Soc. Netw. Anal. Min.
DOI 10.1007/s13278-012-0062-z

ORIGINAL ARTICLE

# On minimizing budget and time in influence propagation over social networks

**Amit Goyal · Francesco Bonchi · Laks V. S. Lakshmanan · Suresh Venkatasubramanian**

Received: 8 November 2011 / Revised: 21 February 2012 / Accepted: 28 February 2012

**Abstract** In recent years, study of influence propagation in social networks has gained tremendous attention. In this context, we can identify three orthogonal dimensions—the number of *seed* nodes activated at the beginning (known as *budget*), the expected number of activated nodes at the end of the propagation (known as *expected spread* or *coverage*), and the *time* taken for the propagation. We can constrain one or two of these and try to optimize the third. In their seminal paper, Kempe et al. constrained the budget, left time unconstrained, and maximized the coverage: this problem is known as *Influence Maximization* (or MAXINF for short). In this paper, we study alternative optimization problems which are naturally motivated by resource and time constraints on viral marketing campaigns. In the first problem, termed *minimum target set selection* (or MINTSS for short), a coverage threshold $\eta$ is given and the task is to find the *minimum size seed set* such that by activating it, at least $\eta$ nodes are eventually activated in the expected sense. This naturally captures the problem of deploying a viral campaign on a budget. In the second problem, termed MINTIME, the goal is to minimize the time in which a predefined coverage is achieved. More precisely, in MINTIME, a coverage threshold $\eta$ and a budget threshold $k$ are given, and the task is to find a seed set of size at most $k$ such that by activating it, at least $\eta$ nodes are activated in the expected sense, *in the minimum possible time*. This problem addresses the issue of *timing* when deploying viral campaigns. Both these problems are **NP**-hard, which motivates our interest in their approximation. For MINTSS, we develop a simple greedy algorithm and show that it provides a bicriteria approximation. We also establish a generic hardness result suggesting that improving this bicriteria approximation is likely to be hard. For MINTIME, we show that even bicriteria and tricriteria approximations are hard under several conditions. We show, however, that if we allow the budget for number of seeds $k$ to be boosted by a logarithmic factor and allow the coverage to fall short, then the problem can be solved *exactly* in PTIME, i.e., we can achieve the required coverage within the time achieved by the optimal solution to MINTIME with budget $k$ and coverage threshold $\eta$. Finally, we establish the value of the approximation algorithms, by conducting an experimental evaluation, comparing their quality against that achieved by various heuristics.

**Keywords** Social networks · Social influence · Influence propagation · Viral marketing · Approximation analysis · MINTSS · MINTIME

A. Goyal (✉) · L. V. S. Lakshmanan
University of British Columbia, Vancouver, BC, Canada
e-mail: goyal@cs.ubc.ca

L. V. S. Lakshmanan
e-mail: laks@cs.ubc.ca

F. Bonchi
Yahoo! Research, Barcelona, Spain
e-mail: bonchi@yahoo-inc.com

S. Venkatasubramanian
University of Utah, Salt Lake City, UT, USA
e-mail: suresh@cs.utah.edu

## 1 Introduction

The study of how influence and information propagate in social networks has recently received a great deal of attention (Domingos and Richardson 2001; Richardson and Domingos 2002; Kempe et al. 2003, 2005; Kimura and Saito 2006; Goyal et al. 2008; Chen et al. 2009, 2010a, b;

Published online: 21 March 2012

Goyal et al. 2010; Weng et al. 2010; Bross et al. 2011; Bakshy et al. 2011; Agarwal et al. 2011; Cha et al. 2011). One of the central problems in this domain is the problem of influence maximization (Kempe et al. 2003). Consider a social network in which we have accurate estimates of influence among users. Suppose we want to launch a new product in the market by targeting a set of influential users (e.g., by offering them the product at a discounted price), with the goal of starting a word-of-mouth viral propagation, exploiting the power of social connectivity. The idea is that by observing its neighbors adopting the product, or more generally, performing an action, a user may be influenced to perform the same action, with some probability. Influence thus propagates in steps according to one of the propagation models studied in the literature, e.g., the *independent cascade* (IC) or the *linear threshold* (LT) models (Kempe et al. 2003). The propagation stops when no new user gets activated.

In this context, we can identify three main dimensions—the number of *seed* nodes (or users) activated at the beginning (known as the *budget*), the expected number of nodes that eventually get activated (known as *coverage* or *expected spread*)[1], and the number of *time* steps required for the propagation. In their seminal paper, Kempe et al. (2003) introduced the problem of *influence maximization* (MAXINF) that asks for a seed set with a budget threshold $k$ that maximizes the expected spread (time being left unconstrained). They showed that under the standard propagation models IC and LT, MAXINF is **NP**-hard, but that a simple greedy algorithm that exploits properties of the propagation function yields a $(1 - 1/e - \phi)$ approximation, for any $\phi > 0$ (as discussed in detail in Sect. 2).

In this paper, we explore the other dimensions of influence propagation. The problem of minimum target set selection (MINTSS) is motivated by the observation that in a viral marketing campaign, we may be interested in the smallest budget that will achieve a desired outcome. The problem can therefore be defined as follows. We are given a threshold $\eta$ for the expected spread and the problem is to find a seed set of *minimum size* such that activating the set yields an expected spread of at least $\eta$.

In both MINTSS and MAXINF, the time for propagation is not considered. Indeed, with the exception of a few papers (see e.g., Leskovec et al. 2007), the temporal dimension of the social propagation phenomenon has been largely overlooked. This is surprising as the timeliness of a *viral marketing* campaign is a key ingredient for its success. Beyond viral marketing, many other applications in time-critical domains can exploit social networks as a means of communication to spread information quickly. This motivates the problem of minimum propagation time (MINTIME), defined as follows: given a budget $k$ and a coverage threshold $\eta$, find a seed set that satisfies the given budget and achieves the desired coverage in *as little time as possible*. Thus, MINTIME tries to optimize the propagation time required to achieve a desired coverage under a given budget.

### 1.1 Our contributions

We now summarize the main results in this paper.

- First, we show (Sect. 4, Theorem 1) that for all instances of MINTSS, where the coverage function is submodular, a simple greedy algorithm yields a bicriteria approximation: given a coverage threshold $\eta$ and a shortfall parameter $\epsilon > 0$, the greedy algorithm will produce a solution $S : \sigma(S) \geq \eta - \epsilon$ and $|S| \leq (1 + \ln(\eta/\epsilon))$OPT, where OPT is the optimal size of a seed set whose coverage is at least $\eta$; that is, the greedy solution exceeds the optimal solution in terms of size (budget) by a logarithmic factor while achieving a coverage that falls short of the required coverage by the shortfall parameter. We prove a generic hardness result (Sect. 4, Theorem 3) suggesting that improving this approximation factor is likely to be hard.
- For MINTIME under IC and LT model (or any model with monotone submodular coverage functions), we show that when we allow the coverage achieved to fall short of the threshold and the budget $k$ for number of seed nodes to be overrun by a logarithmic factor, then we can achieve the required coverage in the minimum possible propagation time, i.e., in the time achieved by the optimal solution to MINTIME with budget threshold $k$ and coverage threshold $\eta$ (Sect. 5, Theorem 6).
- On the other hand, for MINTIME under the IC model, we show that even bicriteria and tricriteria approximations are hard. More precisely, let $R_{\mathrm{OPT}}$ be the optimal propagation time required for achieving a coverage $\geq \eta$ within a budget of $k$. Then we show the following (Sect. 5, Theorem 4): there is unlikely to be a PTIME algorithm that finds a seed set with size under the budget, which achieves a coverage better than $(1 - 1/e)\eta$. Similarly, if we limit the budget overrun factor to less than $\ln(\eta)$, then it is unlikely that there is a PTIME algorithm that finds a seed set of size within the overrun budget which achieves a coverage better than $(1 - 1/e)\eta$. In both cases, the result holds even when we permit any amount of slack in the resulting propagation time.
- The above results are bicriteria bounds, in that they allow slack in two of the three parameters governing MINTIME problems. We also show a tricriteria hardness result (Sect. 5, Theorem 5); namely, if we

[1] We use the terms coverage and expected spread interchangeably throughout the article.

limit the budget overrun factor to be $\beta < \ln(\eta)$, then it is unlikely that there is a PTIME algorithm that finds a seed set with a size within a factor $\beta$ of the budget that achieves a coverage better than $(1 - 1/e^{\beta})\eta$. Similar bounds hold if we place hard limits on the coverage approximation and try to balance overrun in the other parameters.

- Often, the coverage function can be hard to compute exactly. This is the case for both IC and LT models (Kempe et al. 2003). All our results are robust in that they carry over even when only estimates of the coverage function are available.
- We show the value of our approximation algorithms by experimentally comparing their quality with that of several heuristics proposed in other contexts, using two real data sets. We discuss our findings in Sect. 6.

The necessary background is given in Sect. 2 while related work is discussed in Sect. 3. Section 7 concludes the paper and discusses interesting open problems.

## 2 Preliminaries

Suppose we are given a social network together with the estimates of mutual influence between individuals in the network, and suppose that we want to push a new product in the market. The mining problem of *influence maximization* is the following: given such a network with influence estimates, how to select the set of initial users so that they eventually influence the largest number of users in the social network. This problem has received a good deal of attention in the data mining and the theoretical computer science communities in the last decade.

The first to consider the propagation of influence and the problem of identification of influential users from a data mining perspective are Domingos and Richardson (2001), Richardson and Domingos (2002). The problem is modeled by means of *Markov random fields* and heuristics are given for choosing the users to target. In particular, the marketing objective function to maximize is the global expected lift in profit, that is, intuitively, the difference between the expected profit obtained by employing a marketing strategy and the expected profit obtained using no marketing at all. A Markov random field, is an undirected graphical model representing the joint distribution over a set of random variables, where nodes are variables, and edges represent dependencies between variables. It is adopted in the context of influence propagation by modeling only the final state of the network at convergence as one large global set of interdependent random variables.

Kempe et al. (2003) tackle roughly the same problem as a problem in discrete optimization. They obtain provable approximation guarantees under various propagation models studied in mathematical sociology, as we describe next.

A social network can be represented as a directed graph $G = (V, E)$. Every node is in one of two states—*active* or *inactive*. Here, "active" may correspond to a user buying a product or getting infected. In progressive models, it is assumed once a node becomes active, it remains active. Influence is assumed to propagate from nodes to their neighbors according to a *propagation model*, and a node's tendency to become active increases monotonically as more of its neighbors become active.

In the *independent cascade* (IC) model, each active neighbor $v$ of a node $u$ has one shot at influencing $u$ and succeeds with probability $p_{v,u}$, the probability with which $v$ influences $u$. In the *linear threshold* (LT) model, each node $u$ is influenced by each neighbor $v$ according to a weight $b_{v,u}$, such that the sum of incoming weights to $u$ is no larger than 1. Each node $u$ chooses a threshold $\theta_u$ uniformly at random from the interval [0,1]. If at timestamp $t$, the total weight from the active neighbors of $u$ attains the threshold $\theta_u$, then $u$ will become active at timestamp $t + 1$. In both the models, the process repeats until no new node becomes active.

For any propagation model, the *expected influence spread* of a seed set $S$ is the expected number of nodes that eventually get activated by initially activating the nodes $S$. We denote this number by $\sigma_m(S)$, where $m$ stands for the underlying propagation model. Then the *influence maximization problem* is defined as follows. Given a directed and edge-weighted social graph $G = (V, E)$, a propagation model $m$, and a number $k \leq |V|$, find a set $S \subseteq V$, $|S| = k$, such that $\sigma_m(S)$ is maximum.

Under both the IC and LT propagation models, this problem is shown to be **NP**-hard (Kempe et al. 2003). However, for both the propagation models described above, the expected influence spread function $\sigma_m(\cdot)$ is *monotone* and *submodular*. Monotonicity says as the set of activated nodes grows, the likelihood of a node getting activated should not decrease. More precisely, a A function $f$ from sets to reals is monotone if $f(S) \leq f(T)$ whenever $S \subseteq T$. A function $f$ is submodular if $f(S \cup \{w\}) - f(S) \geq f(T \cup \{w\}) - f(T)$ whenever $S \subseteq T$. Submodularity intuitively says an active node's probability of activating some inactive node $u$ does not increase if more nodes have already attempted to activate $u$ and $u$ is hence more "marketing-saturated". It is also called *the law of "diminishing returns"*.[2]

[2] A variant of the linear threshold model, where a *deterministic* threshold $\theta_u$ is chosen for each node, has also been studied (Chen 2008; Ben-Zwi et al. 2009). Coverage under this variant is not submodular.

Due to these two properties, we can have a simple greedy algorithm (see Algorithm 1) for influence maximization which provides an approximation guarantee. In fact, for any monotone submodular function $f$ with $f(\emptyset) = 0$ the problem of finding a set $S$ of size $k$ such that $f(S)$ is maximum, can be approximated to within a factor of $(1 - 1/e)$ by the greedy algorithm Nemhauser et al. (1978). This result carries over to the influence maximization problem Kempe et al. (2003), meaning that the seed set we produce using Algorithm 1 is guaranteed to have an expected spread $(1 - 1/e)$ i.e., >63 %, of the expected spread of the optimal seed set.

The complex step of the greedy algorithm is in line 3, where we select the node that provides the largest marginal gain $\sigma_m(S \cup \{v\}) - \sigma_m(S)$ with respect to the expected spread of the current seed set $S$. Computing the expected spread given a seed set is #**P**-hard under both the IC model (2010a) and the LT model (2010b). In their paper, Kempe et al. run Monte Carlo (MC) simulations of the propagation model for sufficiently many times (the authors report 10,000 trials) to obtain an accurate estimate of the expected spread, resulting in a very long computation time. In particular, they show that for any $\phi > 0$, there is a $\delta > 0$ such that using $(1 + \delta)$-approximate values of the expected spread, we can obtain a $(1 - 1/e - \phi)$-approximation for the influence maximization problem.

```
Algorithm 1 Greedy MAXINF
Input: $G, k, \sigma_m$
Output: seed set $S$
1: $S \leftarrow \emptyset$
2: while $|S| < k$ do
3:     $u \leftarrow \arg\max_{w \in V \setminus S}(\sigma_m(S \cup \{w\}) - \sigma_m(S))$;
4:     $S \leftarrow S \cup \{u\}$
```

We now define the problems we study in this paper. Let $m$ stand for any propagation model with a submodular coverage function $\sigma_m(\cdot)$.

**Problem 1** (MINTSS) Let $G = (V, E)$ be a social graph. Given a real number $\eta \leq |V|$, find a set $S \subseteq V$ of the smallest size $|S|$, such that the expected spread, denoted $\sigma_m(S)$, is no less than $\eta$.

**Problem 2** (MINTIME) Let $G = (V, E)$ be a social graph. Given an integer $k$, and a real number $\eta \leq |V|$, find a set $S \subseteq V, |S| \leq k$, and the smallest $t \in \mathbb{N}$, such that the expected spread at time $t$, denoted $\sigma_m^t(S)$, is no less than $\eta$.

The MINTSS problem is closely related to the real-valued submodular set cover (RSSC) problem. In the submodular set cover (SSC) problem we are given a monotone submodular set function $f$ over a ground set $\mathcal{X}$ and a modular cost function $c$ over $\mathcal{X}$ and are asked to find a set $S \subseteq \mathcal{X}$ minimizing $c(S) := \sum_{s \in S} c(s)$ subject to the constraint that $f(S) = f(\mathcal{X})$. The RSSC problem, which generalizes SSC, is defined as follows: given a submodular function $f : 2^{\mathcal{X}} \rightarrow \mathbb{R}$ and a threshold $\eta$, find a set $S \subseteq \mathcal{X}$ of the least size (or minimum cost, when elements of $\mathcal{X}$ are weighted) such that $f(S) \geq \eta$. MINTSS under any propagation model such as IC and LT, for which the coverage function is submodular is clearly a special case of RSSC, an observation we exploit in Sect. 4.

MINTIME is closely related to the robust asymmetric $k$-center (RAKC) problem in directed graphs, defined as follows: given a digraph $G = (V, E)$, a (possibly empty) set of forbidden nodes and thresholds $k$ and $\eta$, find $k$ or fewer nodes $S$ such that they cover at least $\eta$ non-forbidden nodes in the minimum possible radius, i.e., each of the $\eta$ nodes are reachable from some node in $S$ in the minimum possible distance.

## 3 Related work

Although to the best of our knowledge, MINTIME has never been studied before, some work has been devoted to MINTSS (Chen 2008; Ben-Zwi et al. 2009). Both studies focus on a variant of LT model (DLT model) where a deterministic threshold $\theta_u$ is chosen for each node. Although coverage under the LT model is submodular, it is not submodular under the DLT model. There have been some recent works that evaluate classical propagation models like IC and LT against real world datasets (Goyal et al. 2011; Bhagat et al. 2012). We are not aware of any such studies on the DLT model. In this paper, we study both MINTSS and MINTIME under the classic propagation models IC and LT, under which the coverage function is submodular.

Chen (2008) shows that under the DLT propagation model, MINTSS cannot be approximated within a factor of $O(2^{\log^{1-\delta} n})$ unless NP $\subseteq$ DTIME$(n^{O(\log\log n)})$, and also gives a polynomial time algorithm for MINTSS on trees. Ben-Zwi et al. (2009) build upon Chen (2008) and develop a $O(n^{O(w)})$ algorithm for solving MINTSS exactly under the DLT model, where $w$ is the tree width of the graph. They show the problem cannot be solved in $n^{O(\sqrt{w})}$ time unless all problems in SNP can be solved in sub-exponential time.

A few classical cover problems are related to the problems we study. One such problem is maximum coverage (MC): given a collection of sets $\mathcal{S}$ over a ground set $\mathcal{U}$ and budget $k$, find a subcollection $\mathcal{C} \subseteq \mathcal{S}$ such that $|\mathcal{C}| \leq k$ and $|\bigcup \mathcal{C}|$ is maximized. The problem can be approximated within a factor of $(1 - 1/e)$ and it cannot be improved (Feige 1998; Khuller et al. 1999). Similar results by

Khuller et al. (1999) and Sviridenko (2004) exist for the weighted case.

Another relevant problem is partial set cover (PSC): given a collection of sets $\mathcal{S}$ over the ground set $\mathcal{U}$ and a threshold $\eta$, the goal is to find a subcollection $\mathcal{C} \subseteq \mathcal{S}$ such that $|\bigcup \mathcal{C}| \geq \eta$ and $|\mathcal{C}|$ is minimized. Although PSC can be approximated within a factor of $\lceil \ln \eta \rceil$, Feige (1998) showed that it cannot be approximated within a factor of $(1-\delta)\ln \eta$, for any fixed $\delta > 0$, unless NP $\subseteq$ DTIME$(n^{O(\log \log n)})$.

Our results on MINTSS exploit its connection to the real-valued submodular set cover (RSSC) problem. There has been substantial work on submodular set cover (SSC) in the presence of integer-valued submodular functions (Fujito 1999, 2000; Feige 1998; Slaví k 1997; Bar-Ilan et al. 2001). Relatively much less work has been done on real-valued SSC. For non-decreasing real-valued submodular functions, Wolsey (1982) has shown, among other things, that a simple greedy algorithm yields a solution to a special case of SSC where $\eta = f(\mathcal{X})$, that is within a factor of $\ln[\eta/(\eta - f(S_{t-1}))]$ of the optimal solution, where $t$ is the number of iterations needed by the greedy algorithm to achieve a coverage of $\eta$ and $S_i$ denotes the greedy solution after $i$ iterations. Unfortunately, this result by itself does *not* yield an approximation algorithm with any guaranteed bounds: in Appendix B we give an example to show that the greedy solution can be arbitrarily worse than the optimal one. Furthermore, Wolsey's analysis is restricted to the case $\eta = f(\mathcal{X})$. Along the way to establish our results on MINTSS, we show the greedy algorithm yields a bicriteria approximation for real-valued SSC that extends to the general case of partial cover with $\eta \leq f(\mathcal{X})$, and where elements are weighted.

Our results on MINTIME leverage its connection to the robust asymmetric $k$-center problem (RAKC). It has been shown that, while asymmetric $k$-center problem can be approximated within a factor of $O(\log^* n)$ (Panigrahy and Vishwanathan 1998), RAKC cannot be approximated within any factor unless $P = \text{NP}$ (Li Gørtz and Wirth 2006).

## 4 Minimum target set selection

### 4.1 A bicriteria approximation

Our main result of this section is that a simple greedy algorithm, Algorithm Greedy-Mintss, yields a bicriteria approximation to (weighted) MINTSS, for any propagation model whose coverage function is monotone and submodular.

**Algorithm 2** Greedy-Mintss

```
Input: $G, \eta, \epsilon, \sigma_m$
Output: seed set $S$
$S \leftarrow \emptyset$
while $\sigma_m(S) < \eta - \epsilon$ do
    $u \leftarrow \arg\max_{w \in V \setminus S}(\frac{\min(\sigma_m(S \cup \{w\}), \eta) - \sigma_m(S)}{c(w)})$;
    $S \leftarrow S \cup \{u\}$
```

To prove the results in the most general setting, we consider digraphs $G = (V, E)$ which have non-negative node weights: we are given a cost function $c : V \rightarrow \mathbb{R}^+$ in addition to the coverage threshold $\eta$, and need to find a seed set $S$ such that $\sigma_m(S) \geq \eta$ and $c(S) = \sum_{x \in S} c(x)$ is minimum. Clearly, this generalizes the unweighted case.

**Theorem 1** *Let $G = (V, E)$ be a social graph, with node weights given by $c : V \rightarrow \mathbb{R}^+$. Let $m$ be any propagation model whose coverage function $\sigma_m(\cdot)$ is monotone and submodular. Let $S^*$ be a seed set of minimum cost such that $\sigma_m(S^*) \geq \eta$. Let $\epsilon > 0$ be any shortfall and let $S$ be the greedy solution with chosen threshold $\eta - \epsilon$. Then, $c(S) \leq c(S^*) \cdot (1 + \ln(\eta/\epsilon))$.*

In the rest of this section, we prove this result. We first observe that every instance of MINTSS where the coverage function $\sigma_m(\cdot)$ is monotone and submodular is an instance of RSSC. Thus, it suffices to prove Theorem 1 for RSSC, for which we adapt a bicriterion approximation technique by Slaví k (1997).

Let $\mathcal{X} = \{x_1, x_2, \ldots, x_m\}$ be a ground set, $c : \mathcal{X} \rightarrow \mathbb{R}^+$ be a cost function, $f : 2^{\mathcal{X}} \rightarrow \mathbb{R}$ a non-negative monotone submodular function and $\eta$ a given threshold. Apply the greedy algorithm above to this instance of RSSC. Let $S_i$ be the (partial) solution obtained by the greedy algorithm after $i$ iterations. Let $t$ be the smallest number such that $f(S_t) \geq \eta$. We define $g(S) = \min(f(S), \eta)$. Clearly, $g$ is also monotone and submodular. In each iteration, the greedy algorithm picks an element which provides the maximum marginal gain per unit cost (with respect to $g$), i.e., it picks an element $x$ for which $\frac{g(S \cup \{x\}) - g(S)}{c(x)}$ is positive and is maximum.

Let $c(S^*) = \kappa$ and define $\eta_i = \eta - g(S_i)$, i.e., the shortfall in coverage after $i$ iterations of the greedy algorithm.

**Lemma 1** *At the end of iteration $i$, there is an element $x \in \mathcal{X} \setminus S_i : \frac{g(S_i \cup \{x\}) - g(S_i)}{c(x)} \geq \frac{\eta_i}{\kappa}$.*

*Proof* Let $S_i^* = S^* - S_i$. Let $S_i^* = \{y_1, \ldots, y_t\}$ and $c(S_i^*) = \kappa_i$. Suppose $\forall x \in \mathcal{X} \setminus S_i : \frac{g(S_i \cup \{x\}) - g(S_i)}{c(x)} < \frac{\eta_i}{\kappa}$. Consider adding the elements in $S_i^*$ to $S_i$ one by one. Clearly, at any step $j \leq t$, we have by submodularity that

$$g(S_i \cup \{y_1, \ldots, y_j\}) - g(S_i \cup \{y_1, \ldots, y_{j-1}\}) \leq g(S_i \cup \{y_j\}) \\ - g(S_i) < c(y_j) \cdot \frac{\eta_i}{\kappa}$$

Iterating over all $j$, this yields $g(S_i \cup \{y_1, \ldots, y_j\}) - g(S_i) < \frac{\eta_i}{\kappa} \cdot (c(y_1) + \cdots + c(y_j))$ resulting in $g(S_i \cup \{y_1, \ldots, y_t\}) < g(S_i) + \frac{\eta_i}{\kappa} \cdot \sum_{1 \leq j \leq t} c(y_j) \leq \eta$ which is a contradiction since the left hand side is no less than the optimal coverage. □

*Proof of Theorem 1* It follows from Lemma 1 that $\eta_i \leq \eta_{i-1}$ $(1 - c_i/\kappa)$ where $c_i$ is the cost of the element added in iteration $i$. Using the well known inequality $(1 + z) \leq e^z$, $\forall$ $z$, we get $\eta_i \leq \eta_{i-1} \cdot e^{-c_i/k}$. Expanding, $\eta_i \leq \eta \cdot e^{-\frac{1}{k} \cdot \sum_i c_i}$. Let the algorithm take $l$ iterations to achieve coverage $g(S_l) \geq \eta - \epsilon$ such that $g(S_{l-1}) < \eta - \epsilon$. At any step, $g(S_{i+1}) - g(S_i) \leq \eta_i$. Thus, $c_i \leq \kappa$, and in particular, the cost of the last element picked can be at most $\kappa$. So, $c(S_l) \leq \kappa + c(S_{l-1})$. $g(S_{l-1}) < \eta - \epsilon$ implies $\eta_{l-1} > \epsilon$. Hence, we have $\eta e^{-\frac{1}{\kappa} c(S_{l-1})} > \epsilon$ which implies $c(S_{l-1}) < \kappa \ln(\eta/\epsilon)$. Thus, $c(S_l) \leq \kappa(1 + \ln(\eta/\epsilon))$. □

Using a similar analysis, it can be shown that when the costs are uniform, the approximation factor can be improved to $\lceil \ln(\eta/\epsilon) \rceil$.

For propagation models like IC and LT, computing the coverage $\sigma_m(S)$ exactly is **#P**-hard (Chen et al. 2010a, b) and thus we must settle for estimates. To address this, we "lift" the above theorem to the case where only estimates of the function $f(.)$ are available. We can show

**Theorem 2** *For any* $\phi > 0$, *there exists a* $\delta \in (0, 1)$ *such that using* $(1 - \delta)$*-approximate values for the coverage function* $\sigma_m(\cdot)$, *the greedy algorithm approximates MINTSS under IC and LT models within a factor of* $(1 + \phi) \cdot (1 + \ln(\eta/\epsilon))$.

*Proof* The proof involves a more careful analysis of how error propagates in the greedy algorithm if, because of errors, the greedy algorithm picks the wrong point.

Here, we give the proof for the unit cost version only. Consider any monotone, submodular function $f(\cdot)$. Thus, in the statement of theorem, $\sigma_m(\cdot) = f(\cdot)$. Let $f'(\cdot)$ be its approximated value. In any iteration, the (standard) greedy algorithm picks an element which provides maximum marginal gain. Let $S_i$ be the set formed after iteration $i$.

As we did in Lemma 1, it is straightforward to show that there must exists an element $x \in \mathcal{X} \setminus S_i$ such that $f(S_i \cup \{x\}) - f(S_i) \geq \eta_i/k$ where $\eta_i = \eta - f'(S_i)$. Without the loss of generality, let $x$ be the element which provides the maximum marginal gain. Suppose that due to the error in computing $f(.)$, some other element $y$ is picked instead. Then,

$$(1 - \delta)f(S_i \cup \{x\}) \leq f'(S_i \cup \{x\}) \leq f'(S_i \cup \{y\})$$

Moreover, $f'(S_i) \leq f(S_i)$. Thus,

$$\frac{\eta_i}{k} \leq f(S_i \cup \{x\}) - f(S_i) \leq \frac{f'(S_i \cup \{y\})}{1 - \delta} - f'(S_i)$$
$$\Rightarrow \frac{\eta_i}{k} \leq \frac{\eta - \eta_{i+1}}{1 - \delta} - \eta + \eta_i$$
$$\Rightarrow \eta_{i+1} \leq \eta_i \cdot (1 - \delta) \cdot \left(1 - \frac{1}{k}\right) + \delta \cdot \eta$$
$$\Rightarrow \eta_{i+1} \leq \eta \cdot (1 - \delta)^{i+1} \cdot \left(1 - \frac{1}{k}\right)^{i+1} \\ + \delta \cdot \eta \cdot \left(\frac{1 - (1 - \delta)^{i+1}(1 - 1/k)^{i+1}}{1 - (1 - \delta)(1 - 1/k)}\right)$$

Let $\delta' = \delta/(1 - (1 - \delta)(1 - 1/k))$. Let the greedy algorithm takes 1 iterations. Then,

$$\eta_l \leq \eta \cdot (1 - \delta)^l \cdot \left(1 - \frac{1}{k}\right)^l \\ + \delta' \cdot \eta \cdot \left(1 - (1 - \delta)^l \cdot \left(1 - \frac{1}{k}\right)^l\right) \\ = \eta \cdot (1 - \delta)^l \cdot \left(1 - \frac{1}{k}\right)^l (1 - \delta') + \delta' \cdot \eta$$

Using $(1 - \delta)^l \leq 1$ and $(1 - 1/k)^l \leq e^{-l/k}$,

$$\eta_l \leq \eta e^{-l/k}(1 - \delta') + \delta' \cdot \eta$$

The algorithm stops when $\eta_l \leq \epsilon$. The maximum number of iterations needed to ensure this are

$$l \leq k\left(1 + \ln \frac{\eta(1 - \delta')}{\epsilon(1 - \delta'\eta/\epsilon)}\right)$$

Let $x = \eta/\epsilon$. To prove the lemma, we need to prove that for any $\phi > 0$, there exists $\delta \in [0, 1)$ such that

$$x^{1+\phi} = x\frac{1 - \delta'}{1 - \delta' x} \Rightarrow \delta' = \frac{x^{\phi} - 1}{x^{1+\phi} - 1}$$

Clearly, for any $\phi \geq 0, \delta' \in [0, 1)$. Hence,

$$0 \leq \delta < 1 - (1 - \delta)(1 - 1/k) \\ \iff 0 \leq \delta < 1$$

This completes the proof for unit cost case. Using the slight modification in the greedy algorithm (as we did in proving Theorem 1), the same result can be obtained for weighted version. □

## 4.2 An inapproximability result

Recall that every instance of MINTSS where the coverage function is monotone and submodular is an instance of RSSC. Consider the unweighted version of the RSSC problem. Let $S^*$ denote an optimal solution and let $\mathrm{OPT} = |S^*|$.

**Theorem 3** *For any fixed $\delta > 0$, there does not exist a PTIME algorithm for RSSC that guarantees a solution $S$ : $|S| \leq \mathrm{OPT}(1-\delta)\ln(\eta/\epsilon)$, and $f(S) \geq \eta - \epsilon$ for any $\epsilon > 0$ unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$.*

*Proof* **Case 1:** $\epsilon \geq 1$. Suppose there exists an algorithm $\mathcal{A}$ that finds a solution $S$ of size $\leq \mathrm{OPT}(1-\delta)\ln(\eta/\epsilon)$ such that $f(S) \geq \eta - \epsilon$ for any $\epsilon \geq 1$. Consider an arbitrary instance $\mathcal{I} = \langle \mathcal{U}, \mathcal{S}, \eta \rangle$ of PSC, which is a special case of RSSC. Apply the algorithm $\mathcal{A}$ to $\mathcal{I}$. It outputs a collection of sets $\mathcal{C}_1 : |\mathcal{C}_1| \leq \mathrm{OPT}(1-\delta)\ln(\eta/\epsilon)$ that covers $\geq \eta - \epsilon$ elements in $\mathcal{U}$.

Create a new instance $\mathcal{J} = \langle \mathcal{U}', \mathcal{S}', \eta' \rangle$ of PSC as follows. Let $T = \bigcup \mathcal{C}_1$ be the set of elements of $\mathcal{U}$ covered by $\mathcal{C}_1$. Define $\mathcal{S}' = \{S \setminus T \mid S \in \mathcal{S} \setminus \mathcal{C}_1\}, \mathcal{U}' = \mathcal{U} \setminus T$ and $\eta' = \epsilon$. Set the new shortfall $\epsilon' = 1$. Apply the algorithm $\mathcal{A}$ to $\mathcal{J}$. It will output another collection of sets $\mathcal{C}_2 : |\mathcal{C}_2| \leq \mathrm{OPT}(1-\delta)\ln\epsilon$ which covers $\geq \epsilon - 1$ elements in $\mathcal{U}'$.[3] Let $\mathcal{C} = \mathcal{C}_1 \cup \mathcal{C}_2$. The number of elements covered by $\mathcal{C}$ is $\geq \eta - \epsilon + \epsilon - 1 = \eta - 1$. Clearly, $|\mathcal{C}| = |\mathcal{C}_1| + |\mathcal{C}_2| \leq \mathrm{OPT}(1-\delta)\ln(\eta/\epsilon) + \mathrm{OPT}(1-\delta)\ln(\epsilon) = \mathrm{OPT}(1-\delta)\ln(\eta)$. Thus, we have a solution for PSC with the approximation factor of $(1-\delta)\ln(\eta)$, which is not possible unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$ (Feige 1998). This proves Case 1.

**Case 2:** $\epsilon < 1$. Assume an arbitrary instance $\mathcal{I}$ of RSSC with monotone submodular function $f : 2^{\mathcal{X}} \to \mathbb{R}$. Let $\eta'$ be the coverage threshold and $\epsilon' \geq 1$ be any given shortfall. We now construct another instance $\mathcal{J}$ of RSSC as follows: Set the coverage function $g(S) = f(S)/x$, coverage threshold $\eta = \eta'/x$ and shortfall $\epsilon = \epsilon'/x$. Choose any value of $x > 1$ such that $\epsilon = \epsilon'/x < 1$. We now show that if a solution is a $(1-\delta)\ln(\eta/\epsilon)$-approximation to the optimal solution for $\mathcal{J}$ then it is a $(1-\delta)\ln(\eta'/\epsilon')$-approximation to the optimal solution for $\mathcal{I}$. Clearly, the optimal solution for both the instances are identical, so $\mathrm{OPT}_{\mathcal{I}} = \mathrm{OPT}_{\mathcal{J}}$.[4] Suppose there exists an algorithm for RSSC when the shortfall is $\epsilon \in (0,1)$, that guarantees a solution $S$ : $|S| \leq \mathrm{OPT}(1-\delta)\ln(\eta/\epsilon)$ and $f(S) \geq \eta - \epsilon$. Apply this algorithm to instance $\mathcal{J}$ to obtain a solution $S_{\mathcal{J}}$. We have: $g(S_{\mathcal{J}}) \geq \eta - \epsilon = (\eta' - \epsilon')/x$. It implies $f(S_{\mathcal{J}}) = x \cdot g(S_{\mathcal{J}}) \geq \eta' - \epsilon'$. Moreover, $|S_{\mathcal{J}}| \leq \mathrm{OPT}_{\mathcal{J}}(1-\delta)\ln(\eta/\epsilon)$, implying $|S_{\mathcal{J}}| \leq \mathrm{OPT}_{\mathcal{I}}(1-\delta)\ln(\eta'/\epsilon')$. Thus we have the solution $S_{\mathcal{J}}$ for instance $\mathcal{I}$ whose size is $\leq \mathrm{OPT}_{\mathcal{I}}(1-\delta)\ln(\eta'/\epsilon')$. The theorem follows. □

In view of this generic result, we conjecture that improving the approximation factor for MINTSS to $(1-\delta)\ln(\eta/\epsilon)$ for IC and LT is likely to be hard.

## 5 MINTIME

In this section, we study MINTIME under the IC model. Denote by $\sigma_m^R(S)$ the expected number of nodes activated under model $m$ within time $R$, and let $\eta$ be the desired coverage and $k$ be the desired budget. Let $R_{\mathrm{OPT}}$ denote the optimal propagation time under these budget and coverage constraints. Our first result says that efficient approximation algorithms are unlikely to exist under two scenarios: (1) when we allow a coverage shortfall of less than $\eta/e$ and (2) when we allow a budget overrun less than $\ln\eta$. In the former scenario, we have a strict budget threshold and in the latter we have a strict coverage threshold. In both cases, we allow any amount of slack in propagation time.

**Theorem 4** *Unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$, there does not exist a PTIME algorithm for MINTIME that guarantees (for any $\alpha \geq 1$):*

1. a $(\alpha, \gamma)$-approximation, such that $|S| \leq k, R = \alpha \cdot R_{\mathrm{OPT}}$ and $\sigma_m^R(S) \geq \gamma \cdot \eta$ where $\gamma = (1 - 1/e + \delta)$ for any fixed $\delta > 0$; or
2. a $(\alpha, \beta)$-approximation, such that $|S| \leq \beta \cdot k, R = \alpha \cdot R_{\mathrm{OPT}}$ and $\sigma_m^R(S) \geq \eta$ where $\beta = (1-\delta)\ln\eta$ for any fixed $\delta > 0$.

Our second theorem says efficient approximation algorithms are unlikely to exist under more liberal scenarios than those given above: (1) when for a given budget overrun factor $\beta < \eta$, the fraction of the coverage we want to achieve is more than $1 - 1/e^{\beta}$ and (2) when for a given fraction $\gamma \in (0, 1 - 1/\eta]$ of the coverage we want to achieve, the budget overrun factor we allow is less than $\ln(1/(1-\gamma))$. As before, we allow any amount of slack in propagation time.

**Theorem 5** *Unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$ there does not exist a PTIME algorithm for MINTIME that guarantees $(\alpha, \beta, \gamma)$-approximation factor (for any $\alpha \geq 1$) such that $|S| \leq \beta \cdot k, R = \alpha \cdot R_{\mathrm{OPT}}$ and $\sigma_m^R(S) \geq \gamma \cdot \eta$ where*

1. $\beta \in [1, \ln\eta)$ and $\gamma = 1 - 1/e^{\beta} + \delta$ for any fixed $\delta > 0$; or
2. $\gamma \in \left(0, 1 - \frac{1}{\eta}\right]$ and $\beta = (1-\delta)\ln\left(\frac{1}{1-\gamma}\right)$ for any fixed $\delta > 0$.

Finally, on the positive side, we show that when a coverage shortfall of $\epsilon > 0$ is allowed and a budget boost of $(1 + \ln(\eta/\epsilon))$ is allowed, we can in PTIME find a solution which achieves the relaxed coverage under the relaxed

[3] If $\epsilon = 1$, $\mathcal{A}$ outputs an empty collection.

[4] Here, $\mathrm{OPT}_{\mathcal{I}}$ and $\mathrm{OPT}_{\mathcal{J}}$ represent the size of the optimal solution for instances $\mathcal{I}$ and $\mathcal{J}$ respectively.

budget in optimal propagation time. More precisely, we have:

**Theorem 6** *Let the chosen coverage threshold be $\eta - \epsilon$, for $\epsilon > 0$ and chosen budget threshold be $k(1 + \ln(\eta/\epsilon))$. If the coverage function $\sigma_m^R(\cdot)$ can be computed exactly, then there is a greedy algorithm that approximates the MINTIME problem within a $(\alpha, \beta, \gamma)$ factor where $\alpha = 1, \beta = 1 + \ln(\eta/\epsilon)$ and $\gamma = 1 - \epsilon/\eta$ for any $\epsilon > 0$. Furthermore, for every $\phi > 0$, there is a $\delta > 0$ such that using a $(1 - \delta)$-approximate values for the coverage function $\sigma_m^R(\cdot)$, the greedy algorithm approximates the MINTIME problem within a $(\alpha, \beta, \gamma)$ factor where $\alpha = 1, \beta = (1 + \phi)(1 + \ln(\eta/\epsilon))$ and $\gamma = 1 - \epsilon/\eta$.*

### 5.1 Inapproximability proofs

We next prove Theorems 4 and 5. We first show that MINTIME under the IC model generalizes the RAKC problem. In a digraph $G = (V, E)$ and sets of nodes $S, T \subset V$, say that *R-covers* $T$ if for every $y \in T$, there is a $x \in S$ such that there is a path of length $\leq R$ from $x$ to $y$. Given an instance of RAKC, create an instance of MINTIME by labeling each arc in the digraph with a probability 1. Now, it is easy to see that for any set of nodes $S$ and any $0 \leq R \leq n - 1$, $SR$-covers a set of nodes $T$ iff activating the seed nodes $S$ will result in the set of nodes $T$ being activated within $R$ time steps. Notice that since all the arcs are labeled with probability 1, all influence attempts are successful by construction. It follows that RAKC is a special case of MINTIME under the IC model.

The tricriteria inapproximability results of Theorem 5 subsume the bicriteria inapproximability results of Theorem 4. Still, in our presentation, we find it convenient to develop the proofs first for bicriteria. Since we showed that MINTIME under IC generalizes RAKC, it suffices to prove the theorems in the context of RAKC. It is worth pointing out Li Gørtz and Wirth (2006) proved that it is hard to approximate *RAKC* within any factor unless $P = \text{NP}$. Their proof only applies to (the standard) unicriterion approximation.

For a set of nodes $S$ in a digraph we denote by $f^R(S)$ the number of nodes that are $R$-covered by $S$. Recall the problems MC and PSC (see Sect. 3).

*Proof of Theorem 4* It suffices to prove the theorem for RAKC. For claim 1, we reduce Maximum Coverage (MC) to RAKC and for claim 2, we reduce PSC to RAKC. The reduction is similar and is as follows: Consider an instance of the decision version of MC (equivalently PSC) $\mathcal{I} = \langle \mathcal{U}, \mathcal{S}, k, \eta \rangle$, where we ask whether there exists a subcollection $\mathcal{C} \subseteq \mathcal{S}$ of size $\leq k$ such that $|\cup_{S \in \mathcal{S}} S| \geq \eta$ Construct an instance $\mathcal{J} = \langle \mathcal{G}, k', \eta' \rangle$ of RAKC as follows: the graph $\mathcal{G}$ consists of two classes of nodes – $A$ and $B$. For each $S \in \mathcal{S}$, create a class A node $v_S$ and for each $u \in U$, create a class B node $v_u$. There is a directed edge $(v_S, v_u)$ of unit length iff $u \in S$. Notice, a set of nodes $S$ in $\mathcal{G}R$-covers another non-empty set of nodes iff $S1$-covers the latter set. Moreover, $x$ sets in $\mathcal{S}$ cover $y$ elements in $\mathcal{U}$ iff $\mathcal{G}$ has a set of $x$ nodes which 1-covers $y + x$ nodes. The only-if direction is trivial. For the if direction, the only way $x$ nodes can 1-covers $y + x$ nodes in $\mathcal{G}$ is when the $x$ nodes are from class A.

Next, we prove the first claim. Set $k' = k$ and $\eta' = \eta + k$. Assume there exists a PTIME $(\alpha, \gamma)$-approximation algorithm $\mathcal{A}$ for RAKC such that $f^R(S) \geq (1 - 1/e + \delta) \cdot (\eta')$ for any fixed $\delta > 0$, for some $R \leq \alpha\ R_{\text{OPT}}$. Apply algorithm $\mathcal{A}$ to the instance $\mathcal{J}$. Notice, for our instance, $R_{\text{OPT}} = 1$. The coverage by the output seed set $S$ will be $f^R(S) \geq (1 - 1/e + \delta) \cdot (\eta + k)$ nodes, for some $R \leq \alpha \cdot 1$, implying that the number of class B nodes covered is $\geq (1 - 1/e + \delta) \cdot (\eta + k) - k = (1 - 1/e + \delta - (1/e - \delta) k/\eta)\eta$. Thus the algorithm approximates MC within a factor of $\left(1 - \frac{1}{e} + \delta - \left(\frac{1}{e} - \delta\right)\frac{k}{\eta}\right)$. Let $\delta' = \delta - \left(\frac{1}{e} - \delta\right)\frac{k}{\eta}$. If we show $\delta' > 0$, we are done, since MC cannot be approximated within a factor of $(1 - 1/e + \delta')$ for any $\delta' > 0$ unless $\text{NP} \subseteq \text{DTIME}(n^{O(\log \log n)})$ (Feige 1998; Khuller et al. 1999). Clearly, $\delta'$ is not always positive. However, for a given $\delta$ and $k$, $\delta'$ is an increasing function of $\eta$ and reaches $\delta$ in the limit. Hence there is a value $\eta_0 : \forall \eta \geq \eta_0, \delta' > 0$. That is, there are infinitely many instances of PSC for which $\mathcal{A}$ is a $(1 - 1/e + \delta')$-approximation algorithm, where $\delta' > 0$, which proves the first claim.

Next, we prove the second claim. Set $k' = k$ and $\eta' = \eta + x$. The value of $x$ will be decided later. Assume there exists a PTIME $(\alpha, \beta)$-approximation algorithm $\mathcal{A}$ for RAKC where $\beta = (1 - \delta)\ln(\eta')$ for any fixed $\delta > 0$. Apply the algorithm to $\mathcal{J}$. It gives a solution $S$ such that $|S| \leq k \cdot (1 - \delta)\ln(\eta + x)$ that covers $\geq \eta + x$ nodes. A difficulty arises here since $\delta$ can be arbitrarily close to 1 making $k \cdot (1 - \delta)\ln(\eta + x)$ arbitrarily small, for any given $\eta$ and $k$. However, as we argued in the proof of claim 1, for sufficiently large $\eta$, we can always find an $x : k \leq x \leq k \cdot (1 - \delta)\ln(\eta + x)$. That is, on infinitely many instances of PSC, algorithm $\mathcal{A}$ finds a set of $|S|$ class A nodes which $R$-covers $\eta + x$ nodes, for some $R \leq \alpha \cdot 1$. Without loss of generality, we can assume $x \leq \eta$. Choose the smallest value of $x$ such that the solution $S$ covers $\geq \eta$ class B nodes. This implies the number of class A nodes covered is $\leq x$ and so $|S| \leq x$. Thus, on all such instances, algorithm $\mathcal{A}$ gives a solution $S$ of size $\leq x : k \leq x \leq k \cdot (1 - \delta)\ln(\eta + x)$ that covers $\geq \eta$ nodes. If we show that the upper bound is equal to $k \cdot (1 - \delta')\ln \eta$ for some $\delta' > 0$, we are done, since PSC cannot be approximated within a factor of $(1-\delta')\ln\eta$ unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$ (Feige 1998).

Let $(1-\delta')\ln\eta = (1-\delta)\ln(\eta+x)$, which yields $\delta' = 1-(1-\delta)\frac{\ln(\eta+x)}{\ln\eta}$. It is easy to see that by choosing sufficiently large $\eta$, we can make the gap between $\delta$ and $\delta'$ arbitrarily small and thus can always ensure $\delta' > 0$ on infinitely many instances of PSC, on each of which algorithm $\mathcal{A}$ will serve as an $(1-\delta')\ln\eta$-approximation algorithm proving claim 2. □

Note, in the proofs of both claims 1 and 2 in the above theorem, by choosing $\eta$ sufficiently large, we can always ensure for any given $k$ and $\delta > 0$, the corresponding $\delta'$ is always greater than 0. To prove the tricriteria hardness results, we need the following lemma.

**Lemma 2** *In the MC (or PSC) problem, let $k$ be the minimum number of sets needed to cover $\geq \eta$ elements. Then, unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$, there does not exist a PTIME algorithm that is guaranteed to select $\beta\, k$ sets covering $\geq \gamma\, \eta$ elements where*

1. $\beta \in [1, \ln\eta)$ and $\gamma > 1 - 1/e^{\beta}$; or
2. $\gamma \in \left(0, 1-\frac{1}{\eta}\right]$ and $\beta = (1-\delta)\ln\left(\frac{1}{1-\gamma}\right)$ for any fixed $\delta > 0$.

Lemma 2 is proved in Appendix A. We are ready to prove Theorem 5.

*Proof of Theorem 5* Again, it suffices to prove the theorem for RAKC. For claim 1, we reduce MC to RAKC and for claim 2, we reduce PSC to RAKC. The reduction is the same as in the proof of Theorem 4 and we skip the details here. Below, we refer to instances $\mathcal{I}$ and $\mathcal{J}$ as in that proof.

We first prove claim 1. Given any $\beta$, set $k' = k$ and $\eta' = \eta + \beta k$. Assume there exists a PTIME ($\alpha$, $\beta$, $\gamma$)-approximation algorithm $\mathcal{A}$ for RAKC which approximates the problem within the factors as mentioned in claim 1. Apply algorithm $\mathcal{A}$ to the instance $\mathcal{J}$. The coverage by the output seed set $S$ will be $f^R(S) \geq (1 - 1/e^{\beta} + \delta)\cdot(\eta + \beta k)$ nodes, implying the number of class B nodes covered is $\geq (1-1/e^{\beta}+\delta)\cdot(\eta+\beta k) - \beta k = (1 - 1/e^{\beta} + \delta - (1/e^{\beta} - \delta)\beta k/\eta)\eta$. Thus the algorithm approximates MC within a factor of $\left(1 - \frac{1}{e^{\beta}} + \delta - \left(\frac{1}{e^{\beta}} - \delta\right)\frac{\beta k}{\eta}\right)$.

If we show $\delta - \left(\frac{1}{e^{\beta}} - \delta\right)\frac{\beta k}{\eta} > 0$, then the claim follows, since MC cannot be approximated within a factor of $(1 - 1/e^{\beta} + \delta')$ for any $\delta' > 0$ unless $\mathrm{NP} \subseteq \mathrm{DTIME}(n^{O(\log\log n)})$, by Lemma 2. Let $\delta' = \delta - \left(\frac{1}{e^{\beta}} - \delta\right)\frac{\beta k}{\eta}$. For any $\beta \in [1, \ln\eta)$, $\delta'$ is an increasing function of $\eta$ which approaches $\delta$ in the limit. Thus, given any fixed $\delta > 0$, there must exist some $\eta_o$ such that for any $\eta \geq \eta_o, \delta' > 0$. This proves the first claim (by an argument similar to that in Theorem 4).

Next, we prove the second claim. Set $k' = k$ and $\eta' = \eta + x$. The value of $x$ will be decided later. Assume that there exists a PTIME ($\alpha$, $\beta$, $\gamma$)-approximation algorithm $\mathcal{A}$ for RAKC where the factors $\alpha$, $\beta$ and $\gamma$ satisfy the conditions as mentioned in claim 2. Apply the algorithm to instance $\mathcal{J}$. For any $\gamma_j \in (0, 1 - 1/(\eta + x)]$, it gives a solution of size $\leq k\cdot(1-\delta)\ln(1/(1-\gamma_j))$ that covers $\gamma_j\cdot(\eta+x)$ nodes. There can be $|\mathcal{S}|$ possible choices of $x$. Pick the smallest $x$ such that number of nodes covered in class B is at least $\gamma_j\, \eta$, implying that the number of nodes picked from class A is $\gamma_j\, x$. Thus, $\gamma_j x \leq k\cdot(1-\delta)\ln(1/(1-\gamma_j))$. The existence of $x$ satisfying this inequality can be established as done for claim 2 in Theorem 4.

Thus, algorithm $\mathcal{A}$ gives the solution instance $\mathcal{I}$ of size $\leq k\cdot(1-\delta)\ln(1/(1-\gamma_j))$ that covers $\gamma_j\, \eta$ elements in $\mathcal{U}$ where $\gamma_j \in (0, 1-1/(\eta+x)]$. If we show that for any given $\delta > 0$ and $\gamma_j$ in the range, there exists some $\delta' > 0$ and $\gamma_i \in (0, 1-1/\eta]$ such that $\gamma_i\eta \geq \gamma_j(\eta + x)$ and $(1-\delta')\ln(1/(1-\gamma_i)) = (1-\delta)\ln(1/(1-\gamma_j))$, then the claim follows. Let $Z = \left(\ln\frac{1}{1-\gamma_j}\right)/\left(\ln\frac{1}{1-\gamma_i}\right)$, then $\delta' = 1-(1-\delta)Z$.

Whenever $\gamma_j \leq 1 - 1/\eta$, we can always choose $\gamma_i \geq \gamma_j$ such that $\delta' > 0$. The non-trivial case is when $\gamma_j \in (1 - 1/\eta, 1 - 1/(\eta + x)]$. In this case, by choosing a large enough $\eta$, we can make $Z$ arbitrarily close to 1 and make $\delta' > 0$. In other words, there exists some $\eta_0$: for all $\eta \geq \eta_0, \delta' > 0$, and by an argument similar to that for claim 2 in Theorem 4, the claim follows. □

### 5.2 A tri-criteria approximation

We now consider upper bounds for MINTIME. It is interesting to ask what happens when either the budget overrun or the coverage shortfall is increased. We show that under these conditions, a greedy strategy combined with linear search yields a solution with optimal propagation time. This proves Theorem 6.

Algorithm Greedy-Mintss computes a small seed set $S$ that achieves coverage $\sigma_m(S) = \eta - \epsilon$. Recall that $\sigma_m^R(S)$ denotes the coverage of $S$ under propagation model $m$ within $R$ time steps. It is easy to see that Greedy-Mintss can be adapted to instead compute a seed set that yields coverage $\eta - \epsilon$ within $R$ time steps: we call this algorithm Greedy-Mintss$^R$.

Given such an algorithm, a simple linear search over $R = 0..n-1$ yields the bounds specified in Theorem 6, after setting coverage threshold as $\eta - \epsilon$ and the chosen budget threshold as $\mathrm{budg} = k(1 + \ln(\eta/\epsilon))$. The approximation factors in the theorem follow from Theorem 1 and Lemma 2. These bounds continue to hold if we can only provide estimates for the coverage function (rather than computing it exactly) and also extend to weighted nodes.

We conclude this section by noting that the algorithm above can be naturally adapted to the RAKC problem. The bounds in Theorem 6 apply to RAKC as well, since MINTIME under IC generalizes RAKC.

## 6 Empirical assessment

We conducted several experiments to assess the value of the approximation algorithms by comparing their quality against that achieved by several well-known heuristics, as well as against the state-of-the-art methods developed for MAXINF that we adapt in order to deal with MINTSS and MINTIME. In particular, the goals of experimental evaluation are twofold. First, we have previously established from theoretical analysis that the Greedy algorithm (Greedy-Mintss for MINTSS and Greedy-Mintss$^R$ for MINTIME) provides the best possible solution that can be obtained in PTIME, which we would like to validate empirically. Second, we study the gap between the solutions obtained from various heuristics against the Greedy algorithm, the upper bound, in terms of quality.

In what follows we assume the IC propagation model.

Datasets, probabilities and methods used. We use two real-world networks, whose statistics are reported in Table 1.

The first network, called NetHEPT, is the same used in Chen et al. (2009, 2010a, b). It is an academic collaboration network extracted from "High Energy Physics—Theory" section of arXiv[5], with nodes representing authors and edges representing coauthorship. This is clearly an undirected graph, but we consider it directed by taking for each edge the arcs in both the directions. Following Kempe et al. (2003), Chen et al. (2009, 2010a), we assign probabilities to the arcs in two different ways: *uniform*, where each arc has probability 0.1 (or probability 0.01) and *weighted cascade* (WC), i.e, the probability of an arc $(v, u)$ is $p_{v,u} = 1/d_{\text{in}}(u)$, where $d_{\text{in}}(\cdot)$ indicates in-degree (Kempe et al. 2003). Note that WC is a special case of IC where probabilities on arcs are not necessarily uniform.

The second one, called Meme, is a sample of the social network underlying the Yahoo! Meme[6] microblogging platform. Nodes are users, and directed arcs from a node $u$ to a node $v$ indicate that $v$ *"follows"* $u$. For this dataset, we also have the log of posts propagations during 2009. We sampled a connected sub-graph of the social network containing the users that participated in the most re-posted items. The availability of posts propagations is significant since it allows us to directly estimate actual influence.

**Table 1** Networks statistics: number of nodes and directed arcs with non-null probability, average degree, number of (strongly) connected components, size of the largest one, and clustering coefficient

| | NetHEPT | Meme |
|---|---|---|
| Number of nodes | 15,233 | 7,418 |
| Number of Arcs | 62,794 | 39,170 |
| Average degree | 4.12 | 5.28 |
| Number of CC (strong) | 1,781 | 4,552 |
| max CC (strong) | 6,794 (44.601 %) | 2,851 (38.434 %) |
| Clustering coefficient | 0.31372 | 0.06763 |

In particular, here a propagation is defined based on reposts: a user posts a meme, and if other users like it, they repost it, thus creating cascades. For each meme $m$ and for each user $u$, we know exactly from which other user she reposted, that is we have a relation *repost*$(u, v, m, t)$ where $t$ is the time at which the repost occurs, and $v$ is the user from which the information flowed to user $u$. The maximum likelihood estimator of the probability of influence corresponding to an arc is $p_{v,u} = M_{v2u} \ / \ M_{vu}$ where $M_{vu}$ denotes the number of memes that $v$ posted before $u$, and $M_{v2u}$ denotes the number of memes $m$ such that *repost* $(u, v, m, t)$.

For the sake of comparison, we adapt the state-of-the-art methods developed for MAXINF (also see Sect. 3) to deal with MINTSS and MINTIME. For most of the techniques the adaptation is straightforward. The methods that we use in the experimentation are succinctly summarized in Table 2. It is noteworthy that PMIA is one of the state-of-the-art heuristic algorithms proposed for MAXINF under the IC model by Chen et al. (2010a). In all our experiments, we run 10,000 Monte Carlo simulations for estimating coverage.

**MINTSS**—Our experimental results on the MINTSS problem are reported in Fig. 1. In each of the three plots, we report, for a given coverage threshold ($x$-axis), the minimum size of a seed-set (budget, reported on $y$-axis) achieving such coverage. As Greedy provides the upper bound on the quality that can be achieved in PTIME, in all the experiments it outperforms the other methods, with Random and HighDegree consistently performing the worst.

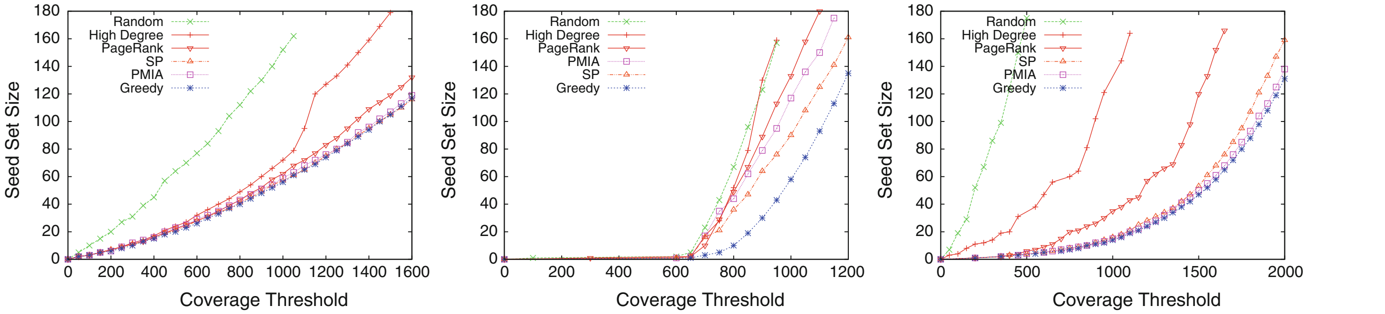
We analyzed the probability distributions of the various data sets we experimented with. At one extreme is the model with uniformly low probabilities (0.01). In Meme, about 80 % of the probabilities are ≤0.05. In NetHEPT WC, on the other hand, approximately 83 % of the

[5] http://www.arXiv.org

[6] http://www.meme.yahoo.com/

**Table 2** The methods used in our experiments

| | |
|---|---|
| Random | Simply add nodes at random to the seed set, until the stopping condition is met |
| HighDegree | Greedily add the highest degree node to the, seed set, until the stopping condition is met |
| PageRank | The popular index of nodes' importance<br>We run it with the same setting used in Chen et al. (2010a) |
| Sp | The shortest-path based heuristic for the greedy algorithm introduced in Kimura and Saito (2006) |
| Pmia | The maximum influence arborescence method of Chen et al. (2010a) with parameter $\theta = 1/320$. |
| Greedy | Algorithm *Greedy-Mintss* for MINTSS and Algorithm *Greedy-Mintss*$^R$ for MINTIME |



**(a)** NetHEPT- WC **(b)** NetHEPT- uniform **(c)** Meme

**Fig. 1** Experimental results on MINTSS

probabilities are ≥0.05 and about 66 % of the probabilities are ≥0.1. However, the combination of a power law distribution of node degrees in NetHEPT together with assignment of low probabilities for high degree nodes (since it's the reciprocal of in-degree) has the effect of rendering central nodes act as poor influence spreaders. And the arcs with high influence probability are precisely those that are incident to nodes with a very low degree. This makes for a low influence graph overall, i.e., propagation of influence is limited. Finally, at the other extreme is the model with uniformly high probabilities (0.1) which corresponds to a high influence graph.

We tested uniformly low probabilities (0.01), and we observed that with such low probabilities, there is limited propagation happening: for instance, in order to achieve a coverage of 150, even the best method requires more than 100 seeds. This forces the quality of all algorithms to look similar.

On data sets where there is a non-uniform mix of low and high probabilities, but the probabilities being predominantly low, as well as on data sets corresponding to low influence graphs, the Pmia method of Chen et al. (2010a, b) and the Sp method of Kimura and Saito (2006), originally developed as efficient heuristics for the MAXI-NF problem, when adapted to the MINTSS problem, continue to provide a good approximation of the results achieved by the Greedy algorithm (Fig. 1a, c). In these situations, the Random and HighDegree heuristics provide seed sets much larger than Greedy. In NetHEPT WC (Fig. 1a), PageRank has a performance that is close to the Greedy solution, while in Meme (Fig. 1c), the seed set generated by PageRank is much larger than Greedy. In data sets with uniformly high probabilities (0.1), the gap between between Greedy and other heuristics is substantial (Fig. 1b). Greedy can achieve a target coverage $\eta = 750$, with just 5 seeds, while Pmia and Sp need 35 and 21 seeds respectively; similarly Greedy can achieve a target coverage $\eta = 1{,}000$, with just 58 seeds, while Pmia and Sp need 117 and 90 seeds respectively. It is worth noting that Random, HighDegree, and the PageRank heuristic all generate seed sets much larger than Greedy on this data set. To sum, the gap between the sizes of the seed sets obtained by the heuristics one the one hand and the Greedy algorithm on the other, varies depending on the influence probabilities on the edges. In general, on graphs with high influence, the gap can be substantial.

**MINTIME**—Our experimental results on the MINTIME problem are reported in Figs. 2 and 3. In Fig. 2, we report, for a coverage threshold given on the $x$-axis, and a fixed budget (75 for NetHEPT, 150 for Meme), the minimum time steps needed to achieve such coverage with the given budget ($y$-axis). As expected, Greedy outperforms all the heuristics. All the plots show that after a certain time, there is no further gain in the coverage, indicating the influence decays over time. Figure 2a compares the various heuristics with the Greedy on the NetHEPT dataset under

WC model. On this data set, Pmia, Sp and Greedy exhibit comparable performance. The PageRank heuristic comes close to them.

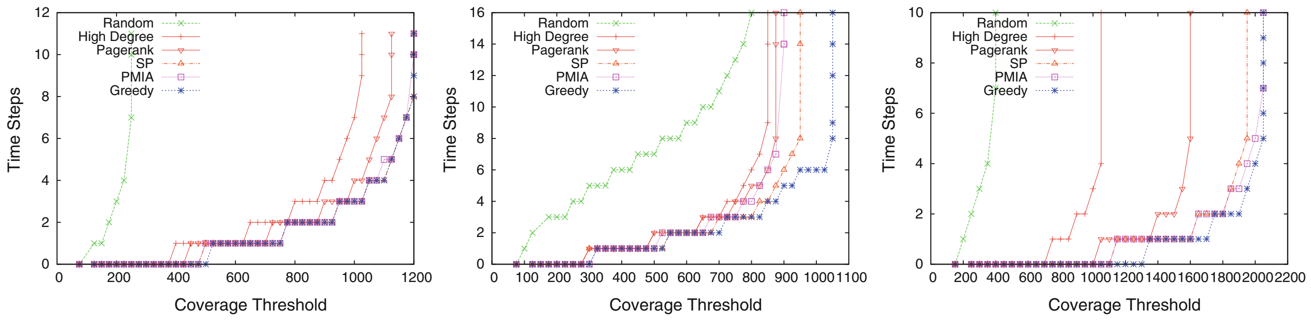
Figure 2b shows the results for the NetHEPT dataset under IC model with uniform probability 0.1. Here, Greedy outperforms all the other heuristics. For instance, when coverage threshold $\eta$ is 900 and budget is 75, Greedy achieves the coverage in 5 time steps, and Sp in 6 time steps, Pmia in 14 time steps. Random, HighDegree and PageRank fail to find a solution. Similarly, when coverage threshold is 1,000 and budget is 75, Greedy achieves the coverage in 6 steps whereas all other heuristics fail to find a solution with this coverage.

Finally, Fig. 2c shows the results on Meme dataset. As we increase the target coverage, the other heuristics fail to give a solution, one by one. Beyond $\eta = 1{,}600$, all but Sp, and Pmia fail and beyond $\eta = 2{,}000$, all but Pmia fail. On this data set, Pmia provides a good approximation to the performance of Greedy.

In Fig. 3, we fix the coverage threshold ($\eta = 1{,}000$ for all the plots). The plots show the minimum time steps needed to achieve the coverage with respect to different seed set sizes (budget). In all the cases, Random fails to find a solution and hence is not shown in the plots. The performance of the HighDegree algorithm is poor as well and it fails to find a solution in case of NetHEPT with uniform probabilities 0.1. As expected, Greedy outperforms all the heuristics and provides us the lower bound on time needed to achieve the required coverage with a given budget.

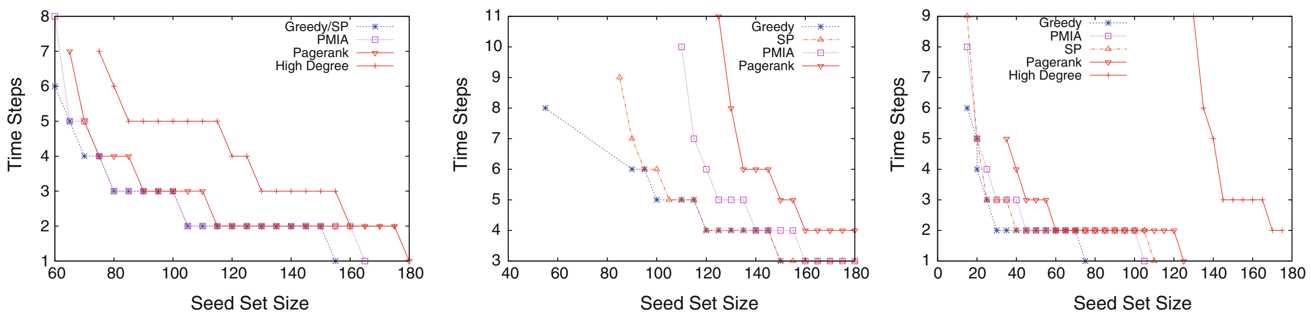
Overall, we notice that the performance quality of all other heuristics compared to Greedy follows a similar pattern to that observed in case of MINTSS: as the graph changes from a low influence graph to a high influence graph, the heuristics' performance drops substantially compared Greedy.

Another key takeaway from the MINTIME plots is the following. *For a given budget, as observed above, the choice of the seed set plays a key role in determining whether a given coverage threshold can be reached or not, no matter how much time we allow for the influence to propagate. Even if the given coverage threshold is achieved, the choice of the seed set can make a big difference to the number of time steps in which the coverage threshold is reached*. Often, for a given budget, relaxing the coverage threshold can dramatically change the propagation time. Example in Fig. 2a (budget fixed to 75), while Greedy takes eight time steps to achieve a coverage



**(a)** NetHEPT- WC, Budget=75 **(b)** NetHEPT- Uniform, Budget=75 **(c)** Meme, Budget=150

**Fig. 2** Experimental results on MINTIME with fixed budget



**(a)** NetHEPT-WC, $\eta$=1000 **(b)** NetHEPT-Uniform, $\eta$=1000 **(c)** Meme, $\eta$=1000

**Fig. 3** Experimental results on MINTIME with fixed coverage threshold

of 1,200, when we relax the threshold to 1,100, the propagation time decreases by 50 %, that is, to just four time steps. A similar phenomenon is observed when the budget is boosted with respect to a fixed coverage threshold. For instance, in Fig. 3c, while using 15 seeds, Greedy takes six time steps to achieve a coverage of 1,000, it achieves the same coverage by 30 seeds in 33 % of the time, that is, in two time steps. These findings further highlight the importance of the MINTIME problem.

## 7 Conclusions

In this paper, we study two optimization problems in social influence propagation: MINTSS and MINTIME. We present a bicriteria approximation for MINTSS which delivers a seed set larger than the optimal seed set by a logarithmic factor $(1 + \ln(\eta/\epsilon))$, that achieves a coverage of $\eta - \epsilon$, which falls short of the coverage threshold by $\epsilon$. We also show a generic tightness result that indicates improving the above approximation factor is likely to be hard.

Turning to MINTIME, we give a greedy algorithm that provides a tricriteria approximation when allowed a budget overrun by a factor of $(1 + \ln(\eta/\epsilon))$ and a coverage shortfall by $\epsilon$, and achieves the optimal propagation time under these conditions. We also provide hardness results for this problem. We conduct experiments on two real-world networks to compare the quality of various popular heuristics proposed in a different context (with necessary adaptations) with that the greedy approximation algorithms. Our results show that the greedy algorithms outperform the other methods in all the settings (as expected), but depending on the characteristics of the data, some of the heuristics perform competitively. These include the recently proposed heuristics Pmia Chen et al. (2010a, b) and Sp Kimura and Saito (2006) which we adapted to MINTSS and MINTIME.

Several questions remain open, including proving optimal approximation bounds for MINTSS and MINTIME, as well as complexity results for these problems under other propagation models.

## Appendix

### A Proof of Lemma 2

Suppose there exists an algorithm $\mathcal{A}$ that selects $\beta$ $k$ sets which covers $\gamma$ $\eta$ elements. Apply $\mathcal{A}$ to an arbitrary instance $\langle \mathcal{U}, \mathcal{S}, \eta \rangle$ of *PSC*. The output is a collection of sets $\mathcal{C}_1$ such that $|\mathcal{C}_1| \leq \beta k$ and $|\cup_{s \in c_1} S| \geq \gamma\eta$ Next, discard the sets that have been selected and the elements they cover, and apply again the algorithm $\mathcal{A}$ on the remaining universe. Repeat this process until 1 or fewer elements are left uncovered.[7]

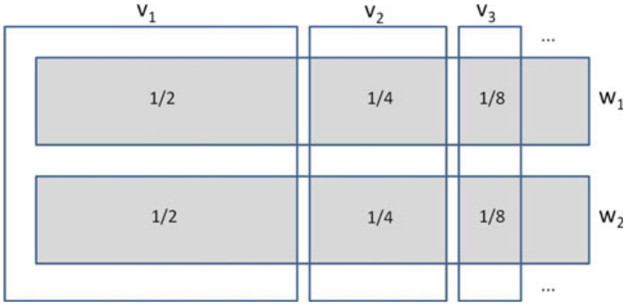


**Fig. 4** Example. *Rectangles* represent the elements in the universe. The *shaded area* within a *rectangle* represents the coverage function $f$ for the element. e.g., $f(v_1) = 1/2 + 1/2 = 1$

Let $\eta_i$ denote the number of elements uncovered after iteration $i$. In iteration $i$, the algorithm picks $\beta$ $k$ sets and covers at least $\gamma$ $\eta_{i-1}$ elements. Hence, $\eta_i \leq \eta_{i-1} \cdot (1 - \gamma)$. Expanding, $\eta_i \leq \eta \cdot (1 - \gamma)^i$. Suppose after $l$ iterations, $\eta_l = 1$. The total number of sets picked is $l\beta k$.$\eta \cdot (1 - \gamma)^l = 1$ implies $l = \frac{\ln \eta}{\ln \frac{1}{1-\gamma}}$.

We now prove the first claim. Let $\gamma > 1 - 1/e^{\beta}$, then $\ln\left(\frac{1}{1-\gamma}\right) > \beta$. This yields a PTIME algorithm for PSC which outputs a solution of size $l\beta k = \beta k \cdot \ln \eta / \ln \frac{1}{1-\gamma} \leq c \cdot k \ln \eta$ (for some $c < 1$) This yields an $c \cdot \ln \eta$-approximation for PSC for some $c < 1$, which is not possible unless NP $\subseteq$ DTIME$(n^{O(\log\log n)})$ (Feige 1998).

To prove the second claim, assume $\beta \leq (1 - \delta) \ln\left(\frac{1}{1-\gamma}\right)$. This gives a PTIME algorithm for *PSC* which outputs a solution of size $l\beta k = \beta k \cdot \ln \eta / \ln \frac{1}{1-\gamma} \leq (1 - \delta) k \cdot \ln \eta$ which is not possible unless NP $\subseteq$ DTIME$(n^{O(\log\log n)})$. □

### B Example illustrating performance of Wolsey's solution

Wolsey (1982) studied the RSSC problem and showed, among many things, that the greedy algorithm provides a solution that is within a factor of $1 + \ln(\eta/(\eta - f(S_{t-1}))$ of the optimal solution. Unfortunately, this does not yield an approximation algorithm with any guaranteed bounds. The following example shows the greedy solution with threshold $\eta$ can be arbitrarily worse than the optimum.

*Example* (Illustrated also in Fig. 4). Consider a ground set $\mathcal{X} = \{w_1, w_2, v_1, v_2, \ldots, v_l\}$ with elements having unit costs. Figure 4 geometrically depicts the definition of a

[7] Instead of 1, we could be left with a constant number of elements. Asymptotically, it does not make a difference.

function $f : 2^{\mathcal{X}} \rightarrow \mathbb{R}$, where for any set $S \subset \mathcal{X}$, $f(S)$ is defined to be the area (shown shaded) covered by the elements of $S$. Specifically, $f(w_1) = f(w_2) = 1 - 1/2^{l+1}$ and $f(v_i) = 1/2^{i-1}$, $1 \leq i \leq l$. Notice, $f(\{v_1, \ldots, v_l\}) = \Sigma_{i=1}^{l} 1/2^{i-1} = 2 - 1/2^{l-1} < 2 - 1/2^{l} = f(\{w_1, w_2\})$. The greedy algorithm will first pick $v_1$. Suppose it picks $S = \{v_1, \ldots, v_i\}$ in $i$ rounds. Then $f(S \cup \{v_{i+1}\}) - f(S) = 1/2^{i} > 1 - 1/2^{l+1} - 1 + 1/2^{i} = 1 - 1/2^{l+1} - 1/2(2 - 1/2^{i-1}) = f(S \cup \{w_1\}) - f(S)$. Thus, greedy will never pick $w_1$ or $w_2$ before it picks $v_1, \ldots, v_l$. Suppose $\eta = 2 - 1/2^{l}$. Clearly, the greedy solution is $\mathcal{X}$ whereas the optimal solution is $\{w_1, w_2\}$. Here $l$ can be arbitrarily large.

## References

Agarwal N, Liu H, Tang L, Yu P (2011) Modeling blogger influence in a community. Social Netw Anal Min 1–24. doi:10.1007/s13278-011-0039-3

Bakshy E, Hofman JM, Mason WA, Watts DJ (2011) Everyone's an influencer: quantifying influence on twitter. In: Proceedings of the fourth ACM international conference on Web search and data mining, ACM, WSDM '11, pp 65–74

Bar-Ilan J, Kortsarz G, Peleg D (2001) Generalized submodular cover problems and applications. Theor Comput Sci 250(1–2): 179–200

Ben-Zwi O, Hermelin D, Lokshtanov D, Newman I (2009) An exact almost optimal algorithm for target set selection in social networks. In: EC '09: Proceedings of the tenth ACM conference on electronic commerce, ACM, New York, NY, USA, pp 355–362

Bhagat S, Goyal A, Lakshmanan LVS (2012) Maximizing product adoption in social networks. In: Web search and data mining, WSDM

Bross J, Richly K, Kohnen M, Meinel C (2011) Identifying the top-dogs of the blogosphere. Social Netw Anal Min 1–15. doi: 10.1007/s13278-011-0027-7

Cha M, Trez JP, Haddadi H (2011) The spread of media content through blogs. Social Netw Anal Min 1–16. doi:10.1007/s13278-011-0040-x

Chen N (2008) On the approximability of influence in social networks. In: SODA '08: Proceedings of the nineteenth annual ACM–SIAM symposium on discrete algorithms, pp 1029–1037

Chen W, Wang Y, Yang S (2009) Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'09)

Chen W, Wang C, Wang Y (2010a) Scalable influence maximization for prevalent viral marketing in large-scale social networks. In: Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'10)

Chen W, Yuan Y, Zhang L (2010b) Scalable influence maximization in social networks under the linear threshold model. In: Proceedings of the 10th IEEE international conference on data mining (ICDM'2010)

Domingos P, Richardson M (2001) Mining the network value of customers. In: Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining, ACM, New York, NY, USA, KDD '01, pp 57–66

Feige U (1998) A threshold of XXX for approximating set cover. J ACM 45(4):634–652

Fujito T (1999) On approximation of the submodular set cover problem. Oper Res Lett 25(4):169–174

Fujito T (2000) Approximation algorithms for submodular set cover with applications. IEICE Trans Inf Syst 83

Goyal A, Bonchi F, Lakshmanan LVS (2008) Discovering leaders from community actions. In: Proceeding of the 17th ACM conference on information and knowledge management, ACM, New York, NY, USA, CIKM '08, pp 499–508

Goyal A, Bonchi F, Lakshmanan LVS (2010) Learning influence probabilities in social networks. In: Proceedings of the third ACM international conference on web search and data mining, ACM, New York, NY, USA, WSDM '10, pp 241–250

Goyal A, Bonchi F, Lakshmanan LVS (2011) A data-based approach to social influence maximization. PVLDB 5(1)

Kempe D, Kleinberg JM, Tardos É (2003) Maximizing the spread of influence through a social network. In: Proceedings of the ninth ACM SIGKDD international conference on knowledge discovery and data mining (KDD'03)

Kempe D, Kleinberg J, Tardos É (2005) Influential nodes in a diffusion model for social networks. In: ICALP, Springer, Berlin, pp 1127–1138

Khuller S, Moss A, Naor JS (1999) The budgeted maximum coverage problem. Inf Process Lett 70(1):39–45

Kimura M, Saito K (2006) Tractable models for information diffusion in social networks. In: Proceedings of PKDD 2006, Lecture notes in computer science, vol 4213

Leskovec J, Krause A, Guestrin C, Faloutsos C, VanBriesen J, Glance NS (2007) Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'07)

Li Gørtz I, Wirth A (2006) Asymmetry in $k$-center variants. Theor Comput Sci 361(2):188–199

Nemhauser GL, Wolsey LA, Fisher ML (1978) An analysis of approximations for maximizing submodular set functions-I. Math Program 14(1):265–294

Panigrahy R, Vishwanathan S (1998) An $O(\log^{*} n)$ approximation algorithm for the asymmetric $p$-center problem. J Algorithms 27(2):259–268

Richardson M, Domingos P (2002) Mining knowledge-sharing sites for viral marketing. In: Proceedings of the eighth ACM SIGKDD international conference on knowledge discovery and data mining, ACM, New York, NY, USA, KDD '02, pp 61–70

Slaví k P (1997) Improved performance of the greedy algorithm for partial cover. Inform Process Lett 64(5):251–254

Sviridenko M (2004) A note on maximizing a submodular set function subject to a knapsack constraint. Oper Res Lett 32(1):41–43

Weng J, Lim EP, Jiang J, He Q (2010) Twitterrank: finding topic-sensitive influential twitterers. In: Proceedings of the third ACM international conference on web search and data mining, ACM, New York, NY, USA, WSDM '10, pp 261–270

Wolsey LA (1982) An analysis of the greedy algorithm for the submodular set covering problem. Combinatorica 2(4):385–393