Maximizing Influence in a Competitive Social Network: A Follower's Perspective

[Extended Abstract]

Tim Carnes[*], Chandrashekhar Nagarajan[†], Stefan M. Wild[‡], and Anke van Zuylen[†]
School of Operations Research and Information Engineering
Cornell University
Ithaca, NY 14853
{tcarnes, chandra, stefan, anke}@orie.cornell.edu

ABSTRACT

We consider the problem faced by a company that wants to use viral marketing to introduce a new product into a market where a competing product is already being introduced. We assume that consumers will use only one of the two products and will influence their friends in their decision of which product to use. We propose two models for the spread of influence of competing technologies through a social network and consider the influence maximization problem from the follower's perspective. In particular we assume the follower has a fixed budget available that can be used to target a subset of consumers and show that, although it is NP-hard to select the most influential subset to target, it is possible to give an efficient algorithm that is within 63% of optimal. Our computational experiments show that by using knowledge of the social network and the set of consumers targeted by the competitor, the follower may in fact capture a majority of the market by targeting a relatively small set of the right consumers.

Categories and Subject Descriptors

F.2 [**Analysis of Algorithms & Problem Complexity**]: Nonnumerical Algorithms and Problems; G.2.1 [**Discrete Mathematics**]: Combinatorics—*combinatorial algorithms*

General Terms

Algorithms, Performance, Theory

Keywords

Approximation Algorithms, Social Networks, Viral Marketing, Network Analysis, Targeted Marketing

[*]Research supported by NSF grants CCR-0635121 & DMI-0500263.

[†]Research supported by NSF grant CCF-0514628.

[‡]Research supported by a DOE Computational Science Graduate Fellowship under grant number DE-FG02-97ER25308 and by NSF grant CCF-0305583.

ICEC'07, August 19–22, 2007, Minneapolis, Minnesota, USA.

1. INTRODUCTION

The spread of a new idea or product is often studied by modeling a social network as a graph where the nodes represent individuals, and edges represent interactions between individuals. These interactions could include the recommendation of a particular product and such recommendation networks and their effects on consumer purchasing have recently been analyzed in [15] and [16]. Further, there has been recent statistical support that such network linkage can directly affect product adoption [9]. Based on these empirical studies, we can formulate assumptions on how people affect the people they interact with. We can then use these graphs to answer questions such as: "If customers influence each other in their decisions to buy products, which customers should be targeted to maximize the expected profit of a new product?" and "How large of a consumer base needs to be targeted for a new technology, product, or idea to capture a significant share of the market?"

Motivated by the declining effectiveness of traditional mass marketing techniques [15], many recent papers have studied these and similar types of questions. The algorithmic problem of designing viral marketing strategies, marketing techniques which exploit pre-existing social networks to reach consumers, was studied by Richardson and Domingos [21], and Kempe, Kleinberg and Tardos [12, 13]. Their research builds on a "word-of-mouth" approach examined in a marketing context by Goldenberg et al. in [8]. In the aforementioned works, the producer of a new product is assumed to have the ability to "influence" a particular set of consumers within the social network – either through targeted advertising, providing free samples, or adding monetary incentive – to adopt the new product. If these people influence some of their friends to also try the product, and these friends in turn recommend it to others, and so forth, the producer can create a cascade of recommendations. The question then becomes how to choose an initial subset of so-called *early adopters* to maximize the number of people that will eventually be reached, and hence be likely to purchase the product. The size of the subsets allowed is assumed to be limited due to marketing budget constraints. Kempe et al. develop general models for the spreading of influence, show that finding the most influential set of nodes is NP-hard, and give an approximation algorithm for finding a set of nodes that approximately maximizes the expected influence.

The models developed by Kempe et al. assume that there

is only one company introducing a product. However, producers of consumer technologies often must introduce a new product into a market where a competitor will offer a comparable product. The introduction of Nintendo's Wii, to compete with Sony's Playstation 3, and Blu-ray discs, competing with Toshiba's HD DVD, are recent canonical examples of such behavior. When adoption of the technology is not free, it is unlikely that a typical consumer will use both products. Furthermore, even if a competing product is superior, consumers are often reluctant to switch technologies if they must bear a cost of transition which may outweigh any direct benefits of the technology [6]. The question whether in this setting a competing product can survive and will be adopted by a significant fraction of the market, or if it will eventually disappear, has been studied in numerous works, including [10], [17], and [23].

It is not always the case that the product with the largest number of early adopters can translate this initial edge into market dominance. A classical example where such *tipping* did occur is the demise of the BETA format due to the VHS format's initial popularity. However, Katz and Shapiro note that consumer heterogeneity coupled with distinct features of rival products tends to limit tipping in markets where consumers care more about a product's features than its overall prevalence [11]. Hence it is an interesting question to consider how a company with a smaller marketing budget may effectively infiltrate a market in which a stronger competing company is also present.

Historically, competition between two products has largely been addressed from an economic modeling perspective and focused on areas such as market equilibrium. For example, in [2] and [3], primarily network-independent properties are employed to model the propagation of two technologies through a market. Tomochi et al. [23] offer a more game-theoretic approach which relies on the network for spatial coordination games. However, they do not address the problem of taking advantage of the social network and viral marketing when introducing a new technology into a market.

In this paper, we study the algorithmic problem of how to introduce a new product into an environment where a competing product is also being introduced. We focus on the case when a company can keep itself hidden from a competitor until the moment of introduction. We assume that the company has a fixed budget for targeting consumers and knows who its competitor's early adopters are – either through extensive market research or industrial espionage. We first develop two models for the spread of adoption of the two products through the network. We show that finding the most influential set of a given size for the company to target – the set that maximizes the expected number of people that will adopt the new product – is NP-hard under the proposed models in this setting. From a game theoretic point-of-view, this can also be viewed as calculating the company's *best response* to a competitor's move in a Stackelberg game [7].

Following Kempe et al. we show that using well known results on submodular functions [18], we can give a $(1-\frac{1}{e}-\varepsilon)$-approximation algorithm for finding the most influential set of nodes. Additionally, using a result of Sviridenko [22], we generalize the allowed subsets to be limited based upon cost rather than simply size, hence allowing different costs to be associated with targeting different subsets of customers. We will empirically show that a company can obtain a larger market share than its unsuspecting competitor even if the competitor has a much larger marketing budget. Further, we show that knowing who the competitor's early adopters are, hence being able to apply our algorithm, will allow the company to capture a given percentage of the market using a much smaller marketing budget.

In the sequel we use the words technology and product synonymously. We discuss useful results from related work in Section 2. Building on these results, we describe the models we developed for the spread of two competing technologies in Section 3 and the results derived for these models in Section 4. In Section 5 we give the results of some numerical simulations of the behavior of these models, and we present conclusions and further research directions in Section 6.

2. BACKGROUND

We begin by recalling some existing results central to the present work.

Submodular function maximization: Given a ground set V, a function $f : 2^V \to \mathbb{R}$ is said to be *submodular* if $f(S \cup \{v\}) - f(S) \geq f(T \cup \{v\}) - f(T)$ for all $v \in V$ and sets $S \subseteq T \subseteq V$. We further say that f is *monotone* if $f(S \cup \{v\}) \geq f(S)$ for all $v \in V$ and subsets $S \subseteq V$.

For a non-negative, submodular, and monotone function f, and the optimization problem

$$\max \{f(I) : |I| = k, I \subseteq V\}, \tag{1}$$

the greedy *Hill Climbing Algorithm* repeatedly adds the element from V that gives the greatest improvement, by solving

$$\max \{f(I \cup \{v\}) : v \in V - I\} \tag{2}$$

until $|I| = k$. In [18] Nemhauser et al. show that hill climbing yields a $\left(1-\frac{1}{e}\right)$-approximation: if I is the set found by the Hill Climbing Algorithm, and I^* maximizes (1), then $f(I) \geq \left(1-\frac{1}{e}\right) f(I^*)$. This result has been extended [12] to show that for any $\varepsilon > 0$, there is a $\gamma > 0$ such that when using a $(1+\gamma)$-approximation of $f(\cdot)$ in (2), we obtain a $(1-\frac{1}{e}-\varepsilon)$-approximation.

Sviridenko recently generalized the result from Nemhauser et al. to include problems of the form (1) with an additional knapsack-type constraint [22]. In particular, for a set of non-negative weights $\{c_i : i \in V\}$ and a budget $B \geq 0$, we now consider the problem:

$$\max \left\{ f(I) : \sum_{i \in I} c_i \leq B, I \subseteq V \right\}, \tag{3}$$

where f is again a non-negative, submodular, and monotone set function. An extension of hill climbing, iteratively adding to I elements $v \in V - I$ which maximize

$$\max \left\{ \frac{f(I \cup \{v\}) - f(I)}{c_v} : c_v + \sum_{i \in I} c_i \leq B \right\}, \tag{4}$$

until $c_v > B - \sum_{i \in I} c_i$ for all $v \in V - I$, is described in [14]. Sviridenko showed that this version of hill climbing yields a $\left(1-\frac{1}{e}\right)$-approximation to (3).

Influence maximization on a network in the single technology case: The spread of a single technology through a network has been approached using different diffusion models (see for example [12, 13, 21, 24]). Here we describe the *independent cascade model* introduced by Kempe et al. [12] which resembles the models we will develop in the case of competing technologies.

We assume some set of nodes I initially uses the technology. The diffusion process then unfolds in discrete time steps. When a node u first adopts the technology, it gets a single chance to make its neighbor v adopt the technology. It succeeds with probability p_{uv} independently of the history so far. In the next time step, the nodes which just adopted the technology get a chance to influence their neighbors and so on. Note that the process is *progressive*: once a node has adopted the technology, it will not go back to the state of not having adopted it.

The quantity of interest is then the *influence function* $\sigma(I)$, signifying the expected number of nodes that eventually adopt the technology given the initial set of adopters I. In [12] Kempe et al. address how to choose an initial set I of some fixed size k to maximize $\sigma(I)$. Kempe et al. previously showed that solving (1) when f is the influence function σ is NP-hard but that σ is submodular. Hence if σ can be approximated (say with numerical simulations) arbitrarily well, then for any given $\varepsilon > 0$, hill climbing gives a $\left(1 - \frac{1}{e} - \varepsilon\right)$-approximation algorithm for finding an influential initial k-set I.

Our results: We propose two models for the simultaneous diffusion of two competing technologies on any network given an initial set of early adopters for each technology. Influence functions $\sigma(I_A|I_B)$ are defined to quantify the success of a technology's choice of initial adopters. While the proposed models for diffusion are conceptually simple, we show that maximizing such influence functions subject to a budget is computationally intractable. However, in each case we are able to show that the influence function is non-negative, submodular, and monotone, and hence hill climbing provides an approximation algorithm. We have also generalized these results to address heterogeneous costs for targeting consumers.

3. MODELING THE DIFFUSION OF TWO TECHNOLOGIES

We now extend the independent cascade model to the case of two competing technologies. In particular, we propose two models for describing how two technologies simultaneously diffuse over a given network.

Consumers are again modeled as nodes in a network and links between nodes represent interaction between consumers. We assume that our network is an undirected[1] graph $G = (V, E)$ with $|V| = n$ nodes and $|E| = m$ edges. Nodes can take on one of three states – A and B referring to the two technologies of interest, and C denoting that neither technology is adopted. We can specify two initial sets of nodes – a set of initial adopters of A, $I_A \subseteq V$, and a set of initial adopters of B, $I_B \subseteq V$ (with the implicit assumption that $I_C = V - (I_A \cup I_B)$). We assume that $I_A \cap I_B = \emptyset$. We assume that once a node has chosen a technology, it will not change to another technology, but that nodes that are using one of the two technologies can influence their neighbors that are not using either technology in their decisions to adopt one of the two technologies.

As in the independent cascade model for a single technology, we assume that if u has adopted a particular technology, then u influences neighbor v with probability p_{uv}. Henceforth, we say that an edge is "active" with probability p_{uv}. However, it is now possible that v is influenced by multiple neighbors that use *different* technologies. We will propose two models that govern this "diffusion" of technologies A and B, starting from the sets of initial adopters, given the set of active edges E_a of the network. In other words, the models we develop operate on a random subgraph of the social network G, where each edge is included independently with probability p_{uv}.

Our first model will describe the diffusion of a technology where the product/technology itself can only be obtained from an initial adopter, and a consumer who becomes interested in the technology (i.e. the corresponding node is influenced by a neighbor via an active edge) will pick one of the closest early adopters at random. In the second model, the technology availability is not tied to the network, and the consumer who becomes interested in the technology will choose one of its neighbors and adopt the same technology as this neighbor.

Given such a diffusion model, and the assumption that initially a set of consumers, I_B, is already using technology B, company A would like to choose a set of k consumers, I_A, to target so as to maximize the expected number of consumers reached eventually.

Let the influence function $f(I_A|I_B)$ be the expected number of consumers that will adopt technology A, given that initially the set I_A is using technology A and the set I_B is using technology B. We are hence after a solution of the influence maximization problem:

$$\max \{f(I_A|I_B) : I_A \subseteq V - I_B, |I_A| = k\}. \quad (5)$$

If the cost of targeting consumers varies from consumer to consumer, a company may instead wish to maximize its revenue subject to some budget B. Given non-negative costs $\{c_i : i \in V\}$, the more general from of (5) is then:

$$\max \left\{ f(I_A|I_B) : I_A \subseteq V - I_B, \sum_{i \in I_A} c_i \leq B \right\}. \quad (6)$$

For ease of exposition, throughout the sequel we suppress these costs and will assume that $c_v = 1$ for all $v \in V$.

3.1 A Distance-based model

Our first model is related to competitive facility location [5] on a network. In this model, the location of a node in the network is important, as well as the connectivity of a node. The idea is that a consumer will be more likely to mimic the behavior of an early adopter if their distance in the social network is small.

We assume that for each edge $(u, v) \in E$, we are also given a length d_{uv}. If no length is specified we assume that all edges have length 1. In the following we will assume all edges have length 1, however the results can easily be extended for arbitrary non-negative edge lengths. We let $I = I_A \cup I_B$ be the set of all initial adopters.

Let $d_u(I, E_a)$ denote the shortest distance from u to I along the edges in E_a, with the notation $d_u(I, E_a) = \infty$[2] if and only if u is not connected to any node of I using only active edges. If $d_u(I, E_a) < \infty$, let $\nu_u(I_A, d_u(I, E_a))$

[1]The results in this paper can be easily extended to the more general directed case.

[2]If $d_u(I, E_a) = \infty$, we say that u will adopt neither technology (state C) because it is not connected to any of the initial adopters by active edges. Henceforth, we assume that any node u under consideration is connected to some $v \in I$ in G.

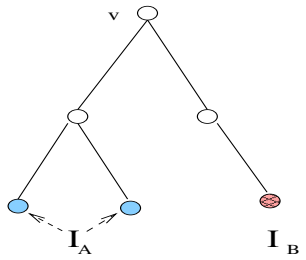

Figure 1: *Given the set of active edges drawn, the probability that node v adopts technology A is $\frac{2}{3}$ in the distance-based model, and $\frac{1}{2}$ in the wave propagation model.*

and $\nu_u(I_B, d_u(I, E_a))$ be the number of nodes in I_A and I_B, respectively, at distance $d_u(I, E_a)$ from u along edges in E_a. Given that $d_u(I, E_a)$ is the shortest distance from u to I along the active edges of G, we will say that node u adopts technology $i \in \{A, B\}$ with probability

$$\frac{\nu_u(I_i, d_u(I, E_a))}{\nu_u(I_A, d_u(I, E_a)) + \nu_u(I_B, d_u(I, E_a))}. \quad (7)$$

Note that – conditioned on set E_a – node u is thus only influenced by nodes in I_A and I_B that are at distance $d_u(I, E_a)$. We note that these are well-defined (conditional) probabilities that sum to one, since at least one of $\nu_u(I_A, d_u(I, E_a))$ and $\nu_u(I_B, d_u(I, E_a))$ is strictly positive.

In this model the expected number of nodes which adopt A will be denoted by

$$\rho(I_A|I_B) = \mathbb{E}\left[\sum_{u\in V} \frac{\nu_u(I_A, d_u(I, E_a))}{\nu_u(I_A, d_u(I, E_a)) + \nu_u(I_B, d_u(I, E_a))}\right],$$

where the expectation is over the set of active edges. We fix I_B and try to determine I_A so as to maximize the expected number of nodes that adopt technology A:

$$\max\{\rho(I_A|I_B) : I_A \subseteq (V - I_B), |I_A| = k\}. \quad (8)$$

3.2 Wave propagation model

Although both of the models we propose here reduce to the independent cascade model of Kempe et al. [12] if there is no competition (if we let $I_B = \emptyset$), our second model for propagation is closer in spirit to the independent cascade model. We motivate this model through the example shown in Figure 1.

In this example, with the edges shown being active, our previous distance-based model gives node v a probability of $\frac{2}{3}$ of adopting technology A, even though it has only two neighbors, one of which adopts technology A and one of which adopts technology B. Under the alternative model presented here a node copies the technology adoption of a neighboring node randomly chosen from the set of its neighbors that are closest to the initial sets (I_A, I_B). In the example, and given the set of active edges shown, this corresponds to giving node v a probability $\frac{1}{2}$ of adopting A and $\frac{1}{2}$ of adopting B.

We can think of the propagation as happening in discrete steps. In step d, all nodes that are at distance at most $d-1$ from some node in the initial sets have adopted technology A or B, and all nodes for which the closest initial node is farther than $d-1$ do not have a technology yet (where the distance is again with respect to active edges). The nodes at a distance d from the initial sets now choose one of their neighbors that are at distance $d-1$ independently at random, and adopt the same technology as this neighbor. As in the previous section, we assume that the node under consideration is in the same connected component of at least one of the nodes $u \in I$.

Formally, let $P(v|I_A, I_B, E_a)$ be the probability that node v adopts technology A when the initial sets for technologies A and B are I_A and I_B, respectively, and the set of active edges is E_a. Let u be a node for which the closest node in $I = I_A \cup I_B$ is at distance d. Let $\mathcal{S}$ be the set of neighbors of u that are at distance $d-1$ from I, where all distances are again with respect to active edges. Then

$$P(u|I_A, I_B, E_a) = \frac{\sum_{v\in\mathcal{S}} P(v|I_A, I_B, E_a)}{|\mathcal{S}|}. \quad (9)$$

For initial sets I_A, I_B, let

$$\pi(I_A|I_B) = \mathbb{E}\left[\sum_{v\in V} P(v|I_A, I_B, E_a)\right] \quad (10)$$

denote the expected number of nodes that adopt technology A. For fixed I_B, we seek a solution to:

$$\max\{\pi(I_A|I_B) : I_A \subseteq (V - I_B), |I_A| = k\}. \quad (11)$$

4. APPROXIMATION ALGORITHMS FOR INFLUENCE MAXIMIZATION

For each of the diffusion models proposed in Section 3 we now show that the decision versions of (8) and (11) are NP-hard but that the corresponding influence functions are nonnegative, monotone and submodular. It will then follow from [18] and [22] that we can use a greedy hill-climbing algorithm to get a $(1 - \frac{1}{e})$-approximation algorithm for these problems. In general it will not be possible to exactly solve the subproblems (2) and (4), as this requires exact evaluation of $\rho(\cdot|I_B)$ and $\pi(\cdot|I_B)$. However, using sampling we can get arbitrarily close approximations of the values needed in (2) and (4). This then allows us to obtain $(1 - \frac{1}{e} - \varepsilon)$-approximation algorithms for both models [12].

THEOREM 1. *For any given I_B with $|V - I_B| \geq k$, the Hill Climbing Algorithm gives a $\left(1 - \frac{1}{e} - \varepsilon\right)$-approximation algorithm for (8).*

PROOF. Given inputs (I_A, I_B), and a set of active edges E_a, $\rho(I_A|I_B)$ can be efficiently evaluated using an algorithm which relies on a single all-pairs shortest paths computation and has overall complexity $\mathcal{O}(|V|^3)$. Using sampling, we can then approximate $\rho(I_A|I_B) = \mathbb{E}[\rho(I_A|I_B)|E_a]$ to within $(1+\gamma)$ for any $\gamma > 0$ (where the running time depends on $\frac{1}{\gamma}$). Hence we can implement the greedy hill-climbing algorithm using $(1+\gamma)$-approximate values for $\rho(I_A \cup \{v\}|I_B)$ in polynomial time. Monotonicity and submodularity of $\rho(\cdot|I_B)$ will be shown in Lemma 2 and Lemma 3, respectively. The approximation guarantee is then an immediate consequence of the results in Section 2. □

THEOREM 2. *For any given I_B with $|V - I_B| \geq k$, the Hill Climbing Algorithm gives a $\left(1 - \frac{1}{e} - \varepsilon\right)$-approximation algorithm for (11).*

PROOF. Given inputs (I_A, I_B) and a set of active edges E_a, $\pi(I_A|I_B)$ can be efficiently evaluated using an algorithm which relies on a single all-pairs shortest path computation

and has overall complexity $\mathcal{O}(|V|^3)$. We can approximate $\pi(I_A|I_B) = \mathbb{E}[\pi(I_A|I_B)|E_a]$ to within $(1+\gamma)$ for any $\gamma > 0$, hence we can implement the greedy hill-climbing algorithm using $(1+\gamma)$-approximate values for $\pi(I_A \cup \{v\}|I_B)$ in polynomial time. Monotonicity and submodularity of $\pi(\cdot|I_B)$ will be shown Lemma 5 and Lemma 6, respectively. The approximation guarantee is again an immediate consequence of the results in Section 2. □

Before proceeding, we note that to show NP-hardness and the desired properties of the influence functions, it suffices to consider the case when $p_{uv} = 1$ for all edges (or equivalently, $E_a = E$): NP-hardness of a special case clearly implies NP-hardness of the more general case, and the expected value of a function of E_a is nonnegative, monotone, and submodular if for any E_a the function is nonnegative, monotone, and submodular. For ease of exposition, we therefore restrict ourselves to this special case in the remainder of this section.

4.1 A Distance-based model

Let the decision version of (8) be to determine if there is a set I_A of size k with $\rho(I_A|I_B) \geq M$ for any $M \in \mathbb{Q}$. We then have the following result.

THEOREM 3. *The decision version of (8) is NP-hard* [3].

PROOF. Given a ground set of elements $E = \{e_i : 1 \leq i \leq n\}$ and a collection of sets $S = \{s_i : 1 \leq i \leq m\}$ such that each $s_i \subseteq E$, the decision version of the *set cover* problem asks if there is a collection of k sets covering all elements in E. Without loss of generality we assume that every element is covered by at least one set and that $k < \min(m, n)$. We reduce the NP-hard set cover decision problem to the decision version of (8).

Given an instance (S, E, k) of set cover, we construct a graph H as follows. We add a node s_i for each set $s_i \in S$ and a node e_j for each element $e_j \in E$. We add an edge (s_i, e_j) if and only if $e_j \in s_i$. We add an additional node x and connect it to each e_j through another node d_j (see Figure 2). Lastly, for a constant $\kappa > 0$ to be specified in the subsequent lemma, we construct a cluster C_j of κ nodes for each $j = 1, \ldots, n$ and connect each of the nodes in C_j to e_j. The following lemma completes the reduction by specifying the value of κ. □

LEMMA 1. *Let $\kappa > (k+1)(m+n)$ and $I_B = \{x\}$. There is a collection of k sets which cover E if and only if there is a set I_A of k nodes in the graph H such that $\rho(I_A|I_B) \geq n(\kappa+1)$.*

PROOF. If there a collection of k sets covering E then take I_A to be the k nodes corresponding to those sets. This gives $\rho(I_A|I_B) \geq n(\kappa+1)$ by the following argument. Each of the nodes e_j is adjacent to one of the nodes in I_A since the corresponding sets form a cover. Each e_j and the nodes in each cluster C_j adopt A with probability one since initial adopters of A are the closest for these nodes. This results in at least $n(\kappa+1)$ nodes with technology A.

If there is no collection of k sets that cover E then we prove that no set I_A of k nodes can be the initial adopters of A and still achieve $\rho(I_A|I_B) \geq n(\kappa+1)$. Consider the nodes $e_1, \ldots, e_n$. There exists a node e_j which adopts technology B

[3]Kempe et al. showed that this problem is NP-hard for the *directed* case when $I_B = \emptyset$. However the problem is not NP-hard for the *undirected* case when $I_B = \emptyset$.

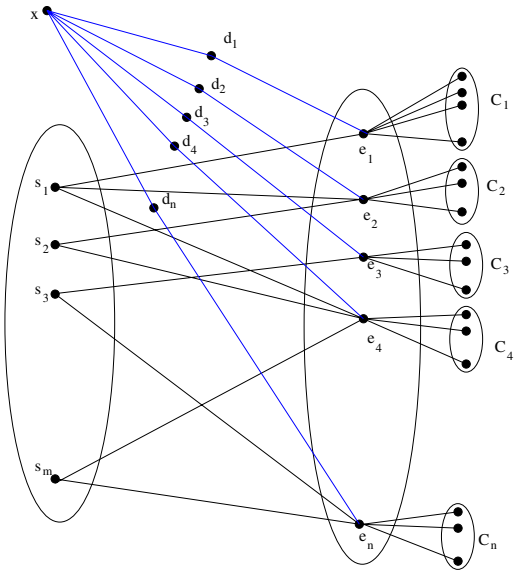

Figure 2: Graph H for set cover reduction.

with probability at least $\frac{1}{k+1}$ since any set I_A of size k cannot be within a distance of 1 from all e_i (by the construction of H and lack of a cover). This implies that e_j and all nodes in C_j adopt technology B with probability at least $\frac{1}{k+1}$. So

$$\begin{aligned}\rho(I_A|I_B) &\leq \underbrace{(n - \frac{1}{k+1})(\kappa+1)}_{\text{from } e_1,\ldots,e_n \text{ and } C_1,\ldots,C_n} + \underbrace{m+n}_{\text{from } s_1,\ldots,s_m \text{ and } d_1,\ldots,d_n} \\ &= n(\kappa+1) + (m+n) - \frac{1}{k+1}(\kappa+1) < n(\kappa+1). \quad \square\end{aligned}$$

Having shown the hardness of (8), we now turn to the Lemmas required in Theorem 1 and show that the influence function ρ is both monotone and submodular. We assume without loss of generality that every edge is active with probability 1, and for ease of notation, we will write $d_u(I)$ instead of $d_u(I, E_a)$. Furthermore, we will drop the subscript u when u is clear from the context.

LEMMA 2. *For any I_B, $\rho(I_A|I_B)$ is a monotone function of I_A.*

PROOF. For a fixed $u \in V$ and initial set I_B, it suffices to show for any $v \in V - I_B$, and any $I_A \subseteq V$, the probability that u adopts A given the initial set I_A is at most the probability that u adopts A when the initial set is $I_A \cup \{v\}$, that is:

$$\frac{\nu(I_A, d(I))}{\nu(I_A, d(I)) + \nu(I_B, d(I))} \leq \frac{\nu(I_A \cup \{v\}, d(I \cup \{v\}))}{\nu(I_A \cup \{v\}, d(I \cup \{v\})) + \nu(I_B, d(I \cup \{v\}))}.$$

We note that the shortest distance from u to a node in I is not smaller than the shortest distance from u to a node in $I \cup \{v\}$, so $d(I) \geq d(I \cup \{v\})$.

Now, if $d(I \cup \{v\}) < d(I)$, then $\nu(I_B, d(I \cup \{v\})) = 0$, so the right hand side is 1, and the inequality clearly holds. Otherwise, $\nu(I_B, d(I \cup \{v\})) = \nu(I_B, d(I))$, and $\nu(I_A, d(I)) =$

$\nu(I_A, d(I \cup \{v\})) \leq \nu(I_A \cup \{v\}, d(I \cup \{v\})$ and the inequality holds since for real numbers $c \geq a \geq 0$ and $b > 0$, $\frac{c}{c+b} \geq \frac{a}{a+b}$. □

LEMMA 3. *For any I_B, $\rho(I_A|I_B)$ is a submodular function of I_A.*

PROOF. For a set of initial adopters of A, S, and a node $x \in V - (S \cup I_B)$, we define the increase in the probability that node u adopts technology A when adding x to the initial set S as:

$$P(u,S,x) = \frac{\nu(S \cup \{x\}, d(S \cup \{x\} \cup I_B))}{\nu(S \cup \{x\}, d(S \cup \{x\} \cup I_B)) + \nu(I_B, d(S \cup \{x\} \cup I_B))} - \frac{\nu(S, d(S \cup I_B))}{\nu(S, d(S \cup I_B)) + \nu(I_B, d(S \cup I_B))}.$$

We need to show that for any node $u \in V$ and $S \subseteq T \subseteq V$, $P(u,S,x) \geq P(u,T,x)$.

Let $d_1 = d(S)$, $d_2 = d(T)$, $d_3 = d(I_B)$. Since $S \subseteq T$, $d_1 \geq d_2$. We analyze three cases:

Case 1 ($d_1 \geq d_2 \geq d_3$): If $d(u,x) > d_3$, adding x does not change the probability of u adopting A. So $P(u,S,x) = P(u,T,x) = 0$. If $d(u,x) < d_3$, then adding x makes u adopt A with probability 1. It then follows from the monotonicity of ρ that

$$\begin{aligned} P(u,S,x) &= 1 - \frac{\nu(S, d(S \cup I_B))}{\nu(S, d(S \cup I_B)) + \nu(I_B, d(S \cup I_B))} \\ &\geq 1 - \frac{\nu(T, d(T \cup I_B))}{\nu(T, d(T \cup I_B)) + \nu(I_B, d(T \cup I_B))} \\ &= P(u,T,x). \end{aligned}$$

If $d(u,x) = d_3$, then $\nu(S, d(S \cup I_B)) = \nu(S, d_3)$ and $\nu(T, d(T \cup I_B)) = \nu(T, d_3)$ and these both increase by 1 if x is added. Furthermore $\nu(I_B, d(X \cup I_B)) = \nu(I_B, d_3)$ for $X \in \{S, S \cup \{x\}, T, T \cup \{x\}\}$. So we need to show that

$$\frac{\nu(S,d_3)+1}{\nu(S,d_3)+1+\nu(I_B,d_3)} - \frac{\nu(S,d_3)}{\nu(S,d_3)+\nu(I_B,d_3)} \geq \frac{\nu(T,d_3)+1}{\nu(T,d_3)+1+\nu(I_B,d_3)} - \frac{\nu(T,d_3)}{\nu(T,d_3)+\nu(I_B,d_3)}.$$

This equation can be easily checked to be true using the fact that $\nu(S,d_3) \leq \nu(T,d_3)$.

Case 2 ($d_3 > d_1 \geq d_2$): In this case $\nu(I_B, d(X \cup I_B)) = 0$ for $X \in \{S, S \cup \{x\}, T, T \cup \{x\}\}$. In this case the probability that u adopts A is 1 for initial sets $X \in \{S, S \cup \{x\}, T, T \cup \{x\}\}$, so $P(u,S,x) = P(u,T,x) = 0$.

Case 3 ($d_1 \geq d_3 > d_2$): Since $d_3 > d_2$, u will adopt technology A with probability 1 if the initial set is T or $T \cup \{x\}$. So $P(u,T,x) = 0$, and $P(u,S,x) \geq 0$ holds by Lemma 2. □

4.2 Wave propagation model

Since it suffices to show that $\pi(I_A|I_B)$ is monotone and submodular in the special case that every edge in the graph is active with probability 1, we will restrict ourselves to this case and write $P(u, I_A, I_B)$ instead of $P(u|I_A, I_B, E_a)$ for the probability that node u adopts technology A when the initial sets for technology A and B are I_A and I_B, respectively and the set of active edges is E_a.

Let the decision version of (11) be to determine if there is a set I_A of size k with $\pi(I_A|I_B) \geq M$ for any $M \in \mathbb{Q}$. We then have the following result.

THEOREM 4. *The decision version of (11) is NP-hard.*

PROOF. We reduce the NP-hard set cover decision problem to the decision version of (11) as in Theorem 3. Given an instance (S, E, k) of set cover, we construct the same graph H constructed in the proof of Theorem 3. The following lemma completes the proof. □

LEMMA 4. *Let $\kappa > (m+1)(m+n)$ and $I_B = \{x\}$. There is a collection of k sets which cover E if and only if there is a set I_A of k nodes in the graph H such that $\pi(I_A|I_B) \geq n(\kappa+1)$.*

PROOF. If there a collection of k sets covering E then take I_A to be the k nodes corresponding to those sets. This gives $\pi(I_A|I_B) \geq n(\kappa+1)$ by the same argument given in the proof of Lemma 1.

If there is no collection of k sets that cover E then we prove that no set I_A of k nodes can be the initial adopters of A and still achieve $\pi(I_A|I_B) \geq n(\kappa+1)$. Consider the nodes $e_1, \ldots, e_n$. Any set I_A of size k cannot be within a distance of 1 from all e_j (by the construction of H and lack of a cover). So there exists a node e_j which adopts technology B with probability at least $\frac{1}{m+1}$ because one of its neighbors d_j has $P(d_j|I_A, I_B) = 0$ and is at distance 1 from I and at most $m+1$ of its neighbors are at distance 1 from I. So $P(e_j|I_A, I_B) \leq \frac{m}{m+1}$. This implies that e_j and all the nodes in C_j adopt technology A with probability at most $\frac{m}{m+1}$. So for any initial set I_A of size k,

$$\begin{aligned} \pi(I_A|I_B) &\leq \sum_{v \in V} P(v, I_A, I_B) \\ &\leq \underbrace{(m+n)}_{\text{for } v \in \{s_1,\ldots,s_m\} \cup \{d_1,\ldots,d_n\}} + \underbrace{(n-1)(\kappa+1)}_{\text{for } v \in e_i \cup C_i, i \neq j} \\ &\quad + \underbrace{(\kappa+1)P(e_j|I_A,I_B)}_{v \in e_j \cup C_j} \\ &\leq (m+n) + (n-1)(\kappa+1) + (\kappa+1)\frac{m}{m+1} \\ &< (n-1)(\kappa+1) + (m+1)(m+n) \\ &< n(\kappa+1). \quad \square \end{aligned}$$

We again benefit from the valuable properties of monotonicity and submodularity.

LEMMA 5. *For any I_B, $\pi(I_A|I_B)$ is a monotone function of I_A.*

PROOF. To prove monotonicity we need to show that $P(u|S \cup x, I_B) \geq P(u|S, I_B)$ for all $x \in V - I_B$. We employ the same notation as in Section 4.1 and let $n(v) = \{u : (u,v) \in E\}$ denote the neighbors of node v. Note that $d(u, S \cup x \cup I_B) \leq d(u, S \cup I_B)$. If $d(u, S \cup x \cup I_B) < d(u, S \cup I_B)$ then $P(u|S \cup x, I_B) = 1 \geq P(u|S, I_B)$ which proves monotonicity. So the interesting case is when $d(u, S \cup x \cup I_B) = d(u, S \cup I_B)$. We prove $P(u|S \cup x, I_B) \geq P(u|S, I_B)$ for this case by induction on the distance $d = d(u, S \cup I_B)$.

Base case: $d = 1$. If x is not a neighbor of u then $P(u|S \cup x, I_B) = P(u|S, I_B)$. If x is a neighbor of u then $P(u|S \cup x, I_B) = \frac{1+|n(u) \cap S|}{1+|n(u) \cap (S \cup I_B)|} \geq \frac{|n(u) \cap S|}{|n(u) \cap (S \cup I_B)|} = P(u|S, I_B)$.

Induction step: Now we prove monotonicity for nodes u such that $d(u, S \cup I_B) = d$ assuming monotonicity for all the nodes v with $d(v, S \cup I_B) < d$. Let $\mathcal{S}$ be the set of neighbors

of u which are at a distance $d-1$ from $S \cup I_B$. Let $\mathcal{K}$ be the set of neighbors of u which are at a distance $d-1$ from x but at a distance greater than $d-1$ from $S \cup I_B$. Let $K = |\mathcal{K}|$. Note that all $v \in \mathcal{K}$ have $P(v|S \cup x, I_B) = 1$. The probability of u accepting technology A is then:

$$\begin{aligned} P(u|S \cup x, I_B) &= \frac{K + \sum_{v \in \mathcal{S}} P(v|S \cup x, I_B)}{K + |\mathcal{S}|} \\ &\geq \frac{\sum_{v \in \mathcal{S}} P(v|S \cup x, I_B)}{|\mathcal{S}|} \\ &\geq \frac{\sum_{v \in \mathcal{S}} P(v|S, I_B)}{|\mathcal{S}|} = P(u|S, I_B). \end{aligned}$$

The second inequality follows from the induction assumption that monotonicity holds for the nodes v with $d(v, S \cup I_B) < d$ and the fact that all nodes in $\mathcal{S}$ are at a distance $d-1$ from $S \cup I_B$. □

LEMMA 6. *For any I_B, $\pi(I_A|I_B)$ is a submodular function of I_A.*

PROOF. We will show that for two sets $S \subseteq T \subseteq V - I_B$, and a node $x \in V - I_B$, we have that $\forall u \in V$

$$P(u|S \cup x, I_B) - P(u|S, I_B) \geq P(u|T \cup x, I_B) - P(u|T, I_B)$$

by induction on $d = d(u, x)$. If $d = 0$, then clearly the inequality holds. Suppose it holds for any v such that $d(v, x) = d - 1$.

As in the proof of Lemma 3, we consider different cases for the distance from u to the closest node in S, T and I_B. Let $d_1 = d(u, S), d_2 = d(u, T), d_3 = d(u, I_B)$. It is easy to see that the proof of Lemma 3 also works for our alternative model, except for the case when $d_1 \geq d_2 \geq d_3$ and $d_3 = d(u, x)$.

Let $\mathcal{S}$ be the set of neighbors of u for whom the closest node from $S \cup I_B$ is at distance $d-1$ so that:

$$P(u|S, I_B) = \frac{\sum_{v \in \mathcal{S}} P(v|S, I_B)}{|\mathcal{S}|}.$$

Note that each neighbor of u that is at distance $d-1$ from x but is at distance greater than $d-1$ from the nodes in $S \cup I_B$, adopts A with probability 1. Let K be the number of such nodes, then:

$$P(u|S \cup x, I_B) = \frac{K + \sum_{v \in \mathcal{S}} P(v|S \cup x, I_B)}{K + |\mathcal{S}|}.$$

Therefore the difference in the probability of u adopting A is:

$$\begin{aligned} &P(u|S \cup x, I_B) - P(u|S, I_B) \\ &= \frac{K + \sum_{v \in \mathcal{S}} P(v|S \cup x, I_B)}{K + |\mathcal{S}|} - \frac{\sum_{v \in \mathcal{S}} P(v|S, I_B)}{|\mathcal{S}|} \\ &= \frac{K}{K + |\mathcal{S}|} + \frac{\sum_{v \in \mathcal{S}} P(v|S \cup x, I_B) - \sum_{v \in \mathcal{S}} P(v|S, I_B)}{K + |\mathcal{S}|} \\ &\quad - \frac{K \sum_{v \in \mathcal{S}} P(v|S, I_B)}{(K + |\mathcal{S}|)|\mathcal{S}|} \\ &= \frac{\sum_{v \in \mathcal{S}} (P(v|S \cup x, I_B) - P(v|S, I_B))}{K + |\mathcal{S}|} \\ &\quad + \frac{K}{K + |\mathcal{S}|} \frac{\sum_{v \in \mathcal{S}} (1 - P(v|S, I_B))}{|\mathcal{S}|}. \end{aligned}$$

Similarly, let $\mathcal{T}$ be the set of neighbors of u for whom the closest node from $T \cup I_B$ is at distance $d-1$, and let L be the number of neighbors of u that are at distance $d-1$ from x, and at distance greater than $d-1$ from $T \cup I_B$. Then:

$$\begin{aligned} &P(u|T \cup x, I_B) - P(u|T, I_B) \\ &= \frac{\sum_{v \in \mathcal{T}} (P(v|T \cup x, I_B) - P(v|T, I_B))}{L + |\mathcal{T}|} \\ &\quad + \frac{L}{L + |\mathcal{T}|} \frac{\sum_{v \in \mathcal{T}} (1 - P(v|T, I_B))}{|\mathcal{T}|} \end{aligned}$$

We now establish the following three inequalities:

$$\frac{K}{K + |\mathcal{S}|} \geq \frac{L}{L + |\mathcal{T}|} \tag{12}$$

$$\begin{aligned} &\frac{\sum_{v \in \mathcal{S}} (P(v|S \cup x, I_B) - P(v|S, I_B))}{K + |\mathcal{S}|} \\ &\geq \frac{\sum_{v \in \mathcal{T}} (P(v|T \cup x, I_B) - P(v|T, I_B))}{L + |\mathcal{T}|} \end{aligned} \tag{13}$$

$$\frac{\sum_{v \in \mathcal{S}} (1 - P(v|S, I_B))}{|\mathcal{S}|} \geq \frac{\sum_{v \in \mathcal{T}} (1 - P(v|T, I_B))}{|\mathcal{T}|} \tag{14}$$

Clearly these inequalities imply that

$$P(u|S \cup x, I_B) - P(u|S, I_B) \geq P(u|T \cup x, I_B) - P(u|T, I_B).$$

To prove (12), let $\mathcal{K}$ and $\mathcal{L}$ be the set of neighbors of u that are at distance $d-1$ from x, and at distance greater than $d-1$ from $S \cup I_B$ and $T \cup I_B$, respectively. (So $K = |\mathcal{K}|, L = |\mathcal{L}|$). Since $S \subseteq T$, we have $\mathcal{K} \supseteq \mathcal{L}$ and hence $K \geq L$. Now, $\mathcal{T} \cup \mathcal{L}$ is the set of neighbors of u that are at distance $d-1$ from $T \cup x \cup I_B$, and $\mathcal{S} \cup \mathcal{K}$ is the set of neighbors of u that are at distance $d-1$ from $S \cup x \cup I_B$, so $\mathcal{S} \cup \mathcal{K} \subseteq \mathcal{T} \cup \mathcal{L}$. Since $\mathcal{T} \cap \mathcal{L} = \mathcal{S} \cap \mathcal{K} = \emptyset$, we get that $K + |\mathcal{S}| \leq L + |\mathcal{T}|$. Combining this with $K \geq L$ we obtain (12).

To prove (13), we note that for $v \in \mathcal{T} - \mathcal{S}$, we must have $P(v|T, I_B) = P(v|T \cup x, I_B) = 1$. Since $v \notin \mathcal{S}$, the shortest distance from v to any node in I_B is greater than $d-1$, and since $v \in \mathcal{T}$, there must be a node in T that is at distance $d-1$ from v. Hence:

$$\begin{aligned} &\sum_{v \in \mathcal{T}} [P(v|T \cup x, I_B) - P(v|T, I_B)] \\ &= \sum_{v \in \mathcal{S}} [P(v|T \cup x, I_B) - P(v|T, I_B)] \\ &\geq \sum_{v \in \mathcal{S}} [P(v|S \cup x, I_B) - P(v|S, I_B)], \end{aligned}$$

where the inequality follows from induction. We established above that $K + |\mathcal{S}| \leq L + |\mathcal{T}|$, which completes the proof of (13).

For (14), we again use the fact that $P(v|T, I_B) = 0$ for $v \in \mathcal{T} - \mathcal{S}$ and obtain:

$$\begin{aligned} \sum_{v \in \mathcal{T}} [1 - P(v|T, I_B)] &= \sum_{v \in \mathcal{S}} [1 - P(v|T, I_B)] \\ &\leq \sum_{v \in \mathcal{S}} [1 - P(v|S, I_B)], \end{aligned}$$

where the inequality follows from monotonicity. The fact that $|\mathcal{T}| \geq |\mathcal{S}|$ gives (14). □

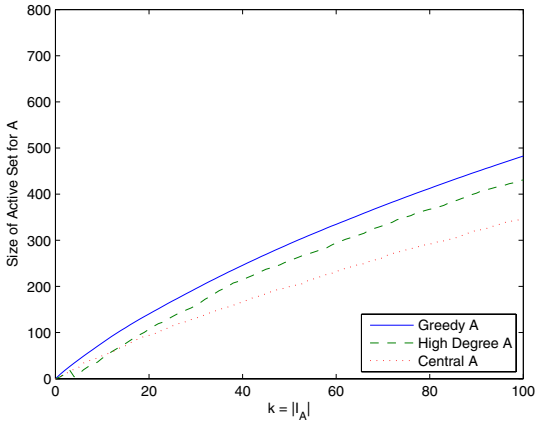

Figure 3: Distance-based model: high-degree I_B

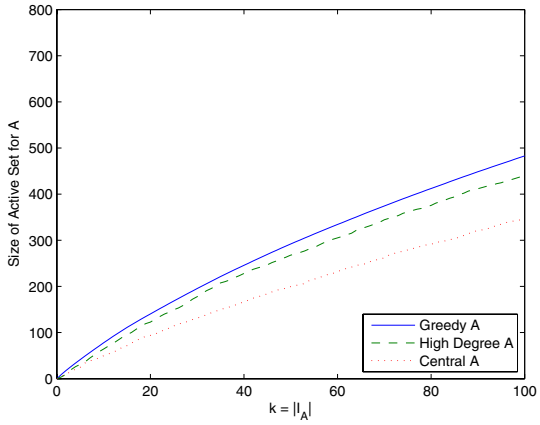

Figure 4: Wave propagation model: high-degree I_B

5. NUMERICAL SIMULATIONS

In this section we analyze the behavior of both models and the resulting influence sets of each on a real network – the coauthorship graph based on papers in theoretical high-energy physics. Empirical evidence suggests that coauthorship graphs are representative of typical social networks [19]. By choosing to run our experiments on the data from an actual social network as opposed to generating random graphs, we are able to obtain results that are more specifically applicable to the motivations for our models.

The specific dataset we employed was the PROXIMITY HEP-Th database based on data from the arXiv archive and the Stanford Linear Accelerator Center SPIRES-HEP database provided for the 2003 KDD Cup competition with additional preparation performed by the Knowledge Discovery Laboratory, University of Massachusetts Amherst [20]. After minor preprocessing, the network consisted of 8392 distinct authors and 461 separate connected components (of size at least 2), the largest of which contained 7034 authors.

We compared different choices for companies A and B, where company B first chooses a certain subset of the nodes, I_B, unaware that company A will also try to enter the market, and company A subsequently targets a subset I_A, after which we look at the spreading of influence from I_A and I_B according to the processes described in Section 3.

We ran simulations where the set I_B was chosen according to several different heuristics. As discussed in [12] the heuristics of choosing high-degree nodes and central nodes are often used in the sociology literature to find influential sets of nodes. Here the high-degree heuristic chooses nodes in order of highest degree, while the central node heuristic chooses nodes with low average distance to other nodes. The average distance is calculated by taking the average of a node's distance to all other nodes, where the distance between unconnected nodes is the number of nodes in the graph. In addition to these two heuristics, we also ran simulations where the nodes of I_B were chosen according to the greedy Hill Climbing Algorithm for the single technology case [12].

We used each of these three heuristics to choose an initial set I_B of fixed size $|I_B| = 100$ corresponding to a little more than 1% of the nodes in the network. For each of these I_B sets, and for both diffusion models from Section 3, we ran the greedy Hill Climbing Algorithm to determine the most influential I_A set, where $|I_A|$ ranged from 1 to 100 nodes. Since the problem of finding the best I_A of a fixed size is NP-complete, we compare the results of the algorithm against two heuristics for choosing the I_A set from $V - I_B$: high-degree nodes and central nodes.

As in Kempe et al. [12], we begin by giving each edge (u, v) in the network a probability $p_{uv} = .1$ of being active. We suppress the details of the simulation procedure employed, but note here that when $p_{uv} \in (0, 1)$, many random subgraphs must be generated to both obtain a node with the largest marginal expected influence and to evaluate the overall influence of all methods.

Figure 3 compares the size of the market which product A captures for increasing values of $k = |I_A|$ in the distance-based model from Section 4.1 if A uses different heuristics, where the 100 nodes of I_B are chosen according to the high degree heuristic. Figure 4 is similar to Figure 3 but uses the wave-propagation model from Section 3.2. Due to space limitations, we only present results for the wave propagation result in the sequel. In all of the experiments which we conducted, the Hill Climbing Algorithm for company A outperformed the other heuristics. This can be attributed to the fact that the Hill Climbing Algorithm takes into account the effect of both the nodes in I_B and the nodes already selected in I_A.

In Figure 5, we fix company A's strategy to the greedy Hill Climbing Algorithm and compare the strategies of company B. We see that for large enough $|I_A|$, company B's market share is smallest when B used the Hill Climbing Algorithm for a single technology, even though Kempe et al. [12] experimentally showed that in the absence of competition, this algorithm performed best. This can be explained by the fact that the greedy algorithm for a single technology iteratively adds the node to I_B that maximizes the expected number of additional nodes influenced when there is no competition, which inherently does not make the solution very robust when an unexpected competitor also tries to influence nodes. When $|I_A| > 5$, the high degree heuristic helped company B maintain the largest possible market share in all of our experiments.

Figure 6 shows the percentage of the market captured by A and B, respectively and in total, for increasing sizes

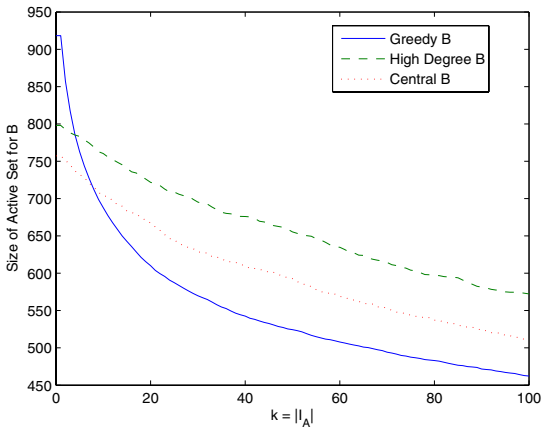

Figure 5: B's share of the market against a greedy I_A (Wave propagation model)

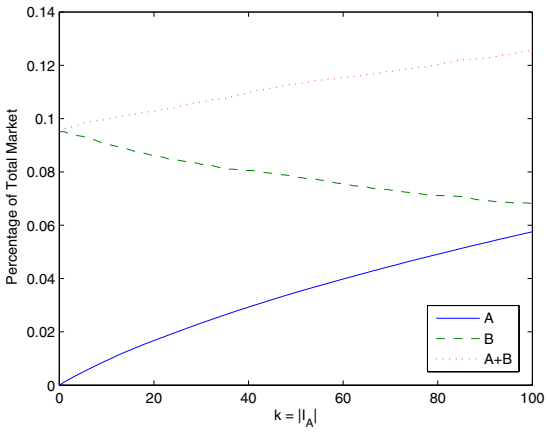

Figure 6: Percentage of the total market captured (Wave propagation model: high-degree I_B)

of I_A when using our greedy algorithm for choosing I_A in the wave propagation model. Since the growth in the total number of adopters of A and B is slower than the growth of A's influence, this figure shows that A's increase in market share is due to both reaching new consumers and drawing consumers away from its competitor. For I_B chosen using the high-degree heuristic, we see that B maintains a larger market share even when $|I_A| = |I_B| = 100$. When B chooses I_B with either of the other two methods considered, A is able to obtain a lead with a considerably smaller initial set.

Figure 7 shows how much larger an initial set each of the heuristics require relative to the greedy algorithm's initial set to attain some specified level of influence. Here we see that to influence 300-400 consumers, the high-degree nodes heuristic requires an initial set which is approximately 15% larger than that required by the greedy algorithm. In particular, this quantifies how much better the greedy algorithm performs relative to the popular heuristics. This information could also be used by A to determine the value of knowing precisely what consumers B will target, since the other heuristics do not require this knowledge.

Figure 8 shows the marginal gain in influence which A enjoys from targeting an additional consumer versus the three different strategies for B. Here we observe that simply greedily targeting the most influential consumer yields an approximate expected return of more than 90 eventual adopters if B chose greedily and more than 10 eventual adopters if B used the high-degree heuristic. We note that given costs for targeting consumers, this figure could help a company decide at what point the cost of targeting an additional consumer outweighs the marginal return expected from this action.

6. CONCLUDING REMARKS

In this paper we studied the spreading of two competing technologies, A and B, in a social network. We addressed the question of finding an initial set of nodes to target for technology A, given that the initial set of nodes adopting technology B is known. To our knowledge, this work represents the first treatment of such questions. We proposed two basic models for the spreading of technologies through a network, in which the two technologies propagate in exactly the same way. These models and our results can be easily extended to handle additional competitors. Furthermore, by adding dummy nodes to the network, our model easily allows for the case when companies can target customers, but the targeted customer adopts the technology only with a certain probability.

We believe our results could also be extended to include more general cases, for example by having different acceptance probabilities for the two technologies, or by allowing the rate at which influence travels in the graph differ for the two technologies.

From a game-theoretic perspective, the question we study is that of finding a best response to the first player's move in a Stackelberg game. A natural next step would be to study the optimal behavior of the first player, given that she knows that the second player will use our approximate best response, and ultimately to study the Nash equilibria of this Stackelberg game. We have shown that A can obtain a significant portion of a market despite choosing consumers second. In our experiments, the heuristic that chooses high degree nodes was the best strategy for the first player among the three strategies we compared. If B doesn't choose their initial set in this way, A can in fact outperform B with a relatively small budget. It would be interesting to find a provably good heuristic for the first player. Other interesting games that could be considered using our models are the simultaneous version of this game, and the game where the two players take turns in targeting nodes.

Lastly, using our first model with edge probabilities equal to 1, these problems can also be seen in the context of competitive facility location [1, 4] on a network, but we are not aware of any previous results for competitive location games on a network.

7. ACKNOWLEDGMENTS

We are grateful to Eric Friedman, Jon Kleinberg, David Williamson, and an anonymous referee for their helpful comments on an earlier version of this paper. We also thank David Shmoys, Christine Shoemaker, and David Williamson for financial support.

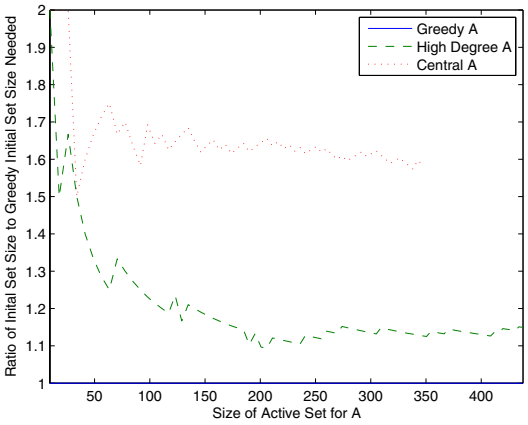

Figure 7: The benefit of knowing I_B, (Wave propagation model: high-degree I_B)

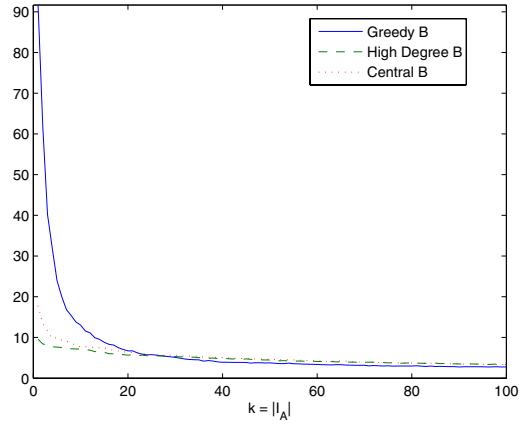

Figure 8: A's gain in adding an early adopter (Wave propagation model: greedy I_A)

8. REFERENCES

[1] H.-K. Ahn, S.-W. Cheng, O. Cheong, M. Golin, and R. van Oostrum. Competitive facility location: the Voronoi game. *Theoret. Comput. Sci.*, 310(1-3):457–467, 2004.

[2] W. B. Arthur. Competing technologies, increasing returns, and lock-in by historical events. *Economic Journal*, 99(394):116–31, March 1989.

[3] P. A. David. *Technical Choice, Innovation and Economic Growth*. Cambridge University Press, 1975.

[4] G. Dobson and U. S. Karmarkar. Competitive location on a network. *Oper. Res.*, 35(4):565–574, 1987.

[5] H. Eiselt and G. Laporte. Competitive spatial models. *European Journal of Operational Research*, 39:231–242, 1989.

[6] J. Farrell and G. Saloner. Installed base and compatibility: Innovation, product preannouncements, and predation. *American Economic Review*, 76:940–955, 1986.

[7] D. Fudenberg and J. Tirole. *Game Theory*. MIT Press, 1991.

[8] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001.

[9] S. Hill, F. Provost, and C. Volinsky. Network-based marketing: Identifying likely adopters via consumer networks. *Journal of Computational and Graphical Statistics*, 21(2):256–276, 2006.

[10] M. O. Jackson and L. Yariv. Diffusion on social networks. *Économie publique*, 16(1):69–82, 2006.

[11] M. L. Katz and C. Shapiro. Systems competition and network effects. *Journal of Economic Perspectives*, 8(2):93–115, 1994.

[12] D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 246–257, 2003.

[13] D. Kempe, J. Kleinberg, and É. Tardos. Influential nodes in a diffusion model for social networks. In *32nd International Colloquium on Automata, Languages and Programming (ICALP)*, pages 1127–1138, 2005.

[14] S. Khuller, A. Moss, and J. S. Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70(1):39–45, 1999.

[15] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *ACM Conference on Electronic Commerce*, 2006.

[16] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. In *Proc. Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2006.

[17] D. López-Pintado. Contagion and coordination in random networks. Technical report, Columbia, September 9, 2005.

[18] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher. An analysis of approximations for maximizing submodular set functions-1. *Mathematical Programming*, 14:265–294, 1978.

[19] M. E. J. Newman. The structure of scientific collaboration networks. *Proc. National Academy of Sciences USA*, 98(2):404–409, 2001.

[20] U. of M-A. Knowledge discovery lab proximity databases. `http://kdl.cs.umass.edu/data`. Accessed April 17, 2006.

[21] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Eighth Intl. Conf. on Knowledge Discovery and Data Mining*, pages 61–70, 2002.

[22] M. Sviridenko. A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters*, 32(1):41–43, 2004.

[23] M. Tomochi, H. Murata, and M. Kono. A consumer-based model of competitive diffusion: the multiplicative effects of global and local network externalities. *Journal of Evolutionary Economics*, 15:273–295, 2005.

[24] T. W. Valente. *Network Models of the Diffusion of Innovations*. Quantitative Methods in Communication Subseries. Hampton Press, New York, NY, 1995.