Theoretical Computer Science 550 (2014) 36–50

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On non-progressive spread of influence through social networks

MohammadAmin Fazli [a,*], Mohammad Ghodsi [a,b], Jafar Habibi [a], Pooya Jalaly [c], Vahab Mirrokni [e], Sina Sadeghian [d]

[a] Sharif University of Technology, Tehran, Iran
[b] Institute for Research in Fundamental Sciences (IPM), Niavaran Square, Tehran, Iran
[c] Department of Computer Science, Cornell University, Ithaca, NY 14853, United States
[d] Department of Combinatorics and Optimization, University of Waterloo, 200 University Ave, West Waterloo, ON N2L 3G1, Canada
[e] Google Research, 111 8th Ave, 4th floor, New York, NY 10011, United States

ARTICLE INFO

Article history:
Received 11 December 2012
Received in revised form 15 June 2014
Accepted 8 July 2014
Available online 17 July 2014
Communicated by R. Gavaldà

Keywords:
Spread of influence
Non-progressive
MinPTS problem
NP-hardness
Approximation algorithms

ABSTRACT

The spread of influence in social networks is studied in two main categories: progressive models and non-progressive models (see, e.g., the seminal work of Kempe et al. [8]). While the progressive models are suitable for modeling the spread of influence in monopolistic settings, non-progressive models are more appropriate for non-monopolistic settings, e.g., modeling diffusion of two competing technologies over a social network. Despite the extensive work on progressive models, non-progressive models have not been considered as much. In this paper, we study the spread of influence in the non-progressive model under the strict majority threshold: given a graph G with a set of initially infected nodes, each of which gets infected at time τ iff a majority of its neighbors are infected at time $\tau - 1$. Our goal in the *MinPTS* problem is to find a minimum-cardinality initial set of infected nodes that would eventually converge to the steady state where all nodes of G are infected.

We prove that while the MinPTS problem is NP-complete for a restricted family of graphs, it admits a constant-factor approximation algorithm for power-law graphs. We do so by proving the lower and upper bounds on the optimal solution of the MinPTS problem in terms of the minimum and maximum degrees of nodes in the graph. The upper bound is achieved in turn by applying a natural greedy algorithm. Our experimental evaluation of the greedy algorithm also shows its superior performance compared to other algorithms for a set of real-world graphs as well as the random power-law graphs. Finally, we study the convergence properties of these algorithms and show that the non-progressive model converges in at most $O(|E(G)|)$ steps.

1. Introduction

Studying the spread of social influence in networks under various propagation models is a central issue in social network analysis [1–4]. This issue plays an important role in several real-world applications including the viral marketing [5–8]. As categorized by Kempe et al. [8], there are two main types of influence propagation models: the progressive and the

* Corresponding author. Tel.: +98 21 66166691.
E-mail address: fazli@ce.sharif.edu (M. Fazli).

http://dx.doi.org/10.1016/j.tcs.2014.07.009
0304-3975/

non-progressive models. In the progressive models, infected (or influenced) nodes will remain infected forever, but in the non-progressive models, under some conditions, infected nodes may become uninfected again. In the context of the viral marketing and diffusion of technologies over social networks, the progressive model captures the monopolistic settings where one new service is propagated among nodes of the social network. On the other hand, in the non-monopolistic settings, multiple service providers might be competing to get people adopting their services, and thus users may switch among two or more services back and forth. As a result, in these non-monopolistic settings, the more relevant model to capture the spread of influence is the non-progressive model [9–12].

While the progressive model has been studied extensively in the literature [8,13–18], the non-progressive model has not received much attention. In this paper, we study non-progressive influence models, and report both theoretical and experimental results for them. We focus on the strict majority propagation rule in which the state of each node at time τ is determined by the states of the majority of its neighbors at time $\tau - 1$. As an application of this propagation model, consider two competing technologies (e.g., IM service) that are competing in attracting nodes of a social network to adopt their services, and nodes tend to adopt a service that the majority of their neighbors have already adopted. This type of influence propagation process can be captured by applying the strict majority rule. Moreover, as an illustrative example of the linear threshold model [8], the strict majority propagation model is suitable for modeling the transient faults in fault tolerant systems [19–21], and also used in verifying convergence of consensus problems on social networks [22]. Here, we study the non-progressive influence model under the strict majority rule. In particular, we are mainly interested in the minimum perfect target set problem where the goal is to identify a target set of nodes to infect at the beginning of the process so that all nodes get infected at the end of the process. We will present approximation algorithms and prove hardness results for the problem as well as experimental evaluation to validate our results. As our main contributions, we provide the improved upper and lower bounds on the size of the minimum perfect target set, which in turn, result in a constant-factor approximations for power-law graphs. Finally, we also study the convergence rate of our algorithms and report some preliminary results. Before stating our results, we define the problem and the model formally.

Problem formulations. Consider a graph $G(V,E)$. Let $N(v)$ denote the set of neighbors of node v, and $d(v) = |N(v)|$. Also, let $\Delta(G)$ and $\delta(G)$ denote the maximum and minimum degrees of nodes in G respectively.

A *0/1 initial assignment* is a function $f_0 : V(G) \to \{0,1\}$. For any 0/1 initial assignment f_0, let $f_\tau : V(G) \to \{0,1\}$ ($\tau \geq 1$) be the state of nodes at time τ and $t(v)$ be the threshold associated with node v. For the strict majority model, the threshold $t(v) = \lceil \frac{d(v)+1}{2} \rceil$ for each node v.

In the non-progressive strict majority model:

$$f_\tau(v) = \begin{cases} 0 & \text{if } \sum_{u \in N(v)} f_{\tau-1}(u) < t(v) \\ 1 & \text{if } \sum_{u \in N(v)} f_{\tau-1}(u) \geq t(v). \end{cases} \tag{1}$$

In the progressive strict majority model:

$$f_\tau(v) = \begin{cases} 0 & \text{if } f_{\tau-1}(v) = 0 \text{ and } \sum_{u \in N(v)} f_{\tau-1}(u) < t(v) \\ 1 & \text{if } f_{\tau-1}(v) = 1 \text{ or } \sum_{u \in N(v)} f_{\tau-1}(u) \geq t(v). \end{cases} \tag{2}$$

The strict majority model is related to the linear threshold model in which $t(v)$ is chosen at random and is not necessarily equal to $\lceil \frac{d(v)+1}{2} \rceil$.

A 0/1 initial assignment f_0 is called a perfect target set (PTS) if for a finite τ, $f_\tau(v) = 1$ for all $v \in V(G)$, i.e., the dynamics will converge to the steady state of all 1's. The cost of a target set f_0, denoted by $cost(f_0)$, is the number of nodes v with $f_0(v) = 1$. The minimum perfect target set (MinPTS) problem is to find a perfect target set with the minimum cost. The cost of this minimum PTS is denoted by $PPTS(G)$ and $NPPTS(G)$ respectively for the progressive and non-progressive models. This problem is also called target set selection [23]. Another variant of this problem is the maximum active set problem [23] where the goal is to find at most k nodes to activate (or infect) at time zero such that the number of finally infected nodes is maximized.

A graph is power-law if and only if its degree distribution follows a power-law distribution asymptotically. That is, the fraction $P(x)$ of nodes in the network having the degree x goes for large number of nodes as $E[P(x)] = \alpha x^{-\gamma}$ where α is a constant and $\gamma > 1$ is called power-law coefficient. It is widely observed that most social networks are power-law [24].

Our results and techniques. In this paper, we study the spread of influence in the non-progressive model under the strict majority threshold. We present approximation algorithms and hardness results for the problem as well as experimental evaluation of our results. As our main contribution, we provide improved upper and lower bounds on the size of the minimum perfect target set, which in turn, result in improved constant-factor approximations for power-law graphs. In addition, we prove that the MinPTS problem (or computing $NPPTS(G)$) is NP-hard for a restricted family of graphs. In particular, we prove lower and upper bounds on $NPPTS(G)$ in terms of the minimum degree ($\delta(G)$) and the maximum degree ($\Delta(G)$) of nodes in the graph, i.e., we show that

$$\frac{2n}{\Delta(G)+1} \leq NPPTS(G) \leq \frac{n\Delta(G)(\delta(G)+2)}{4\Delta(G) + (\Delta(G)+1)(\delta(G)-2)}.$$

The proofs of these bounds are combinatorial and start by observing that in order to bound $NPPTS(G)$ for general graphs, one can bound it for bipartite graphs. The upper bound is achieved in turn by applying a natural greedy algorithm which can be easily implemented. Our experimental evaluation of the greedy algorithm also shows its superior performance compared to other algorithms for a set of real-world graphs as well as the random power-law graphs. Finally, we study the convergence properties of this process. We first observe that the process will always converge to a fixed point or a cycle of size two (an easy corollary of a result in [25]). Then we focus on the convergence time and prove that for a given graph G, it takes at most $O(|E(G)|)$ rounds for the process to converge. We also evaluate the convergence rate of the non-progressive influence models on some real-world social networks, and report the average convergence time for a randomly chosen set of initially infected nodes.

More related work. The non-progressive spread of influence under the strict majority rule is related to the diffusion of two or more competing technologies over a social network [9–12]. As an example, an active line of research in economics and mathematical sociology is concerned with modeling these types of diffusion processes as a coordination game played on a social network [9–12]. Note that none of these prior works provides a bound for the perfect target set problem.

In epidemic modeling, the SIS (Susceptible-Infected-Susceptible) model [3] follows the non-progressive paradigm, where infected nodes may recover from a disease but not get lifelong immunity and still be susceptible. This model is different from ours, because firstly, the SIS model studies the propagation process at a different granularity (nodes are not studied individually). Secondly, different mathematical toolsets are used for modeling. Finally to our knowledge, they have not been examined from an optimization perspective in the context of influence maximization.

It has been brought to our attention that in a relevant work by Chang and Wang [26], the MinPTS problem on power-law graphs is studied and the bound of $NPPTS(G) = O(\lceil \frac{|V|}{2^{\gamma-1}} \rceil)$ is proved under non-progressive majority models in a power-law graph. But his results do not practically provide any bound for the strict majority model. We will show that our upper bound is better and practically applicable for different values of γ under the strict majority threshold. Reference [26] also includes bounds for Erdös–Rényi random graphs.

Tight or nearly tight bounds on $PPTS(G)$ are known for general and special types of graphs such as torus, hypercube, butterfly and chordal ring [27,19,28,20,29,30,16,23,26]. The best bounds for progressive strict majority model in general graphs are due to Chang and Lyuu. In [16], they showed that for a directed graph G, $PPTS(G) \le \frac{23}{27}|V(G)|$. In [30], they improved their upper bound to $\frac{2}{3}|V(G)|$ for directed graphs and $\frac{|V(G)|}{2}$ for undirected graphs. In a work of Khoshkhat et al. [31], the upper bound for directed orientation of undirected graphs with no vertex of in-degree zero (such as strongly connected graphs) is improved to $\lfloor \frac{|V(G)|}{2} \rfloor$.

However, to the best of our knowledge, there is no known tight bound for $NPPTS(G)$ for any type of graphs, which was a problem asked by Peleg [32]. Peleg proposed a generalized version of our model which adopts a tie breaking parameter that can have four values: Prefer-White, Prefer-Black, Prefer-Current and Prefer-Flip. If at any round r, node v has exactly half of its neighbors colored white (infected) and half black (uninfected), then at round $r+1$, v's state will be specified by this parameter (thus our model is Prefer-Black, since in our model when there is a tie for a node v, v becomes uninfected in the next round; see Eq. (1)). Peleg asked how small a PTS may be. Three years later, Berger [33] found examples for Prefer-Current case with PTS of size $O(1)$. In this paper, we will combinatorially prove that for each graph G, $\frac{2n}{\Delta(G)+1} \le NPPTS(G)$ and this bound is tight. We will also show that $NPPTS(G) \le \frac{n\Delta(G)(\delta(G)+2)}{4\Delta(G)+(\Delta(G)+1)(\delta(G)-2)}$.

It is known that the target set selection problem and the maximum active set problem are both NP-hard in the linear threshold model [8], and approximation algorithms have been developed for these problems. Kempe et al. [8] and Mossel and Roch [34] present a $(1-\frac{1}{e})$-approximation algorithm for the maximum active set problem by showing that the set of finally influenced nodes as a function of the originally influenced nodes is submodular. On the other hand, it has been shown that the target set selection problem is not approximable for different propagation models [14,15,30,17]. The inapproximability result of Chen in [17] on the progressive strict majority threshold model is the most relevant result to our results. They show that unless $NP \subseteq TIME(n^{polylog(n)})$, no polynomial time $O(2^{\log^{1-\epsilon} n})$-approximation algorithm exists for computing $PPTS(G)$. To the best of our knowledge, no complexity theoretic result has been obtained for the non-progressive models.

The optimization problems related to the non-progressive influence models are not well-studied in the literature. The one result in the area is due to Kempe et al. [8] who presented a general reduction from the non-progressive models to the progressive ones. Their reduction, however, is not applicable to the perfect target set selection problem.

2. Non-progressive spread of influence in general graphs

In this section, we prove a lower bound and an upper bound for the minimum PTS in graphs, and finally show that finding the minimum PTS in general graphs is NP-hard.

The lower bound. Theorem 1 shows that if we have some lower bound and upper bound for the minimum perfect target set in the bipartite graphs then these bounds could be generalized to all graphs.

Theorem 1. *If $\alpha|V(H)| \le NPPTS(H) \le \beta|V(H)|$ for every bipartite graph H under strict majority threshold, then $\alpha|V(G)| \le NPPTS(G) \le \beta|V(G)|$ under strict majority threshold for every graph G.*

Proof. Consider a graph G with n nodes and node set $V(G) = \{v_1, v_2, \dots, v_n\}$. Define t to be the threshold function. Assume that there exists a perfect target set f_0 for G such that $cost(f_0) < \alpha|V(G)|$. Let H be a bipartite graph whose partition has the parts X and Y with $X = \{x_1, \dots, x_n\}$ and $Y = \{y_1, \dots, y_n\}$ and t' be the threshold function of nodes of H such that for every $1 \le i \le n$, $t'(x_i) = t'(y_i) = t(v_i)$. Define $E(H) = \{x_i y_j | v_i v_j \in E(G)\}$. Let g_0 be a target set for H such that $g_0(x_i) = g_0(y_i) = f_0(v_i)$ for every $1 \le i \le n$. We claim that g_0 is a PTS for H. By induction on τ, we prove that $g_\tau(x_i) = g_\tau(y_i) = f_\tau(v_i)$ for every $1 \le i \le n$. By definition, the assertion is true for $\tau = 0$. Now let the assertion be true at time τ. Consider a node $x_i \in X$. We have $\sum_{y \in N(x_i)} g_\tau(y) = \sum_{v \in N(v_i)} f_\tau(v)$ and also $t(x_i) = t(v_i)$, thus x_i is influenced at time $\tau + 1$ by g_0 iff v_i is influenced at time $\tau + 1$ by f_0. By similar justification we can show that $g_{\tau+1}(y_i) = f_{\tau+1}(v_i)$ too. Therefore g_0 is a PTS for H iff f_0 is a PTS for G. This is a contradiction, since $NPPTS(H) \ge \alpha|V(H)|$ and g_0 is a PTS for H with $cost(g_0) < \alpha|V(H)|$.

Now we prove that $NPPTS(G) \le \beta|V(G)|$. Consider the bipartite graph H with the aforementioned definition. By assumption there is a perfect target set g'_0 with weight at most $\beta|V(H)|$ for H. With no loss of generality assume that the number of nodes v in X with $g'_0(v) = 1$ (initially infected nodes) is less than the number of initially infected nodes of Y. Let f'_0 be a PTS for G such that $f'_0(v_i) = g'_0(x_i)$ for every $1 \le i \le n$. We have $cost(g'_0) \le \beta|V(G)|$ since $|V(H)| = 2|V(G)|$. By induction on τ, we show that $f'_{2\tau}(v_i) = g'_{2\tau}(x_i)$ and $f'_{2\tau+1}(v_i) = g'_{2\tau+1}(y_i)$ for every $1 \le i \le n$ and every $\tau \ge 0$. The assertion is trivial for $\tau = 0$. Now let the assertion be true for time 2τ. Consider a node $v_i \in V(G)$. We have $\sum_{v \in N(v_i)} f'_{2\tau}(v) = \sum_{x \in N(y_i)} g'_{2\tau}(x)$ and also $t(v_i) = t(y_i)$, thus v_i is influenced at time $2\tau + 1$ by f'_0 iff y_i is influenced at time $2\tau + 1$ by g'_0. By similar justification we can show that $f'_{2\tau+2}(v_i) = g'_{2\tau+2}(x_i)$ too. Therefore g'_0 is a PTS for H iff f'_0 is a PTS for G, thus $NPPTS(G) \le \beta|V(G)|$. □

Lemma 1 shows characteristics of PTSs in some special cases. This lemma will be used in proof of our theorems.

Lemma 1. *Consider the non-progressive model and let G be a bipartite graph whose partition has the parts X and Y. Assume that f_0 is a perfect target set under threshold function t. For every $S \subseteq V(G)$ if $\sum_{v \in S \cap X} f_0(v) = 0$ or $\sum_{v \in S \cap Y} f_0(v) = 0$, then there exists at least one node u in S such that $d_S(u) \le d(u) - t(u)$, where $d_S(u)$ denotes the number of neighbors of u in S.*

Proof. With no loss of generality, suppose that $f_0(v) = 0$ for every $v \in S \cap X$. We prove the lemma by contradiction. Assume that for every $u \in S$, $d_S(u) > d(u) - t(u)$. For every $y \in S \cap Y$, $f_1(y) = 0$, since y has at least $d(y) - t(y) + 1$ adjacent nodes $r \in S \cap X$ for which $f_0(r) = 0$. Similarly, for every $x \in S \cap X$, $f_2(x) = 0$ since x has at least $d(x) - t(x) + 1$ adjacent nodes in $S \cap Y$ for which f_1 is zero, and so on. Thus f_0 is not a perfect target set. □

If the conditions of the previous lemma hold, we can obtain an upper bound for the number of edges in the graph. Lemma 2 provides this upper bound. This will help finding a lower bound for NPPTS of graphs. The function $t : V(G) \to \mathbb{N}$ may be any arbitrary function but it is interpreted here as the threshold function.

Lemma 2. *Consider a graph G with n nodes and a perfect target set f_0 on G under threshold function t. If for $S \subseteq V(G)$, all nodes of one side are white with respect to f_0, then $|E(G[S])| \le \sum_{u \in S}(d(u) - t(u))$.*

Proof. We prove the lemma by induction on $|S|$. For $|S| = 1$ the assertion is trivial. Assume that it is also true for $|S| < k$ and we want to prove that for $|S| = k$, $|E(G[S])| \le \sum_{u \in S}(d(u) - t(u))$. By Lemma 1 we know that there exists a node $v \in S$ such that $d_S(v) \le d(v) - t(v)$. Assume that $S' = S \setminus \{v\}$ and apply the induction hypothesis on S', so we have

$$\begin{aligned} \left|E\left(G[S]\right)\right| &= d_S(v) + \left|E\left(G\left[S'\right]\right)\right| \\ &\le d(v) - t(v) + \sum_{u \in S'} d(u) - t(u) \\ &\le \sum_{u \in S} d(u) - t(u) \end{aligned}$$

and we are done. □

Theorem 2 shows that for every bipartite graph G, $NPPTS(G) \ge \frac{2|V(G)|}{\Delta(G)+1}$. Theorem 1 generalizes this theorem to all graphs. More than that, Theorem 3 shows that this bound is tight. In the following, the induced subgraph of G with a node set $S \subseteq V(G)$ is denoted by $G[S]$. From now on, we denote $\Delta(G)$ by Δ.

Theorem 2. *For every bipartite graph G of order n whose partition has the parts X and Y, $NPPTS(G) \ge \frac{2n}{\Delta+1}$.*

Proof. Let f_0 be an arbitrary PTS for G. Partition the nodes of graph G into three subsets B_X, B_Y, and W as follows.

$$B_X = \{v \in X \mid f_0(v) = 1\}$$
$$B_Y = \{v \in Y \mid f_0(v) = 1\}$$
$$W = \{v \in V(G) \mid f_0(v) = 0\}$$

Consider the induced subgraph of G with node set $B_X \cup W$ and assume that $S \subseteq B_X \cup W$. For every node $v \in Y \cap S$, we have $f_0(v) = 0$. By Lemma 2, this implies that $G[B_X \cup W]$ has at most $\sum_{u \in B_X \cup W}(d(u) - t(u))$ edges. Similarly, we can prove that $G[B_Y \cup W]$ has at most $\sum_{u \in B_Y \cup W}(d(u) - t(u))$ edges. Let e_W be the number of edges in $G[W]$, e_{WX} be the number of edges with one end in B_X and the other one in W and e_{WY} be the number of edges with one end in B_Y and one end in W. We have

$$e_{WX} + e_W \le \sum_{v \in B_X \cup W} \bigl(d(v) - t(v)\bigr)$$
$$e_{WY} + e_W \le \sum_{v \in B_Y \cup W} \bigl(d(v) - t(v)\bigr)$$

and therefore,

$$e_{WX} + e_{WY} + 2e_W \le \sum_{v \in V(G)} \bigl(d(v) - t(v)\bigr) + \sum_{v \in W} \bigl(d(v) - t(v)\bigr).$$

The total degree of nodes in W is $\sum_{v \in W} d(v) = e_{WX} + e_{WY} + 2e_W$. Thus

$$\sum_{v \in W} d(v) \le \sum_{v \in V(G)} \bigl(d(v) - t(v)\bigr) + \sum_{v \in W} \bigl(d(v) - t(v)\bigr).$$

If we denote the set of nodes v with $f_0(v) = 1$ by B, we have

$$\sum_{v \in W} \bigl(2t(v) - d(v)\bigr) \le \sum_{v \in B} \bigl(d(v) - t(v)\bigr).$$

For every node v, $t(v) \ge \frac{d(v)+1}{2}$, therefore

$$|W| \le \sum_{v \in B} \frac{d(v) - 1}{2}. \tag{3}$$

Therefore,

$$|W| \le \frac{\Delta - 1}{2}\bigl(|B|\bigr).$$

Since $|B| + |W| = n$, we have

$$|B| \ge \frac{2n}{\Delta + 1}. \qquad \square$$

We now show that the bound in Theorem 2 is tight.

Theorem 3. *For infinitely many n and for any $0 < \epsilon < 1$ there exists a graph with n nodes such that NPPTS$(G) < \frac{1}{1-\epsilon}(\frac{2n}{\Delta+1})$ under strict majority rule.*

Proof. Suppose that $k \in \mathbf{N}$ is given. Consider a $(d+1)$-regular graph G_1 with $m_1 = c(d+1)^{(k+1)}$ nodes where $c \in \mathbb{N}$. In step i $(1 \le i \le k)$, add $m_{i+1} = \frac{d}{d+1} m_i$ nodes to the graph and connect each of them to G_i by $d+1$ edges. Each node of G_i must receive exactly d new edges. Name the subgraph formed by these newly added nodes as G_{i+1}. This process is shown in Fig. 1. Repeat k iterations of this process. The final graph has

$$n = \sum_{i=1}^{k+1} m_i = m_1 \left[\frac{1 - (\frac{d}{d+1})^{k+1}}{1 - \frac{d}{d+1}}\right] = m_1 \left[(d+1) - (d+1)\left(\frac{d}{d+1}\right)^{k+1}\right]$$

nodes. By taking a large value k such that $(\frac{d}{d+1})^{k+1} < \epsilon$, we have

$$n > m_1\bigl[(d+1) - (d+1)\epsilon\bigr] = m_1(d+1)(1 - \epsilon).$$

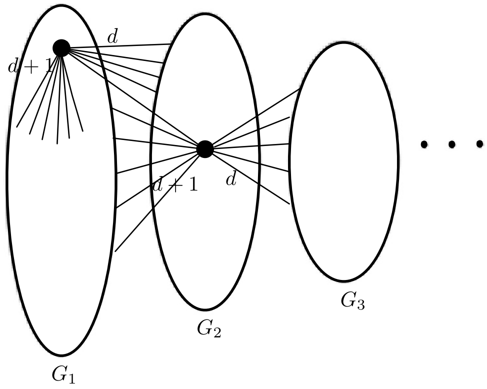

Fig. 1. A tight example for *NPPTS*(G)'s lower bound.

Algorithm 1 Greedy NPPTS.

Input: G: The graph, t: Threshold function
Output: f_0: A PTS

sort the nodes in G in ascending order of their degrees as the sequence $v_1, \ldots, v_n$.
for $i = 1$ to n **do**
  `whiteadj`$[v_i] = 0$
  `blocked`$[v_i] = 0$
end for
for $i = 1$ to n **do**
  for each $u \in N(v_i)$ **do**
    if `whiteadj`$[u] = d(u) - t(u)$ **then**
      `blocked`$[v_i] = 1$
    end if
  end for
  if `blocked`$[v_i] = 1$ **then**
    $f_0(v_i) = 1$
  else
    $f_0(v_i) = 0$
    for each $u \in N(v_i)$ **do**
      `whiteadj`$[u] += 1$
    end for
  end if
end for

$V(G_1)$ is a PTS. So, with $\Delta = 2d + 1$, we have

$$NPPTS(G) \le \left|V(G_1)\right| = m_1 < \frac{2n}{2(d+1)(1-\epsilon)} = \frac{1}{1-\epsilon}\left(\frac{2n}{\Delta+1}\right). \qquad \square$$

The upper bound. In this section, we present a greedy algorithm (Algorithm 1) which gives an upper bound for *NPPTS*(G). Algorithm 1 guarantees this upper bound. This algorithm reads a graph G of order n and the threshold function t as input and determines the value of f_0 for each node.

Lemma 3. *The algorithm Greedy NPPTS finds a perfect target set for the non-progressive spread of influence. More than that, initially infecting the nodes of this set will infect all the nodes of the graph after one step of propagation.*

Proof. By induction on the number of nodes v with determined $f_0(v)$, we prove that f_0 remains a PTS after each step of algorithm if we assume that f_0 is 1 for undetermined values. It is clear that the claim is true at the beginning. Consider a set of values of f_0 which forms a PTS and let v be a node for which the value of $f_0(v)$ is set to 0 by the algorithm in the next step. By induction hypothesis, f_0 is a PTS if $f_0(v)$ is assumed to be 1. According to the algorithm, $f_0(v)$ is set to 0 iff the value of `blocked`$[v]$ is zero, i.e., no adjacent node of v, say u, has exactly $d(u) - t(u)$ adjacent initially uninfected nodes. Therefore by setting $f_0(v)$ to 0, each initially infected node w still has at least $t(w)$ infected nodes and also v has at least $t(v)$ initially infected neighbors itself. Thus, after one step of propagation, all initially infected nodes plus v are infected and by induction hypothesis, all nodes will be infected eventually. Thus f_0 remains a PTS. □

One important fact proved in Lemma 3 is that the initially infected vertices picked by Algorithm 1 actually infect all vertices within one time step, thus the set of picked vertices forms the so-called "strict-majority static monopoly" in some earlier papers (see, e.g., [33]).

Theorem 4. *For every graph G of order n, Greedy NPPTS guarantees the upper bound of $\frac{n\Delta(\delta+2)}{4\Delta+(\Delta+1)(\delta-2)}$ for $NPPTS(G)$ under strict majority threshold where Δ and δ are maximum and minimum degrees of nodes respectively.*

Proof. For $\delta \le 2$, $\frac{n\Delta(\delta+2)}{4\Delta+(\Delta+1)(\delta-2)} \ge n$ and the bound is trivial. So we assume that $\delta > 2$. According to Algorithm 1, for each node v, the value of $f_0(v)$ is set to 1 iff $\texttt{whiteadj}[u] = d(u) - t(u)$ for some $u \in N(v)$. Let S be the set of nodes u with $\texttt{whiteadj}[u] = d(u) - t(u)$. B and W denote the set of infected and uninfected nodes at the end of the algorithm respectively. We have:

$$\sum_{v\in S}\bigl(d(v) - t(v)\bigr) \le \sum_{v\in W} d(v). \tag{4}$$

Therefore,

$$\left(\frac{\delta}{2} - 1\right)|S| \le \Delta|W|,$$

then for $\delta \ge 2$, we have

$$|S| \le \frac{2\Delta}{\delta - 2}|W|.$$

Each node in B has at least one adjacent node in S and each node $v \in S$ has at least $d(v) - t(v)$ adjacent edges to W. So it has at most $t(v)$ adjacent edges to B, thus:

$$|B| \le \sum_{v\in S}\bigl(t(v)\bigr) \le \sum_{v\in S}\left(\frac{d(v)}{2} + 1\right) \le 2|S| + \sum_{v\in W} d(v) \le 2|S| + \Delta|W|.$$

Thus,

$$|B| \le \frac{\Delta(\delta+2)}{4\Delta + (\Delta+1)(\delta-2)}n. \qquad \square$$

The approximation factor of the algorithm follows from previous lemma and the lower bound provided by Theorem 2:

Corollary 1. *The Greedy NPPTS algorithm is a $\frac{\Delta(\Delta+1)(\delta+2)}{8\Delta+2(\Delta+1)(\delta-2)}$ approximation algorithm for NPPTS problem.*

NP-hardness. In this section, we use a reduction from the minimum dominating set problem (MDS) [35] to prove the NP-hardness of computing $NPPTS(G)$.

Theorem 5. *If there exists a polynomial-time algorithm for computing $NPPTS(G)$ for a given graph G under the strict majority threshold, then $P = NP$.*

Proof. In an instance of the minimum dominating set problem (MDS), given a graph $G(V, E)$, our goal is to find a subset $S \subseteq V(G)$ of minimum cardinality such that for any node $v \notin S$, we have $S \cap N(v) \ne \emptyset$. We give a reduction from this NP-hard problem to our problem. Given an instance G of MDS with $V(G) = \{u_1, u_2, \ldots, u_n\}$ and $|E(G)| = e$, we define an undirected graph H as follows (see Fig. 2). First, let

$$X_0 = \{g_1, g_2\}, \qquad X_1 = \{a_i \mid 1 \le i \le 2e+1\},$$
$$X_2 = \{b_i \mid 1 \le i \le 2e+1\}, \qquad X_3 = \{c_i \mid 1 \le i \le 2e\},$$
$$X_4 = \{w_i \mid 1 \le i \le n\}, \qquad X_5 = \{v_i \mid 1 \le i \le n\},$$
$$X_6 = \{d_i \mid 1 \le i \le 2e\}.$$

Now, let $H(V, E)$ be

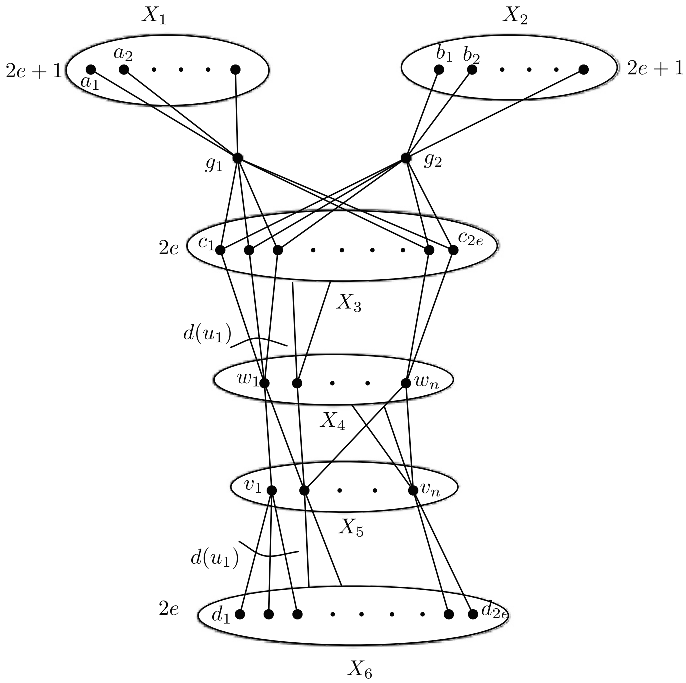

Fig. 2. The graph H.

$$V(H) = \bigcup_{i=0}^{6} X_i,$$

$$\begin{aligned} E(H) = {} & \{g_1 a_i \mid 1 \le i \le 2e+1\} \\ & \cup \{g_2 b_i \mid 1 \le i \le 2e+1\} \\ & \cup \{g_1 c_i \mid 1 \le i \le 2e\} \\ & \cup \{g_2 c_i \mid 1 \le i \le 2e\} \\ & \cup \left\{ w_i c_j \;\middle|\; 1 \le i \le n,\ \sum_{k=1}^{i-1} d(u_k) \le j \le \sum_{k=1}^{i} d(u_k) \right\} \\ & \cup \left\{ v_i w_j \;\middle|\; u_i u_j \in E(G) \vee i = j \right\} \\ & \cup \left\{ v_i d_j \;\middle|\; 1 \le i \le n,\ \sum_{k=1}^{i-1} d(u_k) \le j \le \sum_{k=1}^{i} d(u_k) \right\}. \end{aligned}$$

Suppose that D is a minimum dominating set for G. Define $D^H = \{v_i \mid u_i \in D\}$. We show that $NPPTS(H) = 2e + n + 4 + |D|$. The set $X_0 \cup X_3 \cup X_4 \cup D^H$ plus one node from each of X_1 and X_2 form a perfect target set for the graph H. Therefore, we have

$$NPPTS(H) \le |X_0| + |X_3| + |X_4| + \left|D^H\right| + 2 = 2e + n + 4 + |D|.$$

It remains to prove that $NPPTS(H) \ge 2e + n + 4 + |D|$. Suppose that $S \subseteq V(H)$ is a PTS for H with minimum cardinality. Consider node g_1 in time τ. If $f_\tau(g_1) = 0$, in time $\tau+1$ for every node $a_i \in X_1$ we have $f_{\tau+1}(a_i) = 0$ and then $f_{\tau+2}(g_1) = 0$. Therefore, $g_1 \in S$. Similarly, we have $g_2 \in S$. Moreover, at least $2e+1$ nodes from each of g_1 or g_2's neighbors must be in S, so w.l.o.g suppose that X_3's members plus at least one node from each of X_1 and X_2 are in S. By this setting, the nodes of $X_0 \cup X_1 \cup X_2 \cup X_3$ become infected and remain infected for every $\tau > 0$.

Consider a node $w_k \in X_4$. Let $B(w_k) = \{d_i \in X_6 \mid dis_H(d_i, w_k) = 2\}$, where $dis_H(u, v)$ is the distance between u and v in H. Suppose that $w_k \notin S$. If there exists a $d_i \in B(w_k) \cap S$, we replace it by w_k in S. This modification does not prevent S from being a PTS and also does not increase $|S|$. Thus, we may assume that when $w_k \notin S$, $B(w_k) \cap S = \emptyset$. Assume that $w_k \notin S$ and consider one of w_k's neighbors in X_5 such as v_p. None of v_p's neighbors in X_6 are infected initially. Therefore v_p has at most $d(u_p)$ initially infected neighbors. This implies that $f_1(v_p) = 0$ and it is true for all other w_k's neighbors in X_5. Similarly, $f_2(w_k) = 0$ and $f_2(d_j) = 0$ for all $d_j \in B(w_k)$. Similar to this argument, one can show that for every $\tau > 0$,

$f_{2\tau}(w_k) = 0$ and $f_{2\tau}(d_j) = 0$ for all $d_j \in B(w_k)$. Therefore, after the modification, for each vertex $w_k \in X_4$, we should have $f_0(w_k) = 1$. With almost the same argument we can show that for each vertex $w \in X_4$, at least one of its neighbors in X_5 should be infected as well. This comes from the fact that otherwise, in the next step, neither w nor any of vertices of $B(w)$ will be infected, and the same arguments as before can be applied. So, at least $|D|$ vertices of X_5 should be infected and we have $|S \cap (X_4 \cup X_5)| \geq n + |D|$.

By summing up the above arguments we have

$$\begin{aligned}|S| &= S \cap V(G)\\ &= \left|S \cap \left(\bigcup_{i=0}^{6} X_i\right)\right|\\ &\geq 2 + 1 + 1 + 2e + n + |D|\\ &= 2e + n + 4 + |D|\end{aligned}$$

and so $|S| = 2e + n + 4 + |D|$ and we are done. □

3. Non-progressive spread of influence in connected power-law graphs

In this section, we investigate the non-progressive spread of influence in connected power-law graphs and show that the greedy algorithm presented in the previous section is indeed a constant-factor approximation algorithm for connected power-law graphs. For each natural number x, we assume that the expected number of nodes with degree x is proportional to $x^{-\gamma}$ and use α as the normalization coefficient. The value of γ, known as power-law coefficient, is known to be between 2 and 3 in real-world social networks [24]. We denote the number of nodes of degree x by $P(x)$, so $E[P(x)] = \alpha x^{-\gamma}$ where $E[\cdot]$ denotes the expectation operator.

The lower bound. Consider a connected power-law graph G with a threshold function t and a perfect target set f_0. Denoting the set of initially influenced nodes by B and the rest of the nodes by W, from Eq. (3), we have

$$|W| \leq \sum_{v \in B} \frac{d(v) - 1}{2}.$$

The term $\sum_{v \in B} \frac{d(v)-1}{2}$ is maximized when B consists of higher degree nodes, therefore the maximum cardinality of W is achieved when the degree of all nodes in B is greater than or equal to the degree of all nodes in W. We take the expected value of the number of vertices in W over all connected power-law graphs for which the minimum degree of nodes in B is k and $100p$ percent of nodes of degree k is in B ($0 \leq p \leq 1$). We have

$$\begin{aligned}\sum_{x=1}^{k-1} \alpha x^{-\gamma} + (1-p)\alpha k^{-\gamma} \leq E\big[|W|\big] \leq E\left[\sum_{v \in B} \frac{d(v)-1}{2}\right]\\ \leq \sum_{x=k+1}^{\infty} \alpha x^{-\gamma}\left(\frac{x-1}{2}\right) + p\alpha k^{-\gamma}\frac{k-1}{2}.\end{aligned} \tag{5}$$

Therefore,

$$\sum_{x=1}^{k-1} x^{-\gamma} + (1-p)k^{-\gamma} \leq \frac{\sum_{x=k+1}^{\infty}(x^{1-\gamma} - x^{-\gamma}) + pk^{-\gamma}(k-1)}{2}.$$

Thus,

$$\zeta(\gamma) - \zeta(\gamma, k-1) + (1-p)k^{-\gamma} \leq \frac{\zeta(\gamma-1, k) - \zeta(\gamma, k) + pk^{-\gamma}(k-1)}{2}. \tag{6}$$

From (5), we have the following lower bound for $E[|B|]$:

$$E\big[|B|\big] \geq \sum_{x=k+1}^{\infty} \alpha x^{-\gamma} + \alpha pk^{-\gamma} = \frac{\zeta(\gamma, k) + pk^{-\gamma}}{\zeta(\gamma)} n. \tag{7}$$

For all (k, p) pairs satisfying constraint (6), we calculate the value of the lower bound of $E[|B|]$ and the minimum value is depicted in Fig. 3 as the lower bound of $NPPTS(G)$ for all connected power-law graphs.

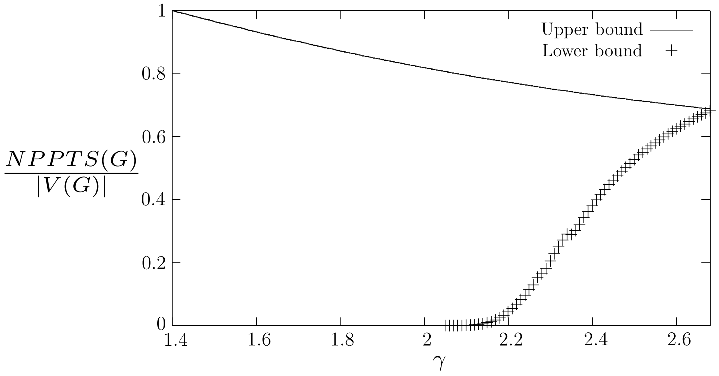

Fig. 3. Values of the upper bound and the lower bound in power-law graphs.

The upper bound. Suppose that one has run Greedy NPPTS algorithm on a connected graph with power-law degree distribution. The following theorem shows that unlike general graphs, the Greedy NPPTS algorithm guarantees a constant factor approximation algorithm on connected power-law graphs.

Theorem 6. *The expected number of initially influenced nodes by Algorithm Greedy NPPTS under the strict majority threshold on connected power-law graphs of order n is at most* $(1 + \frac{1}{2^{\gamma+1}} - \frac{1}{2\zeta(\gamma)})n$.

Proof. We prove that the number of uninfected nodes of degree 1 are sufficient for this upper bound. Let v be a node of degree more than 1 with k adjacent nodes of degree 1 say $u_1, u_2 \dots u_k$. If $d(v)$ is odd, it is clear that at least $\frac{k}{2}$ of the nodes $u_1, u_2 \dots u_k$ will be uninfected since $k \le d(v)$. Note that according to the greedy algorithm, the value of f_0 for degree 1 nodes are determined before any other node. If $d(v)$ is even, at least $\frac{k}{2} - 1$ of nodes $u_1, u_2 \dots u_k$ will be uninfected. Therefore the expected number of nodes infected by Algorithm Greedy NPPTS is less than or equal to:

$$\begin{aligned} & n - \frac{1}{2}\left(P(1) - \sum_{x=1}^{\infty} P(2x)\right) \\ & \quad \le n - \frac{1}{2}\left(\alpha \frac{1}{1^\gamma} - \alpha \sum_{x=1}^{\infty} \frac{1}{(2x)^\gamma}\right) \\ & \quad = n - \frac{\alpha}{2}\left(1 - \frac{1}{2^\gamma}\zeta(\gamma)\right) = n\left(1 + \frac{1}{2^{\gamma+1}} - \frac{1}{2\zeta(\gamma)}\right). \qquad \square \end{aligned}$$

By previous theorem, we conclude that the Greedy NPPTS algorithm is a constant-factor approximation algorithm on connected power-law graphs under strict majority threshold. The lower bound and the upper bound for different values of γ are shown in Fig. 3. As you can see our algorithm acts optimally on social networks with large value of power-law coefficient (values greater than 2.68) since upper and lower bound diagrams meet each other for these values of the power-law coefficient. In Section 5, we will compare the optimality of our algorithm with typical heuristics by running the algorithm on real network data.

4. Convergence issues

Let the state graph H of a non-progressive spread of influence process for graph G be as follows: Each node of this graph represents one of possible states of the graph. An edge between two states A and B in H models the fact that applying one step of the influence process on state A changes the state to state B. First of all, one can easily see that there exists a dynamics for which the non-progressive model may not result in a singleton steady state. To see this, consider the following example: a cycle with $2k$ nodes $C = v_1 v_2 \dots v_{2k}$ and at time 0 infect nodes with odd indices. In this case, the process will oscillate between exactly two states. In fact, one can show a general theorem that any dynamics will converge to either one or two states:

Theorem 7. *The non-progressive spread of influence process on a graph reaches a cycle of length of at most two.*

Proof. In [25], it is shown that, for a function Δ from $\{0,1\}^n$ to $\{0,1\}^n$ whose components form a symmetric set of threshold functions, the repeated application of Δ, leads either to a fixed point or to a cycle of length two. More formally for any $y \in \mathbb{R}$ there exists $s \in \mathbb{N}$ such that $\Delta^{s+2} y = \Delta^s y$. Since the set of functions f_τ (defined in Section 1) are symmetric threshold functions, the lemma follows immediately from this fact. $\square$

Using this intuition, one can define the convergence time of a non-progressive influence process under the strict majority rule as the time it takes to converge to a cycle of size of at most two states, i.e., the convergence time is the minimum time T at which $f_T(v) = f_{T+2}(v)$ for all nodes $v \in V(G)$. For a set S of initially infected nodes, let $ct_G(S)$ be the convergence time of the non-progressive process under the strict majority model (T). In the following theorem, we formally prove an upper bound of $O(|E(G)|)$ for this convergence time:

Theorem 8. *For a given graph G and any set $S \subseteq V(G)$, we have $ct_G(S) = O(|E(G)|)$.*

Proof. In each time step τ of the non-progressive spread of influence, all the nodes apply the function f_τ concurrently. In order to prove the theorem for such concurrent dynamics, we first define a simplified sequential dynamics, prove the convergence time for this simplified dynamics, and finally give a reduction from the concurrent to the sequential dynamics. In the sequential dynamics, the nodes apply the influence process one by one in a sequence of rounds, where in each step one node applies the influence process exactly once.

We first show that the sequential dynamics on every graph G and under the strict majority model converges after at most $O(|E(G)||V(G)|)$ steps. To see this bound, consider the following potential function for a graph G: the number of edges whose endpoints have different states (Like it is used in [19]). One can see that whenever a node changes its state from uninfected to infected the potential of G will decrease at least by one and otherwise it remains unchanged. Consider a node which has k state changes during the process until its final convergence. At least $k/2$ of these changes were from uninfected state to infected and so they cause one decrement in the potential function. The initial amount of G's potential is at most $|E(G)|$ and in each step (or $|V(G)|$ consecutive steps), we have at least one state change. So after at most $2|E(G)||V(G)|$ steps the potential of G would reach its minimum, and the proof for the sequential dynamics is complete.

Now using the above observation, we show that the concurrent dynamics convergences fast. Consider graph $H = (X, Y)$ built from G in Lemma 1. We show that for every concurrent dynamics in G with convergence time of T, there is an equivalent sequential dynamics in H with convergence time of $c|V(G)|T$ for some constant c. This will prove $ct_G \in O(|E(G)|)$, since we know that the convergence time of the sequential dynamics in graph H is at most $2|V(H)||E(H)| = 8|V(G)||E(G)| = cT|V(G)|$. Therefore $T \in O(|E(G)|)$. The main claim follows from the proof of Lemma 1. By induction on the number of steps, we can show that the state of nodes in G is equal to the state of nodes in X at odd steps and is equal to the state of nodes in Y at even steps (as we did in the proof of Lemma 1). Now order nodes of X and Y with numbers $1, 2, \cdots, |V(G)|$ and from $|V(G)| + 1$ to $2|V(G)|$. It is easy to see that the sequential dynamics with this ordering, after $|V(H)|$ steps, has the same outcome under the concurrent dynamics in this graph. □

The above theorem is tight, i.e., there exists a set of graphs and initial states with convergence time of $\Omega(|E(G)|)$. In power-law graphs since average degree is constant, the number of edges is $O(|V|)$ and thus the convergence time of these graphs is $O(|V|)$. From Theorem 8, we can also conclude that checking whether a set S is PTS or not is verifiable is polynomial time. So one can extend Theorem 5 by showing that the MinPTS problem is NP-complete.

In Section 5, we study convergence time of non-progressive dynamics on several real-world graphs, and observe the fast convergence of such dynamics on those graphs.

5. Experimental evaluations

In this section, we run our algorithm on real-world social networks as well as random power-law graphs with a wide range of power-law coefficients. Following the method used in [8], we compare the performance of our algorithm to other heuristics for identifying influential individuals.

5.1. Greedy NPPTS

We evaluate the performance of the Greedy NPPTS algorithm (Algorithm 1) on graphs with various values of power-law coefficients. Following a previously developed way of generating power-law graphs from [36], we set two parameters α and γ defined as follows: α is the logarithm of the graph size and γ is the log–log growth rate (power-law coefficient).

We also run these algorithms over four social networks' data: Who-trusts-whom network of Epinions.com, Slashdot social network, collaboration network of Arxiv Astro Physics, Arxiv High Energy Physics paper citation network, Amazon product co-purchasing network. In cases where graph is not connected we select the graphs' giant component.

Generating random power-law networks. We evaluate the performance of the greedy algorithm on graphs with various amount of power-law coefficient. Following a previously developed way of generating the power-law graphs from [36], we set two parameters α and γ defined as follows: α is the logarithm of the graph size and γ is the log–log growth rate (power-law coefficient). Let y be the number of nodes with degree x. y and x must satisfy

$$\log y = \alpha - \gamma \log x.$$

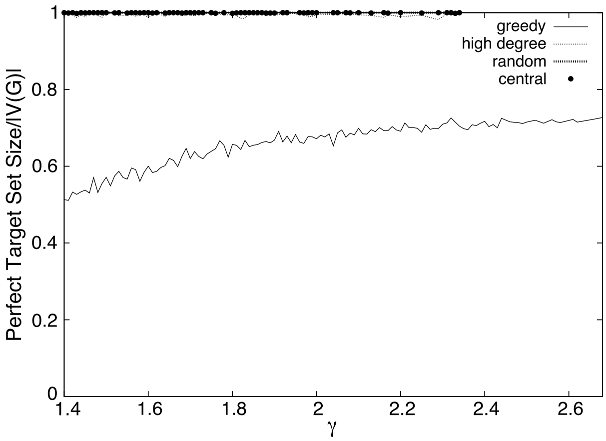

Fig. 4. Results on the random power-law graphs.

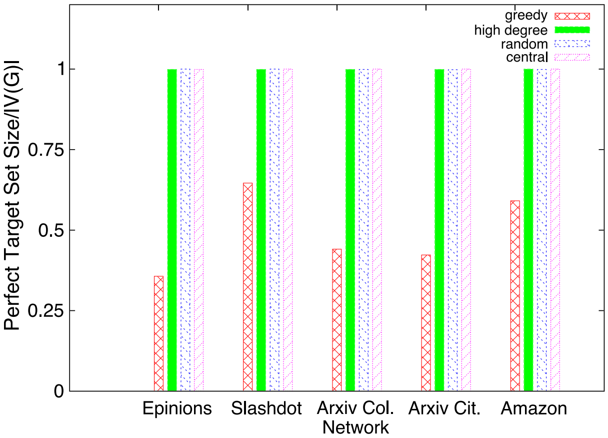

Fig. 5. Results on the real network data.

The random power-law graph model is defined as follows: given n weighted nodes with weights $w_1, w_2, \cdots, w_n$, a pair (i, j) of nodes appears as an edge with probability $w_i w_j p$ independently. These parameters p and $w_1, w_2, \cdots, w_n$ must satisfy

- $|\{i|w_i = 1\}| = \lfloor e^\alpha \rfloor - r$ and $|\{i|w_i = k\}| = \lfloor \frac{e^\alpha}{k^\gamma} \rfloor$ for $k = 2, 3, \ldots, \lfloor e^{\frac{\alpha}{\gamma}} \rfloor$. Here α is a value minimizing $|n - \sum_{k=1}^{\lfloor e^{\frac{\alpha}{\gamma}} \rfloor} \lfloor \frac{e^\alpha}{k^\gamma} \rfloor|$ and $r = n - \sum_{k=1}^{\lfloor e^{\frac{\alpha}{\gamma}} \rfloor} \lfloor \frac{e^\alpha}{k^\gamma} \rfloor$.
- $p = \frac{1}{\sum_{i=1}^{n} w_i}$.

One can easily see the expected degree of ith node would be w_i and also nodes' weights follow power-law.

Setup. We compare our greedy algorithm with heuristics based on nodes' degrees and centrality within the network, as well as the baseline of choosing random nodes to target. High-degree and distance-centrality heuristics choose nodes in the order of decreasing degree and decreasing average distances to other nodes, respectively. These heuristics are commonly used in the social science literature as estimates of a node's influence in the social network [37,8].

In each of these cases, in each step, we check whether the selected nodes are a perfect target set or not. This can be easily verified by simulating the dynamics until the states of nodes become stable. The simulation process ends at a polynomially bounded time τ when for each $v \in V(G)$ we have $f_\tau(v) = f_{\tau-2}(v)$ (see Theorem 7 and Theorem 8).

Notice that because the optimization problem is NP-hard (Theorem 5), and the testbed graphs are prohibitively large, we are not able to compute the optimum value to verify the actual quality of approximations.

Results. Fig. 4 shows the performance of our algorithm in comparison to the introduced heuristics on random power-law graphs. For any value of γ (power-law coefficient), all heuristics pick almost the entire nodes of the graph while our algorithm picks a number of them between the proved lower-bound and upper-bound. The same phenomenon happens for the four real-world social networks data. The results are depicted in Fig. 5.

Table 1
Results on the real networks.

Network	No. of nodes	γ	No. of nodes selected by algorithm			
			Greedy	High degree	Central	Random
Who-trusts-whom network of Epinions.com	75 888	1.50	27 131	75 878	75 879	75 888
Slashdot social network	77 360	1.68	49 978	77 327	77 360	77 360
Collaboration network of Arxiv Astro Physics	18 772	1.84	8287	18 771	18 772	18 763
Arxiv High Energy Physics paper citation network	34 546	2.05	14 647	34 539	34 546	34 505
Amazon product co-purchasing network	262 111	2.54	155 085	262 111	262 005	262 026

Table 1 includes the exact amount of greedy NPPTS's output compared to the output of other heuristics.

5.2. Convergence

In Section 4, we showed that the non-progressive spread of influence under strict majority threshold converge in asymptotically linear time over power-law graphs such as real social networks. In this part, we compute the average convergence time of this process in various networks and observe the fast convergence of such dynamics on those graphs. The average convergence time of a network G is the average number of rounds to reach a steady state when the initial infected set of nodes is picked uniformly, i.e., for a given network G, we calculate $E[ct_G(S)] = \frac{1}{2^{|V|}} \sum_{S \subseteq V} ct_G(S)$.

Computing this average in a brute force manner needs $\Omega(|V(G)|2^{|V(G)|})$ time, but the following theorem shows that bearing some error reduces this time to polynomial time.

Theorem 9. *Computing the average convergence time of the non-progressive spread of influence over strict majority threshold on graph G, with an additive error of ϵ is possible in time $O(\frac{e^3 \log(n)}{\epsilon^2})$ where $e = |E(G)|$ and $n = |V(G)|$.*

Proof. Define the random variable $X_S = ct_G(S)$. We uniformly select some of $V(G)$'s subsets $S_1, S_2, \ldots, S_m$ and take the average of X_{S_i}s. In [38], Hoeffding shows that with the large value of m and if X_{S_i} are bounded between a_i and b_i, $\overline{X_S}$ would be a good estimation (with an error less than ϵ) for $E[X_S]$ that is our desired target:

$$Pr\left(\left|\overline{X_S} - \mathrm{E}[\overline{X_S}]\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2\epsilon^2 m^2}{\sum_{i=1}^{m}(b_i - a_i)^2}\right).$$

From Theorem 8 we know putting $a_i = 0$ and $b_i = 8e$ for all $1 \leq i \leq m$, meets the preconditions of the above inequality. To have $Pr(|\overline{X_S} - \mathrm{E}[\overline{X_S}]| \geq \epsilon) \leq \frac{2}{n}$, we can set

$$m^2 \geq \frac{\ln(n)\sum_{i=1}^{m}(b_i - a_i)^2}{\epsilon^2} = \frac{64me^2\ln(n)}{\epsilon^2}.$$

Thus:

$$m \geq \frac{64e^2\ln(n)}{\epsilon^2}.$$

Since computing each X_{S_i} needs $O(e)$ (Theorem 8) the total time will be at most $O(me) = O(\frac{e^3 \log(n)}{\epsilon^2})$. $\square$

In a power-law graph G, since $|E(G)| \in O(|V(G)|)$, the average convergence time can be computed in time $O(\frac{n^3 \log(n)}{\epsilon^2})$ where $n = |V(G)|$.

As a result of Theorem 9, we can perform experimental evaluation of the convergence time in several families of graphs. In particular, through experimental evaluations, we show the average time of convergence for random power-law graphs with $\epsilon = 0.1$. Fig. 6 shows the average convergence time calculated by sampling for 500 random power-law graphs with an average of 100 nodes.

6. Conclusions

In this paper, we study the minimum target set selection problem in the non-progressive influence model under the strict majority rule and provide theoretical and practical results for this model. Our main results include upper bound and lower bounds for these graphs, hardness and an approximation algorithm for this problem. We also apply our techniques on power-law graphs and derive improved constant-factor approximation algorithms for this kind of graphs.

An important follow-up work is to study the minimum perfect target set problem for the non-progressive models under other influence propagation rules, e.g., the general linear threshold model. It is also interesting to design approximation algorithms for other special kinds of complex graphs such as small-world graphs. Another interesting research direction is to study maximum active set problem for non-progressive models.

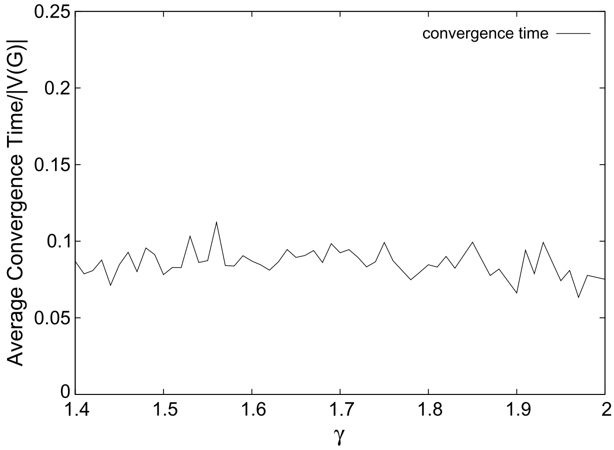

Fig. 6. The average convergence time on random power-law graphs.

Acknowledgements

The authors are thankful to the anonymous referees for carefully reading the manuscript and their many useful comments.

References

[1] L. Freeman, The Development of Social Network Analysis, Empirical Press Vancouver, British Columbia, 2004.
[2] Z. Dezső, A. Barabási, Halting viruses in scale-free networks, Phys. Rev. 65 (2002) 55103.
[3] R. Pastor-Satorras, A. Vespignani, Epidemic spreading in scale-free networks, Phys. Rev. Lett. 86 (2001) 3200–3203.
[4] D. Wilson, Levels of selection: an alternative to individualism in biology and the human sciences, Soc. Netw. 11 (1989) 257–272.
[5] J. Brown, P. Reingen, Social ties and word-of-mouth referral behavior, J. Consum. Res. 14 (1987) 350–362.
[6] P. Domingos, M. Richardson, Mining the network value of customers, in: Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26–29, 2001, San Francisco, CA, USA, Association for Computing Machinery, 2001, p. 57.
[7] M. Richardson, P. Domingos, Mining knowledge-sharing sites for viral marketing, in: Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2002, pp. 61–70.
[8] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2003, pp. 137–146.
[9] N. Immorlica, J. Kleinberg, M. Mahdian, T. Wexler, The role of compatibility in the diffusion of technologies through social networks, in: Proceedings of the 8th ACM Conference on Electronic Commerce, ACM, 2007, pp. 75–83.
[10] L. Blume, The statistical mechanics of strategic interaction, Games Econom. Behav. 5 (1993) 387–424.
[11] G. Ellison, Learning, local interaction, and coordination, Econometrica 61 (1993) 1047–1071.
[12] M. Jackson, L. Yariv, Diffusion on social networks, Écon. Publique 16 (2005) 69–82.
[13] J. Tang, J. Sun, C. Wang, Z. Yang, Social influence analysis in large-scale networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 807–816.
[14] A. Goyal, F. Bonchi, L. Lakshmanan, M. Balcan, N. Harvey, R. Lapus, F. Simon, P. Tittmann, S. Ben-Shimon, A. Ferber, et al., Approximation Analysis of Influence Spread in Social Networks, Arxiv preprint, arXiv:1008.2005, 2010.
[15] O. Ben-Zwi, D. Hermelin, D. Lokshtanov, I. Newman, An exact almost optimal algorithm for target set selection in social networks, in: Proceedings of the Tenth ACM Conference on Electronic Commerce, ACM, 2009, pp. 355–362.
[16] C. Chang, Y. Lyuu, Spreading messages, Theoret. Comput. Sci. 410 (2009) 2714–2724.
[17] N. Chen, On the approximability of influence in social networks, in: Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, Society for Industrial and Applied Mathematics, 2008, pp. 1029–1037.
[18] W. Chen, Y. Wang, S. Yang, Efficient influence maximization in social networks, in: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2009, pp. 199–208.
[19] P. Flocchini, R. Královi, P. Ruika, A. Roncato, N. Santoro, On time versus size for monotone dynamic monopolies in regular topologies, J. Discrete Algorithms 1 (2003) 129–150.
[20] D. Peleg, Local majorities, coalitions and monopolies in graphs: a review, Theoret. Comput. Sci. 282 (2002) 231–257.
[21] P. Flocchini, E. Lodi, F. Luccio, L. Pagli, N. Santoro, Dynamic monopolies in tori, Discrete Appl. Math. 137 (2004) 197–212.
[22] E. Mossel, G. Schoenebeck, Reaching consensus on social networks, in: First Symposium on Innovations in Computer Science, Beijing, China, 2010.
[23] E. Ackerman, O. Ben-Zwi, G. Wolfovitz, Combinatorial model and bounds for target set selection, Theoret. Comput. Sci. (2010).
[24] A. Clauset, C. Shalizi, M. Newman, Power-law distributions in empirical data, SIAM Rev. 51 (2009) 661–703.
[25] J. Goles, et al., Periodic behaviour of generalized threshold functions, Discrete Math. 30 (1980) 187–189.
[26] C.-L. Chang, C.-H. Wang, On reversible cascades in scale-free and Erdős–Rényi random graphs, Theory Comput. Syst. 52 (2013) 303–318.
[27] P. Flocchini, F. Geurts, N. Santoro, Optimal irreversible dynamos in chordal rings, Discrete Appl. Math. 113 (2001) 23–42.
[28] F. Luccio, L. Pagli, H. Sanossian, Irreversible dynamos in butterflies, in: Proceedings of the 6th Colloquium on Structural Information and Communication Complexity, Citeseer, 1999, pp. 204–218.
[29] D. Pike, Y. Zou, Decycling Cartesian products of two cycles, SIAM J. Discrete Math. 19 (2005) 651.
[30] C. Chang, Y. Lyuu, On irreversible dynamic monopolies in general graphs, Arxiv preprint arXiv:0904.2306, 2009.
[31] K. Khoshkhah, H. Soltani, M. Zaker, Dynamic monopolies in directed graphs: the spread of unilateral influence in social networks, Arxiv preprint arXiv:1212.3682, 2012.

[32] D. Peleg, Size bounds for dynamic monopolies, Discrete Appl. Math. 86 (1998) 263–273.
[33] E. Berger, Dynamic monopolies of constant size, J. Combin. Theory Ser. B 83 (2001) 191–200.
[34] E. Mossel, S. Roch, On the submodularity of influence in social networks, in: Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing, ACM, 2007, pp. 128–134.
[35] R. Allan, R. Laskar, On domination and independent domination numbers of a graph, Discrete Math. 23 (1978) 73–76.
[36] W. Aiello, F. Chung, L. Lu, A random graph model for massive graphs, in: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing, ACM, 2000, pp. 171–180.
[37] S. Wasserman, Social Network Analysis: Methods and Applications, Cambridge University Press, 1994.
[38] W. Hoeffding, Probability inequalities for sums of bounded random variables, J. Amer. Statist. Assoc. 58 (1963) 13–30.