

Threshold-Bounded Influence Dominating Sets for Recommendations in Social Networks

Magdalini Eirinaki¹ and Nuno Moniz² and Katerina Potika³

Abstract—The process of decision making in humans involves a combination of the genuine information held by the individual, and the external influence from their social network connections. This helps individuals to make decisions or adopt behaviors, opinions or products. In this work, we seek to investigate under which conditions and with what cost we can form neighborhoods of influence within a social network, in order to assist individuals with little or no prior genuine information through a two-phase recommendation process. Most of the existing approaches regard the problem of identifying influentials as a long-term, network diffusion process, where information cascading occurs in several rounds and has fixed number of influentials. In our approach we consider only one round of influence, which finds applications in settings where timely influence is vital. We tackle the problem by proposing a two-phase framework that aims at identifying influentials in the first phase and form influential neighborhoods to generate recommendations to users with no prior knowledge in the second phase. The difference of the proposed framework with most social recommender systems is that we need to generate recommendations including more than one item and in the absence of explicit ratings, solely relying on the social network’s graph.

Index Terms—social network, greedy algorithm, dominating set, influentials, social recommender systems

I. INTRODUCTION

Humans make decisions, adopt behaviors and form opinions through a process that takes as input a combination of the genuine information held by the individual, and the external influence from their peers. This leads to a common phenomenon, where behaviors and opinions deviate depending on the context, whether this is a demographic group, a geographic location, or a different point in time. One way to explain this phenomenon is that this behavior is adopted by a portion of the individual’s peers, henceforth referred to as their *neighborhood*.

For example, consider the case where a University wants to increase the number of female students in STEM fields by reaching out to the student population in local schools. This is often done with summer camps or one-day events, where female students are introduced to STEM disciplines in the University hoping they will be inspired to follow a related career path and thus consider applying to one of the programs offered. However, Universities have limited resources, both in terms of time and budget. Such events run once per year

and only a few students can attend. Ideally, we need to identify and invite the *influential* students of each social neighborhood, who will not only get convinced to follow such a career path, but will potentially convince their peers to do so. Taking this one step further, the influence happens in two levels of abstraction: a high-level one that guides to a potential STEM career, and a more specific one that guides the decision to a specific major. In this case, and viewed by the “influenced” student’s point of view, her active peers’ major preferences will also influence her own.

The problem of identifying influentials in a social network has gained a lot of attention from various research communities, as it has applications in viral marketing, disease prevention, disease propagation, politics etc. This line of work is drawing inspiration by social correlation theories such as homophily and social influence [21], [25]. However most of the existing approaches regard this as a long-term, network diffusion process, where information cascading occurs in several rounds [4], [15], [13]. Additionally, the assumption is that there is a fixed number of influential neighbors each node must have [8]. Such models do not fit situations like the one described above, where only one round of influence is assumed (e.g. only one group of prospective female students can be invited each academic year).

In this work, we seek to investigate under which conditions and with what cost we can form neighborhoods of influence within a social network, to assist individuals with little or no prior genuine information. In more technical terms, given a social network, we model it as a graph, where the vertices correspond to individuals and the edges are the corresponding social ties. The graph theoretic problem is to identify which vertices (i.e. people) in the social graph need to be “activated” (i.e. targeted) such that the remaining vertices can make decisions by taking into consideration the opinion of active vertices in their immediate neighborhood, given specific constraints. This is a generalized version of the classical minimum dominating set [10]. We assume that the edges are undirected and consider different models of influence that don’t include cascading. We describe and solve a graph theoretical problem similar to the self-monopolies of [23]. Rather than requiring the majority the model requires vertex-depended thresholds. We view this model as a component of a social graph-based recommendation framework, aimed at assisting users with no prior information to make decisions. This two-phase framework aims at identifying influential neighborhoods (i.e. active vertices) in the first phase, and, in the absence of explicit item ratings, generating recommendations to the users with no prior knowledge in

¹San Jose State University, Computer Engineering Department, San Jose, California, U.S.A. magdalini.eirinaki@sjsu.edu

²LIAAD - INESC Tec, FCUP-DCC, University of Porto, Porto, Portugal nmmoniz@inescporto.pt

³San Jose State University, Computer Science Department, San Jose, California, U.S.A. katerina.potika@sjsu.edu

the second phase. The difference of the proposed framework with most social recommender systems is that we need to generate recommendations including more than one item without the use of explicit ratings, solely based on the social network’s graph.

The rest of the paper is organized as follows. The two-phase social recommendation framework is outlined in Section II. We formally describe the model related to neighborhood formation and recommendation process and propose an algorithmic solution to the first phase in Section III. We evaluate the proposed algorithm in different types of networks and discuss our findings in Section IV. An overview of the related work is provided in Section V. Our conclusions and plans for future work are included in Section VI.

II. A TWO-PHASE SOCIAL RECOMMENDATION FRAMEWORK

Social recommender systems leverage social relations to improve the rating-based recommendation process, based on the assumption that a user’s preferences are likely to be similar to, or influenced by these of her friends. This assumption roots from the concepts of homophily and social influence ([21], [25]). In this paper, we address a slightly different problem, that of generating recommendations for users with no or little prior knowledge, in the absence of a rating system that could enable a more traditional, collaborative filtering approach.

Consider, for example, the scenario described before, where a student is first influenced to follow a STEM career (phase 1) and then has to decide which specific major to pursue (phase 2). Another example highlighting the need of a two-phase social recommendation framework is the election process. Consider, for instance, the U.S. primaries, or any local election procedure with multiple candidates. The candidates face two hurdles: convincing the citizens to vote for them, but most importantly, convincing them to vote (abstention reached an abysmal 68% in the 2014 U.S. elections¹). Thus candidates need to first identify influential active voters who can convince their peers to show up on election day (phase 1). Each of them holds their own preference list of candidates and conveys it to the ones who haven’t formed an opinion yet and should, in turn decide who to vote for (phase 2). Moreover, the time is usually limited (especially for local elections, when the campaign period is short), and thus we cannot rely on those newly informed citizens to influence others in turn. While such scenarios do not follow the typical set-up of a recommender system (with users, items, and explicit ratings), the ultimate objective is the same, that of providing a user with some form of recommendation².

¹Source: <http://www.idea.int/vt/countryview.cfm?CountryCode=US>

²Note that, in the real-world the second phase that is primarily a mental process can be enhanced by the outcomes of such a system, while both examples transfer to online communities and social networks, where a recommender system may be employed.

A. Influence-based Social Recommendations

In this work, we propose an influence-based social recommendation framework that enables recommendations in the presence of specific constraints and characteristics that set it apart from typical social recommender systems:

- *Two-phase recommendations.* Recommendations happen in two levels of abstraction, mapped to a two-phase process: a) a higher level of abstraction, where an “inactive” user gets influenced by “active” users in making a decision (e.g. follow a STEM career or vote), and b) a finer-grained level of abstraction, where the influenced user is provided with explicit item recommendations (e.g. which major to follow or candidate to support).
- *Recommendations only for cold-start users.* We assume that the active, influential users whom we target first, have already formed an opinion/made a decision (during the first phase). The focus is on the “cold-start” users, who have little or no prior knowledge and for whom the system only knows their social connections. The second phase of the recommendation framework addresses only these inactive users.
- *Single round of influence.* Because of time constraints, there can be only one full round of influence (i.e. active users can only influence their direct connections) and thus further influence diffusion is not guaranteed.
- *Absence of ratings.* The user similarity can no longer be defined in terms of similar ratings, as in collaborative filtering systems. Instead, only the social connections can be leveraged and used as input to the recommendation process.
- *Personalized, preference-based recommendations.* Each of the influential users maintains a preference list of items (e.g. majors or candidates). The final recommendation to each inactive user should be a personalized aggregated list and not a single item.

In what follows, we define a two-phase social recommendation framework with no influence propagation. The framework consists of two main modules, the neighborhood formation module, and the recommendations’ generation module. An example of this two-phase process is illustrated in Figures 1 and 2. We formulate the social network as a graph, where the vertices correspond to individuals and the edges correspond to the social relationships between them. We should also note that such a framework is context-aware since, depending on the context, the social graph of the users is different (e.g. use the school social graph to choose a major and the friends social graph to decide on a vacation destination).

1) *Phase 1 - Neighborhood formation:* During the neighborhood formation phase, the goal is to identify, given a social graph, which vertices need to be *activated* such that the remaining vertices can make decisions given some threshold. For instance, if the threshold is 50%, each inactive node can make decisions if at least half of their connected vertices are active. The threshold is a parameter of the framework and changes depending on the application domain (e.g. in spam

networks, only 30% of the vertices need to be malicious to influence the rest). We describe the active connections of each inactive node as its *neighborhood* and these are the only vertices that will be used as input in the recommendations' generation process in the next phase. Note that the measures that are broadly used in recommender systems to identify similar users (e.g. Pearson correlation or latent factor models) are not applicable here since there are no explicit item ratings.

The first phase is depicted in Figure 1. The influential users are first identified, given a specific threshold (in this example threshold is 50%, and the influentials are shown in white shirts). The influentials are targeted (for example, they are invited to attend a STEM-related activity). Only the direct connections of each user (example user is shown enlarged) are part of her neighborhood (non-neighborhood users are shown transparent) and influence her (e.g. to follow a STEM career).

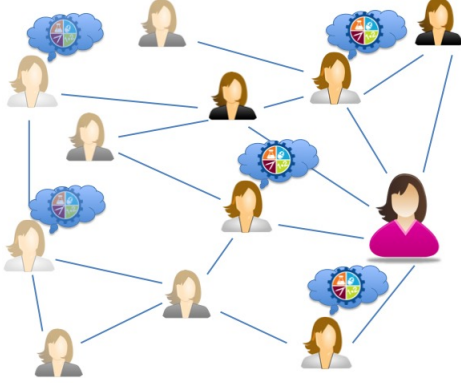


Fig. 1. Phase 1 - Identifying influentials to form neighborhoods of influence.

2) *Phase 2 - Generating Recommendations:* Once the neighborhood of each inactive node has been established, the next phase involves generating specific recommendations. We assume that each of the influential (active) vertices has already formed an opinion/made a decision in the previous phase. Thus, contrary to collaborative filtering recommender systems, where recommendations are generated for all the users, the focus of this phase is solely on the inactive users. Depending on the input items and the objective of the recommendation process, it can be modelled as a classification or a ranking/preference aggregation problem.

The second phase is depicted in Figure 2. Assuming that the user (shown enlarged) has been convinced by their neighboring influentials to make a decision/form an opinion in the previous phase (in this example, follow a STEM career), they now need to make a finer-grained decision (e.g. which STEM major to follow). In the case where the influentials hold preference lists, the recommendation needs to be the result of some sort of voting mechanism.

In the Section that follows, we define in more technical terms the problems of neighborhood formation and prefer-

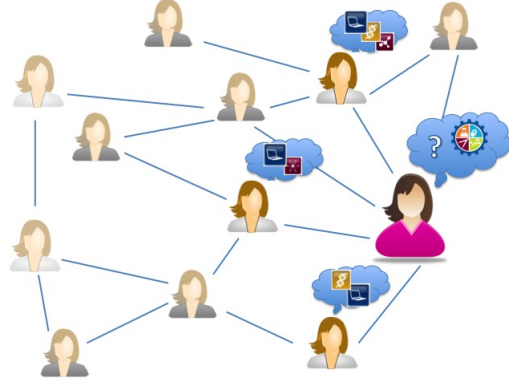


Fig. 2. Phase 2 - A voting mechanism used to generate recommendations.

ence aggregation. We propose an algorithm for a generalized dominating set problem in a social network that identifies influential vertices given specific constraints. The second module of the recommendation framework, involving the preference aggregation process, is out of the scope of this paper and will be discussed in detail in future work.

III. FORMAL PROBLEM DEFINITION

In the first phase of the neighborhood formations, we seek to find an appropriate subset of individuals that will act as influentials. A social network is modeled as a graph $G = (V, E)$, where V is the set of individuals and E is the set of their connections. We consider the model of influence without cascading. The problem that result from this model is a generalized versions of the classical minimum dominating set. A vertex is influenced if a portion of her immediate neighbors are influentials (active). The portion that each vertex v requires in order to get influenced is a threshold function $thr(v)$. We assume that the edges are undirected.

We define and solve a graph theoretical problem similar to the self-monopolies of [23], but the threshold is not fixed to 0.5. Rather than requiring the majority of neighbors to be active so that a vertex is influenced, in our model each vertex has its own threshold.

Before we move to the formal definitions, let us define some notation: let D be the set of influential vertices, $N(v)$ be the set of the neighboring vertices of vertex v , $degree(v)$ be the degree of v , $thr(v) \in [0, 1]$ be the threshold of v , W be the set of influenced vertices and A be all vertices that are neither influentials or influenced. Additionally, let $h(v)$ denote the number of neighbors of v that should be in D so that v is influenced:

$$h(v) = thr(v) \cdot degree(v).$$

The value of the domination of a node $v \in V$ by D is defined as:

- $value(D) = \min\{h(v), |N(v) \cap D|\}$, if $v \in V \setminus D$.
- $value(D) = h(u)$, if $v \in D$.

The function of the remaining cost $f(G, D)$ in a graph G with a set D is defined as: $f(G, D) = \sum_{v \in V} h(v) - \sum_{v \in V} \text{value}(v)$.

The remaining cost quantifies the difference between having all vertices either as an influencer or as influenced and the total value of all vertices with set D .

For a vertex v we say that:

- (a) v is an influential, if $v \in D$.
- (b) v is influenced, if $|D \cap N(v)| \geq \text{thr}(v) \cdot \text{degree}(v)$.

Recall, that the goal of the first problem is to select an appropriate subset of vertices as active so that all inactive vertices have a proportion of $\text{thr}(v)$ active neighbors. We are ready to define the first problem MIN-TBIDS as follows:

[PROBLEM:] THRESHOLD-BOUNDED INFLUENCE DOMINATING SET, MIN-TBIDS

Instance: $G = (V, E)$ is an undirected graph, and a threshold function $\text{thr} : V \rightarrow [0, 1]$.

Feasible solution: find a subset $D \subseteq V$ such that all $v \in V \setminus D$ are influenced, i.e., $W = V \setminus D$.

Goal: find a minimum D .

A. New Greedy Algorithm

The following Greedy algorithm that was proposed in [23] will be the basis of our new (enhanced) Greedy algorithm shown in Algorithm 1: we are looking for a minimum subset D of MIN-TBIDS, and let A denote the subset of vertices not influenced yet. Initially $D = \emptyset$ and $A = V$. While $W \neq \emptyset$, pick a vertex $v \in V \setminus D$ maximizing $|N(v) \cap A|$, add v to D and remove from A any vertex that is now influenced by D . Since the problem is submodular, it is easy to see, using an analysis similar for submodular problems that the resulting set D is at most $\ln|E| + 1$ times greater than the minimum one. In Algorithm 1, we add a preprocessing step to deal with all vertices of degree one. Figure 3 provides an instance of MIN-TBIDS with $\text{thr} = 0.5$, in which the new greedy algorithm makes a better choice with set $D = \{4, 7, 8, 15\}$ (colored red) than the greedy one that takes $D = \{2, 5, 7, 8, 15\}$ (colored red). Note that all remaining blue vertices have at least half of their neighbors red.

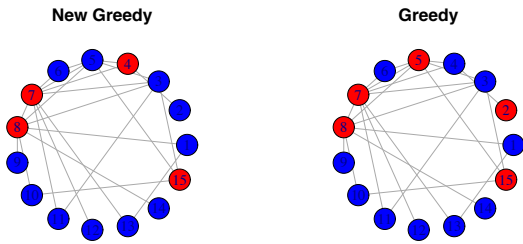


Fig. 3. Example of Greedy vs New Greedy algorithm for MIN-TBIDS with $\text{thr} = 0.5$.

The running time using Fibonacci heaps is $O(n \lg n + m)$. The new greedy algorithm is slightly slower in practice but with the same running time.

Algorithm 1 New Greedy for MIN-TBIDS.

Input: Graph $G = (V, E)$ and threshold $\text{thr} = \{0, \dots, 1\}$
Output: Subset D of V such that all vertices in $V \setminus D$ are influenced.

```

1: initialize  $D = \emptyset$ 
2: for each  $v \in V$  do
3:   compute degree  $d(v)$ 
4:    $dd(v) = d(v)$ 
5:    $\gamma(v) = 0$  ▷ proportion of neighbors in  $D$ 
6: end for
7: while  $\exists v \in V \setminus D: \gamma(v) < \text{thr}$  do
8:   if  $\exists v \in V \setminus D: d(v) = 1$  and  $\gamma(v) < \text{thr}$  then
9:     select  $s = \text{argmax}_v \{d(v) | v \in V \setminus D\}$ 
10:     $D = D \cup s$ 
11:     $dd(s) = 0$ 
12:     $\gamma(s) = 1$ 
13:   else
14:     select  $u = \text{argmax}_v \{dd(v) | v \in V \setminus D\}$ 
15:      $D = D \cup u$ 
16:     for each neighbor  $v$  of  $u$  and  $v \in V \setminus D$  do
17:        $dd(v) = dd(v) - 1$ 
18:        $\gamma(v) = \gamma(v) + 1/d(v)$ 
19:     end if
20:   end while
21: output  $D$ 

```

B. Recommendation Process

Once the neighborhood $N(v)$ of user v has been established, it can be used as input to the next step of the recommendation process. Our objective is to estimate the quantity $R(v, t_j)$ that represents the rating/rank/preference of user v for each item $t_j \in T$, where T represents the set of items/options available. We consider two variations of the same problem, depending on the desired outcome of the recommendation process.

1) *Recommendation of single item:* In the first case, we assume that each influential node contributes a singular opinion/vote (out of a selection of many). Referring to our previous examples, this could amount to each influential supporting “Yes” on a specific ballot measure, or selecting one particular major. In this setup, a simple majority rule of the user’s neighborhood can be held to generate the final recommendation. This process is very similar to K-NN classification, with the difference that similarity is not defined in the Euclidean space, and K is not common for all the vertices, but varies depending on each node’s threshold $\text{thr}(v)$ and number of influential neighbors. In this context, $R(v, t_j)$ is defined as follows:

$$R(v, t_j) = \frac{\sum_{v_i \in N(v) \cap D} \text{sim}(v, v_i) \cdot R(v_i, t_j)}{\sum_{v_i \in N(v) \cap D} \text{sim}(v, v_i)}$$

In other words, the preference score for each item t_j for the user v is the aggregation of the preference scores of the influentials in v ’s neighborhood, weighted by their similarity. In networks where the edges are all of the same strength,

$\text{sim}(v, v_i) = 1$. When the graph has weighted edges, in other words when we differentiate between strong and weak ties in the social network, the similarities can be updated accordingly. The outcome of this recommendation process is to recommend the item t_j that maximizes the preference score R for user v : $\max_j R(v, t_j)$.

2) *Recommendation of ranked list*: When we assume that each influential node maintains a preference list of items, the objective of the recommendation process is to generate a preference (i.e. ranked) list as recommendation to each user. Referring back to our examples, this would be a ranked list of majors or a ranked list of local government officials. In this scenario, there is no natural way to average the preferences of the neighborhood influentials as the input cannot be mathematically averaged. Therefore a more elaborate voting process needs to be designed. Voting is particularly useful when the participating influentials disagree because of genuine divergence of their subjective evaluations [7].

More formally, the outcome of this recommendation process for user v , is a personalized preference list $P_v(T)$ of the set T , where $t_j \succ_v t_k$ signifies that alternative $t_j \in T$ comes before $t_k \in T$ in v 's ranked list.

Calculating the preference list $P_v(T)$ is not a straightforward procedure. A naive approach would be to rely again on simple majority ruling, where the preference scores $R(v, t_j)$ are calculated using only the first preference in each influential neighbor's list. Then the preference list is created such that $t_j \succ_v t_k$ if $R(v, t_j) > R(v, t_k)$ (weak preferences can be defined in a similar way). However, this approach disregards the subsequent preference order of the neighboring nodes, and might rank higher items that would be in lower ranks if they were taken into consideration.

In general, when ranked lists are provided, voting systems based on majority rule suffer from several pathologies, as identified by the Condorcet Paradox, as they become susceptible to strategic agenda-setting [7]. Positional voting (such as the Borda Count) could also be considered but these also present their own pathologies, with many plurality voting election systems, including that of some U.S. states, demonstrating them. When weights are introduced to the edges of the graph, the problem becomes even more complicated and alternative voting systems and preference aggregation mechanisms must be devised. Further analysis of this problem is out of the scope of this paper and part of our future work.

IV. EXPERIMENTAL EVALUATION

The model of neighbourhood formation that was previously presented is portraying a distinct situation. This section presents the results on both synthetic and real-world data sets of this model. The objective is to assess the best approach, given different conditions and under different circumstances, in order to solve the problem presented in Section III, which refers to the first phase of the proposed social recommendation framework: the identification of influential vertices.

The MIN-TBIDS problem is a minimization problem, where the objective is to find the minimum set D of vertices

in order to influence all the remaining vertices in the graph.

To solve the problem we will base our experiments in two algorithms. The first is the greedy algorithm proposed in [23]. The second is the new greedy algorithm presented in Section III-A. The following sections detail the experiments and the evaluation of each set of used data.

A. Synthetic Data

Given the examples which serve as a motivation for this paper, we designed a set of experiments based on synthetic data in order to evaluate the algorithms used to solve the problems set forth. This decision lies on the ability of enabling a more controlled setting of the constraints and therefore a more objective study of the performance of each algorithm.

A significant amount of observed networks are usually categorized as scale-free networks, which relates to the fact they possess a power-law (or scale-free) degree distribution. These types of networks are associated with examples of Internet topology, neural networks and even the Web [22]. Small-world networks [30], which are characterized by their small diameter and high clustering coefficient, are usually observed in scientific collaboration networks. These types of networks are also used to characterize social networks. As such, we based our synthetic evaluation on the generation of multiple networks of three different types: 1) Barabasi-Albert, 2) Small-world and 3) degree sequence networks. The latter is a power-law network, characterized by its monotonic non-increasing sequence of the vertex degrees.

For each of these types of networks, 250 examples were generated with different sizes: for each size $s \in (100, 200, 300, 400, 500)$ we generated 50 undirected networks. As for parameters, in the case of the Barabasi networks, we defined a power of preferential attachment of 1.5. In small-world networks we set nei , the neighborhood within which the vertices of the lattice will be connected to 2 and the rewiring probability p as 0.3. Finally, concerning the degree sequence networks, we used a sequence of degrees following an exponential function of power $-0.5 * N$, where N is the sequence of the number of nodes n , $N = (1, \dots, n)$.

We evaluate the application of both algorithm as to the proportion of the number of vertices chosen to solve the problem. Regarding the constraints we defined a set of distinct thresholds (0.1, 0.2, 0.25, 0.33, 0.5) in order to evaluate the performance of each algorithm in different constraint scenarios.

The results presented in Figure 4 show that our proposed algorithm provides some advantage to the standard algorithm proposed in [23]. In the small-world networks most of the results show a tie or a residual improvement by our proposal. However, in scale-free networks such as the Barabasi-Albert and degree sequence our proposal shows a clear improvement.

Reporting to the evaluation concerning the scale-free networks, results show that in the MIN-TBIDS model, the algorithms approximate their results as the threshold grows.

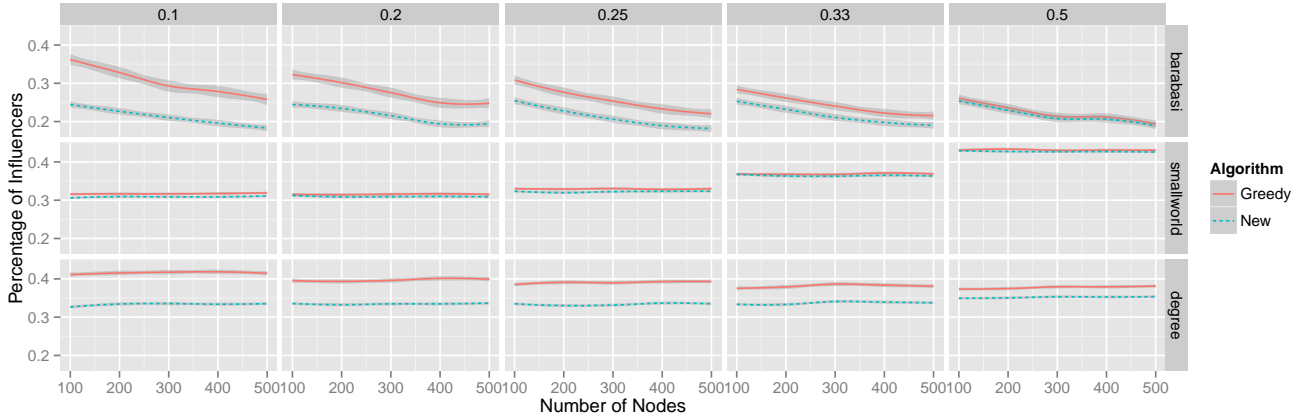


Fig. 4. Threshold-bounded Influence Dominating Set (TBIDS) with Synthetic Graphs. The top values of the grid represent threshold values.

B. Real-World Data

In addition to the synthetic data described in the former section, we evaluated the application of the greedy algorithms with real-world data sets. Our objective is the same: to evaluate the performance of the algorithms given the model of neighborhood formation.

TABLE I
REAL-WORLD DATA SETS

Dataset	AstroPhy	CondMat	DBLP
# Nodes	18772	23133	317080
# Edges	198050	93439	1049866
Avg. Degree	21.10	8.08	6.62
Max Degree	504	279	343

We used three datasets, described in Table I: the Astro Physics and Condense Matter collaboration networks [17] and the DBLP collaboration network [32]. For each of these data sets we induced subgraphs by finding a set of vertices of a size no bigger than 600 and no smaller than 100, through the following procedure: beginning with the addition of a random vertex of the graph, we added its neighbours; if necessary the neighbours of the neighbours were also included.

Concerning constraints we used the same as in the former evaluation of synthetic data sets: a set of thresholds (0.1, 0.2, 0.25, 0.33, 0.5) in the first problem; this same set of values in the second problem to represent the size of the partial sets of vertices in the network that we wish to influence; and 5 budgets k , where $k \in (0.05, 0.1, 0.15, 0.2, 0.25) * n$, where n is the size of the graph. Also, for the second and third problem the threshold for all vertices was set at 0.1. Results are presented in Figure 5.

In comparison to the previous evaluation, the evaluation of real-world data sets show that, overall, our approach provides an advantage, while it is not of the magnitude observed in the former results. This evaluation shows that, besides a residual gain in the model MIN-TBIDS, when the threshold is 0.1, the application of both algorithms shows a very similar outcome.

V. RELATED WORK

A. Identification of influential users

The problem of identifying influentials has gained a lot of attention from several research communities as it has applications in viral marketing [16], [3], [31], disease prevention [14] and propagation [24], politics [11], [6], etc.

Identifying the most influential users of a market in order to propagate influence, was first studied as an algorithmic problem by Domingos and Richardson [5]. In their work they apply data mining techniques to viral marketing, by modeling markets as social networks. They study the spread of influence using probabilistic models of interactions. Each vertex is associated with a value that quantifies how much she can influence other vertices and is used to optimally determine which vertices to choose as influentials. In their empirical study, using the EachMovie database, their proposed market strategy performs much better than two simple existing strategies.

Later, Kempe, et al. [13] formulate that problem as an maximization problem and propose three models of propagation. They prove that the problem is NP-hard and design a greedy approximation algorithm based upon submodular maximization with an $(1 - 1/e)$ -approximation algorithm. Consecutive work is focused on proposing optimizations of the greedy algorithm for better efficiency, see [18], [1]. In our model we don't consider cascading, the influence is in one round and each influencer can influence only vertices in the immediate neighborhood, i.e., one hop neighbors.

Motivated by failures in distributed systems Peleg [23] uses the idea of majority ruling to study problems related to discovering an optimal subset of controlling vertices, called coalition. A vertex v is controlled if the majority of its neighbors belong to the coalition. The two special types of coalitions are considered: the monopoly, if it controls every vertex in the graph and the self-ignoring monopoly M , if it controls only every vertex in $V \setminus M$. Problems regarding the amount of vertices that can be controlled, given the size of the coalition and how small a monopoly can be are investigated. The self-ignoring monopoly problem

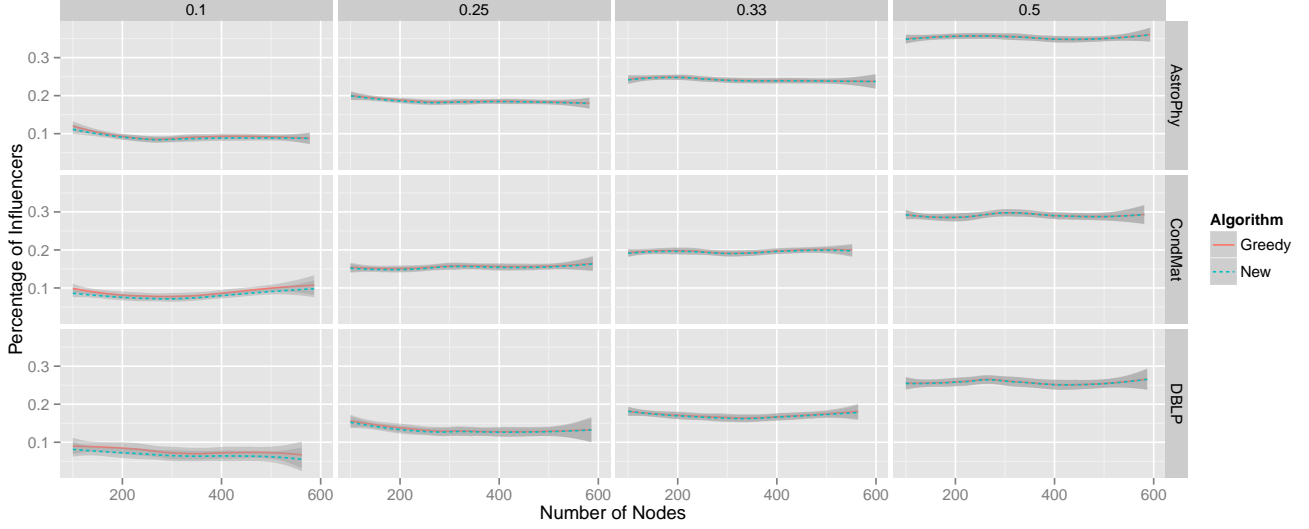


Fig. 5. Threshold-bounded Influence Dominating Set (TBIDS) with Real-World Data Sets. The top values of the grid represent threshold values.

is NP-hard and since it is submodular, the simple greedy algorithm yields to $(\ln|E|) + 1$ -approximation ratio. This work focuses in providing theoretical lower bounds for the general problems as well as for specific instances. The self-ignoring monopoly is close to our model and for that type no lower bound is given, when a vertex can control only its immediate neighbors. We extend this problem with different thresholds and enhance the greedy algorithm proposed there and through experimental results manage to outperform it. Chen [2] shows that MIN-TBIDS is hard to approximate within a polylogarithmic factor, even for some special cases such as bounded-degree graphs or majority thresholds. However, for trees this problem is in P .

The k -dominating set is a generalization of the classical dominating set. In this problem each vertex not in the k -dominating set has at least k neighbors in that set. Compared to our problem this is a special case with the threshold of every vertex set to $\lceil k/\text{degree}(v) \rceil$. An efficient algorithm with a $(1.7 + \lg \Delta)$ -approximation ratio, where Δ is the maximum degree, is given in [8].

In Wang [29], the Positive Influence Dominating Set (PIDS) problem is introduced in which a minimum subset D of vertices is sought so that every vertex (even in D) has at least half of its neighbors in D . They propose and algorithm which iteratively adds classic dominating sets with an $H(d)$, where H is the harmonic function and d is the maximum vertex degree of the graph. Additionally, they prove that the problem is APX hard. For power law graphs the approximation factor is a constant [33].

B. Social recommender systems

Social recommender systems have gained a lot of attention from the research in an effort to leverage social relationships to improve the recommendation process. This line of work is based on the assumption that users' preferences are influenced more by these of their connected friends, than these

of unknown users [31], rooted in the sociology concepts of homophily and social influence [21]. Tang et al. [28] give a narrow definition of social recommendation as "any recommendation with online social relations as an additional input, i.e., augmenting an existing recommendation engine with additional social signals" (a broader definition, not applicable to this work, refers to recommender systems targeting social media domains [9]).

The various proposed approaches can be categorized depending on the type of social relationship (trust, friendship etc.), the type of the underlying recommendation algorithm (model-based, memory-based, etc.), and the level of integration of the social information in the recommendation process. A common approach is to enhance the memory-based collaborative filtering process by forming the user's neighborhood using similarities deriving from the users' ratings and/or their social relationships, focusing on trust. An alternative line of work involves ways to enhance model-based recommender systems with social connections, again most often expressed as trust. This can be done through co-factorization, where the assumption is that the users share the same preference vector in both the rating and the social spaces (e.g. [26]), ensemble methods, where the resulting recommendation is derived by the linear combination of two systems (e.g. [19], [27]), or regularization, where priority is given to the social-based ratings (e.g. [12], [20]).

Most of the work in social recommender systems assumes some form of influence/trust propagation. Moreover, these are attempts to enhance the typical recommendation process with social data, assuming that item ratings are also available. In our work we assume that limited time that prevents influence propagation to affect the recommendation process and that explicit ratings are not available.

VI. CONCLUSIONS AND FUTURE WORK

In this work we present a graph algorithm based on threshold-bounded dominating sets and employ it to identify influential individuals in a social network. This process consists the neighborhood formation phase of a social recommender system, that addresses applications where influence cannot propagate and there are no explicit item ratings. The experimental evaluation of the proposed algorithm, considering our model, and using both synthetic and real-world datasets, have shown that our approach outperforms the previously proposed algorithm in most cases. As part of our future work we plan to design more efficient algorithms, explore the effect of weak and strong ties in the social network, and develop a preference mechanism for the recommendation generation phase.

REFERENCES

- [1] C. Borgs, M. Brautbar, J. T. Chayes, and B. Lucier. Maximizing social influence in nearly optimal time. In *Proc. of SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 946–957, 2014.
- [2] N. Chen. On the approximability of influence in social networks. *SIAM J. Discrete Math.*, 23(3):1400–1415, 2009.
- [3] S. Chen, J. Fan, G. Li, J. Feng, K.-I. Tan, and J. Tang. Online topic-aware influence maximization. *Proc. VLDB Endow.*, 8(6):666–677, Feb. 2015.
- [4] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *Proc. of KDD’09, KDD ’09*, pages 199–208. ACM, 2009.
- [5] P. M. Domingos and M. Richardson. Mining the network value of customers. In *Proc. of KDD’01*, pages 57–66, 2001.
- [6] E. Dubois and D. Gaffney. The multiple facets of influence: Identifying political influentials and opinion leaders on twitter. *American Behavioral Scientist*, 58(10):1260–1277, 2014.
- [7] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge Press, 2010.
- [8] K. Förster. Approximating fault-tolerant domination in general graphs. In *Proc. of ANALCO*, 2013.
- [9] I. Guy and D. Carmel. Social recommender systems. In *Proc. of WWW’11*, pages 283–284. ACM, 2011.
- [10] S. T. Hedetniemi and R. C. Laskar. Bibliography on domination in graphs and some basic definitions of domination parameters. *Discrete Mathematics*, 86(1-3):257–277, 1990.
- [11] R. Huckfeldt and J. Sprague. Networks in context: The social flow of political information. *American Political Science Review*, 81:1197–1216, 12 1987.
- [12] M. Jamali and M. Ester. A matrix factorization technique with trust propagation for recommendation in social networks. In *Proc. of RecSys’10*, pages 135–142. ACM, 2010.
- [13] D. Kempe, J. Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4):105–147, 2015.
- [14] D. A. Kim, A. R. Hwang, D. Stafford, D. A. Hughes, A. J. O’Malley, J. H. Fowler, and N. A. Christakis. Social network targeting to maximise population behaviour change: a cluster randomised controlled trial. *The Lancet*, 386:145–153, 5 2015.
- [15] G. Kossinets, J. Kleinberg, and D. Watts. The structure of information pathways in a social communication network. In *Proc. of KDD’08*, pages 435–443. ACM, 2008.
- [16] S. Lei, S. Maniu, L. Mo, R. Cheng, and P. Senellart. Online influence maximization. In *Proc. of KDD’15*, 2015.
- [17] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graph evolution: Densefication and shrinking diameters. *ACM Trans. Knowl. Discov. Data*, 1(1), Mar. 2007.
- [18] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In *Proc. of KDD’07*, pages 420–429, 2007.
- [19] H. Ma, I. King, and M. R. Lyu. Learning to recommend with social trust ensemble. In *Proc. of SIGIR ’09*, pages 203–210. ACM, 2009.
- [20] H. Ma, D. Zhou, C. Liu, M. R. Lyu, and I. King. Recommender systems with social regularization. In *Proc. of WSDM ’11*, pages 287–296. ACM, 2011.
- [21] J. M. C. Miller McPherson, Lynn Smith-Lovin. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27:415–444, 2001.
- [22] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proc. of IMC ’07*, pages 29–42. ACM, 2007.
- [23] D. Peleg. Local majorities, coalitions and monopolies in graphs: a review. *Theor. Comput. Sci.*, 282(2):231–257, 2002.
- [24] M.-E. G. Rossi, F. D. Malliaros, and M. Vazirgiannis. Spread it good, spread it fast: Identification of influential nodes in social networks. In *Proc. of WWW’15 Companion*, pages 101–102. ACM, 2015.
- [25] C. R. Shalizi and A. C. Thomas. Homophily and contagion are generically confounded in observational social network studies. *Sociological Methods Research*, 40(2):211–239, 5 2011.
- [26] J. Tang, H. Gao, X. Hu, and H. Liu. Exploiting homophily effect for trust prediction. In *Proc. of WSDM ’13*, pages 53–62. ACM, 2013.
- [27] J. Tang, H. Gao, and H. Liu. mtrust: Discerning multi-faceted trust in a connected world. In *Proc. of WSDM ’12*, pages 93–102. ACM, 2012.
- [28] J. Tang, X. Hu, and H. Liu. Social recommendation: a review. *Social Netw. Analys. Mining*, 3(4):1113–1133, 2013.
- [29] F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan. On positive influence dominating sets in social networks. *Theor. Comput. Sci.*, 412(3):265–269, 2011.
- [30] D. J. Watts and S. H. Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):409–10, 1998.
- [31] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twittrank: Finding topic-sensitive influential twitterers. In *Proc. of WSDM’10*, pages 261–270, 2010.
- [32] J. Yang and J. Leskovec. Defining and evaluating network communities based on ground-truth. In *Proc. of MDS ’12*, pages 3:1–3:8. ACM, 2012.
- [33] W. Zhang, W. Wu, F. Wang, and K. Xu. Positive influence dominating sets in power-law graphs. *Social Netw. Analys. Mining*, 2(1):31–37, 2012.