arXiv:1106.3826v2 [cs.SI] 2 Aug 2011

On the Non-progressive Spread of Influence through Social Networks

MohammadAmin Fazli[1], Mohammad Ghodsi[1], Jafar Habibi[1]
Pooya Jalaly Khalilabadi[1], Vahab Mirrokni[2], and Sina Sadeghian Sadeghabad[1]

[1] Computer Engineering Department, Sharif University of Technology, Tehran, Iran
{fazli,jalaly,s_sadeghian}@ce.sharif.edu
{jhabibi,ghodsi}@sharif.edu
[2] Google Research NYC, 76 9th Ave, NewYork, NY 10011

Abstract. The spread of influence in social networks is studied in two main categories: the progressive model and the non-progressive model (see e.g. the seminal work of Kempe, Kleinberg, and Tardos in KDD 2003). While the progressive models are suitable for modeling the spread of influence in monopolistic settings, non-progressive are more appropriate for modeling non-monopolistic settings, e.g., modeling diffusion of two competing technologies over a social network. Despite the extensive work on the progressive model, non-progressive models have not been studied well. 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 node 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 a steady state where all nodes of G are infected.

We prove that while the MinPTS is NP-hard for a restricted family of graphs, it admits an improved constant-factor approximation algorithm for power-law graphs. We do so by proving lower and upper bounds in terms of the minimum and maximum degree 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* over in networks under various *propagation models* is a central issue in social network analysis[19,**?**,30,36]. This issue plays an important role in several real-world applications including viral marketing [6,14,33,26]. As categorized by Kempe et al. [26], there are two main types of influence propagation models: the progressive and the non-progressive models. In *progressive models*, infected (or influenced) vertices will remain infected forever, but in the *non-progressive model*, under some conditions, infected vertices may become uninfected again. In the context of 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 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 [23,5,15,25].

While the progressive model has been studied extensively in the literature [26,34,21,4,9,10,11], the non-progressive model has not received much attention in the literature. In this paper, we study the non-progressive influence models, and report both theoretical and experimental results for our models. We focus on the *the strict majority* propagation rule in which the state of each vertex at time τ is determined by the state 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 service, and nodes tend to adopt a service that the majority of their neighbors 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 [26], the strict majority propagation model is suitable for modeling transient faults in fault tolerant systems [17,31,18], and also used in verifying convergence of consensus problems on social networks [29]. Here we study the non-progressive influence models under the strict majority rule. In particular, we are mainly interested in *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 hardness results for the problem as well experimental evaluation of our results. As our main contributions, 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. Finally, we also study the convergence rate of our algorithms and report preliminary results. Before stating our results, we define the problems and models formally.

Problem Formulations. Consider a graph $G(V, E)$. Let $N(v)$ denote the set of neighbors of vertex v, and $d(v) = |N(v)|$. Also, let $\Delta(G)$ and $\delta(G)$ denote the maximum and minimum degree of vertices in G respectively. The induced

subgraph of G with a vertex set $S \subseteq V(G)$ is denoted by $G[S]$. Also $d_S(v)$ denotes the number of neighbors of v in subset S.

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 vertices at time τ and $t(v)$ be the threshold associated with vertex v. For the strict majority model, the threshold $t(v) = \lceil \frac{d(v)+1}{2} \rceil$ for each vertex v.

In the *non-progressive strict majority* model:

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

In *progressive strict majority* model:

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

Strict majority model is related to the *linear threshold model* in which $t(v)$ is chosen at random and 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 influence will converge to a steady state of all 1's. The cost of a target set f_0, denoted by $cost(f_0)$, is the number of vertices v with $f_0(v) = 1$. The minimum *perfect target set* (MinPTS) problem is to find a perfect target set with minimum cost. The cost of this minimum PTS is denoted by $PPTS(G)$ and $NPPTS(G)$ respectively for progressive and non-progressive models. This problem is also called *target set selection* [1]. Another variant of this problem is the *maximum active set* problem [1] where the goal is to find at most k nodes to activate (or infect) at time zero such that the number of finally infected vertices 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 degree x goes for large number of nodes as $P(x) = \alpha x^{-\gamma}$ where α is a constant and $\gamma > 1$ is called power-law coefficient. It is widely observed that most social networks are power-law [12].

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 experimental evaluation of our results. As our main contributions, 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 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 these algorithms. We first observe that the process will always converges to a fixed point or a cycle of size two. 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 [23,5,15,25]. 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 [23,5,15,25]. Note that none of these previous prior work provide a bound for the perfect target set problem.

It has been brought to our attention that in a relevant unpublished work by Chang [7], the MinPTS problem on pawer-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 amounts of γ under strict majority threshold.

Tight or nearly tight bounds on the $PPTS(G)$ are known for special types of graphs such as torus, hypercube, butterfly and chordal rings [16,17,27,31,32]. The best bounds for progressive strict majority model in general graphs are due to Chang and Lyuu. In [9], they showed that for a directed graph G, $PPTS(G) \leq \frac{23}{27}|V(G)|$. In [8], they improved their upper bound to $\frac{2}{3}|V(G)|$ for directed graphs and $\frac{|V(G)|}{2}$ for undirected graphs. However, to the best of our knowledge, there is no known bound for $NPPTS(G)$ for any type of graphs. In this paper, we will combinatorially prove that $\frac{2n}{\Delta(G)+1} \leq NPPTS(G) \leq \frac{n\Delta(G)(\delta(G)+2)}{4\Delta(G)+(\Delta(G)+1)(\delta(G)-2)}$.

It is known that the Target Set Selection problem and Maximum Active Set problem are both NP-hard in the linear threshold model [26], and approximation algorithms have been developed for these problems. Kempe et al. [26] and Mossel and Roch [28] present a $(1-\frac{1}{e})$-approximation algorithm for the maximum active set problem by showing that the set of finally influenced vertices 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 [21,4,8,10]. The inapproximability result of Chang and Lyuu in [8] on progressive strict majority threshold model is the most relevant result to our results. They show that unless $NP \subseteq TIME(n^{O(\ln \ln n)})$, no polynomial time $((1/2-\epsilon)\ln|V|)$-approximation algorithm exists for computing $PPTS(G)$. To the best of our knowledge, no complexity theoretic results have been obtained for the non-progressive models.

The problem of maximizing social influence for specific family of graphs has been studied under the name of *dynamic monopolies* in the combinatorics literature [16,17,27,31,32,8,1,7]. All these results are for the progressive model. 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. [26] who presented a general reduction from non-progressive models to progressive models. 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 lower bound and upper bounds for minimum PTS in graphs, and finally show that finding the minimum PTS in general graphs is NP-Hard.

Lower bound. The following theorem shows that if we have some lower bound and upper bound for minimum Perfect Target Set in bipartite graphs then these bounds could be generalized to all graphs (Theorem 1).

Lemma 1. *If $\alpha|V(H)| \leq NPPTS(H) \leq \beta|V(H)|$ for every bipartite graph H under strict majority threshold, then $\alpha|V(G)| \leq NPPTS(G) \leq \beta|V(G)|$ under strict majority threshold for every graph G (see appendix A).*

The following lemma shows characteristics of PTSs in some special cases. These will be used in proof of our theorems.

Lemma 2. *Consider the non-progressive model and let $G = (X, Y)$ be a bipartite graph and f_0 be a perfect target set under strict majority threshold. 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 vertex u in S such that $d_S(u) \leq d(u) - t(u)$ (see appendix A).*

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

Lemma 3. *Consider a graph G with n vertices. If for every $S \subseteq V(G)$ there exists at least one vertex v for which $d_S(v) \leq d(v) - t(v)$, then $|E(G)| \leq \sum_{u \in V(G)}(d(u) - t(u))$ (see appendix A).*

The following theorem shows that for every bipartite graph G, $NPPTS(G) \geq \frac{2|V(G)|}{\Delta(G)+1}$. Lemma 1 generalizes this theorem to all graphs. Also, Theorem 4 shows that this bound is tight. In the following, the induced subgraph of G with a vertex set $S \subseteq V(G)$ is denoted by $G[S]$.

Theorem 1. *For every bipartite graph $G = (X, Y)$ of order n, $NPPTS(G) \geq \frac{2n}{\Delta(G)+1}$.*

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

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

Consider the induced subgraph of G with vertex set $B_X \cup W$ and suppose that $S \subseteq B_X \cup W$. For every vertex $v \in Y \cap S$, we have $f_0(v) = 0$. So By Lemma 2, for every $S \subseteq B_X \cup W$ there is at least one vertex u such that $d_S(u) \leq d(u) - t(u)$. By Lemma 3, 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 point in B_X and the other end point in W and e_{WY} be the number of edges with one end point in B_Y and the other end point in W. we have:

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

and so,

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

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

$$\sum_{v \in W} d(v) \leq \sum_{v \in V(G)} (d(v) - t(v)) + \sum_{v \in W} (d(v) - t(v))$$

If we denote the set of vertices for which f_0 is equal to 1 by B, we have

$$\sum_{v \in W} (2t(v) - d(v)) \leq \sum_{v \in B} (d(v) - t(v)) \tag{1}$$

For every vertex v, $t(v) \geq \frac{d(v)+1}{2}$, so

$$|W| \leq \sum_{v \in B} \frac{d(v) - 1}{2} \Rightarrow |W| \leq \frac{\Delta - 1}{2}(|B|)$$
$$\Rightarrow |B| \geq \frac{2n}{\Delta + 1}$$

And the proof is complete.

We now show that the bound in Theorem 4 is tight.

Algorithm 1 Greedy NPPTS

```
sort the vertices in G in ascending order of their degrees as the sequence v_1, ..., 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 ∈ 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) = 1
  else
    f_0(v) = 0
    for each u ∈ N(v_i) do
      whiteadj[u] += 1
    end for
  end if
end for
```

Lemma 4. *For infinitely many n's, there exists a $2d+1$-regular graph with n vertices such that $NPPTS(G) = \frac{n}{d+1}$ under strict majority rule (see appendix A).*

Upper bound. In this section, we present a greedy algorithm which gives an upper bound for $NPPTS(G)$.

Theorem 2. *For every graph G of order n, $NPPTS(G) \leq \frac{n\Delta(\delta+2)}{4\Delta+(\Delta+1)(\delta-2)}$ under strict majority threshold.*

Algorithm 1 guarantees this upper bound. This algorithm gets a graph G of order n and the threshold function t as input and determines the values of f_0 for each vertex.

Lemma 5. *The algorithm Greedy NPPTS finds a Perfect Target Set for non-progressive spread of influence.*

Proof. By induction on the number of vertices for which f_0 is determined, 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 vertex for which 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 vertex of v, say u, has exactly $d(u)-t(u)$ adjacent initially uninfected vertices. So by setting $f_0(v)$ to 0, each initially infected vertex w still has at least $t(w)$ infected vertices

and also v has at least $t(v)$ initially infected neighbors itself. Thus, after one step of propagation, all initially infected vertices plus v are infected and by induction hypothesis, all vertices will be infected eventually and so f_0 remains a PTS.

Lemma 6. *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 degree of vertices respectively.*

Proof. According to the algorithm, for each vertex 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 vertices u for which $\texttt{whiteadj}[u] = d(u) - t(u)$. B and W denote the set of infected and uninfected vertices respectively. We have:

$$\sum_{v\in S}(d(v) - t(v)) \leq \sum_{v\in W} d(v) \Rightarrow \sum_{v\in S}(\frac{d(v)}{2} - 1) \leq \sum_{v\in W} d(v)$$

Therefore,

$$(\frac{\delta}{2} - 1)|S| \leq \Delta|W|. \Rightarrow |S| \leq \frac{2\Delta}{\delta - 2}|W|$$

Each vertex in B has at least one adjacent vertex in S and each vertex $v \in S$ has at least $d(v) - t(v)$ adjacent edges to W and so at most $t(v)$ adjavent edges to B, thus:

$$\begin{aligned}|B| &\leq \sum_{v\in S}(t(v)) \leq \sum_{v\in S}(\frac{d(v)}{2} + 1) \leq 2|S| + \sum_{v\in W} d(v)\\ &\leq 2|S| + \Delta|W| \leq (2\frac{2\Delta}{\delta - 2} + \Delta)|W| \leq \frac{4\Delta + \Delta(\delta - 2)}{\delta - 2}|W|\end{aligned}$$

Thus,

$$|B| \leq \frac{\Delta(\delta + 2)}{4\Delta + (\Delta + 1)(\delta - 2)}n$$

The approximation factor of the algorithm follows from previous claim and the lower bound provided by Theorem 1:

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) [3] to prove the NP-hardness of computing $NPPTS(G)$. The proof of following theorem is provided in appendix A.

Theorem 3. *If there exists a polynomial-time algorithm for computing $NPPTS(G)$ for a given graph G under the strict majority threshold, then $P = NP$.*

3 Non-Progressive Spread of Influence in Power-law graphs

In this section, we investigate the non-progressive spread of influence in power-law graphs, and show that the greedy algorithm presented in the previous section is indeed a constant-factor approximation algorithm for power-law graphs. For each natural number x, we assume that the number of vertices 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 . We denote the number of vertices of degree x by $P(x) = \alpha x^{-\gamma}$. Let n be the number of vertices of graph, so we have:

$$n = \sum_{x=1}^{\infty} \alpha x^{-\gamma} = \alpha\zeta(\gamma) \Rightarrow \alpha = \frac{n}{\zeta(\gamma)},$$

where ζ is the Riemann Zeta function [24].

Lower bound. Consider a power-law graph G with a threshold function t and a perfect target set f_0. Denoting the set of initially influenced vertices by B and the rest of the vertices by W, from the Equation 1, we have:

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

The maximum cardinality of W is achieved when the degree of all vertices in B is greater than or equal to the degree of all vertices in W. In this case, assume that the minimum degree of vertices in B is k and $0 \leq p \leq 1$ is the proportion of the vertices of degree k in B, so under strict majority threshold we have:

$$\begin{aligned}
&\sum_{x=1}^{k-1} \alpha x^{-\gamma} + (1-p)\alpha k^{-\gamma} \leq |W| \leq \sum_{v\in W}(2t(v) - d(v)) \\
&\leq \sum_{v\in B}(d(v) - t(v)) \leq \sum_{x=k+1}^{\infty} \alpha x^{-\gamma}(\frac{x-1}{2}) + p\alpha k^{-\gamma}\frac{k-1}{2} \\
\Rightarrow &\sum_{x=1}^{k-1} x^{-\gamma} + (1-p)k^{-\gamma} \leq \frac{\sum_{x=k+1}^{\infty}(x^{1-\gamma} - x^{-\gamma}) + pk^{-\gamma}(k-1)}{2} \\
\Rightarrow &\zeta(\gamma) - \zeta(\gamma, k-1) + (1-p)k^{-\gamma} \\
&\leq \frac{\zeta(\gamma-1, k) - \zeta(\gamma, k) + pk^{-\gamma}(k-1)}{2}.
\end{aligned}$$

By estimating the value of Riemann Zeta function, we can estimate the upper bound of k and lower bound of p for that k to provide a lower bound for $|B|$. Assuming that we have the maximum possible value of k and minimum value of p for that k, then:

$$|B| \geq \sum_{x=k+1}^{\infty} \alpha x^{-\gamma} + \alpha p k^{-\gamma} = \frac{\zeta(\gamma, k) + pk^{-\gamma}}{\zeta(\gamma)} n.$$

The estimated values of lower bound for $2 \leq \gamma \leq 2.8$ is shown in Figure 1.
Upper bound Suppose that one has run Greedy NPPTS algorithm under strict majority threshold on a graph with power-law degree distribution. The following theorem shows that unlike general graphs, the Greedy NPPTS algorithm guarantees a constant factor upper bound on power-law graphs.

Theorem 4. *Algorithm Greedy NPPTS initially influences at most* $(1 + \frac{1}{2^{\gamma+1}} - \frac{1}{2\zeta(\gamma)})n$ *vertices under the strict majority threshold on a power-law graphs of order* n.

Proof. We may assume that the input graph is connected. We prove that the number of uninfected vertices of degree 1 are sufficient for this upper bound. Let v be a vertex of degree more than 1 with k adjacent vertices of degree 1 say $u_1, u_2 \ldots u_k$. If $d(v)$ is odd, it is clear that at least $\frac{k}{2}$ of the vertices $u_1, u_2 \ldots u_k$ will be uninfected since $k \leq d(v)$. Note that according to the greedy algorithm, the value of f_0 for degree 1 vertices are determined before any other vertex. If $d(v)$ is even, at least $\frac{k}{2} - 1$ of vertices $u_1, u_2 \ldots u_k$ will be uninfected. Therefore we have:

$$\begin{aligned} NPPTS(G) &\leq n - \frac{1}{2}(P(1) - \sum_{x=1}^{\infty} P(2x)) \\ &\leq n - \frac{1}{2}(\alpha \frac{1}{1^{\gamma}} - \alpha \sum_{x=1}^{\infty} \frac{1}{(2x)^{\gamma}}) \\ &= n - \frac{\alpha}{2}(1 - \frac{1}{2^{\gamma}}\zeta(\gamma)) = n(1 + \frac{1}{2^{\gamma+1}} - \frac{1}{2\zeta(\gamma)}) \end{aligned}$$

By previous lemma, we conclude that the Greedy NPPTS algorithm is a constant-factor approximation algorithm on power-law graphs under strict majority threshold. The lower bound and upper bound for different values of γ are shown in Figure 1. As you can see our algorithm acts optimally on social networks with large value of power-law coefficient since upper and lower bound diagram meet each other for these values of power-law coefficient.

4 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 [26], we compare the performance of our algorithm to other heuristics for identifying influential individuals.
Random power-law and real-world 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 power-law graphs from [2], 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). For details of the way to generate random power-law graphs, see Appendix [**?**]. We also

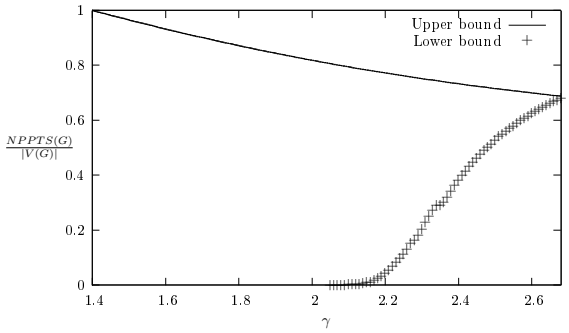


Fig. 1. Values of upper bound and lower bound in power-law graphs

run our 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 graphs' giant component.

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 vertices in the order of decreasing degree and decreasing average distance to other nodes. These heuristics are commonly used in the social science literature as estimates of a node's influence in the social network [35,26].

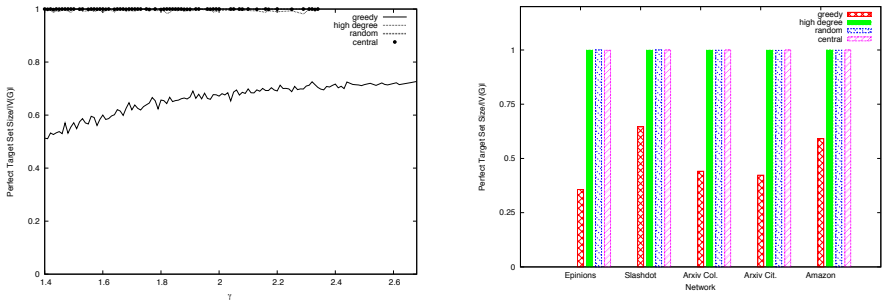
In each of these cases, in each step, we check whether the selected vertices are a perfect target set or not. This can be easily verified by simulating spread of influence process until the states of vertices become stable. The simulation process ends at a polynomially bounded time τ when for each $v \in V(G)$ we have $f_\tau(v) = f_{\tau-2}(v)$ (see Theorem 5 and Theorem 6).

Notice that because the optimization problem is NP-hard (Theorem 3), and the testbed graphs are prohibitively large, we are not able to compute the optimum value to verify the actual quality of approximations.

Experimental Results. Figure 2(a) shows the performance of our algorithm in comparison to introduced heuristics on random power-law graphs. For any value of γ (power-law coefficient), all heuristics pick almost entire vertices of the graph while our algorithm pick a number of them between proved lower-bound and upper-bound. The same phenomena happens for the four real-world social networks data. The results are depicted in Figure 2(b).

5 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



(a) Results on the random power-law graphs (b) Results on the real network data

applying one step of the influence process on state A changes the state to state B. First of all, one can easily see that the non-progressive model may not result to a singleton steady state for any dynamics. To see this, consider the following example: a cycle with $2k$ vertices $C = v_1v_2...v_{2k}$ and at time 0 infect vertices 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 5. *The non-progressive spread of influence process on a graph reaches a cycle of length of at most two (see appendix C).*

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 two states, i.e., the convergence time is the minimum time T at which $f_T(v) = f_{T+2}(v)$ for all vertices $v \in V(G)$. For a set S of initially infected vertices, let $ct_G(S)$ to 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 6. *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 vertices 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 sequential dynamics, the vertices apply the influence process one by one in a sequence of rounds, where in each step one vertex applies the influence process exactly once.

We first show that the sequential dynamic 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. One can see that whenever a vertex

changes its state from uninfected to infected the potential of G will decrease at least one and otherwise it remains unchanged. Consider a vertex which has k state changes during the process until it 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 dynamic in graph H is at most $2|V(H)|.|E(H)| = 8|V(G)|.|E(G)| = c|V(G)|.T$. So $T \in O(|E(G)|)$. The main claim follows from the proof of Lemma 1. By induction on the number of steps, we can show that the state of vertices in G is equal to the state of vertices in X at odd steps and is equal to the state of vertices in Y at even steps (as we did in the proof of Lemma 1). Now order vertices 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 $O(|E(G)|)$. 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 $O(|V|)$.

Finally, we study convergence time of non-progressive dynamics on several real-world graphs, and observe the fast convergence of such dynamics on those graphs. See Appendix C for details.

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 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 set problem for 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.

Acknowledgments

The authors are thankful to Soroush Hosseini and Morteza Saghafian for their ideas and their helps.

References

1. E. Ackerman, O. Ben-Zwi, and G. Wolfovitz. Combinatorial Model and Bounds for Target Set Selection. *Theoretical Computer Science*, 2010.
2. W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 171–180. Acm, 2000.
3. R. Allan and R. Laskar. On domination and independent domination numbers of a graph. *Discrete Mathematics*, 23(2):73–76, 1978.
4. O. Ben-Zwi, D. Hermelin, D. Lokshtanov, and I. Newman. An exact almost optimal algorithm for target set selection in social networks. In *Proceedings of the tenth ACM conference on Electronic commerce*, pages 355–362. ACM, 2009.
5. L. Blume. The statistical mechanics of strategic interaction. *Games and Economic Behavior*, 5:387–424, 1993.
6. J. Brown and P. Reingen. Social ties and word-of-mouth referral behavior. *The Journal of Consumer Research*, 14(3):350–362, 1987.
7. C. Chang. On reversible cascades in scale-free and Erdos Renyi random graphs. *Arxiv preprint arXiv:1011.0653*, 2010.
8. C. Chang and Y. Lyuu. On irreversible dynamic monopolies in general graphs. *Arxiv preprint arXiv:0904.2306*, 2009.
9. C. Chang and Y. Lyuu. Spreading messages. *Theoretical Computer Science*, 410(27-29):2714–2724, 2009.
10. N. Chen. On the approximability of influence in social networks. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1029–1037. Society for Industrial and Applied Mathematics, 2008.
11. W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 199–208. ACM, 2009.
12. A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.
13. Z. Dezső and A. Barabási. Halting viruses in scale-free networks. *Physical Review E*, 65(5):55103, 2002.
14. P. Domingos and M. Richardson. Mining the network value of customers. In *KDD-2001: proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 26-29, 2001, San Francisco, CA, USA*, page 57. Assn for Computing Machinery, 2001.
15. G. Ellison. Learning, local interaction, and coordination. *Econometrica*, 61:1047–1071, 1993.
16. P. Flocchini, F. Geurts, and N. Santoro. Optimal irreversible dynamos in chordal rings. *Discrete Applied Mathematics*, 113(1):23–42, 2001.
17. P. Flocchini, R. Královi, P. Ruika, A. Roncato, and N. Santoro. On time versus size for monotone dynamic monopolies in regular topologies. *Journal of Discrete Algorithms*, 1(2):129–150, 2003.

18. P. Flocchini, E. Lodi, F. Luccio, L. Pagli, and N. Santoro. Dynamic monopolies in tori. *Discrete applied mathematics*, 137(2):197–212, 2004.
19. L. Freeman. *The development of social network analysis.* Empirical Press Vancouver, British Columbia, 2004.
20. J. Goles et al. Periodic behaviour of generalized threshold functions. *Discrete Mathematics*, 30(2):187–189, 1980.
21. 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.
22. W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
23. N. Immorlica, J. M. Kleinberg, M. Mahdian, and T. Wexler. The role of compatibility in the diffusion of technologies through social networks. pages 75–83, 2007.
24. A. Ivic. Riemann zeta-function. *JOHN WILEY & SONS, INC., ONE WILEY DRIVE, SOMERSET, NJ 08873(USA), 1985, 340*, 1985.
25. M. Jackson and L. Yariv. Diffusion on social networks. *EconomiePublique*, 16:69–82, 2005.
26. D. Kempe, J. Kleinberg, and 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*, pages 137–146. ACM, 2003.
27. F. Luccio, L. Pagli, and H. Sanossian. Irreversible dynamos in butterflies. In *Proc. of 6th Colloquium on Structural Information and Communication Complexity*, pages 204–218. Citeseer, 1999.
28. E. Mossel and S. Roch. On the submodularity of influence in social networks. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 128–134. ACM, 2007.
29. E. Mossel and G. Schoenebeck. Reaching consensus on social networks. *Innovations in Computer Science, ICS*, 2009.
30. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200–3203, 2001.
31. D. Peleg. Local majorities, coalitions and monopolies in graphs: a review. *Theoretical Computer Science*, 282(2):231–257, 2002.
32. D. Pike and Y. Zou. Decycling Cartesian products of two cycles. *SIAM Journal on Discrete Mathematics*, 19:651, 2005.
33. M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 61–70. ACM, 2002.
34. J. Tang, J. Sun, C. Wang, and Z. Yang. Social influence analysis in large-scale networks. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 807–816. ACM, 2009.
35. S. Wasserman. *Social network analysis: Methods and applications.* Cambridge university press, 1994.
36. D. Wilson. Levels of selection: An alternative to individualism in biology and the human sciences. *Social Networks*, 11(3):257–272, 1989.

A Proofs from Section 2

In this section we give the missing proofs from Section 2.

Proof of Lemma 1. Consider a graph G with n vertices and vertex set $V(G) = \{v_1, v_2, \dots v_n\}$ and threshold function t. Assume that there is a Perfect Target Set f_0 for G such that $cost(f_0) < \alpha|V(G)|$. Let $H = (X, Y)$ be a bipartite graph such that $X = \{x_1, \dots x_n\}$ and $Y = \{y_1, \dots y_n\}$ and t' be the threshold function of vertices of H such that for every $1 \leq i \leq n$, $t'(x_i) = t'(y_i) = t(v_i)$. 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 $1 \leq i \leq n$. 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 $1 \leq i \leq n$. By the definition, the assertion is true for $\tau = 0$. Now let the assertion be true for time τ. Consider a vertex $x_i \in X$. 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. So g_0 is a PTS for H iff f_0 is a PTS for G, which is a contradiction since by assumption $NPPTS(H) \geq \alpha|V(H)|$ but $cost(g_0) < \alpha|V(H)|$.

Now we prove that $NPPTS(G) \leq \beta|V(G)|$. 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 vertices in X for which g'_0 is one (initially infected vertices) is less than the number of initially infected vertices of Y. Let f'_0 be a PTS for G such that $f'_0(v_i) = g'_0(x_i)$ for every $1 \leq i \leq n$. We have $cost(g'_0) \leq \beta|V(G)|$ 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 $1 \leq i \leq n$ and every $\tau \geq 0$. The assertion is trivial for $\tau = 0$. Now let the assertion be true for time 2τ. Consider a vertex $v_i \in V(G)$. 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. So g'_0 is a PTS for H iff f'_0 is a PTS for G and so $NPPTS(G) \leq \beta|V(G)|$.

In the following, $d_S(v)$ denotes the number of neighbors of v in subset.

Proof of Lemma 2. Consider a set $S \subseteq V(G)$. 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 vertices in $S \cap X$ for which f_0 is zero. Similarly, for every $x \in S \cap X$, $f_2(x) = 0$ since x has at least $d(x) - t(x) + 1$ adjacent vertices in $S \cap Y$ for which f_1 is zero, and so on. Thus f_0 is not a Perfect Target Set, a contradiction.

Proof of Lemma 3. We prove the lemma by induction on n. For $n = 1$ the assertion is trivial. Consider a graph G with n vertices. Let $S = V(G)$. By assumption, there is at least one vertex v, such that $d(v) \leq d(v) - t(v)$. Remove the vertex v from G. By induction hypothesis $G - v$ has at most $\sum_{u \in V(G-v)}(d(u) - t(u))$ edges, so G has at most $\sum_{u \in V(G)}(d(u) - t(u))$ edges.

Proof of Theorem 4. Consider a $(d+1)$-regular graph G_1 with m_1 vertices . In

step i ($1 \leq i \leq \infty$), Add $m_{i+1} = \frac{d}{d+1} m_i$ vertices to the graph and connect each of them to G_i by $d+1$ edges. Each vertex of G_i must receive exactly d newly edges. Name the subgraph formed by these vertices G_{i+1}. This process is shown in Figure 2. The final graph has $n = \sum_{i=1}^{\infty} m_i = m_1(d+1)$ vertices. It is easy to show that $V(G_1)$ is a PTS, so $NPPTS(G) \leq |V(G_1)| = m_1 = \frac{2n}{2d+2} = \frac{2n}{\Delta+1}$.

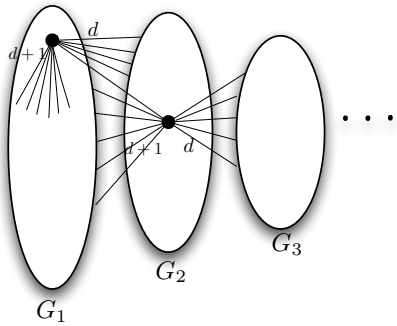


Fig. 2. A tight example for $NPPTS(G)$'s lower bound

Proof of Theorem 3. In an instance of the minimum dominating set problem (MDS), given a graph $G(V, E)$, our goal is to find find a subset $S \subseteq V(G)$ of minimum cardinality such that for any node $v \notin S$, we have $S \cap N(v) \neq \emptyset$. We give reduction from this NP-hard problem to our problem. Given an instance G of MDS with $V(G) = \{u_1, u_2, ..., u_n\}$ and $|E(G)| = e$, we define an undirected graph H as follows (See Figure 3): First, let

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

Now let $H(V, E)$ be

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

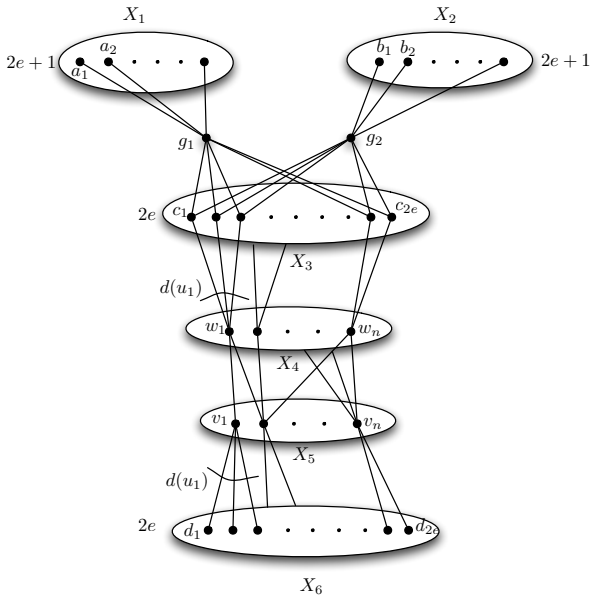


Fig. 3. The graph H

$$
\begin{aligned}
E(H) = &\{g_1a_i | 1 \le i \le 2e+1\} \\
&\cup \{g_2b_i | 1 \le i \le 2e+1\} \\
&\cup \{g_1c_i | 1 \le i \le 2e\} \\
&\cup \{g_2c_i | 1 \le i \le 2e\} \\
&\cup \{w_ic_j | 1 \le i \le n, \sum_{k=1}^{i-1} d(u_k) \le j \le \sum_{k=1}^{i} d(u_k)\} \\
&\cup \{v_iw_j | u_iu_j \in E(G) \vee i = j\} \\
&\cup \{v_id_j | 1 \le i \le n, \sum_{k=1}^{i-1} d(u_k) \le j \le \sum_{k=1}^{i} d(u_k)\}.
\end{aligned}
$$

Suppose that D is a minimum dominating set for G. Define $D^H = \{v_i | u_i \in D\}$. We show that $NPPTS(G) = 2e + n + 4 + |D|$. It is easy to see that vertices in $X_0 \cup X_3 \cup X_4 \cup D^H$ plus one vertex from each of X_1 and X_2 form a Perfect Target Set for the graph H. So, we have $NPPTS(H) \le |X_0| + |X_3| + |X_4| + |D^H| + 2 = 2e + n + 4 + |D|$.

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 vertex g_1 in time τ. If $f_\tau(g_1) = 0$, in time $\tau + 1$ for every vertex $a_i \in X_1$ we will have $f_{\tau+1}(a_i) = 0$

and then $f_{\tau+2}(g_1) = 0$. So we have, $g_1 \in S$. Similarly, we have $g_2 \in S$. Moreover, at least $2e+1$ vertices 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 vertex from each of X_1 and X_2 are in S. By this setting, the vertices of $X_0 \cup X_1 \cup X_2 \cup X_3$ become infected and keep this infection for every $\tau > 0$.

Consider a vertex $w_k \in X_4$. Let $B(w_k) = \{d_i \in X_6 | d_i$ is reachable from w_k by a path of length 2$\}$. 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 $|S|$. So, we may assume that $B(w_k) \cap S = \emptyset$ when $w_k \notin S$. Now, 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. Thus 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, for every $w_k \in X_4$, at least one of its neighbors in X_5 must be in S. This means that $S \cap X_5$ must have at least $|D|$ vertices and the proof is complete.

B Experimental Evaluation Data on Real Networks

Generating andom 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 power-law graphs from [2], 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). The number of vertices with degree x, y satisfies

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

The random power-law graph model is defined as follows: given n weighted vertices with weights $w_1, w_2, \cdots, w_n$, a pair (i, j) of vertices 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

- $\sharp\{i | w_i = 1\} = \lfloor e^{\alpha} \rfloor - r$ and $\sharp\{i | w_i = k\} = \lfloor \frac{e^{\alpha}}{k^{\gamma}} \rfloor$ for $k = 2, 3, .., \lfloor e^{\frac{\alpha}{\gamma}} \rfloor$. 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 i'th vertex would be w_i and also vertices' weights follow power-law.

Experimental results for four real-world networks. Table 1 includes the exact amount of greedy NPPTS's output compared to the output of other heuristics.

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	75888	1.50	27131	75878	75879	75888
Slashdot social network	77360	1.68	49978	77327	77360	77360
Collaboration network of Arxiv Astro Physics	18772	1.84	8287	18771	18772	18763
Arxiv High Energy Physics paper citation network	34546	2.05	14647	34539	34546	34505
Amazon product co-purchasing network	262111	2.54	155085	262111	262005	262026

C Missed Things from Section 5

Proof of theorem 5. In [20], it is shown that, for a function Δ from $\{0,1\}^n$ to $\{0,1\}^n$ whose components from a symmetric set of threshold functions, the repeated application of Δ, leads either to a fixed point or to a cycle of length two. Since the set of functions f_τ (defined in Section 1) are symmetric threshold functions, the lemma follows immediately from this fact.

Average convergence time of the process on social networks. Applying a sampling technique and simple concentration inequalities, one can easily show that the average convergence time of the non-progressive process on graph G can be approximated with an additive error of ϵ in time $O(\frac{e^2.n\log(n)}{\epsilon^2})$ where $e = |E(G)|$ and $n = |V(G)|$.

Theorem 7. *Computing the average convergence time of the non-progressive process on graph G, with an error of ϵ is possible in time $O(\frac{e^2.n\log(n)}{\epsilon^2})$ where $e = |E(G)|$ and $n = |V(G)|$.*

Proof. Define random variable $X_S = ct_G(S)$. We uniformly select some of the $V(G)$'s subsets $S_1, S_2, ..., S_m$ and take the average of X_{S_i}s. In [22], Hoeffding shows that with large value of m and if X_{S_i}s 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:

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

From Theorem 6 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 $Pr(|\overline{X_S} - \mathrm{E}[\overline{X_S}]|) \leq \frac{2}{n}$, we can set

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

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

Corollary 2. *Computing the average convergence time of the process on a power-law graph G, with an error of ϵ is possible in time $O(\frac{n^3 \log(n)}{\epsilon^2})$ where $n = |V(G)|$.*

As a result, we can perform experimental evaluation of 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 $\epsilon = 0.1$. Figure 4 shows average convergence time calculated by sampling for 500 random power law graphs with average of 100 vertices.

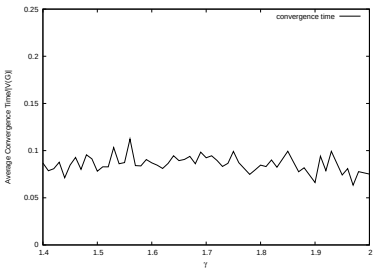


Fig. 4. Average convergence time on random power-law graphs