THEORY OF COMPUTING, Volume 11 (4), 2015, pp. 105–147
www.theoryofcomputing.org

Maximizing the Spread of Influence through a Social Network

David Kempe* Jon Kleinberg† Éva Tardos‡

Received April 17, 2014; Revised September 7, 2014; Published April 22, 2015

Abstract: Models for the processes by which ideas and influence propagate through a social network have been studied in a number of domains, including the diffusion of medical and technological innovations, the sudden and widespread adoption of various strategies in game-theoretic settings, and the effects of "word of mouth" in the promotion of new products. Motivated by the design of viral marketing strategies, Domingos and Richardson posed a fundamental algorithmic problem for such social network processes: if we can try to convince a subset of individuals to adopt a new product or innovation, and the goal is to trigger a large cascade of further adoptions, which set of individuals should we target?

We consider this problem in several of the most widely studied models in social network analysis. The optimization problem of selecting the most influential nodes is NP-hard here. The two conference papers upon which this article is based (KDD 2003 and ICALP 2005) provide the first provable approximation guarantees for efficient algorithms. Using an

The present article is an expanded version of two conference papers [51, 52], which appeared in KDD 2003 and ICALP 2005, respectively.

*Supported in part by an Intel Graduate Fellowship and an NSF Graduate Research Fellowship, an NSF CAREER Award, an ONR Young Investigator Award, and a Sloan Fellowship.

†Supported in part by a David and Lucile Packard Foundation Fellowship and NSF ITR/IM Grant IIS-0081334.

‡Supported in part by NSF ITR grant CCR-011337, and ONR grant N00014-98-1-0589.

ACM Classification: F.2.2, G.3

AMS Classification: 68W25, 90C59, 68Q25, 68Q17

Key words and phrases: social networks, approximation algorithms, influence, viral marketing, diffusion, submodular function

 DOI: 10.4086/toc.2015.v011a004

analysis framework based on submodular functions, we show that a natural greedy strategy obtains a solution that is provably within 63% of optimal for several classes of models; our framework suggests a general approach for reasoning about the performance guarantees of algorithms for these types of influence problems in social networks.

We also provide computational experiments on large collaboration networks, showing that in addition to their provable guarantees, our approximation algorithms significantly out-perform node-selection heuristics based on the well-studied notions of degree centrality and distance centrality from the field of social networks.

1 Introduction

A social network—the graph of relationships and interactions within a group of individuals—plays a fundamental role as a medium for the spread of information, ideas, and influence among its members. An idea or innovation will appear—for example, the use of cell phones among college students, the adoption of a new drug within the medical profession, or the rise of a political movement in an unstable society—and it can either die out quickly or make significant inroads into the population. If we want to understand the extent to which such ideas are adopted, it can be important to understand how the dynamics of adoption are likely to unfold within the underlying social network: the extent to which people are likely to be affected by decisions of their friends and colleagues, or the extent to which "word-of-mouth" effects will take hold.

Such network diffusion processes have a long history of study in the social sciences. Some of the earliest systematic investigations focused on data pertaining to the adoption of medical and agricultural innovations in both developed and developing parts of the world [23, 69, 79]; in other contexts, research has investigated diffusion processes for "word-of-mouth" and "viral marketing" effects in the success of new products [8, 14, 26, 33, 32, 59, 68], the sudden and widespread adoption of various strategies in game-theoretic settings [11, 29, 61, 85, 86], and the problem of cascading failures in power systems [7, 6].

Motivated by applications to marketing, Domingos and Richardson posed a fundamental algorithmic problem for such systems [26, 68]. Suppose that we have data on a social network, with estimates for the extent to which individuals influence one another, and we would like to market a new product that we hope will be adopted by a large fraction of the network. The premise of viral marketing is that by initially targeting a few "influential" members of the network—say, giving them free samples of the product—we can trigger a cascade of influence by which friends will recommend the product to other friends, and many individuals will ultimately try it. But how should we choose the few key individuals to use for seeding this process? In [26, 68], this question was considered in a probabilistic model of interaction; heuristics were given for choosing customers with a large overall effect on the network, and methods were also developed to infer the influence data necessary for posing these types of problems.

In this article, we consider the issue of choosing influential sets of individuals as a problem in discrete optimization. The optimal solution is NP-hard for most models that have been studied, including the model of [26]. The framework proposed in [68], on the other hand, is based on a simple linear model where the solution to the optimization problem can be obtained by solving a system of linear equations. Here we focus on a collection of related, NP-hard models that have been extensively studied in the social networks community, and obtain the first provable approximation guarantees for efficient algorithms in a number of general cases. The generality of the models we consider lies between that of the polynomial-time solvable model of [68] and the very general model of [26], where the optimization problem cannot even be approximated to within a non-trivial factor.

We define the concrete classes of models for the diffusion of innovations in Section 2 below, departing somewhat from the Domingos-Richardson framework: where their models are essentially *descriptive*, specifying a joint distribution over all nodes' behavior in a global sense, we focus on more *operational* models from mathematical sociology [41, 74] and interacting particle systems [33, 32, 28, 56] that explicitly represent the step-by-step dynamics of adoption. Concretely, our models assume that a certain subset A_0 of nodes is targeted for activation by the algorithm, and starts *active*. In turn, the active nodes A_t in time step t may activate further nodes in the next time step according to a known probabilistic rule. In *cascade models*, this rule is based on (independent) edge-wise decisions [33, 32, 28, 56], while *threshold models* posit cumulative effects on nodes [41, 74] which lead to activation once a certain threshold is exceeded. When no activations occur from round t to $t+1$, we say that the process has *quiesced*, and the outcome is the set A_t of active nodes in the end. Notice that our models explicitly assume that the activation process is *progressive*, in the sense that once activated, a node will never become inactive subsequently. A more detailed discussion of this assumption, as well as ways to remove it, is given in Section 6.

Given the above model, we can formally express the Domingos-Richardson style of optimization problem—choosing a good initial set of nodes to target—as follows: the algorithm chooses an initial set $A_0 = A$ of active nodes that start the diffusion process. The *influence* of a set A of nodes, denoted $\sigma(A)$, is the expected number of active nodes at the end of the process, given that A is this initial active set A_0. Formally, $\sigma(A) = \mathbb{E}[|A_t|]$, where t is the (random) time of quiescence. The *influence maximization problem* asks, for a parameter k, to find a k-node set of maximum influence. We will show below that for the models we consider, it is NP-hard to determine the optimum set for influence maximization.

1.1 Our results

Our first main result is that the optimal solution for influence maximization can be efficiently approximated to within a factor of $(1 - 1/\mathrm{e} - \varepsilon)$.

Theorem 1.1. *In the Linear Threshold and Independent Cascade models (defined formally in Section 2), there is a polynomial-time algorithm approximating the maximum influence to within a factor of* $(1 - 1/\mathrm{e} - \varepsilon)$*, where* e *is the base of the natural logarithm and* ε *is any positive real number.*

This is a performance guarantee slightly better than 63%. The algorithm that achieves this performance guarantee is a natural greedy hill-climbing strategy, and so the main content of this result is the analysis framework needed for obtaining a provable performance guarantee, and the fairly surprising fact that hill-climbing is always within a factor of at least 63% of optimal for this problem. We prove this result in Section 4 using techniques from the theory of submodular functions [24, 65]. The key ingredient of our proofs is to exhibit, for both models, distributions over graphs with the following property: the expected activation $\sigma(A)$ equals the expected number of nodes reachable from A if a graph G is chosen according to the distribution. We call this technique the *triggering set technique*.

While the triggering set technique leads to approximation guarantees for several models studied in mathematical sociology, we show in Section 5 that the *Decreasing Cascade* model (a natural generalization of the Independent Cascade model, also defined in Section 2) leads to a submodular function $\sigma(A)$, yet does not admit a proof using triggering sets. Instead, we present a more elaborate proof, based on a step-by-step analysis of the activation process.

The analysis framework of submodular functions also allows us to design and prove guarantees for approximation algorithms in much richer and more realistic models of the processes by which we market to nodes. The deterministic activation of individual nodes is a highly simplified model; an issue also considered in [26, 68] is that we may in reality have a large number of different marketing actions available, each of which may influence nodes in different ways. The available budget can be divided arbitrarily between these actions. In Section 7, we show how to extend the analysis to this substantially more general framework. Our main result here is that a generalization of the hill-climbing algorithm still provides approximation guarantees arbitrarily close to $(1-1/\mathrm{e})$.

Our theoretical results are complemented by an experimental evaluation of our algorithms. In Section 8, we report on the results of computational experiments with both the Linear Threshold and Independent Cascade models. These experiments show that in addition to its provable guarantees, the hill-climbing algorithm significantly out-performs strategies based on targeting nodes of high degree or distance centrality [83], i. e., the nodes with high degrees or small average distances to all other nodes.

1.2 Subsequent work

Since the original publication of the conference papers that this article is based on [51, 52], there has been a large amount of follow-up work. We review some of the main directions and several representative papers for each; a complete review of subsequent work is beyond the scope of this article. Several papers addressing concrete questions directly related to the present discussions are also discussed in those contexts. An interested reader can find a more detailed survey in the recent monograph of Chen, Lakshmanan and Castillo [19].

The paper by Mossel and Roch [62] positively resolves a conjecture from [51], showing that for a broader class of local influence functions, the resulting objective function is submodular. (See the discussion in Section 2.1.2.) Their result implies that the $(1-1/\mathrm{e})$-approximation guarantee for the greedy algorithm extends to a broader class of influence models. On the other hand, Even-Dar and Shapira [30] showed that for the much more restricted *Voter model* [22, 48], the objective function decomposes linearly across nodes, allowing an optimal algorithm or FPTAS, depending on the precise problem definition.

A large number of papers have been written with the aim of reducing the running time for the influence maximization problem. These broadly fall into two categories: (1) heuristics which retain the $(1-1/\mathrm{e})$ approximation guarantee of the greedy algorithm while speeding up the computation in practice (without any worst-case running time improvements, however), and (2) heuristics which guarantee a faster running time, but sacrifice the $(1-1/\mathrm{e})$ approximation guarantee, either providing no guarantees or significantly weaker ones. We give a review of some papers representative of the directions pursued in this domain—a complete review of this literature is impossible due to its sheer volume.

In the first category, the paper by Leskovec et al. [55] proposed the CELF heuristic, based on a lazy evaluation of the objective function. The main insight is the following. If a node's marginal contribution to the objective function in the previous iteration of the greedy algorithm was already smaller than the current best node's, then it need not be reevaluated: by submodularity, its contribution in the current iteration can only be lower. Goyal et al. [38] proposed several additional heuristic optimizations to add to CELF, calling the resulting algorithm CELF++.

In the second category, much of the focus is on speeding up the computation of the objective function via approximations. Several papers use approximations of the influence in terms of the influence only among simple local structures, such as shortest paths [53], local arborescences [81], local neighborhoods [39] or local DAGs [21]. Chen et al. [20] propose reusing previous randomly drawn structures (and the computations on them), as well as a discounted high-degree heuristic. Jung et al. [50] speed up computation of influence by setting up a recurrence relation between the influence of different nodes and linearizing it. Algorithms differing from the greedy addition of one node at a time have been investigated as well. Jiang et al. [49] report significant speedups by using simulated annealing, while Wang et al. [82] propose a pre-processing step breaking the graph into communities, which can then be treated practically separately.

A heuristic which is in spirit similar to those of [20, 21, 39, 81], but comes with provable guarantees, was recently proposed by Borgs et al. [12]. In order to speed up the repeated computation of the influence of sets of nodes, they propose a preprocessing step which generates a random hypergraph sampled according to reverse reachability probabilities in the original graph. Subsequently, the greedy algorithm can be run to solve the maximum coverage problem on the sampled hypergraph, which leads to a near-linear running time of $O((m+n)\varepsilon^{-3}\log n)$ to achieve a $1-1/\mathrm{e}-\varepsilon$ approximation. The (constant) correctness probability can be boosted with repeated invocations. [12] also gives a quantitative tradeoff between faster (sublinear) running time and the deterioration of the approximation guarantee.

In the present article (Section 7), we consider more general marketing strategies than simply selecting a seed set of vertices to activate. In general, a company intent on exploiting social network effects in marketing can combine such effects with differential pricing, as well as possibly offering products to individuals at specific times. This approach has been pursued by several recent papers [1, 2, 5, 43, 45].

The first work along this line was by Hartline et al. [45]. They show that the following "Influence and Exploit" strategies are within a factor $1/4$ of optimal: first, give the product for free to a most influential subset; then, choose prices for the remaining bidders in random order to maximize revenue, ignoring their network effects. By showing that the most influential set in this sense can be approximated within a constant factor, they overall obtain a constant-factor approximation. A slightly modified model was studied by Arthur et al. [5]: in their model, the seller cannot choose arbitrarily whom to offer the product to. Instead, the product spreads by recommendations, but the seller can decide on a price for each individual once a recommendation occurs. In this model, Arthur et al. prove competitive guarantees based on the type of buyer model. Akhlaghpour et al. [2] study a Bayesian model wherein a price trajectory is offered to all buyers (whose values are private), and each individual buys myopically the first time his value for the item exceeds its current price. Under some restrictions, such as about possible price trajectories, they provide algorithms with provable guarantees. Haghpanah et al. [43] consider the design not just of fixed prices, but general auction schemes, and provide a (Bayesian) auction over social networks which achieves a constant approximation to optimal revenue under restricted (all-or-nothing) valuations among the buyers.

The related question of extracting truthfully from individuals how much incentivization they would require to become early adopters and recommenders was studied by Singer [76]. He provides a truthful constant-factor approximation for the problem of choosing and paying individuals to start a cascade.

A different generalization from pricing is suggested in a recent paper by Seeman and Singer [75]. They base their model on the well-known observation that friends of random nodes tend to have higher degrees in expectation than the nodes themselves. Based on this observation, they suggest a two-stage process: first choose several nodes to invite; then, choose a seed set among their friends if they actually are available. They give a constant-factor approximation for this two-stage adaptive optimization process.

The motivations for studying influence maximization also naturally suggest that there will frequently be competition between influences. This competition could occur between companies marketing similar products or between different political movements or ideologies seeking societal acceptance. Indeed, several papers [4, 10, 13, 16, 27, 40, 46, 78] have proposed and analyzed models for competition between multiple cascades.

To our knowledge, Dubey et al. [27] were the first to explicitly propose a model for competing cascades, extending the simple linear model of Richardson and Domingos [68] to competing cascades, and characterizing its equilibria. Subsequently, Carnes et al. [16] and Bharathi et al. [10] proposed extensions of the Independent Cascade model. The main challenge in extending diffusion models to multiple influences is deciding *which* of the influences will convince a node if multiple influences try at once. Carnes et al. [16] propose tie breaking rules under which the best response of the last player remains submodular. Bharathi et al. [10] side-step the issue by introducing a continuous timing component into diffusion, which guarantees that with probability 1, no two influences reach a node at the exact same time. The result is again submodularity of best responses; Bharathi et al. also prove that this makes the competitive cascade game a valid utility game [80], leading to a low Price of Anarchy.

Borodin et al. [13] show that such an extension is not as straightforward for the Linear Threshold model: most natural extensions to multiple influences do not preserve submodularity of the objective function. By adding a fixed sequence in which nodes are considered for state updates, and having them draw new independent thresholds each time, Goyal and Kearns [40] define a threshold-like model that retains many of the desirable properties; in particular, they achieve a Price of Anarchy of 4 for 2 players. He and Kempe [46] subsequently showed that a Price of Anarchy of 2 for arbitrarily many players follows analogously to the result of Bharathi et al. [10], by reducing the game to a valid utility game [80]. Alon et al. [4] and Tzoumas et al. [78] focus on the existence of pure Nash Equilibria in different versions of the game, with 2 [78] or N [4] players. Tzoumas et al. [78] also consider the Price of Anarchy, but in their (deterministic) version of the game, it is unbounded in general, even for 2 players.

Motivated by applications in which the competing cascade is considered "harmful" (such as a harmful rumor, computer virus, or dangerous conviction), several papers [15, 47, 77] have considered the objective of minimizing the spread of a cascade, or maximizing the number of uninfluenced nodes. In this context, like in the previous one, the algorithmic questions studied include the efficient heuristic computation of equilibrium strategies [77] as well as the optimization problem for a single first mover [47]. A slightly different, but related situation arises when negative opinions about a product may emerge on their own as a result of the adoption of a product. The optimization problem under this variant of the model has been studied by Chen et al. [18].

2 Models for the diffusion of an innovation

The social network is represented by a directed graph G, and we write $u \to v$ to denote the existence of a directed edge from u to v. In considering operational models for the spread of an idea or innovation, we will speak of each individual node as being either active (an adopter of the innovation) or inactive. We will focus on settings, guided by the motivation discussed above, in which each node's tendency to become active increases monotonically as more of its neighbors become active.

Also, we will first focus on the *progressive* case in which nodes can switch from being inactive to being active, but do not switch in the other direction. (In Section 6, we show how this assumption can be lifted.) Thus, the process will look roughly as follows from the perspective of an initially inactive node v: as time unfolds, more and more of v's neighbors become active; at some point, this may cause v to become active, and v's decision may in turn trigger further decisions by nodes to which v is connected. We let A_t denote the set of nodes active at time t.

2.1 The Threshold model

Granovetter and Schelling were among the first to propose models that capture such a process; their approach was based on the use of node-specific *thresholds* [41, 74]. Many models of this flavor were subsequently investigated (see, e. g., [9, 41, 57, 58, 61, 67, 74, 79, 84, 85, 86]), but the following *Linear Threshold model* lies at the core of most generalizations. In this model, a node v is influenced by each incoming neighbor w according to a *weight* $b_{v,w} \in [0,1]$. Each node v has a *threshold* θ_v in the interval $[0,1]$; this represents the total weight which has to be exerted upon v by its active neighbors in order for v to become active. The implications of how these thresholds are chosen will be discussed momentarily.

Given a choice of all nodes' thresholds, and an initial set of active nodes A_0 (with all other nodes inactive), the diffusion process unfolds deterministically in discrete *steps*: in step t, all nodes that were active in step $t-1$ remain active; furthermore, each currently inactive node v becomes active if and only if the total weight of its active neighbors is at least θ_v:

$$\sum_{w \to v, w \text{ active}} b_{v,w} \geq \theta_v \,.$$

Thus, the thresholds θ_v intuitively represent the different latent tendencies of nodes to adopt the innovation when their neighbors do.

There are three natural ways to model nodes' thresholds. They could be either hard-wired at a known value (such as 1/2), be part of the input to the optimization problem, or assumed to be random. Models with hard-wired thresholds were studied by Berger [9], Morris [61], and Peleg [67], for example. Unfortunately, we show in Section 3.2 that hard-wired thresholds make the optimization problem hard to approximate to within a multiplicative factor of $n^{1-\varepsilon}$ for any $\varepsilon > 0$; the same hardness results naturally apply to the case in which the thresholds are part of the input. For this reason, and to model our lack of knowledge of the thresholds, we instead assume that the thresholds θ_v are chosen independently and uniformly at random from the interval $[0,1]$; thus, in effect, we average over all possible threshold values.

Given the Linear Threshold model with random node thresholds, the influence maximization problem is now defined as in Section 1. Even with random node thresholds, we can prove that finding the best set A to target is NP-hard.

Theorem 2.1. *The influence maximization problem is* NP*-hard for the Linear Threshold model.*

Proof. Consider an instance of the NP-complete VERTEX COVER problem, defined by an undirected n-node graph $G=(V,E)$ and an integer k; we want to know if there is a set S of k nodes in G so that every edge has at least one endpoint in S. We show that this can be viewed as a special case of the influence maximization problem.

We define a corresponding instance of the influence maximization problem by directing all edges of G in both directions, assigning each directed edge $e=(u,v)$ a weight of $1/\deg v$. If there is a vertex cover S of size k in G, then one can deterministically make $\sigma(A)=n$ by targeting the nodes in the set $A=S$, since each node is either in A or has all its neighbors in A. Conversely, this is the only way to get a set A with $\sigma(A)=n$; for if a pair of adjacent nodes u,v had neither node in A, then with random thresholds close enough to 1, neither node would become active. □

2.1.1 A General Threshold model

The Linear Threshold model can be naturally generalized, by lifting the assumption that influences from individual nodes can only be aggregated linearly. In the general Threshold model, a node v's decision to become active can be based on an *arbitrary monotone function* of the set of neighbors of v that are already active. Each node v has associated with it a monotone *threshold function* f_v mapping subsets of v's neighbor set to real numbers in $[0,1]$, subject to the condition that $f_v(\emptyset)=0$. The diffusion process follows the general structure of the Linear Threshold model. Each node v initially chooses θ_v uniformly at random from the interval $[0,1]$. Now, however, v becomes active in step t iff $f_v(S)\geq\theta_v$, where S is the set of neighbors of v that are active in step $t-1$. Note that the Linear Threshold model is the special case in which each threshold function has the form

$$f_v(S)=\min\Big(1,\sum_{u\in S}b_{v,u}\Big)$$

for parameters $b_{v,u}\in[0,1]$.

2.1.2 The Submodular Threshold model

As we will see below, the influence maximization problem under the general Threshold model is highly intractable. A natural restriction is to require the local influence functions at nodes to have diminishing influence as the size of the influencing set increases. Diminishing influence of individual additions is naturally modeled by *submodularity*, which also plays a central role in our proofs of approximation guarantees. Recall that a function f is *submodular* if it satisfies the following property: the marginal gain from adding an element to a set S is at least as high as the marginal gain from adding the same element to a superset of S. Formally, a submodular function satisfies

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

for all elements v and all pairs of sets $S\subseteq T$.

The Submodular Threshold model is obtained from the general Threshold model by requiring that each $f_v(S)$ be submodular. In the original version of this paper [51], we conjectured that the Submodular

Threshold model would yield the same $1-1/\mathrm{e}$ approximation guarantee for the greedy hill-climbing algorithm as the Linear Threshold model. This conjecture has since been resolved positively by Mossel and Roch [62].

2.2 The Cascade model

Based on work in interacting particle systems [28, 56] from probability theory, we can also consider dynamic *cascade* models for diffusion processes. The conceptually simplest model of this type is what one could call the *Independent Cascade model*, investigated in the context of marketing by Goldenberg, Libai, and Muller [32, 33]. We again start with an initial set of active nodes A_0, and the process unfolds in discrete steps according to the following randomized rule. When node v first becomes active in step t, it is given a single chance to activate each currently inactive neighbor w; it succeeds with a probability $p_{v,w}$—a parameter of the system—independently of the history thus far. (If w has multiple newly activated neighbors, their attempts are sequenced in an arbitrary order.) If v succeeds, then w will become active in step $t+1$; but whether or not v succeeds, it cannot make any further attempts to activate w in subsequent rounds. Again, the process runs until no more activations are possible.

For the Independent Cascade model, too, the problem of finding the best set to target is NP-hard. In fact, for this model, we can even prove an approximation hardness result.

Theorem 2.2. *For the Independent Cascade model, it is* NP*-hard to approximate the most influential set to within better than a factor* $1-1/\mathrm{e}$.

Proof. Consider an instance of the MAXIMUM COVERAGE problem, defined by a collection of subsets $S_1, S_2, \ldots, S_m$ of a ground set $U = \{u_1, u_2, \ldots, u_n\}$; the goal is to select k of the subsets to maximize the size of their union. (We can assume that $k < n < m$.) We show that this can be viewed as a special case of the influence maximization problem.

We define a corresponding directed bipartite graph with $m+n^2$ nodes: for each set S_i, there is a corresponding node i, and for each element u_j, there are n corresponding nodes $j_1, \ldots, j_n$. Whenever $u_j \in S_i$, there is a directed edge (i, j_ℓ) with activation probability $p_{i,j_\ell} = 1$ for all $\ell = 1, \ldots, n$.

If X is a set of k of the subsets S_i, and $T \subseteq U$ the union of the elements covered by the $S_i \in X$, then choosing the k nodes corresponding to X will deterministically activate those nodes and the nodes corresponding to elements in T, for a total of $k+n|T|$ active nodes.

Next, we consider the converse direction, and an initially active set A. Without loss of generality, the set A contains only nodes i corresponding to sets S_i; otherwise, choosing any set covering element j would activate strictly more nodes. Now, let X be the set of sets S_i corresponding to the nodes in A, and $T \subseteq U$ the set of elements covered by X. The number of activated nodes in the IC instance is $k+n|T|$.

In summary, we have shown that in the MAXIMUM COVERAGE instance, r nodes can be covered by k sets if and only if $k+nr$ nodes can be activated by a size-k seed set in the IC instance. As $k \leq n$, we obtain (up to lower order terms) an approximation preserving reduction from MAXIMUM COVERAGE. Since MAXIMUM COVERAGE is known to be hard to approximate better than $1-1/\mathrm{e}$ [31], we obtain the same approximation hardness for influence maximization. □

2.2.1 A General Cascade model

The Independent Cascade model, too, can be naturally generalized, by allowing the probability that u succeeds in activating a neighbor v to depend on the set of v's neighbors that have already tried. Thus, we define an *incremental function* $p_v(u,S) \in [0,1]$, where S is a subset of v's neighbors, and $u \notin S$. A general cascade process works by analogy with the independent case: when u attempts to activate v, it succeeds with probability $p_v(u,S)$, where S is the set of neighbors that have *already* tried (and failed) to activate v. The Independent Cascade model is the special case where $p_v(u,S)$ is a constant $p_{u,v}$, independent of S. We will only be interested in cascade models defined by incremental functions that are *order-independent* in the following sense: if neighbors $u_1, u_2, \ldots, u_\ell$ try to activate v, then the probability that v is activated at the end of these ℓ attempts does not depend on the order in which the attempts are made. More formally, let $S = \{u_1, \ldots, u_{|S|}\}$, and π, ψ two arbitrary permutations of $\{1, \ldots, |S|\}$. Then, we require that

$$\prod_{i=1}^{|S|}(1 - p_v(u_{\pi(i)}, \{u_{\pi(1)}, u_{\pi(2)}, \ldots, u_{\pi(i-1)}\})) = \prod_{i=1}^{|S|}(1 - p_v(u_{\psi(i)}, \{u_{\psi(1)}, u_{\psi(2)}, \ldots, u_{\psi(i-1)}\}))\,.$$

Otherwise, even the definition of the outcome of the activation process would depend on the specific tie-breaking rules for multiple simultaneous activations.

2.2.2 The Decreasing Cascade model

Much like the general Threshold model, the general Cascade model makes influence maximization highly intractable—indeed, we will show in Section 3 that the two general models are equivalent. However, as for the threshold model, a natural "diminishing returns" condition leads to a tractable model again.

In the *Decreasing Cascade model*, the probability of a node u influencing v is non-increasing as a function of the set of nodes that have previously tried to influence v. This means that $p_v(u,S) \geq p_v(u,T)$ whenever $S \subseteq T$. We call this the *diminishing influence condition*. Notice that the Decreasing Cascade model is a generalization of the Independent Cascade model. (On the other hand, in Section 3.1, we show that it is a special case of the Submodular Threshold model.)

Whenever the incremental functions satisfy the diminishing influence condition, we will be able to show that the greedy hill climbing algorithm is a $1-1/\mathrm{e}$ approximation. This is despite the fact that there are functions $p_v(u,S)$ satisfying the diminishing influence condition which do not admit an equivalent formulation in terms of the Triggering Set model at the heart of most of our proofs.

2.3 Node weights

The discussion so far has been treating all nodes as equal, in that the stated goal was to simply maximize the *number* of nodes that are eventually activated in expectation. The model is naturally extended by associating with each node v a non-negative *weight* ω_v, capturing how important it is that v be activated in the final outcome. (For instance, if we are marketing textbooks to college teachers, then the weight could be the number of students in the teacher's class, resulting in a larger or smaller number of sales.) If we let t be the time at which the diffusion process quiesces, we can define the weighted influence function $\sigma_w(A) = \mathbb{E}\left[\sum_{v \in A_t} \omega_v\right] = \sum_v \omega_v \cdot \text{Prob}\left[v \text{ active at time of quiescence}\right]$. The influence function defined previously is the special case obtained by setting $\omega_v = 1$ for all nodes v. As we will see below, most results we obtain for the influence maximization problem carry over directly to the weighted version.

3 Equivalence of models and hardness

The Linear Threshold and Independent Cascade models are two specific, widely studied models for the diffusion of influence. Above, we proposed natural ways of generalizing the models. The goal of such generalizations is to prove approximation guarantees for a broader class of influence models, and to explore the limits of models in which strong approximation guarantees can be obtained. Perhaps somewhat surprisingly, the general Threshold model and the general Cascade model are in fact equivalent. Not only does this equivalence imply that both models can be considered "natural" descriptions of real-world processes, but the two different views of the same random process will also be a key ingredient in our proof of approximation guarantees in Section 5.

We give an explicit method to convert between the two models. Given success probabilities $p_v(u,S)$, we define the activation functions

$$f_v(S) = 1 - \prod_{i=1}^{r} (1 - p_v(u_i, S_i)), \tag{3.1}$$

where $S = \{u_1, u_2, \ldots, u_r\}$, and $S_i = \{u_1, \ldots, u_{i-1}\}$. That f_v is well-defined follows from the order-independence assumption on the $p_v(u,S)$. Conversely, given activation functions f_v, we define success probabilities

$$p_v(u,S) = \frac{f_v(S \cup \{u\}) - f_v(S)}{1 - f_v(S)}. \tag{3.2}$$

It is straightforward to verify that the activation functions defined via Equation (3.1) satisfy Equation (3.2), and the success probabilities defined via Equation (3.2) satisfy Equation (3.1). The equivalence of the models is then captured by the following lemma.

Lemma 3.1. *Assume that the success probabilities $p_v(u,S)$ and activation functions $f_v(S)$ satisfy Equation (3.2). Then, for each node set T and each time t, the probability that exactly the nodes of set T are active at time t is the same under the order-independent cascade process with success probabilities $p_v(u,S)$ and the general threshold process with activation functions $f_v(S)$.*

Proof. We show, by induction, a slightly stronger statement: namely that for each time t and any pair (T, T'), the probability that exactly the nodes of T are active at time $t-1$, and exactly those of T' are active at time t, is the same under both views. By summing over all sets T', this clearly implies the lemma.

At time $t = 0$, the inductive claim holds trivially, as the probability is 1 for the pair $(\emptyset, A_0)$ and 0 for all other pairs, for both processes. For the inductive step to time t, we first condition on the event that the nodes of T are active at time $t-1$, and those of T' at time t.

Consider a node $v \notin T'$. Under the cascade process, v will become active at time $t+1$ with probability

$$1 - \prod_{i=1}^{r} \big(1 - p_v(u_i, T \cup T'_i)\big),$$

where we write $T' \setminus T = \{u_1, \ldots, u_r\}$ and $T'_i = \{u_1, \ldots, u_{i-1}\}$. Under the threshold process, node v becomes active at time $t+1$ iff $f_v(T) < \theta_v \leq f_v(T')$. Because node v is not active at time t, and by the Principle of Deferred Decisions, θ_v is uniformly distributed in $(f_v(T), 1]$ at time t, so the probability that v becomes active is

$$\frac{f_v(T') - f_v(T)}{1 - f_v(T)} .$$

Substituting Equation (3.1) for $f_v(T)$ and $f_v(T')$, a simple calculation shows that

$$\frac{f_v(T') - f_v(T)}{1 - f_v(T)} = 1 - \prod_{i=1}^{r} \big(1 - p_v(u_i, T \cup T'_i)\big) .$$

Thus, each individual node becomes active with the same probability under both processes. As both the thresholds θ_v and activation attempts are independent for distinct nodes, the probability for any set T'' to be the set of active nodes at time $t+1$ is the same under both processes. Finally, as the probability distribution over active sets T'' is the same conditioned on any pair (T, T') of previously active sets, the overall distribution over pairs (T', T'') is the same in both the cascade and threshold processes. $\square$

Lemma 3.1 shows that the general Threshold model is a non-trivial reparametrization of the general Cascade model.[1] In a natural way, it allows us to make all random choices at time 0, before the process starts. An alternate way of attempting to pre-flip all coins, for instance by providing a sequence of random numbers from $[0,1]$ for use in deciding the success of activation attempts, would not preserve order-independence.

For our later proofs, it will be useful to allow delaying the activation of a node whose activation criterion has been met. We therefore generalize the Threshold and Cascade models as follows: each node v has a finite *waiting time* τ_v, meaning that when v's criterion for activation has been met at time t (i. e., an influence attempt was successful in the Cascade model, or $f_v(S) \geq \theta_v$ in the Threshold model), v only becomes active at time $t + \tau_v$. Notice that when $\tau_v = 0$ for all nodes, this is the original Threshold/Cascade model. The following lemma shows that delaying the activation of nodes does not change the eventual outcome under the general Threshold model. Because Lemma 3.1 can be extended straightforwardly to include delays in activation, we obtain the same result for the general Cascade model.

Lemma 3.2. *Under the general Threshold model, the distribution $\varphi(A)$ over active sets at the time of quiescence is the same regardless of the waiting times τ_v. This even holds conditioned upon any random event $\mathcal{E}$.*

Proof. We prove the stronger statement that for every choice of thresholds θ_v, and every vector $\boldsymbol{\tau}$ of waiting times τ_v, the set $S_{\boldsymbol{\tau}}$ of nodes active at the time of quiescence is the same as the set $S_{\mathbf{0}}$ of nodes active at quiescence when all waiting times are 0. This will clearly imply the claim, by integrating over all thresholds that form the event $\mathcal{E}$. So from now on, fix the thresholds θ_v.

Let $A_{0,t}$ denote the set of nodes active at time t when all waiting times are 0, and $A_{\boldsymbol{\tau},t}$ the set of nodes active at time t with waiting times $\boldsymbol{\tau}$. A simple inductive proof using the monotonicity of the activation

[1] Note, however that the Independent Cascade model and Linear Threshold model are different subclasses of this general model, and are not themselves reparametrizations of each other.

functions f_v shows that $A_{\boldsymbol{\tau},t} \subseteq A_{0,t}$ for all times t, which, by setting t to be the time of quiescence of the process with waiting times $\boldsymbol{\tau}$, implies that $S_{\boldsymbol{\tau}} \subseteq S_{\mathbf{0}}$.

Assume now that $S_{\boldsymbol{\tau}} \neq S_{\mathbf{0}}$, and let $T = S_{\mathbf{0}} \setminus S_{\boldsymbol{\tau}} \neq \emptyset$. Among the nodes in T, let v be one that was activated earliest in the process without waiting times, i. e., $T \cap A_{0,t} = \emptyset$, and $v \in T \cap A_{0,t+1}$ for some time t. Because v was activated, we know that $\theta_v \leq f_v(A_{0,t})$, and by definition of v, no previously active nodes are in T, i. e., $A_{0,t} \subseteq S_{\boldsymbol{\tau}}$. But then, the monotonicity of f_v implies that $\theta_v \leq f_v(S_{\boldsymbol{\tau}})$, so v should be active in the process with waiting times $\boldsymbol{\tau}$, a contradiction. □

3.1 Decreasing Cascade model and Submodular Threshold model

We use the insights of the preceding reduction to show that the Decreasing Cascade model is a special case of the Submodular Threshold model.

By Lemma 3.1, the influence function defined via Equation (3.1) gives rise to an instance of the general Threshold model, and the success probabilities $p_v(u,S)$ and activation functions f_v satisfy Equation (3.2), i. e.,

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

The condition that $p_v(u,S) \geq p_v(u,T)$ whenever $S \subseteq T$ now translates to

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

whenever $S \subseteq T$ and $u \notin T$. This is in a sense a "normalized submodularity" property; it is stronger than submodularity, which would consist of the same inequality on just the numerators. (Note that by monotonicity of f_v, the denominator on the left is larger.) Hence, the Decreasing Cascade model is a special case of the Submodular Threshold model.

3.2 Inapproximability results and discussion

The general model proposed above includes large families of instances for which the influence maximization problem is not tractable. Indeed, it may become NP-hard to approximate the optimization problem to within any non-trivial factor.

Theorem 3.3. *In general, it is* NP*-hard to approximate the influence maximization problem to within a factor of $n^{1-\varepsilon}$, for any $\varepsilon > 0$.*

Proof. To prove this result, we reduce from the SET COVER problem. We start with the construction from the proof of Theorem 2.2; however, we do not need duplicates of the element nodes. Specifically, we let $u_1, \ldots, u_n$ denote the nodes corresponding to the n elements; i. e., u_i becomes active when at least one of the nodes corresponding to sets containing u_i is active. Next, for an arbitrarily large constant c, we add $N = n^c$ more nodes $x_1, \ldots, x_N$; each x_j is connected to all of the nodes u_i, and it becomes active only when *all* of the u_i are.

If there are at most k sets that cover all elements, then activating the nodes corresponding to these k sets will activate all of the nodes u_i, and thus also all of the x_j. In total, at least $N + n + k$ nodes will be active. Conversely, assume that there is no set cover of size k, and consider a seed set A. Without loss of generality, A contains no node u_i, since such a node can always be replaced with the node corresponding to any set containing element i, which will still result in activating u_i, and possibly other nodes as well. Because there is no set cover of size k, A cannot activate all of the u_i, and hence none of the x_j will become active (unless targeted). In particular, fewer than $n+k$ nodes are active in the end. If an algorithm could approximate the problem within $n^{1-\varepsilon}$ for any ε, it could distinguish between the cases where $N+n+k$ nodes are active in the end, and where fewer than $n+k$ are. But this would solve the underlying instance of SET COVER, and therefore is impossible assuming $\mathsf{P} \neq \mathsf{NP}$. □

Note that our inapproximability result holds in a very simple model, in which each node is "hard-wired" with a fixed threshold.[2] It is thus worth briefly considering the general issue of performance guarantees for algorithms under the above models. For both the Linear Threshold and the Independent Cascade models, the influence maximization problem is NP-complete, but, as we prove in the next section, it can be approximated well. In the linear model of Richardson and Domingos [68], on the other hand, both the propagation of influence *as well as* the effect of the initial targeting are linear. Initial marketing decisions here are thus limited in their effect on node activations; each node's probability of activation is obtained as a linear combination of the effect of targeting and the effect of the neighbors. In this fully linear model, the influence can be maximized by solving a system of linear equations.

In contrast, generalizing Theorem 3.3, we can show that general models like that of Domingos and Richardson [26], and even simple models that build in a fixed threshold (like $1/2$) at all nodes [9, 61, 67], lead to influence maximization problems that cannot be approximated to within any non-trivial factor, assuming $\mathsf{P} \neq \mathsf{NP}$. Our analysis of approximability thus suggests a way of tracing out a more delicate boundary of tractability through the set of possible models, by helping to distinguish among those for which simple heuristics provide strong performance guarantees and those for which they can be arbitrarily far from optimal. This in turn can suggest the development of both more powerful algorithms, and the design of accurate models that simultaneously allow for tractable optimization.

4 Approximation algorithm and analysis

In this section, we describe and analyze the approximation algorithm for the influence maximization problem. Using approximation results for submodular functions, we will establish Theorem 1.1, our main theorem.

The approximation algorithm we use and analyze is a simple greedy hill-climbing algorithm. Starting from the empty set, it repeatedly adds to A the node x maximizing the *marginal gain* $\sigma(A \cup \{x\}) - \sigma(A)$. Note that it is not clear how to evaluate $\sigma(\cdot)$ exactly in polynomial time; indeed, subsequent to the initial publication of this work [51], it has been proven that evaluating $\sigma(A)$ is generally #P-complete both for the Linear Threshold model [21] and the Independent Cascade model [81].

However, it is possible to obtain arbitrarily good approximations to $\sigma(\cdot)$ in polynomial time, simply by simulating the random choices and diffusion process sufficiently many times. Because $1 \leq \sigma(A) \leq n$

[2] Chen [17] showed, subsequent to the conference version of the present work, that with fixed node thresholds, the goal of minimizing the number of seed nodes required to reach all or a large fraction of the network, is hard to approximate within poly-logarithmic factors.

for all sets A, standard Chernoff–Hoeffding bounds (e. g., Theorem 2.3 from [60]) imply the following proposition.[3]

Proposition 4.1. *If the diffusion process starting with A is simulated independently at least*

$$\Omega\Big(\frac{n^2}{\varepsilon^2}\ln(1/\delta)\Big)$$

times, then the average number of activated nodes over these simulations is a $(1 \pm \varepsilon)$-approximation to $\sigma(A)$, with probability at least $1-\delta$.

Proof. Assume that the sampling process is repeated $T = \Omega((n^2/\varepsilon^2)\ln(1/\delta))$ times, and let $X_1,\ldots,X_T \in [0,1]$ be the *fraction* of nodes activated in each of the runs. The estimate of $\sigma(A)$ is then

$$\frac{n}{T}\cdot\sum_{i=1}^{T} X_i\,.$$

Writing $X := \sum_{i=1}^{T} X_i$, the expectation of X is exactly $(T/n)\cdot\sigma(A)$, and standard bounds (e. g., Theorem 2.3 from [60]) give that

$$\operatorname{Prob}\left[\left|X - \frac{T}{n}\cdot\sigma(A)\right| \geq T\gamma\right] \leq 2e^{-2T\gamma^2}\,.$$

Substituting $\gamma = (2\varepsilon/n)\sigma(A) \geq 2\varepsilon/n$ bounds the right-hand side by δ. □

We can now specify the greedy approximation algorithm formally as Algorithm 1:

Algorithm 1 Greedy Approximation Algorithm

1: Start with $A = \emptyset$.
2: **while** $|A| \leq k$ **do**
3: For each node x, use repeated sampling to approximate $\sigma(A \cup \{x\})$ to within $(1 \pm \varepsilon)$ with probability $1-\delta$.
4: Add the node with largest estimate for $\sigma(A \cup \{x\})$ to A.
5: **end while**
6: Output the set A of nodes.

To obtain good approximation guarantees for the influence maximization problem in a particular model of diffusion, we use a two-step strategy. First, we show that the influence function $\sigma(\cdot)$ is a submodular function of the initial set of active nodes A. (Recall the definition of submodularity from Section 2.1.2.)

Once we have proved submodularity, we can apply the following theorem of Nemhauser, Wolsey and Fisher [24, 65].

[3] The assumption that $\sigma(A) \leq n$ holds for the *unweighted model*. When there are nodes with very high (exponential in n) weight and very small influence probability leading to them, sampling may lead to very inaccurate results unless the number of samples is polynomially large in the weights, which is *pseudo-polynomial* in the input size. How well $\sigma_\omega(A)$ can be approximated for an influence model with arbitrary node weights remains an interesting open question.

Theorem 4.2 ([24, 65])**.** *Let $\sigma(\cdot)$ be a non-negative monotone submodular function. Then the greedy algorithm that (for k iterations) adds an element with the largest marginal increase in $\sigma(\cdot)$ produces a k-element set A such that $\sigma(A) \geq (1-1/\mathrm{e}) \cdot \max_{|B|=k} \sigma(B)$.*

Due to its generality, this result has found applications in a number of areas of discrete optimization (see, e. g., [44, 64]). In particular, it has become a very popular technique for optimization with provable approximation guarantees since the original publication of the present article; see the survey by Krause and Golovin [54] for a recent overview.

Theorem 4.2 assumes that the function $\sigma(\cdot)$ to be optimized can be evaluated exactly at any point. As we argued above, $\sigma(\cdot)$ can in general not be evaluated exactly in polynomial time. However, as shown above, by simulating the diffusion process and sampling the resulting active sets, we are able to obtain arbitrarily close approximations to $\sigma(A)$, with high probability. It is a fairly straightforward extension of Theorem 4.2 that by using $(1 \pm \delta)$-approximate values for the function to be optimized, we obtain a $(1-1/\mathrm{e}-\varepsilon)$-approximation, where ε depends on δ and goes to 0 as $\delta \to 0$. To finish the proof of Theorem 1.1, it thus remains to show that the objective function $\sigma(\cdot)$ is non-negative, monotone and submodular.

4.1 The triggering set technique

Primarily for the purpose of analysis,[4] we define the following *Triggering model*:

Definition 4.3 (Triggering model)**.** Each node v independently chooses a random "triggering set" T_v according to some distribution over subsets of its incoming neighbors. Initially, a set A is activated. Subsequently, an inactive node v becomes active in step t if it has a neighbor in its chosen triggering set T_v that is active in step $t-1$.

Thus, compared to, say, the Threshold model, v's threshold has been replaced by a latent subset T_v of neighbors whose behavior *actually* affects v. It is useful to think of the triggering sets T_v in terms of "live" and "blocked" edges: if node u belongs to the triggering set T_v of v, then we declare the edge (u,v) to be live, and otherwise we declare it to be blocked. Since we are only interested in the node set active at the end of the process, Definition 4.3 is equivalent to stating that A_n is the set of nodes v such that v is reachable from A via a path consisting entirely of live edges.

The key fact about the Triggering model is that any of its instances will directly lead to a submodular influence function:

Lemma 4.4. *In every instance of the Triggering model, the influence function $\sigma(\cdot)$ is submodular.*

Proof. For any outcome $X = (T_v)_v$ of the choices of live and blocked edges, let $\sigma_X(\cdot)$ denote the number of nodes reached with this particular outcome; note that $\sigma_X(S)$ is a deterministic quantity. First, we claim that for each fixed outcome X, the function $\sigma_X(\cdot)$ is submodular. To see this, let S and T be two sets of nodes such that $S \subseteq T$, and consider the quantity $\sigma_X(S \cup \{v\}) - \sigma_X(S)$. Let $R(v,X)$ be the set of nodes that are reachable from v given the edge choices X. Then, $\sigma_X(S \cup \{v\}) - \sigma_X(S)$ is the number of elements in $R(v,X)$ that are not already in the union $\bigcup_{u \in S} R(u,X)$; it is at least as large

[4] The technique can also be helpful for evaluating the objective function $\sigma(A)$ more efficiently.

as the number of elements in $R(v,X)$ that are not in the (bigger) union $\bigcup_{u\in T} R(u,X)$. It follows that $\sigma_X(S\cup\{v\}) - \sigma_X(S) \geq \sigma_X(T\cup\{v\}) - \sigma_X(T)$, which is the defining inequality for submodularity.[5]

When we instead consider distributions over outcomes X, we can write

$$\sigma(A) = \sum_{\text{outcomes } X} \text{Prob}\,[X]\cdot\sigma_X(A)\,,$$

since the expected number of nodes activated is just the weighted average over all outcomes. But a non-negative linear combination of submodular functions is also submodular, and hence $\sigma(\cdot)$ is submodular, which concludes the proof. □

With Lemma 4.4 in hand, our strategy for establishing submodularity of the Independent Cascade and Linear Threshold models will be to exhibit instances of the Triggering model which have identical distributions over activated sets. We will also see other natural models that can be analyzed using this technique, but (in Section 5), we will see that for the Decreasing Cascade model, no equivalent instance of the Triggering model exists, even though the activation functions are submodular.

4.2 Independent Cascade

In this section, we will establish the following theorem:

Theorem 4.5. *For an arbitrary instance of the Independent Cascade model, the resulting influence function $\sigma(\cdot)$ is submodular.*

Proof. We establish this theorem by giving an equivalent instance of the Triggering model; the result then follows directly from Lemma 4.4. To derive the Triggering model instance, consider a point in the cascade process when node v has just become active, and it attempts to activate its neighbor w, succeeding with probability $p_{v,w}$. We can view the outcome of this random event as being determined by flipping a coin of bias $p_{v,w}$. From the point of view of the process, it clearly does not matter whether the coin was flipped at the moment that v became active, or whether it was flipped at the very beginning of the whole process and is only being revealed now. Continuing this reasoning, we can in fact assume that for *each* pair of neighbors (v,w) in the graph, a coin of bias $p_{v,w}$ is flipped at the very beginning of the process (independently of the coins for all other pairs of neighbors), and the result is stored so that it can be later checked *in the event* that v is activated while w is still inactive.

Thus, an equivalent view of the process is as follows: each edge (v,w) is present in the graph G of live edges independently with probability $p_{v,w}$, and absent otherwise. The preceding argument shows that conditioned on the outcome of the coin flips for each edge (v,w), a node u is reachable from A in the resulting graph if and only if it is activated by the Independent Cascade process with the *same* coin flip outcomes. Since the outcomes are pointwise the same, the expected number (or weight) of nodes reachable from A in the Triggering model is the same as the expected number of nodes activated in the Independent Cascade model. □

[5]Notice that the argument in this paragraph did not make use of the fact that each node contributed exactly 1 to the objective function. It therefore carries over verbatim to the weighted version with non-negative weights, establishing the same result.

4.3 Linear Threshold

We now prove an analogous result for the Linear Threshold model.

Theorem 4.6. *For instances of the Linear Threshold model in which $\sum_{w:w\to v} b_{v,w} \leq 1$ for all nodes v, the resulting influence function $\sigma(\cdot)$ is submodular.*

Proof. We again use Lemma 4.4, exhibiting an equivalent instance of the Triggering model. In this case, the analysis is a bit more intricate than in the proof of Theorem 4.5, because there is no clear "pointwise" equivalent process.

Recall that each node v has an influence weight $b_{v,w} \geq 0$ from each of its neighbors w, and we are assuming that $\sum_w b_{v,w} \leq 1$. (We can extend the notation by writing $b_{v,w} = 0$ when w is not a neighbor of v.) The equivalent instance of the Triggering model we analyze is the following: v picks at most one of its incoming edges at random, selecting the edge from w with probability $b_{v,w}$ and selecting no edge with probability $1-\sum_w b_{v,w}$. Thus, each set T_v is either empty or contains exactly one edge. Note the contrast with the proof of Theorem 4.5: there, we determined whether an edge was live independently of the decision for each other edge; here, we negatively correlate the decisions so that at most one live edge enters each node.

To show the equivalence of the models formally, we will establish that for any given targeted set A, the following two distributions over sets of nodes are the same:

1. The distribution over active sets obtained by running the Linear Threshold process to completion starting from A; and

2. The distribution over sets reachable from A via live-edge paths, under the random selection of live edges defined above.

To obtain some intuition, it is useful to first analyze the special case in which the underlying graph G is directed and acyclic. In this case, we can fix a topological ordering of the nodes $v_1, v_2, \ldots, v_n$ (so that all edges go from earlier nodes to later nodes in the order), and build up the distribution of active sets by following this order. For each node v_i, suppose that we have already determined the distribution over active subsets of its neighbors. Then under the Linear Threshold process, the probability that v_i will become active, given that a subset S_i of its neighbors is active, is $\sum_{w\in S_i} b_{v_i,w}$. This is precisely the probability that the live incoming edge selected by v_i lies in S_i, and so inductively, we see that the two processes define the same distribution over active sets.

To prove the claim generally, consider a graph G that is not acyclic. It becomes trickier to show the equivalence, because there is no natural ordering of the nodes over which to perform induction. Instead, we argue by induction over the iterations of the Linear Threshold process. We define A_t to be the set of active nodes at the end of iteration t, for $t = 0, 1, 2, \ldots$. (Note that A_0 is the set initially targeted.) If node v has not become active by the end of iteration t, then the probability that it becomes active in iteration $t+1$ is equal to the probability that the influence weights in $A_t \setminus A_{t-1}$ push it over its threshold, *given* that its threshold was not exceeded already; this probability is

$$\frac{\sum_{u\in A_t\setminus A_{t-1}} b_{v,u}}{1-\sum_{u\in A_{t-1}} b_{v,u}}.$$

On the other hand, we can run the Triggering process by revealing the identities of the live edges gradually as follows. We start with the targeted set A. For each node v with at least one edge from the set A, we determine whether v's live edge comes from A. If so, then v is reachable; but if not, we keep the source of v's live edge unknown, subject to the condition that it comes from outside A. Having now exposed a new set of reachable nodes A'_1 in the first stage, we proceed to identify further reachable nodes by performing the same process on edges from A'_1, and in this way produce sets $A'_2, A'_3, \ldots$. If node v has not been determined to be reachable by the end of stage t, then the probability that it is determined to be reachable in stage $t+1$ is equal to the probability that its live edge comes from $A_t \setminus A_{t-1}$, given that its live edge has not come from any of the earlier sets. But this is

$$\frac{\sum_{u \in A_t \setminus A_{t-1}} b_{v,u}}{1 - \sum_{u \in A_{t-1}} b_{v,u}},$$

which is the same as in the Linear Threshold process of the previous paragraph. Thus, by induction over these stages, we see that the Triggering process produces the same distribution over active sets as the Linear Threshold process. □

Even when $\sum_{w: w \to v} b_{v,w} > 1$ for some nodes v, the result of Mossel and Roch [62] implies that the objective function $\sigma(\cdot)$ is submodular, albeit not via a Triggering process proof. The reason is that the function

$$f(S) = \min\Big(1, \sum_{w \in S: w \to v} b_{v,w}\Big)$$

is submodular for every node v. Mossel and Roch proved as their main result that this is sufficient for $\sigma(\cdot)$ to be submodular. Thus, we obtain submodularity for all instances of the Linear Threshold model.

4.4 Other models

Beyond the Independent Cascade and Linear Threshold, there are other natural special cases of the Triggering model. One example is the "Only-Listen-Once" model. Here, each node v has a parameter p_v so that the first neighbor of v to be activated causes v to become active with probability p_v, and all subsequent attempts to activate v deterministically fail. (In other words, v only listens to the first neighbor that tries to activate it.) This process has an equivalent formulation in the Triggering Set model, with an edge distribution defined as follows: for any node v, the triggering set T_v is either the entire neighbor set of v (with probability p_v) or the empty set.

As a result, the influence function in the Only-Listen-Once model is also submodular, and we can obtain a $(1-1/\mathrm{e}-\varepsilon)$-approximation here as well.

5 The Decreasing Cascade model

In this section, we investigate the Decreasing Cascade model (defined in Section 2.2.2) in detail. First, we show that the Decreasing Cascade model is *not* an instance of the Triggering model, implying that the Triggering Set technique cannot be used to establish the submodularity of the objective function in this case. Nonetheless, using a more involved coupling argument and the Principle of Deferred Decisions, we will be able to show that the objective function is in fact submodular, and hence, the $1-1/e$ approximation guarantee of the greedy algorithm applies to the Decreasing Cascade model as well.

5.1 Relationship to the Triggering Model

We first show an instance of the Decreasing Cascade model that is not equivalent to any instance of the Triggering model. This example shows that the Triggering Set Technique cannot be applied to show submodularity of the Decreasing Cascade model in general.

Our example has five nodes. Node v could potentially be influenced by four nodes $u_1,\dots,u_4$. The first two nodes to try activating v have a probability of $1/2$ each to succeed, whereas all subsequent attempts fail. The influences are thus $p_v(u_i,S)=1/2$ whenever $|S|<2$, and $p_v(u_i,S)=0$ otherwise. Notice that this is indeed an instance of the Decreasing Cascade model, and order-independent.

Assume, for contradiction, that there is a distribution on graphs such that node v is reachable from a set S with the same probability that S will activate v in the Cascade model. For any set $S\subseteq\{1,2,3,4\}$, let q_S denote the probability that in this distribution over graphs, exactly the edges from u_i to v for $i\in S$ are present. Because with probability $1/4$, v does not become active even if all u_i are, we know that $q_\emptyset=1/4$. If u_1,u_2,u_3 are active, then v is also active with probability $3/4$, so the edge (u_4,v) can never be present all by itself. (If it were, then the set $\{u_1,u_2,u_3,u_4\}$ together would have higher probability of reaching v than the set $\{u_1,u_2,u_3\}$.) Thus, we have that $q_{\{i\}}=0$ for all i. The same argument shows that $q_{\{i,j\}}=0$ for all i,j.

Thus, the only non-empty edge sets with non-zero probabilities can be those of size three or four. If node u_1 is the only active node, then v will become active with probability $1/2$, so the edge (u_1,v) is present with probability $1/2$. Hence,

$$q_{\{1,2,3\}}+q_{\{1,2,4\}}+q_{\{1,3,4\}}+q_{\{1,2,3,4\}}=\frac{1}{2},$$

while

$$q_{\{1,2,3\}}+q_{\{1,2,4\}}+q_{\{1,3,4\}}+q_{\{2,3,4\}}+q_{\{1,2,3,4\}}=1-q_\emptyset=\frac{3}{4}.$$

Therefore, $q_{\{2,3,4\}}=1/4$, and a similar argument for nodes u_2,u_3,u_4 gives that $q_S=1/4$ for each set S of cardinality 3. But then, the total probability mass on edge sets is at least $5/4$, as there are four such sets S, and the empty set also has probability $1/4$. This is a contradiction, so there is no such distribution over graphs.

The example above raises the interesting question of which instances of the General Threshold or Cascade model are in fact instances of the Triggering model, for a suitably defined distribution over graphs. In other words, for which models does the Triggering Set technique imply submodularity? Subsequent to the publication of the conference versions of the present work, a complete characterization of such models was provided by Salek et al. [73]: an instance of the general threshold model has an equivalent Triggering model formulation if and only if each node's activation function f_v has discrete derivatives of alternating signs. More specifically, the derivative of a set function g with respect to an element v is defined as $g_v(S)=g(S\cup\{v\})-g(S)$; higher derivatives are defined as derivatives of derivatives. The precise necessary and sufficient condition for the Triggering model is then that all functions f_v be non-negative and decreasing, all even derivatives be non-positive, and all odd derivatives be non-negative.

5.2 Proof of submodularity

We now establish the following theorem:

Theorem 5.1. *If all influence functions $p_u(v,S)$ are non-increasing in S, then $\sigma(\cdot)$ is monotone and submodular.*

Proof. The monotonicity is an immediate consequence of Lemma 5.2 below, applied with $V = V'$ and $p'_v(u,S) = p_v(u,S)$ for all S,v,u. So we focus on submodularity for the remainder of the proof. We have to show that, whenever $A \subseteq A'$, we have $\sigma(A \cup \{x\}) - \sigma(A) \geq \sigma(A' \cup \{x\}) - \sigma(A')$, for any node $x \notin A'$.

The basic idea of the proof is to characterize $\sigma(A \cup \{x\}) - \sigma(A)$ in terms of a *residual process* which targets only the node x, and has appropriately modified success probabilities (similarly for $\sigma(A' \cup \{x\}) - \sigma(A')$). Let $\varphi(A)$ denote the *set* of nodes active at the time of quiescence (a random variable). To show that the residual processes indeed have the same distributions over final active sets $\varphi(\{x\})$ as the original processes, we use Lemma 3.2.

Given a node set B, we define the *residual process* on the set $V \setminus B$: the success probabilities are

$$p_v^{(B)}(u,S) := p_v(u, S \cup B),$$

and the only node targeted is x, targeted at time 1. Let $\Delta_B(x)$ denote the set of nodes active at the time of quiescence of the residual process; notice that this is a random variable. We claim that, conditioned on the event that $[\varphi(A) = B]$, the variable $\Delta_B(x)$ has the same distribution as the variable $\varphi(A \cup \{x\}) \setminus \varphi(A)$.

In order to prove this fact, we focus on the threshold interpretation of the process, and assign node x a waiting time of $\tau_x = n+1$. By Lemma 3.2, this view does not change the distribution of $\varphi(A \cup \{x\}) \setminus \varphi(A)$. Then, x is the only node at time $n+1$ which has not exhausted its activation attempts; by the conditioning, the other active nodes (which have exhausted their activation attempts) are those from B. This implies that only nodes from $V \setminus B$ will make activation attempts after time $n+1$. By using the same order of activation attempts, and the same coin flips for each pair $u,v \in V \setminus B$, a simple inductive proof on the time t shows that the set S of nodes is active in the residual process at time t if and only if the set $S \cup B$ is active in the original process at time $n+t$. In particular, this shows that the two random variables have the same distributions.

Having shown this equivalence, we want to compare the expected sizes of $\Delta_B(x)$ and $\Delta_{B'}(x)$, when $B \subseteq B'$. We write

$$\sigma_B(x) = \mathbb{E}[|\Delta_B(x)|] \qquad \text{as well as} \qquad \sigma_{B'}(x) = \mathbb{E}[|\Delta_{B'}(x)|].$$

First off, notice that the node set $V \setminus B$ of the former process is a superset of $V \setminus B'$. Furthermore, for all nodes u,v and node sets S, the decreasing cascade condition implies that

$$p_v^{(B)}(u,S) = p_v(u, S \cup B) \;\geq\; p_v(u, S \cup B') \;=\; p_v^{(B')}(u,S).$$

Lemma 5.2 below proves the intuitively obvious fact that the combination of a larger ground set of nodes and larger success probabilities results in a larger set of activated nodes, i. e.,

$$\sigma_x(B) \geq \sigma_x(B'). \tag{5.1}$$

Finally, we can rewrite the expected number of active nodes as

$$\begin{aligned}
\sigma(A \cup \{x\}) - \sigma(A) &= \sum_{B} \sigma_x(B) \cdot \text{Prob}\,[\varphi(A) = B] \\
&= \sum_{B} \sum_{B' \supseteq B} \sigma_x(B) \cdot \text{Prob}\left[\varphi(A) = B, \varphi(A') = B'\right] \\
&\geq \sum_{B} \sum_{B' \supseteq B} \sigma_x(B') \cdot \text{Prob}\left[\varphi(A) = B, \varphi(A') = B'\right] \\
&= \sum_{B'} \sigma_x(B') \cdot \text{Prob}\left[\varphi(A') = B'\right] \\
&= \sigma(A' \cup \{x\}) - \sigma(A').
\end{aligned}$$

The inequality followed by applying Inequality (5.1) under the sum. In both of the steps surrounding the inequality, we used that $\text{Prob}\,[\varphi(A) = B, \varphi(A') = B'] = 0$ whenever $B \not\subseteq B'$, by the monotonicity of the cascade process. This completes the proof of submodularity.[6] □

Lemma 5.2. *Let $V' \subseteq V$, and assume that $p'_v(u,S) \leq p_v(u,S)$ for all nodes $u, v \in V$ and all sets S. If $A' \subseteq A$ are the targeted sets for cascade processes on V' and V, then the expected size of the active set at the end of the process on V is no smaller than the corresponding expected size for the process on V'.*

Proof. This claim is most easily seen in the threshold view of the process. Equation (3.1) shows that the activation functions f'_v, f_v corresponding to the success probabilities $p'_v(u,S)$ and $p_v(u,S)$ satisfy $f'_v(S) \leq f_v(S)$, for all nodes v and sets S. Then, for any fixed thresholds θ_v, a simple inductive proof on time steps t shows that the set of active nodes in the former process (with functions f'_v) is always a subset of the set of active notes in the latter one (with functions f_v). Since the inequality thus holds for every point of the probability space, it holds in expectation. □

We remark here again that subsequent to the original publication of our work, a stronger version of Theorem 5.1 was established by Mossel and Roch [62], who showed that in the general Threshold model, whenever the activation functions $f_v(\cdot)$ are monotone and submodular, so is $\sigma(\cdot)$. As we showed in Section 3.1 that the Decreasing Cascade model is equivalent to a (special case of the) Threshold model with submodular activation functions, the theorem of [62] implies Theorem 5.1.

6 Non-progressive processes

We have thus far been concerned with the *progressive* case, in which nodes only go from inactivity to activity, but not vice versa. When considering the *non-progressive* case, in which nodes can switch in

[6] In fact, Lemma 5.2 shows a stronger result: pointwise over random choices, the activated set under B is a superset of the one under B'. Using this fact in the preceding calculations shows that submodularity also holds with non-negative node weights, as we established earlier using the Triggering Set technique for the Independent Cascade and Linear Threshold models.

both directions, the first question is how to exactly define the process and the objective function, since the "number of active nodes" is not a well-defined quantity any more. In this section, we consider what is likely the most natural choice of the process and objective function, and show that with these choices, the non-progressive case can in fact be reduced to the progressive case.

Because the local threshold functions f_v are monotone, the Threshold model is inherently progressive, as is the Cascade model. Perhaps the simplest way of defining a non-progressive version of the process is to have each node choose a *new* threshold $\theta_v^{(t)}$ in each time step t.[7] Thus, a previously active node can become inactive as a result of choosing a larger threshold. Formally, at each step t, each node v chooses a new value $\theta_v^{(t)}$ independently and uniformly at random from the interval $[0,1]$. Node v will be active in step t iff $f_v(S_{t-1}) \geq \theta_v^{(t)}$, where S_{t-1} is the set of neighbors of v that are active in step $t-1$.

When nodes can switch from active to inactive, care must be taken in defining the objective function. A natural choice is to define $\sigma(A)$ as the sum, over all nodes v, of the number of time steps during which v is active. This definition is motivated by considering an active node as a potential customer in that time step, such as an instructor adopting a textbook during a particular semester. This definition is naturally generalized by assigning different weights to nodes in different time steps, such as exponential time-discounting common in economics.

Compared to the progressive model, it may now also be beneficial to choose *times* for particular activations. More formally, we assume that there is a time horizon τ for which the process will be run. We define an *intervention* as the activation of a particular node v at a particular time $t \leq \tau$. (Notice that v itself may quickly de-activate, but we hope to create a large "ripple effect.") Simple examples show that to maximize influence, one should not necessarily perform all k interventions at time 0; for example, G may not even have k nodes. Thus, from the perspective of influence maximization, we now ask the following question: Suppose that we have a non-progressive model that is going to run for τ steps, and during this process, we are allowed to make up to k interventions. Which k interventions should we perform? The *influence maximization problem* in the non-progressive Threshold model is to find the k interventions with maximum influence.

We can show that the non-progressive influence maximization problem reduces to the progressive case in a different graph. Given a graph $G=(V,E)$ and a time limit τ, we build a layered graph G^τ on $\tau \cdot |V|$ nodes: there is a copy v_t for each node v in G and each time step $t \leq \tau$. Let

$$L_t = \{v_t \mid v \in V\}$$

be the t^{th} layer of G^τ. For each node v_t, we define its influence function to be

$$f'_{v_t}(S) = f_v(\{u \mid u_{t-1} \in S\})\,;$$

this function simply applies v's original influence function to the set of nodes that are active in step $t-1$.[8]

[7] At the other extreme would be for each node to have the same threshold throughout. This model would be identical to the Linear Threshold model as we have defined it previously. A more realistic model would posit a strong but not perfect correlation between thresholds for different time steps; we leave the exploration of such models to future work.

[8] If different time steps carry different weights, then we simply assign each node v_t the corresponding weight; see the discussion in Section 2.3.

Theorem 6.1. *The non-progressive influence maximization problem on G over a time horizon τ is equivalent to the progressive influence maximization problem on the layered graph G^τ. Node v is active at time t in the non-progressive process if and only if v_t is activated in the progressive process.*

Proof. The proof simply couples the thresholds of nodes v_t with the threshold chosen by node v in step t in the non-progressive model. Then, the theorem follows by a straightforward induction proof on t. □

Thus, models where we have approximation algorithms for the progressive case carry over. Theorem 6.1 also implies approximation results for certain non-progressive models used by Asavathiratham et al. to model cascading failures in power grids [6, 7].

Note that the non-progressive model discussed here differs from the model of Domingos and Richardson [26, 68] in two ways. We are concerned with the sum over all time steps $t \leq \tau$ of the (possibly weighted) expected number of active nodes at time t, for a *given a time limit* τ, while [26, 68] study the limit of this process: the expected number of nodes active at time t as t goes to infinity. Further, we consider interventions for a particular node v, at a particular time $t \leq \tau$, while the interventions considered by [26, 68] permanently affect the activation probability function of the targeted nodes.

7 General marketing strategies

In the formulation of the problem, we have so far assumed that for one unit of budget, we can deterministically target any node v for activation. This is clearly a highly simplified view. In a more realistic scenario, we may have a number m of different *marketing actions* M_i available, each of which may affect some subset of nodes by *increasing* their probabilities of becoming active, without necessarily making them active deterministically. The more we spend on any one action, the stronger its effect will be; however, different nodes may respond to marketing actions in different ways.

In a general model, we choose *investments* x_i into marketing actions M_i, such that the total investments do not exceed the budget. A *marketing strategy* is then an m-dimensional vector $\boldsymbol{x}$ of investments. The probability that node v will become active is determined by the strategy, and denoted by $h_v(\boldsymbol{x})$. We assume that this function is non-decreasing and satisfies the following "diminishing returns" property for all $\boldsymbol{x} \geq \boldsymbol{y}$ and $\boldsymbol{a} \geq \mathbf{0}$ (where we write $\boldsymbol{x} \geq \boldsymbol{y}$ or $\boldsymbol{a} \geq \mathbf{0}$ to denote that the inequalities hold in all coordinates):

$$h_v(\boldsymbol{x}+\boldsymbol{a}) - h_v(\boldsymbol{x}) \leq h_v(\boldsymbol{y}+\boldsymbol{a}) - h_v(\boldsymbol{y})\,. \tag{7.1}$$

Intuitively, inequality (7.1) states that any marketing action is more effective when the targeted individual is less "marketing-saturated" at that point; it captures concavity of the activation functions in non-negative directions.

We are trying to maximize the expected size of the final active set. As a function of the marketing strategy $\boldsymbol{x}$, each node v becomes active independently with probability $h_v(\boldsymbol{x})$, resulting in a (random) set of initial active nodes A. Given the initial set A, the expected size of the final active set is $\sigma(A)$.[9]

[9] Recently, a slightly different model of partial influence by a marketer has been proposed by Günneç and Raghavan (see [42]) and Demaine et al. [25]. In their model, a marketer also divides a budget among actions, but each action only affects a single individual. Their model is best understood in the threshold model: a node v on whom the marketer spent x_v units of budget becomes active when $\theta_v \leq x_v + f_v(S)$. Thus, the effort of the marketer not only creates a possibility of initial activation, but permanently increases a node's propensity for being activated. The models are technically incomparable.

The expected revenue of the marketing strategy $\boldsymbol{x}$ is therefore

$$g(\boldsymbol{x}) = \sum_{A \subseteq V} \sigma(A) \cdot \prod_{u \in A} h_u(\boldsymbol{x}) \cdot \prod_{v \notin A} (1 - h_v(\boldsymbol{x}))\,.$$

In order to (approximately) maximize g, we assume that we can evaluate the function at any point $\boldsymbol{x}$ approximately, and find a direction i with approximately maximal gradient. Specifically, let $\boldsymbol{e}_i$ denote the unit vector along the i^{th} coordinate axis, and δ be some constant. We assume that there exists some $\gamma \leq 1$ such that we can find an i with

$$g(\boldsymbol{x} + \delta \cdot \boldsymbol{e}_i) - g(\boldsymbol{x}) \geq \gamma \cdot \big(g(\boldsymbol{x} + \delta \cdot \boldsymbol{e}_j) - g(\boldsymbol{x})\big)$$

for each j. We divide each unit of the total budget k into equal parts of size δ. Starting with an all-0 investment $\mathbf{0}$, we perform an approximate gradient ascent, by repeatedly (a total of k/δ times) adding δ units of budget to the investment in the action M_i that approximately maximizes the gradient. Formally, the algorithm is defined as Algorithm 2.

Algorithm 2 The Hill Climbing algorithm

```
Start with x^(0) = 0.
for all rounds t = 0, ..., k·δ^(-1) − 1 do
    Let i_t be a direction maximizing g(x^(t) + δ·e_{i_t}) − g(x^(t)).
    Set x^(t+1) = x^(t) + δ·e_{i_t}.
end for
```

The proof that this algorithm gives a good approximation consists of two steps. First, we show that whenever a function g is non-negative, non-decreasing, and satisfies the "diminishing returns" condition (7.1), the hill-climbing algorithm gives a constant-factor approximation. Then, we show that the specific function g we are trying to optimize is indeed non-negative, non-decreasing, and satisfies Condition (7.1).

Theorem 7.1. *Let g be a non-negative and monotone non-decreasing function satisfying the "diminishing returns" condition (7.1). Assume that the hill climbing algorithm is run with a total budget of k, divided into pieces of size δ each, and a direction i_t with γ-approximately largest gradient*[10] *is chosen in each step t.*

When the hill-climbing algorithm finishes with strategy $\mathbf{x}$, it guarantees that

$$g(\boldsymbol{x}) \geq \left(1 - \mathrm{e}^{-\frac{k \cdot \gamma}{k + \delta \cdot m}}\right) \cdot g(\hat{\boldsymbol{x}})\,,$$

where $\hat{\mathbf{x}}$ denotes the optimal solution subject to $\sum_i \hat{x}_i \leq k$.

[10] A direction i has γ-approximately largest gradient from a point $\boldsymbol{x}$ if

$$g(\boldsymbol{x} + \delta \cdot \boldsymbol{e}_i) - g(\boldsymbol{x}) \geq (1 - \gamma) \cdot (g(\boldsymbol{x} + \delta \cdot \boldsymbol{e}_i) - g(\boldsymbol{x}))$$

for all j.

Proof. The proof is quite similar to, and builds on, the analysis used by Nemhauser et al. [65].

The investments of the optimal solution $\hat{\boldsymbol{x}}$ may not be multiples of δ, so we round them up to the nearest multiple of δ, resulting in a vector $\boldsymbol{y} = (y_1, \ldots, y_m) \geq \hat{\boldsymbol{x}}$. The increase in each coordinate is at most δ, so $\sum_i y_i \leq k + m\delta$.

Consider iteration t of the algorithm. Because g is non-decreasing, we know that $\boldsymbol{x}^{(t)} \geq \mathbf{0}$, so $\boldsymbol{x}^{(t)} + \boldsymbol{y} \geq \boldsymbol{y}$, and therefore (again by monotonicity of g) $g(\boldsymbol{x}^{(t)} + \boldsymbol{y}) \geq g(\boldsymbol{y}) \geq g(\hat{\boldsymbol{x}})$. In turn, we now derive an upper bound on $g(\boldsymbol{x}^{(t)} + \boldsymbol{y})$ in terms of our solution $\boldsymbol{x}^{(t)}$.

Because g has diminishing returns, we know that

$$g(\boldsymbol{x}^{(t)} + \boldsymbol{y}) - g(\boldsymbol{x}^{(t)}) \leq \sum_{i=1}^{m} \left(g(\boldsymbol{x}^{(t)} + y_i \cdot \boldsymbol{e}_i) - g(\boldsymbol{x}^{(t)}) \right).$$

Because y_i is a multiple of δ, we obtain, again from the diminishing returns property of g, that

$$g(\boldsymbol{x}^{(t)} + y_i \cdot \boldsymbol{e}_i) \leq g(\boldsymbol{x}^{(t)}) + \frac{y_i}{\delta} \cdot (g(\boldsymbol{x}^{(t)} + \delta \cdot \boldsymbol{e}_i) - g(\boldsymbol{x}^{(t)})).$$

By the choice of the direction i_t,

$$\begin{aligned} g(\boldsymbol{x}^{(t)} + \boldsymbol{y}) &\leq g(\boldsymbol{x}^{(t)}) + \sum_i \frac{y_i}{\delta} \cdot \left(g(\boldsymbol{x}^{(t)} + \delta \cdot \boldsymbol{e}_i) - g(\boldsymbol{x}^{(t)}) \right) \\ &\leq g(\boldsymbol{x}^{(t)}) + \frac{k + \delta m}{\delta\gamma} \cdot \left(g(\boldsymbol{x}^{(t+1)}) - g(\boldsymbol{x}^{(t)}) \right). \end{aligned}$$

Defining $\Delta_t = g(\boldsymbol{x}^{(t+1)}) - g(\boldsymbol{x}^{(t)})$, and rewriting $\boldsymbol{x}^{(t)}$ as a telescoping series now shows that

$$g(\hat{\boldsymbol{x}}) \leq g(\boldsymbol{x}^{(t)} + \boldsymbol{y}) \ \leq \ g(\mathbf{0}) + \sum_{j<t} \Delta_j + \frac{k + \delta m}{\delta\gamma} \cdot \Delta_t .$$

Multiplying both sides of the t^{th} inequality by

$$\left(1 - \frac{\delta\gamma}{k + \delta m}\right)^{k \cdot \delta^{-1} - t},$$

and summing them all up yields that the term Δ_t appears with coefficient

$$\frac{k + \delta m}{\delta\gamma} \cdot \left(1 - \frac{\delta\gamma}{k + \delta m}\right)^{k \cdot \delta^{-1} - t}$$

in the t^{th} inequality, and with coefficient

$$\left(1 - \frac{\delta\gamma}{k + \delta m}\right)^{k \cdot \delta^{-1} - j}$$

in the j^{th} inequality for $j > t$, so it appears a total of

$$\frac{k + \delta m}{\delta\gamma} \cdot \left(1 - \frac{\delta\gamma}{k + \delta m}\right)^{k \cdot \delta^{-1} - t} + \sum_{j=t+1}^{k \cdot \delta^{-1}} \left(1 - \frac{\delta\gamma}{k + \delta m}\right)^{k \cdot \delta^{-1} - j} = \frac{k + \delta m}{\delta\gamma}$$

times. On the left-hand side of the inequality obtained by summing up, the coefficient for $g(\hat{\boldsymbol{x}})$ is

$$\sum_{t=1}^{k\cdot\delta^{-1}} \left(1-\frac{\delta\gamma}{k+\delta m}\right)^{k\cdot\delta^{-1}-t} = \frac{k+\delta m}{\delta\gamma}\cdot\left(1-\left(1-\frac{\delta\gamma}{k+\delta m}\right)^{k\cdot\delta^{-1}}\right),$$

and similarly for the coefficient of $g(\mathbf{0})$. Hence, the inequality obtained by summing is

$$\frac{k+\delta m}{\delta\gamma}\cdot\left(1-\left(1-\frac{\delta\gamma}{k+\delta m}\right)^{k\cdot\delta^{-1}}\right)\cdot\big(g(\hat{\boldsymbol{x}})-g(\mathbf{0})\big) \le \frac{k+\delta m}{\delta\gamma}\cdot\sum_{t=0}^{k\cdot\delta^{-1}}\Delta_t \;=\; \frac{k+\delta m}{\delta\gamma}\cdot\big(g(\boldsymbol{x})-g(\mathbf{0})\big).$$

Dividing both sides by $(k+\delta m)/(\delta\gamma)$, and bounding that

$$\left(1-\frac{\delta\gamma}{k+\delta m}\right)^{k\cdot\delta^{-1}} \le \mathrm{e}^{-\frac{\gamma k}{k+\delta m}}$$

now shows that

$$g(\boldsymbol{x}) \;\ge\; (1-\mathrm{e}^{-\frac{\gamma\cdot k}{k+\delta m}})\cdot g(\hat{\boldsymbol{x}})+\mathrm{e}^{-\frac{\gamma\cdot k}{k+\delta m}}\cdot g(\mathbf{0}) \;\ge\; (1-\mathrm{e}^{-\frac{\gamma\cdot k}{k+\delta m}})\cdot g(\hat{\boldsymbol{x}}),$$

by the non-negativity of g. This completes the proof. □

With Theorem 7.1 in hand, it remains to show that g is non-negative, monotone, and satisfies Condition 7.1. The first two are clear, so we only prove the third.

First, we determine an expression for the difference $g(\boldsymbol{x}+\boldsymbol{a})-g(\boldsymbol{x})$ when $\boldsymbol{a}\ge\mathbf{0}$, and then show that this difference is non-increasing as a function of $\boldsymbol{x}$. Using Lemma 7.2 below, and changing order of summations, we can write (for an arbitrary, but fixed, ordering of vertices)

$$\begin{aligned}
g(\boldsymbol{x}+\boldsymbol{a})-g(\boldsymbol{x}) &= \sum_A \sigma(A)\cdot\Big(\prod_{i\in A} h_i(\boldsymbol{x}+\boldsymbol{a})\cdot\prod_{i\notin A}(1-h_i(\boldsymbol{x}+\boldsymbol{a})) - \prod_{i\in A} h_i(\boldsymbol{x})\cdot\prod_{i\notin A}(1-h_i(\boldsymbol{x}))\Big)\\
&= \sum_A \sigma(A)\cdot\Big(\sum_u (h_u(\boldsymbol{x}+\boldsymbol{a})-h_u(\boldsymbol{x}))\cdot\\
&\qquad \prod_{i<u,i\in A} h_i(\boldsymbol{x}+\boldsymbol{a})\cdot\prod_{i<u,i\notin A}(1-h_i(\boldsymbol{x}+\boldsymbol{a}))\cdot\prod_{i>u,i\in A} h_i(\boldsymbol{x})\cdot\prod_{i>u,i\notin A}(1-h_i(\boldsymbol{x}))\Big)\\
&= \sum_u \Big((h_u(\boldsymbol{x}+\boldsymbol{a})-h_u(\boldsymbol{x}))\cdot\sum_{A:u\notin A}(\sigma(A\cup\{u\})-\sigma(A))\cdot\\
&\qquad \prod_{i<u,i\in A} h_i(\boldsymbol{x}+\boldsymbol{a})\cdot\prod_{i<u,i\notin A}(1-h_i(\boldsymbol{x}+\boldsymbol{a}))\cdot\prod_{i>u,i\in A} h_i(\boldsymbol{x})\cdot\prod_{i>u,i\notin A}(1-h_i(\boldsymbol{x}))\Big).
\end{aligned}$$

Next, we study the difference

$$\big(g(\boldsymbol{x}+\boldsymbol{a})-g(\boldsymbol{x})\big)-\big(g(\boldsymbol{y}+\boldsymbol{a})-g(\boldsymbol{y})\big)$$

for $\boldsymbol{y}\le\boldsymbol{x}$, and show that it is non-positive. In order to do so, we first use the diminishing returns property of $h_u(\cdot)$, to bound $h_u(\boldsymbol{x}+\boldsymbol{a})-h_u(\boldsymbol{x})\le h_u(\boldsymbol{y}+\boldsymbol{a})-h_u(\boldsymbol{y})$, and then apply Lemma 7.2 again, and change the order of summation once more, to obtain that

$$
\begin{aligned}
&(g(\boldsymbol{x}+\boldsymbol{a})-g(\boldsymbol{x}))-(g(\boldsymbol{y}+\boldsymbol{a})-g(\boldsymbol{y}))\\
&\leq \sum_u \Big((h_u(\boldsymbol{y}+\boldsymbol{a})-h_u(\boldsymbol{y}))\cdot \sum_{A:u\notin A} (\sigma(A\cup\{u\})-\sigma(A))\cdot\\
&\qquad \prod_{i<u,i\in A} h_i(\boldsymbol{x}+\boldsymbol{a})\cdot \prod_{i<u,i\notin A}(1-h_i(\boldsymbol{x}+\boldsymbol{a}))\cdot \prod_{i>u,i\in A} h_i(\boldsymbol{x})\cdot \prod_{i>u,i\notin A}(1-h_i(\boldsymbol{x}))\Big)\\
&-\sum_u \Big((h_u(\boldsymbol{y}+\boldsymbol{a})-h_u(\boldsymbol{y}))\cdot \sum_{A:u\notin A} (\sigma(A\cup\{u\})-\sigma(A))\cdot\\
&\qquad \prod_{i<u,i\in A} h_i(\boldsymbol{y}+\boldsymbol{a})\cdot \prod_{i<u,i\notin A}(1-h_i(\boldsymbol{y}+\boldsymbol{a}))\cdot \prod_{i>u,i\in A} h_i(\boldsymbol{y})\cdot \prod_{i>u,i\notin A}(1-h_i(\boldsymbol{y}))\Big)\\
&= \sum_{u,v:v<u} \Big((h_u(\boldsymbol{y}+\boldsymbol{a})-h_u(\boldsymbol{y}))(h_v(\boldsymbol{x}+\boldsymbol{a})-h_v(\boldsymbol{y}+\boldsymbol{a}))\cdot\\
&\quad \sum_{A:u,v\notin A} (\sigma(A\cup\{u,v\})-\sigma(A\cup\{v\})-\sigma(A\cup\{u\})+\sigma(A))\cdot\\
&\qquad \prod_{i<v,i\in A} h_i(\boldsymbol{x}+\boldsymbol{a})\cdot \prod_{i<v,i\notin A}(1-h_i(\boldsymbol{x}+\boldsymbol{a}))\cdot \prod_{v<i<u,i\in A} h_i(\boldsymbol{y}+\boldsymbol{a})\cdot\\
&\qquad \prod_{v<i<u,i\notin A}(1-h_i(\boldsymbol{y}+\boldsymbol{a}))\cdot \prod_{i>u,i\in A} h_i(\boldsymbol{y})\cdot \prod_{i>u,i\notin A}(1-h_i(\boldsymbol{y}))\Big)\\
&+ \sum_{u,v:v>u} \Big((h_u(\boldsymbol{y}+\boldsymbol{a})-h_u(\boldsymbol{y}))(h_v(\boldsymbol{x})-h_v(\boldsymbol{y}))\cdot\\
&\quad \sum_{A:u,v\notin A} (\sigma(A\cup\{u,v\})-\sigma(A\cup\{v\})-\sigma(A\cup\{u\})+\sigma(A))\cdot\\
&\qquad \prod_{i<u,i\in A} h_i(\boldsymbol{x}+\boldsymbol{a})\cdot \prod_{i<u,i\notin A}(1-h_i(\boldsymbol{x}+\boldsymbol{a}))\cdot \prod_{u<i<v,i\in A} h_i(\boldsymbol{x})\cdot\\
&\qquad \prod_{u<i<v,i\notin A}(1-h_i(\boldsymbol{x}))\cdot \prod_{i>v,i\in A} h_i(\boldsymbol{y})\cdot \prod_{i>v,i\notin A}(1-h_i(\boldsymbol{y}))\Big).
\end{aligned}
$$

Let us consider each of the sums separately. All products are non-negative, as are all of the differences of the form $h_u(\boldsymbol{y}+\boldsymbol{a})-h_u(\boldsymbol{y})$ (and similar ones), by monotonicity and the diminishing returns property of the $h_u(\cdot)$. That leaves the terms

$$\sigma(A\cup\{u,v\})-\sigma(A\cup\{v\})-\sigma(A\cup\{u\})+\sigma(A),$$

which are non-positive by submodularity of $\sigma(\cdot)$. Hence, g does indeed have the diminishing returns property.

Lemma 7.2. *If $a_1,\ldots,a_n$ and $b_1,\ldots,b_n$ are any numbers, then*

$$\prod_{i=1}^n a_i - \prod_{i=1}^n b_i = \sum_{i=1}^n (a_i-b_i)\cdot \prod_{j=1}^{i-1} a_j \cdot \prod_{j=i+1}^n b_j.$$

Proof. The proof is by induction on n. For $n = 1$, the claim obviously holds. For $n > 1$, we can write

$$\begin{aligned}
\prod_{i=1}^{n} a_i - \prod_{i=1}^{n} b_i &= (a_n - b_n) \cdot \prod_{i=1}^{n-1} a_i + b_n \cdot \Big(\prod_{i=1}^{n-1} a_i - \prod_{i=1}^{n-1} b_i\Big) \\
&\stackrel{I.H.}{=} (a_n - b_n) \cdot \prod_{i=1}^{n-1} a_i + b_n \cdot \sum_{i=1}^{n-1} (a_i - b_i) \cdot \prod_{j=1}^{i-1} a_j \cdot \prod_{j=i+1}^{n-1} b_j \\
&= \sum_{i=1}^{n} (a_i - b_i) \cdot \prod_{j=1}^{i-1} a_j \cdot \prod_{j=i+1}^{n} b_j \,,
\end{aligned}$$

completing the proof. □

8 Experiments

In addition to obtaining worst-case guarantees on the performance of the greedy approximation algorithm, we are interested in understanding its behavior in practice, and comparing its performance to other heuristics for identifying influential individuals. We find that the greedy algorithm achieves significant performance gains over several widely used structural measures of influence in social networks [83].

8.1 The network data

For evaluation, it is desirable to use a network data set that exhibits many of the structural features of large-scale social networks. At the same time, we do not address the issue of inferring actual influence parameters from network observations (see, e. g., [26, 68], or [34, 35, 36, 37, 63, 70, 71, 72] for a string of more recent work on this question). Thus, for our testbed, we employ a collaboration graph obtained from co-authorships in physics publications, with simple settings of the influence parameters. It has been argued extensively that co-authorship networks capture many of the key features of social networks more generally [66]. The co-authorship data set was compiled from the complete list of papers in the high-energy physics theory section of the e-print arXiv (`www.arxiv.org`), as of winter of 2002.[11]

The collaboration graph contains a node for each researcher who has at least one paper with co-author(s) in the arXiv database. For each paper with two or more authors, we inserted an edge for each pair of authors. (Single-author papers were ignored.) Notice that this results in parallel edges when two researchers have co-authored multiple papers—we kept these parallel edges as they can be interpreted to indicate stronger social ties between the researchers involved. The resulting graph has 10748 nodes, and edges between about 53000 pairs of nodes.

While processing the data, we corrected many common types of mistakes automatically or manually. In order to deal with aliasing problems at least partially, we abbreviated first names, and unified spellings for foreign characters. We believe that the resulting graph is a good approximation to the actual collaboration graph. (The sheer volume of data prohibits a complete manual cleaning pass.)

[11] We also ran experiments on the co-authorship graphs induced by theoretical computer science papers. We do not report on the results here, as they are very similar to the ones for high-energy physics.

8.2 The influence models

We compared the algorithms in three different models of influence. In the Linear Threshold model, we treated the multiplicity of edges as weights. If nodes u, v have $c_{u,v}$ parallel edges between them and degrees d_u, d_v, then the edge (u, v) has weight $c_{u,v}/d_v$, and the edge (v, u) has weight $c_{u,v}/d_u$.

In the Independent Cascade model, we assigned a uniform probability of p to each edge of the graph, choosing p to be 1% and 10% in separate trials. If nodes u and v have $c_{u,v}$ parallel edges, then we assume that for each of those $c_{u,v}$ edges, u has a chance of p to activate v, i. e., u has a total probability of $1-(1-p)^{c_{u,v}}$ of activating v once it becomes active.

The Independent Cascade model with uniform probabilities p on the edges has the property that high-degree nodes not only have a chance to influence many other nodes, but also to be influenced by them. Whether or not this is a desirable interpretation of the influence data is an application-specific issue. Motivated by this issue, we chose to also consider an alternative interpretation, where edges into high-degree nodes are assigned smaller probabilities. We study a special case of the Independent Cascade model that we term "weighted cascade," in which each edge from node u to v is assigned probability $1/d_v$ of activating v. The weighted Cascade model resembles the Linear Threshold model in that the expected number of neighbors who would succeed in activating a node v is 1 in both models.

8.3 The algorithms and implementation

We compare our greedy algorithm with heuristics based on nodes' degrees and centrality within the network, as well as the crude baseline of choosing random nodes to target. The degree and centrality-based heuristics are commonly used in the sociology literature as estimates of a node's influence [83].

The high-degree heuristic chooses nodes v in order of decreasing degrees d_v. Considering high-degree nodes as influential has long been a standard approach for social and other networks [3, 83], and is known in the sociology literature as "degree centrality."

"Distance centrality" is another commonly used influence measure in sociology, building on the assumption that a node with short paths to other nodes in a network will have a higher chance of influencing them. Hence, we select nodes in order of increasing average distance to other nodes in the network. As the arXiv collaboration graph is not connected, we assigned a distance of n—the number of nodes in the graph—for any pair of unconnected nodes. This value is significantly larger than any actual distance, and thus can be considered to play the role of an infinite distance. In particular, nodes in the largest connected component will have smallest average distance.

Finally, we consider, as a baseline, the result of choosing nodes uniformly at random. Notice that because the optimization problem is NP-hard, and the collaboration graph is prohibitively large, we cannot compute the optimum value to verify the *actual* quality of approximations.

Both in choosing the nodes to target with the greedy algorithm, and in evaluating the performance of the algorithms, we need to compute the value $\sigma(A)$. As discussed in Section 4, computing $\sigma(A)$ exactly is #P-complete; however, very good estimates can be obtained by simulating the random process. More specifically, we simulate the process 10000 times for each targeted set, re-choosing thresholds or edge outcomes pseudo-randomly from the interval $[0, 1]$ every time. Previous runs indicate that the quality of approximation after 10000 iterations is comparable to that after 300000 or more iterations.

In all of the experiments, we vary k from 1 to 30. This is in part for computational reasons, and in part because according to our experiments, the number of activated nodes as a function of k grows very close to linearly beyond $k = 30$.

8.4 The results

Figure 1 shows the performance of the algorithms in the Linear Threshold model. The greedy algorithm outperforms the high-degree node heuristic by about 18%, and the central node heuristic by over 40%. (As expected, choosing random nodes is not a good idea.) This shows that significantly better marketing results can be obtained by explicitly considering the dynamics of information in a network, rather than relying solely on structural properties of the graph.

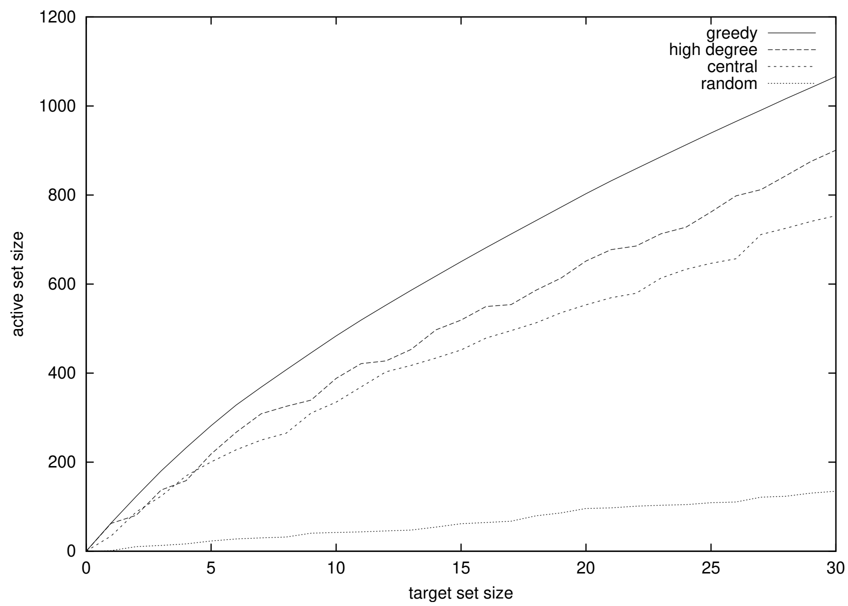

Figure 1: Results for the Linear Threshold model.

When investigating the reason why the high-degree and centrality heuristics do not perform as well, one sees that they ignore such network effects. In particular, neither of the heuristics incorporates the fact that many of the most central (or highest-degree) nodes may be clustered, so that targeting all of them is unnecessary. In fact, the uneven nature of these curves suggests that the network influence of many nodes is not accurately reflected by their degree or centrality.

Figure 2 shows the results for the weighted Cascade model. Notice the striking similarity to the Linear Threshold model. The scale is slightly different (all values are about 25% smaller), but the behavior is qualitatively the same, even with respect to the exact nodes whose network influence is not reflected accurately by their degree or centrality. The reason is that in expectation, each node is influenced by the same number of other nodes in both models (see Section 4), and the degrees are relatively concentrated around their expectation of 1.

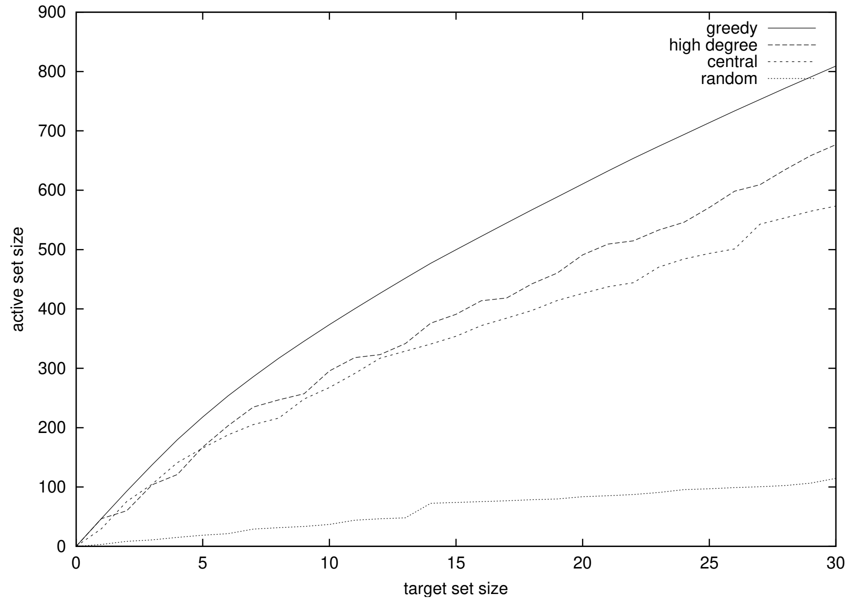

Figure 2: Results for the Weighted Cascade model.

The graph for the Independent Cascade model with probability 1%, given in Figure 3, seems very similar to the previous two at first glance. Notice, however, the very different scale: on average, each targeted node only activates three additional nodes. Hence, the network effects in the Independent Cascade model with very small probabilities are much weaker than in the other models. Several nodes have degrees well exceeding 100, so the probabilities on their incoming edges are even smaller than 1% in the weighted Cascade model. This suggests that the network effects observed for the Linear Threshold and weighted Cascade models rely heavily on low-degree nodes as multipliers, even though targeting high-degree nodes is a reasonable heuristic. Also notice that in the Independent Cascade model, the heuristic of choosing random nodes performs significantly better than in the previous two models.

The improvement in performance of the "random nodes" heuristic is even more pronounced for the Independent Cascade model with probabilities equal to 10%, depicted in Figure 4. In that model, the "random nodes" heuristic starts to outperform both the high-degree and the central nodes heuristics when more than 12 nodes are targeted. It is initially surprising that random targeting for this model should lead to more activations than centrality-based targeting, but in fact there is a natural underlying reason that we explore now.

The first targeted node, if chosen somewhat judiciously, will activate a large fraction of the network, in our case almost 25%. However, any additional nodes will only reach a small additional fraction of the network. In particular, other central or high-degree nodes are very likely to be activated by the initially chosen one, and thus have hardly any marginal gain. This explains the shapes of the curves for the high-degree and distance centrality heuristics, which leap up to about 2415 activated nodes, but make virtually no progress afterwards. The greedy algorithm, on the other hand, takes the effect of the first

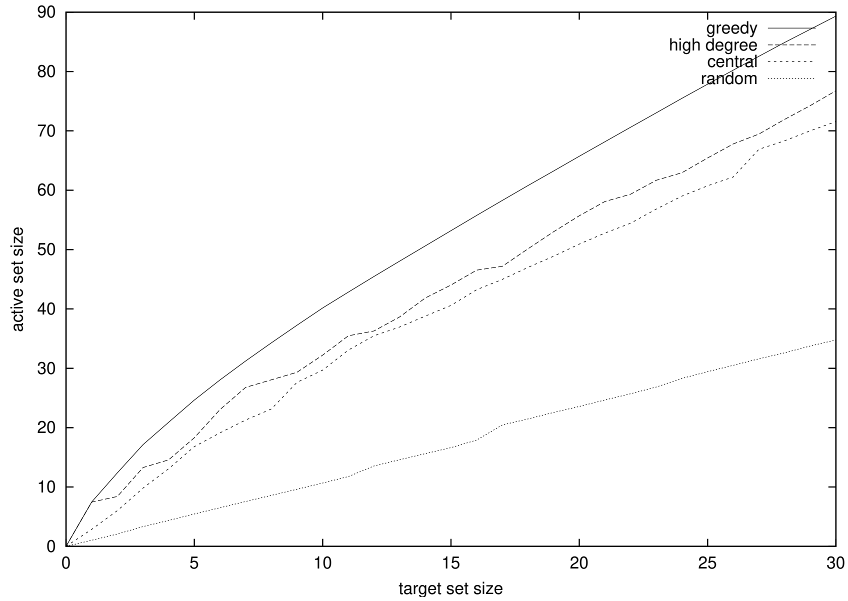

Figure 3: Independent Cascade model with probability 1%.

chosen node into account, and targets nodes with smaller marginal gain afterwards. Hence, its active set keeps growing, although at a much smaller slope than in other models.

The random heuristic does not do as well initially as the other heuristics, but with sufficiently many attempts, it eventually hits some highly influential nodes and becomes competitive with the centrality-based node choices. Because it does not focus exclusively on central nodes, it eventually targets nodes with additional marginal gain, and surpasses the two centrality-based heuristics.

In summary, our experiments show that on a data set with many of the features of a real social network, a greedy algorithm motivated by theoretical insights significantly outperforms several standard heuristics. This is not to say that better algorithms do not exist: for example, a local search step after the greedy algorithm would likely improve the results further. Similarly, there may well be other, more efficient, heuristics with similar or better performance in practice—see the discussion in Section 1.2. Our main goal here was primarily to establish that theoretical guarantees and practical performance are not mutually exclusive.

9 Conclusions

Peer influence and word-of-mouth effects play an important role in the dissemination of ideas and innovations and the adoption of new products. When designing campaigns to promote the adoption of ideas or products, it is therefore important to take these network effects into account. In the present work, we studied this optimization problem formally, under several widely used models of influence and cascading behavior. We showed that under these models, a simple greedy algorithm gives a $1-1/e$

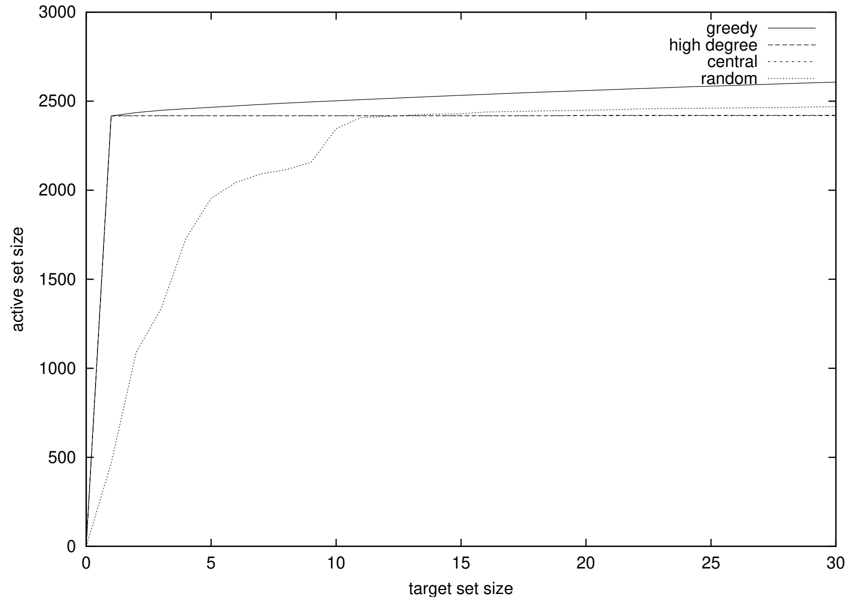

Figure 4: Independent Cascade model with probability 10%.

approximation for the influence maximization objective, and that these results extend to a more general setting in which a marketing budget is to be divided among different available marketing actions, each of which may affect subsets of nodes. Our theoretical analysis is complemented by experiments performed on a collaboration network extracted from the arXiv site.

Many important directions remain for future work. While Mossel and Roch [62] generalize the submodularity results of our work to more general Threshold models, we do not as of yet have a complete understanding of the full range of models for which approximate influence maximization is tractable. Even more immediately, while the factor $1-1/e$ in the approximation guarantee is matched by an approximation hardness result for the Independent Cascade model, no approximation hardness whatsoever is known for the Linear Threshold model. Establishing an improved approximation guarantee, or any kind of approximation hardness, would therefore be of high interest.

Several natural modeling extensions have been studied in the literature recently, including competition between multiple cascades, negative opinions, and negative tie strengths. An important property shared by these extensions is that they need to introduce a timing component, wherein the decision of a node for a particular state depends on which neighbor influences the node first. This stands in contrast with the strong timing independence for our basic models, captured by Lemma 3.2. Defining a consistent model for negative influence or competition which is independent of timing components would be of interest; a promising direction could be a return to graphical models in the vein of those studied by Domingos and Richardson [26].

In order to apply the models and algorithms of the present (or subsequent) work in the real world, it is necessary to estimate the models' parameters (such as influence probabilities in the Cascade model, or edge weights in the Threshold model). A string of recent papers, e. g., [34, 35, 36, 37, 63, 70, 71, 72], begin to address the inference problem. The general approach is to use multiple instances of cascades, usually with information on the times at which nodes became active, and to perform a Maximum Likelihood or similar estimation. Thus, these inference algorithms assume that the model parameters are invariant across all observations. This is a strong assumption: in practice, the influence of one node on another will depend on the context of the product or innovation that is being recommended. Inferring parameters under weaker assumptions, such as a latent low-dimensional space on products which would explain the parameters, is a promising direction. At an even more fundamental level, a thorough validation of the present (and other) models of social influence is clearly necessary.

Acknowledgments

We would like to thank anonymous reviewers of the conference and journal versions of this paper for useful feedback.

References

[1] NIMA AHMADI POUR ANARI, SHAYAN EHSANI, MOHAMMAD GHODSI, NIMA HAGHPANAH, NICOLE IMMORLICA, HAMID MAHINI, AND VAHAB S. MIRROKNI: Equilibrium pricing with positive externalities. *Theor. Comput. Sci.*, 476:1–15, 2013. Preliminary version in WINE'10. [doi:10.1016/j.tcs.2013.01.014] 109

[2] HESSAMEDDIN AKHLAGHPOUR, MOHAMMAD GHODSI, NIMA HAGHPANAH, HAMID MAHINI, VAHAB S. MIRROKNI, AND AFSHIN NIKZAD: Optimal iterative pricing over social networks. In *Proc. 6th Workshop on Internet and Network Economics (WINE'10)*, pp. 415–423, 2010. [doi:10.1007/978-3-642-17572-5_34] 109

[3] RÉKA ALBERT, HAWOONG JEONG, AND ALBERT-LÁSZLÓ BARABÁSI: Error and attack tolerance of complex networks. *Nature*, 406(406):378–382, 2000. [doi:10.1038/35019019] 134

[4] NOGA ALON, MICHAL FELDMAN, ARIEL D. PROCACCIA, AND MOSHE TENNENHOLTZ: A note on competitive diffusion through social networks. *Inform. Process. Lett.*, 110(4):221–225, 2010. [doi:10.1016/j.ipl.2009.12.009] 110

[5] DAVID ARTHUR, RAJEEV MOTWANI, ANEESH SHARMA, AND YING XU: Pricing strategies for viral marketing on social networks. In *Proc. 5th Workshop on Internet and Network Economics (WINE'09)*, pp. 101–112, 2009. [doi:10.1007/978-3-642-10841-9_11, arXiv:0902.3485] 109

[6] CHALEE ASAVATHIRATHAM: *The Influence Model: A Tractable Representation for the Dynamics of Networked Markov Chains*. Ph. D. thesis, Massachusetts Institute of Technology, 2001. Available at MIT DSPACE, Subsequent paper available at IEEE Control Systems. 106, 128

[7] CHALEE ASAVATHIRATHAM, SANDIP ROY, BERNARD C. LESIEUTRE, AND GEORGE C. VERGHESE: The influence model. *IEEE Comp. Soc. Press*, 21(6):52–64, 2001. [doi:10.1109/37.969135] 106, 128

[8] FRANK M. BASS: A new product growth model for consumer durables. *Management Science*, 15(5):215–227, 1969. [doi:10.1287/mnsc.1040.0264] 106

[9] ELI BERGER: Dynamic monopolies of constant size. *J. Combin. Theory Ser. B*, 83(2):191–200, 2001. [doi:10.1006/jctb.2001.2045] 111, 118

[10] SHISHIR BHARATHI, DAVID KEMPE, AND MAHYAR SALEK: Competitive influence maximization in social networks. In *Proc. 3rd Workshop on Internet and Network Economics (WINE'07)*, pp. 306–311, 2007. [doi:10.1007/978-3-540-77105-0_31] 110

[11] LAWRENCE E. BLUME: The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5(3):387–424, 1993. [doi:10.1006/game.1993.1023] 106

[12] CHRISTIAN BORGS, MICHAEL BRAUTBAR, JENNIFER CHAYES, AND BRENDAN LUCIER: Maximizing social influence in nearly optimal time. In *Proc. 25th Ann. ACM-SIAM Symp. on Discrete Algorithms (SODA'14)*, pp. 946–957, 2014. [doi:10.1137/1.9781611973402.70] 109

[13] ALLAN BORODIN, YUVAL FILMUS, AND JOEL OREN: Threshold models for competitive influence in social networks. In *Proc. 6th Workshop on Internet and Network Economics (WINE'10)*, pp. 539–550, 2010. [doi:10.1007/978-3-642-17572-5_48] 110

[14] JACQUELINE J. BROWN AND PETER H. REINEGEN: Social ties and word-of-mouth referral behavior. *Journal of Consumer Research*, 14(3):350–362, 1987. [doi:10.1086/209118] 106

[15] CEREN BUDAK, DIVYAKANT AGRAWAL, AND AMR EL ABBADI: Limiting the spread of misinformation in social networks. In *20th Internat. World Wide Web Conf. (WWW'11)*, pp. 665–674, 2011. [doi:10.1145/1963405.1963499] 110

[16] TIM CARNES, CHANDRASHEKAR NAGARAJAN, STEFAN M. WILD, AND ANKE VAN ZUYLEN: Maximizing influence in a competitive social network: A follower's perspective. In *Proc. Internat. Conf. on Electronic Commerce (ICEC)*, pp. 351–360, 2007. [doi:10.1145/1282100.1282167] 110

[17] NING CHEN: On the approximability of influence in social networks. *SIAM J. Discr. Math.*, 23(3):1400–1415, 2009. Conference version in SODA'08. [doi:10.1137/08073617X] 118

[18] WEI CHEN, ALEX COLLINS, RACHEL CUMMINGS, TE KE, ZHENMING LIU, DAVID RINCON, XIAORUI SUN, YAJUN WANG, WEI WEI, AND YIFEI YUAN: Influence maximization in social networks when negative opinions may emerge and propagate. In *Proc. 11th SIAM Internat. Conf. on Data Mining (SDM'11)*, pp. 379–390, 2011. [doi:10.1137/1.9781611972818.33] 110

[19] WEI CHEN, LAKS V.S. LAKSHMANAN, AND CARLOS CASTILLO: *Information and Influence Propagation in Social Networks*. Synthesis Lectures on Data Management. Morgan & Claypool, 2013. [doi:10.2200/S00527ED1V01Y201308DTM037] 108

[20] WEI CHEN, YAJUN WANG, AND SIYU YANG: Efficient influence maximization in social networks. In *Proc. 15th Internat. Conf. on Knowledge Discovery and Data Mining (KDD'09)*, pp. 199–208, 2009. [doi:10.1145/1557019.1557047] 109

[21] WEI CHEN, YIFEI YUAN, AND LI ZHANG: Scalable influence maximization in social networks under the linear threshold model. In *Proc. 10th Internat. Conf. on Data Mining (ICDM'12)*, pp. 88–97, 2010. [doi:10.1109/ICDM.2010.118] 109, 118

[22] PETER CLIFFORD AND AIDAN SUDBURY: A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973. [doi:10.1093/biomet/60.3.581] 108

[23] JAMES S. COLEMAN, HERBERT MENZEL, AND ELIHU KATZ: *Medical Innovations: A Diffusion Study*. Bobbs Merrill, 1966. 106

[24] GÉRARD CORNUÉJOLS, MARSHALL L. FISHER, AND GEORGE L. NEMHAUSER: Location of bank accounts to optimize float. *Management Science*, 23(8):789–810, 1977. [doi:10.1287/mnsc.23.8.789] 107, 119, 120

[25] ERIK D. DEMAINE, MOHAMMADTAGHI HAJIAGHAYI, HAMID MAHINI, DAVID L. MALEC, S. RAGHAVAN, ANSHUL SAWANT, AND MORTEZA ZADIMOGHADAM: How to influence people with partial incentives. In *23rd Internat. World Wide Web Conf. (WWW'14)*, pp. 937–948, 2014. [doi:10.1145/2566486.2568039, arXiv:1401.7970] 128

[26] PEDRO DOMINGOS AND MATTHEW RICHARDSON: Mining the network value of customers. In *Proc. 7th Internat. Conf. on Knowledge Discovery and Data Mining (KDD'01)*, pp. 57–66, 2001. [doi:10.1007/978-3-540-77105-0_27] 106, 107, 108, 118, 128, 133, 138

[27] PRADEEP DUBEY, RAHUL GARG, AND BERNARD DE MEYER: Competing for customers in a social network: The quasi-linear case. In *Proc. 2nd Workshop on Internet and Network Economics (WINE'06)*, pp. 162–173, 2006. [doi:10.1007/11944874_16] 110

[28] RICHARD DURRETT: *Lecture Notes on Particle Systems and Percolation*. Wadsworth Publishing, 1988. 107, 113

[29] GLENN ELLISON: Learning, local interaction, and coordination. *Econometrica*, 61(5):1047–1071, 1993. [doi:10.2307/2951493] 106

[30] EYAL EVEN-DAR AND ASAF SHAPIRA: A note on maximizing the spread of influence in social networks. *Inform. Process. Lett.*, 111(4):184–187, 2011. Preliminary version in WINE'11. [doi:10.1016/j.ipl.2010.11.015] 108

[31] URIEL FEIGE: A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45(4):634–652, 1998. Preliminary version in STOC'96. [doi:10.1145/285055.285059] 113

[32] JACOB GOLDENBERG, BARAK LIBAI, AND EITAN MULLER: Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters*, 12(3):211–223, 2001. [doi:10.1023/A:1011122126881] 106, 107, 113

[33] JACOB GOLDENBERG, BARAK LIBAI, AND EITAN 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. 106, 107, 113

[34] MANUEL GOMEZ-RODRIGUEZ, DAVID BALDUZZI, AND BERNHARD SCHÖLKOPF: Uncovering the temporal dynamics of diffusion networks. In *Proc. 28th Int. Conf. on Machine Learning (ICML'11)*, pp. 561–568, 2011. [arXiv:1105.0697] 133, 139

[35] MANUEL GOMEZ-RODRIGUEZ, JURE LESKOVEC, AND ANDREASE KRAUSE: Inferring networks of diffusion and influence. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 5(4):21, 2012. Preliminary version in KDD'10. [doi:10.1145/2086737.2086741] 133, 139

[36] MANUEL GOMEZ-RODRIGUEZ AND BERNHARD SCHÖLKOPF: Submodular inference of diffusion networks from multiple trees. In *Proc. 29th Int. Conf. on Machine Learning (ICML'12)*, 2012. Available at ICML. [arXiv:1205.1671] 133, 139

[37] AMIT GOYAL, FRANCESCO BONCHI, AND LAKS V. S. LAKSHMANAN: Learning influence probabilities in social networks. In *Proc. 3rd ACM Internat. Conf. on Web Search and Data Mining (WSDM'10)*, pp. 241–250, 2010. [doi:10.1145/1718487.1718518] 133, 139

[38] AMIT GOYAL, WEI LU, AND LAKS V. S. LAKSHMANAN: CELF++: Optimizing the greedy algorithm for influence maximization in social networks. In *20th Internat. World Wide Web Conf. (WWW'11)*, pp. 47–48, 2011. [doi:10.1145/1963192.1963217] 109

[39] AMIT GOYAL, WEI LU, AND LAKS V. S. LAKSHMANAN: SIMPATH: An efficient algorithm for influence maximization under the linear threshold model. In *Proc. 11th Internat. Conf. on Data Mining (ICDM'11)*, pp. 211–220, 2011. [doi:10.1109/ICDM.2011.132] 109

[40] SANJEEV GOYAL, HODA HEIDARI, AND MICHAEL KEARNS: Competitive contagion in networks. *Games and Economic Behavior*, 2014 (online). Preliminary version in STOC'12. [doi:10.1016/j.geb.2014.09.002] 110

[41] MARK GRANOVETTER: Threshold models of collective behavior. *American Journal of Sociology*, 83(6):1420–1443, 1978. [doi:10.1086/226707] 107, 111

[42] DILEK GÜNNEÇ: *Integrating Social Network Effects in Product Design and Diffusion*. Ph. D. thesis, University of Maryland, 2012. Available at DRUM. 128

[43] NIMA HAGHPANAH, NICOLE IMMORLICA, VAHAB S. MIRROKNI, AND KAMESH MUNAGALA: Optimal auctions with positive network externalities. *ACM Trans. Economics and Comput.*, 1(2):13, 2013. Conference version in EC'11. [doi:10.1145/2465769.2465778] 109

[44] VENKY HARINARAYAN, ANAND RAJARAMAN, AND JEFFREY D. ULLMAN: Implementing data cubes efficiently. In *Proc. 16th ACM SIGMOD Internat. Conf. on Managment of Data*, pp. 205–216, 1996. [doi:10.1145/233269.233333] 120

[45] JASON D. HARTLINE, VAHAB S. MIRROKNI, AND MUKUND SUNDARARAJAN: Optimal marketing strategies over social networks. In *17th Internat. World Wide Web Conf. (WWW'08)*, pp. 189–198, 2008. [doi:10.1145/1367497.1367524] 109

[46] XINRAN HE AND DAVID KEMPE: Price of anarchy for the n-player competitive cascade game with submodular activation functions. In *Proc. 9th Workshop on Internet and Network Economics (WINE'13)*, pp. 232–248, 2013. [doi:10.1007/978-3-642-45046-4_20] 110

[47] XINRAN HE, GUOJIE SONG, WEI CHEN, AND QINGYE JIANG: Influence blocking maximization in social networks under the competitive linear threshold model. In *Proc. 12th SIAM Internat. Conf. on Data Mining (SDM'12)*, pp. 463–474, 2012. [doi:10.1137/1.9781611972825.40, arXiv:1110.4723] 110

[48] RICHARD A. HOLLEY AND THOMAS M. LIGGETT: Ergodic theorems for weakly interacting infinite systems and the voter model. *Annals of Probability*, 3(4):643–663, 1975. Available at Project Euclid. 108

[49] QINGYE JIANG, GUOJIE SONG, GAO CONG, YU WANG, WENJUN SI, AND KUNQING XIE: Simulated annealing based influence maximization in social networks. In *Proc. 26th AAAI Conf. on Artificial Intelligence*, 2011. Available at AAAI. 109

[50] KYOMIN JUNG, WOORAM HEO, AND WEI CHEN: IRIE: Scalable and robust influence maximization in social networks. In *Proc. 12th Internat. Conf. on Data Mining (ICDM'12)*, pp. 918–923, 2012. [doi:10.1109/ICDM.2012.79] 109

[51] DAVID KEMPE, JON KLEINBERG, AND ÉVA TARDOS: Maximizing the spread of influence in a social network. In *Proc. 9th Internat. Conf. on Knowledge Discovery and Data Mining (KDD'03)*, pp. 137–146, 2003. [doi:10.1145/956750.956769] 105, 108, 112, 118

[52] DAVID KEMPE, JON KLEINBERG, AND ÉVA TARDOS: Influential nodes in a diffusion model for social networks. In *Proc. 32th Internat. Colloq. on Automata, Languages and Programming (ICALP'05)*, pp. 1127–1138, 2005. [doi:10.1007/11523468_91] 105, 108

[53] MASAHIRO KIMURA AND KAZUMI SAITO: Tractable models for information diffusion in social networks. In *Proc. 10th European Conf. on Principles and Practice of Knowledge Discovery in Databases (PKDD'06)*, pp. 259–271, 2006. [doi:10.1007/11871637_27] 109

[54] ANDREAS KRAUSE AND DANIEL GOLOVIN: Submodular function maximization. In LUCAS BORDEAUX, YOUSSEF HAMADI, AND PUSHMEET KOHLI, editors, *Tractability: Practical Approaches to Hard Problems*, pp. 71–104. Cambridge University Press, 2014. 120

[55] JURE LESKOVEC, ANDREAS KRAUSE, CARLOS GUESTRIN, CHRISTOS FALOUTSOS, JEANNE VANBRIESEN, AND NATALIE S. GLANCE: Cost-effective outbreak detection in networks. In *Proc. 13th Internat. Conf. on Knowledge Discovery and Data Mining (KDD'07)*, pp. 420–429, 2007. [doi:10.1145/1281192.1281239] 108

[56] THOMAS M. LIGGETT: *Interacting Particle Systems*. Springer, 1985. 107, 113

[57] MICHAEL MACY: Chains of cooperation: Threshold effects in collective action. *American Sociological Review*, 56(6):730–747, 1991. [doi:10.2307/2096252] 111

[58] MICHAEL MACY AND ROBERT WILLER: From factors to actors: Computational sociology and agent-based modeling. *Annual Review of Sociology*, 28:143–166, 2002. [doi:10.1146/annurev.soc.28.110601.141117] 111

[59] VIJAY MAHAJAN, EITAN MULLER, AND FRANK M. BASS: New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1):1–26, 1990. 106

[60] COLIN MCDIARMID: Concentration. In MICHEL HABIB, COLIN MCDIARMID, JORGE RAMIREZ-ALFONSIN, AND BRUCE REED, editors, *Probabilistic Methods for Algorithmic Discrete Mathematics*, pp. 195–247. Springer, 1998. [doi:10.1007/978-3-662-12788-9_6] 119

[61] STEPHEN MORRIS: Contagion. *Review of Economic Studies*, 67(1):57–78, 2000. [doi:10.1111/1467-937X.00121] 106, 111, 118

[62] ELCHANAN MOSSEL AND SEBASTIEN ROCH: Submodularity of influence in social networks: From local to global. *SIAM J. Comput.*, 39(6):2176–2188, 2010. Preliminary version in STOC'07. [doi:10.1137/080714452] 108, 113, 123, 126, 138

[63] SETH A. MYERS AND JURE LESKOVEC: On the convexity of latent social network inference. In *Proc. 22nd Advances in Neural Information Processing Systems Conf. (NIPS'10)*, pp. 1741–1749, 2010. Available at ML Repository. [arXiv:1010.5504] 133, 139

[64] GEORGE L. NEMHAUSER AND LAURENCE A. WOLSEY: *Integer and Combinatorial Optimization*. Wiley, 1988. 120

[65] GEORGE L. NEMHAUSER, LAURENCE A. WOLSEY, AND MARSHALL L. FISHER: An analysis of the approximations for maximizing submodular set functions—I. *Mathematical Programming*, 14(1):265–294, 1978. [doi:10.1007/BF01588971] 107, 119, 120, 130

[66] MARK E. J. NEWMAN: The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA*, 98(2):404–409, 2001. [doi:10.1073/pnas.98.2.404] 133

[67] DAVID PELEG: Local majorities, coalitions and monopolies in graphs: a review. *Theoret. Comput. Sci.*, 282(2):231–257, 2002. Preliminary version in SIROCCO'96. [doi:10.1016/S0304-3975(01)00055-X] 111, 118

[68] MATTHEW RICHARDSON AND PEDRO DOMINGOS: Mining knowledge-sharing sites for viral marketing. In *Proc. 8th Internat. Conf. on Knowledge Discovery and Data Mining (KDD'02)*, pp. 61–70, 2002. [doi:10.1145/775047.775057] 106, 107, 108, 110, 118, 128, 133

[69] EVERETT ROGERS: *Diffusion of Innovations*. Free Press, 4th edition, 1995. 106

[70] ELDAR SADIKOV, MONTSERRAT MEDINA, JURE LESKOVEC, AND HECTOR GARCIA-MOLINA: Correcting for missing data in information cascades. In *Proc. 4th ACM Internat. Conf. on Web Search and Data Mining (WSDM'11)*, pp. 55–64, 2011. [doi:10.1145/1935826.1935844] 133, 139

[71] KAZUMI SAITO, MASAHIRO KIMURA, KOUZOU OHARA, AND HIROSHI MOTODA: Selecting information diffusion models over social networks for behavioral analysis. In *Proc. European Conf. on Machine Learning and Knowledge Discovery in Databases: Part III (ECML/PKDD'10)*, volume 6323 of *LNCS*, pp. 180–195. Springer, 2010. [doi:10.1007/978-3-642-15939-8_12] 133, 139

[72] KAZUMI SAITO, RYOHEI NAKANO, AND MASAHIRO KIMURA: Prediction of information diffusion probabilities for independent cascade model. In *Proc. 12th Internat. Conf. on Knowledge-Based and Intelligent Information & Engineering Systems (KES'08)*, volume 5179 of *LNCS*, pp. 67–75. Springer, 2008. [doi:10.1007/978-3-540-85567-5_9] 133, 139

[73] MAHYAR SALEK, SHAHIN SHAYANDEH, AND DAVID KEMPE: You share, I share: Network effects and economic incentives in P2P file-sharing systems. In *Proc. 6th Workshop on Internet and Network Economics (WINE'10)*, pp. 354–365, 2010. [doi:10.1007/978-3-642-17572-5_29, arXiv:1107.5559] 124

[74] THOMAS SCHELLING: *Micromotives and Macrobehavior*. Norton, 1978. 107, 111

[75] LIOR SEEMAN AND YARON SINGER: Adaptive seeding in social networks. In *Proc. 54th FOCS*, pp. 459–468. IEEE Comp. Soc. Press, 2013. [doi:10.1109/FOCS.2013.56] 110

[76] YARON SINGER: How to win friends and influence people, truthfully: Influence maximization mechanisms for social networks. In *Proc. 5th ACM Internat. Conf. on Web Search and Data Mining (WSDM'12)*, pp. 733–742, 2012. [doi:10.1145/2124295.2124381] 109

[77] JASON TSAI, THANH H. NGUYEN, AND MILIND TAMBE: Security games for controlling contagion. In *Proc. 27th AAAI Conf. on Artificial Intelligence (AAAI'12)*, pp. 1464–1470, 2012. Available at AAAI. 110

[78] VASILEIOS TZOUMAS, CHRISTOS AMANATIDIS, AND EVANGELOS MARKAKIS: A game-theoretic analysis of a competitive diffusion process over social networks. In *Proc. 8th Workshop on Internet and Network Economics (WINE'12)*, pp. 1–14, 2012. [doi:10.1007/978-3-642-35311-6_1] 110

[79] THOMAS W. VALENTE: *Network Models of the Diffusion of Innovations*. Hampton Press, 1995. Available at Springer. 106, 111

[80] ADRIAN VETTA: Nash equlibria in competitive societies with applications to facility location, traffic routing and auctions. In *Proc. 43rd FOCS*, pp. 416–425. IEEE Comp. Soc. Press, 2002. [doi:10.1109/SFCS.2002.1181966] 110

[81] CHI WANG, WEI CHEN, AND YAJUN WANG: Scalable influence maximization for independent cascade model in large-scale social networks. *Data Mining and Knowledge Discovery Journal*, 25(3):545–576, 2012. Preliminary version in KDD'10. [doi:10.1007/s10618-012-0262-1] 109, 118

[82] YU WANG, GAO CONG, GUOJIE SONG, AND KUNQING XIE: Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proc. 16th Internat. Conf. on Knowledge Discovery and Data Mining (KDD'10)*, pp. 1039–1048, 2010. [doi:10.1145/1835804.1835935] 109

[83] STANLEY S. WASSERMAN AND KATHERINE FAUST: *Social Network Analysis*. Cambridge University Press, 1994. 108, 133, 134

[84] DUNCAN J. WATTS: A simple model of global cascades in random networks. *Proc. Natl. Acad. Sci. USA*, 99(9):5766–5771, 2002. [doi:10.1073/pnas.082090499] 111

[85] H. PEYTON YOUNG: *Individual Strategy and Social Structure: An Evolutionary Theory of Institutions*. Princeton University Press, 1998. 106, 111

[86] H. PEYTON YOUNG: The diffusion of innovations in social networks. Technical Report 02-14-018, Santa Fe Institute Working Paper, 2002. Available at IDEAS. 106, 111

AUTHORS

David Kempe
Associate Professor
Department of Computer Science
University of Southern California
Los Angeles, CA 90089
dkempe@usc.edu
http://www-bcf.usc.edu/~dkempe

Jon Kleinberg
Tisch University Professor
Departments of Computer Science and Information Science
Cornell University
Ithaca, NY 14853
kleinber@cs.cornell.edu
http://www.cs.cornell.edu/home/kleinber/kleinber.html

Éva Tardos
Jacob Gould Schurman Professor
Department of Computer Science
Cornell University
Ithaca, NY 14853
eva.tardos@cornell.edu
http://www.cs.cornell.edu/people/eva/eva.html

ABOUT THE AUTHORS

DAVID KEMPE received his Ph. D. from Cornell University in 2003, and has been on the faculty in the Computer Science Department at USC since the Fall of 2004, where he is currently an Associate Professor.

His primary research interests are in computer science theory and the design and analysis of algorithms, with a particular emphasis on social networks, algorithms for feature selection, and game-theoretic and pricing questions. He is a recipient of the NSF CAREER award, the VSoE Junior Research Award, the ONR Young Investigator Award, a Sloan Fellowship, and an Okawa Fellowship, in addition to several USC mentoring awards.

JON KLEINBERG is the Tisch University Professor in the Departments of Computer Science and Information Science at Cornell University. His research focuses on algorithmic issues at the interface of networks and information, with an emphasis on the social and information networks that underpin the Web and other on-line media.

He is a member of the National Academy of Sciences and the National Academy of Engineering, and is the recipient of awards including a MacArthur Fellowship, the Nevanlinna Prize, the Harvey Prize, the ACM SIGKDD Innovation Award, and the ACM-Infosys Foundation Award in the Computing Sciences.

ÉVA TARDOS is a Jacob Gould Schurman Professor of Computer Science at Cornell University. Her research interest is algorithms and algorithmic game theory, the subarea of theoretical computer science theory of designing systems and algorithms for selfish users. Her research focuses on algorithms and games on networks.

She has been elected to the National Academy of Engineering, the National Academy of Sciences, is an external member of the Hungarian Academy of Sciences, and is the recipient of a number of fellowships and awards including the Packard Fellowship, the Gödel Prize, Dantzig Prize, Fulkerson Prize, and the IEEE Technical Achievement Award. She was editor editor-in-chief of the SIAM Journal of Computing 2004–2009, and is currently an editor of several other journals including the *Journal of the ACM* and *Combinatorica.*