

Better Approximation Algorithms for Technology Diffusion

Jochen Könemann¹, Sina Sadeghian¹, and Laura Sanità¹

University of Waterloo, Faculty of Mathematics, Department of Combinatorics & Optimization, Waterloo, ON N2L 3G1,
{jochen, s3sadegh, laura.sanita}@uwaterloo.ca

Abstract. Motivated by cascade effects arising in network technology upgrade processes in the Internet, Goldberg and Liu [SODA, 2013] recently introduced the following natural *technology diffusion* problem. Given a graph $G = (V, E)$, and thresholds $\theta(v)$, for all $v \in V$. A vertex u *activates* if it is adjacent to a connected component of active nodes of size at least $\theta(v)$. The goal is to find a seed set \mathcal{A} whose initial activation would trigger a cascade activating the entire graph. Goldberg and Liu presented an algorithm for this problem that returns a seed set of size $O(rl \log(n))$ times that of an optimum seed set, where r is the diameter of the given graph, and l is the number of distinct thresholds used in the instance. We improve upon this result by presenting an $O(\min\{r, l\} \log(n))$ -approximation algorithm. Our algorithm is simple and combinatorial, in contrast with the previous approach that is based on randomized rounding applied to the solution of a linear program.

Keywords: Approximation Algorithms, Technology Diffusion, Combinatorial Optimization

1 Introduction

Networks connecting autonomous entities are pervasive in today’s world, and it is not surprising that their various properties are the subject of a vast and growing body of research (e.g., see the two books [2, 9]). In this paper, we focus on the study of algorithmic aspects of diffusion processes and cascade effects in such networks. How does a virus spread through a population of individuals, and how quickly is a rumor propagated through the members of a group of friends? Modeling dynamic network aspects like this has been an increasingly active sub-area in its own right and we refer the reader to Kleinberg’s survey in [14] for an introduction.

Our work here is specifically motivated by a question addressed by Domingos and Richardson [1, 15] and their work on viral marketing. The authors studied “word-of-mouth” strategies in advertising a new product. The authors posed the following question: given a social network that connects potential customers, can we identify a small *seed* set of influential individuals that, if initially convinced to adopt a product, will eventually persuade all other customers to follow? The authors propose a probabilistic model, and heuristics to address this question.

Kempe et al. [10] later recast the questions of Domingos and Richardson in the language of discrete optimization. The authors single out two diffusion models: the *linear threshold* [7, 16], and *independent cascade* [6] models. In this paper, we will focus on adaptations of the former, where an individual’s inclination to adopt the product is a function of the behaviour of her *immediate* network neighbours. Each edge $uv \in E$ has a weight b_{uv} such that the sum of weights incident to a node v is at most 1. A vertex v adopts the product (*activates*) if the total weight of active neighbours is at least her *threshold* θ_v . The goal in the *influence maximization* problem is to find a minimum-size set of initially active *seed* vertices \mathcal{A} that cause all vertices to eventually adopt the product. In [10], the authors proved that this problem is NP-hard, and presented several approximation algorithms.

The threshold model described above is inherently *local* as a player’s behaviour is only impacted by immediate network neighbours. Goldberg and Liu [5] recently pointed out that locality may not adequately model network externalities [12]; the authors argue that the standard threshold model used by Kempe et al. is a particularly unsuitable model for cascade effects arising in technology upgrade processes in the Internet [4, 8]. The authors propose a generalized threshold model in which a vertex’ utility is influenced by the size of its connected component in the graph of active nodes. We describe this model formally next.

1.1 Non-local threshold model

Consider a network $G = (V, E)$ connecting a population of individuals, each of which has a *threshold* $\theta(v) \in \{\theta_1, \dots, \theta_l\}$. Choose a *seed set* $\mathcal{A}_0 \subseteq V$ of vertices that are initially *active*. Goldberg and Liu describe the following diffusion process: in any step $i \geq 1$, the set of active vertices \mathcal{A}_i consists of all previously active vertices \mathcal{A}_{i-1} and vertices v whose connected component in the graph $G[\mathcal{A}_{i-1} \cup \{v\}]$ induced by \mathcal{A}_{i-1} and v has size at least $\theta(v)$. The smallest t such that $v \in \mathcal{A}_t$ is called the *activation time* of v .

In the *technology diffusion* (TD) problem, we want to find a minimum-cardinality seed set \mathcal{A}_0 such that the above process yields the activation of all vertices in V .

1.2 Goldberg-Liu and our results

Following the notation of [5], we let r be the *diameter* of G ; i.e., if $P(u, v)$ is the smallest number of edges on any u, v -path in G , then we let r be the maximum of $P(u, v)$ over all pairs u, v of vertices. We also use l for the number of thresholds of the given TD instance, and assume that

$$\theta_1 < \theta_2 < \dots < \theta_l.$$

We use \mathcal{A}_0^* for the seed set of an optimum solution. Goldberg and Liu showed that TD is as hard to approximate as *set-cover*, and hence, no $o(\log(n))$ -approximation

may exist unless NP has $n^{O(\log \log(n))}$ -time deterministic algorithms [3]. In the same paper the authors propose an $O(r l \log(n))$ -approximation. Our main result is the following improvement over Goldberg and Liu's work.

Theorem 1. *There is an $O(\min\{r, l\} \log(n))$ -approximation algorithm for TD.*

The algorithm is obtained in two steps. We first describe an $O(r \log(n))$ -approximation by reducing an instance of TD to one of *submodular set-cover* [17]. In order to obtain an $O(l \log(n))$ -approximation, we reduce the problem to an instance of the *quota version of the node-weighted Steiner tree* problem [11, 13]. Both of our algorithms are substantially simpler than those presented in [5], and are deterministic and combinatorial while Goldberg and Liu's methods relied on randomized rounding applied to the solution of a linear program. We complement our algorithmic improvements with the following negative result.

Theorem 2. *TD is as hard to approximate as the quota-version of the unit-weight node-weighted Steiner tree problem.*

The above theorem is as strong as that given in [5] as the quota-version of the unit-weight node-weighted Steiner tree problem generalizes set cover, and at the same time it is based on a simpler reduction. Although the above theorem does not immediately imply any new hardness result for TD, it shows that TD with only two thresholds is already as hard to approximate as the quota-version of the unit-weight node-weighted Steiner tree problem, and this may indicate the need for new ideas to shave off the $\min\{r, l\}$ factor in our result in Theorem 1.

2 An $O(r \log(n))$ approximation algorithm for TD

In this section we develop an $O(r \log n)$ approximation algorithm for TD. In the following, we use \mathcal{A}_0^* for an optimum seed set, and let \mathcal{A}_t^* be the set of vertices that are activated at the end of round t of the dynamics described in Section 1.1.

We begin by obtaining a good lower-bound on the cardinality of \mathcal{A}_0^* . For a threshold θ_i and a vertex v , let $G_v^{\theta_i}$ be the subgraph induced by v and all vertices of G with threshold at most θ_i :

$$G_v^{\theta_i} = G[\{v\} \cup \{u \mid \theta(u) \leq \theta_i\}].$$

Note that $G_v^{\theta_i}$ may in general be disconnected. In the following, let $\Gamma(\theta_i, v)$ be the vertex set of the connected component of $G_v^{\theta_i}$ containing v . Note that, by definition, $v \in \Gamma(\theta_i, v)$. Similarly, for a general set of vertices $S \subseteq V$, we define $\Gamma(\theta_i, S) = \bigcup_{v \in S} \Gamma(\theta_i, v)$. For ease of notation, we let $\theta_0 = 0$, so that $\Gamma(\theta_0, S) = S$ for every $S \subseteq V$.

Lemma 21 *For all $\theta_i \in \{\theta_0, \dots, \theta_{l-1}\}$, $|\Gamma(\theta_i, \mathcal{A}_0^*)| \geq \theta_{i+1} - 1$.*

Proof. Consider a threshold θ_i . If for all vertices $v \notin \mathcal{A}_0^*$ we have $\theta(v) \leq \theta_i$, then by definition $\Gamma(\theta_i, \mathcal{A}_0^*) = V$ and therefore $|\Gamma(\theta_i, \mathcal{A}_0^*)| = |V| \geq \theta_{i+1} - 1$.

Now assume that there is a vertex $v \notin \mathcal{A}_0^*$ with threshold at least θ_{i+1} . In particular, among such vertices, let v be one with smallest activation time t ; i.e., $v \in \mathcal{A}_t^* \setminus \mathcal{A}_{t-1}^*$. Directly from the definition we now see that v 's connected component in $G[\mathcal{A}_{t-1}^* \cup \{v\}]$ has at least $\theta(v) = \theta_{i+1}$ vertices, and among these, only v has threshold larger than θ_i . Thus $\mathcal{A}_{t-1}^* \subseteq \Gamma(\theta_i, \mathcal{A}_0^*)$, and $|\mathcal{A}_{t-1}^*| \geq \theta_{i+1} - 1$. The lemma follows.

By the previous lemma, we can conclude that the size of the minimum cardinality subset of vertices S satisfying $|\Gamma(\theta_i, S)| \geq \theta_{i+1} - 1$ for all $1 \leq i < l$, gives us a lower bound on $|\mathcal{A}_0^*|$.

Corollary 22 *An optimal solution S^* to the following minimum threshold problem*

$$\min_{S \subseteq V} \{|S| : |\Gamma(\theta_i, S)| \geq \theta_{i+1} - 1, \forall 0 \leq i < l\}. \quad (\text{MT})$$

has size at most \mathcal{A}_0^ .*

The above corollary suggests the strategy to follow in designing an approximation algorithm. First, we will search for a vertex set that is a good approximate solution to the minimization problem (MT), and then we will slightly adjust it to turn it into a feasible solution for the technology diffusion problem. As a first step, we present an $O(\log(n))$ -approximation algorithm for (MT) by reducing it to the *submodular set-cover* (SSC) problem. The input to an instance of SSC is a universe U , and a *submodular* function f defined over the subsets of U . Recall that f is submodular, whenever it satisfies

$$f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

for all $A \subseteq B \subseteq U$ and for all $e \in U \setminus B$. Given a cost c_e for all $e \in U$, the goal is now to find a minimum-cost set $T \subseteq U$ such that $f(T) = f(U)$. The problem clearly generalizes the set-cover problem, and admits an $O(\log(\max_{e \in U} f(\{e\})))$ -approximation [17].

Theorem 23 *There is an $O(\log n)$ -approximation algorithm for (MT).*

Proof. As promised, we achieve the result by reducing (MT) to SSC. Let

$$U = V = \{v_1, \dots, v_n\}$$

be the universe of the SSC instance. For each threshold θ_i , $0 \leq i \leq l$, we define function f_i by letting

$$f_i(D) = \min\{|\Gamma(\theta_i, D)|, \theta_{i+1} - 1\},$$

for all $D \subseteq U$. It is an easy exercise to show that f_i is indeed submodular for all i , and hence, so is their sum f , defined by

$$f(D) = \sum_{i=0}^{l-1} f_i(D),$$

ALGORITHM 1: SelectSeedSet1

Input: Graph $G = (V, E)$, thresholds θ .

Output: A seed set \mathcal{A}_0^r .

$\mathcal{A}_0^r = \emptyset$

Compute a solution D to the instance of (MT) according to Theorem 23

While $D \neq \emptyset$ Choose $v \in D$ and add it to \mathcal{A}_0^r

If ($G[S]$ is not connected)

 Compute a shortest v, S -path P in G

 Add the vertices of the path P to \mathcal{A}_0^r

 Remove v from D

for all $D \subseteq U$. Consider the SSC instance with groundset U , submodular function f , and unit cost for each $e \in U$. Suppose that S^* is an optimal solution to (MT) and note that

$$f(S^*) = \sum_{i=0}^{l-1} f_i(S^*) = \sum_{i=0}^{l-1} \theta_{i+1} - 1 = f(U).$$

Thus, an optimal solution D^* of the SSC instance defined above has cardinality at most $|S^*|$. Finally note that

$$f(U) = \sum_{i=0}^{l-1} (\theta_{i+1} - 1) \leq n^2,$$

and the algorithm of [17] therefore returns a set $D \subseteq U$ of size at most $O(\log(n))|S^*|$. By definition, $f(D) = f(U) = \sum_{i=0}^{l-1} \theta_{i+1} - 1$, and therefore $f_i(D) = \theta_{i+1} - 1$, for all $i = 0, \dots, l-1$. But this implies that $|\Gamma(\theta_i, D)| \geq \theta_{i+1} - 1$ for all such i , and thus, D is a feasible solution to the instance of our minimum threshold problem.

We are now ready to give our $O(r \log n)$ -approximation algorithm for the technology diffusion problem.

Theorem 24 *Algorithm 1 is an $O(r \log n)$ -approximation algorithm for TD.*

Proof. Given an instance of the technology diffusion problem, Algorithm 1 first computes a solution D to the instance of the minimum threshold problem defined by G and θ , according to Theorem 23. Note that Theorem 23 and Corollary 22 imply $|D| = O(\log n)|\mathcal{A}_0^*|$. Then, the algorithm constructs the connected seed set \mathcal{A}_0^r from D as follows.

The algorithm adds to \mathcal{A}_0^r , one by one, each vertex $v \in D$: after each addition, if the vertex v is not adjacent to any vertex in the current set \mathcal{A}_0^r , the algorithm also adds to \mathcal{A}_0^r the vertices of a shortest v, S -path P . Note that P has at most r new vertices. Therefore, eventually the set \mathcal{A}_0^r output by the algorithm has size $\leq r|D| \leq O(r \log n)|\mathcal{A}_0^*|$.

Finally, we argue that \mathcal{A}_0^r is a feasible solution to our instance of the technology diffusion problem.

By induction, we prove that if we start the technology diffusion process with the seed set \mathcal{A}_0^r , all of the vertices in $\Gamma(\theta_i, \mathcal{A}_0^r)$ will be activated at some time $t < \infty$, for all $1 < i \leq l$. Since $\Gamma(\theta_l, \mathcal{A}_0^r) = V$, this will prove feasibility for our seed set \mathcal{A}_0^r .

By definition, $\Gamma(\theta_0, \mathcal{A}_0^r) = \mathcal{A}_0^r$ and $G[\mathcal{A}_0^r]$ is connected and contains at least $\theta_1 - 1$ vertices. Assume now that all vertices in $\Gamma(\theta_i, \mathcal{A}_0^r)$ are activated, for some $i \geq 0$. By the definition of our minimum threshold problem, the set D computed by the algorithm in its first step is such that $|\Gamma(\theta_i, D)| \geq \theta_{i+1} - 1$ and since $D \subseteq S$, $|\Gamma(\theta_i, \mathcal{A}_0^r)| \geq \theta_{i+1} - 1$. Therefore, $\Gamma(\theta_i, \mathcal{A}_0^r)$ is a set that induces a connected subgraph of activated vertices of size at least $\theta_{i+1} - 1$. This implies that all the vertices in $\Gamma(\theta_{i+1}, \mathcal{A}_0^r)$ will be activated after some finite time.

3 An $O(l \log(n))$ approximation algorithm

In this section, we give an $O(l \log n)$ -approximation algorithm for the technology diffusion problem. We recall from [5] that we may, w.l.o.g., limit our search to seed sets that *induce connected activation sequences*.

Definition 1. *A seed set \mathcal{A}_0 induces a connected activation sequence with anchor s if there is a permutation v_1, \dots, v_n of vertices so that (i) $s = v_1$, (ii) for all $1 \leq t \leq n$, vertices v_1, \dots, v_{t-1} induce a connected component in G and v_t is adjacent to it, and (iii) v_t is in \mathcal{A}_0 whenever $t < \theta(v_t)$.*

[5] showed that there is a choice of anchor such that the size of a minimum cardinality seed set inducing a connected activation sequence is at most twice the size of an optimal seed set.

Lemma 31 ([5]) *Given an instance of TD, let \mathcal{A}_0^* be the optimal seed set. There is a choice of anchor $s \in \mathcal{A}_0^*$ such that the minimum-cardinality seed set \mathcal{A}^c that induces a connected activation sequence anchored at s has cardinality at most $2|\mathcal{A}_0^*|$.*

Given the above lemma, we will from now on assume that we know anchor vertex s , and will search for a seed set that induces a connected activation sequence anchored at s .

The key insight in our algorithm is its connection with the *quota-constrained node-weighted Steiner Tree* (qNST) problem. In an instance of qNST we are given an undirected graph $G = (V, E)$, a root vertex $s \in V$, vertex weights $w(v)$ for all $v \in V$, and a quota $Q \in \mathbb{Z}_+$. The goal is to find a tree T containing $s \in T$, that spans at least Q vertices, and has smallest weight $w(T) = \sum_{v \in V(T)} w(v)$. This problem is known to have an $O(\log n)$ -approximation algorithm.

Theorem 3 ([11, 13]). *There is a polynomial time $O(\log n)$ approximation algorithm for qNST.*

ALGORITHM 2: SelectSeedSet2

Input: Graph $G = (V, E)$, thresholds θ and a vertex s .

Output: A seed set \mathcal{A}_0^l .

$\mathcal{A}_0^l = \emptyset$

for $i = 1 \dots l$ **do**

For each $v \in V$, set $w_i(v) = 0$ if $\theta(v) < \theta_i$ and $w_i(v) = 1$, otherwise

Find an $O(\log n)$ -approximate tree T_i for the instance $(G, s, w_i, \theta_i - 1)$ of qNST problem

Add to \mathcal{A}_0^l each vertex $v \in T_i$ with $w_i(v) = 1$

end

We now show how to use the above theorem to give an $O(l \log n)$ -approximation for the technology diffusion problem. Algorithm 2 runs in l steps. In each step i , we define a weight function $w_i(v) = 1$ for all $v : \theta(v) \geq \theta_i$, and $w_i(v) = 0$ otherwise, and find an $O(\log n)$ -approximate minimum w_i -weight tree containing s and covering at least $\theta_i - 1$ vertices. This can be done in polynomial time as stated in Theorem 3.

We now argue that (i) the set \mathcal{A}_0^l output by the algorithm is a feasible seed set and (ii) the cardinality of \mathcal{A}_0^l is $O(l \log n)|\mathcal{A}_0^*|$.

Lemma 32 *Algorithm 2 outputs a feasible seed set \mathcal{A}_0^l .*

Proof. By induction on i , we prove that \mathcal{A}_0^l activates all the vertices in T_i at some time $t < \infty$. This implies that \mathcal{A}_0^l activates $\theta_i - 1$ vertices which form a connected component. For $i = l$, this would lead to the result since if there is a connected component C of activated vertices of size at least $\theta_l - 1$ at some time t , clearly all non-activated vertices adjacent to C will become activated at time $t + 1$, and therefore eventually all vertices in V will be activated.

In the first step, the algorithm finds a set T_1 of $\theta_1 - 1$ vertices and adds all of them to the seed set \mathcal{A}_0^l , since $w_1(v) = 1$ for every vertex v . For $i > 1$, assume that all vertices in T_{i-1} are activated at time t . Clearly they form a connected component of size at least $\theta_{i-1} - 1$. Consider the set W_t of non-activated vertices in $T_i \setminus T_{i-1}$ at time t . If $W_t = \emptyset$, we are done. If not, clearly there is a subset $W' \subseteq W_t$ of vertices that are adjacent to the connected component formed by the vertices in T_{i-1} : this is because $T_i \cap T_{i-1} \neq \emptyset$, since both contains the vertex s . Note that a vertex $u \in W'$ cannot have $w_i(u) = 1$, since otherwise the algorithm would have added it to \mathcal{A}_0^l , and therefore u would be activated, a contradiction. It follows that $w_i(u) = 0$, i.e. $\theta(v) \leq \theta_{i-1}$ and therefore it becomes activated at time $t + 1$. If we now consider the set W_{t+1} we have that $W_{t+1} \subset W_t$: so we can repeat this process till eventually all vertices in W_t become activated.

Lemma 33 *In each phase i , Algorithm 2 adds to \mathcal{A}_0^l a number of vertices equals to $w_i(T_i) = O(\log n)|\mathcal{A}_0^*|$.*

Proof. Consider the connected activation process starting with the seed set \mathcal{A}^c . Let t be the first time in which at least $\theta_i - 1$ vertices forming a connected

component are activated. If a vertex v with $\theta(v) \geq \theta_i$ is active at time t , then v must be in the seed set \mathcal{A}^c . Note that \mathcal{A}_t^c contains a tree T_i^* of at least $\theta_i - 1$ vertices and $w_i(T_i^*)$ is exactly the number of vertices in T_i^* with threshold $\geq \theta_i$. Since T_i is an $O(\log n)$ -approximate minimum w_i -weight tree, it follows that $w_i(T_i) \leq O(\log n)w_i(T_i^*) \leq O(\log n)|\mathcal{A}^c|$. The lemma follows from Lemma 31.

Lemma 33, Lemma 32 and Lemma 31 imply:

Theorem 34 *Algorithm 2 is an $O(l \log n)$ approximation algorithm for technology diffusion problem.*

Theorems 24 and 34 together provide a proof of Theorem 1.

4 Complexity

We now provide a proof of Theorem 2, and show that TD with only two thresholds is as hard to approximate as 0, 1-cost qNST. Our reduction is simpler and at the same time as strong as that given in [5] as 0, 1-cost qNST generalizes set cover.

Consider an unrooted instance of qNST on graph $G = (V, E)$ with weights $w(v) \in \{0, 1\}$ for all $v \in V$, and quota Q . We define an instance of TD as follows. For every vertex $v \in V(G)$, let $\theta(v) = Q$ if $w(v) = 1$ and let $\theta(v) = 1$ if $w(v) = 0$.

First assume that T is a solution for the given qNST instance of cost k . Then let \mathcal{A}_0 be the set of k weight 1 vertices of T . Since all other vertices of T have threshold 1, using seed set \mathcal{A}_0 will lead to the activation of all vertices of T . Since T spans at least Q vertices, and all thresholds are at most Q , all other vertices will eventually be activated. Hence, the constructed TD instance has a solution of size at most k .

Now let \mathcal{A}_0 be a valid seed set of size k for the TD instance. If there is no vertex in $V \setminus \mathcal{A}_0$ with threshold Q then all vertices not in \mathcal{A}_0 have weight 0. Thus, any spanning tree T of G has weight at most $|\mathcal{A}_0| = k$.

Now assume that there is a vertex $v \in V \setminus \mathcal{A}_0$ with threshold Q . Pick such a vertex with smallest activation time t . By definition, the connected component T containing v in $G[\mathcal{A}_{t-1} \cup \{v\}]$ has at least Q vertices. Moreover, all but one of the weight-1 vertices in T are also in \mathcal{A}_0 . Thus, the weight of T is at most $k + 1$.

As we discussed in previous section, we can solve the technology diffusion problem using an algorithm for qNST which together with the discussion above shows that the technology diffusion problem with two threshold is essentially equivalent to 0, 1-weight qNST.

5 Further Work

Using the threshold model proposed by Goldberg and Liu, we studied the problem of finding a smallest seed set whose activation triggers a cascade that eventually activates the entire population. We presented improved approximation

algorithms for this problem, and pointed out challenges standing in the way of further improvements.

Many open problems remain. For example, given a target value Q , can we find a small seed set whose activation triggers the activation of at least Q vertices? Or, given a parameter k , we could be looking for a seed set of size at most k such that the largest number of vertices are activated.

Another natural generalization is the weighted version of TD where each vertex $v \in V$ has an associated cost c_v , and the goal is to find a minimum-cost seed set activating all vertices. Unlike the unit-cost version in which the optimal solution inducing connected activation sequence approximates the optimal solution within factor of two, there is an arbitrarily large gap between these two for version of the problem with costs.

This can be easily seen by way of the following example. Consider a star graph with $n+1$ vertices such that the threshold of every vertex is $n+1$. Let the cost of each leaf vertex be $\frac{1}{n}$ while the cost of the vertex at center of the star be an arbitrary large value C . Then, the optimal solution is the set of leaf vertices with total cost of 1 while the cost of any optimal solution inducing a connected activation sequence is at least C as it must contain the vertex at the center. This suggests that one should seek other approaches for solving the version of problem with vertex costs.

Bibliography

- [1] P. Domingos and M. Richardson. Mining the network value of customers. In *In Proceedings, Knowledge discovery and data mining*, pages 57–66, 2001.
- [2] D. A. Easley and J. M. Kleinberg. *Networks, Crowds, and Markets - Reasoning About a Highly Connected World*. Cambridge University Press, 2010.
- [3] U. Feige. A threshold of $\ln n$ for approximating set cover. *J. ACM*, 45, 1998.
- [4] P. Gill, M. Schapira, and S. Goldberg. Let the market drive deployment: a strategy for transitioning to bgp security. *SIGCOMM Comput. Commun. Rev.*, 41(4):14–25, 2011.
- [5] S. Goldberg and Z. Liu. Technology diffusion in communication networks. In *Proceedings, ACM-SIAM Symposium on Discrete Algorithms*, pages 233–240, 2013.
- [6] J. Goldenberg, B. Libai, and E. Muller. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing letters*, 12(3):211–223, 2001.
- [7] M. Granovetter. Threshold models of collective behavior. *American journal of sociology*, pages 1420–1443, 1978.
- [8] R. Gu erin and K. Hosanagar. Fostering ipv6 migration through network quality differentials. *ACM SIGCOMM Computer Communication Review*, 40(3):17–25, 2010.
- [9] M. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.
- [10] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *In Proceedings, Knowledge discovery and data mining, KDD '03*, pages 137–146, 2003.
- [11] J. K onemann, S. Sadeghian, and L. Sanit a. An LMP $O(\log n)$ -approximation algorithm for node weighted prize collecting Steiner tree. Technical Report 1302.2127, arXiv, 2013.
- [12] B. Metcalfe. Metcalfe’s law: A network becomes more valuable as it reaches more users. *InfoWorld*, 1995.
- [13] A. Moss and Y. Rabani. Approximation algorithms for constrained node weighted steiner tree problems. *SIAM J. Comput.*, 37(2):460–481, 2007.
- [14] N. Nisan, T. Roughgarden, E. Tardos, and V. Vazirani. *Algorithmic Game Theory*. Cambridge University Press, 2007.
- [15] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *In Proceedings, Knowledge discovery and data mining*, 2002.
- [16] T. Schelling. *Micromotives and Macrobehavior*. Norton, 1978.
- [17] L.A. Wolsey. An analysis of the greedy algorithm for the submodular set covering problem. *Combinatorica*, 2(4):385–393, 1982.