

Cascading Behavior in Networks: Algorithmic and Economic Issues ^{*}

Jon Kleinberg [†]

Abstract

The flow of information or influence through a large social network can be thought of as unfolding with the dynamics of an epidemic: as individuals become aware of new ideas, technologies, fads, rumors, or gossip, they have the potential to pass them on to their friends and colleagues, causing the resulting behavior to cascade through the network.

We consider a collection of probabilistic and game-theoretic models for such phenomena proposed in the mathematical social sciences, as well as recent algorithmic work on the problem by computer scientists. Building on this, we discuss the implications of cascading behavior in a number of on-line settings, including word-of-mouth effects (also known as “viral marketing”) in the success of new products, and the influence of social networks in the growth of on-line communities.

1 Introduction

The process by which new ideas and new behaviors spread through a population has long been a fundamental question in the social sciences. New religious beliefs or political movements; shifts in society that lead to greater tolerance or greater polarization; the adoption of new technological, medical, or agricultural innovations; the sudden success of a new product; the rise to prominence of a celebrity or political candidate; the emergence of bubbles in financial markets and their subsequent implosion — these phenomena all share some important qualitative properties. They tend to begin on a small scale with a few “early adopters”; more and more people begin to adopt them as they observe their friends, neighbors, or colleagues doing so; and the resulting new behaviors may eventually spread through the population contagiously, from person to person, with the dynamics of an epidemic.

People have long been aware of such processes at an anecdotal level; the systematic study of them developed, in the middle of the 20th century, into an area of sociology known as the *diffusion of innovations*. The initial research on this topic was empirical (see e.g. [CMK66, Rog95, SS98] for

^{*}Appears as Chapter 24 of *Algorithmic Game Theory*, edited by Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay Vazirani (Cambridge University Press, 2007).

[†]Department of Computer Science, Cornell University, Ithaca NY 14853. Email: kleinber@cs.cornell.edu. Supported in part by a John D. and Catherine T. MacArthur Foundation Fellowship, a Google Research Grant, a Yahoo! Research Alliance Grant, and NSF grants CCF-0325453, IIS-0329064, CNS-0403340, and BCS-0537606.

background), but in the 1970s economists and mathematical sociologists such as Thomas Schelling and Mark Granovetter [Sch78, Gra78] began formulating basic mathematical models for the mechanisms by which ideas and behaviors diffuse through a population. There are several reasons to seek models that capture observed data on diffusion: in addition to helping us understand, at a fundamental level, how the spread of new ideas “works,” such models have the potential to help us predict the success or failure of new innovations in their early stages, and potentially to shape the underlying process so as to increase (or reduce) the chances of success.

In this chapter, we discuss some of the basic models in this area, as well as suggesting some current applications to on-line information systems. While the overall topic is much too vast even to survey in a brief setting such as this, we hope to convey some of the game-theoretic and algorithmic grounding of the area, and to highlight some directions for future work. We also indicate some of the ways in which large-scale on-line communities provide rich data for observing social diffusion processes as they unfold, thus providing the opportunity to develop richer models. Further related work is discussed briefly in the Notes at the end of the chapter.

2 A First Model: Networked Coordination Games

One of the simplest models for social diffusion can be motivated by game-theoretic considerations. To set the stage for this, notice that many of the motivating scenarios considered above have the following general flavor: each individual v has certain friends, acquaintances, or colleagues, and the benefits to v of adopting the new behavior increase as more and more of these other individuals adopt it. In such a case, simple self-interest will dictate that v should adopt the new behavior once a sufficient fraction of v 's neighbors have done so. For example, many new technological, economic, or social practices become more valuable as the number of people using them increases: two organizations may find it easier to collaborate on a joint project if they are using compatible technologies; two people may find it easier to engage in social interaction — all else being equal — if their beliefs and opinions are similar.

Defining the game. Specifically, here is a first model for such situations, based on work of Morris [Mor00] that in turn builds on earlier work by Blume [Blu93], Ellison [Ell93], and Young [You98]. Consider a graph $G = (V, E)$ in which the nodes are the individuals in the population, and there is an edge (v, w) if v and w are friends, or otherwise engaged in some kind of social interaction. Sociologists refer to such a graph as a *social network*, a structure in which the nodes are individuals or other social entities (such as organizations), and the edges represent some type of social tie.

We will study a situation in which each node has a choice between two possible behaviors: the “old” behavior, labeled A , and the “new” behavior, labeled B . On each edge (v, w) , there is an incentive for v and w to have their behaviors match, which we model as the following coordination game parametrized by a real number q , with $0 < q < 1$.

- If v and w both choose behavior A , they each receive a payoff of q .
- If v and w both choose behavior B , they each receive a payoff of $1 - q$.
- If v and w choose opposite behaviors, they each receive a payoff of 0.

Of course, it is easy to imagine many possible generalizations of this simple game, and we will explore some of these in the next section as well as in the exercises at the end of the chapter. But for now, we will keep things deliberately simple.

Node v is playing this game with each of its neighbors in G , and its overall payoff is simply the sum of the payoffs from these separate games. Notice how q (specifically its relation to $1 - q$) captures the extent to which the new behavior is preferable to the old behavior at a purely “local” level, taking into account only pairwise interactions.

Suppose the behaviors of all other nodes are fixed, and node v is trying to select a behavior for itself. If the degree of node v is d_v , and d_v^A of its neighbors have behavior A and d_v^B have behavior B , then the payoff to v from choosing behavior A is qd_v^A while the payoff from choosing behavior B is $(1 - q)(d_v^B)$. A simple computation shows that v should adopt behavior B if $d_v^B > qd_v$, and behavior A if $d_v^B < qd_v$. (To handle ties, we will say that v adopts behavior B if $d_v^B = qd_v$.) In other words, q is a threshold: a node should adopt the new behavior if at least a q fraction of its neighbors have done so. Note that new behaviors for which q is small spread more easily — a node is more receptive to switching to a new behavior B when q is small.

Cascading behavior and the contagion threshold. We can now study a basic model of cascading behavior in G , simply assuming that each node repeatedly updates its choice of A or B in response to the current behaviors of its neighbors. Keeping the model as simple as possible, we assume that each node simultaneously updates its behavior in each of discrete time steps $t = 1, 2, 3, \dots$. If S is the set of nodes initially adopting the new behavior B , we let $h_q(S)$ denote the set of nodes adopting B after one round of updating with threshold q ; we let $h_q^k(S)$ denote the result of applying h_q to S a total of k times in succession — in other words, this is the set of nodes adopting B after k rounds of updating. Note that nodes may switch from A to B or from B to A , depending on what their neighbors are doing; it is not necessarily the case, for example, that S is a subset of $h(S)$.

One of the central questions in such a model is to determine when a small set of nodes initially adopting a new behavior can eventually convert all (or almost all) of the population. We formalize this as follows. First, we will assume the node set of G is countably infinite, with each node having a finite number of neighbors. (Anything we refer to as a “graph” in this section will have this property.) We say that a node w is *converted* by a set S if, for some k , the node w belongs to $h_q^j(S)$ for all $j \geq k$. We say that a set S is *contagious* (with respect to h_q) if every node is converted by S — that is, if a new behavior originating at S eventually spreads to the full set of nodes.

Now, it is easier for a set S to be contagious when the threshold q is small, so the interesting question is how large a threshold we can have and still observe small sets spreading a new behavior to everyone. We therefore define the *contagion threshold* of the social network G to be the maximum q for which there exists a finite contagious set. Note that the contagion threshold is a property purely of the topology of G — a network with large contagion threshold enables even behaviors that spread sluggishly to potentially reach the full population.

An Example and a Question. Here is an example to make the definitions more concrete. Suppose G is a (two-way) infinite path, with nodes labeled $\{\dots, -2, -1, 0, 1, 2, \dots\}$, and there is a new behavior B with threshold $q = 1/2$. Now, first, suppose that the single node 0 initially adopts B . Then in time step $t = 1$, nodes 1 and -1 will adopt B , but 0 (observing both 1 and -1 in their

initial behaviors A) will switch to A . As a result, in time step $t = 2$, nodes 1 and -1 will switch back to behavior A , while nodes 0, 2, and -2 will switch to B . The system will continue alternating in this “blinking” pattern between increasingly long intervals of even-numbered and odd-numbered nodes, but no node will ever permanently switch to B .

On the other hand, suppose that the set $S = \{-1, 0, 1\}$ initially adopts B . Then in time step $t = 1$, these three nodes will stay with B , and nodes -2 and 2 will switch to B . More generally, in time step $t = k$, nodes $\{-k, -(k-1), \dots, k-1, k\}$ will already be following behavior B , and nodes $-(k+1)$ and $k+1$ will switch to B . Thus, every node is converted by $S = \{-1, 0, 1\}$, the set S is contagious, and hence the contagion threshold of G is at least $q = 1/2$. (Note that it would in fact have been sufficient to start with the smaller set $S' = \{0, 1\}$.)

In fact, $1/2$ is the contagion threshold of G : given any finite set S adopting a new behavior B with threshold $q > 1/2$, it is easy to see that B will never spread past the rightmost member of S .

It is instructive to try this oneself on other graphs; if one does, it quickly becomes clear that while a number of simple graphs have contagion threshold $1/2$, it is hard to find one with a contagion threshold strictly above $1/2$. This suggests the following question: *Does there exist a graph G with contagion threshold $q > 1/2$?* We will shortly answer this question, after first resolving a useful technical issue in the model.

Progressive vs. Non-progressive processes. Our model thus far has the property that as time progresses, nodes can switch from A and B or from B to A , depending on the states of their neighbors. Many behaviors that one may want to model, however, are *progressive*, in the sense that once a node switches from A to B , it remains with B in all subsequent time steps. (Consider, for example, a professional community in which the behavior is that of returning to graduate school to receive an advanced degree. For all intents and purposes, this is a progressive process.) It’s worth considering a variation on our model that incorporates this notion of monotonicity, for two reasons. First, it’s useful to be able to capture these types of settings; and second, it will turn out to yield useful ways of thinking about the non-progressive case as well.

We model the progressive contagion process as follows. As before, time moves in discrete steps $t = 1, 2, 3, \dots$. In step t , each node v currently following behavior A switches to B if at least a q fraction of its neighbors are currently following B . Any node following behavior B continues to follow it in all subsequent time steps. Now, if S is the set of nodes initially adopting B , we let $\bar{h}_q(S)$ denote the set of nodes adopting B after one round of updating in this progressive process, and we let $\bar{h}_q^k(S)$ denote the result of applying \bar{h}_q to S a total of k times in succession. We can then define the notion of *converted* and *contagious* with respect to \bar{h}_q exactly as we did for h_q .

With a progressive process, it seems intuitively that it should be easier to find finite contagious sets — after all, in the progressive process, one doesn’t have to worry about early adopters switching back to the old behavior A and thereby killing the spread of B . In view of this intuition, it’s perhaps a bit surprising that for any graph G , the progressive and non-progressive models have the same contagion threshold [Mor00].

Theorem 2.1 *For any graph G , there exists a finite contagious set with respect to h_q if and only if there exists one with respect to \bar{h}_q .*

Proof. Clearly, if S is contagious with respect to h_q , then it is also contagious with respect to \bar{h}_q .

Hence the crux of the proof is the following: given a set S that is contagious with respect to \bar{h}_q , we need to identify a set S' that is contagious with respect to h_q .

Thus, let S be contagious with respect to \bar{h}_q . The main observation behind the proof is the following. Since all nodes of G have finite degree, there is a finite set \bar{S} that consists of S together with every node that has a neighbor in S . Since $\bar{h}_q^k(S)$ eventually grows to include every node of G , there exists some ℓ such that $\bar{h}_q^\ell(S)$ contains \bar{S} . We define $T = \bar{h}_q^\ell(S)$, and we claim that T is contagious with respect to h_q , which will complete the proof. Thus, intuitively, we watch the non-progressive process until it “engulfs” the set of initial adopters S , surrounding them with all their possible neighbors; this larger set is then a robust enough point that the process would spread even under the progressive rule from here on.

So why is the set T contagious with respect to h_q ? This requires a bit of manipulation of the definitions of h_q and \bar{h}_q , although the details are not that complicated. We first note the following fact, whose proof is by induction on j is left an exercise to the reader:

$$\text{For all } X \text{ and all } j, \text{ we have } \bar{h}_q^j(X) = X \cup h_q(\bar{h}_q^{j-1}(X)). \quad (1)$$

In other words, to get $\bar{h}_q^j(X)$, rather than applying \bar{h}_q to $\bar{h}_q^{j-1}(X)$, we can instead apply h_q and then add in X .

For ease of notation, let S_j denote $\bar{h}_q^j(S)$, and let T_j denote $\bar{h}_q^j(T)$. (Recall also that $T = S_\ell$.) Now, suppose $j > \ell$. Then by (1) above, we have

$$S_j = S \cup h_q(S_{j-1}). \quad (2)$$

But since S_{j-1} includes T and hence all the neighbors of S , we have $S \subseteq h_q(S_{j-1})$. Hence the “ $S \cup$ ” in (2) is superfluous, and we can write $S_j = h_q(S_{j-1})$. By induction, it now follows that for all $j > \ell$, we have

$$h_q^{j-\ell}(T) = h_q^{j-\ell}(S_\ell) = S_j,$$

and hence T is contagious with respect to h_q . ■

The Contagion Threshold is at Most $1/2$. We now return to our question: does there exist a graph G whose contagion threshold exceeds $1/2$? Thanks to Theorem 2.1, this question has the same answer regardless of whether we consider the progressive or non-progressive process, and it turns out that the analysis is very easy if we consider the progressive version.

We now show that $1/2$ is in fact an upper bound for all graphs [Mor00], which can be read as a general statement about contagion on networks: a behavior can’t spread very far if it requires a strict majority of your friends to convince you to adopt it.

Theorem 2.2 *The contagion threshold of any graph G is at most $1/2$.*

Proof. Let $q > 1/2$, and let S be any finite subset of the nodes of G . We show that S is not contagious with respect to \bar{h}_q .

Recall our notation that $S_j = \bar{h}_q^j(S)$. For a set of nodes X , we let $\delta(X)$ denote the set of edges with one end in X and the other end not in X , and we let $d(X)$ be the cardinality of $\delta(X)$. Since all nodes in G have finite degree, $d(X)$ is a natural number for any finite set of nodes X .

We now claim that for all $j > 0$ for which $S_{j-1} \subsetneq S_j$, we have $d(S_j) < d(S_{j-1})$. To see this, we account for the difference between the sets $\delta(S_{j-1})$ and $\delta(S_j)$ by allocating it over the separate contributions of the nodes in $S_j - S_{j-1}$. For each node v in $S_j - S_{j-1}$, its edges into S_{j-1} contribute to $\delta(S_{j-1})$ but not $\delta(S_j)$, and its edges into $V - S_j$ contribute to $\delta(S_j)$ but not $\delta(S_{j-1})$. But since $q > 1/2$, and since v decided to switch to B in iteration j , it has strictly more edges into S_{j-1} (i.e. the nodes that had already adopted B) than it has into $V - S_j$. Summing these strict inequalities over all nodes v in $S_j - S_{j-1}$, we have $d(S_j) < d(S_{j-1})$.

Finally, we argue that S is not contagious with respect to \bar{h}_q . Indeed, the sequence of numbers $d(S), d(S_1), d(S_2), d(S_3), \dots$ is strictly decreasing as long as the sets S, S_1, S_2, \dots remain distinct from one another. But since $d(S)$ is a natural number, and $d(S_j) \geq 0$ for all j , there must be some value k for which the sets stop growing, and $S_k = S_{k+1} = S_{k+2} = \dots$ from then on. Since S_j is finite for any j , the set S_k in particular is finite, and hence S is not contagious. ■

3 More General Models of Social Contagion

Thus far, we have been considering a very simple model for cascading behavior in a social network: people switch to a new behavior when a certain threshold fraction of neighbors have already switched, but in our first model this threshold was the same for all nodes, and all neighbors had equal “weight” in raising a node toward its threshold. Clearly a more general model could capture a greater degree of heterogeneity in the population.

It is useful to mention a few preliminary points here. In this section, we will consider graphs that may be either finite or infinite. Also, we will work with directed graphs, enabling us to encode the notion that for two nodes v and w , the influence of v on w may be different from the influence of w on v . (One can model symmetric relationships between a pair of nodes in this case by including edges in both directions between them.) Also, we will consider contagion processes that are progressive, in that a node never switches back from a new behavior to an old behavior; at this end of this section, we will discuss a way to encode non-progressive processes by a reduction to the progressive case, though by a different means than we saw earlier.

A Linear Threshold Models. In a first generalization of the model, we allow nodes to weigh the influences of their neighbors differently.¹ Further, we assume that each node’s threshold — the fraction of neighbors required for it to adopt the new behavior — is chosen uniformly at random. Thus, in the *Linear Threshold Model*, we introduce the following two ingredients.

- We have a non-negative weight b_{vw} on each edge (w, v) , indicating the influence that w exerts on v . We will require that $\sum_{w \in N(v)} b_{vw} \leq 1$, where $N(v)$ denotes the set of nodes with edges to v .
- Each node v choose a threshold θ_v uniformly at random from $[0, 1]$; this indicates the weighted fraction of v ’s neighbors that must adopt the behavior before v does.

Now, the (progressive) dynamics of the behavior operates just as it did in the previous section. Some set of nodes S starts out adopting the new behavior B ; all other nodes start out adopting

¹Given the directed nature of the graph, we adopt the following terminological convention: we say here that w is a *neighbor* of v if there is an edge (w, v) , and we say that w is an *out-neighbor* if there is an edge (v, w) .

A. We will say that a node is *active* if it is following B , and accordingly will say that it has been *activated* when it switches from A to B .

Time operates in discrete steps $t = 1, 2, 3, \dots$. At a given time t , any inactive node v becomes active if its fraction of active neighbors exceeds its threshold:

$$\sum_{\text{active } w \in N(v)} b_{vw} \geq \theta_v.$$

This in turn may cause other nodes to become active in subsequent time steps, as it did in the model of the previous section, leading to potentially cascading adoption of behavior B .

Note how the different thresholds for nodes indicate different predispositions to adopt B — small θ_v indicates a more liberal approach toward adoption, while a node with large θ_v waits until a greater fraction of its neighbors have already adopted. While we have motivated the model directly in terms of the thresholds, one can also easily derive it from a networked coordination game, in which nodes have different payoff matrices, and different “stakes” in the games with their various neighbors.

A General Threshold Model. Of course, the Linear Threshold Model is still very simple, in that it assumes influences of neighbors are strictly additive. It would be nice to express the notion that an individual will adopt a behavior when, for example, two of her relatives and three of her co-workers do so — a rule that can’t be expressed as a simple weighted sum.

To handle this richer type of model, we consider the following *General Threshold Model*. Each node v now has an arbitrary function $g_v(\cdot)$ defined on subsets of its neighbor set $N(v)$: for any set of neighbors $X \subseteq N(v)$, there is a value $g_v(X)$ between 0 and 1. We will assume here that this function is *monotone* in the sense that if $X \subseteq Y$, then $g_v(X) \leq g_v(Y)$.²

Now, the dynamics of adoption proceed just as in the Linear Threshold Model, but with $g_v(\cdot)$ playing the role of the weighed sum. Specifically, each node v chooses a threshold θ_v uniformly at random in $[0, 1]$; there is an initial set S of active nodes; and for time steps $t = 1, 2, 3, \dots$, each v becomes active if its set of currently active neighbors satisfies $g_v(X) \geq \theta_v$.

This new model is extremely general — it encodes essentially any threshold rule in which influence increases (or remains constant) as more friends adopt. Moreover, the assumption that the threshold is selected uniformly at random (rather than from some other distribution) is essentially without loss of generality, since other distributions can be represented by appropriately modifying the function g_v . We now consider one final class of models, and then discuss some of the intermediate classes of special cases that may hold particular interest.

A Cascade Model. Thus far, we have formulated models for the spread of a behavior strictly in terms of node thresholds — as some people adopt the behavior, the thresholds of others are exceeded, they too adopt, and the process spreads. It is natural, however, to ask whether we can pose a different model based more directly on the notion that new behaviors are contagious: a probabilistic model in which you “catch” the behavior from your friends. It turns out not to be hard to do this, and moreover, the resulting model is equivalent to the General Threshold Model.

²An interesting issue, which has been the subject of qualitative investigation but much less theoretical modeling, is the question of non-monotone influence — a node may be motivated to adopt a new behavior once a few friends have done so, but then motivated to abandon it once too many have done so (see e.g. [Gra78]).

We define the *Cascade Model* to incorporate these ideas as follows. Again, there is an initial active set S , but now the dynamics proceeds as follows: whenever there is an edge (u, v) such that u is active and v is not, the node u is given one chance to activate v . This activation succeeds with some probability that depends not just on u and v , but also on the *set* of nodes that have already tried and failed to activate v . If u succeeds, then v may now in turn try to activate some of its (currently inactive) out-neighbors; if u fails, then u joins the set of nodes who have tried and failed to activate v .

This model thus captures the notion of contagion more directly, and also allows us to incorporate the idea that a node's receptiveness to influence depends on the past history of interactions with its neighbors. We make the model concrete as follows. In place of a function g_v , each node v now has an *incremental function* that takes the form $p_v(u, X)$, where u is a neighbor of v and X is a set of neighbors of v not containing u . The value $p_v(u, X)$ is the probability that u succeeds in activating v , given that the set X of neighbors has already tried and failed. For our purposes here, we will only consider functions p_v that are *order-independent*: if a set of neighbors u_1, u_2, \dots, u_k all try to influence v , then the overall probability of success (as determined by successive applications of p_v) does not depend on the order in which they try.

While the Cascade Model is syntactically different from the General Threshold Model, we now argue that the two are in fact equivalent: One can translate from a set of incremental functions p_v to a set of threshold functions g_v , and vice versa, so that the resulting processes produce the same distributions on outcomes [KKT03].

We now describe the translations in both directions; further detail behind the proofs can be found in [KKT03]. First, suppose we are given an instance of the General Threshold Model with functions g_v ; we define corresponding functions p_v as follows. If a set of nodes X has already tried and failed to activate v , then we know that v 's threshold θ_v lies in the interval $(g_v(X), 1]$; subject to this constraint, it is uniformly distributed. In order for u to succeed after all the nodes in X have tried and failed, we must further have $\theta_v \leq g_v(X \cup \{u\})$. Hence we should define the incremental function

$$p_v(u, X) = \frac{g_v(X \cup \{u\}) - g_v(X)}{1 - g_v(X)}.$$

Conversely, suppose we have incremental functions p_v . Then the probability v is not activated by a set of neighbors $X = \{u_1, u_2, \dots, u_k\}$ is $\prod_{i=1}^k (1 - p_v(u_i, X_{i-1}))$, where we write $X_{i-1} = \{u_1, \dots, u_{i-1}\}$. Note that order-independence is crucial here, to ensure that this quantity is independent of the way in which we label the elements of X . Hence we can define a threshold function g_v by setting

$$g_v(X) = 1 - \prod_{i=1}^k (1 - p_v(u_i, X_{i-1})).$$

This completes the translations in both directions, and hence establishes the equivalence of the two models.

Next we consider some special cases of the Cascade Model that will be of particular interest to us. (Given the equivalence to the General Threshold Model, these could also be written in that framework, though not always as simply.)

- (i) First, it is easy to encode the notion that v will deterministically activate once it has k active neighbors: we simply define $p_v(u, X) = 0$ if $|X| \neq k - 1$, and $p_v(u, X) = 1$ if $|X| = k - 1$.

- (ii) In contrast, the influence of a node’s neighbors exhibits *diminishing returns* if it attenuates as more and more people try and fail to influence it. Thus, we say that a set of incremental functions p_v exhibits diminishing returns if $p_v(u, X) \geq p_v(u, Y)$ whenever $X \subseteq Y$.
- (iii) A particularly simple special case that exhibits diminishing returns is the *Independent Cascade Model*, in which u ’s influence on v is independent of the set of nodes that have already tried and failed: $p_v(u, X) = p_{uv}$ for some parameter p_{uv} that depends only on u and v .

We will see that the contrast between (i) and (ii) above will emerge as a particularly important qualitative distinction: whether the influence of one’s neighbors in the social network incorporates some notion of “critical mass” (as in (i)), with a crucial number of adopters needed for successful influence; or whether the strength of influence simply decreases steadily (as in (ii)) as one is exposed more and more to the new behavior. In the next section, we will discuss an algorithmic problem whose computational complexity is strongly affected by this distinction; and following that, we will discuss some recent empirical studies that seek to identify the two sides of this dichotomy in on-line influence data.

Before this, we briefly discuss a useful way of translating between the progressive and non-progressive versions of these cascade processes.

Progressive vs. Non-progressive processes (redux). The discussion in this section has been entirely in terms of progressive processes, where nodes switching from the old behavior A to the new behavior B never switch back. There is a useful construction that allows one to study the non-progressive version of the process by translation to a progressive one on a different graph [KKT03]. As it is a very general construction, essentially independent of the particular influence rules being used, we describe it at this level of generality.

Given a graph G on which we have a non-progressive process that may run for up to T steps, we create a larger graph G' built from T copies of G , labeled G_1, G_2, \dots, G_T . Now, let $v^{(i)}$ be the copy of node v in the graph G_i ; we construct edges $(u^{(i-1)}, v^{(i)})$ for each neighbor u of v . As a result, the neighbors of $v^{(i)}$ in G' are just the copies of v ’s neighbors that “live” in the previous time-step. In this way, we can define the same influence rules on G' , node-by-node, that we had in G , and study the non-progressive process in G as a progressive process in G' : some copies of v in G' will be an active, and other won’t, reflecting precisely the time steps in which v was active in G .

4 Finding Influential Sets of Nodes

A number of current Internet applications concern domains in which cascading behavior is a crucial phenomenon, and where a better understanding of cascades could lead to important insights. One example is *viral marketing*, where a company tries to use word-of-mouth effects to market a product with a limited advertising budget, relying on the fact that early adopters may convince friends and colleagues to use the product, creating a large wave of adoptions. While word-of-mouth effects have a history in the area of marketing that long predates the Internet, viral marketing has become a particularly powerful force in on-line domains, given the ease with which information spreads, and the rich data on customer behavior that can be used to facilitate the process.

A second example is the design of search tools to track news, blogs, and other forms of on-line discussion about current events. When news of an event first appears, it generally spreads rapidly

through a network of both mainstream news sources and the larger population of bloggers — as news organizations and individuals learn of the event, they write their own commentary or versions of the story, and subsequent waves can occur as new developments take place. As with our first example, news was studied as a diffusion process long before the appearance of the Internet, but the fact that news sources are now on-line provides large-scale, time-resolved data for studying this diffusion, as well as the opportunity to build tools that can help people track the development of a news story in real-time as it evolves.

There are a number of interesting algorithmic questions related to these processes, and here we focus on a particular one, posed by Domingos and Richardson [DR01] — the identification of influential sets of nodes. While this is a natural question in the context of both our examples above, we describe the problem in terms of the viral marketing framework, where it is particularly easy to express the underlying motivation.

The Most Influential Set of Nodes. Suppose we are a firm trying to market a new product, and we want to take advantage of word-of-mouth effects. One strategy would be as follows: we collect data on the social network interactions among our potential customers, we choose a set S of initial adopters, and we market the product directly to them. Then, assuming they adopt the product, we rely on their influence to generate a large cascade of adoptions, without our having to rely on any further direct promotion of the product. The question of how one goes about inferring the social network interactions and building a cascade model for this purpose is an interesting and largely unexplored topic. Here, however, we assume that all this data is provided to us, and we focus on the algorithmic problem that comes next: how do we choose the set S ?

We can formulate this question concretely as follows. For any instance of the General Threshold or Cascade Models, there is a natural *influence function* $f(\cdot)$ defined as follows: for a set S of nodes, $f(S)$ is the expected number of active nodes at the end of the process, assuming that S is the set of nodes that are initially active. (We will assume in this section that all graphs are finite, so the processes we are considering terminate in a number of steps that is bounded by the total number of nodes n .) From the marketer’s point of view, $f(S)$ is the expected number of total sales if they get S to be the set of initial adopters. Now, given a budget k , how large can we make $f(S)$ if we are allowed to choose a set S of k initial adopters? In other words, we wish to maximize $f(S)$ over all sets S of size k .

This turns out to be a hard computational problem. First, for almost any special case of the models in the previous section — even very simple special cases — it is NP-hard to find the optimal set S . Moreover, one can construct instances of the model for which it is NP-hard even to approximate the optimal value of $f(S)$ to within a factor of $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where again n is the number of nodes [KKT03]. We leave the proofs of these statements as exercises to the reader.

Since NP-hardness applies to almost all versions of the model, there is not much we can do about it; instead, we will try to identify broad sub-classes of the models that are not susceptible to strong inapproximability results, and for which good approximation results can be obtained.

Submodularity as a Route to Good Approximations. While we will not go into the details of the inapproximability proofs, they rely on constructing a cascade process for which the resulting influence function f has a “knife-edge” property: as one adds nodes to S , one initially gets very little spreading, but once exactly the right set has been added, the process suddenly spreads very

widely. And as we will see shortly, this is actually crucial, since influence functions f that grow in a less pathological way allow for good approximation algorithms. The key to this is a property called *submodularity*.

We say that a function f is *submodular* if adding an element to a set Y causes a smaller marginal improvement than adding the same element to a subset of Y . Specifically, f is submodular if for all sets $X \subseteq Y$ and all elements $v \notin Y$, we have

$$f(X \cup \{v\}) - f(X) \geq f(Y \cup \{v\}) - f(Y).$$

Thus, submodularity is a type of diminishing returns property: the benefit of adding elements decreases as the set to which they are being added grows. We also note that all the influence functions arising from our models here are *monotone*, in the sense that $f(\phi) = 0$ and $f(X) \leq f(Y)$ whenever $X \subseteq Y$.

For such a function f , it is natural to hope that a simple “hill-climbing” approach might lead to a good approximation for the optimal value over all k -element sets: since the marginal benefits only decrease as elements are added, it is hard to “hide” a very large optimum. In fact, this intuition is formalized in the following theorem of Nemhauser, Wolsey, and Fisher [NWF78].

Theorem 4.1 *Let f be a monotone submodular function, and let S^* be the k -element set achieving the maximum possible value of f . Let S be a k -element set obtained by repeatedly, for k iterations, including the element producing the largest marginal increase in f . (We can think of S as the result of straightforward hill-climbing with respect to the function f .) Then $f(S) \geq (1 - \frac{1}{e})f(S^*) \geq .63f(S^*)$.*

This theorem will be our main vehicle for obtaining approximation algorithms for influence maximization: we will identify instances of the Cascade Model for which the influence function is submodular (it is always monotone for the models we consider), and this will allow us to obtain good results from hill-climbing. There is one further wrinkle in our use of Theorem 4.1, however: it is computationally intractable, even in simple special cases, to evaluate the function f exactly. Fortunately, one can adapt Theorem 4.1 to show that for any $\varepsilon > 0$, if one can approximately evaluate f sufficiently accurately (relative to the granularity taken by the values of f , which are integers in our case), this approximate evaluation of f produces a set S such that $f(S) \geq (1 - \frac{1}{e} - \varepsilon)f(S^*)$. As we can achieve such approximate evaluation for the influence functions we are considering here, this will allow us to apply this approximate form of Theorem 4.1.

We now proceed with the search for broad special cases of the Cascade Model for which the influence function f is submodular.

Diminishing Returns and Submodularity. It is perhaps easier to think first of instances for which the resulting function f is *not* submodular. For example, suppose every node v has a sharp threshold $\ell > 1$: v requires at least ℓ active neighbors before v itself becomes active. Then it is easy to construct graphs G on which $f(S)$ remains small until a sufficient number of nodes have been included in S , and then it abruptly jumps up; such a function is not submodular.

More generally, instances based on critical mass effects at the level of individual nodes tend not to yield influence functions that are submodular. But is the opposite true? Does diminishing

returns at the level of individual nodes imply that the influence function f is submodular? In fact, the following theorem of Kempe, Kleinberg, and Tardos [KKT03, KKT05] establishes a general sense in which this is the case:

Theorem 4.2 *For any instance of the Cascade Model in which all the incremental functions p_v exhibit diminishing returns, the resulting influence function f is submodular.*

An appealing feature of this theorem is the way in which it establishes a “local-to-global” link: if each individual node experiences influence in a way that exhibits diminishing returns, then the network as a whole experiences influence (from an external marketer) in the same way. The search for such cases in which local behavior implies analogous global behavior is a theme throughout social network research, and finding common principles underlying such effects is an interesting general research issue.

The proof of Theorem 4.2, while not long, is a bit intricate; so instead we describe the proof of a special case of this theorem, where the analysis is conceptually very clean, yet still illustrative of some of the issues involved in proving submodularity. Recall that in the Independent Cascade Model, defined in the previous section, we define the incremental functions via $p_v(u, X) = p_{uv}$: in other words, the influence of u on v depends only on u and v , not on the set of other nodes that have already tried and failed to influence v .

Theorem 4.3 *For any instance of the Independent Cascade Model, the resulting influence function f is submodular.*

Proof. Much of the challenge in proving submodularity here is that the computational intractability of our function f also translates into a kind of conceptual intractability: we don’t know enough about its structure to simply “plug it in” to the submodular inequality and hope to verify it directly. A much more powerful approach is instead to decompose f into simpler functions, and check submodularity for the parts of this decomposition.

To lay the groundwork for this plan, we start by discussing two useful facts about submodularity. The first is this: if f_1, f_2, \dots, f_r are submodular functions, and c_1, \dots, c_r are non-negative real numbers, then the function f defined by the weighted sum $f(X) = \sum_{i=1}^r c_i f_i(X)$ is also submodular. The second fact is actually the identification of a simple, useful class of submodular functions, the following. Suppose we have a collection of sets C_1, C_2, \dots, C_r , and for a set $X \subseteq \{1, 2, \dots, r\}$, we define $f(X) = |\cup_{i \in X} C_i|$. For obvious reasons, we will call such a function f a *size-of-union* function. Then it is not hard to verify that any size-of-union function is submodular. As we will see below, such functions will arise naturally in our decomposition of the influence function f .

Now, consider the following alternate way of viewing the Independent Cascade Process. In the standard view, each time a node u becomes active, it flips a coin of bias p_{uv} to determine whether it succeeds in activating v . Now, in the alternate, equivalent view of the process, suppose that for each edge (u, v) , we flip a coin of bias p_{uv} *in advance*, planning to only consult the outcome of the coin flip if we ever need to, when u becomes active.

If there are m edges in the graph, then there are 2^m possible collective outcomes of the coin flips. Let α denote a particular one of these 2^m outcomes, and let $f_\alpha(S)$ denote the eventual number of activated nodes, given that S is the initial active set *and* α is the outcome of the coin flips. Unlike

f , the function f_α is easy to understand, as follows. For each edge (u, v) , we say that it is *live* (with respect to α) if the advance coin flip came up heads. For a node s , we let $R_s^{(\alpha)}$ denote the set of all nodes that are reachable from s on paths consisting entirely of live edges. It is now easy to check that a node is eventually activated if and only if it is reachable from some node in S by some path consisting entirely of live edges, and hence

$$f_\alpha(S) = \left| \bigcup_{s \in S} R_s^{(\alpha)} \right|.$$

In other words, each f_α is a size-of-union function, and hence submodular.

But this is essentially all we need, since by definition

$$f(S) = \sum_{\alpha} \text{Prob}[\alpha] \cdot f_\alpha(S).$$

That is, f is a non-negative weighted sum of submodular functions, and hence f is submodular, as desired. ■

We close this particular discussion by noting that Theorem 4.2 may not be the most general possible formulation of the local-to-global principle we have been discussing. In particular, if all the incremental functions p_v in an instance of the Cascade Model exhibit diminishing returns, then in the equivalent instance of the General Threshold Model, the resulting threshold functions g_v are submodular (a related notion of diminishing returns). However, the converse does not hold: there exist instances of the General Threshold Model with submodular threshold functions g_v for which the equivalent instance of the Cascade Model has incremental functions p_v that do *not* satisfy diminishing returns. (An important example is the Linear Threshold Model, which does not translate into an instance of the Cascade Model with diminishing returns. Despite this, one can show via a separate analysis that influence functions f arising from the Linear Threshold Model are always submodular.)

Thus, the General Threshold Model with submodular threshold functions is strictly more general than the Cascade Model with incremental functions satisfying diminishing returns. Hence the following very recent result of Mossel and Roch [MR07], proving a conjecture of Kempe et al. [KKT03], generalizes Theorem 4.2.

Theorem 4.4 *For any instance of the General Threshold Model in which all the threshold functions g_v are submodular, the resulting influence function f is submodular.*

Further Direction: Alternate Marketing Strategies. We conclude this section by briefly discussing a different strategy through which a marketer could try to take advantage of word-of-mouth effects. Rather than targeting nodes, as we have been discussing thus far, one could instead target edges: each time an individual u buys a product, an incentive is offered for u to recommend the product to a friend v . A number of on-line retailers have constructed recommendation incentive programs around this idea: for example, each time you buy a product, you are given the opportunity to send e-mail to a friend with a special offer to buy the product as well; if the friend goes on to buy it, each of you receives a small cash refund [LAH06, LSK06].

Strategies of this type have a different flavor from the targeting of nodes: rather than trying to create a large cascade by influencing initial adopters, one tries to create a large cascade in effect by amplifying the force with which influence is transmitted across edges. (Clearly, one still needs some initial adopters as well for this to succeed.) While there has been empirical work on the outcomes of particular recommendation incentive programs, it is an open question to analyze theoretical models for the result of such incentives for the problem of maximizing influence.

5 Empirical Studies of Cascades in On-Line Data

As we noted at the outset, there has been a huge amount of empirical work on diffusion in social networks over the past half century. A crucial challenge in this research is the fact that the phenomena being studied — the ways in which social network links affect adoption of innovations — are very hard to measure and assess. Most research of this type has focused on small or moderately-sized groups, which are then studied in detail, and the resulting analysis has provided fundamental insights into subtle issues such as the characteristics of adopters at different stages of the innovation.

The theoretical models we have been discussing thus far are motivated at a qualitative level by this empirical work, but it remains an important problem to relate the models to real diffusion data at a more precise, quantitative level. One reason why it has been difficult to do this stems from the type of data available: While existing empirical studies can address fairly rich questions at the scale at which they operate, the resulting datasets tend to be too small to provide accurate estimates of basic quantities needed for assessing the validity of the theoretical models — for example, how adoption probability depends on structural properties of a node’s network neighbors. What is needed for this task are large datasets tailored to provide answers to such questions with limited noise.

Diffusion Data from On-Line Communities. Very recently, large on-line datasets from several sources have produced measurements that raise interesting connections to the theoretical models. One such study was performed on the on-line blogging and social networking site LiveJournal [BHKL06]. LiveJournal has several million members and several hundred thousand user-defined communities; members maintain individual Web pages with personal information, blogs, and — most importantly for our purposes here — lists of their friends in the system and the communities to which they belong.

From the lists of friends, we can construct an underlying social network, with an edge (v, w) if v lists w as a friend. We then treat each community as a behavior that diffuses through this network: since communities grow by acquiring members over time, we can study how a member’s likelihood of joining a group depends on the number of friends he or she has in the group.

Here is a concrete way of formulating this question. At two times t_1 and t_2 (a few months apart), snapshots are taken of the social network and community memberships on LiveJournal. Now, for each number $k \geq 0$, consider the set U_k of all pairs (u, C) such that user u did not belong to community C at time t_1 , but had u had k friends in C at t_1 . We let $P(k)$ denote the fraction of pairs (u, C) in the set U_k for which u had joined C by time t_2 . In this way, $P(k)$ serves as an answer to the question: what is the probability of joining a LiveJournal community, given that you had k friends in it at an earlier time?

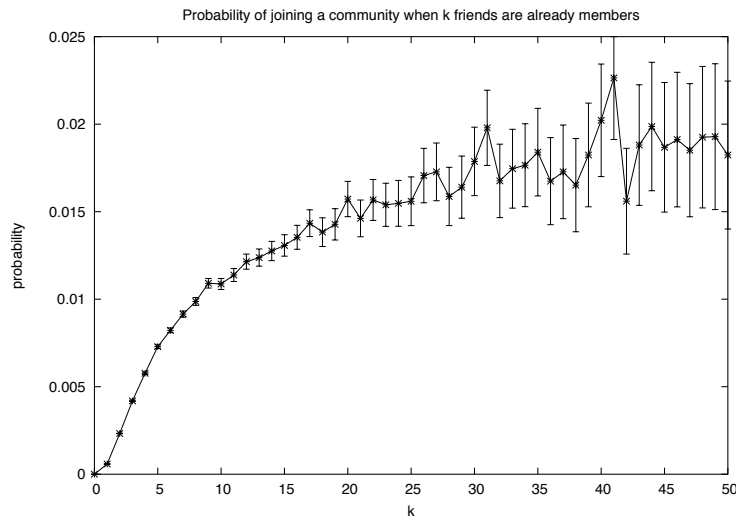


Figure 1: The probability of joining a LiveJournal community as a function of the number k of friends in the community at an earlier point in time. Error bars represent two standard errors.

Figure 1 shows a plot of $P(k)$ as a function of k for the LiveJournal data. A few things are quickly apparent from the plot. First, the dependence on k is clearly dominated by a diminishing returns effect, in which $P(k)$ grows less quickly as k increases. Indeed, this dependence is quite smooth, with a good fit to a function of the form $P(k) = \varepsilon \log k$ up to moderately large values of k — in particular, this means that $P(k)$ continues increasing even as k becomes fairly large. Finally, there is an initial but significant deviation from diminishing returns for $k = 0, 1, 2$, with $P(2) > 2P(1)$. In other words, having a second friend in a community gives a significant boost to the probability of joining, but after that the diminishing returns effect takes over.

Similar diminishing returns effects have been observed recently in other large-scale datasets that exhibit diffusion-style processes — for example, the probability of publishing a paper at a computer science conference as a function of the number of co-authors who have previously published there [BHKL06]; or the probability of purchasing a DVD in a recommendation incentive program run by a large on-line retailer, as a function of the number of friends who sent an e-mail recommendation [LAH06]. Given how a common effect — diminishing returns — is appearing in large-scale data from such diverse sources, it is an interesting open question to try finding a reasonable mechanism to explain this, potentially including the approximately logarithmic functional form of $P(k)$ in terms of k .

Closer analysis of the LiveJournal data also reveals more subtle effects that contribute to diffusion in a significant way. One that is particularly striking is the connectedness of a person’s friends. If we look at the pairs in U_k — recall that this consists of users with k friends in a community they don’t initially belong to — it turns out that a user is significantly more likely to join a community

if his or her k friends have many edges among themselves than if they don't. In other words, a highly connected set of friends in a community exerts a greater "gravitational force" than a comparable number of friends who are not as well connected among themselves. It is an interesting open question to understand the underlying factors that lead to this effect, to investigate how broadly it applies by considering related datasets, and to think about ways of extending the theoretical models to incorporate effects such as the connectedness of a node's neighbor set.

Relating the Empirical and Theoretical Models. The results shown in Figure 1 provide a good means of elaborating on the point made at the start of this section, that diffusion data from on-line sources, at very large scales, provides both more and less than one finds in classical diffusion studies. It provides less in the sense that we know very little about who individual users on LiveJournal are, what the links between them mean, or what motivates them to join communities. A LiveJournal user may have a link to a close friend, or to someone they've barely met, or simply because they're trying to accumulate as many links on the system as they can. Given this, it is very hard to ask the kind of nuanced questions that one sees in more traditional diffusion studies, which deal with smaller datasets for which they have assembled (often at great effort) a much clearer picture. On the other hand, the fact that the curve in Figure 1 is so smooth is precisely the result of having a dataset large enough to contain hundreds of thousands of communities diffusing across a network of millions of users — on a dataset containing just hundreds of individuals, any curve representing $P(k)$ will be extremely noisy. Indeed, given how many different things the links and community memberships on LiveJournal mean to different users, the clean logarithmic form of the resulting curve is perhaps all the more striking, and in need of deeper explanation.

That is a first caveat. A second is that there remains a significant challenge in relating curves like the one in Figure 1 to the theoretical models in the earlier sections. The models we discussed there all had a discrete, operational flavor: each node follows a fixed probabilistic rule, and it uses this rule to incorporate information from its neighbors over time. In contrast, the curve in Figure 1 is produced by observing the full system at one point in time, and then returning to see what has changed at a later point in time. The dependence of $P(k)$ on k expressed in this way reflects an aggregate property of the full population, and does not imply anything about any particular individual's response to their friends' behaviors. Even for a specific individual, we do not know when (or even if) they became aware of their friends' behavior between these two points of time, nor when this translated into a decision by them to act. In particular, this makes it hard to determine how the notion of diminishing returns captured in Figure 1 is actually aligned with the formal definition of diminishing returns in Sections 3 and 4. It is a general and interesting question to explore frameworks that can incorporate this asynchrony and uncertainty about the way in which information flows and is acted on by nodes in a social network, leading to a closer integration of the theoretical and empirical results.

As novel kinds of on-line information systems continue to proliferate, one sees diffusion processes not just forming part of the underpinnings of the system, but in many cases built directly into the design as well, in settings such as social search, media sharing, or community-based question-answering. As part of this process, extremely rich data is becoming available for studying diffusion processes in on-line environments, at a large scale and with very fine time resolution. These developments are forming a feedback loop that will inevitably drive the formulation of richer theories, and enable us to pose more incisive questions about the ways in which information, influence, and

behaviors of all kinds spread through human social networks.

6 Notes and Further Reading

The general topic of diffusion in social networks is discussed in the books by Coleman, Menzel, and Katz [CMK66], Rogers [Rog95], and Valente [Val95], and the survey by Strang and Soule [SS98]. Granovetter [Gra78] and Schelling [Sch78] provide some of the early mathematical models for these processes. Early game-theoretic formulations of diffusion models were proposed by Blume [Blu93] and Ellison [Ell93], and they form part of the focus of a book by Young [You98]. The specific model and results in Section 2 are from Morris [Mor00], and further game-theoretic models of diffusion have been explored by Jackson and Yariv [JY05]. There are also connections at a technical level between the models used in studying diffusion and some of the more graph-theoretic techniques that have been applied to evolutionary game theory; see for example the survey by Lieberman, Hauert, and Nowak [LHN05] and the paper by Kearns and Suri [KS06].

Models for diffusion are also closely related to work the topic of contact processes and particle systems studied in the area of probability [Dur88, Lig85], as well as to the long history of work in mathematical epidemiology, which studies the dynamics of biological (as opposed to social) contagion [Bai75]. There has been a recent line of work aimed at relating such probabilistic contagion models more closely to the underlying network structure; see for example the recent work of by Pastor-Satorras and Vespignani [PSV00], Newman [New02], and Alon, Benjamini, and Stacey [ABS04].

The problem of finding the most influential set of k individuals, as discussed in Section 4, was posed, together with the viral marketing motivation, by Domingos and Richardson [DR01]. The search for influential nodes in networks of blogs and news sources has been considered by Adar et al. [AZAL04], Gruhl et al. [GLNGT04], and Kumar et al. [KNRT04].

The approximation results in Section 4 are due to Kempe, Kleinberg, and Tardos [KKT03, KKT05] and Mossel and Roch [MR07]; the general theorem on which they depend, that hill-climbing provides a good approximation for arbitrary monotone submodular functions, is due to Fisher, Nemhauser, and Wolsey [NWF78]. The formulations of the models in Section 3 are also from Kempe et al. [KKT03], with closely related models proposed independently by Dodds and Watts [DW04].

Recommendation incentive programs, discussed at the end of Section 4, are studied empirically by Leskovec, Adamic, and Huberman [LAH06] and by Leskovec, Singh, and Kleinberg [LSK06]; the development of theoretical models for such systems remains largely an open question. The study of diffusion processes on LiveJournal discussed in Section 5 is from Backstrom, Huttenlocher, Kleinberg, and Lan [BHKL06], where a number of more subtle features of diffusion on LiveJournal are investigated as well. The identification of diminishing returns effects in recommendation incentives is from Leskovec, Adamic, and Huberman [LAH06].

7 Exercises

1. Prove Claim 1 in the proof of Theorem 2.1.

2. The first model we considered in Section 2 was based on a networked coordination game in which, across each edge (v, w) , nodes v and w each receive a payoff of q if they both choose behavior A , they each receive a payoff of $1 - q$ if they both choose behavior B , and they each receive a payoff of 0 if they choose opposite behaviors. Let us call this a *coordination game with parameter q* .

It is natural to ask what happens if we consider a more general kind of coordination game on each edge. Suppose in particular that, on each edge (v, w) , node v receives payoff u_{XY} if it chooses strategy X while w chooses strategy Y , for any choice of $X \in \{A, B\}$ and $Y \in \{A, B\}$. Moreover, to preserve the coordination aspect, we assume that it is still better to play matching strategies: $u_{AA} > u_{BA}$ and $u_{BB} > u_{AB}$.

While this is indeed a more general kind of game, prove that the results on the contagion threshold remain the same. Specifically, prove that for any infinite graph G with finite node degrees, and for any choice of payoffs $\{u_{AA}, u_{BA}, u_{AB}, u_{BB}\}$ satisfying $u_{AA} > u_{BA}$ and $u_{BB} > u_{AB}$, there exists a real number q such that the following holds: A finite set S is contagious in G with respect to the coordination game defined by $\{u_{AA}, u_{BA}, u_{AB}, u_{BB}\}$ if and only if it is contagious in G with respect to the coordination game with parameter q .

3. **(a)** In Section 4, we considered the problem of finding a set S of k nodes that maximizes the expected number of activated nodes $f(S)$. Show that for some class of instances of the General Threshold or Cascade Model, finding the optimal set S is NP-hard.

(b) For some class of instances of the General Threshold or Cascade Model, show that in fact it is NP-hard to approximate the optimal value of $f(S)$ to within a factor of $n^{1-\varepsilon}$ for any $\varepsilon > 0$, where n is the number of nodes.

References

- [ABS04] Noga Alon, Itai Benjamini, and A. Stacey. Percolation on finite graphs and isoperimetric inequalities. *Annals of Probability*, 32:1727–1745, 2004.
- [AZAL04] Eytan Adar, Li Zhang, Lada A. Adamic, and Rajan M. Lukose. Implicit structure and the dynamics of blogspace. In *Workshop on the Weblogging Ecosystem*, 2004.
- [Bai75] N. Bailey. *The Mathematical Theory of Infectious Diseases and its Applications*. Hafner Press, 1975.
- [BHKL06] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. Group formation in large social networks: Membership, growth, and evolution. In *Proc. 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2006.
- [Blu93] Lawrence Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.
- [CMK66] James Coleman, Herbert Menzel, and Elihu Katz. *Medical Innovations: A Diffusion Study*. Bobbs Merrill, 1966.

- [DR01] Pedro Domingos and Matt Richardson. Mining the network value of customers. In *Proc. 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 57–66, 2001.
- [Dur88] Richard Durrett. *Lecture Notes on Particle Systems and Percolation*. Wadsworth Publishing, 1988.
- [DW04] Peter Dodds and Duncan Watts. Universal behavior in a generalized model of contagion. *Physical Review Letters*, 92(218701), 2004.
- [Ell93] Glenn Ellison. Learning, local interaction, and coordination. *Econometrica*, 61:1047–1071, 1993.
- [GLNGT04] Daniel Gruhl, David Liben-Nowell, R. V. Guha, and Andrew Tomkins. Information diffusion through blogspace. In *Proc. 13th International World Wide Web Conference*, 2004.
- [Gra78] Mark Granovetter. Threshold models of collective behavior. *American Journal of Sociology*, 83:1420–1443, 1978.
- [JY05] Matthew Jackson and Leeat Yariv. Diffusion on social networks. *Economie Publique*, 16:69–82, 2005.
- [KKT03] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence in a social network. In *Proc. 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.
- [KKT05] David Kempe, Jon Kleinberg, and Éva Tardos. Influential nodes in a diffusion model for social networks. In *Proc. 32nd Intl. Colloq. on Automata, Languages and Programming*, pages 1127–1138, 2005.
- [KNRT04] Ravi Kumar, Jasmine Novak, Prabhakar Raghavan, and Andrew Tomkins. Structure and evolution of blogspace. *CACM*, 47(12):35–39, 2004.
- [KS06] Michael Kearns and Siddharth Suri. Networks preserving evolutionary equilibria and the power of randomization. In *Proc. 7th ACM Conference on Electronic Commerce*, 2006.
- [LAH06] Jure Leskovec, Lada Adamic, and Bernardo Huberman. The dynamics of viral marketing. In *Proc. 7th ACM Conference on Electronic Commerce*, 2006.
- [LHN05] E. Lieberman, C. Hauert, and Martin Nowak. Evolutionary dynamics on graphs. *Nature*, 433:312–316, 2005.
- [Lig85] Thomas Liggett. *Interacting Particle Systems*. Springer, 1985.
- [LSK06] Jure Leskovec, Ajit Singh, and Jon M. Kleinberg. Patterns of influence in a recommendation network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 380–389, 2006.

- [Mor00] Stephen Morris. Contagion. *Review of Economic Studies*, 67:57–78, 2000.
- [MR07] Elchanan Mossel and Sebastien Roch. On the submodularity of influence in social networks. In *Proc. 39th ACM Symposium on Theory of Computing*, 2007.
- [New02] Mark E. J. Newman. The spread of epidemic disease on networks. *Physical Review E*, 66:016128, 2002.
- [NWF78] George Nemhauser, Laurence Wolsey, and Marshall Fisher. An analysis of the approximations for maximizing submodular set functions. *Mathematical Programming*, 14:265–294, 1978.
- [PSV00] R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical Review Letters*, 86:3200–3203, 2000.
- [Rog95] Everett Rogers. *Diffusion of Innovations*. Free Press, fourth edition, 1995.
- [Sch78] Thomas Schelling. *Micromotives and Macrobehavior*. Norton, 1978.
- [SS98] David Strang and Sarah Soule. Diffusion in organizations and social movements: From hybrid corn to poison pills. *Annual Review of Sociology*, 24:265–290, 1998.
- [Val95] Thomas Valente. *Network Models of the Diffusion of Innovations*. Hampton Press, 1995.
- [You98] H. Peyton Young. *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, 1998.