Socially Embedded Prediction Markets

Yiling Chen   David M. Pennock

Yahoo! Research

45 West 18th Street, 6th Floor

New York, NY 10011

{cheny, pennockd}@yahoo-inc.com

Abstract

We propose a model of prediction markets where participants are biased according to their social relationships. We relax the standard assumption of complete rationality and adopt an arguably more realistic model where agents are disproportionally influenced by their neighbors in a social network. We conduct extensive agent-based simulations of our model. We find that prices in prediction markets remain accurate even when participants are biased and irrational. Moreover, accuracy is robust to changes in many factors, including how individuals are motivated to participate in the market, the way that individuals use public information, individual utility functions, the topology of the social network, and the strength of social influences. Our model can explain the high volume of trade often observed in speculative markets that is hard or impossible to explain under standard market rationality assumptions. Our model can also explain the documented ability of prediction markets to succeed even in the face of biased and irrational participants.

Keywords: prediction market, social network, biased agent, multi-agent simulation, bounded rationality.

1 Introduction

A prediction market is a type of financial market designed to elicit a forecast, for example the probability that a Democratic Party candidate wins the 2008 US presidential election. Participants buy and sell contracts that are tied to the outcome of the forecast variable, for example contracts that pay \$1 if and only if a Democrat actually wins the election. The price that each participant is willing to pay—some amount less than \$1—reflects her estimated probability of the outcome, and the market's equilibrium price reflects the consensus judgment among all participating agents.

In a broad and diverse number of settings in the lab and in the field, prediction markets seem to yield equal or better forecasts than other methods. For example, markets like the Iowa Electronic Markets (IEM)[1] predict political election outcomes better than polls [11, 12, 23, 3, 4]. Futures, options, and other derivative financial markets rapidly incorporate information, providing accurate forecasts of their underlying commodities or securities [17, 31]. Sports betting markets provide accurate forecasts of game outcomes [13, 35, 8, 32]. Laboratory experiments confirm information aggregation [28, 29, 30, 10, 6]. Field tests at Hewlett-Packard beat managerial forecasts 15 out of 16 times [7]; similar tests at Google[2] and Microsoft show promise. Even play-money market games like the Hollywood Stock Exchange[3] and NewsFutures.com can yield accurate forecasts [33, 26].

While the empirical success of prediction markets is encouraging, there remains some gap between theory and practice. Starting from the so-called *efficient market hypothesis* [9], many authors have shown that under ideal conditions market prices fully incorporate all available information, properly accounting even for dependent information [14, 19, 22]. Standard theories assume that all participants share common prior probabilities and perform perfect Bayesian updating on a joint probability distribution over the outcome variable(s) crossed with all possible realizations of evidence variables among all participants. The theories assume that rationality is *common knowledge* (i.e., all participants are rational, know that everyone else is rational, know that everyone else knows that everyone is rational, etc.). In contrast, real prediction markets like IEM are populated by a non-representative sample of mistake-prone

[1] http://www.biz.uiowa.edu/iem/

[2] http://googleblog.blogspot.com/2005/09/putting-crowd-wisdom-to-work.html

[3] http://www.hsx.com/about/press/060307.htm

and biased traders [11, 23]. For example, political persuasions do seem to unduly effect participants' opinions[4] about election outcomes. Traders often behave irrationality by leaving arbitrage opportunities on the table or not choosing the best available price among equivalent options [23]. Despite these mistakes and biases at the individual level, the aggregate view at the market level seems to reflect an accurate and relatively unbiased picture. Hanson [15] shows that in a market microstructure model the presence of one type of biased agent—namely a price manipulator—can actually improve overall accuracy of the market. Still, most theory is at a loss to explain the observed robustness of prediction markets to individual biases.

Moreover, so-called *no-trade theorems* [21] assert that a group of entirely rational participants should never engage in speculative trade,[5] a consequence that seems strikingly at odds with reality, where a wide variety of functioning prediction markets exhibit quite high volumes of trade. Some market microstructure theories circumvent the no-trade dilemma by assuming a certain proportion of *noise traders* that behave erratically [18]. Other theories sidestep the issue by asserting that participants have different prior probabilities, or other unexplained sources of differences of opinion [37, 27]. In this paper, we propose that disagreements among participants persist due to mutual reinforcement along social connections.

We model prediction markets together with the social context in which they are embedded. People are linked to one another according to a social network. The information each person possesses is partially correlated with her neighbors in the social network. We institute a form of bounded rationality by presuming that individuals are not fully aware of the bias in their neighborhoods. Instead, they (incorrectly) treat information available in their neighborhoods as a representative sample of global information. We believe this abstraction captures some of the conventional wisdom about how people naturally cluster according to similarity of beliefs, yet tend to see their corner of the world as more representative than perhaps it is. For example, voters in a heavily Democratic region of the country whose friends are mostly Democratic may overweight the probability of a Democratic candidate winning the election. In fact, evidence of such biases appear in surveys of IEM participants [11, 12].

We conduct a variety of simulations of agents in our model trading in a prediction market. We show that the

[4]We will use the terms "information", "belief", and "opinion" interchangeably, depending on context.

[5]The no-trade argument goes loosely as follows: A rational agent should know that another rational agent's willingness to trade signals a new piece of information that revises the first agent's beliefs, eliminating the utility of actually executing the trade.

forecast accuracy of market prices remains robust across a variety of factors, including how agents are motivated to participate in the market, influences of public information, differences in agent utility functions, the topology of the social network, and the strength of social influences. In almost all cases, the market predictions are more accurate than small sample opinion polls. This reliability across so many factors appears as a positive attribute for prediction markets, and may help explain why, in practice, markets seem to function well even in the face of participants' mistakes, biases, and irrationality. Our simulations exhibit significant trading volumes driven by differences of opinion reinforced through social proximity. We argue that social biases offer a natural explanation of high trading volumes observed in the field.

The remainder of the paper is organized as follows. Section 2 discusses properties of social networks that are relevant to our model and reviews related network evolution models. In Section 3, we introduce our model of prediction markets. Simulation results are presented in Section 4. Section 5 concludes our findings.

2 Properties and Models of Social Networks

A social network is a social structure that indicates the ways in which individuals, organizations, or other social entities are connected through various social familiarities. Usually, graphs are used to represent social networks, with nodes representing social entities and edges representing social relations. Social relations are so universal that almost all naturally-evolved networks can be considered social networks in a broad sense. Every individual is part of some social network, depending on how social relations are defined.

When it comes to information about some future event, it is natural to think that information is distributed over a social network. A node in such a social network represents an individual with a piece of information or personal opinion about the future event. An (undirected) edge represents that the connected two nodes have access to each others' information.

Studies have found some common properties of social networks, including small diameter [20], power law distribution of degree [16], and local clustering of edges [36]. For the purpose of this paper, we discuss two properties of social networks and review some generative network models.

2.1 Power Law Degree Distribution and Generative Models

It has been found that the connectivity or degree distribution of naturally evolved networks approximately conforms to the power law [16]. The number of nodes whose degree is k (i.e. nodes that have k neighbors) is inversely proportional to the degree k raised to a constant power. That is,

$$f_k \propto k^{-\gamma}, \tag{1}$$

where f_k is the frequency of nodes of degree k, and γ is a constant usually close to one. A power law degree distribution indicates that a few individuals have many social relations, while most individuals only have a few social relations. Due to this property of social networks, accessibility of information in a society should display a similar characteristic—a few individuals have access to a lot of information via their social relations, but most individuals only have access to a little information.

There has been an explosion of interest in statistical generative models that explain the observed power law distribution of degree. Not intending to be a complete review, we briefly introduce three network evolution models that we believe are representative. Readers can refer to [24] for a more complete comparison of network models. All three models introduced below are evolution models. They all add one node and m edges to the existing network at each time step. But they differ in how to choose endpoints of new edges.

One of the simplest network generation models is the exponential growth model [24] and their variations [5]. At each time step, the m edges all connect from the new node and end with existing nodes that are randomly chosen according to a uniform distribution. Thus, at time t, the probability for an existing node i to be chosen as an edge endpoint is

$$\Pi(i,t) = \frac{1}{|V(t)|}, \tag{2}$$

where $|V(t)|$ is the total number of existing nodes at time t. This way of selecting destination nodes is called *uniform attachment*. Although nodes are randomly selected at each stage, older nodes tend to have more edges due to the growth of the system. But the degree distribution for exponentially grown network deviates from the power law.

Barabási and Albert [2] use *preferential attachment* to generate networks that conform to observed power law distribution. New edges still connect from the new node and end with existing nodes. But preferential attachment requires that the probability for an existing node to be chosen as an edge endpoint is proportional to its degree. That is,

$$\Pi(i,t) = \frac{k_i(t)}{\sum_j k_j(t)}, \tag{3}$$

where $k_i(t)$ is the degree of existing node i at time t. Preferential attachment captures the power law degree distribution in a simple and explicit way, but it is rigid in the sense that it can not fit in different power law exponents. Hence, many generalizations of the model have been proposed [1, 25].

One extension of Barabási and Albert [2]'s model is what is called by Park et. al. [24] the "Pretty Good" (PG) model. PG model, proposed by Pennock et. al. [25], uses a mixture of preferential attachment and uniform attachment when adding edges to the network. More specifically, the probability that a node is chosen to be an edge endpoint is

$$\Pi(i,t) = \alpha \frac{k_i(t)}{\sum_j k_j(t)} + (1-\alpha)\frac{1}{|V(t)|}, \tag{4}$$

where α ranges between 0 and 1. The additional degree of freedom provided by the parameter α provides the PG model with more flexibility to fit different power law exponents and divergences from the strict power law distribution that are seen in practice. The parameter α controls the degree of preferential attachment versus uniform attachment in the growth model.

2.2 Locally Clustered Beliefs

Much evidence has indicated that people have a herding instinct—a social tendency to follow the behaviors or beliefs of a larger group of individuals with whom they identify [34]. Thus, an individual is more likely to have the same or similar beliefs as the majority of his/her neighbors in the social network. As a result of such social influences, individual beliefs are locally clustered. People who know each other tend to have similar beliefs. A good example of this property is that political beliefs or opinions on various policy issues tend to exhibit geographical clusters.

The above mentioned three network evolution models focus on generating networks that are topologically similar to real world networks. However, they do not address distribution of beliefs. We introduce our heuristics for generating locally clustered beliefs in the next section.

3 A Social Network Model of Prediction Markets

In this section, we propose our model of prediction markets that formalizes some realistic social aspects of information structure and individual behavior. We will present our model in the setting of a U.S. presidential election prediction market, assuming that the event we are predicting is whether a political party, for example Democratic Party, will win the presidential election. Our model can easily fit into other settings of winner-takes-all markets for predicting binary events.

3.1 The Social Network of Information

Suppose that there are N individuals in the society, each having a piece of binary information, denoted as $s_i \in \{0, 1\}$ for individual i. Intuitively, s_i can be interpreted in the presidential election setting as individual i's tendency to vote for a candidate. Without lose of generality, we assume that s_i equals 1 if individual i plans to vote for the Democratic candidate, and 0 if the individual favors the Republican candidate. In the absence of social influences, s_i's are independently drawn from a Bernoulli distribution,

$$s_i \sim Bernoulli(p), \tag{5}$$

where p is the probability that s_i will be 1 and is unknown to all individuals. Individuals, together with their information, are embedded in a social network. Each node in the network represents an individual. The label of the node, 0 or 1, is the information of the individual. For simplicity, we refer to nodes with label 0 as 0-nodes and nodes with label 1 as 1-nodes. An edge simply means that the connected two individuals have some social relation. We further assume that any two connected individuals have access to each other's information. Hence, the graph for the social network is undirected.

The topology of the network is generated via a network evolution model. At each time step t, we add a new labeled node and m edges to the network until all N nodes are in the network. The label of the new node is determined by the Bernoulli distribution specified in (5). All m edges connect from the new node. The other endpoint of a new edge is selected from existing nodes according to a mixture of preferential attachment and uniform attachment. The probability that an existing node i will be selected as an endpoint of a new edge at time t is defined by equation (4) in

Section 2.

We model the effect of social influences in the network as the likelihood for a node to change its label. Changing label for a node indicates that the individual changes his/her mind on whom to vote for. Social influences imply that an individual is more likely to convert when his/her belief is different from the majority of his/her neighbors. We use the following heuristics to generate clustered beliefs resulting from social influences.

1. Calculate the percentage of different beliefs in one's neighborhood for each node. For a 0-node, calculate the percentage of 1-nodes in his neighborhood. For a 1-node, calculate the percentage of 0-nodes in his neighborhood.

2. Sort 0-nodes and 1-nodes in descending order respectively, according to the percentage of different beliefs in their neighborhoods.

3. Swap labels for the highest-ranked 0-node and the highest-ranked 1-node.

4. Repeat steps 1 to 3 l times, or until the two nodes being swapped no longer change.

Parameter l allows our model to adjust to different degrees of belief clustering in a social network. After applying the above heuristics, individual information is no longer independent, but correlated.

3.2 Prediction Market and Individual Behavior

A prediction market in the society trades a contract that pays \$1 per share if more than 50% of the nodes are 1-nodes, and pays \$0 otherwise. The contract corresponds to a winner-takes-all security in a presidental election prediction market, which pays off \$1 if and only if the Democratic candidate collects the majority of votes in the election.

An individual i who participates in the prediction market uses information in his neighborhood to make a guess for the prior p in the Bernoulli distribution (5), denoted as p_i. The neighborhood of an individual includes himself and all individuals who are connected with him. Individuals are biased—they treat information in their neighborhoods as an objective representation of all information in the society. Thus,

$$p_i = \frac{\text{Number of 1-nodes in i's neighborhood}}{\text{Total number of nodes in i's neighborhood}}. \tag{6}$$

For individual i, the expected value of the contract in the prediction market, is simply the probability that there are more than N/2 successes for a Binomial(N, p_i) distribution. Thus, individual i's belief about the probability that future contract payoff will be \$1 is,

$$v_i = \sum_{l=\lfloor N/2 \rfloor+1}^{N} \left(\begin{array}{c} N \\ l \end{array} \right) (p_i)^l (1-p_i)^{N-l}. \tag{7}$$

Individuals might also consider some public statistic in addition to their local information. Hence, their belief is a weighted average of v_i and the public statistic,

$$\pi_i = w_i v_i + (1 - w_i)g, \tag{8}$$

where g is some public statistic such as market price or expected contract value derived from some public information. w_i is between 0 and 1. When w_i equals 1, individual i only uses local information. Market participants will use their π_i's to make trading decisions in the prediction market so as to maximize their utility.

We consider risk-averse individuals with either negative exponential utility (CARA) or generalized logarithmic utility (GLU). If an individual i has negative exponential utility function for money,

$$u_i(y) = -e^{-c_i y}, \tag{9}$$

it means that the individual exhibits *constant absolute risk aversion (CARA)*, hence the utility function is also called CARA utility. c_i is agent i's absolute risk aversion coefficient. On the other hand, an individual i with GLU,

$$u_i(y) = \ln(y + b_i) \;\; (b_i > 0), \tag{10}$$

has decreasing absolute risk aversion. b_i usually is interpreted as individual i's initial wealth.

Suppose there is no credit limit and market participants can short sell. Market participant i will maximize his/her expected utility,

$$\max \quad U_i(x_i) = \pi_i u_i((1 - P)x_i) + (1 - \pi_i)u_i(-Px_i), \tag{11}$$

where x_i is participant i's demand (supply if negative), and P is the price of contract.

We can compute the equilibrium demand and prices according to Pennock and Wellman [27]. If all market participants have CARA utility, solving the optimization problem of (11) results in a unique demand function for participant

i,

$$x_i(P) = \frac{1}{c_i} \ln(\frac{\pi_i}{P} \cdot \frac{1-P}{1-\pi_i}) \tag{12}$$

At competitive equilibrium when market clears, equilibrium market price for the contract is

$$P^* = \frac{\prod_{j=1}^{M} \pi_j^{\delta_j}}{\prod_{j=1}^{M} \pi_j^{\delta_j} + \prod_{j=1}^{M} (1-\pi_j)^{\delta_j}}, \tag{13}$$

where $\delta_j = \frac{1/c_j}{\sum_{l=1}^{M}(1/c_l)}$ is the normalized risk tolerance of individual j, and M is the total number of people participating in the prediction market. Equilibrium price with CARA participants is a weighted geometric mean of every participants' beliefs.

Similarly, if all market participants have GLU utility, demand function for participant i is

$$x_i(P) = b_i(\frac{\pi_i}{P} - \frac{1-\pi_i}{1-P}). \tag{14}$$

Equilibrium market price is

$$P^* = \sum_{j=1}^{M} \beta_j \pi_j, \tag{15}$$

where $\beta_j = \frac{b_j}{\sum_{l=1}^{M} b_l}$ is the normalized initial wealth of participant j. Equilibrium price with GLU agents is a weighted arithmetic mean of every participants' beliefs. For both markets with CARA participants and GLU participants, trading volume at equilibrium is

$$V(P^*) = \sum_{j \in \{l:\, x_l(P^*)>0)\}} x_j(P^*). \tag{16}$$

4 Simulation Results

Agent-based simulation of our model is used to examine price accuracy and trading volumes in prediction markets. In every simulation round, we generate a social network with 101 nodes, i.e. $N = 101$. When generating the network, a new node and 3 edges are added to the network at each step. The probability for a node to have a label 1 is 0.55, i.e. $p = 0.55$ in (5). Thus, in our simulations, the real probability that contract payoff will be \$1 or the real value of the contract is

$$v = \sum_{l=51}^{101} \begin{pmatrix} 101 \\ l \end{pmatrix} (0.55)^l (1-0.55)^{101-l} = 0.8438. \tag{17}$$

Except in Section 4.5, all results presented use networks that grow solely via preferential attachment, i.e. $\alpha = 1$, and have the highest degree of belief clustering. For prediction accuracy of markets, we use the measure of absolute error, which is $|P^* - v|$. The lower the absolute error, the more accurate the prediction is. All results reported below are averaged over 1000 simulation rounds, thus we use the term "mean absolute error of probability" in our figures for prediction accuracy.

4.1 Local Information and Motivation to Participate

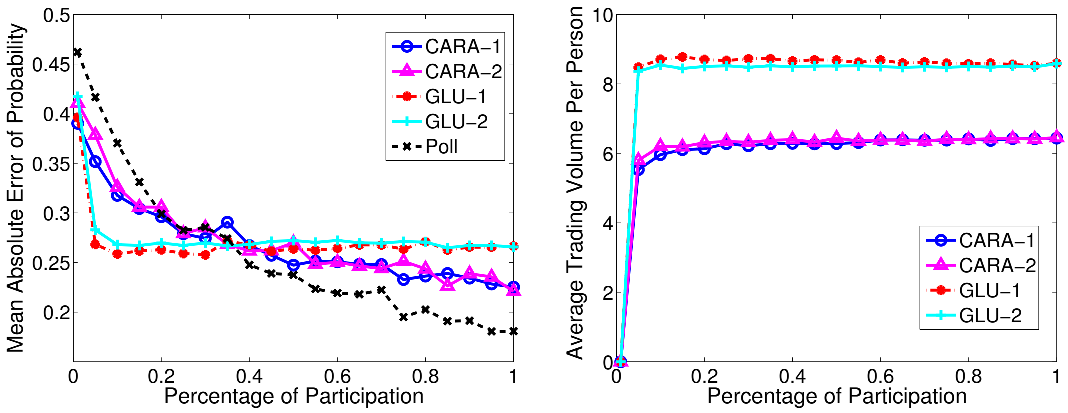
We consider two cases of how agents are motivated to participate in the market. First, individuals who can access more information in the social network are more motivated to participate in the prediction market. Thus, the probability for an individual i to participate in the prediction market is modeled as

$$q_i = \frac{k_i}{\sum_j k_j}, \tag{18}$$

where k_i is the degree of node i. Second, individuals are equally likely to participate in the market regardless of how much information they have.

When market participants only use their local information in forming their beliefs, π_i equals v_i. We compare the prediction accuracy of the market with the prediction accuracy offered by poll when there are the same number of participants in the market as in the poll. For prediction markets, participants are chosen from the population either according to the probability distribution defined in (18) or according to a uniform distribution. Polls select individuals to participate according to a uniform distribution. For a particular poll, the fraction of 1-nodes among all selected nodes is treated as the poll-believed prior p in the Bernoulli distribution (5). Then, the prediction offered by the poll can be calculated in the same way as in (7).

Figure 1 (a) shows the relationship between prediction errors and percentage of people who participate in market or poll. It demonstrates that when there are less participants (less than 40% of the total population), the prediction market is more accurate than the poll. Markets in which individuals have GLU utility outperform markets in which individuals have CARA utility. Figure 1 (b) indicates that trading volume is nonnegligible when there is more than one participant. Both prediction accuracy and trading volume are not significantly affected by how individuals are motivated to participate in the market. For the rest of the paper, we only present results for the case when an individual's likelihood to participate is proportional to the amount of information he/she has.



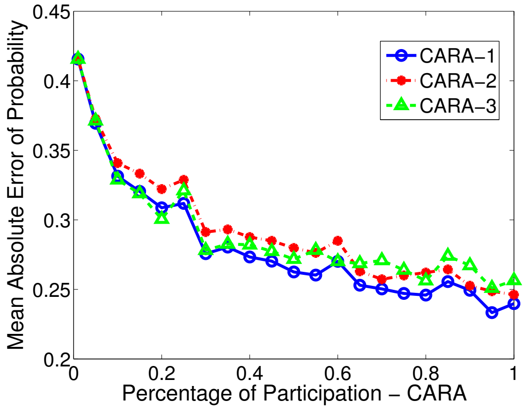
(a) Prediction Error

(b) Equilibrium Trading Volume

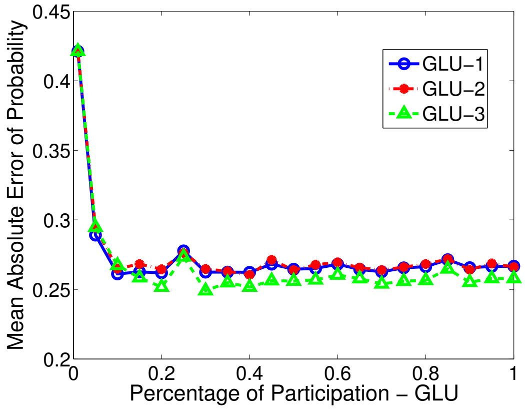
Figure 1: *Participants only use information in their neighborhoods to form their beliefs. X-axis is the percentage of the population that actually participate in the market or poll. The curves labeled CARA are for markets in which participants have CARA utility. The absolute risk aversion coefficient* c_i *for CARA is chosen according to Uniform(0.3, 1.7). The curves labeled GLU are for markets in which participants have GLU utility. The initial wealth* b_i *for GLU is chosen according to Uniform(3, 17). For CARA-1 and GLU-1, market participants are chosen from the population according to the distribution of (18). For CARA-2 and GLU-2, participants are chosen from the population according to a uniform distribution.*

4.2 Getting Feedback from Market Price

When market participants can use market price as a feedback and incorporate it into forming beliefs, how does it affect prediction accuracy and equilibrium trading volume? Figure 2 shows the results that linearly weighing between v_i and market price as defined in (8) results in similar prediction accuracy for both markets with CARA participants and markets with GLU participants. Weight w_i are either randomly chosen according to a uniform distribution or proportional to the amount of information that an individual has. We also found that average trading volume per agent increases with the differences of w_i's among agents. Limited by space, we omit the figures for trading volume.



(c) Prediction Error for CARA



(d) Prediction Error for GLU

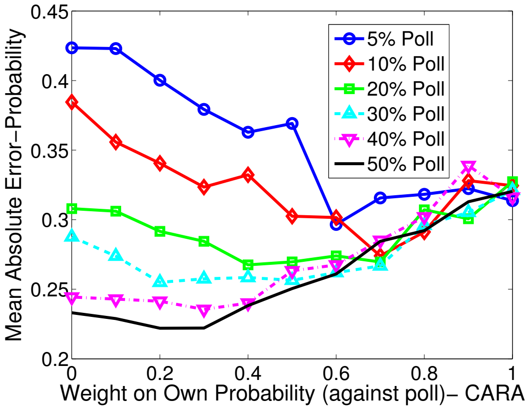
Figure 2: *Participants can use both information in their neighborhoods and market prices in forming their beliefs. CARA-1 and GLU-1: Participants only use information in their neighborhoods, same as in Figure 1. CARA-2 and GLU-2: Participants weight between prediction only based on local information and market price as defined in (8), where g is the market price. w_i ∼Uniform(0,1). CARA-3 and GLU-3: Participants weight between prediction only based on local information and market price. But w_i is proportional to the amount of information that i can access.*

4.3 Getting Public Information Outside the Market

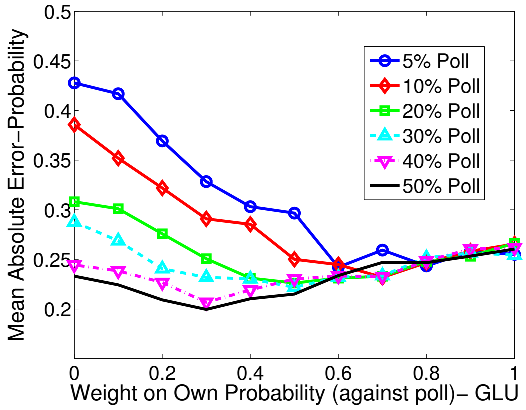
Suppose that a poll statistic is announced publicly, we attempted to simulate the situation when market participants use the poll information in addition to their local information to form beliefs. Figure 3 shows that linearly weighing between v_i and poll-implied prediction (8) may increase the price accuracy of prediction markets, if the poll is informative. If poll is not informative enough, for example if only 5% of population participates in the poll, incorporating poll information is not helpful. In addition, the more informative the poll is, the more weight individuals should put on poll-implied predictions (smaller w_i). Average trading volume monotonically increases with the weight on v_i, indicating that the more different beliefs of market participants are the higher the trading volume.

4.4 Different Risk Attitude or Initial Wealth

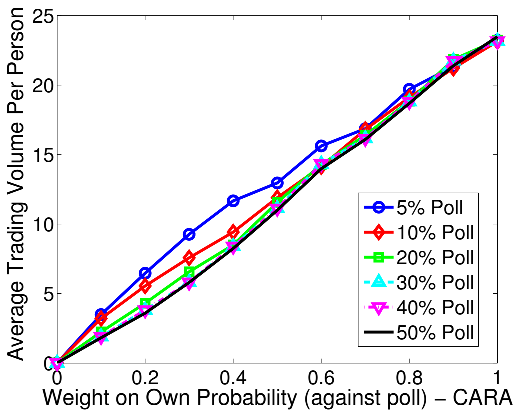
Simulating with different parameters of utility functions, we find that the prediction accuracy for markets with GLU participants is very robust to the distribution of initial wealth b_i's, as shown in Figure 4 (b). For markets with CARA



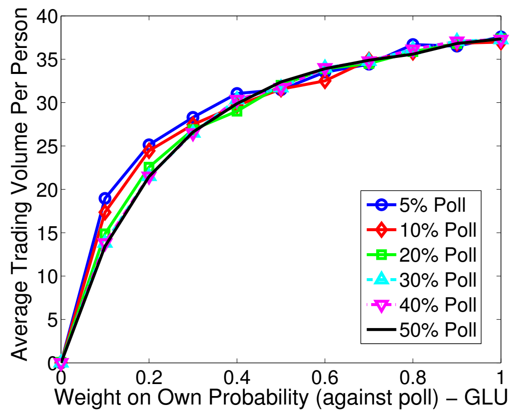
(a) Prediction Error for CARA



(b) Prediction Error for GLU



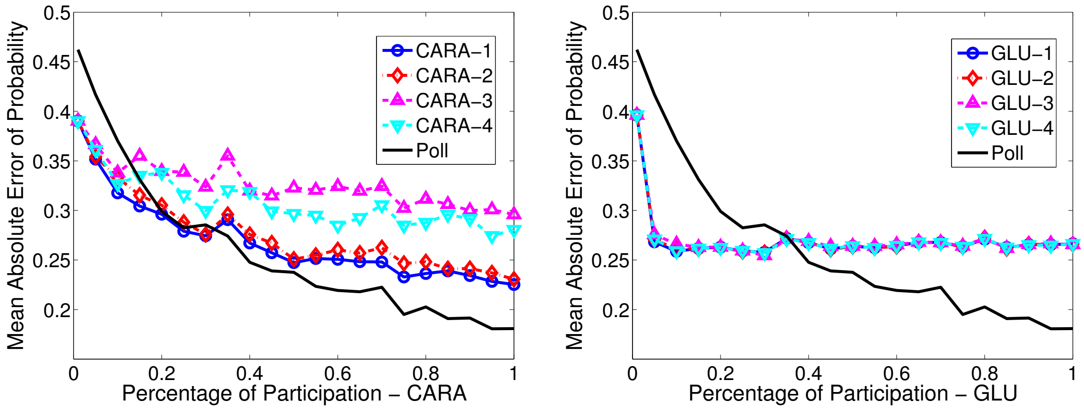
(c) Trading Volume for CARA



(d) Trading Volume for GLU

Figure 3: *Participants use both information in their neighborhoods and publicly announced poll results in forming their beliefs. Participants weight between prediction only based on local information and prediction implied by poll as in (8), where g is the prediction implied by poll. X-axis is changes of w_i. The curve labeled with a% means that the publicly announced poll has a% of total population participated. Results shown are from markets with 15 participants.*

participants, shown in Figure 4 (a), prediction accuracy is slightly affected by the distribution of risk averse coefficient c_i. There is a vague trend, showing that the more similar the risk attitude is, the more accurate the market predictions are. But for all simulated distributions of c_i, markets with CARA participants outperform polls when the percentage of participation is low.

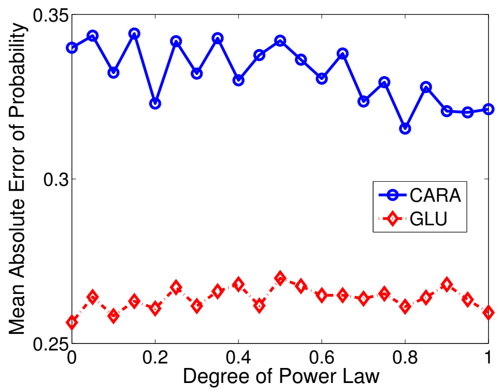


(a) Prediction Error for CARA (b) Prediction Error for GLU

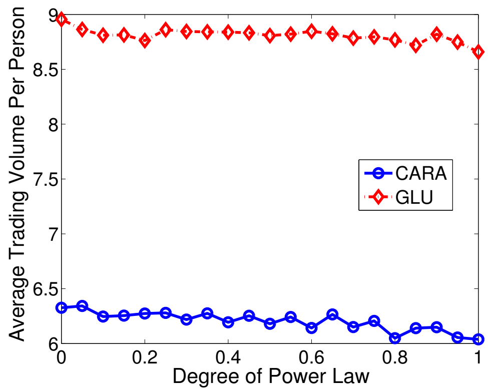
Figure 4: *Different utility parameters. CARA-1:* c_i*=1; GLU-1:* b_i*=10; CARA-2:* c_i *∼Uniform(0.3,1.7); GLU-2:* b_i *∼Uniform(3,17); CARA-3:* c_i *∼Uniform(0.01,2); GLU-3:* b_i *∼Uniform(3,50); CARA-4:* c_i *∼N(1,0.5); GLU-4:* b_i *∼N(10,2).*

4.5 Network Topology and Strength of Social Influences

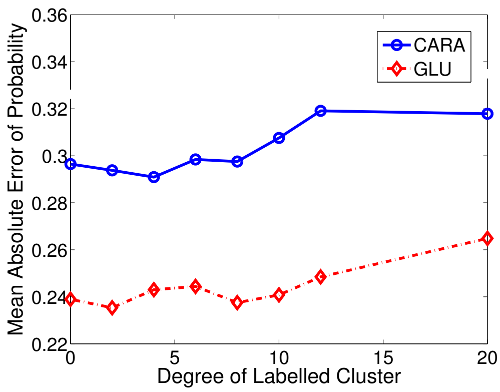
We model the information structure in a society as a social network with clustered beliefs due to social influences. It is natural to ask whether the topology of the network and different degrees of belief clusters will affect price accuracy and trading volumes in prediction markets. Figure 5 (a) and (b) illustrate that both price accuracy and trading volume are stable relative to the changes in topology. But price becomes slightly less accurate and trading volume increases when beliefs become more clustered as shown in Figure 5 (c) and (d). The decreased price accuracy is probably due to increased individual biases as beliefs become more clustered. The increase in trading volume further confirms that there is more trading when individuals disagree more. Market prices are still more accurate than small polls.



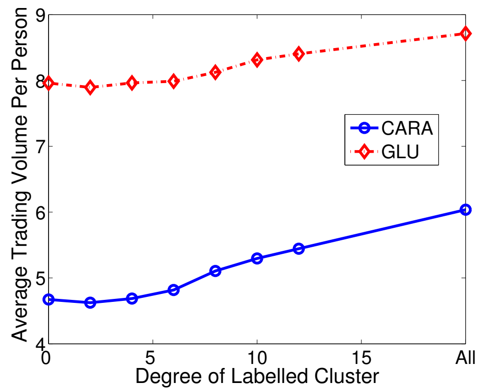
(a) Prediction Error



(b) Trading Volume



c) Prediction Error



(d) Trading Volume

Figure 5: *Effect of Network Structure on prediction error and trading volume. Results shown are from markets with 15 participants. (a) and (b): Effect of topology. X-axis is the parameter α in (4). The larger the α is, the more weight is put on the preferential attachment. (c) and (d): Effect of belief clusters. X-axis is the round parameter l in our heuristics to form clustered beliefs. The larger the l is, the more clustered beliefs are. "All" is corresponding to the time when the swapped two nodes no longer change.*

5 Conclusions

We model prediction markets together with the social context in which they are embedded. Information related to an event is represented as distributed in a social network with power law distribution of degree. Social influences over the network result in belief clusters. Members of the social network participate in a winner-takes-all prediction market for a binary uncertain event. Individuals are socially biased, treating accessible local information in their social neighborhoods as a representative sample of global information and make trading decisions accordingly. Our model captures some realistic social aspects of information structure and individual behavior.

With simulation, our model provides an explanation for the observed prediction accuracy and high trading volume in real-world prediction markets, for which existing theory has not given satisfactory explanations. Specifically, we have shown that:

- With biased social individuals, market price is still an accurate forecast, at least more accurate than polls within a reasonable range.
- Unlike the no-trade theorems, at equilibrium markets generate a significant amount of trading as a result of the bias of participants.
- Price accuracy is very robust to how individuals are motivated to participate in the market, the way that individuals use market price and publicly available information, risk attitude and initial wealth of individuals, topology of the social network, and the strength of social influences in the network.
- Trading volume increases as market participants disagree with each other more.

References

[1] R. Albert and A. Barabási. Topology of evolving networks: Local events and universality. *Physical Review Letters*, 85(24):5234–5237, 2000.

[2] A. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[3] J. E. Berg, R. Forsythe, F. D. Nelson, and T. A. Rietz. Results from a dozen years of election futures markets research. In C. A. Plott and V. Smith, editors, *Handbook of Experimental Economic Results (forthcoming)*. 2001.

[4] J. E. Berg and T. A. Rietz. Prediction markets as decision support systems. *Information Systems Frontier*, 5:79–93, 2003.

[5] D. S. Callaway, J. E. Hopcroft, J. M. Kleinberg, M. E. J. Newman, and S. H. Strogatz. Are randomly grown graphs really random? *Physical Review E*, 64(041902), 2001.

[6] K. Chen, L. Fine, and B. Huberman. Predicting the future. *Information System Frontier*, 5(1):47–61, 2003.

[7] K. Y. Chen and C. R. Plott. Information aggregation mechanisms: Concept, design and implementation for a sales forecasting problem. Working paper No. 1131, California Institute of Technology, Division of the Humanities and Social Sciences, 2002.

[8] S. Debnath, D. M. Pennock, C. L. Giles, and S. Lawrence. Information incorporation in online in-game sports betting markets. In *Proceedings of the Fourth Annual ACM Conference on Electronic Commerce (EC'03)*, San Diego, CA, June 2003.

[9] E. F. Fama. Efficient capital market: A review of theory and empirical work. *Journal of Finance*, 25:383–417, 1970.

[10] R. Forsythe and F. Lundholm. Information aggregation in an experimental market. *Econometrica*, 58:309–47, 1990.

[11] R. Forsythe, F. Nelson, G. R. Neumann, and J. Wright. Anatomy of an experimental political stock market. *American Economic Review*, 82(5):1142–1161, 1992.

[12] R. Forsythe, T. A. Rietz, and T. W. Ross. Wishes, expectations, and actions: A survey on price formation in election stock markets. *Journal of Economic Behavior and Organization*, 39:83–110, 1999.

[13] J. M. Gandar, W. H. Dare, C. R. Brown, and R. A. Zuber. Informed traders and price variations in the betting market for professional basketball games. *Journal of Finance*, LIII(1):385–401, 1999.

[14] S. J. Grossman. On the efficiency of competitive stock markets where traders have diverse information. *The Journal of Finance*, 31(2):573–585, 1976.

[15] R. D. Hanson and R. Oprea. Manipulators increase information market accuracy. Technical report, George Mason University, 2004.

[16] B. A. Huberman and L. A. Adamic. Growth dynamics of the world-wide web. *Nature*, 401(6749):131, 1999.

[17] J. C. Jackwerth and M. Rubinstein. Recovering probability distributions from options prices. *Journal of Finance*, 51(5):1611–1631, 1996.

[18] A. S. Kyle. Continuous auctions and insider trading. *Econometrica*, 53(6):1315–1336, 1985.

[19] R. D. McKelvey and T. Page. Common knowledge, consensus, and aggregate information. *Econometrica*, 54(1):109–127, 1986.

[20] S. Milgram. The small world problem. *Psychology Today*, (22):61–67, 1967.

[21] P. Milgrom and N. Stokey. Information, trade and common knowledge. *Journal of Economic Theory*, 26(1):17–27, 1982.

[22] L. T. Nielsen, A. Brandenburger, J. Geanakoplos, R. McKelvey, and T. Page. Common knowledge of an aggregate of expectations. *Econometrica*, 58(5):1235–1238, 1990.

[23] K. Oliven and T. Rietz. Suckers are born, but markets are made: Individual rationality, arbitrage and market efficiency on an electronic futures market. *Management Science*, 50(3):336–351, 2004.

[24] S.-T. Park, D. M. Pennock, and C. L. Giles. Comparing static and dynamic measurements and models of the internet's AS topology. In *Proceedings of 23rd Conference of the IEEE Communications Society (INFOCOM 2004)*, pages 174–183, Hong Kong, China, 2004.

[25] D. M. Pennock, G. W. Flake, S. Lawrence, E. J. Glover, and C. L. Giles. Winners don't take all: Characterizing the competition for links on the web. In *Proceedings of the National Academy of Sciences (PNAS)*, pages 5207–5211, April 2002.

[26] D. M. Pennock, S. Lawrence, C. L. Giles, and F. A. Nielsen. The real power of artificial markets. *Science*, 291:987–988, February 2002.

[27] D. M. Pennock and M. P. Wellman. A market framework for pooling opinions. Technical Report 2001-081, NEC Research Institute, 2001.

[28] C. Plott and S. Sunder. Efficiency of experimental security markets with insider information: An application of rational expectations models. *Journal of Political Economy*, 90:663–98, 1982.

[29] C. Plott and S. Sunder. Rational expectations and the aggregation of diverse information in laboratory security markets. *Econometrica*, 56:1085–118, 1988.

[30] C. R. Plott, J. Wit, and W. C. Yang. Parimutuel betting markets as information aggregation devices: Experimental results. Technical Report Social Science Working Paper 986, California Institute of Technology, Apr. 1997.

[31] R. Roll. Orange juice and weather. *The American Economic Review*, 74(5):861–880, 1984.

[32] C. Schmidt and A. Werwatz. How accurate do markets predict the outcome of an event? The Euro 2000 soccer championships experiment. Technical Report 09-2002, Max Planck Institute for Research into Economic Systems, 2002.

[33] E. Servan-Schreiber, J. Wolfers, D. M. Pennock, and B. Galebach. Prediction markets: Does money matter? *Electronic Markets*, 14(3):243–251, 2004.

[34] R. J. Shiller. Conversation, information, and herd behavior. *The American Economic Review*, 85:181–185, 1995.

[35] R. H. Thaler and W. T. Ziemba. Anomalies: Parimutuel betting markets: Racetracks and lotteries. *Journal of Economic Perspectives*, 2(2):161–174, 1988.

[36] D. Watts and S. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

[37] J. Wolfers and E. Zitzewitz. Interpreting prediction market prices as probabilities. Technical report, NBER Working Paper No. 10359, 2005.