The Pennsylvania State University
The Graduate School
College of Information Sciences and Technology

PREDICTION MARKET AND ITS COMBINATION WITH SOCIAL NETWORK

A Thesis in
Information Science and Technology

by

Wenying Xiong

© 2011 Wenying Xiong

Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Science

August 2011
The thesis of Wenying Xiong was reviewed and approved* by the following:

Tracy Mullen
Assistant Professor of Information Science and Technology
Thesis Advisor

John Yen
Professor of Information Science and Technology

Xiaolong (Luke) Zhang
Associate Professor of Information Science and Technology

Mary Beth Rosson
Chair of College of Information Sciences and Technology

*Signatures are on file in the Graduate School.
ABSTRACT

A prediction market is a speculative market designed to predict event outcomes. It is an information aggregation method that aggregates participants’ information into a single prediction. Prediction markets have a wide range of real-world applications including predicting political events, sports competitions, current events, entertainment box office results, business sales figures, dates for scientific and technology breakthroughs, and more.

The content of this thesis can be divided into two parts. First, I describe the motivation for and history of prediction markets, along with prediction market applications, and mechanisms. Second, I review the limited research that combines prediction markets with social networks and describe my initial research in this area. The intersection of prediction markets and social networks is a promising area, since in most cases participants in prediction markets are actually interacting in some kind of a social network context, but such a context is often neglected in the modeling of prediction markets. In my research, I build on previous research by Chen et al [1], introducing uncertainty to participant’s beliefs, and analyze the effect on the prediction market, analyzed in particular its influence on market prediction accuracy, trading volume, and social network clustering. Two social network theories that motivate and help explain my experimental results are Festinger’s theory[2] and Burt’s theory[3]. Festinger’s theory states that people with similar attitudes and perceptions tend to be attracted to one another, which changes the social network structure. Our experimental design follows Festinger’s theory that people with similar Party registration tend to cluster together. Burt’s theory asserts that overall network structure as
well as individual network characteristics influence information flows. My results support this theory, showing that when participants are more clustered, the market has better overall prediction.

My preliminary study finds that adding uncertainty causes several effects: First, more uncertainty tends to reduce clustering. Second, and not surprisingly, more uncertainty tends to result in lower accuracy of information aggregation. Third, more uncertainty tends to cause lower trading volumes.

**Keywords:** prediction markets, social network, information aggregation, uncertainty, trading volume, clustering
TABLE OF CONTENTS

List of Figures ................................................................................................................................ vii
List of Tables ................................................................................................................................... viii
Acknowledgement ........................................................................................................................... ix

1 Introduction ................................................................................................................................ 1
  1.1 My main work ............................................................................................................................... 2
  1.2 Thesis framework .......................................................................................................................... 3

2 History and applications of prediction market ............................................................................. 5
  2.1 Market ........................................................................................................................................... 5
  2.2 Prediction Market and its history .................................................................................................. 6
  2.3 Real world applications of prediction market .............................................................................. 12
  2.4 The advantages and challenges of prediction market ................................................................. 18
  2.5 Summary ..................................................................................................................................... 20

3 The principle of prediction market ................................................................................................. 21
  3.1 Types of prediction market .......................................................................................................... 21
  3.2 How prediction markets work ...................................................................................................... 25
    3.2.1 Participant Selection ............................................................................................................... 26
    3.2.2 Preference ............................................................................................................................... 28
    3.2.3 Bidding methods ................................................................................................................. 30
    3.2.4 The meaning of price .......................................................................................................... 33

  3.3 Comparison with other information aggregating methods .......................................................... 34
  3.4 Summary ..................................................................................................................................... 39

4 Intersection of prediction market and social network ................................................................... 42
  4.1 An introduction of social network ............................................................................................... 42
  4.2 Previous work ............................................................................................................................... 46
    4.2.1 Background ............................................................................................................................. 46
    4.2.2 Social network theory ......................................................................................................... 46
    4.2.3 The Google application ....................................................................................................... 47
    4.2.4 Chen’s model ....................................................................................................................... 49

  4.3 My work: Adding voting certainty level ...................................................................................... 56
List of Figures

Figure 1: An illustration of how prediction market works: Step 1 ................................................................. 8
Figure 2: An illustration of how prediction market works: Step 2 ................................................................. 9
Figure 3: An illustration of how prediction market works: Step 3 .............................................................. 10
Figure 4: An illustration of how prediction market works: Step 4 .............................................................. 11
Figure 5: Iowa Electronic Market 2008 Democratic Convention Market ................................................... 14
Figure 6: The Predictalot bulletin, which keeps tracking the most recent user activities ............................ 16
Figure 7: The user interface of Predictalot to make a prediction .................................................................. 17
Figure 8: An example of user’s selections in Predictalot ............................................................................ 17
Figure 9: An illustration about current implementations of ........................................................................ 24
Figure 10: A classification of participant selection methods ...................................................................... 27
Figure 11: Different kinds of bidding methods and the relationships between them ................................. 32
Figure 12: A comparison of prediction market, scoring rules and market scoring rules ............................ 37
Figure 13: An illustration of the relationships between different methods of aggregating opinions .......... 38
Figure 14: An example of social network diagram ...................................................................................... 43
Figure 15: An illustration of the twitter user interface .................................................................................. 45
Figure 16: An example of how agents exchange their positions in a social network according to their beliefs: Step 1 .............................................................................................................................................. 50
Figure 17: An example of how agents exchange their positions in a social network according to their beliefs: Step 2 .............................................................................................................................................. 51
Figure 18: An example of how agents exchange their positions in a social network according to their beliefs: Step 3 .............................................................................................................................................. 52
Figure 19: An example of how agents exchange their positions in a social network according to their beliefs: Step 4 .............................................................................................................................................. 52
Figure 20: An example about participant selection using preferential attachment .................................. 55
Figure 21: Different voting uncertainty causes different level of clustering ............................................... 58
Figure 22: Agents vote at the very beginning ............................................................................................... 63
Figure 23: Decide-first method, CARA utility function. The agents vote before they trade. ..................... 65
Figure 24: Decide-first method, GLU utility function. The agents vote before they trade. ......................... 66
Figure 25: Decide-last method, CARA utility function. The agents vote after they trade. ......................... 67
Figure 26: Decide-last method, GLU utility function. The agents vote after they trade. ......................... 67
Figure 27: Decide-first method, CARA utility function. The agents vote before they trade. ..................... 69
Figure 28: Decide-first method, GLU utility function. The agents vote before they trade. ......................... 70
Figure 29: Decide-last method, CARA utility function. The agents vote after they trade. ......................... 70
Figure 30: Decide-last method, GLU utility function. The agents vote after they trade. ......................... 71
Figure 31: An illustration of the conclusions ............................................................................................... 73
List of Tables

Table 1: The comparison of several mechanisms which aggregate opinions........................................41
Table 2: A comparison of the three methods in decision making ............................................................63
Acknowledgement

This dissertation would not have been possible without the guidance and the help of several individuals who in one way or another contributed and extended their valuable assistance in the preparation and completion of this study.

First and foremost, my utmost gratitude to Dr. Tracy Mullen, my academic and thesis adviser whose sincerity and encouragement I will never forget. Dr. Mullen has been my inspiration as I hurdle all the obstacles in the completion this research work.

Also I need to thank Dr. John Yen and Dr. Luke Zhang, for their unselfish and unfailing support as my thesis committees. They have brought about lots of good suggestions for my research and thesis.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

Wenying Xiong
**Introduction**

Prediction markets originated from betting markets in 1980s and their numbers gradually increased in the 2000s. In the history of prediction markets, there are many different types of prediction markets and they can be divided into categories according to several different methods of classification. According to the complexity of the state space, it can be divided into simple prediction markets and combinatorial prediction markets. According to the application, it can be divided into election prediction markets and sports prediction markets. While election prediction markets and sports prediction markets are not the only applications of prediction markets, they are two main types of application. According to the source of capital, it can be divided into play-money and real-money markets. In essence, a prediction market aims to predict the likelihood of future events based on the market price, where a higher price generally means the event is more likely to occur. The final prediction is an aggregated opinion since all participants trading in the market will have influence on the price.

Freeman et al.’s [4] article provides a detailed description of the development of social network analysis. Social network analysis is a structured approach based on the study of interactions between social actors. It focuses on the relationships that link individual human beings, which can be friendship, kinship, common interest, financial exchange, dislike, or relationships of beliefs or knowledge. Interestingly, important social relationships may link individuals that are not human, such as bees, ants, and deer.

Although this thesis mainly focuses on the prediction market part, my research considered that the participants in a prediction market are actually in a social network context. As a result, it...
might be interesting to see how the interactions between individuals will influence on the market behavior.

1.1 My main work

Most previous prediction market research has not considered the impact of traders’ social networks upon individual trader’s trading performance and upon the overall prediction market forecasting accuracy. Chen et al [1] built a model and performed multi-agent simulations to investigate the impact of social network influences on traders’ behavior and the resultant prediction market accuracy and trading volume. In their model, traders’ beliefs about who will win an election are biased according to the voting activities of their neighbors in their social network. Traders know who both they and their immediate neighbors are voting for (i.e., Democratic or Republican) and use this information to create their election outcome prediction and related prediction market bid. Our work extended the Chen model by introducing the factor of uncertainty. First, in reality, a person might change his/her mind. So he/she has uncertainty in his/her own opinions. Second, a person usually does not have a perfect knowledge about his/her neighborhood, so there is uncertainty in his/her neighbors’ opinions.

Our preliminary study finds that adding uncertainty causes several effects: First, more voting certainty means that it is easier for people with similar attitudes and beliefs to reside, or cluster, near each other in the social network. Our simulation shows that when the agents are totally certain about their beliefs (100% voting certainty level), the clustering reaches the highest level, which means clusters of agents tend to vote in a similar fashion. On the other hand, when the
agents are totally not certain about their beliefs (50% voting certainty level), no clustering occurs. Second, and not surprisingly, more uncertainty tends to result in lower accuracy of information aggregation. We found that when the voting certainty level is low, the aggregated market prediction is less accurate than when the voting certainty is high. Third, more uncertainty tends to cause lower trading volumes. We observed that the trading volume goes down with the decrease of voting certainty level.

1.2 Thesis framework

The objective of this thesis is to review prediction market literature, find related theories, present my research on the combination between prediction market and social network, and link the experimental results to the theories of social network. Section 2 gives an introduction of the history, the motivation, and the challenges of prediction market including several examples about the real-world applications of prediction market. Also section 2 provides a detailed example to explain how prediction market works, and the working principles are expanded through section 3. Section 3 describes prediction market theoretical underpinnings, including how prediction markets are classified and how prediction markets work. It extends the example in section 2 about how prediction market works, such as how to select the participants, how to choose a utility function, and how the market price is formed. In addition, this section includes a brief comparison between prediction market and other models which aggregate opinions. In section 4, I present the limited research available on the intersection between prediction markets and social network, describe my research on this intersection, and in section 5 I make conclusions and point out potential future research.
2 History and applications of prediction market

2.1 Market

A market is an institution in which buyers and sellers are allowed to trade and exchange goods and services. According to [5], the concept of a market is any structure that allows people to buy and sell. And the content to exchange can be any type of goods, services and information.

Markets originated from physical marketplaces, such as farmer’s markets, clothing markets, etc. In those places people can buy and sell goods to meet their needs. The nature of business transactions could define markets. As technology developed, the meaning of market gradually extended to other structures beyond physical marketplaces. An important term called contract appears. A contract in this context is really a security, or a fungible, negotiable financial instrument representing financial value. Instead of trading real goods, people trade contracts in online marketplaces. We can consider the contract as a certain kind of stock. If one thinks that its potential price should be higher than the market price, then a trader will tend to buy it. Otherwise the person is likely to sell it. For example, suppose the market price for the contract “Obama will win the 2012 election” is $0.78. Those who believe the price will rise beyond $0.78 are potential buyers, while those who think the price will fall below $0.78 are potential sellers. Two main types of markets include financial markets and prediction markets. In financial markets people can efficiently exchange liquid assets. Financial markets can be stock markets, bond markets, futures markets, currency markets, etc. Generally speaking, financial markets are real money markets, so people focus on buying and selling contracts and making profits. A prediction market is just a financial market whose aim is to provide predictions rather than trading gains.
Unlike most financial markets though, most prediction markets are play money markets, and only a small number of prediction markets are real money markets. In prediction markets each person is like an expert who can give a prediction for a certain event, such as whether the Democratic Party will win the election. A prediction market can aggregate individual predictions into a single prediction, say, the probability that the Democratic Party will win. Actually, to some extent, a financial market is a kind of prediction market, for people make predictions about whether the price of contracts will rise or fall in a financial market. But in prediction markets the scope of prediction is about general events. Participants predict whether an event will happen or not, such as “President Clinton will win the 2008 election”. So in most of cases, financial markets and prediction markets have different usages.

2.2 Prediction Market and its history

In this section we will talk about what is prediction market, the history of prediction market and some related examples. Berg et al.’s [6] [7] work provides a comprehensive introduction to prediction markets. According to [6], prediction markets are those run for the primary purpose of aggregating the information content in individuals to make predictions about specific future events, and prediction markets are speculative markets.

Consider the following example that will be discussed throughout my thesis. Each individual is registered either for the Democratic Party or for the Republic Party, and each of them is predicting which party will win the presidential election. In designing a contract around this event for a prediction market, the contract content will be tied to the outcome of the event. For example, a contract could be in the form of “The Democratic Party will win the election”,
“Clinton will win the 2008 election”, or, in a complex case, “The percentage of votes received by the Republican nominee conditional on Robert Dole being the Republican nominee”. Let’s consider the case where the contract is “The Democratic Party will win the election”. In this case, a contract that pays $1 if and only if the Democratic Party wins the election and it pays $0 if and only if the Republican Party wins the election. It is a 0-1 value in the example, and this market is actually a winner-takes-all market, which will be explained in section 3.2.3. So we can see that the market price is a value which is always between 0 and 1.

Figure 1 illustrates how a prediction market works in this example. In step 1, as shown in Figure 1, each participant has an expectation about the prediction result according to the information he/she has got. The information can include many aspects, such as news released from internet, news from friends, the current market price of the contract, and more. As a result, the participants have an expectation about the market price, which directly links to the probability that the Democratic Party will win. And those two are numerically equal. For example, if an individual believes that the market price will be $0.8, then his/her prediction of the probability that the Democratic Party will win is 80%. We assume that participants 1-6 expect the probability that the Democratic Party will win the election to be 100%, 25%, 50%, 0%, 65%, 80%, separately, so that predictions of the final market prices for participants from participant 1 to participant 6 are $1, $0.25, $0.5, $0, $0.65, $0.8 separately. Section 3 will give some theoretical explanations about this relationship.
Figure 1: An illustration of how prediction market works: Step 1

In step 2, as shown in Figure 2, we assume that the current price is $0.60. So the three participants whose expectations of the price are $1, $0.65, and $0.8 (meaning their predictions of the probability that the Democratic Party will win are 100%, 65%, 80% correspondingly) become potential buyers, because the current market price is lower than they expected. And the other three participants whose expectations of the price are $0.25, $0.5, and $0 (meaning their predictions of the probability that the Democratic Party will win are 25%, 50%, 0% correspondingly) become potential sellers, since the current market price is lower than they expected.
In step 3, as indicated in Figure 3, buyers and sellers begin to trade contracts. Intuitively, if a person’s expectation of the market price is much higher than the current price, then he/she is likely to trade contracts more than when his/her expectation is only a little higher than the current price. However, we cannot neglect that some people are more aggressive in trading, while some are more conservative. Those who are more aggressive are likely to trade higher volumes than those who are more conservative. One of the measurements about how aggressive they are is the utility function, as a measure of relative satisfaction, which will be introduced in detail in section 3. We can see that in this example, person 1 buys 300 contracts since he is very confident that the Democratic Party will win (The probability is 100%). Person 5 buys less (50 contracts) since she is less confident that the Democratic Party will win (The probability is 65%). Similarly, for sellers we can see that person 4 sells more than person 2, since person 4 is very confident that
there is low probability for the Democratic Party to win (0%), while person 2 is less confident that the Democratic Party will lose.

In step 4, as indicated in Figure 4, buyers and sellers went through steps 1-3 for multiple times until the market reaches an equilibrium price. After the buyers and sellers have done step 3, the market price will change and both buyers and sellers will have a new expectation of the market price, which returns to step 1. The equilibrium price ($0.58) here is also the market prediction, which means the whole market expects that the probability of Democratic will win is 58%. In this case, the prediction market asynchronously aggregates the opinions of the six participants, or “experts”. Note that finally each contract will pay $1 if the Democratic Party really wins the election, and it will pay $0 if the Democratic Party really loses the election.
The example of prediction market above is also an election prediction market which focuses on prediction political event. It is a market in which the values of the contracts being traded are based on the election outcome. The most common applications cited in the prediction market literature tends to be election prediction markets [6-9] and sports prediction markets, which focus on predicting sports competitions [10-13]. Prediction markets are also referred to as information markets, predictive markets, decision markets, or virtual markets. But in this thesis they are referred as prediction markets only.

The history of prediction markets is long and colorful. The origination of prediction market is the betting on elections. After 1940s formal markets on Wall Street gradually took the place of betting markets. Prior to widespread polling, newspapers begin to report market conditions to let readers know some information about elections. According to Rhode et al.’s [14] paper, the markets involved thousands of participants, had millions of dollars in volume in current terms, and had remarkable predictive accuracy. John Brunner's 1975 science fiction story “The
Shockwave Rider”[15] described a prediction market which is called the Delphi Pool. Around 1988, Iowa Electronic Market, which is one of the most famous prediction markets, started to run. In July 2003, the U.S. Department of Defense publicized a Policy Analysis Market. It was criticized very quickly as a “terrorism futures market” and the program was hastily canceled. Several books gave a very good evaluation for prediction markets. They are James Surowiecki's 2004 book “The Wisdom of Crowds”[16], Cass Sunstein's 2006 book “Infotopia”[17], and Douglas Hubbard’s 2007 book “How to Measure Anything: Finding the Value of Intangibles in Business” [18]. In October 2007 the Prediction Market Industry Association was formed, countries included the United States, Ireland, Austria, Germany, and Denmark. This marked that prediction market has become international.

2.3 Real world applications of prediction market

There are many examples of prediction markets open to the public. A lot of them are real-world applications and are used by people all over the world. Generally speaking, the predictions include political events (in the United States, Germany, India, etc), sports competitions (like football games, horse races), entertainment (like Oscar awards, Academic awards), science (will the Higgs boson be observed?), technology (the unit sales of Apple ipad), finance (NASDAQ-100, gold price), climate change, current event (like bird flu), and more. Iowa Electronic Market[19] is an on-line futures market. The contracts are the real world events, such as political elections, stock prices, etc. It has a long history and is one of the most famous prediction markets. It is also an academic market used for research. One famous application of Iowa Electronic Market is the U.S. presidential election, and there are many different kinds of predictions. In a
winner-takes-all market, people can bid a certain Party to win the election, such as Democratic Party, or a certain person, such as Bill Clinton. In a vote-share market, they can bid the percentage of votes got from a certain Party, or a certain person. Berg et al. [6] did some research about U.S. presidential election in Iowa Electronic Market and showed how both prediction and conditional prediction markets can be designed and used for decision support. Compared to ordinary prediction markets, conditional prediction markets could give more information, which can make the predictions more accurate.

Figure 5 is an illustration of how Iowa Electronic Market works in the prediction of the U.S. presidential election in 2008. The data goes from March 3rd, 2007, to August 26th, 2008. The four curves are about the closing price of the contract of “Clinton wins”, “Obama wins”, “Edward wins”, and “Others win”, separately. We can see that the prices of those four contracts vary with time. The price of the Obama contract increases dramatically around January 2008, while the prices of the Clinton contract decreases around January and February 2008. And we can expect that certain events happened at the beginning of 2008, which made people start to buy lots of Obama contract instead of Clinton contract. This shows that the price of prediction market can be linked to real world events.
The Hollywood Stock Exchange (HSX) performs prediction about upcoming films, actors, and film-related options. The creators, Max Keiser and Michael Burns, received U.S. patent in 1999 for the invention. HSX is a web-based multiplayer game. In the game participants use play money to buy and sell the contracts of actors, directors, and more. For example, if a particular movie stock trades at “H$40.00”, the market is predicting that the movie will gross US$40 million in the first four weekends of wide release. Another example is that people can purchase the contracts of musical artists. During the dot-com boom, HSX attracted some private investment and ran TV advertisements to attract players.

Sports prediction market is another important application, which might include football, horse race, etc. Betfair is an online sports market, and it is the world's biggest prediction exchange. In 2007, $28 billion are traded in Betfair. Yoopick[12] is another sports prediction market where people can bet on the results of football games. Similar to Election Prediction markets, the
betting language in the sports prediction markets is also flexible. In the sports prediction markets people might bet on who will win, or how many points a given team will win. In addition, in the sport games, people might be interested not only in who gets the first, but also in who gets the second and the third. Goel et al.[12] provides an implementation in which agents select a point interval, and bet on the final point difference landing in that interval.

Some people might wonder, since in the real money markets people can get money, what is the incentive for them to participate in a play money market? One reasonable explanation is that in a play money market, people treat trading as a game. According to the definition, a game is a structured playing and it is usually taken for enjoyment. It has key components including goals, rules, interactions and challenges. We can see that prediction market matches the definition of game very well. In a prediction market, the participants have to follow the trading rules, such as the maximum or minimum amount to trade. The goal for each participant is to make money in order to win the game. The buying and selling behaviors are interactions between participants. In addition, it is also possible for the participants to communicate with each other. A good example is when we put prediction market into a social network context, as will be discussed in detail later. The participants will face challenges in the game, because each of them has chances to lose money. Based on those key components of game, it is reasonable to describe the trading in a prediction market as a game. Some online applications, such as Predicalot[20], simulate real money markets. Participants initially get $1000 play money, and they can make predictions according to the events happened. Predicalot is a combinatorial market, which means the events people are predicting can be combinations of more than one event. The predictions are like
“Florida State will advance further than Notre Dame”, “Big 12 will win more than Atlantic 10”, “Florida State will get the first place and Duke will get the third place”, etc. Figure 6 shows the Predictalot bulletin. It keeps track of the most recent activities of all the users. Figure 7 shows the user interface of Predictalot to make a prediction. One can predict the champion, such as which team will win the tournament, or predict the team progress, such as what team will advance further than what other team, or other interesting predictions. In Figure 8, we can see that the participant can select teams, and whether they win or lose in the dialog box. And then move the slider to choose how much to invest. Servan et. al’s research [21] showed that in the play money market, the participants’ performance is as good as in the real money market.
Recently, prediction markets have been linked with social networks. The assumption is that people in a prediction market are actually in a social network context. Cowgill et al. [22] ran prediction markets at Google with employees as traders. They collected the trading data and also
some real-world social-network information about the traders including who they sat next to, what group they worked with, and so on. Cowgill noticed that geographical proximity between traders can cause correlated trading, as measured by the number of net shares purchased. In addition, organizational proximity between traders also led to correlated trading. We will extend the Google example in section 4.2.3.

Besides the examples mentioned above, there are many other applications. Intrade is a for-profit company with various kinds of contracts which do not include sports. Actually, most prediction markets are play money markets, such as the Industry Standard's technology industry prediction market, the simExchange, the Popular Science Predictions Exchange, Lumenogic, and the Foresight Exchange Prediction Market. In those markets purchases are made with play money.

2.4 The advantages and challenges of prediction market

A prediction market is an efficient way to predict future events. In other words, it can be used to support people’s decisions. Lots of literature has discussed the advantages of prediction markets. Berg et al.’s [6-7, 23] and Forsythe et al.’s [24] papers have given a good summary of prediction markets’ advantages. First, since the participants can trade continuously, the market can give updated predictions in a continuous way as time goes on. Second, through the price formation process, the markets aggregate information across traders, solving what would otherwise be complex aggregation problems. In other words, we can treat each participant as an expert and let the prediction market efficiently aggregate opinions from multiple experts. Third, evidence suggests that prediction markets have ability to give unbiased and relatively accurate
forecasts[23]. Fourth, these forecasts made by prediction markets can outperform existing alternatives [7, 23-25]. Fifth, evidence suggests that market dynamics can provide a non-biased estimation of future events[24, 26]. Finally, the markets can be designed to forecast a variety of issues and provide a variety of types of information [23, 25, 27-30]. As mentioned, it can almost predict all kinds of future events. According to what was mentioned above, we can see that prediction markets can asynchronously aggregate the opinions of an unlimited number of participants in a very easy way. Because of those reasons, prediction market research is very promising.

There are some challenges for prediction market. First, it is important to know many parameters in prediction market. Especially, when we are doing simulations, we need to consider many factors to make it perform like a real prediction market. Factors such as how to select participants, how participants select their favorite contracts and decide how many contracts to buy need to be considered. Second, since most prediction markets are play money market rather than real money market, we may wonder whether participants have enough incentives to participate in. Third, sometimes the contract in the market is very complex to define, and there are few participants buying or selling such a contract. As a result, the prediction market will suffer from thin market problem, which means the prediction may not be accurate. Fourth, in the real prediction market there are actually many interactions between participants, but most literature fails to consider those interactions in the prediction market model. This thesis will mainly focus on researching the fourth challenge and will discuss the application of prediction market in the social network context in section 4.
2.5 Summary

In chapter 2 we talked about the history, application, motivation, and challenges of prediction market. Prediction markets are developed from betting on elections. The first known corporate prediction market was established in 1990s and many other prediction markets were publicized in the 2000s. Also, lots of research has been done to prediction markets during the 2000s. The history of prediction markets has shown how prediction markets are developed. What is more important, this section described multiple applications of prediction market, such as election prediction market like Iowa Electronic Market, sports prediction market like YooPick, and social network context such as the Google company. Section 3 will talk about how prediction market works in a theoretical way.
3 The principle of prediction market

This section will theoretically introduce the working principles of prediction markets, which extend the example in section 2 and provide a theoretical basis for section 4. Section 2 mentioned a 4-step example of how prediction market works. However, many factors need to be considered when implementing a prediction market, such as how to effectively select participants, how many contracts should a participant buy or sell, how the market forms an equilibrium price. Section 3 gives a detailed explanation of all those issues.

3.1 Types of prediction market

Section 2.2 and 2.3 gave some examples about election prediction market and sports prediction market. Actually they are two main types of prediction market in current literature. In election prediction markets, values of contract traded are based on the outcome of elections. In sports prediction markets, values of contract traded are based on the outcome of sport games, which can be football games, basketball games, horse races, etc. In any type of prediction market, participants invest their own funds to trade contracts. There are some other kinds of prediction markets, such as corporate prediction market, which tries to bring technology into business. One example is the Hewlett-Packard (HP), which performed sales forecasts several years ago. Other examples include Lumenogic, which was born out of the combination of prediction markets and collective intelligence technology pioneer NewsFutures and two managing partners of The New England Consulting Group (NECG), a leading growth strategy consulting firm. A common thing about different kinds of prediction markets is they have opportunities to earn profits, and at the same time they bear the risk of losing money. According to current literatures, the main
difference between election prediction market and sports prediction market is about the complexity of the outcome space. In the election, usually people are only concerned about who wins, and there is likely to be only one winner. In the sport games such as horse race, however, people might be interested not only in who gets the first, but also in who gets the second and the third. As a result, the research on sports prediction markets [10-13] is often a little more complex than the research on election prediction markets. Usually, the betting on sports prediction market is a combinatorial betting, which means the events people are predicting for are combinations for two or more events. For example, “horse A will be in the first place and horse B will be in third place”. The outcome space can be high-dimensional. For combinatorial prediction markets, there are lots of variables that have a joint probability distribution. Combinatorial markets have to aggregate information on the entire joint probability distribution, by allowing bets on all combinations for the variable values. Actually, this literature review has further explanations in section 3.3 about combinatorial prediction markets. One example is the outcome space in a horse race, which might be all n! possible permutations of horses, and the allowable bets are properties of permutations. Since the traders may like to bet on arbitrary properties of the final ordering, the goal of the exchange is to search the entire spaces of possible orderings to find at least two that form an agreeable match. In this case, computational complexity becomes a problem. For a prediction market to work well, contracts must be clear, easily understood and easily adjudicated[31]. This requirement for clarity can also turn out to be complex. Since this thesis focuses more on prediction market itself rather than computational complexity, we will not discuss complexity problem in further details. Usually the simple version of prediction market can be representative because it has most of the characters of
prediction market. For election prediction market, combinatorial betting might also exist since traders can bet on arbitrary properties[10]. For example, they may bet on “candidate F will result in either second place or third place”, “candidates D and R will both defeat candidate L”, etc. But usually people might focus on single outcomes in election prediction market, such as whether Democratic party of Republic party will win the election[1]. In election prediction market, sometimes people may focus on conditional probabilities such as “the percentage of votes received by the Democratic nominee conditional on Robert Dole being the Republican nominee”, and such a prediction market is called conditional prediction market[6, 29]. Conditional estimates are highly linked to our daily life, for we may want the conditional estimates such as the chance of rain given the geographic information, or the chance of war given a particular person selected as president. From these examples we can see that conditional prediction market is a promising area which has potential to be applied to many places.

In election prediction markets, less emphasis is put on combinatorial betting than on single betting. However, there is a trend that more research about combinatorial betting will be conducted in election prediction market. According to their complexity, prediction markets can be divided into combinatorial prediction markets and simple prediction markets. According to the application, election prediction markets and sports prediction markets are the two main types of prediction markets, but there are still some other types of prediction markets, which are not as frequently implemented as those two. Figure 9 has given an illustration of the two perspectives in which prediction markets can be classified, and the relationships between those two perspectives.
The third method to classify prediction markets is that they can be divided into play-money markets and real-money markets. Real-money market means that people actually put their money into the market, so it is like a real betting. But in a play-money market, people do not put their money into the market, and the market is operated by using virtual money. It is widely believed among economists that prediction markets based on play money are not likely to generate credible predictions. Economists think that real money markets can have a better performance of prediction than play money markets, since in real money markets participants risk their money. But in play money markets participants run no financial risk, so they will not have enough incentives to participate in. This belief insists that monetary risk is required in order to obtain valid conclusions about economic behavior[21]. However, the data collected so far from Berg’s [7], Forsythe’s [26] and Pennock’s [32] work has shown that the performances of prediction
markets based on both play money and real money have remarkable accuracy[7, 26, 32]. In addition, sometimes prediction markets with real money have to be restricted. Firstly, since online betting is forbidden by laws in United States, most of prediction markets are operated by play money rather than real money. Play money markets are free to play without purchase, so participants take no risk. The best traders are offered prizes, so they have enough incentives to participate. Intrade is a play money markets and it is an exception, for it is operated from Dublin, Ireland, where betting is legal. Iowa Electronic Markets is also an exception, which operates from the University of Iowa and allows bets up to $500. Another special example is Bet2Give. It is a charity prediction market in which real money is traded, but ultimately all winnings are given away to non-profit organizations of the winner’s choice. Generally speaking, regulators only allow exceptions for certain purposes, and regulatory bodies are devoted to approving some markets for the specific purpose. So we can see that prediction markets are only allowed accidentally, for there are no regulatory bodies devoted to approving markets which serve the purpose of aggregating information[33]. Secondly, even in those countries where betting is legal, setting up a market based on real money incurs huge technical and regulatory costs. Finally, it is difficult for a corporation to require its employees to risk some of their own money on producing better forecasts for the company.

3.2 How prediction markets work

In this section we will talk about how prediction markets work. Section 2.2 has given an example about the working principle of prediction market, and this section will extend it in more detail. Participation is the most important issue, because the participants are usually human beings, and
A prediction market needs humans to participate to operate. The participants’ preferences are the second issue, since participants need some criterion to decide how many contracts to buy or sell. The third issue is bidding methods. There is more than one bidding method for each kind of prediction market. The last, but not the least issue, is the meaning of price. If the market and contracts work well, then the equilibrium price will reflect the consensus probability that an event will happen.

### 3.2.1 Participant Selection

A prediction market is a mechanism which needs participants. In a prediction market, participation is the most important part. So participant selection is the first important issue in prediction market. In most of the cases in an academic environment participants are selected to participate in the market with equal chance, the premise is that there is no strong preference for any participant and each participant is equal in the selection. But when some of the participants are more knowledgeable than others, those participants can be given a higher probability to be selected. In Chen et al.’s [1] paper, the assumption is that people who are well connected can get access to more information, so a preferential selection is adopted, and we will further discuss it in section 4.2.4. Here Figure 10 is used to illustrate different participant selection methods.
Chen et al.’s [34] work compared the differences of participant selection in an academic environment and in a business environment. Usually, in an academic experiment, the amount of information each participant can access is decided and controlled by experiment designers. In addition, in an academic environment it is often hard to assume which participants are experts and which are non-experts. Therefore, there is usually no issue of selecting participants for the information they acquire since all of them are designed equally. The case in Chen et al.’s [1] paper is somewhat like an exception, because all the market participants are in a social network context and preferential attachment is beneficial in this certain context. In a business environment, participants need to be carefully selected. Firstly, a person with much information is very valuable and it is not desirable to miss such a person. Secondly, it is not efficient to include too many participants who have no relevant information. Indeed, it is often difficult to decide a specific value of the information size for effective information aggregation to take place, for such a value is usually got from experiments or previous experience rather than theories.
3.2.2 Preference

In section 2.2, we gave an example about how prediction markets work. The agents start to trade in step 3 based on the difference between the market price and their expected price, but the question is how many contracts to trade. This section will discuss this question. In economics, preference refers to the set of assumptions relating to a real or imagined choice between alternatives and the possibility of rank ordering of these alternatives, based on the degree of happiness, satisfaction, gratification, enjoyment, or utility they provide[35]. Usually, economists are not interested in choices or preferences themselves, instead they are interested in the theory of choice such as “why A is chosen but not B”, which might help market designers to design a more efficient market. In a prediction market it is hard to judge the degree of happiness, satisfaction, or gratification. It is often easier and more reasonable to use utility, which is defined as a measure of relative satisfaction, to measure one’s preference. In a prediction market, every market participant wants to maximize his or her expected utility[5]. Chen et al.’s [36] work has given an introduction of utility theory. According to Chen et al.[36], a utility function is a twice-differentiable function of wealth $U(w)$ defined for $w > 0$ which has the properties of non-satiation (the first derivative $U'(w) > 0$) and risk aversion (the second derivative $U''(w) < 0$). It measures an investor’s relative preference for different levels of total wealth. The first property is the non-satiation property, which states that utility increases with wealth. The more wealth one possesses, the happier one will be. The second property is risk aversion property, which states that the marginal utility of wealth decreases as wealth increases and the utility functions are concave. We can use the example of acquiring one additional dollar to explain why utility functions are concave. For a millionaire, obtaining one more dollar is nearly meaningless. But
for someone who has nothing in his pocket, obtaining one dollar to start with is quite important. So we can see that as wealth increases, the increase in utility decreases, which illustrates the risk aversion property.

Now return to the question of how to decide the exact amount of contract to trade for each individual. The answer is that each individual will choose the amount to trade in order to maximize his or her utility. In a prediction market, we often assume that individuals are risk-averse. So when we design prediction market, risk-averse utility functions such as constant absolute risk aversion (CARA) and generalized logarithmic utility (GLU) can be used. The definition of CARA for the individual $i$ is as follows:

$$u_i(y) = -e^{-c_i y} \quad \text{eq. 1}$$

In this case the individual $i$ has negative exponential utility function for money, where $y$ indicates gain or loss of money. If individual $i$ gains, $y$ is positive, and $y$ is negative if individual $i$ loses. $c_i$ is individual $i$’s absolute risk aversion coefficient.

For an individual $i$ with GLU,

$$u_i(y) = \ln(y + b_i) \quad (b_i > 0) \quad \text{eq. 2}$$

it has decreasing absolute risk aversion. $b_i$ is individual $i$ ’s initial wealth.

The CARA function is more risk-averse than GLU since one’s investment is a constant in CARA, while in GLU one’s investment increases with his/her wealth. One example for CARA is that if a person’s investment is $500 in the risky asset, he/she would still invest the same $500 even
he/she has a wealth of $5,000,000 and nothing with $4,999,500. Chen et al.’s [1] paper made a comparison of CARA and GLU when designing prediction market and found that the trading volume (how many contracts an individual buys or sells) is much higher for GLU than for CARA, as will be further discussed in section 4.2.4. Other utility functions include Iso-Elastic utility function, Cobb-Douglas utility function, von Neumann-Morgenstern utility function, and more.

### 3.2.3 Bidding methods

Most prediction markets use continuous double auction to match buyers to sellers. In a continuous double auction, buyers submit bids and sellers submit asking prices. The mechanism executes a trade, which makes the two sides of the market reach a mutually agreeable price. For election stock markets, there are two basic types. The first type is a winner-takes-all market in which only one contract pays a certain amount, typically $1, and all other contracts pay $0. In winner-takes-all market the one who gets majority votes wins. For example, there is an election in which three parties compete, a Red Party, a Blue Party, and a Green Party, and the Blue Party wins the election. In a winner-takes-all market, if you have bought a contract of Blue Party at the price of 42 cents, you earned 58 cents for buying the contract for currently the value for the contract of Blue Party is $1. However, suppose you have bought a contract of Red Party and Green Party at the price of 36 cents and 51 cents separately, then you lost 87 cents in total since the value for the contract of both Red Party and Green Party are $0.

The other common type of election stock market is a proportional share market, or vote share market. In a vote share market, the payout of each contract is equal to the percentage proportions of a specific outcome, and payouts for all contracts sum up to a fixed amount such as $1. Again
suppose there are the same three parties who participate as in the last example: the Red Party, the Blue Party, and the Green Party. Traders are interested in how much of the vote a given party will win, such as whether Blue Party will win more than 42.3% of the popular vote. The common principle across both types of election stock markets is that the payouts for all contracts sum up to a specific amount, generally related to a probability, such as $1.

Contracts are put into circulation as they are bought and sold by market participants. For the election in which the three parties compete, the share of popular votes for each party must sum to 100%. Traders buy and sell contracts, and they make profits by buying low and selling high. For example, if a trader $X$ estimates that the Green Party will win 45.7% of the vote, the trader $X$ will find it profitable to buy the contract of Green Party at the price of less than 45.7 cents. In addition, when another trader $Y$ is willing to buy the contract for more than 45.7 cents, the trader $X$ will find it profitable to sell the same contract. This idea has also been indicated in step 2 in the example in section 2.2. When a trader’s expectation is higher than current price, he or she will buy some contracts; otherwise, he or she will sell some contracts.

The model for sports prediction markets is different from election prediction markets. Two of the most common types of sports bets are money line bets and point spread bets, and Goel et al.’s [12] work has given a good introduction about those two types. According to Goel et al.[12], in money line bets agents bet on which team will win a given game, with payoffs based on the likelihood of that team winning. In point spread bets, gamblers wager on whether the spread, or the final point difference, between the competing teams will be greater than or less than the value specified by a bookmaker. For example, one possible contract is that “team $A$ wins team $B$ by
more than 4.5 points”. As a result, if team A beats team B by 5 points, the participants who bought the contract will get paid off. The spread is continuously adjusted to guarantee that gamblers are equally likely to bet on either side. Actually money line bet is similar to the betting methods in election prediction market, especially winner-takes-all method. In the case of money line bets, the contract would pay $1 if a given team wins a particular game, which is the same as in the case of winner-takes-all method in election prediction market. Figure 11 has given an illustration of different kinds of bidding methods and the relationships between them.

![Figure 11: Different kinds of bidding methods and the relationships between them](image)

Point spread bet is more complex because one may desire detailed information regarding the entire point spread probability distribution (i.e., the probability Team A beats Team B by exactly k points). Goel et al.’s [12] work gave a good method to build this model. It pointed out that a possible approach is to create separately traded securities, one for each possible point difference. Each security would pay $1 precisely when its corresponding point difference was achieved, and its current price would indicate the market probability of that outcome. Goel et al. also pointed out that the significant disadvantage of this mechanism is that information does not automatically
propagate between securities even though they are clearly related, which means in the model the securities are independent with each other, but actually they are mutually related. This also demonstrates that it is often harder to predict a probability distribution than a single result, as indicated in section 3.1.

3.2.4 The meaning of price

Wolfers et al.’s [37] work proved that the current price of the contract is the market probability of the team winning. The proof is complex, which includes utility functions, differential and integral calculus, etc. This can be also explained in a more intuitive way. Let us imagine an election. Suppose M people think Democratic Party will lose, and N people think Democratic Party will win. As more and more people believe Democratic Party will win, the number of people who are going to buy the contract of the Democratic Party also increases. As a result, the price of the contract of Democratic Party goes up. On the contrary, if more and more people believe Democratic Party will lose, then there should be an increasing number of people who plan to sell the contract of Democratic Party. If they hold the contracts and the Democratic Party loses, they will get nothing in a winner-takes-all market, so it is better for them to sell the contracts before the result comes out. At this time, the price of the contract of Democratic Party falls down. Intuitively, the price is proportional to the percentage of people who believe Democratic Party will win. In practice, the prices of binary prediction markets have proven to be closely related to actual frequencies of events in the real world[21, 32]. We need to notice that when the participants are predicting, whether an event will happen or not is uncertain and it is just a probability. If the market prediction result is that “The probability that Democratic Party
will win is 90%”, it means that there is still a 10% chance that Democratic Party will lose the election.

3.3 Comparison with other information aggregating methods
A scoring rule is a widely used approach for information aggregation. It appeared long before there were prediction markets, and it aggregates estimations from participants. Those estimations are in the form of probability distributions. In decision theory, a score function or a scoring rule is a measure of one's performance when he or she is repeatedly making decisions under uncertainty. One example is the weather forecaster, every day he will give the probability of rain. A score function compares his/her forecast and the actual result to give the forecaster a score. If a large portion of the forecasts match the real case, the weather forecaster will get a high score, otherwise he/she will get a low score and is encouraged to perform better. We can define the reward function as \( u(x, q) \). \( q \) is the rain probability that the forecaster has stated. In the function, \( x = 1 \) if it rains, and \( x = 0 \) if it does not. If the weather forecaster wishes to maximize his/her reward, he/she will choose a forecast \( q \) which maximizes

\[
\hat{u}(u|p) = pu(1,q) + (1-p)u(0,q)
\]

where \( p \) is his/her personal probability that rain will fall. A scoring rule is a proper scoring rule if \( \hat{u}(u|p) \) is maximized when \( q = p \) for any value of \( 0 \leq p \leq 1 \). As a result, every participant will report his or her estimations honestly, in other words, he or she will let \( q = p \).

However, the scoring rule has met some problems. First, the number of states is exponential in the number of variables. When there are \( N \) binary variables, the possible state space is as large as \( 2^N \), which is very complex for the traders. As a result, it is difficult to compute with probability
distributions which have many variables. The second and also big problem with scoring rules is still unsolved. When people are asked to give estimations about a certain random variable, they tend to have different answers. But what we want is a single estimation which aggregate different information from people. This is the so called “thick market problem”. Currently most literature has not solved this problem[38].

Compared to prediction markets, scoring rules do not suffer from thin market problems, while prediction markets do. Scoring rules have the ability to effectively reveal information even if there are only a few participants.

Prediction markets have some different problems when compared to scoring rules. Even when prediction markets are legal, they often suffer from the thin market problem. When trading, traders have to coordinate on the assets they will trade. So it is very complex when there are a large number of possible events to define the asset, this case often exist in combinatorial prediction market because it is very complex to define the asset. In addition, traders must make agreement on when they are going to trade, for new public information might influence the value of asset. As a result, the trading activities are greatly limited, and the traders can only trade in a very small set of assets and are hard to trade other assets. However, prediction markets do not suffer from thick market problems. They have the ability to aggregate different opinions together.

An opinion poll is a survey of public opinion from a particular sample. It is a straightforward method to aggregate information. Berg et al.’s [23] work discovered that the Iowa Electronic Market predictions were more accurate 451 out of 596 times when compared to concurrent major
opinion polls on U.S. presidential elections. Chen et al.’s [34] work found that when compared to official Hewlett-Packard predictions of printer sales, internal corporate markets were more accurate 6 in 8 times. In addition, according to Pennock et al. [39], play money markets forecasted the 2000 Oscar Winners better than 4 in 5 columnists who made concurrent predictions. In a theoretical paper, Wolfers et al. [31] asserts that “In a truly efficient prediction market, the market price will be the best predictor of the event, and no combination of available poll or other information can be used to improve the market-generated forecast”. Chen et al.’s [1] paper compared the prediction accuracy of the market with the prediction accuracy offered by poll, and the number of participants for the two methods are the same. It indicated that if poll is informative enough (a high percentage of participants), poll-implied prediction has a higher accuracy than prediction market. However, if poll is not informative enough (a low percentage of participants, say 5%), the poll-implied prediction has larger prediction error than prediction markets. One possible reason is that when only a small percentage of people are selected to participate in the poll, they cannot successfully represent the whole population. As a result, the performance is likely to be affected by random sampling error.

From the literature above, we can see that both prediction markets and polls are very competitive in predicting future events. However, in most of the cases predictions markets outperform polls, and both of them have some biases.

Market scoring rules have the ability to combine the advantages of standard prediction markets and scoring rules. They are expected to produce an accurate prediction in both thick markets and thin markets. In essence, market scoring rules are sequentially shared scoring rules. As a result, if
many people use a market scoring rule (thick market), it in effect becomes an automated market maker facilitating trades between all participants, which means the function is like a prediction market. On the other hand, it becomes a simple scoring rule when there is only one person participating in the market. We can see a comparison between prediction market, scoring rule and market scoring rule in Figure 12. Here prediction market is indicated as “Simple Info Markets”.

Figure 12: A comparison of prediction market, scoring rules and market scoring rules

Figure 13 has given an illustration of the relationships between different methods of aggregating opinions. We can see that the three main methods are prediction market, scoring rule, and polling. The intersection of prediction market and scoring rule is market scoring rule, since market scoring rule combines the two aggregating methods.
In Figure 12 we have seen that the performance of market scoring rule is good in both thin market case and thick market case. As a result, combinatorial prediction markets are usually implemented by using market scoring rules[12-13]. Because combinatorial prediction markets are likely to suffer from thin market problem, and market scoring rules can effectively solve such a problem. In addition, market scoring rules make the implementation of markets easy, because instead of using several different markets to combine the possible outcomes, only one market is made use of, and traders try to minimize the sum of their errors over all predictions. But many of the simple prediction markets are not implemented by market scoring rules. The first reason is that market scoring rule is complex, and simple prediction markets can be implemented by a simple mechanism. The second reason is that simple prediction markets are using continuous double auctions (CDA), while market scoring rules do not. The CDA is a mechanism which matches buyers and sellers of a particular good or contract. In a CDA, people who want to buy the contract submit their bids, and simultaneously people who want to sell the contract submit
their ask prices to an auctioneer, and the auctioneer matches the buyers and sellers up. So the working principle of a CDA is different from the principle of market scoring rules.

### 3.4 Summary

Prediction markets can be divided by several different criterions. They can be divided into combinatorial prediction markets and simple prediction markets according to the contract complexity. The prediction markets can also be divided according to the application. Currently two main applications are sports prediction markets and election prediction markets. Most of the sports prediction markets are combinatorial markets, and most of the election prediction markets are simple markets. In addition, prediction markets can be divided into play-money markets and real-money markets. It is often believed among economists that play-money markets have a better performance than real-money markets. However, research has shown that play-money markets have the ability to perform as well as real-money markets.

The mechanism of prediction markets includes several issues. The first issue is participant selection. Participants are most important parts, for they are people who can absorb and share information. Participants can be selected uniformly if we do not know who are experts and who are novices, which is usually the case. But preferential attachment can be used when there are parts of people who are experts, or some people get access to more information than others.

The second issue is bidding methods. For election stock markets there are two basic types, which are winner-takes-all and vote-share. For sports prediction markets there are also two basic types,
but they are different from those of election prediction markets. The two basic types of sports prediction markets are money line bets and point spread bets.

The third issue is the utility function. It is important because in a prediction market, people have to decide how many contracts they want to buy. Since all participants are rational, they always want to choose an optimum amount of contracts so that they can get the most from the market. A utility function has the assumption that the participants are risk-averse and at the same time want their wealth to increase. There are many different kinds of utility functions, and the most commonly used are constant absolute risk aversion function (CARA) and generalized logarithmic utility function (GLU). GLU has a very low risk-aversion while CARA has a higher risk-aversion.

The last, but not the least issue is the meaning of price. Research shows that the current price of the contract can be explained as the probability that the event will happen. The proof is complex and abstract, but the problem can be explained by the relationship between supply and demand. The price goes up as more and more people believe the event will happen (want to buy the contract), and it goes down as more and more people believe the event will not happen (want to sell the contract).

Besides prediction market, there are other approaches which can aggregate information. Table 1 has given a summarization of advantages and disadvantages of several mechanisms which aggregate opinions, including prediction markets, scoring rules, polling and market scoring rule. The comparison includes whether participants make use of probability in prediction, whether
there are feedbacks between participants, whether the method suffers from thick-market problem, and whether the method suffers from thin-market problem.

<table>
<thead>
<tr>
<th></th>
<th>Make use of probability</th>
<th>Feedback</th>
<th>Thick-market problem</th>
<th>Thin-market problem</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prediction Markets</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Scoring rule (Opinion Pools)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Polling</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Market Scoring Rule</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>
4 Intersection of prediction market and social network

4.1 An introduction of social network

A social network is a social structure made up of two basic elements. The first element is called a “node”. In a real society nodes are usually individuals or organizations, such as people, schools, and companies. The second element is called a “link”, which is a specific type of relationship between nodes and represent the interdependencies, such as friendship, kinship, co-authorship, etc. In a social network, nodes are connected by one or more types of links.

Figure 14 gives an example of social network diagram. In figure 14 nodes are the black points, and links are the blue lines. The social network diagram shows how the society is partially connected, or clustered, by its relationships. Some of the nodes have no links with other nodes and they are isolated. Links can be unidirectional or bidirectional. For example, if the relation is “Person A is the parent of person B”, then the link is unidirectional. But if the relation is “Person A is a friend of person B on Facebook”, then the link is bidirectional. Because on Facebook, if A is B’s friend, B is also A’s friend.
Social network research focuses on the relationships between individuals, not just the attributes of individuals or group. Examples of relationships include similarity, influence, etc. One example of the social network influence on human behavior is why people are recently saving money. The reasons can come from different levels of scope, that is, from a macroscopic view, it can be the current economic depression; and from a microscopic view, it can be the influence from our friends, parents, neighbors and so on. The butterfly effect is another example of the complex relationship in social network. In Chaos theory, the butterfly effect means the sensitive dependence on initial conditions, and a small change can cause large differences in later stages. In social network, it means one’s behavior can be affected greatly by indirect relations. For example, a study in Japan[40] found that women with small social networks showing more than twice the death rate.
Social network research combines different academic fields, such as computer science, mathematics, sociology, and others. The research methods of those fields are usually different. For example, mathematicians sometimes simplify the relations between individuals so that they can develop some theories, but maybe in the sociologist’s view, this relation is complex.

For both individuals and for structures, one main question is connections. Typically, some individuals have lots of connections, others have fewer. Differences among individuals in how connected they are can be extremely consequential for understanding their attributes and behavior. More connections often mean that individuals are exposed to more diverse information. Highly connected individuals may be more influential, and may be more influenced by others. Differences among whole populations in how connected they are can be quite consequential as well. One application of social network is the spread of disease. Disease and rumors spread more quickly in a highly connected network than in a sparse network.

Besides direct connections, the distance between individuals is the second important factor. Some individuals can reach almost all the other members in the social network since they tell their friends, and their friends can tell their friends’ friends. Finally everyone knows the story. But for some other individuals, although they tell their friends, those friends have very limited connections and the story is not successfully spread.

The social network data can be either directed or undirected. With directed data, it means that if node A is connected to node B in a social network, then B does not have to be connected to A. One real-world example of this directed social network data is twitter. In Twitter, for each
individual we can see a list of “followers”, which means people who follow the individual. We can also see a list of “following”, which means the people the individual follows. We can see that the data is directed, because for each individual the “followers” and “following” can be different. Though it is very likely that the “followers” and “following” might overlap, they do not have to be the same. For example, some people just choose someone they are interested in to follow, such as some movie stars, and they do not have to know each other in person. In this case, some famous people are always followed by many individuals, and they do not have to follow back.

Another type is undirected data, which means in a social network, if node A is connected to node B, then node B must be connected to node A. Facebook can be a good example here. In Facebook, every relationship is undirected, or bi-directed. For example, if person A is person B’s friend on Facebook, then person B is also person A’s friend. The “friendship” in Facebook can be rephrased as a “network”, because it is possible that although the two people who are friends on Facebook did not meet in person, one of them is interested in the other and wants him/her to be included in the network. There is a trend that some people like to include some popular stars
into their networks, but they don’t have to know each other in person. LinkedIn is another good example of undirected data.

In addition, numerous work [4, 41-46] provides a good introduction about social networks. Issues include complex network models, degree distributions and correlations, community identification and measurements, measurement of network dynamics and perturbation. Since this thesis mainly focuses on prediction markets, I will put emphasis on how prediction markets work in the context of social network instead of the social network itself.

4.2 Previous work

4.2.1 Background
Social networks have the potential to understand more about the impact of social structures on the information flows among participants and thereby to potentially improve the performance of prediction markets. Previous work about prediction markets has not considered the mutual relationship between participants, mostly focusing on how participants are influenced, as individuals, by the current market price. Putting participants into a social network context is a promising topic, for prediction markets, especially those oriented towards business applications, may actually operate in a small society where people interact with each other. In this context, people can exchange their opinions within their social network about the upcoming event, such as whether Democratic Party or Republic Party will win the 2012 election.

4.2.2 Social network theory
Festinger et al.’s[2] work found that similar perceptions exist between members of social groups. Such a phenomenon exists especially in the small, face-to-face groups, and the behaviors and
attitudes of the group members tend to be more similar than outside members. One factor is called homophily, which means people with similar attitudes and perceptions tend to be attracted to one another. So on the one hand, the social network structure can change people’s behaviors, for individuals can be influenced by their group members. On the other hand, the social network structure can be adjusted according to people’s behaviors. Once the people with similar attitudes attract each other, the social network structure can change. Festinger et al.’s work also discovered that information, opinions, attitudes and perceptions flow through network ties. Burt’s [3] work discovered that overall network structure as well as individual network characteristics influence information flows. We will further discuss Festinger’s theory and Burt’s theory in Chen’s model[1] and our own model[47].

4.2.3 The Google application

Currently only a few papers[1, 22] have started working in this area. One is Cowgill et al.’s[22] work. For each pair of the 1,463 prediction market traders in Google, they calculate measures of their geographic, organizational, and social proximity and their demographic similarity. The geographic proximity includes several levels: same city, proximity within city, same building, same floor, proximity on floor, and same office. The organizational proximity includes same Senior Vice President, same “2-levels-below-CEO” manager, same “3-levels-below-CEO” manager, 1-2 steps away on organization chart, and 3 steps away from organization chart. The social proximity includes self-reported professional relationship, self-reported friendship, and number of overlapping email lists. The demographic similarity includes the levels of
undergraduate majors (computer science, electrical engineering, social science/law, etc) and graduate degree.

Cowgill’s research found that employees who sit near each other, share an office or their offices are located on the same floor have correlated trading behaviors. In addition, they discovered that organizational proximity, such as sharing a manager, being someone’s manager, being someone’s manager’s manager, leads to correlated trading. They found that measures of personal connections do not explain trading correlations well, but a history of peer-reviewing code or overlapping on a project does. Also, they did not discover a meaningful relationship between demographic similarity and similar trading behaviors.

The results show the importance of geographic and organizational proximity in explaining information flows in companies. Since geographic proximity and organizational proximity are a part of relationships in social network, it reflects the influence of social network on people’s trading behaviors. But it did not consider how such correlations might influence the accuracy of market predictions.

Another recent application of Iowa Electronic Market is the movie box office predictions, which we learned of by personal communications. In the market traders are asked to submit a forecast of the 4-week box office performance supported by a 2-4 page justification. They were also required to submit a survey on the time and energy they spent on the forecast, which measured the reliance of their predictions. In addition, their work combined prediction market and social
network. For example, they tried to use the diversity in people’s information to help with the prediction. Through the communication we find this idea interesting, and plan to see the effect of social network structure on the overall market prediction as the first step in our simulation.

4.2.4 Chen’s model

Chen et al.’s [1] work proposed a prediction market model in which participants are influenced by their social relationships. More specifically, it adopted a more realistic model in which participants are influenced by their neighbors in a social network. Chen used agent-based simulation for the model and did not conduct human experiments.

In Chen’s model, there are N individual agents, each corresponding to a node in the social network. Each individual has a piece of binary information, denoted as $s_i \in \{0, 1\}$ for individual $i$. If individual $i$ plans to vote for the Democratic candidate, then $s_i$ equals 1, while $s_i$ equals 0 if the individual plans to vote Republican. Agents are placed into a generated social network, and then each agent’s position is rearranged in the network according to its voting preference. The model follows Festinger et al.’s [2] theory that people with similar behaviors and attitudes tend to be attracted to one another. So in the model, an agent registered as Democratic will prefer to be located in a neighborhood with more Democratic voters (and similarly for Republican voters), so the model simulates that trading agents end up being highly clustered by voting preference. More specifically, the algorithm keeps switching the positions of agents according to their voting preferences until all people with similar beliefs are clustered. For example, if individual $A$ plans to vote for the Democratic Party and most of its neighbors plan to vote for Republic, and
individual $B$ plans to vote for the Republic Party while most of its neighbors plan to vote for Democratic, then individual $A$ and $B$ are likely to exchange their positions.

Figure 16 through Figure 19 have given us an example about how agents exchange their positions in a social network based on their voting preferences. In step 1, as shown in Figure 16, nodes A, E, F, G, J prefer to vote for the Republic Party, so they are indicated as R-nodes. Nodes B, C, D, H, I prefer to vote for the Democratic Party, so they are referred as D-nodes. The table in Figure 16 has two columns. The first column is about the percentage of D-nodes around each R-node, and the second column is the percentage of R-nodes around each D-node. Each time, the algorithm counts the percentage of different preference in the neighborhood for each node and the highest value from each column is selected. We can see that in the table, node A, as an R-node, has the highest percentage of D-nodes around it (100%). Node B, as a D-node, has the highest percentage of R-nodes around it (100%). So the algorithm changes the position of nodes A and B.

**Figure 16: An example of how agents exchange their positions in a social network according to their beliefs: Step 1**
In step 2 we can see that the positions for nodes A and B have been exchanged. The social network becomes more clustered since we can see from Figure 17 that for each node, the percentage of different preference in the neighborhood becomes less. We can see that in the table, node A, as an R-node, has the highest percentage of D-nodes around it (100%). Node H, as a D-node, has the highest percentage of R-nodes around it (100%). So the algorithm changes the position of nodes A and H.

![Figure 17](image)

*Figure 17: An example of how agents exchange their positions in a social network according to their beliefs: Step 2*

After nodes A and H exchange their positions the algorithm goes to step 3, as in Figure 18. We can see that the social network get even more clustered than in step 2. In this round, node E, as an R-node, has the highest percentage of D-nodes around it (50%). Node I, as a D-node, has the highest percentage of R-nodes around it (100%). So the positions of nodes E and I are changed in this step.
In step 4, as in Figure 19, the social network forms two clusters, so the algorithm ends here. All participants who plan to vote for the Democratic Party (blue nodes), as well as all those who plan to vote for the Republic Party (orange nodes), are clustered together. When the social network structure gets more complex, it might form multiple clusters.
After the individuals exchanged their positions, they began to form their beliefs based on the information in the neighborhood. The model follows Festinger et al.’s [2] theory that information, opinions and attitudes flow through network ties. The neighborhood of an agent includes itself and all individuals connected to it. Each individual agent treats information in its neighborhood as a representative sample of information in the society. Thus, we can get agent $i$’s prediction of the probability that Democratic Party will win based on agent $i$’s neighborhood as:

$$p_i = \frac{\text{# of D nodes in } i's \text{ neighborhood}}{\text{Total # of nodes in } i's \text{ neighborhood}}$$ eq. 4

Here D-node means an agent who is registered for (which means will also vote for) the Democratic Party, R-node means an agent who is registered for (which means will also vote for) the Republic Party. The proportion of Democratic votes in agent $i$’s neighborhood is $p_i$. Similarly, the proportion of Republic votes in agent $i$’s neighborhood is $1 - p_i$. By this step, each individual has its own prediction according to the information it gains from the neighborhood.

After each individual gets its own prediction, it uses a certain utility function to decide how many contracts to buy/sell in order to maximize the profit, or, in other words, it begins to “gamble”, and the market finally reaches an equilibrium price, which is also the aggregated prediction.

Chen’s model chose two utility functions, CARA and GLU, as defined in section 3.2.2, to model the agent’s preference in trading. Comparatively, CARA is more conservative and GLU is more aggressive, which means when facing the same situation, agents tend to trade more when they
are using GLU utility function. This can be reflected in the trading volumes. Chen at el.’s [1] work shows that when the agents use GLU utility function, they have a higher trading volume than when they use CARA utility function. One possible reason is that an investor with a logarithmic utility function (GLU) has very low risk aversion, while an investor with a constant absolute risk aversion function (CARA) has a higher risk aversion. For CARA, the investor’s portfolio becomes more conservative very quickly with any increase of wealth. The agent puts the same absolute amount of investment no matter what its wealth is. Although the trading volume is influenced by the utility function, the prediction is robust under the difference of agent utility functions.

Also, they varied the percentage of participation in the prediction market from 0% (no participants) to 100% (all people in the social network participate). Their work shows that the performance of prediction market is robust even if participate changes.

Another important thing is how the agents are selected to participate in the market. Chen’s model compared two methods: The first one is preferential attachment, which means individuals who can get access to more information are more motivated to participate in the prediction market. Preferential attachment assumes that the degree of nodes follows a Power-Law Degree distribution. The probability that an agent will participate in the prediction market is related to how well it is connected in the social network, which is modeled as:

\[ p_i = \frac{k_i}{\sum_j k_j} \quad \text{eq. 5} \]
Here $k_i$ is the degree of node $i$, $\sum_j k_j$ is the sum of the degree of all the nodes. So the larger the degree of the node is, the more likely that the agent will participate in the market.

The second one is the uniform attachment, which means the participants are equally likely to participate in the prediction market regardless of their characteristics. According to Burt’s [3] theory, overall network structure can influence information flows. However, the experiment result shows that whether to choose preferential attachment or uniform attachment does not change the performance of prediction market.

![Figure 20: An example about participant selection using preferential attachment](image)

Figure 20 has given an example about participant selection using preferential attachment. We can see that on the left side of Figure 20, nodes B, G, M, O, P, Q have relatively higher degrees, so the probability for them to be selected into the prediction market is relatively higher. So in the prediction market, the social network structure is actually indicated on the right side of the figure, which is a subgraph of the left side of the figure. The subgraph includes the nodes selected and all the links which connect those nodes. From the example we can see that there are totally 16 nodes, but only 6 nodes actually participate, so the participation rate is $6/16 = 37.5\%$. 
4.3 My work: Adding voting certainty level

4.3.1 Adding uncertainty

Chen’s [1] social network/prediction market model assumes that agents had perfect certainty of their own opinions or actions, which means if an individual agent is registered with a given Party, then it always votes for that Party’s candidate. But in the real world this is often not true, and for any given election, an individual might not vote for the candidate whose party he or she favors. Also, Chen’s model assumes that agents had perfect knowledge of their contacts' opinions or actions. In the real world case, it is often the case that we cannot get information from the people in our network perfectly.

Our extension considers those two cases and incorporates two kinds of uncertainties into the multi-agent simulation, the uncertainty of the agents’ own opinions and the uncertainty of their neighbors’ opinions. In Chen’s model, agents are registered to either the Democratic or Republican Party, and they will definitely vote for the same Party. However, in our model this does not guarantee that they will vote for the registered Party’s candidate, but only indicates that they will vote for the Party’s candidate with some likelihood.

In our model each individual $i$ has four possible states denoted as $s_i \in \{R_0, R_1, D_0, D_1\}$. Here $R_j$ means an agent is registered with Republican Party and $D_j$ means an agent is registered with Democratic Party. The agent votes for Republican when $j=0$, and votes for Democratic when $j=1$. 
For example, if $s_i$ equals $R_0$, then individual $i$ is registered with the Republican Party and votes Republican (as in the original Chen model).

In addition, we considered the case that the agents don’t know their neighbor’s preference to vote in our simulation, which is referred to as Decide-last method below.

Those two kinds of uncertainties, uncertainty about one’s own opinion and uncertainty about the neighborhood’s information, can cause two effects. The first effect is the influence on the level of clustering. In Chen’s model, agents are randomly placed into a social network and then move to neighborhoods where there are more voters like themselves. In that model, traders know who they will vote for with no uncertainty, resulting in a high level of clustering. However, in our model, while traders are registered to a given Party, there may be some uncertainty about whether they will actually vote for that Party’s candidate. Traders are clustered based on their Party registration similarities, but once they vote, the level of voted-Party clustering may be quite different than that registered-Party clustering.

Figure 21 provides an intuitive illustration of how trader’s voting uncertainty influences the level of clustering. We assume that the voting certainty level, which is the probability an agent will vote for a Party’s candidate given that it is registered for that Party, is the same across all agents. On the left most side of Figure 21 the voting certainty level is 50%, which means the agents are equally likely to pick either a Democratic or Republican candidate regardless of Party affiliation. We can see that people who are going to vote for the Democratic Party are no longer clustered...
together, nor do the people who are going to vote for the Republic Party. In the middle part, the 
voting certainty level is 80%. We can see that their voting decisions are more clustered than 
when the voting certainty level is 50%, but less clustered than their previous registrations. On the 
right most side, the voting certainty level is 100% and the graph is exactly the same as Figure 19. 
The model is effectively Chen’s model where the agents definitely vote for their Party’s 
candidate.

![Figure 21: Different voting uncertainty causes different level of clustering](image)

The second effect is the influence on the accuracy of information aggregation. Not surprisingly, 
a better knowledge of one’s own opinion, as well as a better knowledge of the information from 
neighbors, results in higher prediction accuracy. In section 4.3.2, we quantitatively compare the 
different knowledge of the neighborhood’s opinions (Decide-first method versus Decide-last 
method) with resultant prediction accuracy.

### 4.3.2 Implementation and results

In these preliminary experiments, we compared market prediction accuracy and trading volume 
for different levels of voting certainty and clustering. We used the same two utility functions for 
our traders as used in Chen et al.[1]: constant absolute risk aversion (CARA) and generalized
logarithmic utility (GLU). We also varied the percentage of participation in the prediction market from 0% (no participants) to 100% (all people in the social network participate). For the social network generation, we used the Power-Law Degree distribution, as mentioned in Chen et al.’s paper[1]. Each time, the probability for an agent to be chosen into the prediction market is proportional to how many people it is connected to, which is indicated as the preferential attachment in Chen et al.’s paper. And our method also exactly follows Figure 20. In addition, we also assume that people with similar attitudes and perceptions tend to cluster together, which follows Festinger et al.’s [2] theory. Similar to the Chen model, each agent’s position is rearranged in the network according to its voting preference, and finally the agents get clustered based on their registrations for the Parties.

We extend the Chen model by adding the voting certainty level and deciding when the agents fix their voting decision, as indicated in the middle part of Figure 22. The voting certainty level implements the first kind of uncertainty, which is the uncertainty for an agent about its own opinions. The decision making of when the agents fix their voting decision implements the second kind of uncertainty, indicated as the imperfect knowledge of an agent from its neighbors’ opinions. While agents are always clustered before their voting decisions are fixed, we examine the effects of allowing them to fix their vote choice either before or after they have exchanged vote information with their neighbors. The contents below will describe the possible outcomes when these two factors interact with each other.
For the first factor, the voting certainty level, we varied the level from 50% to 100%. In our experiments, we assume that the voting certainty level to vote for the candidate proposed by their Party is the same across all agents. Thus if the voting certainty level is 75%, then all registered Democrats have a 75% chance of voting Democratic and a 25% chance of voting Republican, while all registered Republicans have a 75% chance of voting Republican and a 25% chance of voting Democratic. Each agent incorporates voting information from its neighbors and then bids on the likelihood of a Democratic winner. This factor models the first kind of uncertainty, which is the uncertainty of participants about their own opinions.

The second factor is when agents fix their voting decisions. One option is to fix agents’ decisions before exchanging neighborhood voting information, so that agents know precisely their neighbors’ votes. The other option is to fix the voting decision after exchanging voting information meaning that agents have only a probabilistic idea of their neighbors’ actual votes. This factor models the second kind of uncertainty, which is the uncertainty of participants about their neighbors’ opinions.

The overall market process is illustrated in Figure 22. In the left part of Figure 22, agents decide on their vote before exchange voting information. We call this method decide-first. Similar to Chen’s model, the neighborhood of an agent includes itself and all individuals connected to it. Thus, we can get agent i’s prediction of the probability that Democratic Party will win based on agent i’s neighborhood as:
whose representation is the same as eq.1. After each agent exchanged their place in the social network based on their Party registrations, they start to get their voting decisions based on their Party registrations and voting certainty level. We can see that in decide-first method, the participants have a good knowledge about their contacts’ opinions. What is different from eq.1 is, here D-node means an agent who will vote for Democratic Party (but not necessarily have registered for Democratic Party before), R-node means an agent who will vote for Republic Party (but not necessarily have registered for Republic Party before). The proportion of Democratic votes in agent \(i\)’s neighborhood is \(p_{i,\text{first}}\). Similarly, the proportion of Republic votes in agent \(i\)’s neighborhood is \(1 - p_{i,\text{first}}\).

We call the method in the right part of Figure 22 the \textit{decide-last} method since agents don’t fix their voting decisions until after exchanging voting information with their neighbors. Here we indicate the agent \(i\)’s prediction of the probability that Democratic Party will win as \(p_{i,\text{last}}\). Thus agents’ assessment of their neighborhood sample, \(p_{i,\text{last}}\), changes when voting information becomes uncertain. In this case, we compute the likelihood rather than the actual proportion of Democratic or Republican voters as in the decide-first case. For example, assume the voting certainty level is 75%. Then suppose in an agent’s neighborhood that 4 agents are registered as Democrats and 2 as Republicans, then the likelihood of Democratic voters in the neighborhood is \(75\% \times 4 + (1-75\%) \times 2 = 3.5\), while the likelihood of Republic voters, in its neighborhood is (1-
75%)*4+75%*2 = 2.5. We can compute the likelihood of all-nodes in i’s neighborhood as the sum of the likelihood of Democratic voters and the likelihood of republic voters. As a result, (eq. 4) becomes:

\[
p_{i,\text{last}} = \frac{\text{The likelihood for D nodes in i’s neighborhood}}{\text{The likelihood for all nodes in i’s neighborhood}} \quad \text{eq. 7}
\]

Clearly the actual timing of when agents make their voting decisions would be expected to have an effect on the resultant information aggregation accuracy. In the decide-first case, agents have perfect voting information both about their own vote and their neighbors and while having only probabilistic information in the decide-last case. So the hypothesis is that in decide-first, both the prediction accuracy and trading volume are higher than in the decide-last scheme because the agents have more precise information.
Figure 22 (Left): Agents vote at the very beginning

(Right): Agents vote at the very end

Table 2 gives a comparison of the three different models of agent decision making. We can see that the sequence of the decide-first model replicates Chen’s model. However in the Chen’s model, if an agent is registered as to a certain Party, then it will vote for the same Party, but in our decide-first model, we have voting certainty levels ranging from [0.5, 1]. So Chen’s model exactly matches the 100% voting certainty level case in our decide-first model.
Chen’s model measures the absolute error as the difference between the real value of the contract \( v \) and the equilibrium price \( P^* \). However, when introducing uncertainty, we had to modify this measurement as agents’ bidding behaviors were highly determined by their utility functions. Even when the agents are 100% certain about their voting behaviors, agents’ with the GLU utility function bid only a little higher or lower than 0.5 that a future contract payoff will be $1 (i.e., more conservative), while agent’s with CARA utility functions tend to bid closer to 0 or 1 that future contract payoff will be $1 (i.e., more aggressively). Since the equilibrium price greatly depends on the voting certainty level, when the voting certainty level is high, the equilibrium price will be near $1, while when the voting certainty level is low, the equilibrium price is near $0.5. As a result, for GLU agents, it appeared they were most accurate when voting certainty levels are around 50% and less accurate when voting certainty levels are 100%.

To avoid this problem, our model adopted a different calculation of absolute error, which compares the contract payoff and the market prediction, since both are 0-1 values. According to Chen’s model, a contract payoff is 1 if more than half of people vote for the Democratic
candidate and 0 otherwise. For the market prediction, the 0-1 value means “Will the market think the Democratic Party will win?”, where the value is 1 if the equilibrium price is larger than 0.5, and 0 otherwise.

Figure 23 to Figure 26 shows the relationship between prediction errors and percentage of agents who participate in market with different voting certainty levels. In each of the four figures, the voting certainty level for each participant changes from 50% to 100%. The X-axis is the percentage of the population that actually participates in the market. The Y-axis is the average prediction error of the market. Figure 23 and Figure 24 are decide-first methods for the CARA and GLU utility functions, while Figure 25 and Figure 26 are decide-last methods for those two functions separately. When we compare the decide-first method and decide-last method, one noticeable difference is that in decide-last method, the prediction error increases very fast with the decrease in voting certainty levels. While in decide-first method, the prediction error does not change as quickly with the decrease of voting certainty levels. And for both methods, CARA and GLU utility functions have similar performance.

![Figure 23: Decide-first method, CARA utility function. The agents vote before they trade.](image_url)
One reason for the difference in those two methods is that the decide-first method only contains the first kind of uncertainty (about their own opinions), but it does not contain the second kind of uncertainty (about their neighbor’s opinions). After they exchanged their beliefs based on their registration and got their voting preference, the participants can have a good knowledge about which Party their neighbors are going to vote. So the uncertainty only has an effect of clustering. However, in the decide-last method, the participants actually have no idea about which Party their neighbors are going to vote. They only know their registrations and the only thing they are do is to take a “guess” about their neighbors’ voting preference. As a result, both of the two kinds of uncertainties are involved. They have an effect of clustering as well as prediction accuracy.
In addition, when we look into the decide-first method and the decide-last method, we observe that in both methods, the market prediction accuracy decreases with the decrease of voting certainty level (and the level of clustering also, according to the analysis in section 4.3.1). Burt’s [3] theory indicates that overall network structure can influence information flows. Here our experimental results match Burt’s theory and we can see that the information has a better aggregation when the social network is more clustered.
To provide a baseline for measuring the market accuracy, we performed a poll across the agents. In the poll each agent votes for either Democratic Party or Republic Party and if more than half of people vote for a given Party, say Democratic, then the market predicts that Democratic Party will win, and the same is true for the Republic Party. This poll was run for the decide-first cases, across all voting uncertainty levels and we found that the prediction market has a slightly better performance than poll. Our future work will include extending our poll to be used in the decide-last method. Since in this case, the actual information aggregation process must deal with uncertain information, such a poll can be particularly useful as a performance baseline.

In Figure 27 through Figure 30, the voting certainty level for each participant changes from 50% to 100%. The X-axis is the percentage of the population that actually participates in the market. The Y-axis is the average trading volume. From Figure 27 and Figure 28 we can see that, for both CARA and GLU, the average trading volume increases with the increase of voting certainty level. Not surprisingly, if individual agents have a better knowledge about who they are going to vote, then they will be likely to buy or sell contracts, and trading volume increases.

From Figure 29 and Figure 30 we see that as the voting certainty level decreases, the trading volume decreases very quickly for the decide-last case. Adopting this explanation to the real world, it makes sense to assume that when all the agents know who their neighbors plan to vote for before they trade, they have more accurate information, which makes them more confident so that they bet more.
In addition, we can see the effect of clustering on trading volume. The more agents are clustered, the higher the trading volume. We can see this effect in both Decide-first case and Decide-last case. When the voting certainty level is 100%, which means the agents’ believes get most clustered, the trading volume is highest. When the voting certainty level is 50%, which means the agents’ believes get least clustered, the trading volume is lowest. So less clustering means more diverse opinions and less trading. In Decide-last case the differences of trading volume between different voting certainty levels are larger than in Decide-first case, because for the agents, a better knowledge of their neighbors’ opinions could lead to more trading.

Figure 27: Decide-first method, CARA utility function. The agents vote before they trade.
Figure 28: Decide-first method, GLU utility function. The agents vote before they trade.

Figure 29: Decide-last method, CARA utility function. The agents vote after they trade.
4.4 Summary

Currently the intersection of prediction market and social network is a promising area, since in most cases, people in a prediction market are also in a social network context. Until now not much work has been done in this area, save for Chen et al.’s [1] work that effectively simulated how people in a social network might trade in the prediction market, given the assumption that agent’s voting information are certain.

Figure 30: Decide-last method, GLU utility function. The agents vote after they trade.
5 Discussion and conclusion

Prediction markets have many advantages, such as it is easy to aggregate information from diverse human (or agent) sources. However, prediction markets face several challenges: First, in an agent-based market designers have to consider many issues in creating a prediction market, including participant selection, preference, utility function. Second, since most prediction markets are play money markets, whether the participants have enough incentives to participate in becomes a question. Third, it is important to solve the thin market problem in prediction markets. Fourth, participants in prediction market are actually in a social network, but there is little research on the effects of different social network structures on prediction market models.

This thesis focuses on the fourth challenge, the intersection of prediction market and social network. The intersection of prediction market and social network is a promising area for two main reasons: First, participants in a prediction market are actually in a social network, so modeling them into a social network matches the real-world case. If we combine prediction markets and social networks, then information can be shared by people explicitly, since people will communicate in a social network. Second, it is a brand new area and no much literature has addressed the socially-embedded prediction market research. This work has used simulation as an effective first step. Obvious next steps would include undertaking human experiments to support the agent-based simulations.

In this thesis, I modeled the interaction between prediction markets and social networks where agent’s voting information may be uncertain, thereby extending Chen’s work into more a general version. Chen’s model assumes that once the participants are registered for a given Party, their
beliefs will not change any more. In our model, participants’ voting preference can change after they are registered, which fits better with the real world case. Agents were placed into the social network so as to represent clustering of beliefs, or in this case, party affiliation. The agents participated in a winner-takes-all prediction market for an election between a Republican and Democratic candidate. Each individual makes a prediction based on its own behavior and the voting information that it learns from its neighbors in the social network. Our model investigates how information is aggregated when the beliefs contain uncertainty, which reflects realistic social aspects of individual behaviors.

Figure 31: An illustration of the conclusions

Discoveries are summarized in Figure 31. (D1) Higher voting certainty level leads to a higher level of clustering. (D2) Higher voting certainty level leads to a higher trading volume. (D3) Higher voting certainty level leads to a higher prediction accuracy of the market. (D4) Higher level of clustering leads to a higher trading volume. (D5) Higher level of clustering leads to a
higher prediction accuracy. (D6) Higher trading volume is likely to cause a higher prediction accuracy.

First, the higher voting certainty level leads to a higher level of clustering (D1). Figure 21 has illustrated this. From the figure we can see that when the agents’ voting certainty level is 100%, the network get clustered at its maximum. And the network structure get less clustered with the decrease of the voting certainty level.

Second, we observed the influence of the voting certainty level on both the information aggregation accuracy and trading volume (D2-3). The higher the voting certainty level, the more precise information people can get, resulting in higher trading volumes as well as information aggregation accuracy.

Third, we found that the structure of social network can change the trading volume and the overall market prediction accuracy, which is the most interesting (D4-6). When the agents are more clustered, the higher confidence in their beliefs lead to higher trading volume as well as higher prediction accuracy, while less clustering means more diverse opinions and less trading. One possible reason is that each individual predict based on its neighbors. If there are an even number of Democratic Party and Republic Party in the neighborhood, the participant might have no idea about which Party to trust based on the neighborhood information. So in this case, the voting certainty level and the level of clustering are mutually correlated. This also matches Burt’s [3] theory that the network structure influence information aggregation.
Future work includes clearly separating the effects of voting certainty levels from the effects caused by clustering. We also plan to determine appropriate error measures for prediction markets that are not win/lose markets. When a market’s equilibrium price is, say $0.50, it is unclear whether agents have accurately assessed the situation to be 50% likely or whether they are simply uncertain about what the situation is going to be.

Another interesting thing is that the result might change when there are multiple sources of information in one’s neighborhood in the social network. In our model, each individual only face two sources of information from the neighborhood: people who are in favor of Democratic Party and people in favor of Republic Party. In the real world, there are always more than two sources of information in one’s network, and the diverse information is helpful for a participant to make predictions, since the participant can get access to more information and thus make more rational decisions accordingly.

Longer term future work includes making a more precise determination of what it means for an agent to be influenced by its neighbors when their beliefs are uncertain since this uncertainty can arise from different causes. For example, it may be that agents have different reliability, which can cause a variation in the certainty of information they provide. These different reliabilities may be the result of an agent simply being less knowledgeable about a given topic or it may be that they are less trustworthy. Allowing individual agents to be represented as having different levels of expertise and/or trust would make this model more realistic. Finally, we would like to ground this model by basing it on real-world combined social network and prediction market
data. Currently no literature has been found to analyze the performance of socially-embedded prediction markets in the real world. So it is promising to apply the socially-embedded prediction markets to real models, such as a social network on the Internet, or a social network in a particular academic area. Potential applications can include Facebook, Twitter, and more.
References