Discrete Optimization 8 (2011) 87–96

Contents lists available at ScienceDirect

Discrete Optimization

journal homepage: www.elsevier.com/locate/disopt

Treewidth governs the complexity of target set selection ☆

Oren Ben-Zwi [a,*], Danny Hermelin [b], Daniel Lokshtanov [c], Ilan Newman [a]

[a] Department of Computer Science, University of Haifa, Haifa 31905, Israel

[b] Max-Planck-Institute for Informatics, 66123 Saarbru"cken, Germany

[c] Institutt for Informatikk, University of Bergen, PB 7803, N-5020 Bergen, Norway

ARTICLE INFO

Article history:
Received 16 November 2009
Received in revised form 14 September 2010
Accepted 20 September 2010
Available online 8 October 2010

Keywords:
Target set selection
Bounded tree-width algorithm
Bounded tree-width lower-bound
Viral marketing
Social networks
Game of life

ABSTRACT

In this paper we study the Target Set Selection problem proposed by Kempe, Kleinberg, and Tardos; a problem which gives a nice clean combinatorial formulation for many applications arising in economy, sociology, and medicine. Its input is a graph with vertex thresholds, the *social network*, and the goal is to find a subset of vertices, the *target set*, that "activates" a pre-specified number of vertices in the graph. Activation of a vertex is defined via a so-called activation process as follows: Initially, all vertices in the target set become active. Then at each step i of the process, each vertex gets activated if the number of its active neighbors at iteration $i-1$ exceeds its threshold. The activation process is "monotone" in the sense that once a vertex is activated, it remains active for the entire process.

Our contribution is as follows: First, we present an algorithm for Target Set Selection running in $n^{O(w)}$ time, for graphs with n vertices and treewidth bounded by w. This algorithm can be adopted to much more general settings, including the case of directed graphs, weighted edges, and weighted vertices. On the other hand, we also show that it is highly unlikely to find an $n^{o(\sqrt{w})}$ time algorithm for Target Set Selection, as this would imply a sub-exponential algorithm for all problems in SNP. Together with our upper bound result, this shows that the treewidth parameter determines the complexity of Target Set Selection to a large extent, and should be taken into consideration when tackling this problem in any scenario. In the last part of the paper we also deal with the "non-monotone" variant of Target Set Selection, and show that this problem becomes #P-hard on graphs with edge weights.

1. Introduction

Consider the following scenario: You are a marketing executive of a huge clothing company given the task of marketing a new line of summer wear. You have at hand a description of the relationship network formed among a sample of teenagers from the district. After some heavy thinking you come up with the following idea: You will identify, or *target*, key social figures of the network and persuade them into adopting the new summer line, by say, handing out substantial amounts of free samples. You then hope that by peer-pressure laws, the friends of those targeted individuals would be persuaded into buying the new products, which in turn will also cause their friends to be persuaded, and so forth, creating a domino-like effect in the network. But how do you find a good set of individuals to target?

☆ This work appeared in *Proceedings of the 10th ACM Conference on Electronic Commerce (EC-2009)*, Stanford, California, USA, July 6–10, 355–362, 2009.

* Corresponding author. Tel.: +972 48288301.

E-mail addresses: nbenzv03@cs.haifa.ac.il (O. Ben-Zwi), hermelin@mpi-inf.mpg.de (D. Hermelin), daniello@ii.uib.no (D. Lokshtanov), ilan@cs.haifa.ac.il (I. Newman).

1572-5286/$ – see front matter
doi:10.1016/j.disopt.2010.09.007

Research in the area of *viral marketing* [1–3] studies questions similar to the one raised above. The key objects under research are *social networks* which are often modeled by graphs with individuals or organizations as vertices, and relationships or interactions as edges. Social networks play a leading role in many scientific fields, including most social sciences [4–6], life sciences [7,8] and medicine [7,5,9]. In viral marketing, one attempts to take advantage of social network properties, in order to enhance revenue in various commercial applications. This is based on the premise that targeting a few key individuals may lead to strong word-of-mouth effects, which in turn, will cause a cascade of influence in the network. Viral marketing has recently become a widespread technique for promoting novel ideas, marketing new products, or spreading innovation [10,11]. Today, in the age of the Internet, the huge amount of available data poses new challenges for this area which are both daunting and extremely profitable at the same time. As an example, QZone, FaceBook, and MySpace, are just three of many social networking websites boasting more than five hundred million users world-wide (November 2009); endlessly engaged in the exchange of news, opinions, gossip, and almost any other thinkable type of information.

One simple way to model the cascade of influence in viral marketing is given by the *threshold model* [12]. The main idea is to associate with each vertex v of the network two states, *active* and *inactive*, which indicate whether v is persuaded into adopting the idea or product that is marketed. Moreover, v is also assigned a *threshold value* $t(v)$, specifying how many neighboring vertices of v need to get persuaded before v itself is persuaded. A cascade of influence, or *activation process*, proceeds in the network as follows: Initially, all vertices are inactive. In phase 0 of the process, we select k initial vertices, the *target set*, that instantly become active. Then, at every phase $i > 0$, a vertex v becomes active if at least $t(v)$ of its neighbors were active in phase $i - 1$. Once a vertex becomes active, it remains active for the entire process, and so in this sense the activation process is *monotone*. The process ends in phase $i_{\text{end}} < n$, where n is the number of vertices in the network, when no more vertices can get activated. Given the rules of this activation process, and knowledge of the thresholds in our network, which individuals should we target so as to persuade as many individuals in the network as possible?

The first to study this question from an algorithmic point of view were Kempe et al. in their seminal paper [13]. They investigated the following maximization problem: Given a social network G with vertex-thresholds, find a target set of size at most k that activates as many vertices in G as possible. This models the situation where there is a pre-specified budget for targeting. We note that Kempe et al. focused mostly on the case where the thresholds of the graph are random. This work was extended in [14,15]. Chen [16] studied the following related minimization problem: Given a social network G, find a target set of smallest possible size that activates at least ℓ vertices of G. This models the case where we have a minimum limit for the number of persuaded individuals overall. We reduce these two optimization problems to a single search problem, which is the main focus of this paper. We refer to this search problem throughout as the Target Set Selection problem.

Unsurprisingly perhaps, the decision version of Target Set Selection is NP-complete. More surprising is the fact that both of its optimization variants turn out to be extremely hard to approximate, even for very restrictive special cases. Kempe et al. show that the maximization problem they introduced cannot be approximated within any non-trivial factor, unless P = NP, even when the given social network is bipartite with bounded degree, and all vertices have equal thresholds [13]. Chen [16] shows a polylogarithmic approximation lower bound for the minimization problem described above, and his bound also holds for bounded degree bipartite graphs, even when the thresholds are taken from the set $\{1, 2\}$. Regarding the parameterized complexity status of the problem, Abrahamson, Downey, and Fellows show that both problems are W[P]-complete, *i.e.* fixed-parameter intractable, when parameterized by the size of the solution target set [17].

The high inapproximability and fixed-parameter intractability results for Target Set Selection mentioned above are a striking blow from the algorithm designer point of view. In light of these results, we must turn our consideration towards special cases of the problem, or otherwise resort to heuristic approaches. When considering special cases, it is desirable to obtain a robust algorithm that behaves relatively well also on more general cases. Furthermore, one must overcome the fact that the problem is already known to be hard for many restricted cases; in particular, for notoriously easy classes of graphs such as bounded degree graphs and bipartite graphs.

In this paper we tackle these difficulties by considering the *treewidth* parameter of graphs. This parameter plays an important role in the design of many exact and approximation algorithms for many NP-hard problems. The notion was introduced by Robertson and Seymour [18] in their celebrated proof of the Graph Minor Theorem. Roughly, it measures the extent a given graph is similar to a tree in a very deep structural sense. For instance, trees have treewidth 1. We will show that the treewidth parameter governs the complexity of the Target Set Selection problem in a very strict sense. The first clue for this was given by Chen [16] who showed that the problem is polynomial-time solvable in trees. We generalize this result substantially. Letting n and w respectively denote the number of vertices and treewidth of our input graph, we prove the following theorem:

Theorem 1.1. Target Set Selection *can be solved in* $n^{O(w)}$ *time.*

It is worth pointing out that the time complexity in the theorem above can be rewritten as $t^{O(w)}n$, where t is the maximum threshold of any vertex in the network. Thus, the algorithm used in proving this theorem also shows that the problem is fixed-parameter tractable when parameterizing by both the treewidth and maximum degree of the graph. Also, this algorithm can be adopted to the more general setting of directed graphs with edge influences and vertex weights.

On the other hand, we will show that we cannot do much better than Theorem 1.1. We prove that, under a well-established complexity-theoretic assumption, the above algorithm is optimal up to a quadratic factor in the exponent dependency on w. This shows that the treewidth of the given network indeed determines to a large extent whether one