



University of Pennsylvania  
ScholarlyCommons

---

Statistics Papers

Wharton Faculty Research

---

2007

# Submodularity of Influence in Social Networks: From Local to Global

Elchanan Mossel  
*University of Pennsylvania*

Sébastien Roch

Follow this and additional works at: [http://repository.upenn.edu/statistics\\_papers](http://repository.upenn.edu/statistics_papers)

 Part of the [Statistics and Probability Commons](#), and the [Theory and Algorithms Commons](#)

---

## Recommended Citation

Mossel, E., & Roch, S. (2007). Submodularity of Influence in Social Networks: From Local to Global. *Proceedings of the Thirty-ninth Annual ACM Symposium on Theory of Computing, STOC '07* 128-134. <http://dx.doi.org/10.1145/1250790.1250811>

At the time of publication, author Elchanan Mossel was affiliated with University of California, Berkeley. Currently, he is a faculty member at the Statistics Department at the University of Pennsylvania.

This paper is posted at ScholarlyCommons. [http://repository.upenn.edu/statistics\\_papers/192](http://repository.upenn.edu/statistics_papers/192)  
For more information, please contact [repository@pobox.upenn.edu](mailto:repository@pobox.upenn.edu).

---

# Submodularity of Influence in Social Networks: From Local to Global

## Abstract

Social networks are often represented as directed graphs 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 Kempe, Kleinberg, and Tardos (KKT) in [KKT03, KKT05] where the authors also impose several natural assumptions: the threshold values are random and the activation functions are monotone and submodular. 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. Roughly, our results demonstrate that “local” submodularity is preserved “globally” under this diffusion process. This is of natural computational interest, as many optimization problems have good approximation algorithms for submodular functions.

## Keywords

coupling, social networks, submodularity, viral marketing

## Disciplines

Statistics and Probability | Theory and Algorithms

## Comments

At the time of publication, author Elchanan Mossel was affiliated with University of California, Berkeley. Currently, he is a faculty member at the Statistics Department at the University of Pennsylvania.

# Submodularity of Influence in Social Networks: From Local to Global \*

Elchanan Mossel<sup>†</sup> and Sebastien Roch<sup>‡</sup>

July 28, 2009

## Abstract

Social networks are often represented as directed graphs 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 Kempe, Kleinberg, and Tardos (KKT) in [KKT03, KKT05] where the authors also impose several natural assumptions: the threshold values are random and the activation functions are monotone and submodular. The monotonicity condition indicates that a node is more likely to become active if more of its neighbors are active,

---

\*Keywords: growth process, coupling method, submodularity, social networks, viral marketing. MSC classification: 60K35, 91D30, 68Q25. E.M. is 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.

<sup>†</sup>Statistics and Computer Science, UC Berkeley and Mathematics and Computer Science, Weizmann Institute.

<sup>‡</sup>Mathematics, UCLA. Work done at UC Berkeley.

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. Roughly, our results demonstrate that “local” submodularity is preserved “globally” under this diffusion process. This is of natural computational interest, as many optimization problems have good approximation algorithms for submodular functions.

## 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 [Mor04, Lig85, Lig99, Dur88, BBCS05, DJ06], much of the recent interest has resulted from applications in sociology, economics, and engineering [BR87, ARLV01, GLM01b, GLM01a, DR01, RD02, KKT03, KKT05].

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 [DR01, RD02, KKT03, KKT05]. Roughly speaking, *viral marketing*—unlike conventional mar-

keting—takes into account the “network value” of potential customers, that is, it seeks to target a set of individuals whose influence on their social network through word-of-mouth effects is high.

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 [WF94]. Typically these heuristics provide no guarantees. Here we prove a structural conjecture of Kempe, Kleinberg, and Tardos (KKT) [KKT03, KKT05] which in turn implies that a natural greedy algorithm has a good performance guarantee. The conjecture can be roughly stated as follows: if a diffusion model is locally submodular, that is, the influence on an individual by its neighbors in the network has “diminishing returns,” then the process is globally submodular (on average). This is relevant in this context because, under the submodularity property, optimization problems—such as the viral marketing problem—are known to have good approximation algorithms [NW88]. In particular, in [KKT03], 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 [KKT03], KKT introduced the *general threshold model*, a broad generalization of a variety of natural diffusion models on networks, including the influential *linear threshold model* of Granovetter in sociology [Gra78]. 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  $v$ , as well as 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 values. KKT actually assume that the thresholds are *uniformly* random. Note however that, given that any distribution can be generated from a uniform random variable, it follows that, by appropriately modifying the activation functions, the threshold values can *effectively* have any distribution. (See Section 3.)
- The *activation functions* are *monotone increasing*. That is, 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” [DR01, RD02]. 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 [KKT03] it was shown that the Influence Maximization Problem is NP-hard to approximate within a factor  $1 - 1/e + \varepsilon$  for all  $\varepsilon > 0$  and that the problem is in fact  $n^{1-\varepsilon}$  hard to approximate without the submodularity condition. (See, e.g., [GJ79, ACG<sup>+</sup>99] for background on NP-completeness and hardness of approximation.) On the other hand, it was shown in [KKT05] that for all  $\varepsilon > 0$  it is possible to 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  $\sigma$  is *itself* submodular. This leads to the following conjecture.

**Conjecture 1 ([KKT03, KKT05])** *The function  $\sigma$  is submodular.*

While the result of [KKT03, KKT05] showed that  $\sigma$  is submodular in special cases and related models (see below), the general case was open prior to our work. 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  $f_v : 2^V \rightarrow [0, 1]$  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

$f : 2^V \rightarrow \mathbb{R}$  is monotone if

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

for all

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

We say that a collection  $\mathcal{F}$  of functions from  $2^V$  to  $\mathbb{R}$  is monotone if all its elements are monotone functions.



**Definition 3 (Submodularity)** *The function*

$f : 2^V \rightarrow \mathbb{R}$  *is submodular if for all*  $S, T \subseteq V$

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

*We say that a collection  $\mathcal{F}$  of functions from  $2^V$  to  $\mathbb{R}$  is submodular if all its elements are submodular functions.*

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 process, which is often referred to as a “diffusion” process in the sociology literature.

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

1. *Associate to each node  $v$  an independent random variable  $\theta_v$  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  $w : 2^V \rightarrow \mathbb{R}_+$ , 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 [KKT03].** 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 [KKT03] for details.

**“Normalized” Submodular Threshold Model [KKT05].** This is the general threshold model with  $f_v$  satisfying the so-called “normalized” submodularity

property:

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

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 [KKT03].** 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 [KKT05].** 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, that is, 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 [KKT03]—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  $S = \{w_1, \dots, w_r\}$  and  $S_i = \{w_1, \dots, w_i\}$ . When

$$p_v(w, S) \geq p_v(w, T) \tag{2}$$

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 [KKT03], it is also shown that these results carry over to the *non-progressive* case where  $\theta_v$  is resampled independently at each time step and to *general marketing strategies* where one can use several marketing actions simultaneously. We refer the reader to [KKT03] for details.

**Contact Process.** A result similar to our Theorem 1 below holds for a related *contact process* [Har74, Lig85, LSS07], where the infection rates are monotone, concave functions of the number of infected neighbours (a special case of monotone, submodular functions) and where vertices also heal at a constant rate. In particular, Harris [Har74] uses a coupling argument, although it appears that Harris' proof does not extend easily to our setting as it uses the special form of the infection rates.

**Subsequent Work.** After the publication of this work in its extended abstract

form [MR07], further work was devoted to viral marketing, particularly in the competitive framework where firms compete for customers [DGM06, CNWvZ07, BKS07, EDS07].

### 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,  $\sigma_w$  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. Furthermore, assume that  $w$  takes values between 1 and  $\text{poly}(n)$ . Then, there exists a (greedy)  $(1 - 1/e - 1/\text{poly}(n))$ -approximation algorithm for maximizing  $\sigma_w(S)$  among all sets  $S$  of size  $k$  [KKT05] in time  $\text{poly}(n)$ . In particular, this is true when  $w$  is the cardinality function.*

The corollary follows from Theorem 1 and Theorem 2 of [KKT05]. KKT’s Greedy Approximation Algorithm is a simple variant of the standard greedy algorithm where sampling is used to estimate  $\sigma_w$ .

**Our proof.** Similarly to [KKT05], 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 [KKT05], we do the following.

- We use a careful combination of cascade and threshold models, which we call the need-to-know representation.
- More importantly, we introduce a novel “antisense” coupling technique based on the intuition that coupling the processes started at arbitrary sets  $A$  and  $B$  by using  $\theta_v$  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., [Lin92] for a general reference on the coupling method.

The rest of the paper is organized as follows. We begin with a few preliminary remarks in Section 2.1. The need-to-know representation and the antisense process are introduced in Section 2.2. The full coupling and the proof of its correctness appear in Section 2.3.

The results of this paper were announced in the form of an extended abstract in [MR07].

## 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

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

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

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

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

in such a way that

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

and

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

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*

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

**Proof:** We have by monotonicity and submodularity of  $w$

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

and therefore, taking expectations we get (5). ■

Our coupling is based on the following ideas:

- **Antisense coupling.** The obvious coupling is to use the same  $\theta_v$ '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  $\theta_v$  for **A** and **B** “maximizes their intersection” while using  $\theta_v$  for **A** and  $(1 - \theta_v)$  for **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 **B** to dominate **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  $\theta_v$ 's at the beginning of the process. Instead, at each step, we uncover as little information as possible about  $\theta_v$ . This is related to the cascade model of [KKT05] 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  $\mathcal{Q}(S | \theta)$  the process  $\mathcal{Q}(S)$  conditioned on  $\theta =$



$(\theta_v)_{v \in V}$ . For a partition  $S^{(1)}, \dots, S^{(K)}$  of  $S$  (we allow some of the  $S^{(k)}$ 's to be empty), consider the process

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

where

1. For each  $v \in V$  pick  $\theta_v$  uniformly and independently in  $[0, 1]$  and set  $T_{-1} = \emptyset$ ;
2. For  $1 \leq k \leq K$ , 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  $\theta_v$ 's for all stages.

The outcomes of the processes  $\mathcal{Q}(S)$  and  $\mathcal{Q}(S^{(1)}, \dots, S^{(K)})$  have the same distribution. This result actually follows from a more general discussion in [KKT05], but we give a proof here for completeness.

**Lemma 2 (Piecemeal Growth)** *Let  $S^{(1)}, \dots, S^{(K)}$  be a partition of  $S \subseteq V$ . Let*

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

and

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

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

**Proof:** Pick  $\theta_v$  uniformly and independently 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

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

Moreover, let

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

and

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

By monotonicity and induction on the  $K$  stages,

$$T''_{K_{n-1}} \subseteq T_{K_{n-1}} \subseteq T'_{K_{n-1}}$$

But clearly

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

so that  $S_{n-1} = T_{K_{n-1}}$ . ■

## 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  $S^{(1)}, \dots, S^{(K)}$  be a partition of  $S$  and let  $T \subseteq V \setminus S$ . We define the process

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

where

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

$$f_v(T_{t-1}) - f_v(T_{Kn-1}) \geq 1 - \theta_v.$$

**Lemma 3 (Antisense Phase)** Assume  $S^{(1)}, \dots, S^{(K)}$  is a partition of  $S$  and  $T \subseteq V \setminus S$ . Let

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

and

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

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  $\theta_v$ '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  $t \geq 1$ ,  $v \notin S_{t-1}$ . 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  $\theta_v$ ? Note that, given the event  $\{f_v(S_{t-2}) \leq \theta_v\}$ ,  $\theta_v$  is uniformly distributed in  $[f_v(S_{t-2}), 1]$  and we have that  $\theta_v$  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  $1 \leq t \leq n - 1$ , 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  $\theta_v$  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  $\mathcal{Q}(S^{(1)}, \dots, S^{(K)})$ .

Up to time  $Kn - 1$ , the processes  $(S_t)$  and  $(T_t)$  have identical transition probabilities. Hence, we can take

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

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

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

For each such  $v$ , we now pick  $\theta_v$  uniformly in  $[f_v(S_{Kn-1}), 1]$ . Moreover, we define for all  $v \in V$

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

Finally, let

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

and

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

That is, we run the last stage of **S** and **T** as before, with  $\theta$  and  $\theta'$  respectively.

It is clear that  $\mathbf{T} \sim \mathcal{Q}_-(S^{(1)}, \dots, S^{(K)}; T)$  by construction. Moreover, it follows

that  $S_{(K+1)n-1}$  and  $T_{(K+1)n-1}$  have the same distribution from the fact that for a uniform variable  $\theta_v$  in  $[f_v(S_{K_{n-1}}), 1]$ , the random variables  $\theta_v$  and  $f_v(S_{K_{n-1}}) + 1 - \theta_v$  have the same distribution. ■

### 2.3 Coupling

We are now ready to prove Theorem 1. We will need the following easy consequence of monotone submodularity.

**Lemma 4** *Let  $f : 2^V \rightarrow \mathbb{R}_+$  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

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

■

We now give a proof of Theorem 1.

**Proof:** We proceed with our coupling of **A**, **B**, **C**, and **D**. In fact, by Lemmas 1, 2, and 3, it suffices instead to couple

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

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

$$C_t \subseteq A_t \cap B_t, \quad D_t \subseteq A_t \cup B_t. \quad (8)$$

Our coupling is as follows. We pick  $\theta_v$  uniformly in  $[0, 1]$  for all  $v \in V$  and use the same  $\theta$  for all four processes above. See Figure 1 for a graphical representation of the proof.

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

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

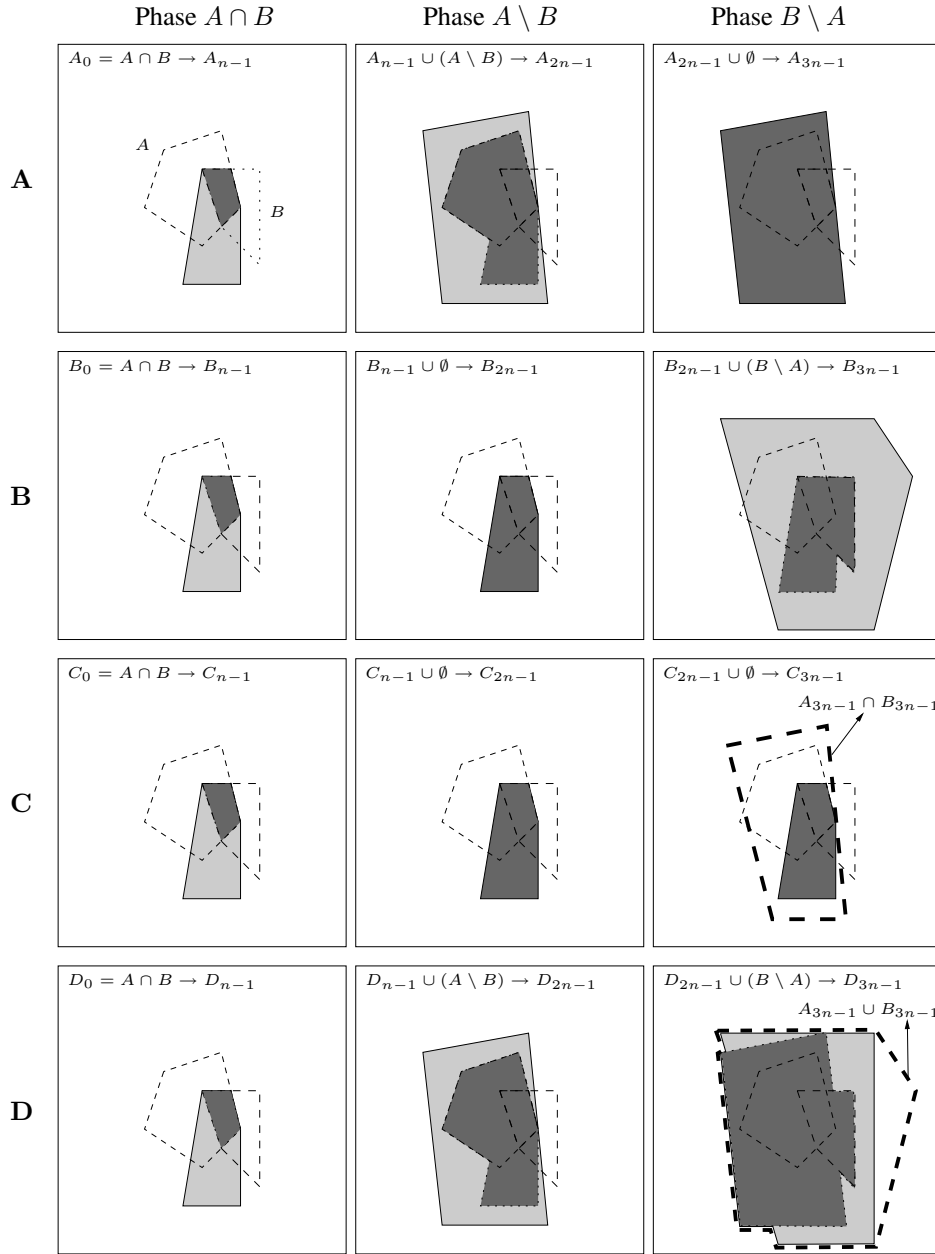


Figure 1: The three phases of the coupling. In each phase, the dark shaded region is the initial set, while the light shaded region is the final set. The sets  $A$  and  $B$  are indicated by dashed lines. The thick dashed lines show that the desired properties are satisfied.



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

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

(See Figure 2 for an illustration of this step.) We proceed by induction. By monotonicity and Lemma 4, we then have for all  $2n \leq t \leq 3n - 1$

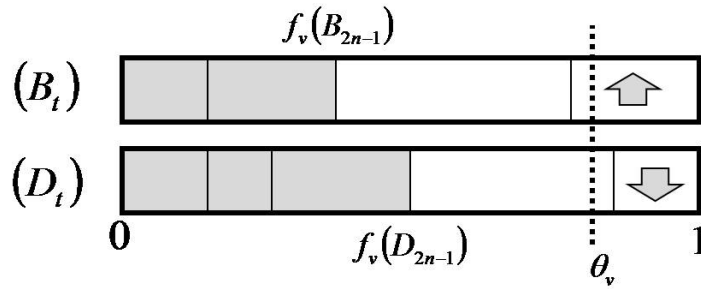
$$(\Omega_{1,t}) \quad D_t \setminus D_{2n-1} \subseteq B_t \setminus B_{2n-1},$$

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

for all  $v \notin D_{2n}$ . Indeed, assume  $(\Omega_{1,t'})$  and  $(\Omega_{2,t'})$  for all  $2n \leq t' \leq t$ . We have already proved the base case  $t = 2n$ . Then  $(\Omega_{2,t})$  implies  $(\Omega_{1,t+1})$  by definition of the process in the antisense phase (Definition 6). In turn,  $(\Omega_{1,t+1})$  implies  $(\Omega_{2,t+1})$  by monotonicity and Lemma 4.

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

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



$$\boxed{\uparrow f_v(B_{2n}) - f_v(B_{2n-1})} \geq \boxed{f_v(D_{2n}) - f_v(D_{2n-1}) \downarrow}$$

Figure 2: Last phase of the coupling seen from a node  $v$ . In the antisense phase, the interval is filled up “from the end.” By submodularity, the process  $(B_t)$  reaches the threshold first.

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  $f : 2^V \rightarrow \mathbb{R}_+$ ,  $A, B \subseteq V$  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:

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

**Other threshold distributions.** As noted in the introduction, our results hold for general threshold distributions. Assume that  $\theta_v$  has cumulative distribution  $F_v$ . Notice that the dynamics can be re-written as

$$f_v(S_{t-1}) \geq \theta_v \equiv F_v^{-1}(U_v) \quad \Leftrightarrow \quad F_v(f_v(S_{t-1})) \geq U_v,$$

where the  $U_v$ 's are independent uniform in  $[0, 1]$ . Hence, if  $F_v \circ f_v$  is increasing submodular for all  $v$  then the influence function is submodular. Furthermore, the example above demonstrates that this condition is necessary.

## Acknowledgments

We thank Jon Kleinberg and Éva Tardos for insightful discussions and encouragements. We also thank Tom Liggett for bringing to our attention the work of Harris.

## References

- [ACG<sup>+</sup>99] G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi. *Complexity and approximation*. Springer-Verlag, Berlin, 1999. Combinatorial optimization problems and their approximability properties, With 1 CD-ROM (Windows and UNIX).
- [ARLV01] C. Asavathiratham, S. Roy, B. Lesieutre, and G. Verghese. The influence model. *IEEE Control Systems*, 2001.
- [BBCS05] Noam Berger, Christian Borgs, Jennifer T. Chayes, and Amin Saberi. On the spread of viruses on the internet. In *SODA*, pages 301–310, 2005.
- [BKS07] Shishir Bharathi, David Kempe, and Mahyar Salek. Competitive influence maximization in social networks. In *WINE*, pages 306–311, 2007.

- [BR87] J. Brown and P. Reinegen. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14:350–362, 1987.
- [CNWvZ07] Tim Carnes, Chandrashekar Nagarajan, Stefan M. Wild, and Anke van Zuylen. Maximizing influence in a competitive social network: a follower’s perspective. In *ICEC ’07: Proceedings of the ninth international conference on Electronic commerce*, pages 351–360, New York, NY, USA, 2007. ACM.
- [DGM06] Pradeep Dubey, Rahul Garg, and Bernard De Meyer. Competing for customers in a social network: The quasi-linear case. In *WINE*, pages 162–173, 2006.
- [DJ06] Rick Durrett and Paul Jung. Two phase transitions for the contact process on small worlds. Submitted, 2006.
- [DR01] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the 7th Intl. Conference on Knowledge Discovery and Data Mining*, pages 57–66, 2001.
- [Dur88] Rick Durrett. Lecture notes on particle systems and percolation. 1988.
- [EDS07] Eyal Even-Dar and Asaf Shapira. A note on maximizing the spread of influence in social networks. In *WINE*, pages 281–286, 2007.
- [GJ79] Michael R. Garey and David S. Johnson. *Computers and intractability*. W. H. Freeman and Co., San Francisco, Calif., 1979.

A guide to the theory of NP-completeness, A Series of Books in the Mathematical Sciences.

- [GLM01a] 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*, 12:211–223, 2001.
- [GLM01b] J. Goldenberg, B. Libai, and E. Muller. Using complex systems analysis to advance marketing theory development: Modeling heterogeneity effects on new product growth through stochastic cellular automata. *Academy of Marketing Science Review*, 2001.
- [Gra78] M. Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978.
- [Har74] T. E. Harris. Contact interactions on a lattice. *Ann. Probability*, 2:969–988, 1974.
- [KKT03] 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.
- [KKT05] 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.
- [Lig85] Thomas M. Liggett. *Interacting particle systems*, volume 276 of *Grundlehren der Mathematischen Wissenschaften [Fundamental*

- Principles of Mathematical Sciences*]. Springer-Verlag, New York, 1985.
- [Lig99] 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.
- [Lin92] T. Lindvall. *Lectures on the Coupling Method*. Wiley, New York, 1992.
- [LSS07] T. M. Liggett, R. B. Schinazi, and J. Schweinsberg. A contact process with mutations on a tree. To appear in *Stochastic Process. Appl.*, 2007.
- [Mor04] M. Morris, editor. *Network Epidemiology*. Oxford University Press, 2004.
- [MR07] Elchanan Mossel and Sébastien Roch. On the submodularity of influence in social networks. In David S. Johnson and Uriel Feige, editors, *STOC*, pages 128–134. ACM, 2007.
- [NW88] G. Nemhauser and L. Wolsey. *Integer and Combinatorial Optimization*. John Wiley, 1988.
- [RD02] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the 8th Intl. Conference on Knowledge Discovery and Data Mining*, pages 61–70, 2002.

[WF94] S. Wasserman and K. Faust. *Social Network Analysis*. Cambridge University Press, 1994.