2012

Go Niche or Go Home: Influence Maximization in the Presence of Strong Opponent

Long Foong LIOW
Singapore Management University, lfliow.2009@msis.smu.edu.sg

GO NICHE OR GO HOME: INFLUENCE MAXIMIZATION IN THE PRESENCE OF STRONG OPPONENT

LIOW LONG FOONG, MIKE

SINGAPORE MANAGEMENT UNIVERSITY

2012

Go Niche or Go Home: Influence Maximization in the Presence of Strong Opponent

by

LIOW Long Foong, Mike

Submitted to School of Information Systems
in partial fulfillment of the requirements for the Degree of
Master of Science in Information Systems

Thesis Committee

LAU Hoong Chuin (Supervisor/ Chair)
Associate Professor of Information Systems
Singapore Management University

CHENG Shih-Fen (Co-Supervisor)
Assistant Professor of Information Systems
Singapore Management University

Robert KAUFFMAN
Professor of Information Systems
Singapore Management University

Singapore Management University
2012

Go Niche or Go Home: Influence Maximization in the Presence of Strong Opponent
by LIOW Long Foong, Mike

Abstract

In hotly contested product categories dominated by a few powerful firms, it is quite common for weaker or late entrants to focus only on particular segments of the whole market. The rationale for such strategy is intuitive: to avoid direct confrontation with heavy-weight firms, and to concentrate in segments where these weaker firms have comparative advantages. In marketing, this is what people called "go niche or go home". The niche-building strategy may rely on "homophily", which implies that consumers in a particular market segment might possess certain set of attributes that cause them to appreciate certain products better (in other words, weaker firms would customize their products to target some particular market segments and not the mass market). On the other hand, the niche-building strategy may also rely on the network effect, which implies that consumers having social relationship would reinforce each other via their respective adoptions. In this case, weaker firms should recognize such inter-customer network and concentrate only on customers belonging to certain set of strategic clusters.

In this thesis, we present the model for building effective niche-seeking strategies as a sequential adversarial search problem (game) on an infinite horizon. For simplicity, we assume that the adoption choice depends only on the network effects (in other words, a customer will

choose the product that is chosen by the majority of her neighbor). The social network is directed, and there will be two firms, one with significantly more marketing budget than the other firm. Firms take turns making investment choices on which customer to convert. For both firms, their budgets are fixed over time and unused budget will not carry over to future time periods. Furthermore, we introduced a *contractual lock-in* constraint to represent an obligatory policy imposed by companies for preventing their customers from defecting to adversary's product choice. With this model, we manage to show that a simple strategy based on the evaluation of individual customer's "value" can effectively identify and secure niches within randomly generated scale-free networks. We also show that such niche-building strategy indeed performs better in the long run than a myopic strategy that only cares about immediate market gains.

We deployed the Minimax approach to strategically reason the two node selection strategy, and employed a simple $\alpha - \beta$ pruning mechanism to improve search performance.

Contents

List of Figures

List of Tables

Acknowledgments

Firstly I would like to thank my advisor, LAU Hoong Chuin for his guidance, advice, motivation and continuous support. The completion of this thesis would not be possible without his encouragement throughout these years. I also like to express my gratitude to my coauthors, mentors and committee members, CHENG Shih Fen and Robert KAUFFMAN for their guidance in my research work, answering all my questions and request for assistance toward the completion of this thesis.

I wish to thank the members of Intelligent Systems and Decisions Analytics group: Aldy GUNAWAN, Fu NA, Lindawati, Sharon LIM Yee Pin, CHEN HuaXing and Nguyen Thi DUONG for their inspirations, great ideas, insights and friendship. I am grateful to Aldy for his good will in guiding and advising me over various topics about research, university and family life.

Finally, I wish to express my deepest gratitude to my dad, LIOW Khoon Pong and my mom, LIM Suat Gnoh for their unconditional love and trust. Special thanks to my siblings, Teik Foong, Yi Foong, Kae Foong, Yuh Foong and Oi Lean for their endless love and advice on everything about life. My sincere gratitude to my wife, WONG Soke Kuen and lovely twin, Jia Hong and Wen Qing for their love and support toward the completion of my thesis.

Dedication

I dedicate this thesis to my family, especially to my understanding and patient wife, Soke Kuen, who has given me the full support over these many years of post-graduate study, and also to our precious twin, Jia Hong and Wen Qing.

Chapter 1

Introduction

Social systems have been shown to be an important factor in affecting consumption behavioral patterns since the 1960s. The seminal work by Bass [2] marks the dawn of an era where researchers begin to explore the significance of networks in explaining or predicting product adoptions or innovation diffusions. The Bass model is closely related to the work on network externality in economics (e.g., see [5]) in that it adopts a macroscopic view, investigating adoptions or diffusions at the industry level. In the Bass model, the impact of networks is aggregated as the count on *previous adopters*, and the future adoption is then a function of this aggregated count. The simple and elegant Bass model was later expanded to model adoptions of products with *successive generations* (e.g., high-tech products like DRAMS or consumer electronics) [16] and diffusion process with decision variables (e.g., price) [18] as well.

With the prevalence of technologies and devices that can accurately capture the digital traces of an individual (e.g., smartphones or social network sites like Facebook) recently, it becomes increasingly plausible to investigate adoption or diffusion processes at microscopic level. Now researchers are able to investigate and infer the micro-structures that are behind these macrooutcomes instead of fitting the observed statistics at macro-levels. Such micro-structural insights can be utilized to explicitly describe ripple effects of adoption or diffusion among interconnected individuals, and this lead to the intensive study of cascading phenomenon. In particular, researchers are studying how one could maximize the influence/ diffusion/ adoption in a given network through a targeted set of individuals which was well defined as the *influence maximization problem* by [12]. More concretely, given a directed weighted graph, G =(V, E, W) with vertices V as users, edges E as relationships with weight function W: E→[0,1] which denotes the influence probabilities, the goal is to select a subset $S \subseteq V$ for initiating the diffusion process so as to maximize $\sigma(S)$, the number of vertices influenced by *S* at the end of diffusion process. The dynamics of influence propagation can be represented by one of many existing models, such as the linear threshold model, general threshold model and utility based model. Most of the propagation models in literature assume progressive activation in which an activated node cannot revert back to inactive mode. This assumption implied a consumer is unable to change his choice after an initial purchase, which is rarely practical in the context of business marketing.

In this thesis, the classical influence maximization problem is extended by introducing an adversary to the model and relaxing the assumption on progressive activation. This research is motivated by an emerging e-Commerce practice known as *influencer marketing*, which channels marketing focus (i.e. invests) on specific key individuals, known as *influencers*, instead of targeting the mass in conventional marketing strategies. The main idea is to generate substantial awareness and subsequently possible sales from potential buyers which surround these influencers. The *influencers* serve as conduit to the entire buyer segment, and are perceived as individuals who shape the purchasing decisions of true potential buyers. Since all marketers are aware of such phenomena and may deploy the same practice to compete in the same market, it might eventually lead marketers to engage the same group of key individuals in a repeated manner. It is therefore essential, from a marketer's perspective, to design a marketing strategy (to perform influence maximization) strategically in the presence of adversary.

In this thesis, the study is narrowed to a duopoly market where two players (called the Incumbent/Adversary and the Entrant) competing for their respective market shares on a single product. Each player is endowed with a marketing budget, and is engaged in investing in customers with that budget over an infinite horizon. Of particular interest in this thesis is the assumption that the adversary is endowed with a higher budget. The natural question to ask is whether the entrant has the ability to target customers strategically to secure a *niche* consumer segment by fortifying the consumers choice through influence propagation effect. Our problem departs from the classical Influence Maximization Problem which studies how one could maximize her consumer segment by targeting a set of individuals to reinforce the consumers product choice.

Our problem is modeled as a two-player influence maximization problem under an infinite horizon. Both the Incumbent and Entrant are assumed to have the information about each other budget. The choice of consumers is governed by its utility, which comprises of network influence effect and investment on it (if any). Furthermore, a *contractual lock-in* constraint which prevents a customer from defecting to opponent's product choice within a pre-determined time period is introduced to the model. A minimax algorithm which allows players to reason strategically on which customers to invest at each time period is used to solve this problem, under the assumption that neither players are aware of the exact node selection policy used by the opponent. The above game is played sequentially, and the goal is to determine whether some form of steady state could be reached, and if so, whether the proposed *value-based* approach will exhibit a niche-seeking behavior that enables the Entrant to secure a larger share by targeting the niche market at steady state. Our results illustrated empirically that, under the tested conditions, the proposed *value-based* approach is niche-seeking in nature, and can indeed secure a larger pool of customers (performs better than the conventional count-based greedy approach) when competing against a stronger opponent.

In general, this research offers a broader view on the dynamics of competition between two players and how a collective behavior will emerge from changes in individual characteristics in an influence network under bounded rationality.

1.1 Problem Definition

The rules of game, players' role, their decision model and information sets are defined as follow. Our game model is comprised of non zero-sum, pure strategies, sequential two-players competition game with network influence. The objectives of both players are to maximize their respective number of adopters within the network. The structure of consumer markets is set to conform to *scale free* network topology which is conjectured to be a good representation for most social network structure. Both players have complete information of the ply depth used by their opponent but incomplete information about the type of node selection policy will be used by each other. The *Incumbent* player assumes her adversary (*Entrant* player) to deploy a simplistic count based approach. Each player gets to take a lead in starting the game in instigating persuasive incentive on a consumer based on her decision model, and her opponent will perform the same subsequently after observing the new game state. The game will repeat in this manner until either one of the following termination criteria is met. The termination criteria are, 1) repetition of game state and 2) four hundred game stage is played with no repetition of game state observed. Both players are forbade from targeting the same consumer within a specified game stages. We termed this constraint as *contractual lock-in* in this thesis.

1.2 Motivation

Our research is motivated by a new business practice known as *influencer marketing* which channels the focus to specific key individuals, *influencers*, within a group instead of targeting the mass in conventional marketing strategies. The main idea is to generate substantial awareness and subsequently possible sales from potential buyers whom surround these *influencer*. The *influencers* serve as conduit to the entire buyers segment, and perceive as people who shape the purchasing decision of true potential buyers. Since all marketers are aware of such phenomena and may deploy the same practice to compete in the same market. This will eventually lead the marketers to engage the same group of key individuals in a repeated manner. Therefore it is essential to explore an alternative strategic approach in focus on securing a niche consumer segment by fortifying the consumers choice through influence propagation effect. Our problem departs from the traditional *Influence Maximization Problems* by studying how one could maximize the niche (*loyal*) consumer segment by targeting a set of individuals to reinforce the consumer adoption choice. Our research amalgamated the two-players game with an extension of the classical influence maximization problem. The main ingredients comprised two competing players, non zero-sum, pure strategy, sequential influence maximization in a special case of oligopoly market structure. Both the *Incumbent* and *Entrant* are assumed to have complete but imperfect information of their opponent. The players are aware of the game tree search algorithm deployed by others but both are unaware of the exact node selection policy used by their adversary. In the context of standard game, we are unaware of any clear formalism that described and modeled our problem in such a manner so far.

Here we wish to illustrate the significance of our work using a simple example as follow. Suppose that the telecommunication service provider, Singapore Telecom ST comes up with a new "cloud" and attempts to sell this service through *influencer marketing* leveraging on the advantage of social network, such as Twitter, Facebook or other form of social networking platform. The conventional influence cascading models (e.g. linear threshold, general threshold, independent cascade, weighted cascade, decreasing cascade, and history sensitive models) provided a mean to identify the nodes that are more "influential" based on their propagation effects within the given social networking community.

However the presence of an adversary such as Starhub SH, another telecommunication service provider in Singapore, wish to launch a similar service in an attempt to capture the same community of users by containing the spread of Singapore Telecom's influencer marketing campaign. Such dynamics and interplay between two competing players presented here cannot be modeled using conventional models. The conventional influence cascading models dictate that each node u within the network should be in one of the following two possible states: 1) active (adopted the service by ST), and 2) inactive (has not adopted the service by ST). With the presence of an adversary like SH to our problem where he is not only trying to dissuade u to adopt the service by ST but actively attempt to persuade u to adopt a competing service by SH. Thus each u in the social networking community can be in one of the following three states: 1) active (adopted the service by ST), 2) active (adopted the service by SH) and 3) inactive (has not adopted either service by ST or SH). Besides this, the dynamic of cascading effects is also differ from the conventional models where propagation is allowed to diffuse through every active nodes, whereas the propagation is only allow to diffuse through activated node of similar type in our problem with the presence of an adversary.

Our findings can be applied in decision problems that require game theoretic modeling to mimic strategic interaction, especially in business marketing and economic studies where an adversary is presence. One specific example is the game of market penetration decision. A new player (*Entrance*) needs to decide whether to compete against a stronger adversary, the *Incumbent* in a new market. If the *Entrance* decides to do so, what strategy should be deployed to secure the largest pool of loyal consumers with strong resiliency against *Incumbent's* strategy. It is more economic viable to maximize the overall market adoption or simply to fortify a niche pool of loyal consumers?

1.3 Research Objectives and Contributions

We are interested to mine the duopoly consumer market using a sequential game model with network influence, for ascertaining an efficient node selection policy in securing the largest pool of loyal customer base given a customer network. We illustrated the viability of using the *value based* approach for achieving such objective by deploying the *alpha-beta pruning* heuristic in Minimax approach to evaluate the game states. Our work offers a broader view on the dynamics of competition between two players and how a collective behavior will emerge from changes in individual characteristics in an influence network under bounded rationality.

Chapter 2

Literature Review

Network influence are observable dispersion effects which normally originated from a small local group to larger interconnected structure through various means of diffusion. A phenomena which resulted from network behavioral interaction. Researchers in business and engineering schools studied the effects of network influence in a variety of domains such as riot control policies in social science context, the decision making process of politicians to reach a common consensus in politics, the outbreak of contagious disease in science, the interactions among yeast protein in biomedical, and viral marketing through online media system in business economics. The field of research can be broadly classified into two categories namely as the influence propagation models, and the algorithms to estimate the propagation effects.

The classical *Influence Maximization Problem* which necessitates an influence propagation model and means to estimate the propagation effects is used as one of the basis for our work. A formal description to express the *Influence Maximization Problem* as an optimization problem is provided by [12] as follow. A graph *G=(V,E)* represents a network of interest with *E* as the set of all binary connections between two vertices in set *V*. The$S \subseteq V$ defines the subset of vertices selected for initiating the diffusion process. The objective is to maximize total influenced vertices. The propagation models considered by the researchers were *Linear Threshold* and *Independent Cascade*, and the complexity of problem was proven to be NP-hard through reduction from the Vertex Cover and the Hitting Set Problems. The solution quality is also guaranteed to be an (1 - 1/e - ϵ) approximate to the optimal solution by exploiting the sub-modularity property inherited in this problem, illustrated through a greedy algorithm which adds node with maximum marginal gain into the seed set. We will cover the literature of former category in greater details in the section below.

On the other hand, research on strategic competitions are typically grounded on sequential search models. Sequential search is an approach used for investigating the dynamics of decision making logics through specified conditions of social interaction (e.g. player characteristics, search rules, informational assumptions and payoffs). Each player is assumed to have their own beliefs and utility function over the outcomes. The players' choices are to maximize (or minimize) their respective expected utility value.

Our problem departs from all earlier models in the following manner: our model allows a consumer to renew their product choice at different stages based on the payoff function, instead of progressive manner where the consumers are unable to change their product choice subsequently. We allow both players to instigate any eligible consumer to change his product choice at each stage in a sequential manner (which is different from the classical single-agent *influence maximization problem*). Thirdly a *contractual lock-in* policy is introduced to deter recent targeted consumer from defecting.

2.1 Influence Propagation Models

Given a social network with estimates of individuals' influence rating, Domingos and Richardson [4, 17] explored a fundamental Set Selection problem with influence propagation effect. The social network is represented as a directed acyclic graph (DAG), G where nodes represent consumers and the edges represent influence relationship. For a binary model, the node is *active* when it adopts a product or service and *inactive* otherwise. The tendency for a node to convert from *inactive* to *active* state is governed by one of the many diffusion models described later in this chapter. For these diffusion models, the node conversion process occurs in a progressive manner where an *inactive* node may convert to become *active* but not otherwise. Heuristics are normally employed to identify the k-cardinal of consumers to target when the propagation effects are probabilistic. This class of problems is known as *single agent Influence Maximization Problem*. More formally, the objective is to target a set of A such that the expected size of converted consumers, $\sigma(A) := E[|\varphi(A)|]$ is maximized [13]. The efficiency in achieving this is largely depends on the performance of the search algorithm used to determine set A over a very large network, which entails the empirical estimation of propagation effects. Some of the common estimation approaches are Mix Greedy Independent Cascade approach [3], Cost Effective Lazy Forward [14], Greedy algorithm [12], Degree Discount heuristic [3], and Stochastic Cellular Automata [7]. Garlick and Chli [6] extended the *Influence Maximization Problem* to an agent based simulation model for investigating the lock-in dynamic within consumer networks. Their proposed influence propagation model is composed of environmental factor, consumers' follower tendency, and products quality. Their simulation model assumed the agents to be aware of everyone else product choices in the network.

2.1.1 Linear Threshold Model

Numerous types of influence propagation models were proposed by researchers to describe various means of cascading effects under different conditions. Granovetter and Schelling [8] were among the first to propose models that capture the progressive change of nodes in networks. The concept of *Linear Threshold* model proposed by [8] was based on node-specific thresholds. An *inactive* node *v* at time *t-1* will become active (and remain active) at time *t* under the following condition:

$$\sum_{u \in N(v)} w_{v,u} X_{u,t-1} \geq Threshold(v) \tag{2.1}$$

The variable $X_{u,t-1}$ is 1 if *u* was active at time *t-1* and *0* otherwise. The variable $w_{v,u}$ denotes the degree of influence on *v* by *u* (the level of which *u* being active will contribute to *v* being active). Intuitively, if a predetermined fraction (*threshold*) of *v* is less than the sum of degree of influence from its *active* neighbors. This model is also known to be the foundation for a large body of work in the Sociology domain. Subsequently the Granovetter's Linear Threshold model is generalized as *General Threshold* [13] model as it can be deduced into a re-parameterization form of node *v* with monotone *activation function* $f_v \colon 2^v \rightarrow [0, 1]$, and activation threshold θ_v which is chosen independently, uniformly and randomly from the interval of (0,1]. A node *v* will become *active* at time *t+1* if and only if $f_v(S) \geq \theta_v$, where *S* denotes the set of active nodes at time *t*. This model is different from the Linear Threshold model as it focuses on the *cumulative influence* of all nodes from a set *S* instead of the individual attempts of nodes $u \in S$.

2.1.2 General Threshold Model

Subsequently the Granovetter's Linear Threshold model is generalized as *General Threshold* [13] model as it can be deduced into a re-parameterization form of node *v* with monotone *activation function* $f_v\colon 2^v \rightarrow [0,1]$, and activation threshold θ_v which is chosen independently, uniformly and randomly from the interval of (0,1]. A node *v* will become *active* at time *t+1* if and only if $f_v(S) \geq \theta_v$, where *S* denotes the set of active nodes at time *t*. This model is different from the Linear Threshold model as it focuses on the *cumulative influence* of all nodes from a set *S* instead of the individual attempts of nodes $u \in S$.

2.1.3 Independent Cascade Model

Given a social network represented as a directed graph G=(V, E). Each edge $e = (u, v) \in E$ has a probability of p_e to be activated. The state of each node is either $active$ or $inactive$. When node u is activated at time t, it will attempt to activate all currently inactive neigbor v. If v is activated by u, v will become active at time $t + T_{uv}$. The time unit of T_{uv} are independent and exponentially distributed continuous random variables. Node v will attempt to activate his inactive neighbors subsequently and so forth. Therefore the activated node always has the same type as the first neighbor that succeeded in activating it. [7]

In the context of influence maximization game with b number of players, each player i selects a set S_i of at most k_i cardinal of nodes. A node selected by more than one player will be assigned to one of the players randomly and uniformly. When S_i is activated for influence cascade by i, the process will proceed as described earlier until no new activations occur. If we let T_1, ..., T_b denote the active sets at that point, then the goal of i is to maximize $E[|T_i|]$. Player

i is indifferent between strategies S_i and S_i' if the expected propagation effects are the same.

2.1.4 Decreasing Cascade Model

The decreasing cascade model is a special case of independent cascade model defined by [13]. Letting the probability of successful activation by a node u in activating v is denote as $p_v(u)$. For the Independent cascade model, this probability is a constant $p_v(u)$ independent of the history of the process. However, v's propensity for being activated may change as a function of its neighbors's attempts (and failed) in influencing it. If we let S denotes the set of v's neighbors that have attempted to influence v, then the success probability for u can be denoted as $p_v(u, S)$. In the decreasing cascade model, the function of $p_v(u, S)$ is non-increasing in S (i.e. $p_v(u, S) \geq p_v(u, T)$ whenever $S \subseteq T$. So this addition restriction limit the a contagious node's probability to activate v.

2.1.5 Utility Function Model

Most relevant to our problem context was the *Utility Function based* propagation model commonly used to evaluate consumer product selection. Janssen and Jager [10] proposed this model which incorporated cognitive behavioral theories, to study the consumer purchasing decision from a psychological perspective. The authors concluded that the behavioral processes which drive the consumers decision are mainly based upon their needs such as low prices, high social comparison, and type of cognitive processing that the consumers utilize. [6] generalized the notion to formulate an utility model to represent the consumer personal pleasure in consuming a product. The value in the utility model is derived based on the consumer's perception of a product quality, and their tendency to follow the trends within a localized community. Here the decision making process of consumers is assumed to be adversarial and responsive to network externalities, so a product with the highest utility value will be selected. A general utility model proposed by [6] is,

$$U_i = (1 - ft)(Q_i - Q_{des}) + ft(N_i/N_p). \tag{2.2}$$

U_i denotes the amount of pleasure a consumer derives from product i. The follower tendency of a consumer, ft has a value in the range of [0,1]. Q_i is the consumer's experienced quality of product i and has a range value of [0,1]. Here the consumers are assumed to have a minimum quality requirement, Q_{des} of 0.5 to be satisfied. The number of consumers who select product i and the total number of consumers are denoted as N_i and N_p respectively.

Numerous variation of propagation models were proposed to represent specific cascading effects for agricultural, citations, on-line community and social media, biological networks and human interactions. However the linear threshold, independent cascade and decreasing cascade model do not have any mechanism to account for the impact of the presence of adversaries in their cascading models. Therefore the strategies obtained based on the effects and outcomes of propagation with active adversaries will be significantly different from the conventional study. Nonetheless, the literature review presented above are relevant to our specific problem to certain extends and provided an overview of the landscape in influence propagation/ cascading research field.

2.2 Sequential Search Algorithm

Research on strategic competitions are typically grounded on a form of sequential search which deal with analytical analysis of strategic play between two or more decision makers typically known as players. Sequential search models entail the anticipation of adversary's next action whilst analyze the next counter-measure which returns the "best" possible expected payoffs. The expected payoff for a player is not only determine by his choice of play, but conjointly with adversary' choice of play. A player's decision is typically governed by an objective function which he tries to maximize (or minimize). Suppose we consider a simple competition between two players where both do not know of each other's payoff function. Then the first player, p_1 strategy choice will depend on what he perceives to be his opponent's, p_2 payoff function U_2, as the nature of the latter will be an important determinant of p_2 action in the game. If we let the expectation about U_2 to be called p_1's $first\ order$ expectation, then p_1's strategy choice will also depend on what he expects to be p_2's $first\ order$ expectation about his own (p_1's) payoff function, U_1. So this is termed as p_1's $second\ order$ expectation because it is an expectation concerning the $first\ order$ expectation. So p_1's strategy choice will depend on what he expects to be p_2's $second\ order$ expectation (what p_1 thinks that p_2 thinks that p_1 thinks about p_2's payoff function U_2. So this can be termed as p_1's $third\ order$ expectation and so on infinitely [9]. In short, it is a logic game for designing tactic to achieve best results, given the adversary's possible approach.

The players in sequential search are typically assumed to behave in a rational manner. A rational player is defined as an individual with consistent preferences if his beliefs and the information of the current state are consistent as well. Researchers use the analogy of a tree to reason decisions made by players. First the search space is defined as a tree where each of the tree nodes corresponds to a decision choice by the player's and the branches at each node denote all possible choices at that decision epoch. The tree root represents the first move from the current state and the tree leaves represent all possible final outcomes at specified ply depth. The expected payoff for each strategy is calculated based on the trace from root to each leave in the tree. As the construction of the decision tree involves the dimension of time, it is essential to consider the player sequence, the given known information, and payoff function for each leave in the decision tree.

2.2.1 Minimax Search

The crucial component in most sequential search analysis is the efficiency of *search* algorithm. A human player will try to find the "best" move to make by "looking ahead" a few moves, predicting the response of his opponent to each of his own possible moves (and the responses to these responses, and so on) and select the move that seems most promising. In short, the entire search space which comprises of all the possible moves by the player is evaluated to determine a strategy with the "best" outcome. Assuming that both players choose the strategy with the highest possibility of winning for them, the payoff of the best strategy is returned to the current search state. Player 1 p_1 tries to maximize the chance of winning the competition, whilst Player 2 p_2 tries to maximize his own chance, which is equivalent to minimizing p_1's chance. Therefore the process of backing up the value of the best move for alternating sides is called *minimaxing*, and hence the two-player sequential search algorithm is said to perform a minimax search.

Minimax search is based on a depth-first algorithm and increasing the search depth generally will improve the quality of the decision. Hence the researchers are focused on improving the efficiency of search, such that they could effectively search a higher number of ply depth within a realistic time constraint. One of such enhancement to the basic Minimax search is alpha-beta pruning heuristic to be described further in Section 2.2.2. For now we will look at a conventional minimax search using tic-tac-toe game as an example. Given the game is in current state, all the possible moves from the current state is depicted in Figure 2.1.

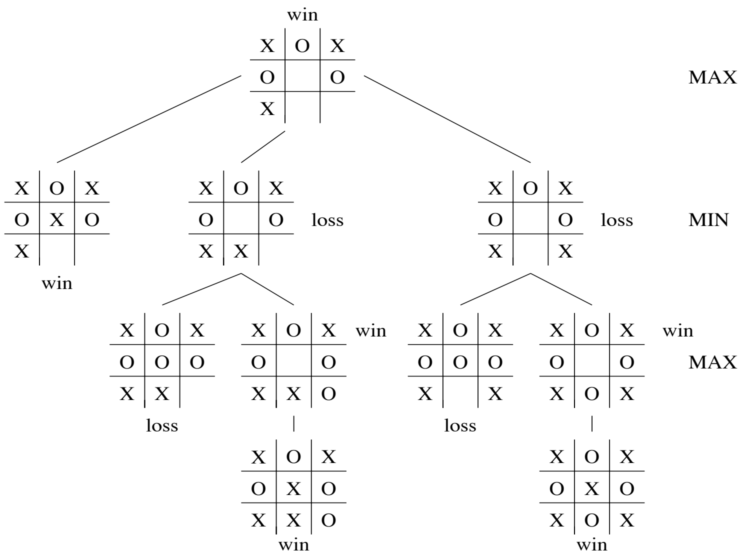

Figure 2.1: Possible Moves for Current Tic-Tac-Toe State

If we denote player X as "Max" and player O as "Min", the entire search space can be represented using a graph as shown in Figure 2.2 where Max player's moves are shown as square, and Min player's moves are shown as circle. The possible moves from each decision state are represented by a link in the graph. The root node represents the current state. In this example, the Max player has three possible moves from the current state which lead to nodes b, c and g. By accounting for the options of Min player in each of these nodes, and the responses by Max player, the entire search space can be constructed as in Figure 2.2.

Nodes a, c, e, g and i are called interior nodes. Nodes b, d, f, h and j are the leaf nodes

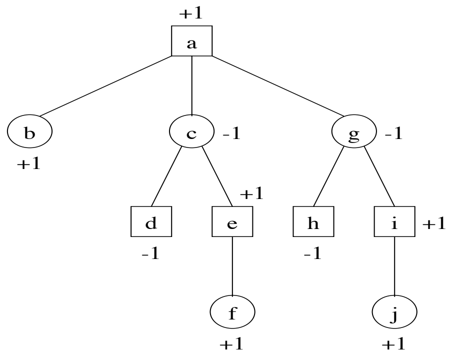

Figure 2.2: Entire State Space for Tic-Tac-Toe

in the tree. Considering all the possible strategies to find the best one is usually referred to as *searching the tree*. A move by a player is often known as a half-move or a *ply*. The leaf nodes of the tree indicate the end of decision choice, and the value of minimax function f can be determined. A win for Max player is denoted as +1 here, whereas a -1 denotes a loss. The value of 0 denotes a draw. The root of the tree is at level 0. Player Max will make a move to maximize the outcome f at all even levels, whilst player Min will do similarly to minimize the outcome f at all odd levels. Thus the f equals the maximum values of its children at all even levels, and f equals the minimum values at all odd levels likewise. The f is recursively defined at all states. By searching for the minimax value, the best strategy can be found.

The minimax search is practical for tic-tac-toe example where the tree is sufficiently small enough to traverse all the possible strategy till the evaluation ends within a reasonable amount of time. Further more the evaluation function is simple and the final outcome is either win, loss or draw. For most practical scenarios, this approach is infeasible because the entire search space is too big. For these complex scenarios (e.g. chess, checkers and othello), the evaluation function is changed to return a heuristic assessment, which allow a payoff value to be returned from any state in the analysis, instead of where the evaluation has ended. From the perspective of search algorithm, the change can be considered as minor. The evaluation function is modified by a different definition of a leaf node (terminal node). The search of minimax tree is stopped at a fixed ply depth from the root and the expected payoff at that state is evaluated. This heuristic approach is known as $Alpha$-$Beta$ pruning.

2.2.2 Minimax with Alpha-Beta Pruning

The Alpha-Beta heuristic enhances the minimax search with pruning. Pruning can yield sizable performance improvement by reducing the number of strategies need to be evaluated, and thus reduce the complexity of search to the square root. The researchers that first describe a form of pruning are [15] and the corresponding pseudo-code is depicted in Figure 2.3. It consists of the minimax function with two additional input parameters and pruning tests. The α and β parameters are known as the search bound. At max nodes, the payoff value is returned as the lower bound to children nodes as the α parameter. Whenever any of the children nodes finds that it can no longer return a payoff value above that lower bound, further search is deem to be unnecessary and therefore pruned. At the min nodes, the payoff value returned as the upper bound β parameter. The β parameter is passed on o that any max children nodes with a lower bound $\geq \beta$ can be pruned as required. Together the α and β form a search bound which can be regarded as the range for a node to return a payoff which lies within this bound. Whenever a node finds that its payoff value is proven to be outside of this bound, the search is terminated. As more nodes are expanded, the bounds become tighter till the min and max payoff value of equal to the minimax value of the root.

Our problem departs from all earlier models in the following manner: our model allows a consumer to renew their product choice at different stages based on a payoff function, instead

```
function alphabeta(n, α, β) → g;
    if n = leaf then return eval(n);
    else if n = max then
        g := −∞;
        c := firstchild(n);
        while g < β and c ≠ ⊥ do
            g := max(g, alphabeta(c, α, β
            α := max(α, g);
            c := nextbrother(c);
    else /* n is a min node */
        g := +∞;
        c := firstchild(n);
        while g > α and c ≠ ⊥ do
            g := min(g, alphabeta(c, α, β
            β := min(β, g);
            c := nextbrother(c);
```

Figure 2.3: Pseudo-code for Alpha-Beta Pruning Heuristic

of progressive manner where the consumer is unable to change their product subsequently. Both players are allow to instigate any eligible consumer at each stage in sequential manner to optimize the respective player's market share which is differ from the single-agent *influence maximization problem*. Thirdly a *contractual lock-in* policy is introduced to deter recent targeted consumer from defecting.

Chapter 3

Methodology Design and Implementation

3.1 Influence Propagation with Investment

According to studies in school of behavior sciences, consumer's decision is affected by both intrinsic factors and extrinsic factors irregardless whether it's a minor or major purchase. For simplicity purpose and without lost of generality, we let the intrinsic factor to comprise of market player's incentive which will alter a consumer's decision choice. Similarly we let the extrinsic factors to cover peers' influence. Here the consumers are assumed to respond positively towards marketing campaign by selecting product which provides higher normalized utility value synonymous to consumer satisfaction level. The proposed normalized utility value is an extension of [6] model, where we incorporated a persuasive incentive, *Inv*, which can be a form of monetary reward. The effects of influence is assumed to propagate instantaneously.

We assume that there are two players competing for their respective market shares in a social network modeled as an acyclic directed graph. The nodes in the directed graph represent individuals and links represent relationships among individuals (since the graph is directed, the impacts are directional). The two players are to take turns making decisions with infinite decision horizon. At each decision epoch, the player will be endowed a fixed player-specific budget and could spend this budget on one of the nodes. As stated earlier, unspent budget cannot be carried over.

To focus on the network effect, we assume that for each node (customer), her decision depends solely on two factors: 1) the influence from her neighbors, and 2) the direct investment from any player. Expressed formally, customer n's utility value for the product owned by player i in time period t is:

$$u_i^t(n) = \sum_{m:a_{m,n}=1} \frac{\mathrm{I}_{\{c^{t-1}(m)=i\}}}{|\{m' : a_{m',n} = 1\}|} + M_i^t(n), \tag{3.1}$$

where $c^{t-1}(m)$ represents the choice of node m in time period $t-1$, $a_{m,n}$ denotes the linkage from m to n (1 if such link exists, 0 otherwise), and this term evaluates the network influence effects (influence from all neighbors which link to him) on node n. $M_i^t(n)$ represents the investment by player i on node n in time period t. With the above utility function definition, node n's choice in time period t is then simply:

$$c^t(n) = \begin{cases} \arg\max_i u_i^t(n), & T^t(n) \geq \tau; \\ c^{t-1}(n), & \text{otherwise.} \end{cases} \tag{3.2}$$

Note that in (3.2), customer n is only allowed to change her decision if she has maintained a particular choice for more than τ time periods ($T^t(n)$ is the number of time periods customer n has maintained its current choice). This particular feature is to emulate the minimum length of contract one has to sign on when a new product is chosen. This design also helps to eliminate simple cycles among players (players keep selecting the most crucial node).

When a new investment is made, or the time period has progressed (thus changing the value of $T^t(n)$), some nodes might end up with new product choices, and these changes will create ripple effects that need to be properly accounted for. Considering the potential interactions among connected nodes, we have to propagate these updates using proper order. The procedure is described as follows:

1. Let $\mathcal{S}$ be the set containing all nodes.
2. Let $\mathcal{R} \equiv \{n | n \in \mathcal{S},\ a_{m,n} = 0, \forall m \in \mathcal{S}\}$ (in other words, $\mathcal{R}$ is the set of *root nodes* in $\mathcal{S}$). Let $\mathcal{S} \leftarrow \mathcal{S} \setminus \mathcal{R}$.
3. For $n \in \mathcal{R}$, compute $u_i^t(n)$ for $i = 1$ and 2 following (3.1). $c^t(n)$ can then be found from (3.2).
4. If $\mathcal{S} = \emptyset$, stop, otherwise, go to step 2.

The above procedure always terminates if the graph is acyclic.

3.2 Node Selection Policies

Wih the above influence propagation model, players are allowed to take turns making decisions on which node to invest in. Given the complexity of the influence propagation model described in 3.1, even with perfect information on $\{c^t(n)\}$ and $\{T^t(n)\}$, a player has to rely on pure enumeration to find the *best* node to invest in. Note that the above problem is only with one time period and without considering any adversary. Although it might be possible to formulate player's decision making problem with infinite horizon and adversary, it will be computationally intractable. As such, when we design player's strategy, we explicitly define number of future time periods to be included in the evaluation, and treat that as a strategy parameter (we call it the *look-ahead time periods*).

Although we can make single-player's strategy tractable by setting a small-enough look-ahead time periods, such limitation would create some unexpected issue in how we evaluate the importance of each node. In most influence maximization problems, the importance of a node can be characterized by the number of converts it can bring in through influence propagation. When adversaries are present and horizon is infinite, we can still estimate the importance of a node by using average or discounted measure (commonly used techniques in infinite horizon decision making problems) and having appropriate opponent model. Unfortunately, if we artificially limit the number of periods that we look ahead, these classical approach will not work anymore. To see this, assume the look-ahead period is just 1, implying that this player would be myopic. In this case, the strategy is essentially an influence maximizer that simply chooses the node that would result in maximum immediate gain; however, as one would expect, given that there is an adversary, such gain might be short-lived, and the choice might turn out to be a *short-term gain, long-term loss*.

One way to deal with such undesirable side effect of limiting planning horizon is to properly define a *value function* that would approximate a node's true value suppose we are able to reason with infinite horizon. In our initial study, we defined one such estimation function, and to distinguish it from the conventional ways of estimating a node's value, we call the strategy that relies on conventional measure the *count-based* approach, and the strategy that relies on the value function as the *value-based* approach which is define formally in the next subsection.

3.2.1 Count Based Approach

In this thesis, the conventional way of estimating a node's value is deem to be myopic and only aims to maximize the immediate market share. As an example, suppose there are two or more nodes that potentially be targeted by the player. The player can evaluate the importance of each of these nodes by counting the number of converts resulted from the influence propagation after each decision epoch by both players. The node with the highest expected number of converts will be selected for conversion. As such the approach is resemble to the greedy approach where the node with the largest marginal gain in number of adopters will be selected. This approach does not strategize any plan to fortify nodes from changing their product choice under the network influence effects. As such it is called the *count-based* approach.

Formally speaking, the count-based approach is the strategy where player i use the following function to evaluate the total value for all nodes under his control in time period t:

$$v_i^t = \sum_n \mathrm{I}_{\{c^t(n)=i\}}. \tag{3.3}$$

With (3.3), the importance value of a particular node m not owned by i is simply $v_i^t(c^t(m) \text{ set to } 1) - v_i^t$.

3.2.2 Value Based Approach

Although the *value-based* approach can take many different forms, a simple function that focuses on the quantity of nodes under the player's control as well as the strength of the control is adopted here. The individual node's value is still computed in a similar fashion as the count based case, but the strength of control is quantified by summing up the utilities for nodes that are controlled by the player i in time t:

$$v_i^t = \sum_{n:c^t(n)=i} u_i^t(n)^2. \tag{3.4}$$

The intuition behind the derivation of this function is originated from the first term in (3.1) which evaluates the network influence effects. Suppose there exist two nodes, n and m with same marginal gain in number of adopters if either one is targeted, and both n and m have more than a neighbor (surrounding node with influence link) connected to them. The node with higher proportion of neighbors that adopt the same product choice as himself will be selected because the strength of control under this case will be higher based on equation (3.4). Therefore we conjecture that the *value-based* approach will exhibit a "*fortifying*" effect on selected nodes under infinite horizon. When this occurs, we conjecture the number of adopters for player that deploys this approach to increase as well.

3.3 Implementation

To account for the adversary who is effectively competing for the market share in a zero-sum fashion, we introduce a minimax procedure to enable players to reason strategically. Given a state tuple $(\{c^t(n)\}, \{T^t(n)\})$, the current player (the maximizing player) will explore all feasible choices, and for each choice, compute the objective value v_i^t by using either Equation (3.3) or (3.4) from the state space at number of look-ahead moves. Now it's adversary's turn to make choice, the assumption is that he will make choice that minimizes the maximizing player's objective value. In general both players are maximizing their own payoff value calculated according to their objective function at their number of look-ahead moves. The number of look-ahead moves allowed will be a player-specific parameter in our model. This process will continue until we reach one of the termination criteria. A complete state space for our problem cannot be illustrated as the tic-tac-toe example shown in Figure 2.2 due to the size of the problem. An illustration of the possible Minimax search state space for our problem is presented in Figure 3.1. The square and round nodes represent the decision points of two different players, and the number within each node denotes the player's choice.

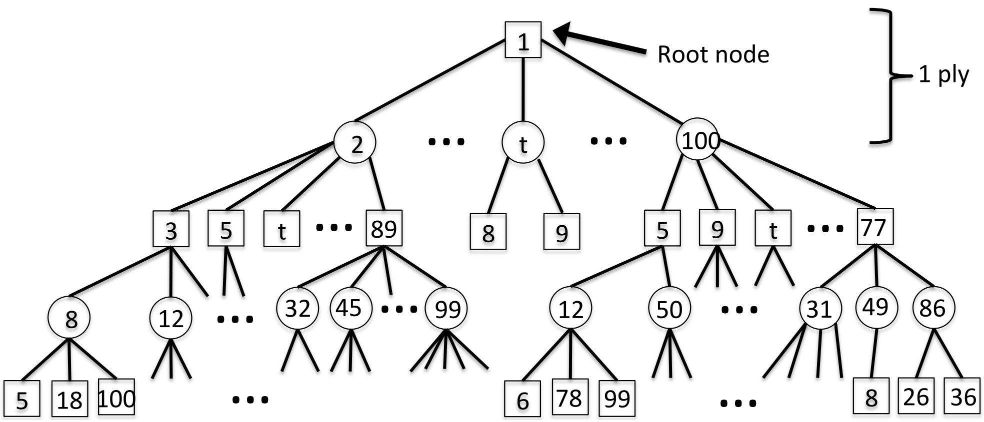

Figure 3.1: An Illustration of the Minimax Search State Space for Our Problem.

To improve the performance of the above minimax search procedure, we apply a standard α-β pruning on the search tree. The α and β values refer to the lower bound for the maximizing player and the upper bound for the minimizing player. All nodes in the search tree that have values lower than α (for the maximizing player) or higher than β (for the minimizing player) will be pruned. The pseudo-code for Minimax with Alpha -Beta pruning heuristic is presented in Section 2.2.2.

Chapter 4

Empirical Analysis

4.1 Experiment Setting

Social influence networks are commonly modeled using a class of graph structures known as the scale-free networks, which exhibit high clustering coefficient, small mean shortest path length properties and power-law degree distribution. Scale-free networks exhibit higher fraction of nodes with large (larger than average) number of in-degree edges connected to them in a network. So networks of any topology that comply to the three properties above can be classified as scale-free model (also known as power-law degree distribution networks). Kempe et. al. [11] used the *Kleinberg's Small World* network structure as the basis to study gossip protocols for spreading information in a communication network. In our experiments, we employed the *JGraphT* Java graph library which contains mathematical graph theory objects and algorithm for generating the synthetic scale-free networks. These networks contain 100 nodes which are a reasonable size for representing influential social network with strong ties. According to Adam et. el., procurement decision of a consumer is influenced mainly by her neighbors

Table 4.1: Experimental Design for Scale Free Network Analysis

Factor	Values		Level
	Incumbent	Entrant	
Node selection policy	Count based	Count , Value based based	2
Budget ratio	1	$\frac{1}{3}$, $\frac{1}{2}$	2
Initial state	Null, Random lock-in		2
Play sequence	First, Second		2
Instances	1, 2, 3, ... 17		17

with *strong ties*, compares to all others [1].

Given the size of our experimental networks, minimax algorithm is employed to evaluate the decision game tree. α-β pruning heuristic is incorporated in the minimax algorithm to reduce the size of state space by pruning decision branches that prove to be less promising. The computation effort for α-β approach is upper bounded by the brute force approach in a complete tree search for each game stage. Our adversarial search problem is emulated using a simulation model written in Java. When the size of influence network is scaled up, the dynamics of our problem will result in an exponential growth in number of state space. However when such situation occurs, the sampling-based approaches can be used to improve the computational efficiency by sacrificing the comprehensiveness of search.

The following design is used to address our conjecture in this research. A total of 272 experimental runs are evaluated using 17 sets of distinct instances of scale free network structure to determine if the proposed *value-based* node selection policy is able to, 1) identify and 2) secure a larger pool of customers via the niche-seeking behavior, as oppose to the conventional count-based (greedy) approach against an adversary who employed the count-based approach.

The nature of our research revolves around two players game where each tries to minimize their maximum lost in the number of adopters. The consumers in the influence networks are initialized under two conditions for the experiments. When the market players enter a new market space untainted by competition, there is ample amount of opportunity for growth given that all the consumers are yet to be explored. This is synonymous to a market of consumers without any obligation (*contractual lock-in*) and every consumer can be targeted by the players. Here we denote this initial condition as *null* state. On the contrary, if the market boundaries are well defined with players' competitive rules. The consumers are normally committed to some form of existing obligations (with various remaining time length) and thus are unavailable to be targeted the market players. This correspond to the metaphor of red ocean used commonly to describe a market condition (denoted as *random* state).

For simplicity purpose, the budget for stronger player is set to 1, and the weaker player is assigned either $\frac{1}{3}$ or $\frac{1}{2}$ in this thesis. Since the customer's utility function is determined by the ratio of her incoming neighbors owning that product, a budget of 1 implies the stronger player can convert *any* customer, while the weaker player can only convert a much smaller fraction of customers. We use Figure 4.1 as an example to illustrate the concept. Here the minimum budget required for the player (*black*) to instigate consumer A is $\frac{1}{3}$.

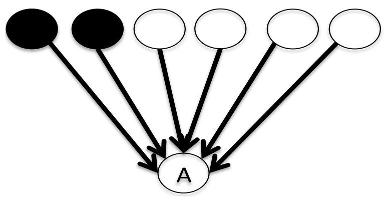

Figure 4.1: Definition of Normalized Marketing Budget

In the following set of experiment, we assumed the *incumbent* to always adopt the myopic *count based* approach and the *entrant* will adopt both node selection policies for comparison purpose. To remove the first-mover's advantage (any advantage from knowing the other player's move before making her own choice), each player will get to play first for each network instance.

4.2 Results and Discussions

The effectiveness of our proposed *value based* node selection approach is investigated using the synthetic influence networks generated as described earlier. The experimental results presented in the following are based on the performance of *entrant* player. The performance of the *incumbent* is not presented because the node selection approach adopted by the *incumbent* player remain the same across all experimental runs.

Given the budget limitation constraint on the *entrant* player, we foresee her market share to be lower than the adversary. However we conjectured that the *entrant* player could garner a larger market share simply by changing her node selection policy to consider long-term node values. Using the synthetic consumer influence networks data, we attested the effectiveness of the proposed *value-based* node selection policy. Without going into details, we would like to illustrate visually the niche-seeking behavior by our value-based strategy. As illustrated in Figure 4.2, a sub-group is quickly identified and captured by the entrant player, and such behavior is consistently observed in the steady state and also for other network instances. For the rest of the section, we will define the steady state and quantitative measures that allow us to quantify the niche-seeking behaviors.

4.2.1 Identifying Set of Steady States

Recall that our model is with infinite horizon, thus the adoption status might change radically from epoch to epoch. Therefore, unless this dynamic adoption process comes to a complete stop

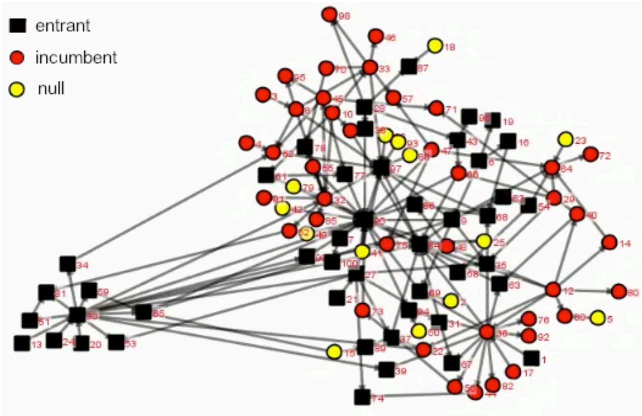

Figure 4.2: The Effects of Value-Based Approach.

(e.g., one player completely dominates the market), it would be unfair to take any snapshot and conclude the performances of player's strategies.

If complete stop is not observe in a particular experiment instance, what we can do instead is to identify the *repeating states*. According to Section 3.1, the state at time t can be described by two vectors: $\{c^t(n)\}$ and $\{T^t(n)\}$. For the purpose of performance evaluation, we only need $\{c^t(n)\}$. Based on the above description, at least one state will eventually appear for the second time, since the set of feasible values of $c^t(n)$ is finite and t is unbounded. Suppose the state in time t is denoted as S_t and $S_t = S'$, if S' is observed again in time $t+\delta$, the experiment can be terminated early and the set $\{S_t, S_{t+1}, \ldots, S_{t+\delta-1}\}$ will contain all the repeating states of this experiment instance. We can make this claim because all strategies we proposed are deterministic and not dependent on $\{T^t(n)\}$.

This is an important observation, as we can now use the set of repeating states to quantify player's performances. When we report the experiment results, we focus on two measures: the average and the niche performance metric over the set of repeating states. The analysis of experiment results is presented in the next subsections.

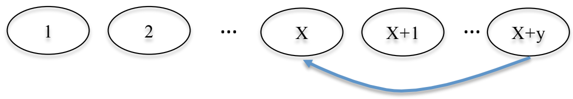

Figure 4.3: Repeating States

4.2.2 Average and Volatility in Number of Adopters

The study of diffusion models typically refer to a successful propagation of an idea or innovation to an individual node within a social network G as *active*, whereas an unsuccessful propagation as *inactive*. Similarly here but with a slight change, the node which adopts either the *entrant* or *incumbent* product choice is denoted as *adopter*. Whereas the node which adopts neither players' product choice is denoted as *null*. The number of *adopter* for each player will be determined at every game stage in each experimental run. For example, the expected number of *adopter* for *entrant* player is calculated by taking the mean value of *entrant* player's *adopter* throughout the steady state for all experimental runs. The effects of play sequence is blocked by computing the mean value of number of *adopter* from both play sequences by the same player. In this way, any effect from the first mover advantage can be blocked and thus isolated.

Besides examining averages, we also compute the volatility of market shares from state to state (in the same repeating set); such measure allows us to judge the stability and robustness of a particular player's strategy as well. The computation of volatility is inspired by and borrowed from the financial literature. To illustrate how it's computed, let $\{m_1, m_2, \ldots, m_\delta\}$ be the sequence of market shares for a particular player over the set of steady states. We can first compute the log market share relatives as: $r_i = \ln(m_{i+1}/m_i)$. The volatility can then be computed by computing the standard deviation of all r_i's. The above two performance measures under different experiment settings (combination of *null* and *random* initialization and budget ratio of $\frac{1}{3}$ and $\frac{1}{2}$ with blocked play sequence) are presented in Table 4.2.

Table 4.2: Summary of experiment results when opponent adopts CB approach (CB and VB stand for count- and value-based respectively).

			# of Adopters (std. dev.) CB	Volatity (Volatity against CB)
null	$\frac{1}{3}$	CB	45.2 (0.015)	0.0625
		VB	67.8 (0.031)	0.0335
		Sig. diff. (p-value)	p<0.001	p<0.001
	$\frac{1}{2}$	CB	46.3 (0.018)	0.0761
		VB	69.1 (0.027)	0.0337
		Sig. diff. (p-value)	p<0.001	p <0.001
random	$\frac{1}{3}$	CB	30.8 (0.027)	0.0170
		VB	40.9 (0.019)	0.0232
		Sig. diff. (p-value)	p<0.001	p<0.001
	$\frac{1}{2}$	CB	32.0 (0.021)	0.0130
		VB	44.4 (0.030)	0.0191
		Sig. diff. (p-value)	p<0.001	p<0.001

A two sample *t*-test is applied to evaluate the statistical difference between the two node selection policies. The null hypothesis, H_0 denotes the means (in number of adopters) of both policies is equal; whereas we let the alternate hypothesis, H_1 to denote the means of *value-based* is greater than the count-based. For convenient purposes, we let the significance level (also known as the alpha level), α = 0.05 arbitrary. The results indicate that it is always better for the entrant player to adopt the *value-based* node selection policy. By simply adopting the *value-based* strategy (which considers a node's long-term value), the entrant player can perform considerably better than the myopic count-based strategy by approximately 50% in average number of adopters for all cases (e.g. initial market conditions and budget levels). Both node selection policies are also observe to be insensitive to budget as both strategies record almost similar number of adopters in the experiments. The average number of adopters increases merely by one when the budget is increase from $\frac{1}{3}$ to $\frac{1}{2}$ of the adversary's.

Figure 4.4 depicts the performance in average number of adopters under different initial condition for both strategies when the entrant's budget is $\frac{1}{3}$ of the incumbent's. The market share of the entrant player gains approximately 10% (for initial *null* condition) and approximately 20% (for initial *random* lock-in condition) respectively, when the node selection policy is changed from count based strategy to value based strategy. Similar comparison at a higher budget level of $\frac{1}{2}$ is illustrate in Figure 4.5. The average number of adopters for entrant player increases approximately 12% and 25% for initial *null* and *random* lock-in condition respectively, when the node selection policy is changed from count based to value based strategy. These observations are intuitive because the dispersion of influence within a network of *null* state is less restrictive (where all the nodes are available to be targeted), and thus allows the influence effects to be cascaded to a larger number of nodes. Therefore the number of adopters is expected to be higher than *random* initial lock-in condition.

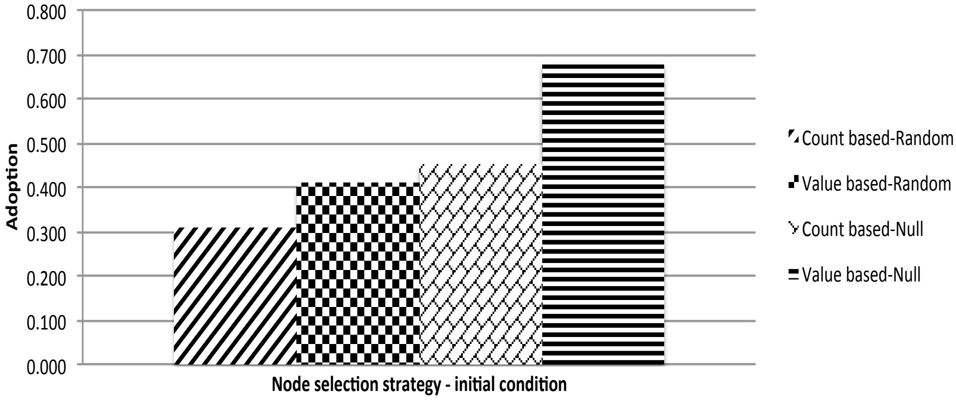

Figure 4.4: Percentage of Adopter For Entrant Product Choice At A Third of Incumbent's Budget Levels With Both Initial Game States

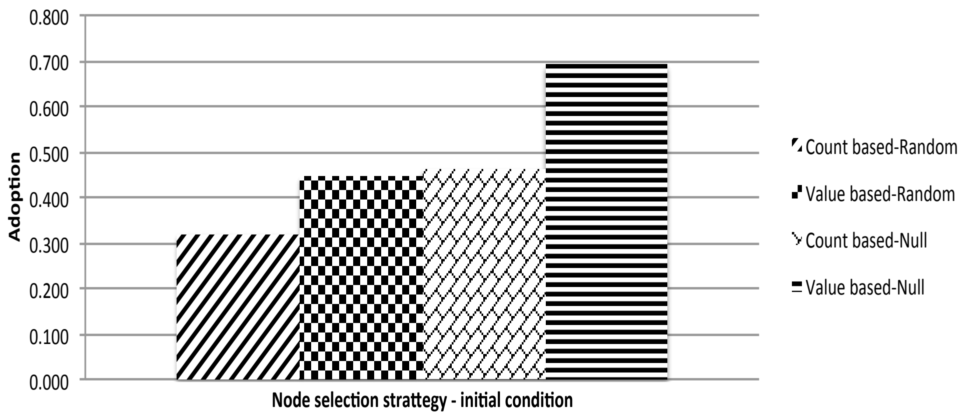

Figure 4.5: Percentage of Adopter For Entrant Product Choice At Half of Incumbent's Budget Levels With Both Initial Game States

Although the budget level will affect the size of potential target customers, the *value-based* strategy is able to estimate the true long term value of each customer more accurately and subsequently translate this finding into an observable result in number of adopters. Therefore the node selection policy is more important factor than budget level for improving the market share.

In summary, the experimental results strongly support our conjecture that to succeed in a competitive influence maximization game with infinite horizon, it's very important to evaluate a node's long-term value correctly. Although our value function is extremely simple, it's still significantly better than the conventional approach that just myopically maximize immediate market gains. In terms of volatility, we can see that against a count-based adversary in a null initialization, the entrant will enter a much stable steady state if he chooses the value-based approach. However, such difference diminishes when the initialization becomes random.

4.2.3 Niche Performance Metric

In the previous subsection, we show that value-based strategy performs better than the count-based approach under all circumstances for the entrant player. To find out whether such value-based strategy would result in players building niche customer base, we define a *niche performance metric* to measure the degree of niche seeking. This metric calculates the proportion of edges where both the source and target nodes adopt the same product type over total number of edges. When the source and target nodes are adopting the same product type, it implies an influence propagation effect from the source to target node. On the contrarily, when the source and target node's product type is different from each other, it symbolizes a non-continuity in influence spread.

The comparison on *niche performance metric* for both initializations is summarized in Table 4.3. It is observed that regardless of setups, the value-based approach always ends up with higher niche performance metric in steady states. The These experiment results confirm our second conjecture: it's indeed more advantageous for weaker player to concentrate on smaller niche in the scale-free networks. However the increments in *niche performance metric* when the budget is increased are statistically insignificant.

4.3 Sensitivity Analysis

Suppose that we have just completed an analysis which will have a major impact on the company's marketing strategy and about to present such results to the management committees. It is important to know if our analysis will remain valid when the basic data and assumptions are slightly inaccurate. The questions normally one could ask are, what impact will those differ-

Table 4.3: Summary of Niche Performance Metric When Opponent Adopts CB Approach (CB and VB Stand for Count- and Value-Based Respectively).

			Average value (std. dev.)
null	$\frac{1}{3}$	CB	0.425 (0.019)
		VB	0.711 (0.033)
		Sig. diff. (p-value)	p<0.001
	$\frac{1}{2}$	CB	0.443 (0.024)
		VB	0.718 (0.038)
		Sig. diff. (p-value)	p<0.001
random	$\frac{1}{3}$	CB	0.247 (0.038)
		VB	0.365 (0.054)
		Sig. diff. (p-value)	p<0.001
	$\frac{1}{2}$	CB	0.262 (0.023)
		VB	0.388 (0.046)
		Sig. diff. (p-value)	p<0.001

ences have on our conclusions? Will the findings be completely different, or just only minor impact? Thus it is crucial to conduct some level of sensitivity analysis to examine the impact of the assumptions or parameter values used in the experiment.

In the following sensitivity analysis, our interest is to address the question of whether the *value-based* node selection policy is still a better strategy for the entrant player, if her opponent adopts the *value-based* approach. How the two primary performance measures (average number of adopters and niche performance metric) will response when the assumption on incumbent's strategy is relaxed. To answer these questions, the same series of experiments as in Table 4.1 are performed with the incumbent adopts the *value-based*.

4.3.1 Average Number of Active Nodes

Table 4.4 highlights the impact of two different node selection policies adopt by the incumbent player (by table row) and whether the earlier observations still valid (by table column). The results indicate a reduction in average number of adopters across all experimental settings when the incumbent's strategy is changed to *value-based*. This observation is intuitive as the incumbent player is leveraging on the advantages of higher budget level and *value-based* strategy to capture the market share. Thus the entrant player has no additional competitive advantage over her opponent when both adopt the *value-based* node selection policy. The effect of increment in budget level is also statistically insignificance regardless of incumbent's node selection policy. Nevertheless the entrant player will still perform better with the *value-based* strategy always (as indicate by the p-value obtain from the t-test).

Table 4.4: Summary of Comparison Results of Both Node Selection Policies by Incumbent Player (CB and VB Stand for Count- and Value-Based Respectively).

			# of Adopters (std. dev.)	
			Adversary Strategy	
			CB	VB
null	$\frac{1}{3}$	CB	45.2 (0.015)	24.3 (0.036)
		VB	67.8 (0.031)	45.1 (0.023)
		Sig. diff. (p-value)	p<0.001	p<0.001
	$\frac{1}{2}$	CB	46.3 (0.018)	25.5 (0.039)
		VB	69.1 (0.027)	45.7 (0.021)
		Sig. diff. (p-value)	p<0.001	p<0.001
random	$\frac{1}{3}$	CB	30.8 (0.027)	17.8 (0.028)
		VB	40.9 (0.019)	28.0 (0.031)
		Sig. diff. (p-value)	p<0.001	p<0.001
	$\frac{1}{2}$	CB	32.0 (0.021)	23.1 (0.023)
		VB	44.4 (0.030)	30.8 (0.037)
		Sig. diff. (p-value)	p<0.001	p<0.001

Figure 4.6 illustrates the effects of both node selection policies at specific budget level across the two initializations. The benefit of *value-based* strategy is more prominent when the initial market condition is in *null* state as depict by a larger increase in number of adopters. The random *contractual lock-in* initial state allows the formation of sub-structures/ sub-groups within the influence network. This in turn constricts the dispersion of influence and subsequently the effects of *value-based* strategy. Similar outcomes are observed (Figure 4.7) when the entrant player budget is increased to $\frac{1}{2}$ of the incumbent's

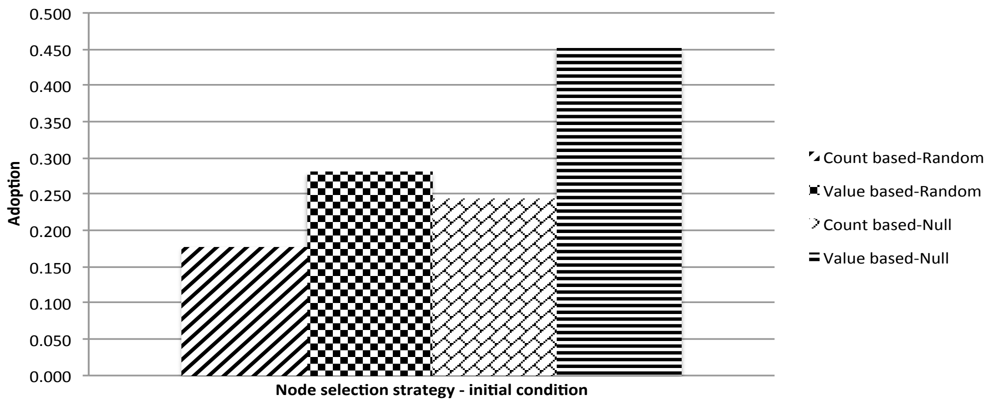

Figure 4.6: Percentage of Adopter For Entrant Product Choice At $\frac{1}{3}$ of Incumbent's Budget Levels With Both Initial Game States

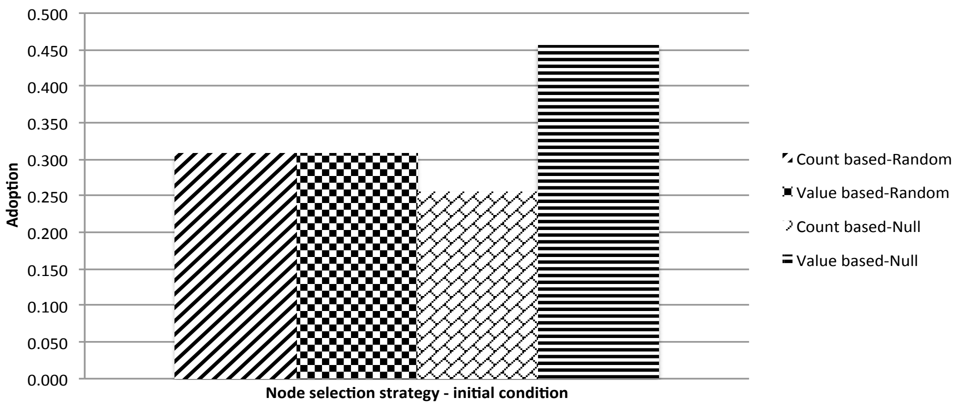

Figure 4.7: Percentage of Adopter For Entrant Product Choice At $\frac{1}{2}$ of Incumbent's Budget Levels With Both Initial Game States

4.3.2 Niche Performance Metric

The results in term of *niche performance metric* indicate likewise that the proposed *value-based* strategy is able to construct a stronger niche customer base than the count-based strategy under *null* initial state (Figure 4.8), as well as under *random* initial state (Figure 4.9). Now It is possible to conclude that the *value-based* strategy is more $niche - seeking$ than then myopic and greedy count-based strategy.

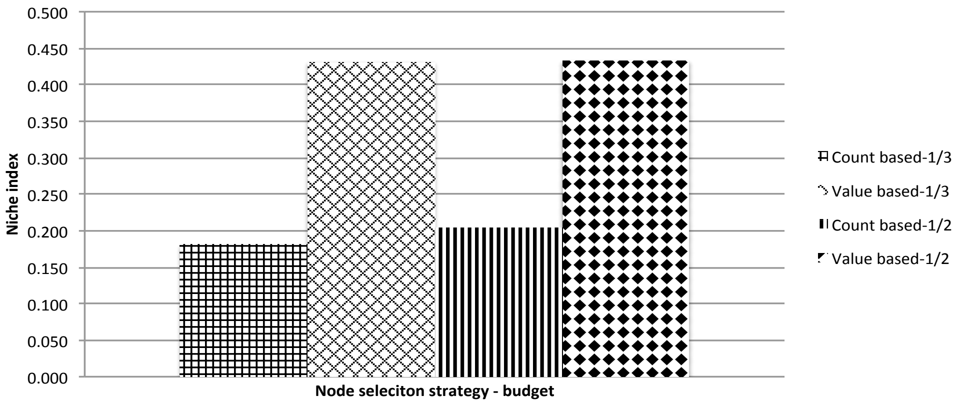

Figure 4.8: Niche Performance of Count Based Versus Value Based Approach under Null Initial Constraint at Two Budget Levels Given Incumbent With Value-based Approach

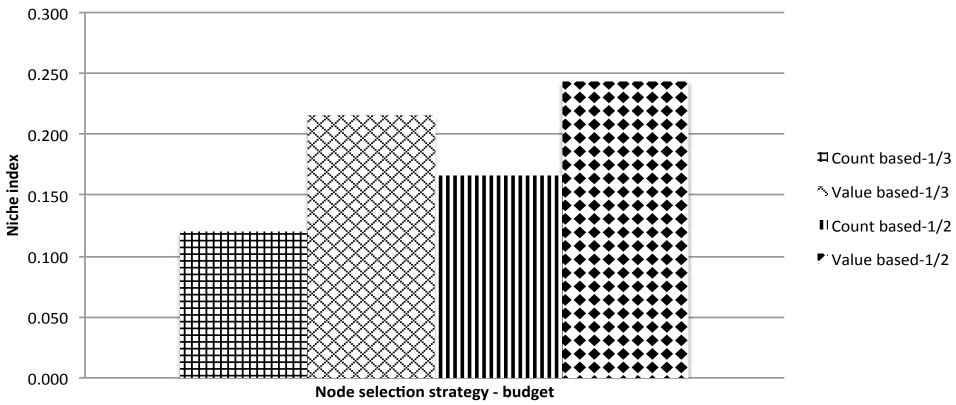

Figure 4.9: Niche Performance of Count Based Versus Value Based Approach under Random Initial Lock-In Constraint at Two Budget Levels Given Incumbent With Value-based Approach

Chapter 5

Conclusion

Product marketing has been very competitive in this era where information (user comments or friends recommendation) propagated almost instantaneously through the interconnected consumers network. Given that consumers purchasing behavior are increasingly influence by their social networks to maximize the consumers' value in product consumption. We constructed a simple two-player, infinite horizon extension of the influence maximization model to illustrate the importance of considering both adversary and time in this thesis. Although influence maximization is a well-studied problem in the literature, it resurfaces to capture researchers' attention recently due to the fact that societal-scale social networks are increasingly ubiquitous. However, the classical research on influence maximization lacks either the explicit modeling of adversaries or the consideration of time.

In this thesis, the count-based approach which represents the conventional influence maximization approach is shown to perform much worse than the value-based approach which approximates the long-term values of individual nodes. Furthermore, we show the value-based approach to achieve a better performance via niche-seeking, which is measured using our niche performance metric. Our model also provides a simple and clean framework for future study on extending well-known results on influence maximization to real-world marketing problems.

5.1 Summary of contributions

We summarize our contributions in this thesis by grouping them according to the main tasks (findings) toward this thesis. This thesis adheres to the following three steps, 1) we start with the conjecture that node selection policy plays a critical part in improving the rate of adoption within a duopoly market besides the marketing budget, 2) The prior arts in this research area led us to design a framework to study an extension of the classical influence maximization problem, by introducing a strong adversary, and 3) Given the intuition observed from the preliminary results, we developed the value based node selection policy which exhibits a niche seeking behavior in this thesis.

1. Observations:

   · We discovered that the budget level is of less significant in improving the number of adopters in the specific network topology that we experimented.

   · We proved that the proposed value based approach outperforms the conventional ”greedy” count based approach because the former exhibits a niche-seeking behavior.

2. Models:

   · An utility based influence propagation model (which extends the classical utility function based model by [6]) is proposed to represent the decision making process of consumers within an influence social network.

   · A game theoretic framework to study a duopoly competition with imbalance budget levels under network influence effects. This framework allows other form of network topology to be evaluate under similar experimental settings to determine if the niche-seeking value approach is applicable to garner a larger amount of adopters.

3. Algorithms:

   · We designed the value based approach, an algorithm for targeting specific nodes via niche-seeking within an influence network that allows the neighboring nodes to reinforce the adoption decision of specific nodes.

5.2 Limitations of Thesis

Although this thesis has reached its objectives, there were some limitations on the current framework and models. Firstly, this research was conducted using directed acyclic influence network of 100 nodes. The current network size may need to be expanded to generalize our current conclusions to cover wider range of social networks. Secondly, the decision of each node (customer) is currently depends solely on the influence from her neighbors and direct investment from the players. Other logical factors such as user requirement, product/ service quality and cost can be included in the model to reflect a more realistic decision making process of consumers. Thirdly, the number future time periods to be included in the adversarial model (termed as the *look-ahead time periods* is currently treated as a strategy parameter. Another viable approach is to define a maximum allowable computation time as player-specific strategy parameter, so the maximum *look-ahead time periods* can be applied by the player at each decision epoch.

5.3 Future Work

Our long term research direction is to design effective node selection policies for market player, which allow him to maximize the number of adopters under the presence of stronger adversaries in large scale networks. Our goal is to develop the node selection policies which will outperform conventional greedy based approaches, independent of the underlying utility function which govern the consumers' decision and the network structures which link the consumer in the community. The rationale of this lies on the process of decision making for consumer is too complex to model and involve other fields of work in psychology, sociology and human behavior which is highly unpredictable. The influence (propagation) networks for different communities could be very distinctive and thus a general node selection policy which can perform relatively well irregardless of network structure is highly desirable.

On the way to achieve this long term research goal, we envisage the work will comprise of,

1. Analyzing the theoretical models of network structure and its evolution process,
2. Developing statistical influence propagation models and algorithms to efficiently evaluate the decision game tree, for pruning less promising game tree branch
3. Extend the model to encapsulate the stochasticity of human behavior in decision making process by incorporating probability to influence propagation models

We believe that this goal is achievable given that game theoretical models can be applied to predict the general decision strategies of players under fictitious play, and this allows various other forms of network structures to be evaluated extensively. An enhancement for the current value based node selection policy can be designed after gaining greater understanding of the characteristics inherited within each specific classes of network structure through theoretical models. In this thesis, we have undertake the first step towards achieving this goal by constructed the fundamental building blocks for studying the competition between market players with different capability under the network influence effects.

Bibliography

[1] Paul. Adams. How your customers' social circles influence what they buy, what they do and where they go. Presented Web 2.0 Expo 2011, San Francisco, California, USA, 2011.

[2] Frank M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[3] Wei Chen, Yajun Wang, and Siyu Yang. Efficient influence maximization in social networks. In John F. Elder IV, Françoise Fogelman-Soulié, Peter A. Flach, and Mohammed Javeed Zaki, editors, *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28 - July 1, 2009*, pages 199–208. Association for Computing Machinery, 2009.

[4] Pedro Domingos and Matthew Richardson. Mining the network value of customers. In *Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, Aug 26 - Aug 29, 2001*, pages 57–66. Association for Computing Machinery, 2001.

[5] Nicholas Economides. The economics of networks. *International Journal of Industrial Organization*, 14(6):673–699, 1996.

[6] Michael Garlick and Maria Chli. An agent-based simulation of lock-in dynamics in a duopoly. In *the 9th International Conference on Autonomous Agents and Multiagent Systems, Toronto, Canada, May 10 - May 14, 2010*, pages 1545–1546, 2010.

[7] Jacob Goldenberg, Barak Libai, and Eitan Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 2001.

[8] Mark Granovetter. Threshold model of collective behavoir. *The American Journal of Sociology*, 83(6), 1978.

[9] J. Harsanyi. Games with incomplete information played by 'Bayesian' players, parts I-III. *Management Science*, 14:159–182, 320–334, 486–502, 1968.

[10] Marco Janssen and Wander Jager. An integrated approach to simulating behavioral processes: A case study of the lock-in of consumption patterns. *Journal of Artificial Societies and Social Simulation*, 2(2), 1999.

[11] Kempe, Kleinberg, and Demers. Spatial gossip and resource location protocols. *JACM: Journal of the ACM*, 51, 2004.

[12] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.

[13] David Kempe, Jon Kleinberg, and Eva Tardos. Influential nodes in a diffusion model for social networks. *Computer*, 3580:1127–1138, 2005.

[14] Jure Leskovec, Andreas Krause, Carlos Guestrin, Christos Faloutsos, Jeanne M. VanBriesen, and Natalie S. Glance. Cost-effective outbreak detection in networks. In Pavel Berkhin, Rich Caruana, and Xindong Wu, editors, *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Jose, California, USA, August 12 - Aug 15, 2007*, pages 420–429. Association for Computing Machinery, 2007.

[15] A. Newell, C. Shaw, and H. Simon. Chess playing programs and the problem of complexity. *IBM Journal of Research and Development*, 2:320–335, 1958.

[16] John A. Norton and Frank M. Bass. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science*, 33(9):1069–1086, 1987.

[17] Matthew Richardson and Pedro Domingos. Mining knowledge-sharing sites for viral marketing. In *The Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, July 23 - July 26, 2002*, pages 61–70. Association for Computing Machinery, 2002.

[18] Bruce Robinson and Chet Lakhani. Dynamic price models for new-product planning. *Management Science*, 21(10):1113–1122, 1975.