A TWITTER-BASED PREDICTION MARKET: SOCIAL NETWORK APPROACH

Completed Research Paper

Liangfei Qiu
University of Texas at Austin
Austin, Texas 78712
liangfeiqiu@mail.utexas.edu

Huaxia Rui
University of Texas at Austin
Austin, Texas 78712
ruihuaxia@mail.utexas.edu

Andrew Whinston
University of Texas at Austin
Austin, Texas 78712
abw@uts.cc.utexas.edu

Abstract

Information aggregation mechanisms are designed explicitly for collecting and aggregating dispersed information. An excellent example of the use of this “wisdom of crowds” is a prediction market. The purpose of our Twitter-based prediction market is to suggest that carefully designed market mechanisms can elicit and gather dispersed information that can improve our predictions. We develop an information system that combines the power of prediction markets with the popularity of Twitter. Simulation results show that our network-embedded prediction market can produce better predictions as a result of the information exchange in social networks and can outperform other non-networked prediction markets. We also demonstrate that forecasting errors decrease with the cost of acquiring information in a network-embedded prediction market.

Keywords: prediction market, social networks, information acquisition
Introduction

Predicting outcomes is essential to business decision making. But how do we assess the probability of future events in a world full of uncertainties? The two most popular approaches for producing reliable forecasts are experts' predictions and "the wisdom of crowds" (i.e., collecting and aggregating dispersed information from a large population). In the experts' predictions approach, a regression model forecasts future events based on known past events. For example, specialists' forecasting in econometrics might predict the opening price of a stock based on its past performance. There are several disadvantages to this approach; for instance, identifying the true experts can be hard work. Even if we can identify the experts, the questions of how to elicit the knowledge quickly and of how to combine experts' differing opinions remain. One can rarely conclusively infer that the expert was intentionally deceptive, even if the realized outcome departs significantly from the expert's prediction.

The other approach, "the wisdom of crowds," is based on the assumption that information and knowledge in social systems frequently exist only as dispersed opinions, and that aggregating dispersed information can produce accurate predictions. A prediction market illustrates effective use of the wisdom of crowds. Assets are created whose final value is tied to a particular event—for example, whether the next U.S. president will be a Republican or a Democrat. People "place bets" on events that they think are most likely to happen, thus revealing in a sense the nature of their private information and subsequent posterior beliefs. The market mechanisms provide a method of "putting your money where your mouth is" (Fang, Stinchcombe, and Whinston, 2007). Prediction markets can also immediately incorporate new information and provide a real-time forecast. The hope is that by aggregating the private information of a large population, a prediction market can generate fairly accurate predictions of future events. Financial Times reported a voting application on Facebook, which has some features of a prediction market. Producer Endemol used Facebook's Credits system of micropayments to let fans vote for their favorite contestants via a Facebook application. The fans must put money into their credits accounts, thus voting for the favorite contestants is a typical way of “putting your money where your mouth is.” By aggregating the information from voting, the company could obtain a more accurate prediction of the contestants’ potentials to be celebrities. It is also important to note that Facebook network plays a role in online voting. Unlike traditional phone and text-message voting, each vote cast is shared with friends on the site, and the fans could exchange views on Facebook network.

In prediction markets, prediction decisions rely on information that participants gather through communication with friends, neighbors, and co-workers. Twitter is a good example of such a network. Launched in 2006, the closely held Twitter now boasts more than 200 million accounts, and more than 130 million unique messages are typically posted per day. Providing an example of how Twitter has been used, CNBC News reported that farmers used their Twitter accounts to post a message, or to tweet, about a particularly robust corn crop. Tweeting with fellow farmers has become a way for the participants in a far-flung and isolated business to compare notes on everything from weather conditions to new fertilizers. The use of tweets by a growing network of farmers and traders is transforming how this multi-trillion dollar industry does business. These tweets are dramatically accelerating the flow of information, giving investors an edge in the commodities market. Generally speaking, people connected by Twitter network create important conduits of information. Thus, natural questions about these information exchanges arise, such as how farmers' trading decisions are affected by the flow of information transmitted through the Twitter network. More precisely, what is the effect on farmer's trading decisions of these connections?

---

1 A number of theoretical papers demonstrate that an empirical test can be passed not only by an expert who knows the future event and reports this knowledge truthfully, but also by an expert who knows nothing about it but who delivers forecasts strategically to pass this particular test. Empirical tests cannot distinguish between these two types of experts (Olszewski and Sandroni (2008, 2009)).

2 See "Endmol and Facebook in online voting tie-up", Financial Times, August 27, 2011.

3 The information is from CNBC News, March 8, 2011. The CNBC reporter called the phenomenon "Trading on Twitter." Grisafi, known as @IndianaGrainCo on Twitter, says he tweets with at least 15 farmers on a regular basis to check on crop conditions.
In this study, a prediction market is a betting market where individuals place bets on the outcome of future events. Players in the prediction market are connected by a Twitter network, and they make forecasts based on the information gathered through the network. It is a socially embedded prediction market based on Twitter network.

This paper studies a network-embedded prediction market, modeled as an incomplete information network game. Each player observes her number of friends, but does not know the structure of the network on which she plays. This setup is motivated by the privacy issues of social media. For instance, on Google Plus, people put their friends into different Circles. Every time they share information and thoughts, they can specify exactly which Circles receive them. People know the number of friends who share information with them, but they don't know the whole network of information sharing. Twitter also has something similar to Circle, called List.

In the network game, the equilibrium action of information acquisition is non-increasing in the player's degree. There is a symmetric Bayes-Nash Equilibrium, where all players use a simple cut-off strategy involving the threshold degree. The inefficiency of information acquisition is caused by free-riding. In the equilibrium, higher degree players exert lower efforts in acquiring information but earn a higher payoff compared to their less connected peers. This implies that social connections confer personal advantage.

We conduct a variety of simulations of agents trading in prediction markets and show that, in prediction performance, the network-embedded prediction market outperforms the non-networked prediction market. Our Twitter-based prediction market has a significant advantage in forecasting and can produce a better prediction as a result of the information exchange in social networks. We also demonstrate that, in a network-embedded prediction market, the forecasting errors decrease along with the cost of acquiring information. The implication is similar to the efficient market hypothesis. As more and more players acquire information, the prediction market prices fully incorporate all available information and thus yield better forecasts. We further demonstrate that, compared to the prediction markets without social networks, the forecasting errors are lower in a network-embedded prediction market.

Literature Review

A number of empirical studies show that prediction markets have been successfully used to predict outcomes in many areas (Wolfers and Zitzewitz, 2004). In the political domain, Berg, Forsythe, Nelson, and Reitz (2008) document that prediction markets outperform polls for longer horizons. In the Iowa Electronic Market, the most famous prediction market, traders buy and sell contracts that pay $1 if a given candidate wins the election. In business practice, Hewlett-Packard uses prediction markets to forecast sales, as well as financial and accounting results. The predictions consistently beat the official HP forecasts (Plott and Chen, 2002). In Berg et al. (2009), prediction markets are designed to forecast market capitalization prior to an initial public offering. Castro and Cramton (2009) propose a prediction market as a way to forecast future demand for electricity. Guo, Fang, and Whinston (2006) propose a macro prediction market to effectively elicit and aggregate useful information about systematic demand risk and show that such information can be used to achieve accurate demand forecast sharing and better channel coordination in the supply chain system. All these examples share the following characteristic: Small bits and pieces of relevant information exist in the opinions and intuitions of diverse individuals. Most previous papers focus on how to elicit dispersed private information. For example, Fang, Stinchcombe, and Whinston (2006) propose a betting mechanism that elicits agents' private information, as well as the precision of the information. In their work, the information of all the players is independent. However, in our network-embedded prediction market, the information players possess is correlated with that of their neighbors.

Burggen et al. (2010) designed an experiment under laboratory conditions to compare the forecasting accuracy of the prediction market approach and the traditional combined judgmental approach. However, lab experiments in isolation have limited relevance in predicting field outcomes. In this experiment by Burggen et al., upper level undergraduates and MBAs participated in the study. Although students are often the standard subject pool because they are a convenient sample for academics, the primary drawback is evident: The information sources of the students are very similar. Our Twitter-based prediction market is different from a lab experiment in many ways. In the field, subjects bring certain information to their trading activities in addition to their knowledge of the trading institution. In abstract
settings, the importance of this information is diminished, which can lead to behavioral changes. Field experience also can play a major role in helping individuals develop heuristics for specific tasks (Harrison and List, 2004). In the Twitter-based prediction market, we are recruiting subjects in the field rather than in the classroom and are using a field context rather than abstract instructions. The environment of this experiment thus can provide a context for suggesting strategies and heuristics that a lab setting might not.

This study is related to the work on social networks. The role of social networks in finding jobs is a leading example of networked markets (Granovetter, 1973). Calvo-Armengol and Jackson (2004) develop a model where players obtain information about job opportunities through a social network to examine how the network structure can affect employment and wage dynamics. Golub and Jackson (2010) consider the wisdom of crowds in social networks. They discuss how network structure influences the spread of information and show that all opinions in a large society converge to the truth if and only if the influence of the most influential agent vanishes as the society grows. Galeotti, Goyal, Jackson, Vega-Rendondo, and Yariv (2010) provide a framework to analyze strategic interactions in an incomplete information network game.

Our model can also be viewed as an extension of Grossman and Stiglitz (1980). Their paper discusses information acquisition in a competitive market. Prices reflect the information of informed individuals. However, in today’s economy, many markets operate through networked interactions. A series of questions arise. Does the use of networked markets affect information acquisition and efficiency of a market? What is the effect of social connections on players’ incentives to acquire information? Do players with more connections earn more, compared to their less connected peers? How does the cost of acquiring information affect forecasting errors in the prediction markets? Our work provides a game-theoretic framework to analyze these questions.

The rest of the paper is organized as follows. Section 3 presents the model of social network-embedded prediction markets. In section 4, we describe the design of our prediction markets. Section 5 discusses how to embed Twitter into the prediction market. We summarize and conclude in Section 6.

A Theoretical Model

Social Networks

In this section, we set up a theoretical model of a network prediction market. Players are linked to each other according to their Twitter network, and information is transmitted through the networks.

We first outline an economic model of prediction markets. A principle wants to forecast a random variable \( V \), and \( v \) is the realization of \( V \). In reality, \( V \) could be movie box office revenue, future demand for electricity, or stock prices. The principal resorts to \( N \) risk-averse agents to obtain an accurate prediction. Principal and agents share a common prior on the distribution of \( V \), given by:

\[
V \sim N(V_0, 1/\rho_t)
\]

Suppose the agents are connected by a Twitter network. The network \( \Gamma = (N, L) \) is given by a finite set of nodes \( N = \{1, 2, \ldots, n\} \) and a set of links \( L \subseteq N \times N \). The connections between the players 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 neighbors of \( i \). The degree of player \( i \) is the number of \( i \)'s neighbors, \( k_i(g) = |N_i(g)| \). Each agent can bet on an asset tied to \( V \) in the prediction market, which

\[^4\text{For simplicity, we assume the network is undirected, but the results also hold for directed networks.}\]
means the agents who know they have more accurate information are generally willing to bet more money on it. The agents’ payoffs depend on the realization of $V$. For simplicity, we assume an agent's payoff from betting is an exponential utility function with a constant coefficient of absolute risk aversion $\gamma$:

$$
E\left[-\exp\left[-\gamma x_i (V - P) - cm_i \right]|I_i \right],
$$

where $P$ is the market price of the asset tied to $V$, and $x_i$ is the demand for the risky asset. If agent $i$ thinks the realization of $V$ is high, she will buy the risky asset; otherwise, she will short. $I_i$ is agent $i$’s information set, which includes the information she acquires and the signals passed by her neighbors.

From the moment-generating function, it is well known that maximizing equation (1) is equivalent to maximizing the following quadratic function of $x_i$:

$$
\text{Max}_{x_i} \gamma x_i \left(E\left[V | I_i \right] - P \right) - \frac{1}{2} Var\left[V | I_i \right] \gamma^2 x_i^2.
$$

The price $P$ is endogenously given by the market clearing condition: $\sum_{i=1}^{n} x_i = 0$.

From the first-order condition, we obtain

$$
x_i = \frac{E\left[V | I_i \right] - P}{\gamma Var\left[V | I_i \right]}.
$$

Each agent can access a private, independent information source at a cost $c$. $m_i$ is an indicator function, which represents whether player $i$ acquires information. Players exchange information over the communication network. If player $i$ does acquire information from her private source ($m_i = 1$), she observes a conditionally independent private signal and passes it to her neighbors:

$$
S_i = v + e_i, e_i \sim N\left(0, 1/\rho_i \right),
$$

where $\rho_i$ is the precision of player $i$’s information for $i = 1, 2, ..., N$. $e_i$ is an error term that is independent across agents. It implies that in arriving at a decision, a player makes use of her information as well as the signals of others, especially of those who are close to her. In this case, player $i$ forms beliefs about the random variables $V$ from her private signal $S_i$, as well as information she obtains from her neighbors. If the player has decided not to acquire information ($m_i = 0$), she forms the beliefs only from her neighbors' signals.

Next, we turn to the relationship between player's strategies and their payoff. Player $i$’s payoff can be written as:

$$
u\left(m_i, \hat{V}\left(m_i, m_{N_i}^{(e)} \right) \right) = u\left(m_i, m_{N_i}^{(e)} \right),
$$

where $m_{N_i}^{(e)} \in \{0, 1\}^k$ is the action profile of player $i$’s neighbors. The payoff function depends on whether player $i$ and her neighbors acquire information. Similar to Galeotti, Goyal, Jackson, Vega-Rendondo, and Yariv (2010), the payoff function depends on the player’s degree $k_i$ but not on her identity $i$. Therefore,

---

5 The assumption is similar to the setup in Calvo-Armengol and Jackson (2004). We assume people exchange information according to reciprocity and norms of fairness instead of discussing the incentives to share information (Akerlof, 1982).
any two players who have the same degree have the same payoff function. It is also evident that \( u \) depends on the vector \( m_{N_i(s)} \) in an anonymous way; thus, a permutation of \( m_{N_i(s)} \) does not change the payoff.

We say a payoff function exhibits strategic substitutes if an increase in others’ actions lowers the marginal returns from a player’s own actions: For all \( k \), \( m_i' > m_i \) and \( m_{N_i(s)}' \geq m_{N_i(s)} \):

\[
 u(m_i', m_{N_i(s)}') - u(m_i, m_{N_i(s)}') \leq u(m_i', m_{N_i(s)}') - u(m_i, m_{N_i(s)}') .
\]

The intuition is that when a payoff function exhibits strategic complements, a player’s incentive to take a given action increases as more neighbors take that action. When a payoff function exhibits strategic substitutes, a player’s incentive to take a given action decreases as more neighbors take that action.

**Lemma 1.** The payoff function in the network-embedded prediction market exhibits strategic substitutes.

**Proof:** Applying the following property, a player’s utility maximization problem is equivalent to a predictor error minimization problem.

**Property 1.** If \( E[V^2] \) is finite and \( \mu(x) = E[V|x] \), then \( \mu \) is a solution to

\[
 \min_{h \in H} E \left[ (V - h(x))^2 \right],
\]

where \( H \) is the set of functions \( h \). \( \mu \) is the best mean square predictor of \( V \) based on information contained in \( x \).

The proof of property 1 is in Wooldridge (2002, p. 30). From property 1, we can obtain the best mean square predictor of \( V \) based on \( S_i \):

\[
 E[V \mid S_i] = \frac{\rho_v}{\rho_v + \rho_x} V_0 + \frac{\rho_x}{\rho_v + \rho_x} S_i .
\]

Similarly, we can obtain the best mean square predictor of \( V \) based on other information sets. Assume that for \( m_{N_i(s)} \), there are \( k_a \) of player \( i \)’s neighbors (among the total number \( k_i \)) who acquire information. In other words, for vector \( m_{N_i(s)} \), there are \( k_a \) elements of 1 and \( k_i - k_a \) elements of 0. For player \( i \)’s action, \( m_i = 0 \), and \( m_i' = 1 \), we can obtain:

\[
 \frac{\partial}{\partial k_a} \left[ u(m_i', m_{N_i(s)}') - u(m_i, m_{N_i(s)}') \right] = \left[ \exp \left( - \frac{(1 + k_a) \rho_x}{2 \rho_v} \right) \right] - \left[ \exp \left( - \frac{k_a \rho_x}{2 \rho_v} \right) \right] \frac{\rho_x}{2 \rho_v} < 0
\]

Therefore, the payoff function exhibits strategic substitutes. ■

In our context, the player’s payoffs depend on the sum of the neighbors’ strategies, and the payoff function satisfies the following general property discussed in Galeotti, Goyal, Jackson, Vega-Rendondo, and Yariv (2010):

**Property 2.** \( u(m_i, m_{N_i(s)}', 0) = u(m_i, m_{N_i(s)}) \) for any \( \left( m_i, m_{N_i(s)} \right) \in \{0,1\}^{k+1} \).

Under Property 2, adding a link to a neighbor who chooses action 0 is payoff equivalent of not having an additional neighbor. Property 2 is true in this prediction network game.
We assume that players do not know the structure of the network on which they play. Thus, we relax the assumption of complete information on the social networks, and in our prediction network game, players do not know the whole network but are informed only of their degrees. Each player observes her own degree $k_i$ (her type), but does not observe the degree or connections of any other player in the network. Her beliefs about the degrees (types) of her neighbors, given her own degree $k_i$, are equal to

$$P(k_i | k) \in \Delta\{1, ..., k_{\text{max}}\},$$

where $k_{\text{max}}$ is the maximal possible degree, and $\Delta\{1, ..., k_{\text{max}}\}$ is the set of probability distribution on \{1, ..., $k_{\text{max}}$\}. For simplicity, we made a strong assumption that neighbors’ degrees are all stochastically independent. This assumption is inappropriate for some networks, including those of scientific collaboration or actor collaboration, which display significant positive degree correlation (Newman, 2003). However, we can argue that the degrees of two neighbors are approximately independently distributed for large networks, such as Twitter.

**Assumption 1.** Degrees of neighboring nodes are independent.

**Symmetric Bayes-Nash Equilibria**

A strategy is a function $\sigma : \{1, ..., k_{\text{max}}\} \rightarrow \Delta\{0, 1\}$, and we focus on symmetric Bayes-Nash equilibria, where all players follow the same strategy $\sigma$. The expected payoff of player $i$ with degree $k_i$ and action $m_i$ is equal to

$$U(m_i, \sigma; k_i) = \sum_{k_{N_i(x)}} P(k_{N_i(x)} | k_i) u_k(m_i, \sigma_{N_i(x)}),$$

where $k_{N_i(x)} = (k_1, ..., k_{k_i})$ is a vector of degrees of player $i$’s neighbors.

Recall that player $i$’s strategy $\sigma$ is non-increasing if $\sigma(k_i)$ first-order stochastically dominates $\sigma(k_i')$ for each $k_i' > k_i$. In our context, this simply implies that high-degree players will randomize their actions with lesser probability in $m_i = 1$ and with greater probability in $m_i = 0$.

**Proposition 1.** If Assumption 1 holds, there exists a symmetric equilibrium that is non-increasing in degrees in the network embedded prediction market: There exists some threshold $k^* \in \{0, 1, 2, ..., k_{\text{max}}\}$, such that the probability $\sigma(m_i = 1 | k_i)$ of choosing to acquire information in the unique, non-increasing symmetric equilibrium strategy $\sigma$ satisfies:

$$\sigma(m_i = 1 | k_i) = \begin{cases} 1 & \text{for } k_i \leq k^*; \\ 0 & \text{for } k_i > k^*. \end{cases}$$

---

6 This particular formulation is not standard. Notice that a player does not have beliefs about all players in the network, but only about the degrees of the players connected to him. We assume that players’ beliefs about the rest of the network are summarized by a probability distribution over the degrees of their neighbors.

7 Actually, our model can allow correlation between neighbors’ degrees. This implies that the conditional distributions concerning neighbors’ degrees can vary with a player’s degree. The results also hold when higher degrees for a given player are correlated with lower degrees of all his neighbors.
Proof: Suppose that Property 2 and Assumption 1 hold. If the payoff function exhibits strategic substitutes, then the expected payoff exhibits degree substitution. If strategy \( \sigma_i \) is non-increasing, the payoffs \( U(m_i, \sigma_i; k_i) \) have decreasing differences in \( m_i \) and \( k_i \). For \( m_i' > m_i \) and \( k_i' \geq k_i \):

\[
U(m_i', \sigma_i; k_i') - U(m_i, \sigma_i; k_i) \leq U(m_i', \sigma_i; k_i) - U(m_i, \sigma_i; k_i')
\]

Then, the existence of a symmetric equilibrium follows from the standard existence proof. Let \( \Sigma^{\text{dec}} \) be the set of non-increasing strategies. By the Topkis Theorem, each best response to all non-increasing strategies is non-increasing. Thus, we can apply the proof of the existence theorem (the fixed-point theorem) to the best response correspondence on \( \Sigma^{\text{dec}} \). The correspondence is non-empty and convex-valued, and it satisfies the standard continuity conditions.

The intuition is that, under independence, degree \( k_i \) and degree \( k_i + 1 \) players have the same beliefs about the degree of each of their neighbors. If the \( k_i + 1 \)th neighbor is choosing \( m_i = 0 \), then because Property 2 holds in our game, the degree \( k_i + 1 \) players still will choose the same best response as the degree \( k_i \) player. If the \( k_i + 1 \)th neighbor chooses \( m_i = 1 \), then strict strategic substitutes imply that the degree \( k_i + 1 \) players’ best response is with a lower action. From Proposition 1, we know there is a unique symmetric equilibrium strategy \( \sigma \) involving thresholds in the information network game.

Proposition 1 has very clear implications. The player’s equilibrium action is weakly decreasing in her degree. In other words, the more neighbors she has, the less willing she will be to acquire information. Higher degree players expect that they will receive more information from their neighbors; thus, they have less incentive to acquire costly information by themselves. The players can “free-ride” on the acquiring actions of their neighbors. If player \( i \) has more neighbors, she is more likely to benefit from the signals passed by her neighbors. Because the marginal effects of signals in forecasting are decreasing, players with more neighbors are less willing to acquire costly information.

Corollary 1. The network-embedded prediction market displays positive externalities; thus, in every non-increasing symmetric equilibrium, the expected payoffs are non-decreasing in degree.

Proof: The result directly follows from Proposition 1.

In the corollary, we emphasize that in our network-embedded prediction market, players with more neighbors earn higher payoffs under the appropriate monotone equilibrium. Here, higher degree players exert lower effort, but they earn a higher payoff compared to their less connected peers because of free riding. The non-increasing property of equilibrium actions implies that social connections create personal advantage. In the network-embedded prediction market, well-connected players earn more than poorly connected players.

Simulation Results

We conduct a variety of agent-based simulations of agents trading in prediction markets to examine the forecasting performance of the Twitter-embedded prediction market. In every simulation round, we generate a random social network with 100 agents, using a 100×100-dimensional matrix. The probability of a link between two agents is \( p=0.7 \). Without loss of generality, we set the realization of future event \( v=50 \), the common prior

\[
V \sim N \left( V_o, 1/\rho_v \right) = N(40,100) ,
\]

and the noise of the signal \( \varepsilon_i \sim N \left( 0, 1/\rho_\varepsilon \right) = N(0,100) \).
For each cost level of information acquisition, we do 100 simulations. Figure 1 shows the dispersion of prediction market prices at different cost levels. The implication is similar to the efficient market hypothesis. As the cost of information decreases and more and more agents acquire information, the prediction market prices fully incorporate all available information, and the price dispersion becomes smaller; the prediction market yields good forecasts around $\nu=50$. It demonstrates that when threshold $k'$ is high (which implies that more agents acquire information), the forecasting performance of the network-embedded prediction market is better. Figure 1 reveals an interesting pattern: When only a small proportion of agents acquires information (because of the high cost to do so), an agent would receive very little information from her neighbors, and she would mainly use the prior information to generate a forecast. Therefore, the price dispersion is small at first and is around the mean of the prior $V_0 = 40$. As more and more agents acquire information, the agent receives differential information from the network, and differences of opinion occur. When most of the agents acquire information, the information is transmitted through the network, and the price dispersion becomes small again.

In Figure 2, the forecasting errors decrease with the cost to acquire information in our network-embedded prediction market. For prediction accuracy, we use the measure of standard forecasting error, which is
\( \sqrt{\sum_{j=1}^{100} (P_j - v)^2} \), and \( P_j \) is the price of every simulation round. We also find that the gain from reducing the prediction error is not significant as the cost decreases. The policy implication is important and straightforward. When the cost to acquire information is not high, a reduction in cost does not have a significant effect on market efficiency.

\[ P_{SN} = \frac{\rho_v}{3\rho_e + \rho_v}V_0 + \frac{\rho_e}{3\rho_e + \rho_v}(S_1 + S_2 + S_3), \]

and the forecasting errors by:

\[ Var[P_{SN} - V] = \frac{1}{3\rho_e + \rho_v}. \]

Now let's consider a non-networked prediction market. Because \( MV_3 > c > MV_4 \), each player acquires information in a non-networked prediction market, and the best forecast for player \( i \) is

\[ \frac{\rho_v}{\rho_e + \rho_v}V + \frac{\rho_e}{\rho_e + \rho_v}S_i. \]

The price can be obtained by the market-clearing condition:

\[ P_n = \frac{\rho_v}{\rho_e + \rho_v}V + \frac{\rho_e}{\rho_e + \rho_v} \frac{1}{n} \sum_{i=1}^{n} S_i. \]
and the forecasting errors

\[ \text{Var} [P_N - V] = \frac{\rho_e + n \rho_v}{n (\rho_e + \rho_v)}. \]

When

\[ \rho_v > \frac{1}{\sqrt{n}} \rho_e, \]

we can obtain:

\[ \text{Var} [P_N - V] > \text{Var} [P_{SN} - V]. \]

In other words, when \( n \) is sufficiently large, the forecasting errors in the network-embedded prediction market are smaller than those in the non-networked prediction market. Example 1 thus shows that the network-embedded prediction market outperforms the non-networked prediction market in the circle network.

**The Design of Prediction Markets**

Because the network-embedded prediction market could outperform the non-networked prediction market, a carefully designed network-based prediction market could generate more precise predictions. In this section, we design a Twitter-based prediction market using a pari-mutuel betting mechanism. In the theoretical model, players are linked to each other as in a social network. In our information system design, a social network refers to a Twitter network. When we began our study, 2009 data showed Twitter to be one of the fastest growing social network sites, with 105 million registered users by April 2010. Through Twitter, users post and read messages known as tweets, which are text-based posts of up to 140 characters. In the model, players exchange the information with their neighbors. In the same way, the information sharing in Twitter relies on free user contributions. On Twitter—a micro-blogging site that allows users to post short messages to their followers—people can find opinions and information on a broad range of topics posted by their peers almost in real time. “Neighbors” in the theoretical model can be interpreted as followers in the Twitter network.

The power of prediction markets derives from the fact that the mechanisms provide incentives for truthful revelation, as well as for research and information discovery, and the market provides an algorithm for aggregating opinions. Blume and Easley (1992, 2006) discuss how a Bayesian learner updates his or her belief over a long period of time. As time goes to infinity, the Bayesian learner assigns a posterior probability of one to the correct hypothesis. However, our question focuses on how we can elicit dispersed information from a large number of people in finite periods. Fang, Stinchcombe, and Whinston (2010),
generalizing their previous results, show that in very general conditions, some strict scoring rules exist that elicit the expert’s true beliefs as probabilistic forecasts.

Many different prediction market mechanisms are available, including double auction, pari-mutuel betting, and a market maker mechanism. For our purposes in the Twitter-based prediction markets, we adopt a pari-mutuel betting mechanism. This option is widely used in sports betting; for instance, all betting on horse races in the United States is pari-mutuel. Here, participants buy bets on each of the S possible states. Each bet costs one virtual dollar, and a participant spends all of his or her virtual endowment on buying bets. The total number of bets of each type that have been purchased is displayed publicly. The odds are updated using budget balance conditions after each transaction. If \( W \) is the total budget and \( h_s \) is the total quantity of state-\( s \) bets purchased, then the odds for state \( s \) are given by \( O_s = W / h_s \), which implies that if a participant buys ten $1 bets on state \( s \), he will receive \( 10 \cdot O_s \) dollars when the realized state is \( s \).

Generally speaking, in pari-mutuel betting the total amount of money bet on all outcomes is placed in a common pool. The money is returned to those who bet on the winning outcome. Pari-mutuel betting differs from fixed-odds betting in that the odds are updated in real time, and the final payout is not determined until the market is closed.

**Embedding Twitter into the Prediction Market**

To combine the popularity of Twitter with the power of a prediction market, we set up a master account to follow all the participants of the prediction market on Twitter and designed a simple and brief language that participants could use to place bets and to query information by sending tweets. We also collected participants’ demographic information using a pre-experiment survey. (See Appendix A, Pre-Experiment Survey.) One possible use of our Twitter-based prediction market—called IBET—is to forecast movie box office revenues, as shown in Figure 5.

![Figure 5. A Twitter-Based Prediction Market](image)

Figure 6 shows the betting intervals and odds for movies. If you think “Sucker Punch” is an excellent movie, you can bet most of your virtual money on interval 4, which means you predict the box office revenue is above $40 million. Like the traditional prediction market, we still offer a website that allows participants to undertake their betting through the web interface. However, with our Twitter-integrated system, Twitter users can easily participate in our prediction market simply by sending us a direct message. When people are waiting or chatting and they come across some new information, they can take out their smart phones and place their bets accordingly. The purpose of such an easily accessed mechanism is to induce people to provide the information as soon as they have time and before they
forget it. As more people participate and provide more information in a more timely manner, our prediction market potentially could outperform the traditional prediction markets.

For example, a participant sends a direct message saying, "bet 10 1 green hornet," which means bet $10 on interval 1 for the movie, "Green Hornet." Our master account receives this tweet through Twitter, processes it, and then updates the participant's betting portfolio accordingly. On the other hand, if a participant sends a direct message saying, "query portfolio green hornet," our master account sends a message back to the participant with her betting portfolio information. Similarly, users can query their balance by sending a message.

Our system has to be able to do some natural language processing to "understand" the instruction from the participants. Because of the complexity and challenge of enabling that processing, we have designed the language between the participants and our master account to be simple and straightforward. Figure 7 shows the prediction market result for the movie, "Hall Pass." The amount bet on interval 1 is $194; on interval 2, $474; on interval 3, $238; on interval 4, $70; and on interval 5, $10. Therefore, participants think the box office revenue is more likely to be between $10 million and $20 million. The true opening
weekend box office revenue was $13,535,374, so the Twitter-based prediction market did a pretty good job. The use of networked markets affects information acquisition and improves the efficiency of prediction markets.

**Conclusions**

The Twitter-embedded prediction market can induce people to provide the information as soon as they have the time and before they forget it. Given that sending direct messages from smart phones is a more convenient mode of communication, we expect that, with more people participating and providing more information in a more timely manner, our prediction market potentially can outperform other prediction markets. Simulation results show that the Twitter-based prediction markets produce better predictions than the prediction markets without social networks.

Using the embedded prediction markets also allows us to conduct a field experiment to study social learning and to measure the social distance of the participants using network data. Coval and Moskowitz (2001) find that social networks help fund managers earn above-normal returns in nearby investments: The average fund manager generates an additional 2.67 percent return per year from local investments, relative to nonlocal holdings. Their results suggest that investors trade local securities at an informational advantage in the social networks, indicating a strong geographic link between mutual fund investment and performance. Generally speaking, investors in the same local, social network structure would exhibit similar local holdings and local performance.

Coval and Moskowitz (2001) assume that investors located near a firm can visit the firm's operations, talk to suppliers and employees, and gain access to private information. For example, mutual fund managers and local corporate executives may run in the same circles, belong to the same service or social organizations, and so forth. The close social networks affect the performance of fund managers significantly. However, Coval and Moskowitz did not have data to measure the networks and social distance between investors and local corporate executives. Still unclear is how social learning of local investors occurs and how it varies in different social networks. Our Twitter-based prediction market exhibits similar characteristics to these social networks. As we collect more and more data from the embedded prediction markets, our future work is to test whether participants in the same local structure exhibit similar bets and prediction performance.

Twitter-based prediction markets also provide us with a field experiment to test the famous "strength of weak ties" hypothesis (Granovetter, 1973). The gist of the hypothesis is that we always get truly new information from acquaintances, rather than from close friends. The groups with which we have strong ties, although they are filled with people eager to help, are also filled with people who know roughly the same things we do. Thus, strong ties usually result in informational redundancy. Weak ties, meanwhile, are much more valuable in terms of contributing genuinely new information.

It is worthy of pointing out that our network framework can also be applied to analyze strategic complements: My friend's adoption of a product raises the marginal return of my adoption. People will have an incentive to be on the sites their friends are using, even when large parts of the rest of the world are using something else. Thus, the basic framework of our paper is also applicable in viral marketing.

Another future research direction is to examine the incentives for sharing information in a social network. In the current setup, we assume people exchange information according to reciprocity and norms of fairness, rather than focusing on the incentives for sharing information. However, whether people have incentives to share information, and what those incentives are, remain open questions. A well-developed literature of "value of information" and "information sharing" has been generated in the setup without networks. For example, Gal-Or (1985) shows that no information sharing is the unique symmetric equilibrium of the game in an oligopolistic market. In addition, Prendergast (1993) demonstrates the disadvantages of shared information: Individuals tend to conform to the opinions of others in a way that can be inefficient. Studying the incentives for sharing information in a social network would be an interesting extension of information sharing.
References


Economics and Value of Information Systems


Appendix A: Pre-Experiment Survey

We collected players' demographic information by a pre-experiment survey.

1. What is your gender?
   a. Female   b. Male

2. How old are you?
   a. under 15   b. 15-19   c. 20-29   d. 30-39   e. 40-49   f. 50 or above

3. How many years of formal education have you completed? (For example, if you have completed high-school, you may write 12; if you have completed college, you may write 16)

4. Please select the occupational category below that best describes your profession:
   a. Executive, administrative, and managerial occupations
   b. Engineers and engineering technicians
   c. Teachers and instructors, counselors, and college faculty
   d. Lawyers and judicial workers or health diagnosticians
   e. Marketing and sales occupations
   f. Service occupations
   g. Students
   h. Others

5. What is your email address?

6. How much time do you think you will devote to playing the prediction game per day?
   a. less than 5 minutes   b. 5 - 10 minutes   c. 10 minutes - 30 minutes   d. 30 minutes - 1 hour   e. more than 1 hour

Thank you for completing this survey! You will now be redirected back to the prediction game.