On the Submodularity of Influence in Social Networks

Elchanan Mossel*
Dept. of Statistics
U.C. Berkeley
mossel@stat.berkeley.edu

Sebastien Roch†
Dept. of Statistics
U.C. Berkeley
sroch@stat.berkeley.edu

ABSTRACT

We prove and extend a conjecture of Kempe, Kleinberg, and Tardos (KKT) on the spread of influence in social networks.

A social network can be represented by a directed graph where the nodes are individuals and the edges indicate a form of social relationship. A simple way to model the diffusion of ideas, innovative behavior, or "word-of-mouth" effects on such a graph is to consider an increasing process of "infected" (or active) nodes: each node becomes infected once an activation function of the set of its infected neighbors crosses a certain threshold value. Such a model was introduced by KKT in [7, 8] where the authors also impose several natural assumptions: the threshold values are (uniformly) random to account for our lack of knowledge of the true values; and the activation functions are monotone and submodular, i.e. have "diminishing returns." The monotonicity condition indicates that a node is more likely to become active if more of its neighbors are active, while the submodularity condition, indicates that the marginal effect of each neighbor is decreasing when the set of active neighbors increases.

For an initial set of active nodes S, let $\sigma(S)$ denote the expected number of active nodes at termination. Here we prove a conjecture of KKT: we show that the function $\sigma(S)$ is submodular under the assumptions above. We prove the same result for the expected value of any monotone, submodular function of the set of active nodes at termination.

In other words, our results demonstrate that "local" submodularity is preserved "globally" under diffusion processes. This is of natural computational interest, as many optimization problems have good approximation algorithms for submodular functions. In particular, our results coupled with an argument in [7] imply that a greedy algorithm gives an $(1 - 1/e - \varepsilon)$-approximation algorithm for maximizing $\sigma(S)$ among all sets S of a given size. This result has important practical implications for many social network analysis problems, notably viral marketing.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Stochastic Processes

General Terms

Algorithms, Economics, Theory

Keywords

Viral marketing, social networks, submodularity, coupling

1. INTRODUCTION

Social Networks. In recent years, diffusion processes on social networks have been the focus of intense study. While traditionally such processes have been of major interest in epidemiology where they model the spread of diseases and immunization, see e.g. [12, 9, 10, 4, 1, 5], much of the recent interest has resulted from applications in sociology, economics, and engineering. (See e.g. [7] for references.)

In computer science, a strong motivation for analyzing diffusion processes has recently emanated from the study of viral marketing strategies in data mining, where various novel algorithmic problems have been considered [2, 3, 7, 8]. Roughly speaking, *viral marketing*—unlike conventional marketing—takes into account the "network value" of potential customers, i.e. it seeks to target a set of individuals whose influence on their social network through word-of-mouth effects is high. (For more background on viral marketing, see [2, 3, 7, 8].)

Commonly-used heuristics to identify influential nodes in social networks include picking individuals of high degree—so-called degree centrality heuristics—or picking individuals with short average distance to the rest of the network—so-called distance centrality heuristics [14]. Here we prove a structural conjecture of Kempe, Kleinberg and Tardos (KKT) [7, 8], which can be roughly stated as follows: if a diffusion model is locally submodular, i.e. the influence of an individual on its neighbors in the network has "diminishing returns," then the process is globally submodular. This is relevant here because, under the submodularity property, optimization problems, such as the viral marketing problem, are known to have good approximation algorithms [13]. In particular, in [7], greedy algorithms based on the above conjecture were shown to achieve significantly better performances in practice than widely-used network analysis heuristics.

General Threshold Model. In [7], KKT introduced the *general threshold model*, a broad generalization of a variety of natural diffusion models on networks, including the influential *linear threshold*

*Supported by an Alfred Sloan fellowship in Mathematics and by NSF grants DMS-0528488, and DMS-0548249 (CAREER) and by ONR grant N0014-07-1-05-06

†Supported by NSERC (Canada), FQRNT (Quebec, Canada) and by NSF grant DMS-0528488.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
STOC'07, June 11–13, 2007, San Diego, California, USA.
Copyright 2007 ACM 978-1-59593-631-8/07/0006 ...$5.00.

model of Granovetter in sociology [6]. Given an initial set of infected or active individuals on a network, the process grows in the following way. (See Section 1.1 for a formal description.) Each individual, say v, has an activation function which measures the effect of its neighbors on v and a threshold value. At any time, if the set of previously infected neighbors of v is such that its activation function crosses its threshold value, then v becomes infected. This process is *progressive*—an active node stays active forever. KKT consider the following natural assumptions:

- The *threshold values* are *random*. This is to account for our lack of knowledge of the exact threshold value of each individual.
- The *activation functions* are *monotone increasing*. This corresponds to the intuition that a node is more likely to become infected if a larger set of its neighbors is infected.
- The *activation functions* are *submodular*. This corresponds to the fact that the marginal effect of each neighbor of v decreases as the set of active nodes increases.

The Influence Maximization Problem. Since the diffusion process defined above is increasing, it terminates after a finite number of steps. For a given initial set of active nodes S we define $\sigma(S)$ to be the expected size of the set of active nodes at the end of the process. In the *Influence Maximization Problem*, we aim to find a set S of a fixed size maximizing $\sigma(S)$.

The Influence Maximization Problem is a natural problem to consider in the context of viral marketing. Given a social network, it is desired to find a small set of "target" individuals so as to maximize the number of customers who will eventually purchase a product following the effects of "word-of-mouth" [2, 3]. The same problem may also be of interest in epidemiology where finding the set S of a fixed size maximizing $\sigma(S)$ is a natural problem both in terms of bounding the spread of a disease and in terms of maximizing the effect of immunization.

In [7] it was shown that the Influence Maximization Problem is NP-hard to approximate within a factor $1 - 1/e + \varepsilon$ for all $\varepsilon > 0$. (The problem is in fact $n^{1-\varepsilon}$ hard to approximate without the submodularity condition.) On the other hand, it was shown in [8] that for all $\varepsilon > 0$ it is possible find a set S of fixed size that is a $(1 - 1/e - \varepsilon)$-approximation of the maximum in random polynomial time if the set function σ is *itself* submodular, which leads to the following conjecture.

CONJECTURE 1 ([7, 8]). *The function σ is submodular.*

While the result of [7, 8] showed that σ is submodular in special cases and related models (see below), the general case was open prior to our work, as highlighted in a recent invited talk of J. Kleinberg at FOCS 2006. In this paper we prove Conjecture 1 and extend it to the case where $\sigma(S)$ is the expected value of any monotone, submodular function of the final active set. This gives a $(1 - 1/e - \varepsilon)$-approximation algorithm for finding a set S of fixed size maximizing $\sigma(S)$.

1.1 The Model

In this section, we define formally the *general threshold model*.

DEFINITION 1 (SOCIAL NETWORK). *A* social network *is given by:*

- *A* ground set V *with* $|V| = n$
- *A collection of* activation functions $\mathcal{F} = (f_v)_{v \in V}$, *where* $f_v : 2^V \to [0, 1]$ *is a* $[0, 1]$*-valued set function on* V.

Typically, we think of V as the individuals of a social network $G = (V, E)$ where each f_v measures the effect of v's neighbors $N(v)$ on v. In particular f_v depends only on neighbors $N(v)$ affecting v, so $f_v(S) = f_v(N(v) \cap S)$ for all S. However, the specification of the graph will not be needed below.

DEFINITION 2 (MONOTONICITY). *The function $f : 2^V \to \mathbb{R}$ is* monotone *if*

$$f_v(S) \leq f_v(T)$$

for all

$$S \subseteq T \subseteq V.$$

DEFINITION 3 (SUBMODULARITY). *The function $f : 2^V \to \mathbb{R}$ is* submodular *if for all $S, T \subseteq V$*

$$f(S) + f(T) \geq f(S \cap T) + f(S \cup T).$$

The monotonicity condition corresponds to the fact that the effect of a larger set on v is stronger than the effect of a smaller set. The submodularity condition is equivalent to the fact that if $S \subseteq T$ and $v \in V$ then:

$$f(T \cup \{v\}) - f(T) \leq f(S \cup \{v\}) - f(S),$$

so the effect of each individual is decreasing when the set increases.

ASSUMPTION 1. *Throughout, we assume that $f_v(\emptyset) = 0$ and that f_v is monotone and submodular for all $v \in V$.*

We will consider the following diffusion process.

DEFINITION 4 (DIFFUSION). *For a given $\mathcal{F}$, consider the following process $\mathbf{S} = (S_t)_{t=0}^{n-1}$ started at $S \subseteq V$:*

1. *Associate to each node v an independent random variable θ_v uniform in $[0, 1]$;*
2. *Set $S_0 = S$;*
3. *At time $t \geq 1$, initialize $S_t = S_{t-1}$ and add to S_t the set of nodes in $V \setminus S_{t-1}$ such that $f_v(S_{t-1}) \geq \theta_v$.*

Clearly the process stops on or before time $n - 1$. We denote by $\mathcal{Q}_{\mathcal{F}}(S)$ the distribution of $\mathbf{S}$ when started at S and write $\mathbf{S} \sim \mathcal{Q}_{\mathcal{F}}(S)$, where we will drop the subscript when $\mathcal{F}$ is clear from the context.

DEFINITION 5 (INFLUENCE). *For a weight function $w : 2^V \to \mathbb{R}_+$, we define the influence $\sigma_w(S)$ of $S \subseteq V$ as*

$$\sigma_w(S) = \mathbb{E}_S[w(S_{n-1})],$$

where $\mathbb{E}_S$ is the expectation under $\mathcal{Q}_{\mathcal{F}}(S)$.

1.2 Previous Results

Conjecture 1 was previously verified in several special cases and related models.

Linear Threshold Model [7]. This is the general threshold model with f_v of the form

$$f_v(S) = \sum_{w \in S} b_{v,w},$$

for nonnegative constants $b_{v,w}$. The proof uses a representation in terms of a related percolation model. See [7] for details.

"Normalized" Submodular Threshold Model [8]. This is the general threshold model with f_v satisfying the so-called "normalized" submodularity property:

$$\frac{f_v(S \cup \{i\}) - f_v(S)}{1 - f_v(S)} \geq \frac{f_v(T \cup \{i\}) - f_v(T)}{1 - f_v(T)}, \qquad (1)$$

for all $S \subseteq T$. Note that this is stronger than submodularity. The proof takes advantage of an equivalence with the *decreasing cascade model* (see below).

Independent Cascade Model [7]. This is a related model where each edge (v, w) has an associated probability $p_{v,w}$ of being *live*, independently of all other edges. Infected nodes are those connected to the initial set through a *live path*. The proof of Conjecture 1 in this case also uses a percolation argument.

Decreasing Cascade Model [8]. A natural generalization of the previous model consists in defining for each v, each neighbor w of v and each subset of neighbors S of v a success probability $p_v(w, S)$ which is the probability that node w will succeed in activating v given that nodes in S are active and have failed to activate v. Each node w gets only one chance to activate each of its neighbors. KKT impose a natural *order-independence* condition on the success probabilities, i.e. the overall success probability of activating v does not depend on the order in which the active neighbors of v try to activate it. This model—called the *general cascade model* in [7]—turns out to be equivalent to the *general threshold model* under the maps

$$p_v(w, S) = \frac{f_v(S \cup \{w\}) - f_v(S)}{1 - f_v(S)},$$

and

$$f_v(S) = 1 - \prod_{i=1}^{r} (1 - p_v(w_i, S_{i-1})),$$

where $S = \{w_1, \ldots, w_r\}$ and $S_i = \{w_1, \ldots, w_i\}$. When

$$p_v(w, S) \geq p_v(w, T) \qquad (2)$$

for all $S \subseteq T$ and all v, w, the model is called the *decreasing cascade model*.

It is easy to check that the decreasing cascade model is equivalent to (1) under the mapping above. The proof of the conjecture for the decreasing cascade model works by coupling the processes started at S and T with $S \subseteq T$ and then adding w in a second phase where condition (2) is used.

In [7], it also shown that these results carry over to the *non-progressive* case where θ_v is resampled independently at each time step and to *general marketing strategies* where one can use several marketing actions simultaneously. See [7] for details.

1.3 Main Result

THEOREM 1 (MAIN RESULT). *Consider the process defined in Definition 4 where $\mathcal{F}$ and w are monotone and submodular, then σ_w is monotone and submodular. In particular, this is true when w is the cardinality function.*

COROLLARY 1. *Consider the process defined in Definition 4 where $\mathcal{F}$ and w are monotone and submodular. Then there exists a (greedy) $(1 - 1/e - \varepsilon)$-approximation algorithm for maximizing $\sigma_w(S)$ among all sets S of size k [8]. In particular, this is true when w is the cardinality function.*

The corollary follows from Theorem 1 and Theorem 2 of [8]. KKT's Greedy Approximation Algorithm is a simple variant of the standard greedy algorithm where sampling is used to estimate σ_w.

Our proof. Similarly to [8], a natural idea is to run the process in stages. Here we use three phases: we first grow $A \cap B$, then $A \setminus B$, and finally $B \setminus A$. See Figure 1 for an illustration. The key difference is in the execution of the last phase. To do away with the "normalized" submodularity condition of [8], we use

- a careful combination of cascade and threshold models, which we call the need-to-know representation;
- and, more importantly, a novel "antisense" coupling technique based on the intuition that coupling the processes started at arbitrary sets A and B by using θ_v and $1-\theta_v$ respectively, in a way, "maximizes their union" (note that $1 - \theta_v$ is also uniform in $[0, 1]$); this has to be implemented carefully to also control the intersection; see Section 2 for details; see e.g. [11] for a general reference on the coupling method.

2. PROOF

Throughout we fix $\mathcal{F}$ and w monotone, submodular. We also fix two arbitrary sets $A, B \subseteq V$ and let $C = A \cap B$ and $D = A \cup B$. The idea of the proof is to couple the four processes

$$\begin{aligned} \mathbf{A} &= (A_t)_{t=0}^{n-1} \sim \mathcal{Q}(A), \\ \mathbf{B} &= (B_t)_{t=0}^{n-1} \sim \mathcal{Q}(B), \\ \mathbf{C} &= (C_t)_{t=0}^{n-1} \sim \mathcal{Q}(C), \\ \mathbf{D} &= (D_t)_{t=0}^{n-1} \sim \mathcal{Q}(D), \end{aligned}$$

in such a way that

$$C_{n-1} \subseteq A_{n-1} \cap B_{n-1}, \qquad (3)$$

and

$$D_{n-1} \subseteq A_{n-1} \cup B_{n-1}. \qquad (4)$$

Indeed, we then have the following lemma.

LEMMA 1. *Suppose there exists a coupling of $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$ satisfying (3), (4). Then*

$$\sigma_w(A) + \sigma_w(B) \geq \sigma_w(A \cap B) + \sigma_w(A \cup B). \qquad (5)$$

PROOF. Indeed, we have by monotonicity and submodularity

$$\begin{aligned} &w(A_{n-1}) + w(B_{n-1}) \\ &\quad \geq w(A_{n-1} \cap B_{n-1}) + w(A_{n-1} \cup B_{n-1}) \\ &\quad \geq w(C_{n-1}) + w(D_{n-1}), \end{aligned} \qquad (6)$$

and therefore, taking expectation we get (5). □

Our coupling is based on the following ideas:

- **Antisense coupling.** The obvious coupling is to use the same θ_v's for all processes. It is easy to see that such a coupling does not satisfy (4). It does however satisfy (3). Intuitively, using the same θ_v for $\mathbf{A}$ and $\mathbf{B}$ "maximizes their intersection" while using θ_v for $\mathbf{A}$ and $(1 - \theta_v)$ for $\mathbf{B}$ "maximizes their union." We call this last coupling, the *antisense coupling*. To dominate both the intersection and the union simultaneously, we combine these two couplings.
- **Piecemeal growth.** The growth of the four processes can be divided in several stages where we add the initial sets progressively. Roughly, the coupling below starts by growing $A \cap B$, then $A \setminus B$ and finally $B \setminus A$. Following our previous comment, the last phase uses the antisense coupling to allow the process $\mathbf{B}$ to dominate $\mathbf{D}$ in that phase.

- **Need-to-know representation.** Finally, to help carry out the previous remarks, we note that it is not necessary to pick the θ_v's at the beginning of the process. Instead, at each step, we uncover as little information as possible about θ_v. This is related to the cascade model of [8] although here we use an explicit combination of cascade and threshold models.

We explain these ideas next. The proof of Theorem 1 follows in Section 2.3.

2.1 Piecemeal growth

We first describe an equivalent representation of the process where the initial set is added in stages. We denote by $\mathcal{Q}(S \mid \theta)$ the process $\mathcal{Q}(S)$ conditioned on $\theta = (\theta_v)_{v \in V}$. For a partition $S^{(1)}, \ldots, S^{(K)}$ of S (we allow some of the $S^{(k)}$'s to be empty), consider the process

$$\mathbf{T} = (T_t)_{t=0}^{Kn-1} \sim \mathcal{Q}(S^{(1)}, \ldots, S^{(K)}),$$

where

1. For each $v \in V$ pick θ_v uniformly in $[0, 1]$ and set $T_{-1} = \emptyset$;
2. For $1 \le k \le K$, we set
$$(T_t)_{t=(k-1)n}^{kn-1} \sim \mathcal{Q}(T_{(k-1)n-1} \cup S^{(k)} \mid \theta);$$
in other words, we add the $S^{(k)}$'s one at a time and use the same θ_v's for all stages.

It is easy to see that the processes $\mathcal{Q}(S)$ and $\mathcal{Q}(S^{(1)}, \ldots, S^{(K)})$ have the same distribution. This result actually follows from a more general discussion in [8], but we give a proof here for completeness.

LEMMA 2 (PIECEMEAL GROWTH). *Let $S^{(1)}, \ldots, S^{(K)}$ be a partition of $S \subseteq V$. Let*

$$\mathbf{S} = (S_t)_{t=0}^{n-1} \sim \mathcal{Q}(S),$$

and

$$\mathbf{T} = (T_t)_{t=0}^{Kn-1} \sim \mathcal{Q}(S^{(1)}, \ldots, S^{(K)}).$$

Then S_{n-1} and T_{Kn-1} have the same distribution.

PROOF. Pick θ_v uniformly in $[0, 1]$ for each $v \in V$ and let

$$\mathbf{S} = (S_t)_{t=0}^{n-1} \sim \mathcal{Q}(S \mid \theta),$$

and

$$\mathbf{T} = (T_t)_{t=0}^{Kn-1} \sim \mathcal{Q}(S^{(1)}, \ldots, S^{(K)} \mid \theta).$$

Moreover, let

$$\mathbf{T}' = (T'_t)_{t=0}^{Kn-1} \sim \mathcal{Q}(S, \emptyset, \ldots, \emptyset \mid \theta),$$

and

$$\mathbf{T}'' = (T''_t)_{t=0}^{Kn-1} \sim \mathcal{Q}(\emptyset, \ldots, \emptyset, S \mid \theta).$$

By monotonicity and induction on the K stages,

$$T''_{Kn-1} \subseteq T_{Kn-1} \subseteq T'_{Kn-1}$$

But clearly

$$T'_{Kn-1} = T''_{Kn-1} = S_{n-1}$$

so that $S_{n-1} = T_{Kn-1}$. □

2.2 Antisense phase and need-to-know representation

To implement the antisense coupling, we define the following variant of the process.

DEFINITION 6. *Let $S^{(1)}, \ldots, S^{(K)}$ be a partition of S and let $T \subseteq V \setminus S$. We define the process*

$$\mathbf{T} = (T_t)_{t=0}^{(K+1)n-1} \sim \mathcal{Q}_-(S^{(1)}, \ldots, S^{(K)}; T),$$

where

1. *For each $v \in V$ pick θ_v uniformly in $[0, 1]$;*
2. *Let $\mathbf{T} = (T_t)_{t=0}^{Kn-1} \sim \mathcal{Q}(S^{(1)}, \ldots, S^{(K)} \mid \theta)$;*
3. *Set $T_{Kn} = T_{Kn-1} \cup T$;*
4. *At time $Kn + 1 \le t \le (K + 1)n - 1$, initialize $T_t = T_{t-1}$, and add to T_t the set of nodes in $V \setminus T_{t-1}$ such that*
$$f_v(T_{t-1}) - f_v(T_{Kn-1}) \ge 1 - \theta_v.$$

LEMMA 3 (ANTISENSE PHASE). *Assume $S^{(1)}, \ldots, S^{(K)}$ is a partition of S and $T \subseteq V \setminus S$. Let*

$$\mathbf{S} = (S_t)_{t=0}^{(K+1)n-1} \sim \mathcal{Q}(S^{(1)}, \ldots, S^{(K)}, T),$$

and

$$\mathbf{T} = (T_t)_{t=0}^{(K+1)n-1} \sim \mathcal{Q}_-(S^{(1)}, \ldots, S^{(K)}; T).$$

Then, $S_{(K+1)n-1}$ and $T_{(K+1)n-1}$ have the same distribution.

PROOF. As was discussed at the beginning of Section 2, rather than picking the θ_v's at the beginning of the process, it is useful to think of them as being progressively uncovered on a need-to-know basis. Consider only the *first* stage of the process $\mathbf{S}$ for the time being. Let $S_{-1} = \emptyset$. Suppose that, at time $t \ge 1$, $v \notin S_{t-1}$. Then we have that $\theta_v \in [f_v(S_{t-2}), 1]$ and all we need to know to decide if v is added to S_t is whether or not $\theta_v \in [f_v(S_{t-2}), f_v(S_{t-1})]$. In other words, was the increase in f_v between time $t - 2$ and $t - 1$ enough to hit θ_v? Note that, given the event $\{f_v(S_{t-2}) \le \theta_v\}$, θ_v is uniformly distributed in $[f_v(S_{t-2}), 1]$ and we have that θ_v is in $[f_v(S_{t-2}), f_v(S_{t-1})]$ with probability

$$\frac{f_v(S_{t-1}) - f_v(S_{t-2})}{1 - f_v(S_{t-2})}.$$

Therefore, we can describe the process $(S_t)_{t=0}^{n-1}$ equivalently as follows. We first set $S_{-1} = \emptyset$, $S_0 = S$. Then, at step $1 \le t \le n - 1$, we initialize $S_t = S_{t-1}$ and for each $v \in V \setminus S_{t-1}$:

- With probability
$$\frac{f_v(S_{t-1}) - f_v(S_{t-2})}{1 - f_v(S_{t-2})}, \tag{7}$$
we add v to S_t and pick θ_v uniformly in
$$[f_v(S_{t-2}), f_v(S_{t-1})];$$
- Otherwise, we do nothing.

By the discussion above, this new version of the process has the same distribution as $\mathcal{Q}(S^{(1)})$. We proceed similarly for the following $K - 1$ stages to get $(S_t)_{t=0}^{Kn-1}$ which is then distributed according to $\mathcal{Q}(S^{(1)}, \ldots, S^{(K)})$.

We can clearly choose

$$(T_t)_{t=0}^{Kn-1} = (S_t)_{t=0}^{Kn-1}.$$

Then note that, at time $t = Kn$, for each $v \notin S_{Kn-1} = T_{Kn-1}$, we have that θ_v is uniformly distributed in

$$[f_v(S_{Kn-1}), 1] = [f_v(T_{Kn-1}), 1].$$

For each such v, we now pick θ_v uniformly in $[f_v(S_{Kn-1}), 1]$ and set

$$\theta'_v = \begin{cases} \theta_v, & v \in S_{Kn-1}, \\ f_v(S_{Kn-1}) + 1 - \theta_v, & v \notin S_{Kn-1}. \end{cases}$$

Finally, let

$$(S_t)_{t=Kn}^{(K+1)n-1} \sim \mathcal{Q}(S_{Kn-1} \cup T \,|\, \theta),$$

and

$$(T_t)_{t=Kn}^{(K+1)n-1} \sim \mathcal{Q}(T_{Kn-1} \cup T \,|\, \theta').$$

That is, we run the last stage of $\mathbf{S}$ and $\mathbf{T}$ as before, with θ and θ' respectively. It is clear that $\mathbf{T} \sim \mathcal{Q}_-(S^{(1)}, \ldots, S^{(K)}; T)$ by construction. Moreover, it follows easily that $S_{(K+1)n-1}$ and $T_{(K+1)n-1}$ have the same distribution from the fact that for a uniform variable θ_v in $[f_v(S_{Kn-1}), 1]$, the random variables θ_v and $f_v(S_{Kn-1}) + 1 - \theta_v$ have the same distribution. □

2.3 Coupling

We are now ready to prove Theorem 1. See Figure 1 for a graphical representation of the proof. We will need the following easy consequence of monotone submodularity.

LEMMA 4. *Let $f : 2^V \to \mathbb{R}_+$ be monotone and submodular. Then if $S \subseteq S' \subseteq V$ and $T \subseteq T' \subseteq V$, we have*

$$f(S \cup T') - f(S) \geq f(S' \cup T) - f(S').$$

PROOF. Note that by monotonicity and submodularity

$$\begin{aligned} f(S \cup T') - f(S) &\geq f(S \cup T) - f(S), \\ &= f(S \cup (T \setminus S)) - f(S), \\ &\geq f(S \cup (S' \setminus (T \cup S)) \cup (T \setminus S)) \\ &\qquad - f(S \cup (S' \setminus (T \cup S))), \\ &\geq f(S' \cup T) - f(S'). \end{aligned}$$

□

PROOF. We proceed with our coupling of $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$, and $\mathbf{D}$. In fact, by Lemmas 1, 2, and 3, it suffices instead to couple

$$\begin{aligned} \mathbf{A} &= (A_t)_{t=0}^{3n-1} \sim \mathcal{Q}(A \cap B, A \setminus B, \emptyset), \\ \mathbf{B} &= (B_t)_{t=0}^{3n-1} \sim \mathcal{Q}_-(A \cap B, \emptyset; A \setminus B), \\ \mathbf{C} &= (C_t)_{t=0}^{3n-1} \sim \mathcal{Q}(A \cap B, \emptyset, \emptyset), \\ \mathbf{D} &= (D_t)_{t=0}^{3n-1} \sim \mathcal{Q}_-(A \cap B, A \setminus B; B \setminus A), \end{aligned}$$

in such a way that for all $0 \leq t \leq 3n-1$

$$C_t \subseteq A_t \cap B_t, \qquad D_t \subseteq A_t \cup B_t. \tag{8}$$

Our coupling is as follows. We pick θ_v uniformly in $[0, 1]$ for all $v \in V$ and use the same θ for all four processes.

By construction, for all $0 \leq t \leq 2n-1$ we have

$$B_t = C_t \subseteq A_t$$

so that

$$C_t = A_t \cap B_t.$$

Similarly for all $0 \leq t \leq 2n-1$ we have $D_t = A_t$ so that

$$D_t \subseteq A_t \cup B_t.$$

Thus (8) is satisfied for $0 \leq t \leq 2n-1$. To see (8) holds also for $2n \leq t \leq 3n-1$, note that by Lemma 4 for all $v \notin D_{2n}$

$$f_v(B_{2n}) - f_v(B_{2n-1}) \geq f_v(D_{2n}) - f_v(D_{2n-1}),$$

since $B_{2n-1} \subseteq D_{2n-1}$,

$$B_{2n} = B_{2n-1} \cup (B \setminus A),$$

and

$$D_{2n} = D_{2n-1} \cup (B \setminus A).$$

We proceed by induction. By monotonicity and Lemma 4, we then have for all $2n \leq t \leq 3n-1$

$$D_t \setminus D_{2n-1} \subseteq B_t \setminus B_{2n-1},$$

and

$$f_v(B_t) - f_v(B_{2n-1}) \geq f_v(D_t) - f_v(D_{2n-1}),$$

$\forall\, v \notin D_{2n}$. This proves the claim since we then have for all $2n \leq t \leq 3n-1$, $A_t = D_{2n-1}$ and

$$D_t \setminus D_{2n-1} \subseteq B_t$$

which implies

$$D_t \subseteq A_t \cup B_t.$$

The condition

$$C_t \subseteq A_t \cap B_t$$

is clear from the construction. □

3. CONCLUDING REMARKS

Necessity. It is easy to see that the submodularity assumption in Theorem 1 is necessary in the following sense: Any function f which is not submodular admits a network with activation function f where the influence is not submodular. Indeed, let $f : 2^V \to \mathbb{R}_+$, $A, B \subseteq V$ such that

$$f(A) + f(B) < f(A \cap B) + f(A \cup B).$$

Let $V^* = V \cup \{v^*\}$ with $f_{v^*} \equiv f$ and $f_v \equiv 1$ for all $v \in V$. It is then immediate to check that:

$$\begin{aligned} \sigma(A) + \sigma(B) &= |A| + |B| + f(A) + f(B) \\ &= |A \cap B| + |A \cup B| \\ &\qquad + f(A) + f(B) \\ &< |A \cap B| + |A \cup B| \\ &\qquad + f(A \cap B) + f(A \cup B) \\ &= \sigma(A \cap B) + \sigma(A \cup B). \end{aligned}$$

Acknowledgments

We thank Jon Kleinberg and Éva Tardos for insightful discussions and encouragements.

4. REFERENCES

[1] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the internet. In *SODA*, pages 301–310, 2005.

[2] P. Domingo and M. Richardson. Minning the network value of customers. In *Proceedings of the 7th Intl. Conference on Knowledge Discovery and Data Minning*, pages 57–66, 2001.

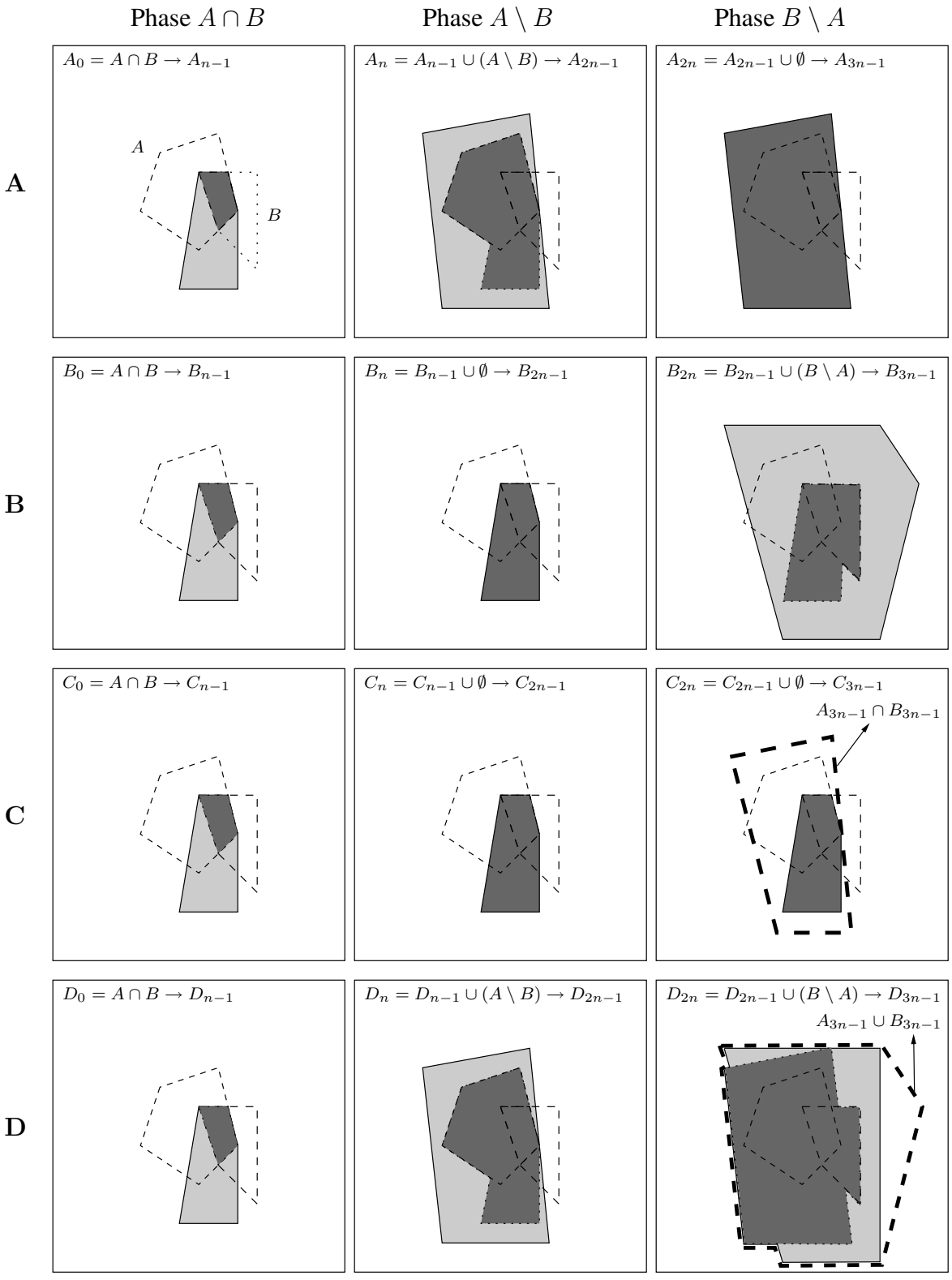

Figure 1: The three phases of the coupling. In each phase, the dark shaded region is the initial set, while the light shaded region is the final set. The sets A and B are indicated by dashed lines. The thick dashed lines show that the desired properties are satisfied.

[3] P. Domingo and M. Richardson. Minning knowledge-sharing sites for viral marketing. In *Proceedings of the 8th Intl. Conference on Knowledge Discovery and Data Minning*, pages 61–70, 2002.

[4] Rick Durrett. Lecture notes on particle systems and percolation. 1988.

[5] Rick Durrett and Paul Jung. Two phase transitions for the contact process on small worlds. Submitted, 2006.

[6] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.

[7] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *Proc. 9th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*, 2003.

[8] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Proc. 32nd International Colloquium on Automata, Languages and Programming (ICALP)*, 2005.

[9] Thomas M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.

[10] Thomas M. Liggett. *Stochastic interacting systems: contact, voter and exclusion processes*, volume 324 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, 1999.

[11] T. Lindvall. *Lectures on the Coupling Method*. Wiley, New York, 1992.

[12] M. Morris, editor. *Network Epidemiology*. Oxford University Press, 2004.

[13] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization*. John Wiley, 1988.

[14] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, 1994.