may be misleading or systematically biased. For example, stocks plunged sharply on April 23, 2013, after a hacker accessed a newswire's account and tweeted about a false White House emergency.[7] The erroneous tweet, which was posted around 1:07 p.m. ET, said "BREAKING: Two Explosions in the White House and Barack

---

[5] Local participants can visit the firm's operations, talk to employees, managers, and suppliers of the firm, and assess the local market conditions in which the firm operates. They may have close personal ties with local executives (e.g., run in the same circles, belong to the same country club).

[6] That is why Muddy Waters Research Group is especially known for its keen eye in spotting fraudulent accounting practices at Chinese companies. See http://blogs.wsj.com/deals/2012/11/28/examining-muddy-waters-track-record/.

[7] See http://buzz.money.cnn.com/2013/04/23/ap-tweet-fake-white-house/?iid=EL.

Obama is injured." This tweet sent shock waves through the stock market and caused the market to tumble.

(a) Small-Bias Signal (b) Large-Bias Signal

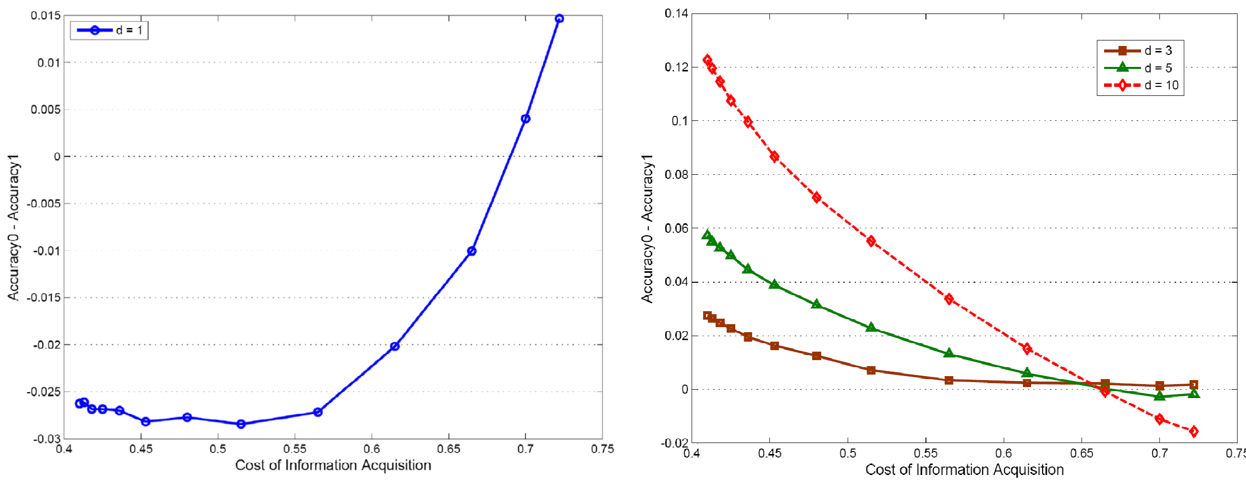


Figure 10. A Comparison between the Performances of the SEPM and the NNPM When the Signals are Systematically Biased

What would happen when the signals in the markets are systematically biased? More formally, we modify equation (2) and assume that the signal is misleading in the sense that $S_i = V + d + \varepsilon_i$, where $d > 0$ or $d < 0$. The absolute value of d measures the systematic bias of the signal. The underlying social network is the Erdős–Rényi random graph. In Figure 10, we redo the simulation analysis when the signal is misleading. Accuracy0 represents the forecast accuracy computed in the NNPM, and Accuracy1 represents the forecast accuracy in the SEPM. In Figure 10(a), we find that when the systematic bias is small (i.e., the absolute value of d is 1), the result is similar: The performance of an SEPM increases compared to an NNPM with decreasing information acquisition costs. However, when the systematic bias is sufficiently large (the absolute value of d is 3, 5, or 10), the result is the opposite (Figure 10(b)). Note that because our problem is symmetric, the simulation results when d = y are the same as the results when d =

- y, y = 1, 3, 5, 10. The intuition is that when the bias is large, receiving more signals is misleading rather than beneficial (the role of signals is just the opposite). When the cost of information acquisition is small, a social network between participants exacerbates the spread of misleading signals. Instead of improving the market performance, the dissemination of information is detrimental in this case. When the cost is high, the existence of a social network impedes the acquisition of biased information because of possible free-riding opportunities. Thus, an SEPM outperforms an NNPM.

5.3 What Happens When Participants Can Observe Their Friends' degrees?

In our theoretical model, we assume that each participant observes her own degree, but does not observe the degrees of her friends. In many situations, a participant has a good forecast of her own degree, but has incomplete information about the degrees of others [17]. However, this is a strong assumption when we talk about social media, such as Facebook and LinkedIn. In this section, we extend our analytical model and relax this assumption by allowing each participant to observe her friends' degrees in the four networks shown in Figure 1. The model setup in this section is similar to the setup in the previous theoretical analysis, except that the social networks are topologies in Figure 1 instead of random graphs. It is a complete information game in the sense that each participant has perfect knowledge about her friends' degrees, so the equilibrium concept is a Nash equilibrium rather than a Bayes-Nash equilibrium. In this section, we show that a non-increasing strategy in information acquisition is also a Nash equilibrium strategy (it might not be the unique equilibrium strategy). In other words, the basic result in Proposition 1 is also valid when each participant can observe the degrees of her friends.

For simplicity, let's consider sixteen participants in total, and each network structure consists of four participants. Participant i's net benefit of acquiring information when k_a (k_a= 0, 1, 2, or 3) of her friends acquire information is:

$$NB_{k_a} = u\left(m_i' = 1, m_{N_i(g)}\right) - u\left(m_i = 0, m_{N_i(g)}\right) = b\left(\frac{1}{k_a\rho_\varepsilon + \rho_V} - \frac{1}{(k_a+1)\rho_\varepsilon + \rho_V}\right) - c,$$

and for vector $m_{N_i(g)}$, there are k_a elements of 1 and $k_i - k_a$ elements of 0. Note that NB_{k_a} is decreasing in k_a, so we have five possible cases:

(1) $NB_3 < NB_2 < NB_1 < NB_0 \leq 0$.

In this case, the cost of information acquisition is too high, and $m_i = 0$ is a Nash equilibrium strategy for all sixteen participants in the four networks. It is trivial to show that the equilibrium strategy is non-increasing in degree.

(2) $NB_3 < NB_2 < NB_1 \leq 0 < NB_0$.

In Case 2, a Nash equilibrium strategy for all four participants in the non-networked environment is $m_i = 1$, because the net benefit of acquiring information when k_a= 0 is positive. A Nash equilibrium strategy for participants in the complete network is that one participant acquires information and the other three participants do not acquire information. For participants in the star network, a Nash equilibrium strategy is that the central participant does not acquire information and the other three participants acquire information. A Nash equilibrium strategy for participants in the circle network is that Participants 1 and 3 in Figure 1 acquire information and the other two participants do not acquire information. Summarizing all the equilibrium strategies, we find that 100% of participants with degree 0 acquire information, 100% of participants with degree 1 acquire information, 50% of participants with degree 2 acquire information, and 20% of participants with degree 3 acquire information. Thus, the equilibrium strategy is non-increasing in degree.

(3) $NB_3 < NB_2 \leq 0 < NB_1 < NB_0$.

Similarly, in Case 3, Nash equilibrium strategies for participants in the non-networked environment, the star network, and the circle network are the same as the strategies in Case 2. For participants in the complete network, a Nash equilibrium strategy is that two participants acquire information and the other two participants do not acquire information. We find that 100% of participants with degree 0 acquire information, 100% of participants with degree 1 acquire information, 50% of participants with degree 2 acquire information, and 40% of participants with degree 3 acquire information. Thus, the equilibrium strategy is non-increasing in degree.

(4) $NB_3 \leq 0 < NB_2 < NB_1 < NB_0$.

In Case 4, Nash equilibrium strategies for participants in the non-networked environment and the star network are the same as the strategies in case (2). For participants in the complete network, a Nash equilibrium strategy is that three participants acquire information and the other one participant does not acquire information. A Nash equilibrium strategy for participants in the circle network is that all of them acquire information. We find that 100% of participants with degree 0 acquire information, 100% of participants with degree 1 acquire information, 100% of participants with degree 2 acquire information, and 60% of participants with degree 3 acquire information. Thus, the equilibrium strategy is non-increasing in degree.

(5) $0 < NB_3 < NB_2 < NB_1 < NB_0$.

In Case 5, the cost of information is low, and the net benefit of acquiring information when k_a= 3 is positive. All participants in the four networks acquire information. The equilibrium strategy is trivially non-increasing in degree.

6. Conclusions

In this paper, we designed and carried out a laboratory experiment to examine the effect of a social network on the performance of a prediction market, as well as on the behavior of its participants. Through randomization in the controlled experiment, we were able to identify the causal relationship between the network degrees of players and their performance in the prediction market as well as their strategic decisions regarding whether to acquire costly information. More importantly, we tested the hypotheses that social-network-embedded prediction markets outperform prediction markets without social network in terms of prediction accuracy, and we found the difference to be significant when the cost of information acquisition is low but insignificant when the cost of information acquisition is high. Further numerical simulations suggest that the existence of a social network in a prediction market lowers the forecasting accuracy when the cost of information acquisition is high. This has a direct managerial implication for the business practice of prediction markets: When the predicted event is simple, promoting social networking among participants is beneficial, whereas if the predicted event involves complicated issues, a social network among participants should be discouraged.

In the past few years, many large firms, such as Google, Microsoft, and HP, have experimented with internal prediction markets to improve business decisions [10]. The primary goal of these markets is to generate predictions that efficiently aggregate many employees' information. It is easier for employees to gain access to private information about the company. Compared to outsiders, the cost of information acquisition is lower for internal employees. In this context of corporate prediction markets, the implication of our results is that an SEPM outperforms an NNPM when participants are internal employees.

Our experiment results also suggest that network structure matters when it comes to the performance of social-network-embedded prediction markets. An important future research direction is to extend our static model to further investigate exactly how the network structure affects prediction market performance as well as the performance and behavior of each participant over time. It would be interesting to create some social network measures that can help explain the variation of performances of prediction markets with different social network structures. Another interesting future research direction is to examine the incentives to share information in a social network through a laboratory experiment. Do participants exchange information according to reciprocity and norms of fairness? Studying the incentives for sharing information or the sale of information in social-network-embedded prediction markets remains an open question.

References

1. Aral, S., and Walker, D. Creating Social Contagion through Viral Product Design: A Randomized Trial of Peer Influence in Networks. Management Science, 57(9), (2011), 1623-1639.
2. Armstrong, J. S. *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Norwell: Kluwer Academic, 2001.
3. Bapna, R., Gupta, A., Rice, S., and Sundararajan, A. Trust, Reciprocity and the Strength of Social Ties: An Online Social Network Based Field Experiment, (2011), working paper.
4. Berg, J., Forsythe, R., Nelson, F., and Rietz, T. Results from a Dozen Years of Election Futures Markets Research. In Plott, C.R., and Smitt V.L. (eds.), *Handbook of Experimental Economics Results*, Amsterdam: Elsevier Science, 2008, 742-751.
5. Berg, J., Neumann, G. R., and Rietz, T. A. Searching for Google's Value: Using Prediction Markets to Forecast Market Capitalization Prior to an Initial Public Offering. *Management Science*, 55(3), (2009), 348-361.
6. Berg, J. E., and Rietz, T. A. Prediction Markets as Decision Support Systems. *Information Systems Frontiers*, 5(1), (2003), 79-93.
7. Brown, M. B., and Forsythe, A. B. Robust Tests for The Equality of Variances. *Journal of the American Statistical Association*, 69(346), (1974), 364-367.
8. Cameron, A. C., Gelbach, J. B., and Miller, D. L. Bootstrap-Based Improvements for Inference with Clustered Errors. *Review of Economics and Statistics*, 90(3), (2008), 414-427.
9. Çelen, B., and Hyndman, K. Social Learning Through Endogenous Information Acquisition: An Experiment. *Management Science*, 58(8), (2012), 1525-1548.
10. Chen, K., and Plott, C. Information Aggregation Mechanisms: Concept, Design and Implementation for a Sales Forecasting Problem. California Institute of Technology Social Science, Working Paper No. 1131, (2002).
11. Chen, J., Xu, H., and Whinston, A.B. Moderated Online Communities and Quality of User-Generated Content. *Journal of Management Information Systems*, 28(2), (2011), 237-268.
12. Cohen, L., Frazzini, A. and Malloy, C.J. The Small World of Investing: Board Connections and Mutual Fund Returns. *Journal of Political Economy*, 116(5), (2008), 951-979.
13. Coval, J.D., and Moskowitz, T.J. The Geography of Investment: Informed Trading and Asset Prices. *Journal of Political Economy,* 109(4), (2001), 811-841.

14. Erdős, P., and Rényi A. On the Evolution of Random Graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, (1960), 17–61.
15. Fang, F., Stinchcombe, M., and Whinston, A.B. Proper Scoring Rules with Arbitrary Value Functions. *Journal of Mathematical Economics*, 46(6), (2010), 1200-1210.
16. Foutz, N. Z., and Jank, W. Prerelease Demand Forecasting for Motion Pictures Using Functional Shape Analysis of Virtual Stock Markets. *Marketing Science*, 29(3), (2010), 568-579.
17. Galeotti, A., Goyal S., Jackson, M.O., Vega-Rendondo F., and Yariv L. Network Games. *Review of Economic Studies,* 77(1), (2010), 218-244.
18. Garg, R., Smith, M. D., and Telang, R. Measuring Information Diffusion in An Online Community. *Journal of Management Information Systems*, 28(2), (2011), 11-38.
19. Gaviria, A., and Raphael, S. School-Based Peer Effects and Juvenile Behavior. *Review of Economics and Statistics*, 83(2), (2001), 257-268.
20. Gillen, B.J., Plott, C.R., and Shum, M. Information Aggregation Mechanisms in the Field: Sales Forecasting Inside Intel. California Institute of Technology Social Science, Working Paper, (2012).
21. Golub, B., and Jackson, M. Naive Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics,* 2(1), (2010), 112-149.
22. Guo, Z., Fang F., and Whinston, A.B. Supply Chain Information Sharing in a Macro Prediction Market. *Decision Support Systems,* 42(3), (2006), 1944-1958.
23. Hanson, R. Combinatorial Information Market Design. *Information Systems Frontiers,* 5(1), (2003), 107-119.
24. Healy, P.J., Linardi S., Lowery, J.R., and Ledyard, J.O. Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders. *Management Science,* 56(11), (2010), 1977-1996.
25. Hinz, O., and Spann, M. The Impact of Information Diffusion on Bidding Behavior in Secret Reserve Price Auctions. *Information Systems Research*, 19(3), (2008), 351-368.
26. Jackson, M.O. *Social and Economic Networks*, Princeton University Press, Princeton, NJ, 2008.
27. Jian, L., and Sami, R. Aggregation and Manipulation in Prediction Markets: Effects of Trading Mechanism and Information Distribution. *Management Science,* 58(1), (2012), 123-140.
28. Lorenz, J., Rauhut, H., Schweitzer, F., and Helbing, D. How Social Influence Can Undermine the Wisdom of Crowd Effect. *Proceedings of the National Academy of Sciences*, 108(22), (2011), 9020-9025.
29. Manski, C.F. Identification of Endogenous Social Effects: The Reflection Problem. *Review of Economic Studies*, 60(3), (1993), 531-542.
30. Qiu, L., Rui, H., and Whinston, A.B. Social Network-Embedded Prediction Markets: The Effects of Information Acquisition and Communication on Predictions. *Decision Support Systems*, (2013), forthcoming.
31. Sacerdote, B. Peer Effects with Random Assigment: Results for Dartmouth Roommates. *Quarterly Journal of Economics*, 116(2), (2001), 681-704.
32. Sorensen, A. T. Social Learning and Health Plan Choice. *RAND Journal of Economics*, 37(4), (2006), 929-945.
33. Spann, M., and Skiera, B. Internet-Based Virtual Stock Markets for Business Forecasting. *Management Science*, 49(10), (2003), 1310-1326.

34. Stieglitz, S., and Dang-Xuan, L. Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29(4), (2013), 217-248.
35. Wattal, S., Racherla, P., and Mandviwalla, M. Network Externalities and Technology Use: A Quantitative Analysis of Intraorganizational Blogs. *Journal of Management Information Systems*, 27(1), (2010), 145-174.
36. Wooldridge, J.M. *Econometrics Analysis of Cross Section and Panel Data*, MIT Press: Cambridge, MA, 2002.
37. Wooldridge, J.M. Cluster-Sample Methods in Applied Econometrics. *American Economic Review*, 93(2), (2003), 133-138.
38. Zhang, X., and Wang, C. Network Positions and Contributions to Online Public Goods: The Case of Chinese Wikipedia. *Journal of Management Information Systems*, 29(2), (2012), 11-40.

Appendix

Appendix A: Proof of Proposition 1

We say that a function u exhibits strategic substitutes if an increase in others' actions lowers the marginal returns from one's own actions: For all $m_i' > m_i$ and $m'_{N_i(g)} \geq m_{N_i(g)}$,

$$u\left(m_i', m'_{N_i(g)}\right) - u\left(m_i, m'_{N_i(g)}\right) \leq u\left(m_i', m_{N_i(g)}\right) - u\left(m_i, m_{N_i(g)}\right).$$

When u exhibits strategic substitutes, a participant's incentive to take a given action decreases as more friends take that action.

Lemma 1. *If the payoff is a quadratic loss function, then* $u\left(m_i, m_{N_i(g)}\right)$ *exhibits strategic substitutes.*

Proof. A participant's utility maximization problem given m_i and $m_{N_i(g)}$ is equivalent to a predictor error minimization problem. We can obtain the best mean square predictor of V based on S_i:

$$E[V|S_i] = \frac{\rho_V}{\rho_\varepsilon + \rho_V} V_0 + \frac{\rho_\varepsilon}{\rho_\varepsilon + \rho_V} S_i.$$

Similarly, we can obtain the best mean square predictor of V based on other information sets. Assume that for $m_{N_i(g)}$, there are k_a of Participant i's friends (among the total number k_i)

who acquire information. In other words, for vector $m_{N_i(g)}$, there are k_a elements of 1 and $k_i - k_a$ elements of 0. Let $A_{N_i(g)}$ be the set of friends who acquire information. If Participant i acquires information, the best mean square predictor is:

$$\frac{\rho_V}{(k_a+1)\rho_\varepsilon+\rho_V}V_0+\frac{\rho_\varepsilon}{(k_a+1)\rho_\varepsilon+\rho_V}\left(\sum_{i\in A_{N_i(g)}} S_i\right).$$

For Participant i's action, $m_i = 0$, and $m_i' = 1$:

$$u\left(m_i, m_{N_i(g)}\right) = a - b\left[\frac{\rho_V^2}{(k_a\rho_\varepsilon+\rho_V)^2}\frac{1}{\rho_V}+\frac{\rho_\varepsilon^2}{(k_a\rho_\varepsilon+\rho_V)^2}\frac{k_a}{\rho_\varepsilon}\right] = a - b\left(\frac{1}{k_a\rho_\varepsilon+\rho_V}\right),$$

and

$$u\left(m_i', m_{N_i(g)}\right) - u\left(m_i, m_{N_i(g)}\right) = -b\left(\frac{1}{(k_a+1)\rho_\varepsilon+\rho_V}-\frac{1}{k_a\rho_\varepsilon+\rho_V}\right) - c.$$

From here, obtaining the following equation is straightforward:

$$\frac{\partial}{\partial k_a}\left[u\left(m_i', m_{N_i(g)}\right) - u\left(m_i, m_{N_i(g)}\right)\right] = -\frac{\rho_\varepsilon}{(k_a\rho_\varepsilon+\rho_V)^2}+\frac{\rho_\varepsilon}{[(k_a+1)\rho_\varepsilon+\rho_V]^2} < 0.$$

Therefore, u exhibits strategic substitutes. ■

If the payoff function exhibits strategic substitutes, then for $m_i' > m_i$ and $k_i' > k_i$,

$$\begin{aligned}
&U(m_i', \sigma; k_i) - U(m_i, \sigma; k_i) \\
&= \textstyle\sum_{k_{N_i(g)}} P\left(k_{N_i(g)}|k_i\right)\left[u\left(m_i', \sigma_{N_i(g)}\right) - u\left(m_i, \sigma_{N_i(g)}\right)\right] \\
&= \textstyle\sum_{k_{N_i(g)}} P\left(k_{N_i(g)}|k_i\right)\left[u\left(m_i', \left(\sigma_{N_i(g)}, 0\right)\right) - u\left(m_i, \left(\sigma_{N_i(g)}, 0\right)\right)\right] \\
&= \textstyle\sum_{k_{N_i(g)}} P\left(k_{N_i(g)}|k_i'\right)\left[u\left(m_i', \left(\sigma_{N_i(g)}, 0\right)\right) - u\left(m_i, \left(\sigma_{N_i(g)}, 0\right)\right)\right] \\
&> \textstyle\sum_{k_{N_i(g)}} P\left(k_{N_i(g)}|k_i'\right)\left[u\left(m_i', \left(\sigma_{N_i(g)}, m_{k+1}\right)\right) - u\left(m_i, \left(\sigma_{N_i(g)}, m_{k+1}\right)\right)\right] \\
&= U(m_i', \sigma; k_i') - U(m_i, \sigma; k_i'),
\end{aligned} \tag{10}$$

where the third equality follows from the assumption that neighbors' degrees are all stochastically independent, and the first inequality follows from strategic substitutes. Then, we show the existence of a decreasing symmetric equilibrium by using two steps: (1) There exists a symmetric equilibrium; and (2) every symmetric equilibrium is non-increasing in degree. First, we want to show a symmetric equilibrium exists (we allow mixed-strategy equilibrium). Our game is a standard symmetric incomplete information game because all participants have identical action sets of information acquisition $\Delta\{0,1\}$; the quadratic payoff functions are also the same; and participant's beliefs concerning networks are ex-ante symmetric. Given that the action set $\Delta\{0,1\}$ is compact, and the payoff function is continuous, then a symmetric mixed strategy Bayes-Nash equilibrium exists according to the fixed-point theorem (see [17]).

Next, we show that every symmetric equilibrium is non-increasing. Let σ_k be a symmetric equilibrium strategy for the participant with degree k. If σ_k is a strategy with all degrees choosing action 1 with probability 1, the equilibrium is obviously non-increasing. Thus, we focus on the non-trivial case. If σ_k is not a trivial strategy, then let $m_k = \min\,[supp(\sigma_k)]$, where $supp(\sigma_k)$ is the support of the mixed strategy σ_k. In our context, the support can be {0}, {1}, or {0,1}. If $m_k = 1$, it is easy to show that $m_{k'} \leq m_k$ for all $m_{k'} \in supp(\sigma_{k'})$ with $k' > k$. If $m_k = 0$, then for any $m > m_k$, we have the following inequality by equation (10):

$$U(m, \sigma; k) - U(m_k, \sigma; k) > U(m, \sigma; k+1) - U(m_k, \sigma; k+1).$$

Note that $m_k \in supp(\sigma_k)$, so we have: $U(m, \sigma; k) - U(m_k, \sigma; k) \leq 0$. Thus, $U(m, \sigma; k+1) - U(m_k, \sigma; k+1) < 0$, for all $m > m_k$. This implies that if $m_{k+1} \in supp(\sigma_{k+1})$ then $m_{k+1} \leq m_k$. Therefore, σ_k FOSD σ_{k+1}. The conclusion follows by iterating this process.

Now let's show that the non-increasing strategy is actually a threshold strategy. Suppose that for degree k_i participant, there is a positive probability of acquiring information, we can prove that $\sigma\left(m_i = 1 | \hat{k}\right) = 1$, for all $\hat{k} < k_i$, by the decreasing difference of $U(m_i, \sigma; k_i)$.

Similarly, we can show that if for degree k_i participant, there is positive probability of not acquiring information, then $\sigma\left(m_i = 1|\hat{k}\right) = 0$, for all $\hat{k} < k_i$. Then the equilibrium strategy is a threshold strategy.

We make a few more remarks here. Because all participants adopt a threshold strategy, Participant i believes that the probability for a randomly chosen neighbor to acquire information is $\theta = \Pr\left(k_j \leq k^*\right),\ j \in N_i(g)$. Participant i's belief about the number of informed neighbors thus follows a binomial distribution given by:

$$f(k_a; k_i, \theta) = \binom{k_i}{k_a}\theta^{k_a}(1-\theta)^{k_i-k_a},$$

where k_a is the number of participants who acquire information, and $f(k_a; k_i, \theta)$ is the density function of the binomial distribution. Knowing the belief of Participant i, we can obtain the expected payoff $U\left(m_i, \sigma_{N_i(g)}; k_i\right)$. Because k^* is a threshold, it is determined by the following inequalities:

$$U\left(m_i = 1, \sigma_{N_i(g)}; k^*\right) \leq U\left(m_i = 0, \sigma_{N_i(g)}; k^*\right),$$

$$U\left(m_i = 1, \sigma_{N_i(g)}; k^* + 1\right) > U\left(m_i = 0, \sigma_{N_i(g)}; k^* + 1\right).$$

These inequalities simply mean that the participant with degree k^* is better off to acquire information, and the participant with degree $k^* + 1$ is better off not to acquire information.

Appendix B: Experimental Instructions

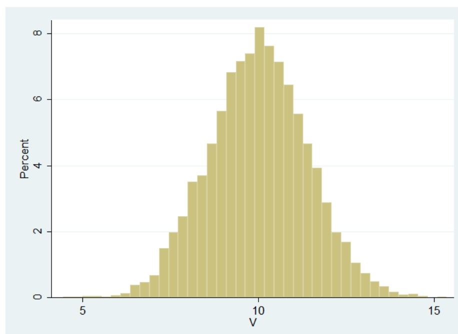
The following are the experimental instructions for an SEPM. The guidelines for an NNPM are similar except that the participants are not allowed to communicate with others.

Experiment Guidelines

General Guideline: This is an economic experiment so it is conducted with Real Money! Your profit is a direct result of your prediction performance during the experiment. The experiment has 2 rounds. The highest cash payoff for you to earn is $5*2=$10! In order to maximize your profits, you need to read the instructions carefully and use your information wisely. The experiment has 2 rounds. Your total payoff is the sum of the payoff in each round. If your total payoff is less than \$5, you will get \$5.

Experiment Description

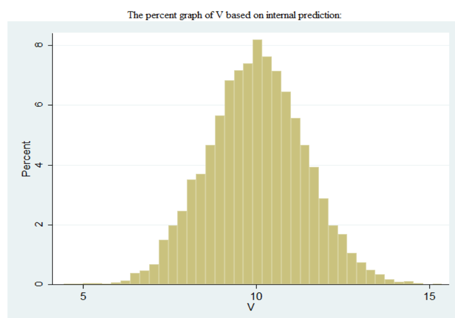
CREC (Central Real Estate Company) needs to predict the size of the rental market, V, in a large metropolitan area. The internal estimation predicted by employees within the company suggests that the market size, V, is probably around \$10 millions. Below is the percent graph of the employees' predictions: Most of them think that the market size $V = 10$.



As the head of the marketing department of CREC, you can also consider purchasing an evaluation of the market size from one of several outside experts. Your experience tells you that each expert's prediction is twice as accurate as the internal prediction. Obtaining the prediction of an outside expert will cost you money in this experiment. If you choose to purchase an expert's opinion, you can combine the internal estimation from employees with the expert prediction to get a more precise estimate. The actual market size, V, in million USD, will be announced right after the experiment (Note that the true value of V in round 1 is different from the value in round 2). Suppose that your prediction is x million USD, if you do not purchase expert opinions, your payoff in this round is: $\$ 5 - (x - V)^2$.

If you choose to purchase an expert's opinion, you have to pay a cost $c = \$1$ in this experiment and your payoff in the first round is \$ $5 - (x - V)^2 - c$.

Apparently, the more precise your estimate is, the higher the payoff you will get. You are free to communicate with other experiment participants on your designated Gmail account after making a decision about whether to purchase an expert's signal. You can discuss with the other group members through Gtalk. If your group members purchase evaluations from outside experts, they may provide useful information to you. After reading the guidelines, you need to make a decision on whether to purchase a signal.



Do you want to purchase a signal at cost $1 ?

Yes

No

Then, you can submit your predictions on the basis of the prior and your signal.

Prior: 10

Signal: 10.8575

Submit