Balancing the Spread of Two Opinions in Sparse Social Networks* (Student Abstract)

Dušan Knop, Šimon Schierreich, Ondřej Suchý

Department of Theoretical Computer Science, Faculty of Information Technology,
Czech Technical University in Prague, Prague, Czechia
{dusan.knop,schiesim,ondrej.suchy}@fit.cvut.cz

Abstract

We propose a new discrete model for simultaneously spreading two opinions within a social network inspired by the famous TARGET SET SELECTION problem. We are given a social network, a seed-set of agents for each opinion, and two thresholds per agent. The first threshold represents the willingness of an agent to adopt an opinion if she has no opinion at all, while the second threshold states the readiness to acquire a second opinion arriving. The goal is to add as few agents as possible to the initial seed-sets such that, once the process started with these seed-set stabilises, each agent has either both opinions or none.

We perform an initial study of its computational complexity. It is not surprising that the problem is NP-hard even in quite restricted settings. Therefore, we investigate the complexity of the problem from the parametrized point-of-view with special focus on sparse networks, which appears often in practice. Among other things, we show that the proposed problem is in FPT if we parametrize by the vertex cover number of the underlying graph.

Introduction

To study the influence of direct marketing in social networks, Domingos and Richardson (2001) introduced the TARGET SET SELECTION problem (TSS for short). We follow the threshold model proposed by Kempe, Kleinberg, and Tardos (2003). In the original TSS problem, we are given a social network G, which is an undirected graph, a threshold function $f\colon V(G) \to \mathbb{N}$, and a budget $\mathscr{b}$. Our goal is to decide whether there is a subset $T \subseteq V(G)$ such that if the following activation process

$$P^0 = T, \qquad P^{i+1} = P^i \cup \{v \mid f(v) \leq |N(v) \cap P_i|\},$$

where $N(v)$ is the set of neighbours of v, stabilizes in ℓ-th round, i.e, $P^\ell = P^{\ell+1}$, then $P^\ell = V(G)$.

We propose a novel model called 2-OPINION TARGET SET SELECTION where we extend the input of the problem by a second threshold function for every agent $v \in V(G)$, we introduce an *initial seed-sets* S_a and S_b for both opinions, and we change the activation process as follows.

$$\begin{aligned} P_a^0 &= S_a \cup T_a, \quad P_b^0 = S_b \cup T_b, \\ P_c^{i+1} &= \{v \in V(G) \setminus (P_a^i \cup P_b^i) \mid f_1(v) \leq |P_c^i \cap N(v)|\} \\ &\quad \cup \{v \in P_{\neg c}^i \mid f_2(v) \leq |P_c^i \cap N(v)|\} \cup P_c^i, \end{aligned}$$

where $c \in \{a, b\}$ and $\neg c \in (\{a, b\} \setminus c)$. In this setting, our goal is not to spread both opinions in the whole network, but to influence the process by selecting an appropriate $T_a, T_b \subseteq V(G)$, with $|T_a| + |T_B| \leq \mathscr{b}$, such that when the process stabilizes, every agent $v \in V(G)$ has either both opinions or none.

The TSS problem with two opinions was already studied by Garimella et al. (2017). In contrast to their model, we have no opinion-specific threshold, but we allow a certain level of interaction between the opinions. We believe that this is an interesting setting as it captures different agent's mindsets. For example, we are able to model the case where an agent v is tougher towards the second opinion (by setting, e.g., $f_2(v) = 5f_1(v)$), which can find an application in, e.g., modeling the following of political leaders. On the other hand, agents can infer the second opinion easily in many applications, e.g., when the first virus decreases the ability of an agent to resist a second virus. One might also ask why to balance the spread of two opinions already presented in a social network. Suppose that there are two experts having a strong opinion, for example, on COVID-19 vaccination. As a social network manager, we do not have enough expertise to recognise the truthful opinion. And if many agents receive only the opinion which later turns out to be deceptive, then these agents might feel deceived by the network. Thus, we can help both opinions to spread evenly.

Hardness Results

Our first result shows that the studied model is NP-hard and thus there is a need to study the problem from different perspectives to hope for tractable cases.

Theorem 1. *It is NP-hard to solve the* 2-OPINION TARGET SET SELECTION *problem.*

*The authors acknowledge the support of the OP VVV MEYS funded project CZ.02.1.01/0.0/0.0/16_019/0000765 "Research Center for Informatics" and the Grant Agency of the CTU in Prague funded grant No. SGS20/208/OHK3/3T/18.

To prove the theorem, we reduce from the original TSS. It is easy to verify that if we are given an instance $\mathcal{I} = (G, f, \flat)$ of the TSS problem, then instance $\tilde{\mathcal{I}} = (G, \emptyset, V(G), f_1, f, \flat)$, where $\forall v \in V(G)\colon f_1(v) = \infty$, is equivalent. The task is then reduced to select a minimal target set $T_a \subseteq V(G)$ such that the process spreads the opinion a to the whole network. Consequently, unless $\mathsf{P} \neq \mathsf{NP}$, there is no hope for a polynomial-time algorithm for the general 2OTSS. Thus, it is reasonable to study the problem under a different lens.

Parametrized Complexity

The traditional computational complexity is focused on the classification of computational problems into complexity classes based on the size of the input. For NP-hard problems, it means that they are considered intractable under reasonable theoretical assumptions. The parametrized complexity perspective (Cygan et al. 2015) is focused on the study of a problem's complexity with respect to not only the instance size, but it takes into account also an additional information about the input instance – a *parameter*.

Formally, for a fixed finite alphabet Σ, we define a *parametrized problem* as a language $L \subseteq \Sigma^* \times \mathbb{N}$. An instance is a pair (x, k), where x is the problem's *input* and $k \in \mathbb{N}$ is the value of a *parameter*. An instance is called a *yes*-instance if and only if $(x, k) \in L$.

Similarly to the classical complexity, we define complexity classes to distinguish "easy" and "hard" problems. The most favourable complexity class is FPT, which contains all problems solvable in $f(k) \cdot n^{\mathcal{O}(1)}$ time, where f is a computable function. In addition, the class $\mathsf{W[1]}$ contains both problems in FPT and problems believed not to be in FPT. For a more comprehensive introduction to the parametrized complexity, we refer the reader to Cygan et al. (2015).

Bounded Vertex Cover Networks

In this section, we show that the studied problem is in FPT with respect to the vertex cover number of the underlying graph. Let C be a minimum size vertex cover in G, i.e., a set such that $\forall \{u, v\} \in E(G)\colon u \in C \vee v \in C$, of size k. Our algorithm is based on multiple reduction rules, which are applied exhaustively in the order of occurrence.

First, if there is a vertex $v \in V(G)$ such that $v \in S_a \cap S_b$, then we remove v from the network and continue with the instance, where for each neighbour u of v we decrease the threshold values by one.

Next, let $v \in V(G)$ be a vertex such that either $v \in S_a$ or $v \in S_b$, and $f_2(v) > \deg(v)$. Since it is not possible to reach v with the second opinion, we add v to the seed-set for both opinions and decrease budget $\flat$ by one.

Then, we observe that there is no difference whether a threshold exceeds the degree of a vertex by one or by a million – in both cases the natural process is not capable to affect this vertex. Thus, for each vertex $v \in V(G)$ we set $f_i(v) = \min\{f_i(v), \deg(v) + 1\}$.

The final rule directly recognises big *yes*- and *no*-instances. It is easy to see that if $\flat < 0$, then we are struggling with *no*-instance. On the other hand, if $\flat \geq |C \setminus S_a| + |C \setminus S_b|$, then we output a trivial *yes*-instance. With the necessary preparation, we obtain our final result.

Theorem 2. *The* 2-OPINION TARGET SET SELECTION *is in FPT when parametrized by the minimum size k of a vertex cover C of the underlying graph.*

Proof. Let $(G, S_a, S_b, f_1, f_2, B)$ be an instance of 2OTSS. We assume the input instance to be reduced with respect to the presented reduction rules. If $\flat \geq 2k$, then the final rule applies. Hence, we assume $\flat < 2k$.

We define an equivalence relation $\sim$ on $V \setminus C$ such that for $u, v \in V \setminus C$ we have $u \sim v$ if and only if $N(u) = N(v)$, and for $i = 1, 2$ it holds that $f_i(u) = f_i(v)$. The cardinality of the quotient set $(V \setminus C)/\sim$ is at most $2^k \cdot (k+1)^2$ and together with the vertices of vertex cover C we have $k + 2^k \cdot (k+1)^2$ different kinds of vertices from which we have to select at most $\min\{B, k\}$ vertices into the seed set T_a and at most $\min\{B, k\}$ into T_b. I.e., at most $2k$ vertices with at most $k + 2^k \cdot (k+1)^2 + 1 \leq 2^k \cdot (k+2)^2$ options each which gives us at most $\left(2^k (k+2)^2\right)^{2k} = 2^{2k^2 + \mathcal{O}(k \log k)}$ options in total that can all be checked in $2^{2k^2 + \mathcal{O}(k \log k)} \cdot n^2$ time. Thus, 2OTSS is in FPT when parametrized by the vertex cover number. $\square$

Conclusions and Future Work

In addition to Theorem 2 presented here, we proved that the 2OTSS is $\mathsf{W[1]}$-hard with respect to the most natural parameter – the sum of sizes of the seed-sets. The problem remains $\mathsf{W[1]}$-hard even if we combine the size of the seed-sets with the tree-width of the underlying graph .

On the positive side, we showed that 2OTSS is in FPT if we parametrize with the number of rounds, maximum threshold, and the tree-width of the underlying graph. Moreover, the same algorithm also applies for the combined parameter tree-depth and maximum threshold. Last but not least, we proved that the problem is in FPT for the parameter 3-path vertex-cover. It is worth mentioning that our algorithm based on N-fold integer programming applies also to TSS and is the fastest currently known.

Identifying and studying important special cases of 2OTSS, such as the majority thresholds version, or other structural limitations, especially those that are bounded in sparse graphs, should be of interest for future research.

References

Cygan, M.; Fomin, F. V.; Kowalik, Ł.; Lokshtanov, D.; Marx, D.; Pilipczuk, M.; Pilipczuk, M.; and Saurabh, S. 2015. *Parameterized Algorithms*. Springer. ISBN 978-3-319-21274-6.

Domingos, P.; and Richardson, M. 2001. Mining the Network Value of Customers. KDD '01, 57–66. New York, NY, USA: ACM. ISBN 978-1-58113-391-2.

Garimella, K.; Gionis, A.; Parotsidis, N.; and Tatti, N. 2017. Balancing information exposure in social networks. In *NIPS '17*, 4663–4671. Curran Associates, Inc.

Kempe, D.; Kleinberg, J.; and Tardos, E. 2003. Maximizing the Spread of Influence through a Social Network. KDD '03, 137–146. ACM. ISBN 978-1-58113-737-8.