On the Approximability of Positive Influence Dominating Set in Social Networks

**Thang N. Dinh · Yilin Shen ·
Dung T. Nguyen · My T. Thai**

Abstract In social networks, there is a tendency for connected users to match each other's behaviors. Moreover, a user likely adopts a behavior, if a certain fraction of his family and friends follows that behavior. Identifying people who have the most influential effect to the others is of great advantages, especially in politics, marketing, behavior correction, and so on. Under a graph-theoretical framework, we study the *positive influence dominating set* (PIDS) problem that seeks for a minimal set of nodes $\mathcal{P}$ such that all other nodes in the network have at least a fraction $\rho > 0$ of their neighbors in $\mathcal{P}$. We also study a different formulation, called *total positive influence dominating set* (TPIDS), in which even nodes in $\mathcal{P}$ are required to have a fraction ρ of neighbors inside $\mathcal{P}$. We show that neither of these problems can be approximated within a factor of $(1-\epsilon)\ln\max\{\Delta, |V|^{1/2}\}$, where Δ is the maximum degree. Moreover, we provide a simple proof that both problems can be approximated within a factor $\ln\Delta + O(1)$. In power-law networks, where the degree sequence follows a power-law distribution, both problems admit constant factor approximation algorithms. Finally, we present a linear-time exact algorithms for trees.

Keywords Hardness of Approximation · Approximation Algorithm · Social Networks · Information Diffusion

1 Introduction

Regularly, individuals tend to be influenced by the opinions/behaviors of their family and friends. For examples, children whose parents smoked are twice as likely to begin smoking between 13 and 21 [10], and peer pressure accounts

T. N. Dinh and Y. Shen and D. T. Nguyen · M. T. Thai
Department of Computer & Information Science & Engineering
University of Florida, Gainesville, FL, USA,
E-mail: {tdinh, yshen, dtnguyen, mythai}@cise.ufl.edu

for 65% reasons for binge drinking, a major health issue, by children and adolescents [16]. Moreover, the tendency of a user to adopts a behavior increases together with the number his neighbors follows that behavior.

Thus, exploiting the relationships and influences among individuals in social networks might offer great benefit to both the economy and society. As an example, positive impacts of intervention and education programs on a properly selected set of initial individuals can diffuse widely into society via various social contacts: face-to-face, phone calls, email, social networks and so on.

The *positive influence dominating set* (PIDS) problem emerges in the context of social networks, in which a set of influential users are sought for to propagate the influence to other users. Formally, let $G = (V, E)$ be an undirected graph modeling a social network and denote by $N(v)$ the set of neighbors of a vertex $v \in V$ and $d(v) = |N(v)|$ the degree of v. We study the following problem:

Definition 1 (Positive Influence Dominating Set (PIDS)) Given an undirected graph $G = (V, E)$, a subset $\mathcal{P} \subset V$ is a PIDS of G, if for all $u \in V \setminus \mathcal{P}$, we have $|N(u) \cap \mathcal{P}| \geq \rho d(u)$ for some constant $0 < \rho < 1$. In the PIDS problem, our goal is to find a PIDS of minimum cardinality.

We say nodes in $\mathcal{P}$ dominate their neighbors in $V \setminus \mathcal{P}$. The constant ρ is called the *influence factor*, since it determines for each node the minimum number of neighbors to include in the PIDS. For convenience, the terms vertex/node and edge/link will be used interchangeably in the rest of the paper.

In an another formulation, we require even nodes in the PIDS $\mathcal{P}$ to be dominated by a fraction ρ of their neighbors. The exact definition is as follow.

Definition 2 (Total Positive Influence Dominating Set (TPIDS)) Given an undirected graph $G = (V, E)$, a subset $\mathcal{T} \subset V$ is a TPIDS of G, if for all $v \in V$, we have $|N(u) \cap \mathcal{T}| \geq \rho d(u)$. The TPIDS problem asks to find in G a TPIDS of minimum cardinality.

Both PIDS and TPIDS problems aim to find the minimum set of nodes that can influence/dominate the rest of nodes in the networks. Thus, they are expected to capture the most influential people in networks. Since both problems are NP-hard [18], research efforts have been spent on studying approximability of the problems.

The first approximation algorithm for PIDS is presented in [21], in which the authors prove the $\ln \Delta + O(1)$ approximation factor using submodular theory. Later Cicalese et al., using the same submodularity technique in [21], show the $\ln \Delta + O(1)$ approximation ratio for both PIDS and TPIDS.

The first hardness of approximation result is also given in [21], in which PIDS is shown to be APX-hard. Cicalese et al. [3] give an improved $c \ln n$ inapproximability factors for both PIDS and TPIDS problems. However, it is unclear in the proof [3] that how small the constant c can be [1]. This paper

---

[1] the larger the constant c, the better the inapproximability result

focuses on improving the inapproximability results for PIDS and TPIDS problems. In addition, we provide better approximation factors for special graph classes such as power-law networks, dense networks, and tree.

Our results. We summarize our contributions as follows.

– We prove the inapproximability factor $(1/2 - \epsilon) \ln n$ for both PIDS and TPIDS, assuming that NP $\not\subseteq$ DTIME$\left(n^{O(\log \log n)}\right)$. We achieve the explicit constant $1/2$ in the approximation factor by adjusting the complex construction of the set cover problem [6]. This differs from previous approaches [3, 18, 21] in which hardness results of set cover and dominating set problems are used in a "blackbox fashion".
– We prove that both PIDS and TPIDS are hard to approximate within $\ln \Delta - O(\ln \ln \Delta)$, unless P=NP. In addition, we provide a new proof for the $\ln \Delta + O(1)$ approximation factors of both TPIDS and PIDS.
– In power-law graphs, including many important online social networks, it has been observed empirically that the greedy method that targets the highest degree vertices, performs extremely well. We show that for this class of graphs, the degree-based selection method actually yields a constant factor approximation ratio. In addition, both problems are also well approximated within a constant factor when graphs are dense i.e. $|E| = \Omega(|V|^2)$.
– Finally, we present a linear-time algorithms to find optimal solutions over trees, which has better running time than that of the $O(n^2)$ dynamic programming algorithm in [3].

Note that it is straightforward to extend our results to the case in which each node v requires an arbitrary domination threshold $0 < r_v \leq d(v)$ instead of the $\rho d(v)$ requirement in PIDS and TPIDS.

Related Work. Domingos and Richardson [5] were the first to study the propagation of influence and the problem of identification of the most influential users in networks. Kempe et al. [11, 12] formulated the influence maximization problem as an optimization problem. Leskovec et al. [13] study the influence propagation in a different perspective in which they aim to find a set of nodes in networks to detect the spread of virus as soon as possible.

Influence propagation with a limited number of hops as well as a special case of TPIDS, when $\rho = 1/2$ were first considered in Wang et al. [18] in which they iteratively add (normal) dominating sets until forming a TPIDS. Feng et al. [19] showed NP-completeness for the PIDS problem, when $\rho = 1/2$. The APX-hardness and an $O(\log n)$ approximation algorithm for TPIDS problem were introduced in [21]. Cicalese et al. [3] extends the approximation algorithm for the TPIDS problem and provide $O(\log n)$ inapproximability factors for both problems.

The multiple-hop version of the PIDS problems are studied in [4, 15, 20]. In Unit Disk Graphs, Zhang et al. [20] devised a Polynomial Time Approximation Scheme (PTAS) for the t-latency bounded information propagation when the maximum degree is bounded by a constant. The approximability of the multiple-hop version is studied in [4]. The same paper also provides theoretical analysis for power-law networks and a scalable algorithm to identify the seed

set. An another efficient heuristic to identify the seed set, focusing on the case $\rho = 1/2$, is presented in [15].

2 Approximability in General Networks

Our tight hardness ratios are obtained by altering the Feige's reduction for the inapproximability of the set cover problem. We first give the hardness result for bounded-degree graphs, which leads to the hardness $\ln \Delta - O(\ln \ln \Delta)$. Then, we present the hardness $(1/2 - \epsilon) \ln |V|$. The combination of two hardness results gives us the inapproximability $(1 - \epsilon) \ln \max\{\Delta, |V|^{1/2}\}$. Finally, we present a unified and simple $\ln \Delta + O(1)$ approximation algorithm for both problems. To understand the proofs in this section, we first present briefly the Feige's reduction for the $\ln n$ approximation threshold for the set cover problem [6].

2.1 Feige's Reduction for Set Cover

Feige presented a reduction from a k-prover proof system for a MAX 3SAT-5 instance ϕ that is a *conjunctive normal form* formula consists of n variables and $\frac{5n}{3}$ clauses of exactly 3 literals. The verifier interacts with k provers, and ask provers different questions based on a random string r; each question involves $l/2$ clauses and $l/2$ variables. If the formula ϕ is satisfiable, then the provers have a strategy that cause the verifier accepts for all random strings. If only a $(1 - \epsilon)$ fraction of the clauses in ϕ are simultaneously satisfiable, then for all strategies of the provers, the probability of having two consistent answers is at most $k^2 \cdot 2^{-cl}$, where c is a constant that depends only on ϵ.

The core of the Set cover gadget is a partition system $B(m, L, k, d)$, where B is a ground set of m points. The partition system is a collection of $L = 2^l$ partitions $P_1, \ldots, P_L$ of B, each partition P_i has exactly k disjoint subsets $p_{i,1}, \ldots, p_{i,k}$. Any cover of m points in B requires at least $d = (1 - \frac{2}{2})k \ln m$ subsets, where k and m are selected so that $k < \frac{\ln m}{3 \ln \ln m}$.

Let $R = (5n)^l$ denote the number of possible random strings for the verifier. We make R copies of partition system B. Let B_r denote the copy of the partition associated with the random string r and $p_{i,j}^r$ the copy of set $p_{i,j}$ in B_r.

We are now ready to describe the instance of Set Cover in the Feige's reduction. The universal set $\mathcal{U} = \bigcup_{r \in R} B_r$ contains $N = |\mathcal{U}| = mR$ points; and the set system is $\mathcal{S} = \{S_{q,a,i}\}_{q,a}$, where i can be deduced from syntax of (q, a). Each set $S_{q,a,i}$ corresponds to a question-answer pair (q, a) of the ith prover and $S_{q,a,i} = \bigcup_{(q,i) \in r} p_{a_r,i}^r$ where $(q, i) \in r$ means on random string r, the ith prover receives question q, and a_r is the assignment of variables extracted from a.

As long as $k^2 2^{-cl} < \frac{8}{k^3 \ln^2 m}$, we obtain the hardness result $(1 - \frac{4}{k}) \ln m$ i.e. if formula ϕ is satisfiable, then mR points in $\mathcal{U}$ can be covered by kQ subsets, and if only $(1-\epsilon)$ fraction of the clauses are simultaneously satisfiable, the minimum set cover has size at least $(1 - \frac{4}{k}) \ln m \; kQ$. Here, Q is the set of all $n^l (5/3)^{l/2}$ possible questions. The condition can be satisfied with $l > \frac{1}{c} (5 \log k + 2 \log \ln m)$.

The hardness ratio $(1 - f(k)) \ln m$ of the set cover is obtained from the following key lemma.

Lemma 1 *(Lemma 4.1 [6]) If ϕ is satisfiable, then the above set of $N = mR$ points can be covered by kQ subsets. If only a $(1 - \epsilon)$ fraction of the clauses in ϕ are simultaneously satisfiable, the above set requires $(1 - 2f(k))kQ \ln m$ subsets in order to be covered, where $f(k) \to 0$ as $k \to \infty$.*

Note that $\ln m = (1 - \epsilon) \ln N$ by the setting of n, l, and m in the proof. Thus, the final hardness ratio is $(1-\epsilon) \ln N$, where $N = |\mathcal{U}|$. However, we can choose different settings of n, l, and m and obtain different hardness ratios.

We finish the present of Feige's reduction by giving upper bounds for quantities that appear later in our proofs.

- The *number of subsets* $|\mathcal{S}| \leq |Q|2^{2l}$. Since, for each question $q \in Q$, there are at most 2^{2l} answers of $2l$ bit length.
- The *maximum size* of a subset $\Delta_{\mathcal{S}} = \max_{S \in \mathcal{S}} |S| \leq m3^{l/2}$. Since each i and $q \in Q$ there are at most $3^{l/2}$ random strings r such that the verifier makes query q to the ith prover and $|p_{a_r,i}^r| \leq m$.
- The *maximum frequency* of a point (element) in $\mathcal{U}$: $f \leq k2^l$. Because, for a pair (q, i), each partition $p_{a_r,i}^r$ is included at most 2^l times, plus each point in B_r appears in exactly k partitions.

We continue with hardness results for the PIDS and TPIDS problems in bounded degree graphs.

2.2 Hardness Results on Bounded-Degree Graphs

We prove that neither PIDS nor TPIDS can be approximated within $\ln \Delta - O(\ln \ln \Delta)$ in graphs of maximum degree Δ, unless P=NP. We use a reduction from an instance of the *Bounded Set Cover* problem (SC_B) to an instance of PIDS problem whose degrees are also bounded by $B' = B \text{ poly} \log B$.

Definition 3 (Bounded Set Cover) Given a set system $(\mathcal{U}, \mathcal{S})$, where $\mathcal{U} = \{e_1, e_2, \ldots, e_N\}$ is a universe and $\mathcal{S}$ is a collection of subsets of $\mathcal{U}$. Each subset in $\mathcal{S}$ has at most B elements and each element belongs to at most B subsets, for a predefined constant $B > 0$. A cover is a subfamily $\mathcal{C} \subseteq \mathcal{S}$ of sets whose union is $\mathcal{U}$. Find a cover which uses the minimum number of subsets.

Recall that Δ_S and f stand for the maximum cardinality of sets in S, and the maximum frequency of elements in $\mathcal{U}$, respectively.
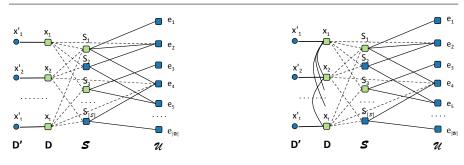
Fig. 1: Reduction from SC_B to PIDS (left) and TPIDS (right)

Lemma 2 *There exist constants $B_0, c_0 > 0$ such that for every $B \geq B_0$ it is NP-hard to approximate the SC_B problem within a factor of $\ln B - c_0 \ln \ln B$.*

Proof For a sufficient large constant $B_0 > 0$, set cover problem where each set has at most $B > B_0$ elements is hard to approximate to within a factor of $\ln B - O(\ln \ln B)$, unless $P = NP$ [17]. The only missing piece is the bound on the frequency of elements in the set cover.

The proof [17] maps an instance of $GAP-SAT_{1,\gamma}$ to an instance $\mathcal{F} = (\mathcal{U}, \mathcal{S})$ of set cover with $\Delta_S \leq B$. Parameters l, m in Feige's construction [6] are fixed to $\theta(\ln \ln B)$ and $\frac{B}{\text{poly} \log(B)}$, respectively. The produced instance has the following properties

- $|\mathcal{U}| = mn^l \text{ poly log} B, |\mathcal{S}| = n^l \text{ poly log B}$
- $\Delta_S \leq B, f \leq \text{ poly log } B$ for sufficient large B.

If we select a sufficient large constant B_0, then we have $f \leq \text{ poly log}(B) \leq B$ for all $B \geq B_0$. $\square$

SC_B-PIDS reduction. For each instance $\mathcal{F} = (\mathcal{U}, \mathcal{S})$ of SC_B, we construct a graph $\mathcal{H} = (V, E)$ as follows (Fig. 1):

- Construct a bipartite graph with the vertex set $\mathcal{U} \cup \mathcal{S}$ and edges between S and all elements $x \in S$, for each $S \in \mathcal{S}$.
- Add a set D consisting of t vertices and a set D' with same number of vertices, say $D = \{x_1, x_2, \ldots, x_t\}$ and $D' = \{x'_1, x'_2, \ldots, x'_t\}$. The value of t will be determined later.
- Connect x_i to $x'_i, \forall i = 1 \ldots t$, to force the selection of x_i in the optimal PIDS.
- Connect each vertex $e_j \in \mathcal{U}$ to $\lceil \frac{\rho}{1-\rho} f(e_j) \rceil - 1$ and each vertex $S_k \in \mathcal{S}$ to $\lceil \frac{\rho}{1-\rho} |S_k| \rceil$ vertices in D, where $f(e_j)$ is the frequency of element e_j. Moreover, we keep the degree differences of vertices in D to be at most one.

Lemma 3 *The size difference between the optimal PIDS of $\mathcal{H}$ and the optimal SC_B of $\mathcal{F}$ is exactly the cardinality of D, i.e., $OPT_{PIDS}(\mathcal{H}) = OPT_{SC}(\mathcal{F}) + t$.*

Proof Let $\mathcal{P}$ be an optimal PIDS of $\mathcal{H}$. Since either x_i or x_i' must be selected into $\mathcal{P}$, and we can always replace $x_i' \in \mathcal{P}$ with x_i inside $\mathcal{P}$. Thus, it is safe to assume that $D' \cap \mathcal{P} = \emptyset$ and $D \subset \mathcal{P}$.

By the construction, each vertex $S_k \in \mathcal{S}$ has enough required neighbors in $\mathcal{P}$, while each vertex $e_i \in \mathcal{U}$ needs at least one more neighbor in $\mathcal{P}$ or it has to be selected. Since all vertices in $\mathcal{U}$ must be adjacent to at least one vertex in $\mathcal{S}$, we can always replace each vertex $e_i \in \mathcal{P}$ with one of its neighbor in $\mathcal{S}$ without increasing the size of $\mathcal{P}$. We therefore can assume that the optimal solution contains vertices in $\mathcal{S}$ but not in $\mathcal{U}$.

Hence, $\mathcal{P} \setminus D$ must induce a cover for $\mathcal{F} = (\mathcal{U}, \mathcal{S})$. In other words, we have $\mathrm{OPT}_{PIDS}(\mathcal{H}) \geq \mathrm{OPT}_{SC}(\mathcal{F}) + t$. Besides, given a cover $\mathcal{C} \subseteq \mathcal{S}$ for $(\mathcal{U}, \mathcal{S})$, it is easy to check that $\mathcal{C} \cup D$ gives a PIDS for $\mathcal{H}$. Thus, $\mathrm{OPT}_{SC}(\mathcal{F}) + t \geq \mathrm{OPT}_{PIDS}(\mathcal{H})$ that completes the proof. □

The key to transfer hardness results of set cover to PIDS problem is to keep the degree of vertices in $\mathcal{H}$ bounded and the gap between the optimal solutions' sizes small. The following lemma states the existence of a construction with such properties.

Lemma 4 *There exists a construction of $\mathcal{H}$ with $t \leq \frac{OPT_{SC}}{\ln^2 B}$ and $B' = \Delta(\mathcal{H}) = O(B \ poly \ log \ B)$.*

Proof We first compute $\mathrm{vol}(D)$, the total degree of vertices in D. For two sets of vertices A and B, we define $\phi(A, B)$ the set of edges crossing between them.

$$\mathrm{vol}(D) = |\phi(D, D')| + |\phi(D, \mathcal{U})| + |\phi(D, \mathcal{S})|$$
$$= |D| + \sum_{S_k \in \mathcal{S}} \lceil \frac{\rho}{1-\rho} |S_k| \rceil + \sum_{e_j \in \mathcal{U}} \lceil \frac{\rho}{1-\rho} f(e_j) - 1 \rceil$$
$$\leq \frac{2\rho}{1-\rho} |\mathcal{S}| B + |\mathcal{S}| + t = \left(\frac{2\rho}{1-\rho} B + 1 \right) |\mathcal{S}| + t \qquad (1)$$

We have used the facts that $\sum_{S_k \in \mathcal{S}} |S_k| = \sum_{e_j \in \mathcal{U}} f(e_j)$ and $|S_k| \leq B, \ \forall S_k \in \mathcal{S}$.

Select $t = \frac{|\mathcal{U}|}{B \ln^2 B}$. Since each set in $\mathcal{S}$ can cover at most B elements, it follows that $\mathrm{OPT}_{SC} \geq \frac{|\mathcal{U}|}{B}$, hence, $\frac{\mathrm{OPT}_{SC}}{\ln^2 B} \geq t$.

To have a valid construction of $\mathcal{H}$, it is sufficient that $t.B' \geq \mathrm{vol}(D)$. Since

$$\frac{\mathrm{vol}(D)}{t} = \frac{1}{t} \left(\left(\frac{2\rho}{1-\rho} B + 1 \right) |\mathcal{S}| + t \right)$$
$$\approx \left(\frac{2\rho}{1-\rho} B + 1 \right) \frac{B \ln^2 B \ n^l \ poly \ log \ B}{mn^l \ poly \ log \ B} \approx B \ poly \ log \ B \qquad (2)$$

Hence, setting $B' = B \ poly \ log \ B$ gives us the desired construction of $\mathcal{H}$. □

Theorem 1 *There exist constants B_1, c_1 such that for every $B' \geq B_1$ it is NP-hard to approximate the PIDS problem in graphs with degrees bounded by B' within a factor of $\ln B' - c_1 \ln \ln B'$.*

Proof We prove by contradiction. Assume we have an algorithm that find a PIDS of size at most $\ln B' - c_1 \ln \ln B'$ the optimal size in graph with degrees bounded by B'. We then show how to approximate the SC_B problem with ratio $\ln B - c_0 \ln \ln B$ in polynomial time. Selecting sufficient large B_1 is not difficult and shall be ignored to make the proof simpler.

Let $\mathcal{F} = (\mathcal{U}, \mathcal{S})$ be an instance of SC_B. Construct an instance $\mathcal{H}$ of PIDS problem using the reduction SC_B-PIDS. From (2), there exists constant $\beta > 0$ so that $B' \leq B \ln^\beta B$. Using the approximation for PIDS, we obtain a solution of size at most $(\ln B' - c_1 \ln \ln B') \mathrm{OPT}_{PIDS}$. We can then convert that to a solution of SC_B by excluding vertices in D (see Lemma 3) and obtain a cover of size at most

$$
\begin{aligned}
&(\ln B' - c_1 \ln \ln B')(\mathrm{OPT}_{SC} + t) - t \\
\leq &(\ln B' - c_1 \ln \ln B')\mathrm{OPT}_{SC} + (\ln B' - c_1 \ln \ln B')\frac{\mathrm{OPT}_{SC}}{\ln^2 B} \\
\leq &\left(\ln B + \beta \ln \ln B - c_1 \ln(\ln B + \theta(\ln \ln B)) + O(\frac{1}{\ln B}) \right) \mathrm{OPT}_{SC}
\end{aligned}
$$

Select $c_1 = c_0 + \beta + 1$. The solution for SC_B problem is then smaller than $\ln B - c_0 \ln \ln B$ times OPT_{SC} which implies P=NP by Lemma 2. $\square$

Theorem 2 *It is NP-hard to approximate TPIDS problem in graphs of bounded degree $B' > B_2$ within a factor of $\ln B' - c_2 \ln \ln B'$ for some constants $B_2, c_2 > 0$.*

Proof We adjust the reduction SC_B-PIDS to achieve the same hardness result. We now need to connect some pairs in D so that they can dominate one another. Fortunately, we can do so with subtly increasing in the degrees of nodes in D and the rest of the proof is essentially the same with that in PIDS.

Specifically, from the gadget obtained in the reduction SC_B-PIDS, we connect a node $x_i \in D$ with $\lceil \frac{\rho}{1-\rho} d(x_i) \rceil$ other nodes in D, balancing nodes' degrees in D. Thus, we roughly multiple the degree of each node in D by a constant. Since $t = \frac{|\mathcal{U}|}{B \ln^2 B} \gg \lceil \frac{\rho}{1-\rho} B' \rceil$, we always have enough vertices in D to connect x_i to. The rest of the proof goes through straightforwardly. $\square$

By a simple argument, we have the following result.

Theorem 3 *Unless P=NP, both PIDS and TPIDS problem cannot be approximated within a factor of $ln\Delta - O(\ln \ln \Delta)$, where Δ is the maximum degree.*

Remark. Instead of using bounded set cover, we can also reduce from the bounded-degree dominating set problem [2] to the PIDS and TPIDS problems. However, it will be difficult to tightly bound the size of D in term of the optimal solutions. As a consequence, the best obtained inapproximability ratios are only $\frac{1}{2} (\ln B - O(\ln \ln B))$.

2.3 Inapproximability in Term of the Network Size

In the same fashion, we can alter the parameter setting in Feige's proof to obtain the following hardness result.

Theorem 4 *PIDS and TPIDS cannot be approximated within $\left(\frac{1}{2} - o(1)\right) \ln |V|$ where n is the number of vertices, unless $\mathrm{NP} \subset \mathrm{DTIME}\left(n^{O(\log \log n)}\right)$.*

Proof We use the same gadget in Fig. 1 to prove the hardness for both PIDS and TPIDS. Since, we no longer need to keep degree of vertices in the gadget bounded, we form a clique with vertices in D.

The sufficient conditions to make the construction feasible are

- (PIDS): We can connect each $v \in (\mathcal{S} \cup \mathcal{U})$ to μ_v vertices in D. That is

$$|D| = O(\max_{v \in (\mathcal{S} \cup \mathcal{U}) \mu_v} \theta(\frac{\rho}{1 - \rho} \Delta_{\mathcal{S}})) = O(\Delta_{\mathcal{S}}) = O(m 3^{l/2})$$

- (TPIDS): Vertices in D have to dominate themselves. Since $\mu_v = \max\{\lceil \frac{\rho}{1-\rho} d(v) \rceil, x_0\}$, this can be satisfied when

$$|D| - 1 = O(\frac{\rho}{1 - \rho} \frac{1}{|D|} \sum_{v \in (\mathcal{S} \cup \mathcal{U})} \mu_v).$$

Or equivalently

$$|D|^2 = O(\sum_{v \in (\mathcal{S} \cup \mathcal{U})} d(v) + x_0(|\mathcal{S}| + |\mathcal{U}|)) = O(2 \sum_{v \in \mathcal{U}} d(v) + |\mathcal{S}| + |\mathcal{U}|) = O(m \, R \, k \, 2^l)$$

To summarize, the sufficient condition for both problems is

$$|D| = O(m 2^{\theta(l)} + (mRk2^l)^{1/2}). \tag{3}$$

By Lemma 1 and the construction, the hardness ratios of our problems are given by

$$\frac{(1 - \frac{4}{k})kQ \ln m + |D|}{kQ + |D|}.$$

Unfortunately, with the same setting in the Feige's reduction, $|D| = O(\Delta_{\mathcal{S}}) = O((5n)^{\frac{2l}{\epsilon}} 2^{\theta(l)})$, the above hardness ratio gets arbitrary close to 1. Hence we use a different setting in which $m = (5n)^{cl}$ with a small constant $c > 0$ to reduce the maximum degree. The consequence is that the inapproximability ratio is reduced accordingly.

The optimal setting to get the best inapproximability ratio is to set $m = (5n)^{l(1-\epsilon)}$ for some $\epsilon > 0$. Then, $N = mR = (5n)^{l(2-\epsilon)}$, or $m = N^{\frac{1-\epsilon}{2-\epsilon}}$. From (3), it is sufficient that

$$|D| = n^l \frac{2^{\theta(l)}}{n^{l\frac{\epsilon}{2}}} = o(Q)$$

Hence, the hardness ratio will be

$$\frac{(1 - \frac{4}{k})kQ \ln m + o(Q)}{kQ + o(Q)} > (1 - \frac{5}{k}) \ln m$$

The number of vertices in the graph, denoted by $n_{\mathcal{H}}$, is

$$n_{\mathcal{H}} = 2|D| + |\mathcal{S}| + |\mathcal{U}| < \theta(m3^{l/2}) + n^l 2^{2l} \left(\frac{5}{3}\right)^{l/2} + (5n)^{2l-\epsilon} < 2|\mathcal{U}| = 2N$$

Finally, the hardness ratio is at least

$$(1 - \frac{5}{k}) \ln \left(\frac{n_{\mathcal{H}}}{2}\right)^{1/2 - \frac{\epsilon}{4-2\epsilon}} > (1 - \frac{5}{k}) \frac{1}{2} \left(1 - \frac{\epsilon}{2-\epsilon}\right) \ln n_{\mathcal{H}} - \theta(1) > \frac{1}{2}(1-\epsilon) \ln n_{\mathcal{H}}.$$

Here, we assume k is sufficiently large and ϵ is sufficiently small. □

2.4 Approximation Algorithm

Approximation algorithms for the PIDS and TPIDS problems have been proposed in [21] by proving the submodularity for the gain functions associated with the selection of nodes into the dominating set. We show that both studied problems can be seen as the instances of the *Constrained Multiset Multicover* (CMM) problem. Thus, both problems inherit the following approximation factors for the CMM problem.

Theorem 5 *Given a graph $G = (V, E)$, there exist $O\left((|V| + |E|) \log \log |V|\right)$ algorithms that approximate PIDS within $H((\rho + 1)\Delta)$ and TPIDS within $H(\Delta)$.*

Proof We first present the Constrained Multiset Multicover problem (CMM).

Definition 4 (Constrained Multiset Multicover) Given a set cover instance $(\mathcal{U}, \mathcal{S})$. Each element e has an integer requirement r_e and occurs in a set S with arbitrary multiplicity, denoted by $m(S, e)$. Moreover, we associate a cost, c_S, with each set $S \in \mathcal{S}$. The Constrained Multiset Multicover problem asks for the minimum cost subcollection which fulfils all elements' cover requirements.

Notice that in CMM, each multiset is picked at most one.

Lemma 5 *[14] There is a natural greedy algorithm that finds a constrained multiset multicover within an H_k factor of the optimal solution, where $k = \max_S \sum_e m(S, e)$.*

The PIDS problem on the graph $G = (V, E)$ can be reduced to the following instance of CMM

– $\mathcal{U} = \{e_u : u \in V\}$
– The cover requirement of e_u is set to $r_u = \lceil \rho d(u) \rceil$

– $\mathcal{S} = \{S_v : v \in V\}$, where S_v contains $\{e_u : u \in N(v)\}$ plus r_v copies of e_v. That is $m(S_v, e_u) = 1, \forall u \in N(v)$ and $m(S_v, e_v) = \lceil \rho d(v) \rceil$.

It follows that the PIDS problem can be approximated within $H_k = H((\rho + 1)\Delta)$.

In case of TPIDS, the only difference in the reduction is that each multiset S_v contains all the neighbors of v, but not a single copy of v. The approximation ratio is, hence, $H(\Delta)$. We note that if we replace $r_u = \lceil \rho d(u) \rceil$ with an arbitrary threshold $0 \le r'_u \le d_u$, we still obtain an $H(\Delta + \max_u r_u) = O(\log \Delta)$ approximation algorithms.   □

Since, $H(n) \approx \ln n + 0.58$ and $1 + \rho < 2$, we can rewrite the approximation ratios for PIDS and TPIDS as $\left(\ln \Delta + \frac{4}{3}\right)$ and $(\ln \Delta + 1)$, respectively.

3 Power-law Networks

Many social, biological, and technology networks including OSNs display a non-trivial topological feature: their degree sequences can be well-approximated by a power-law distribution [8]. Many optimization problems that are hard on general graphs, can be solved much more efficiently in power-law graphs [7,9].

We use the well-known $P(\alpha, \beta)$ model [1] in which there are y vertices of degree x, where x and y satisfy $\log y = \alpha - \beta \log x$. In other words,

$$|\{v : d(v) = x\}| = y = \frac{e^\alpha}{x^\beta}$$

Basically, α is the logarithm of the size of the graph and the constant β is the log-log growth rate of the graph. Without affecting the conclusions, we will simply use real number instead of rounding down to integers. The error terms can be easily bounded and are sufficiently small in our proofs.

The maximum degree in a $P(\alpha, \beta)$ graph is $e^{\frac{\alpha}{\beta}}$. The number of vertices and edges are

$$n = \sum_{x=1}^{e^{\frac{\alpha}{\beta}}} \frac{e^\alpha}{x^\beta} \approx \begin{cases} \zeta(\beta)e^\alpha & \text{if } \beta > 1 \\ \alpha e^\alpha & \text{if } \beta > 1 \\ \frac{e^{\frac{\alpha}{\beta}}}{1-\beta} & \text{if } \beta < 1 \end{cases}, m = \frac{1}{2}\sum_{x=1}^{e^{\frac{\alpha}{\beta}}} x\frac{e^\alpha}{x^\beta} \approx \begin{cases} \frac{1}{2}\zeta(\beta-1)e^\alpha & \text{if } \beta > 2 \\ \frac{1}{4}\alpha e^\alpha & \text{if } \beta = 2 \\ \frac{1}{2}\frac{e^{\frac{2\alpha}{\beta}}}{2-\beta} & \text{if } \beta < 2 \end{cases}$$

where $\zeta(\beta) = \sum_{i=1}^{\infty} \frac{1}{i^\beta}$ is the Riemann Zeta function.

Theorem 6 *In a power-law graph $G \in P(\alpha, \beta)$, the size of the optimal PIDS*

$$OPT_{PIDS} = \begin{cases} \Omega(n^\beta) & \text{if } \beta < 1 \\ \Omega(n/\log n) & \text{if } \beta = 2 \\ \Omega(n) & \text{if } \beta > 2 \end{cases}$$

Proof 1. Let k be the size of the optimal PIDS. Note that all nodes of degree
more than k/ρ must be selected (otherwise the number of selected neighbors
will exceed k). The number of nodes with degree larger than k will be

$$k > \sum_{x=k/\rho}^{e^{\frac{\alpha}{\beta}}} \frac{e^\alpha}{x^\beta} > \sum_{x=k/\rho}^{e^{\frac{\alpha}{\beta}}} \frac{e^\alpha}{x} = e^\alpha(\ln e^{\frac{\alpha}{\beta}} - \ln k) \tag{4}$$

Solving the above relation gives us $k > \Omega(e^\alpha) = \Omega(n^\beta)$.

2. Similarly, when $\beta = 2$, we have $k = \Omega(e^\alpha) = \Omega(n/\log n)$.

3. We use a dual setting approach to obtain the lower bound. Consider the
following linear program and its dual of the PIDS problem

LP: min $\sum\limits_{v \in V} x_v$ **DP**: max $\sum\limits_{u \in V} r_u y_u - \sum\limits_{v \in V} z_v$

s. t. $r_v x_v + \sum\limits_{u \in N(v)} x_u \geq r_v$ s. t. $r_u y_u + \sum\limits_{v \in N(u)} y_v - z_u \leq 1$ (5)

$\quad - x_u \geq -1$ $\quad z_v \geq 0$

$\quad x_u \geq 0$ $\quad y_v \geq 0$

where $r_u = \rho d_u$. We note that for the integral versions of LP(5) and DP(5),
both setting $r_u = \rho d_u$ and $r_u = \lceil \rho d_u \rceil$ yield the same optimal solutions,
however, setting $r_u = \rho d_u$ simplifies the approximation ratio analysis.

Set $y_u = \gamma \; \forall u \in V$. We solve for value of γ to achieve the tightest lower
bound on the size of the optimal PIDS.

To satisfy constraints in the dual, set $z_u = \max\{(\rho + 1)d_u\gamma - 1, 0\}$. Then
the objective value becomes

$$DP = \rho\gamma \sum_{u \in V} d_u - \sum_{u \in V} \max\{(\rho + 1)d_u\gamma - 1, 0\} \tag{6}$$

$$= \rho\gamma \sum_{u \in V} d_u - \sum_{d_u > \tau(\gamma)} ((\rho + 1)d_u\gamma - 1) \tag{7}$$

where $\tau(\gamma)$ denotes $(\rho + 1)^{-1}\gamma^{-1}$.

Substitute $\gamma = \frac{1}{(\rho+1)\tau}$ into (7), we obtain

$$DP = \frac{\rho}{(\rho + 1)\tau} \sum_{x=1}^{e^{\frac{\alpha}{\beta}}} \frac{e^\alpha}{x^\beta} x - \sum_{x > \tau} (\frac{1}{\tau} \frac{e^\alpha}{x^\beta} x - \frac{e^\alpha}{x^\beta}) \tag{8}$$

$$= \frac{\rho}{(\rho + 1)\tau} \zeta(\beta - 1) e^\alpha - \sum_{x > \tau} (\frac{1}{\tau} \frac{e^\alpha}{x^{\beta-1}} - \frac{e^\alpha}{x^\beta}) \tag{9}$$

Except for at most $\lfloor e^{\frac{\alpha}{\beta}} \rfloor$ points $\tau = 1, 2, \dots, \lfloor e^{\frac{\alpha}{\beta}} \rfloor$, the derivatives of the ob-
jective function, $\frac{dDP}{d\tau}$, is defined. Moreover, at those integral points, both one-
sided limits, $\lim\limits_{\tau \to i^-} DP$ and $\lim\limits_{\tau \to i^+} DP$, agree i.e. DP is a continuous function
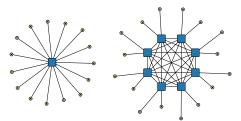everywhere with respect to τ.

Fig. 2: A PIDS(left) may consist of only one node, while a TPIDS(right) must contain at least $O(\sqrt{|V| + |E|})$.

Lemma 6 *For every* $\tau \in (i, i+1), i \in \ltimes^+$, *the derivative* $\frac{dDP}{d\tau}$ *is defined and satisfies*

$$\frac{d}{d\tau}DP = -\frac{1}{\tau^2}\left(\frac{\rho\ \zeta(\beta-1)}{\rho+1}e^\alpha - \sum_{x>\tau}\frac{e^\alpha}{x^{\beta-1}}\right) \tag{10}$$

By (10), there exists a fixed dividing point $x_0 \in \mathbb{N}^+$ that depends only on β, satisfying $\frac{dDP}{d\tau}(\tau) \geq 0,\ \forall \tau < x_0$ and $\frac{dDP}{d\tau}(\tau) < 0,\ \forall \tau > x_0$. Since DP is continuous everywhere, it obtains the global maximum value at $\tau = x_0$.

We show that the value of DP at $\tau = x_0$ is $\Omega(n)$, and since the objective of the primal is lower bounded by DP, it follows that the size of the minimum PIDS will be at least $\Omega(n)$.

$$DP(x_0) = \frac{1}{x_0}\left(\frac{\rho\ \zeta(\beta-1)}{(\rho+1)}e^\alpha - \sum_{x>x_0}\frac{e^\alpha}{x^{\beta-1}}\right) + \sum_{x>x_0}\frac{e^\alpha}{x^\beta}$$

$$\geq \sum_{x>x_0}\frac{e^\alpha}{x^\beta} \approx (\zeta(\beta) - \sum_{x\leq x_0}\frac{1}{x^\beta})e^\alpha \approx (1 - \frac{\sum_{x\leq x_0}\frac{1}{x^\beta}}{\zeta(\beta)})n = \Omega(n) \quad \square$$

Using the same approach in Theorem 6, we have similar bounds for TPIDS.

Theorem 7 *In a power-law graph* $G \in P(\alpha, \beta)$, *the size of the optimal TPIDS,*

$$OPT_{TPIDS} = \begin{cases} \Omega(n) & \text{if } \beta < 1 \text{ or } \beta > 2 \\ \Omega(n/\log n) & \text{if } \beta = 1 \end{cases}$$

If networks have optimal PIDS/TPIDS of $\Omega(n)$ size, clearly, any algorithms that produce valid PIDS/TPIDS will be constant factor approximation algorithms.

4 Dense Graphs

Lemma 7 *If* $\mathcal{T}$ *is a TPIDS of* $G = (V, E)$, *then* $|\mathcal{T}| \geq \Omega(\sqrt{|V| + |E|})$.

Proof Let $k = |\mathcal{T}|$ be the size of an TPIDS. All $v \in V \setminus \mathcal{T}$ must be adjacent to at least one vertex in $\mathcal{T}$. Thus, $|V| \leq |\mathcal{T}| + \bigcup_{v \in \mathcal{T}} |N(v) \setminus \mathcal{T}|$. Moreover, for each vertex $v \in \mathcal{T}, |N(v) \cap \mathcal{T}| \geq \rho|N(v)| \Rightarrow |N(v) \setminus \mathcal{T}| \leq \frac{1-\rho}{\rho}|N(v) \cap \mathcal{T}| \leq \frac{1-\rho}{\rho}(k-1)$. Therefore

$$n \leq k + k\frac{1-\rho}{\rho}(k-1) = \frac{1-\rho}{\rho}k^2 + \frac{2\rho-1}{\rho}k \tag{11}$$

Divide edges in E into three categories: (1) edges whose both ends are in $\mathcal{T}$, (2) edges whose exact one end is in $\mathcal{T}$, (3) edges whose both ends are not in $\mathcal{T}$. We have at most $\binom{k}{2}$ edges of type 1. For a vertex $v \in \mathcal{T}$, the number of type 2 edges incident to v is at most $\frac{1-\rho}{\rho}(k-1)$ since v is adjacent to at most $k-1$ vertices in $\mathcal{T}$. Hence, the number of type 2 edges is upper bounded by $2k\frac{1-\rho}{\rho}(k-1) = \frac{2(1-\rho)}{\rho}\binom{k}{2}$. For each vertex $u \notin \mathcal{T}$, the number of type 3 edges incident to u is at most $\frac{1-\rho}{\rho}$ times the number of type 2 edges incident to u. Therefore, the number of type 3 edges is at most $\frac{(1-\rho)^2}{\rho^2}\binom{k}{2}$.

Adding all three types of edges together, we have

$$|E| \leq \left(1 + \frac{2(1-\rho)}{\rho} + \frac{(1-\rho)^2}{\rho^2}\right)\binom{k}{2} = \frac{1}{\rho^2}\binom{k}{2} \tag{12}$$

It follows from (11) and (12) that $|\mathcal{T}| = k = \Omega(\sqrt{|V| + |E|})$. $\square$

The bound is tight i.e. we can construct a TPIDS of size $\Omega(\sqrt{|V| + |E|})$. For example we construct a 'hairy' clique of $n = k + k \cdot \lfloor\frac{1-\rho}{\rho}(k-1)\rfloor$ vertices and $m = \binom{k}{2} + k \cdot \lfloor\frac{1-\rho}{\rho}(k-1)\rfloor$ edges by connecting each vertex in a clique of size k to $\lfloor\frac{1-\rho}{\rho}(k-1)\rfloor$ leaf nodes (Fig. 2). The minimum TPIDS will be the clique itself that is of size $k = \Omega(\sqrt{n+m})$.

Theorem 8 *For a dense graph $G = (V, E)$ with $|E| = \Omega(|V|^2)$, there exist constant approximation algorithms for both PIDS and TPIDS problems.*

5 Finding Optimal Solutions in Trees

In trees, it is possible to find optimal PIDS and TPIDS in polynomial time. However, designing such algorithms in a linear-time fashion is not too obvious. We present two Depth-first search-based (DFS) algorithms in Algorithms 1 and 2 for PIDS and TPIDS, respectively.

PIDS-TREE(G)
1: $\mathcal{P} = \emptyset$
2: PIDS-VISIT(u), for any $u \in V$
3: **return** $\mathcal{P}$

PIDS-VISIT(u)
1: **for** each unvisited $v \in N(u)$ **do**
2:       PIDS-VISIT(v)
3:       **if** $r_p(v, \mathcal{P}) > 0$ **then**
4:            $\mathcal{P} = \mathcal{P} \cup \{u\}$
5: **if** $r_p(u, \mathcal{P}) > 1$ **then**
6:       $\mathcal{P} = \mathcal{P} \cup \{u\}$

TPIDS-TREE(G)
1: $\mathcal{T} = \emptyset$
2: TPIDS-VISIT(u, u), for any $u \in V$
3: **return** $\mathcal{T}$

TPIDS-VISIT(u, p_u)
1: **for** each unvisited $v \in N(u)$ **do**
2:       TPIDS-VISIT(v, u)
3: **if** $r_t(u, \mathcal{T}) > 0$ **then**
4:       $\mathcal{T} = \mathcal{T} \cup \{p_u\}$
5: Select arbitrary $r_t(u, \mathcal{T})$ unselected
neighbors(children) of u into $\mathcal{T}$.

Algorithm 1: PIDS-TREE(G)   Algorithm 2: TPIDS-TREE(G)

Theorem 9 *Optimal PIDS and TPIDS in trees can be found in linear-time.*

Proof Assume that the given tree is rooted at some vertex u. For an edge (u, v), if u is visited before v, then u is the parent of v and v is a child of u.

At a given step, $\mathcal{P}/\mathcal{T}$ denote the current PIDS/TPIDS. For each $v \in V$, define the functions $r_p(v, \mathcal{P}) = \lceil \rho \cdot d(v) \rceil (1 - \mathbf{1}_{\mathcal{P}}(v)) - |N(v) \cap \mathcal{P}|$ and $r_t(v, \mathcal{T}) = \lceil \rho \cdot d(v) \rceil - |N(v) \cap \mathcal{T}|$, where $\mathbf{1}_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$

Functions $r_p(v, \mathcal{P})$ and $r_t(v, \mathcal{T})$ determine the minimum numbers of v's neighbors to appear in a solution. A node u with $r_p(u, \mathcal{P}) > 0$ or $r_t(u, \mathcal{T}) > 0$ is called *uncovered*, otherwise u is called *covered*.

Correctness. We show by induction that each selection step is optimal.

PIDS: Assume that all the selections made so far are optimal i.e. there exists an optimal solution that contains the selected vertices. In the steps 3 and 4 of Alg. 1, when node u is selected, the following properties follow from the condition $r_p(v, \mathcal{P}) > 0$.

1. $v \notin \mathcal{P}$ (otherwise $r_p(v, \mathcal{P}) \leq 0$) and
2. $r_p(v, \mathcal{P}) = 1$ (otherwise v has already been selected by the end of PIDS-VISIT(v)).

To cover v, we have to either select v, u or some children of v. However, since all nodes in the subtree rooted at v have been covered. There will be no extra benefit in selecting v or its children. Formally, if we have an optimal solution that selects v or its children, we can always replace the selected vertex with u and obtain a new optimal PIDS. In case $r_p(u, \mathcal{P}) > 1$, we are forced to select u.

TPIDS: After TPIDS-VISIT(v, p_v) finishes, v always becomes covered. Assume the selection of vertices into $\mathcal{T}$ is optimal so far. During the visit of a node u, if $r_t(u, \mathcal{T}) > 0$, we select p_u, the parent of u, if $p_u \notin \mathcal{T}$. Since p_u might cover other vertices, while selecting children of u will not affect any uncovered vertices other than u. Finally, we might have select children of u to fully covered u (but only after p_u is selected).

Time complexity. Since the number of edges $|E| = |V| - 1$, the values of $r_p(v, \mathcal{P})$ and $r_t(v, \mathcal{T})$ can be maintained in $O(|V|)$ time. Only when a new vertex is added, we need to update $r_p(.)$ and/or $r_t(.)$ values of that node and all its neighbors. Each node is added at most once, hence the total cost has the same order with the total degree of all vertices i.e. $2|V|$. Hence, the overall time complexities are still $O(|V|)$. □

References

1. Aiello, W., Chung, F., Lu, L.: A random graph model for massive graphs. In: STOC '00, pp. 171–180. ACM, New York, NY, USA (2000). DOI http://doi.acm.org/10.1145/335305.335326

2. Chlebík, M., Chlebíková, J.: Approximation hardness of dominating set problems. In: ESA'04, pp. 192–203 (2004)

3. Cicalese, F., Milanič, M., Vaccaro, U.: Hardness, approximability, and exact algorithms for vector domination and total vector domination in graphs. In: Proceedings of the 18th international conference on Fundamentals of computation theory, FCT'11, pp. 288–297. Springer-Verlag, Berlin, Heidelberg (2011). URL http://dl.acm.org/citation.cfm?id=2034214.2034239

4. Dinh, T.N., Dung, N.T., Thai, M.T.: Cheap, easy, and massively effective viral marketing in social networks: Truth or fiction? In: Proceedings of the 23rd ACM conference on Hypertext and Social Media, HT '12. ACM, Milwaukee, WI, USA (2012)

5. Domingos, P., Richardson, M.: Mining the network value of customers. In: KDD '01, pp. 57–66. ACM, New York, NY, USA (2001). DOI http://doi.acm.org/10.1145/502512.502525

6. Feige, U.: A threshold of ln n for approximating set cover. Journal of ACM **45**(4), 634–652 (1998). DOI http://doi.acm.org/10.1145/285055.285059

7. Ferrante, A., Pandurangan, G., Park, K.: On the hardness of optimization in power-law graphs. Theoretical Computer Science **393**(1-3), 220–230 (2008). DOI http://dx.doi.org/10.1016/j.tcs.2007.12.007

8. Girvan, M., Newman, M.E.: Community structure in social and biological networks. PNAS **99**(12), 7821–7826 (2002). DOI 10.1073/pnas.122653799. URL http://dx.doi.org/10.1073/pnas.122653799

9. Gkantsidis, C., Mihail, M., Saberi, A.: Conductance and congestion in power law graphs. In: SIGMETRICS '03, pp. 148–159. ACM, New York, NY, USA (2003). DOI http://doi.acm.org/10.1145/781027.781046

10. Hill, K.G., Hawkins, J.D., Catalano, R.F., Abbott, R.D., Guo, J.: Family influences on the risk of daily smoking initiation. Journal of Adolescent Health **37**(3), 202 – 210 (2005). URL http://www.sciencedirect.com/science/article/B6T80-4GWPPTF-5/2/3548286c9c2228ce3c7afaa4db17830c

11. Kempe, D., Kleinberg, J., Tardos, É.: Maximizing the spread of influence through a social network. In: KDD'03, pp. 137–146. ACM New York, NY, USA (2003)

12. Kempe, D., Kleinberg, J., Tardos, E.: Influential nodes in a diffusion model for social networks. In: ICALP'05, pp. 1127–1138 (2005)

13. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., VanBriesen, J., Glance, N.: Cost-effective outbreak detection in networks. In: KDD '07, pp. 420–429. ACM, New York, NY, USA (2007). DOI http://doi.acm.org/10.1145/1281192.1281239

14. Rajagopalan, S., Vazirani, V.V.: Primal-dual rnc approximation algorithms for (multi)-set (multi)-cover and covering integer programs. In: STOC '93, pp. 322–331. IEEE Computer Society, Washington, DC, USA (1993). DOI http://dx.doi.org/10.1109/SFCS.1993.366855

15. Shakarian, P., Paulo, D.: Large social networks can be targeted for viral marketing with small seed sets. In: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) (2012)

16. Standridge, J.B., Zylstra, R.G., Adams, S.M.: Alcohol consumption: An overview of benefits and risks. Southern Medical Journal (2004). URL `http://journals.lww.com/smajournalonline/Fulltext/2004/07000/Alcohol_Consumption__An_Overview_of_Benefits_and.12.aspx`

17. Trevisan, L.: Non-approximability results for optimization problems on bounded degree instances. In: ACM Symposium on Theory of Computing '01, pp. 453–461. ACM, New York, NY, USA (2001). DOI http://doi.acm.org/10.1145/380752.380839

18. Wang, F., Camacho, E., Xu, K.: Positive influence dominating set in online social networks. COCOA '09, pp. 313–321. Springer-Verlag, Berlin, Heidelberg (2009). DOI http://dx.doi.org/10.1007/978-3-642-02026-1_29. URL `http://dx.doi.org/10.1007/978-3-642-02026-1_29`

19. Z., F., Z., Z., W., W.: Latency-bounded minimum influential node selection in social networks. In: B. Liu, A. Bestavros, D.Z. Du, J. Wang (eds.) WASA, Lecture Notes in Computer Science, pp. 519–526 (2009). URL `http://dx.doi.org/10.1007/978-3-642-03417-6_51`

20. Zhang, W., Zhang, Z., Wang, W., Zou, F., Lee, W.: Polynomial time approximation scheme for t-latency bounded information propagation problem inwirelessnetworks. Journal of Combinatorial Optimization pp. 1–11 (2010). URL `http://dx.doi.org/10.1007/s10878-010-9359-x`. 10.1007/s10878-010-9359-x

21. Zhu, X., Yu, J., Lee, W., Kim, D., Shan, S., Du, D.Z.: New dominating sets in social networks. Journal of Global Optimization **48**, 633–642 (2010). URL `http://dx.doi.org/10.1007/s10898-009-9511-2`. 10.1007/s10898-009-9511-2