Threshold Models for Competitive Influence in Social Networks

Allan Borodin, Yuval Filmus, and Joel Oren

Department of Computer Science, University of Toronto, Canada.
bor,yuvalf,oren@cs.toronto.edu

Abstract. The problem of influence maximization deals with choosing the optimal set of nodes in a social network so as to maximize the resulting spread of a technology (opinion, product-ownership, etc.), given a model of diffusion of influence in a network. A natural extension is a competitive setting, in which the goal is to maximize the spread of our technology in the presence of one or more competitors.
We suggest several natural extensions to the well-studied linear-threshold model, showing that the original greedy approach cannot be used.
Furthermore, we show that for a broad family of competitive influence models, it is NP-hard to achieve an approximation that is better than a square root of the optimal solution; the same proof can also be applied to give a negative result for a conjecture in [2] about a general cascade model for competitive diffusion.
Finally, we suggest a natural model that is amenable to the greedy approach.

1 Introduction

The problem of influence maximization has long been the focus of study in social science (e.g. [5]). It has been formally defined and addressed in [6, 3] as follows: given a social-network as a directed graph, and a prescribed number k, pick the k most "influential" nodes that will function as early adopters of a particular influence, so as to maximize the final number number of *infected*, or *activated* nodes (the two terms are used in this paper interchangeably), subject to a specified model of influence diffusion.

This problem begs the natural extension of a competitive version: given the competitor's choice of early adopters of technology B, maximize the spread of technology A by choosing a set of early adopters such that the expected spread of technology A will be maximal. Indeed, this problem has been the subject of interest in subsequent papers ([2, 7, 1]) which present competitive extensions for the independent cascade model presented in [6].

In this paper we suggest several natural extensions to a well-studied model that was also given in [6] for the diffusion of social influence in a social network — the linear threshold model. Formally, an instance of the problem would be composed of a directed, edge-weighted graph $G = (V, E)$, a set of technology B's initial adopters $I_B \subseteq V$, and an integer k. The computational problem is

how to choose a set $I_A \subseteq V - I_B$ of k nodes such that the *expected* number of A-active nodes at the end of the process, $\sigma(I_A, I_B)$, is maximized, given the specific model for competitive diffusion of technologies (when I_B is known from context we omit it and simply write $\sigma(I_A)$). It is important to note that all of the presented models can be motivated by natural processes. Our models reduce to the original linear threshold model formalized in [6] whenever I_B is the empty set. For simplicity of notation, the models presented are stated in terms of only two competing technologies. However, all of the models and results can be easily extended for when there are several competing technologies.

A well-known greedy $(1 - e^{-1})$-approximation given in [4] is used extensively for problems of maximizing set-functions, and in particular has been applied to both the original problem and competitive extensions ([2, 1]). The approximation algorithm requires that the set function $\sigma(\cdot)$ at hand, which assigns a real-value to subsets of a ground set U, uphold two basic properties.

- **Monotonicity:** the value of the function increases as more items are added to the set: $\sigma(S) \leq \sigma(T)$ for any two sets $S \subseteq T$;
- **Submodularity:** the impact of adding an element to a set decreases as the set is extended (diminishing returns): $\sigma(S \cup \{x\}) - \sigma(S) \geq \sigma(T \cup \{x\}) - \sigma(T)$, for any $S \subseteq T \subseteq U$ and $x \in U - T$;

Except for the last model, described in section 6 — the *OR* model, all of the models do not satisfy submodularity. In fact, one of them is not even monotone.

Outline The remainder of this document is organized as follows. Sections 2 and 3 describe two competitive threshold models. Section 4 shows that even when applying a final step that A-activates more nodes, the process remains non-submodular. Section 5 shows that the last two models are in general hard to approximate. On a more positive note, in section 6 we suggest a fairly natural and simple model for which the approximation algorithm given in [4] is applicable. Finally, section 7 summarizes our main results along with a few open problems and possible directions for future research.

2 The Weight-Proportional Competitive Linear Threshold Model

As in the non-competitive case, the process unfolds in discrete steps, during which new nodes become "activated" for a single technology[1] The infection of a node by a technology represents an individual in the social network that has assumed the influence of that technology. The process is *progressive*: a node that is infected by a technology remains infected by it. As in the non-competitive case, every edge (u, v) is assigned a weight $w_{u,v} \in [0, 1]$ which roughly characterizes

[1] The term "technology" stands for any concept or influence that spreads in the social-network (car ownership, club membership, voting preference, etc.).

the weight of influence that u has over v (i.e the impact that u's infection will have over v's likelihood to be infected with the same technology as u). Also, the total weight of incoming edges to every node is bounded: for every $v \in V$ we have: $\sum_u w_{u,v} \in [0,1]$. Each node u initially chooses a threshold θ_u which represents the minimum fraction of active neighbours necessary for u's activation. As in [6], in order to make up for our lack of knowledge about each node we assume that $\theta_u \in_R [0,1]$ (uniformly at random), or $\theta_u \in_R [a, a']$ for $0 \leq a \leq a' \leq 1$ to reflect partial knowledge about a node.

In order to describe the process itself, we will use the following notation:

Definition 1. *For a given step t in the process, let Φ^t denote the set of active nodes at the beginning of step t. Furthermore, let Φ^t_A and Φ^t_B be the sets of A-active and B-active nodes in step t, respectively.*

Given the sets I_A, I_B of early technology adopters, the process unfolds as follows. First, each node chooses its threshold value at step 0. Then, in each step t, every inactive node v checks the set of edges incoming from its active neighbours. If their collective weight exceeds the threshold values by having $\sum_{u \in \Phi^t} w_{u,v} \geq \theta_v$, the node becomes active. In that case, the node will adopt technology A with probability equal to the ratio between the collective weight of edges outgoing from A-active neighbours and the total collective weight of edges outgoing from all active neighbours; that is,

$$Pr[v \in \Phi^t_A | v \in \Phi^t \backslash \Phi^{t-1}] = \frac{\sum_{u \in \Phi^t_A} w_{u,v}}{\sum_{u \in \Phi^t} w_{u,v}} \tag{1}$$

Otherwise, it will adopt technology B. Since this problem can be reduced to the single-technology linear threshold model whenever I_B is the empty set, we notice that this problem is NP-hard — as proved in [6].

Intuitively, it appears that by adding a new node to the set of initial A-adopters, the spread of technology A in the social network can only increase (or remain unchanged). However, this is in fact not always the case, even for some binary rooted trees. We will formalize this somewhat counter-intuitive behaviour.

Theorem 1. *There exists an instance of the weight-proportional competitive linear threshold problem for which monotonicity does not hold.*

Also, it can be shown that submodularity fails to hold in some cases, as the following theorem shows:

Theorem 2. *There exists a graph G, for which the expected influence of technology A is not submodular.*

The proof of the above two theorems is given in appendix A.

3 The Separated-Threshold Model for Competing Technologies

In the previous model, the node activation step regarded active nodes as equal, so that a given node is activated by its active neighbours regardless of their technologies. That is, the sum of generally active nodes was used for activating a node. However, one could model the following notion of a spread process. Each individual has separate criteria for becoming active for each technology. A node can be activated by either its A-active or B-active neighbours whenever the sums of their respective edge-weights exceed the required thresholds specified for their technologies.

Formally, consider the following model. For a given network $G = (V, E)$, every edge $(u, v) \in E$ is assigned a real-valued weight corresponding to each technology $w^A_{u,v}, w^B_{u,v} \in [0, 1]$ such that $\sum_u w^A_{u,v}, \sum_u w^B_{u,v} \in [0, 1]$, which reflects node u's impact on v. Two *disjoint* sets $I^0_A, I^0_B \subseteq V$ denote the sets of initially A-active and B-active nodes, respectively. At step 0, each node $v \in V$ picks two threshold values $\theta^A_v, \theta^B_v \in_R [0, 1]$. For step t, denote I^{t-1}_A, I^{t-1}_B as the sets of A-active and B-active nodes. During every step t, an inactive node v will become A-active if $\sum_{u \in I^{t-1}_A} w^A_{u,v} \geq \theta^A_v$, and will become B-active if $\sum_{u \in I^{t-1}_B} w^B_{u,v} \geq \theta^B_v$. If for the node v both thresholds are exceeded during the same step t, then a technology would be picked uniformly at random (we can either use a simple coin-flip or employ a more involved tie-breaking function).

In contrast to the previous model, this model *is* monotone. Its key property, which distinguishes it from the previous model, is that the probability that technology B will propagate cannot increase as a result of A-activating additional nodes. This stems from the definition of the model, in which each set of technology specific neighbours relate to a separate threshold value.

Let us use the following notation:

Definition 2. *Given the sets I_A and I_B, and a node $x \notin I_B$, define $\alpha^t_v, \hat{\alpha}^t_v$ as the probabilities that a given node v will adopt technology A by step t for the initial sets of early adopters (I_A, I_B) and $(I_A \cup \{x\}, I_B)$, respectively. Similarly, define similar probabilities $\beta^t_v, \hat{\beta}^t_v$ for technology B.*

Theorem 3. *For a given instance of the problem and a choice of early adopters: I_A, I_B and node x, $\hat{\alpha}^t_v \geq \alpha^t_v$ for any node v and for any step $t \geq 0$.*

The proof of theorem 3 is fairly straightforward, and is given in appendix B for completeness. The process is not submodular in general.

Theorem 4. *There exist instances of the competitive influence problem where the separated-threshold competitive model is not submodular.*

A corresponding counter-example for this theorem is fairly easy to construct. It appears in the full version of this paper located on the authors' personal websites.

4 Competitive Threshold Model with Forcing

We now introduce a modification which changes the concept of influence in a network: *forcing*. Specifically, at the end of the previous model, each inactive node v will choose a technology randomly (say, it will choose technology A with probability δ). This step is natural for cases where individuals have to eventually decide which influence to adopt (e.g. voting when abstentions are not allowed). For convenience we will assume that the "forcing" step occurs at step n (the spread can take up to $n-1$ steps), whether or not the spread took $n-1$ steps. Clearly this does not have any effect on the outcome of the process. We show that regardless of the forcing step, this variant does not help us achieve submodularity. However, the process remains monotone as the following theorem can be proven by extending lemma 3 (in appendix B) with an additional case analysis for the forcing step.

Theorem 5. *For a given instance of the competitive influence with forcing problem, a choice of early adopters I_A, I_B and node x, $\hat{\alpha}_v^t \geq \alpha_v^t, \hat{\beta}_v^t \leq \beta_v^t$ for any node v and for any t.*

The following theorem shows that not only is the given model non-submodular, but also regardless of the tie-breaking rule and the forcing rule (if any is used), the model remains non-submodular.

Theorem 6. *For any tie-breaking rule, and any forcing rule, the separated-threshold competitive model is non-submodular.*

A corresponding counter-example is given in appendix C.

5 Hardness of Approximation

We show that in any model with separate edge-weights and separate threshold values for each technology the problem is hard to approximate.

Theorem 7. *It is NP-hard to give an approximation with a ratio better than $\Omega(N^{\frac{1}{2}-\epsilon})$, for all $\epsilon > 0$, for the Separated-Threshold Competitive Influence problem, where N is the number of nodes in the graph.*

The proof is supplied in appendix D. It is important to note the proof of theorem can be applied to similar competitive cascade models as well. Namely, in [2] it was conjectured that when allowing 2 sets of edge weights for each edge — one for each technology, the process will remain monotone and submodular. The above hardness of approximation result can be modified in order to apply for the separate edge-weights case of the Wave Propagation model suggested by Carnes et al., thereby giving a negative answer to their conjecture.

Theorem 8. *It is NP-hard to give an approximation with a ratio better than $\Omega(N^{\frac{1}{2}-\epsilon})$, for all $\epsilon > 0$, for the Wave Propagation Competitive Influence problem given by Carnes when edges are allowed to have technology-specific probabilities.*

6 The OR Model

We now introduce a different way of extending the original threshold model, in which each technology diffuses unhindered by the competing technology. Here, the tie-breaking stage will take place after all technologies finish spreading. This model can be considered natural for cases in which individuals have the liberty of being influenced separately and independently by different technologies, but have to choose a single technology eventually.

We will define the *OR* model as follows. As before, an instance of the model is a graph $G = (V, E)$, a set of edge weights for each technology: $W_A = \{w^A_{u,v}\}_{(u,v)\in E}, W_B = \{w^B_{u,v}\}_{(u,v)\in E}$ (with the same constraints as before), and initial adopters: $I_A, I_B \subseteq V$. Additionally, for each node $v \in V$ two "decision" functions $f^A_v : 2^V \times 2^V \to [0,1]$, $f^B_v : 2^V \times 2^V \to [0,1]$ are assigned. Let each technology propagate separately throughout the graph w.r.t the original non-competitive linear threshold propagation process, and let $R_A, R_B \subseteq V$ be the sets of nodes reached by the technologies (independently). As a final step, a node $v \notin I_A \cup I_B$ will assume technology A with probability $f^A_v(R_A, R_B)$, technology B with probability $f^B_v(R_A, R_B)$, and no technology with probability $1 - f^A_v(R_A, R_B) - f^B_v(R_A, R_B)$, respectively ($f^A_v(R_A, R_B) + f^B_v(R_A, R_B) \leq 1$). We only require the functions $f^A_v(\cdot,\cdot)$, for every $v \in V$, to be monotone and submodular with respect to the set of initial A nodes.

The following theorem shows that one can efficiently find an approximation for the set that maximizes the spread of one's own technology, given a competitor[s] choice of initial adopters:

Theorem 9. *Given technology B's early adopters I_B, one can find an $(1 - e^{-1} - \epsilon)$-approximation for the competitive OR process in a polynomial number of steps, for any $\epsilon > 0$.*

The proof follows immediately from the following two lemmas which prove the properties required in [4]. We will show that this process is monotone and submodular whenever the function $f_v(\cdot,\cdot)$ is monotone and submodular w.r.t. technology A, for all $v \in V$.

Lemma 1. *The OR model is monotone with respect to the number of nodes influenced by technology A.*

Proof. Let $r_A(I_A), r_B(I_B)$ denote an outcome for a run of the independent propagation processes of the two technologies. Monotonicity w.r.t technology A is satisified if for any two sets $S \subseteq T \subseteq V - I_B$:

$$\mathbb{E}[f^A_v(r_A(S), r_B(I_B))] \leq \mathbb{E}[f^A_v(r_A(T), r_B(I_B))] \tag{2}$$

Since until the decision step the two technologies' propagations are independent, we can fix the outcome of technology B, and show that the expected propagation of technology A is monotone. This is immediate since first, the propagation of technology A until the decision step is clearly monotone (follows from the non-competitive threshold model in [6]). Second, the decision functions $f^A_v(\cdot,\cdot)$ and

$f_v^B(\cdot,\cdot)$ are monotone with respect to technologies A and B, which along with the previous argument yields monotonicity.

Lemma 2. *The OR model is submodular with respect to the number of nodes influenced by technology A.*

Proof. In order to prove this, we will use a technique given in [6] that suggests an alternative and equivalent model for the propagation of a single technology. For each node $v \in V$, instead of choosing a threshold in $[0,1]$, choose an incoming edge (u,v) with respective probability $w_{u,v}$, and no incoming edge with probability $\sum_u w_{u,v}$. A node will become infected if and only if there is a path from the initially infected nodes that consists strictly of such chosen edges.

Fix an instantiation R_B of the outcome of the propagation of technology B (independent of the propagation of technology A) and a set of chosen edges for the propagation process of technology A. For a set I_A of initial A nodes, as before, let $R_A(I_A)$ denote the set of nodes reachable from I_A in the sub-graph induced by the set of chosen edges. In order to show that the process is submodular, we need to show that for all $S \subseteq T \subseteq V - I_B$:

$$f_v^A(R_A(S\cup\{x\}),R_B)-f_v^A(R_A(S),R_B) \geq f_v^A(R_A(T\cup\{x\}),R_B)-f_v^A(R_A(T),R_B), \tag{3}$$

for all $v \in V$. We will simply use the monotonicity property of the independent propagation process and the submodularity of $f_v^A(\cdot,\cdot)$. Let $R_A(S \cup \{x\}) = R_A(S) \cup \Delta_1$, and similarly, $R_A(T \cup \{x\}) = R_A(T) \cup \Delta_2$. From the monotonicity and submodularity we get that $R_A(S) \subseteq R_A(T)$ and $\Delta_2 \subseteq \Delta_1$. Therefore:

$$\begin{aligned} & f_v^A(R_A(S) \cup \Delta_1, R_B) - f_v^A(R_A(S), R_B) \\ \geq\ & f_v^A(R_A(T) \cup \Delta_1, R_B) - f_v^A(R_A(T), R_B) \\ \geq\ & f_v^A(R_A(T) \cup \Delta_2, R_B) - f_v^A(R_A(T), R_B) \end{aligned} \tag{4}$$

The first inequality and second inequality follow from the submodularity and the monotonicity of $f_v^A(\cdot,\cdot)$, respectively. Taking all possible instantiations gives submodularity since a positive linear combination of submodular functions is submodular.

Mossel et al. [8] show that if we generalize the propagation process by replacing the linear sum (used to decide whether an uninfected node exceeds its threshold) with an arbitrary monotone submodular function, then the resulting process (under any monotone submodular objective function) is again monotone and submodular. This result generalizes to the corresponding competitive process, which we call the *generalized OR process*.

Theorem 10. *Given technology B's early adopters I_B, one can find an $(1 - e^{-1} - \epsilon)$-approximation for the generalized competitive OR process in a polynomial number of steps, for any $\epsilon > 0$.*

Proof. Use the objective function $\varphi(R_A) = \mathbb{E}_{R_B} \sum_v f_v^A(R_A, R_B)$ in the main result of [8]. The function φ counts the expected number of A-adopters at the end of the process. It is monotone and submodular because the f_v^A are.

6.1 Repeating OR processes

Finally, we give a natural extension of the *OR* model. There are cases in which the independent propagation process will repeat several times (e.g. every day, for ℓ days). The process can be thought of as being run iteratively, where during each iteration i the previous iteration's turnouts R_A and R_B are used as the initial adopters for each technology. At the end of the ℓ'th iteration, *and only then*, the decision step is invoked by using the functions $f_v^A(\cdot,\cdot)$ and $f_v^B(\cdot,\cdot)$, for all $v \in V$. One may notice that this formulation simply defines a composition of ℓ *OR* processes (with a single execution of the decision step at the end).

We can give a natural motivation for such a process: during the course of an election race, voters will spread the word each day. However, once in while, an unaffected voter may change her mind (her threshold value) and thus the process of "rumor spread" and social-based recommendation will run again, infecting additional voters as a result.

With this in mind, the following general theorem follows from a simple generalization of the proof in [8].

Theorem 11. *A process based on the repetitive execution of the generalized OR process with a single decision step at the end is monotone and submodular.*

Note that theorem 11 holds even if the edge weights are modified between each iteration.

7 Conclusions

We have presented a number of fairly natural and general approaches for modelling competitive diffusion of influence in a social network, extending the known threshold model for the spread of a single technology. However, most of our suggested approaches have been shown to be unfit for the Nemhauser et al. [4] approximation technique. For some models, we can show NP-hardness of approximation, while for others we only show that they are not submodular (and not even monotone in one case), leaving open the question to whether an efficient approximation algorithm can be found.

We emphasize that all of the suggested models in this paper have reasonable, natural motivations, which implies that there is no single "true" model. Also, as suggested in [2], we believe that these models can be used in a more game theoretic setting, where players are the competing companies, who place bids on strategic nodes in hope for maximizing their outcome. We suggest the following directions for future research:

- Can the hardness-of-approximation result be extended to other models?
- Are there any other natural competitive models which are approximable in polynomial time?
- Study some natural game-theoretic setting for the competitive models.
- Suggest models for cases where nodes may adopt more than one technology.

References

1. S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *WINE*, pages 306–311, 2007.
2. T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *ICEC '07: Proceedings of the ninth international conference on Electronic commerce*, pages 351–360, New York, NY, USA, 2007. ACM.
3. P. Domingos and M. Richardson. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66, New York, NY, USA, 2001. ACM.
4. M. F. G. Nemhauser, L. Wolsey. An analysis of the approximations for maximizing submodular set functions. In *Mathematical Programming, 14*, pages 265–294, 1978.
5. P. R. J. Brown. Social ties and word-of-mouth referral behavior. In *Journal of Consumer Research 14:3*, pages 350–362, 1987.
6. D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146, New York, NY, USA, 2003. ACM.
7. J. Kostka, Y. A. Oswald, and R. Wattenhofer. Word of mouth: Rumor dissemination in social networks. In *SIROCCO*, pages 185–196, 2008.
8. E. Mossel and S. Roch. Submodularity of influence in social networks: From local to global. *SIAM J. Comput.*, 39(6):2176–2188, 2010.

A Counter Examples for The Weight-Proportional Competitive Linear Threshold Model

In section 2 we gave two theorems concerning the monotonicity and submodularity of the model described. These theorems will be proven in this appendix.

Theorem 12. *There exists an instance of the weight-proportional competitive linear threshold problem for which monotonicity does not hold.*

Proof. Consider the tree in figure 1(a). One can verify that $\alpha_r^4 = \frac{11}{40}$, whereas $\hat{\alpha}_r^4 = \frac{103}{400}$, which violates monotonicity.

Theorem 13. *There exists a graph G, for which the expected influence of technology A is not submodular.*

Proof. Consider the tree depicted in figure 1(b). It can be easily shown that for $S = \{w\}, T = \{w, y\}$ (the set of early adopters of technology B is denoted in the diagram) submodularity does not hold as $\alpha_r^4(S) = \frac{3}{10}, \hat{\alpha}_r^4(S) = \frac{17}{60}, \alpha_r^4(T) = \frac{7}{10}, \hat{\alpha}_r^4(T) = \frac{17}{50}$.

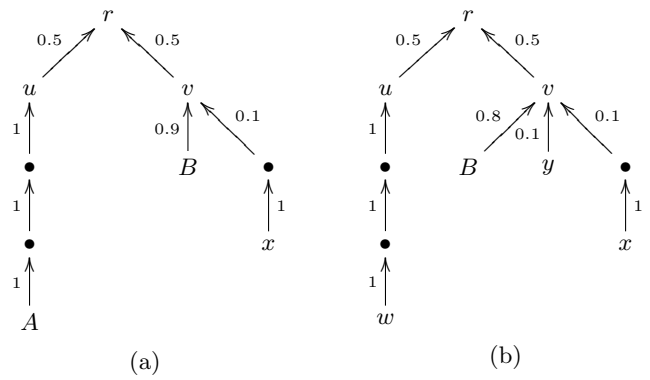

Fig. 1: Counter-examples for (a) monotonicity (b) submodularity

B Proof of Monotonicity of Separated Threshold Model

Theorem 14. *For a given instance of the problem and a choice of early adopters: I_A, I_B and node x, $\hat{\alpha}_v^t \geq \alpha_v^t$ for any node v and for any step $t \geq 0$.*

In order to prove the theorem, we will fix the set I_B of early technology B adopters and consider a set of early technology A adopters I_A and a node x not in I_B.

We prove the monotonicity by fixing an arbitrary instantiation of the thresholds, and by choosing for every node technology A or B with equal probability; these choices will be revealed in cases where the two thresholds chosen for a particular node are exceeded simultaneously. Notice that this defines a deterministic instantiation of the process.

Denote by π_1, π_2 the deterministic processes using the same instantiations of the threshold values and coin-flips, and using (I_A, I_B) and $(I_A \cup \{x\}, I_B)$, respectively. Furthermore, let $N_A^t(\pi), N_B^t(\pi)$ denote the set of A and B active nodes at step t in process π, respectively. The following lemma implies theorem 3.

Lemma 3. *The following holds for each node $v \in V$ and every step $t \geq 0$:*

1. *If v is not B-active at step t in π_1, then it isn't B-active at any step $t' \leq t$ in π_2.*
2. *If v is A-active at step t of π_1, then v is activated in some step $t' \leq t$ in π_2.*

Proof. The straightforward proof by induction is omitted for lack of space. It can be found in the full version.

C Counter-Examples for The Competitive Threshold Model with Forcing

Theorem 15. *For any tie-breaking rule, and any forcing rule, the separated-threshold competitive model is non-submodular.*

Proof. We will give a counter-example in which there are no ties, and the node in question does not remain inactive. Consider the rooted tree in figure 2. Let $S = \{u\}, T = \{u, y\}$. The initially B-activated nodes are given in the diagram.

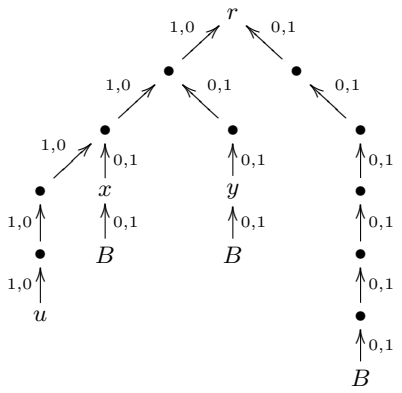

Fig. 2: A Counter-example for submodularity when applying a forcing step

One can check that there are no ties, and that forcing never applies to the root. Also, when using S, $\alpha_r^6 = \hat{\alpha}_r^6 = 0$. On the other hand, when using T, we get $\alpha_r^6(T) = 0$ and $\hat{\alpha}_r^6(T) = 1$, contradicting submodularity.

D Proof of The Hardness of Approximation Result

Theorem 16. *It is NP-hard to approximate the Seperated-Threshold Competitive Influence problem with a ratio better than $\Omega(N^{\frac{1}{2}-\epsilon})$, for all $\epsilon > 0$, where N is the number of nodes in the graph.*

Proof. We are motivated by the counter-example in theorem 6, constructing a reduction from Vertex Cover.

The reduction We are given an instance of Vertex Cover, a graph $G = (V, E)$ and a number k. Let α, β be constants defined later. Our new graph contains a special vertex A_0, a vertex A_v for each node $v \in V$, n^α vertices $B_0^{e,t}, X_0^{e,t}, X_1^{e,t}, M^{e,t}$ for each edge $e \in E$, and an extra n^α vertices B_1^t, P_0^t, P_1^t; here $1 \leq t \leq n^\alpha$. The rest of the graph appears in figure 3, where

- Dotted edges have A-weight 1 and B-weight 0.
- Dashed edges have A-weight 0 and B-weight 1.
- Plain edges have both weights set to 1.
- Edges with a length annotation are paths of that length of the given type.

Finally, I_B is composed of the set of nodes $B_0^{e,t}$ and B_1^t for every $e \in E$ and t.

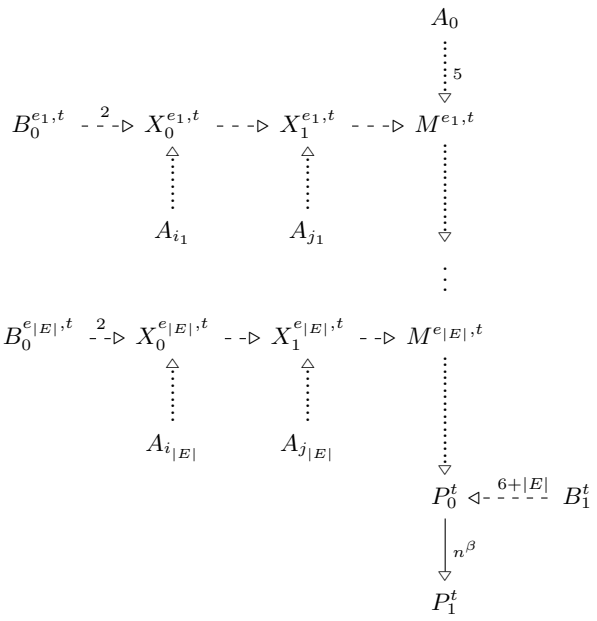

Fig. 3: The reduction. Diagram repeated for each t (except for the A_x).

Claim. If there exists a k-cover for the original graph G there exists a set $I_A \subseteq V - I_B$ of size $k+1$ that yields $\sigma(I_A) \geq n^{\alpha+\beta}$. Otherwise, for every $I_A \subseteq V - I_B$, $\sigma(I_A) = O(\max\{n^{\alpha+3}, n^{\beta+1}\})$.

Proof. Assume first that there is a k-cover S for G. Let $I_A = \{A_v | v \in S\} \cup \{A_0\}$. Since S is a vertex-cover, the spread of technology B emanating from the vertices B_0^t is completely blocked. Thus, every node on the path from A_0 to P_0^t, for all t, will be A-infected. Hence every node on the path from P_0^t to P_1^t will be A-infected. Thus, we have at least $n^{\alpha+\beta}$ A-active nodes, as required.

For the second part of the claim, for any set I_A of $k+1$ initial A-adopters, either $A_0 \notin I_A$ or $I_A \cap \{A_v\}_{v \in V}$ is not a vertex cover. Therefore the best choices for vertices in I_A are: choosing A_0, which contributes at most $(|E|+5)n^\alpha$ nodes; and choosing P_0^t, which contributes n^β nodes. The contribution of vertices of the first type is at most $O(s \cdot n^{\alpha+2}) = O(n^{\alpha+3})$, and the vertices of the second type contribute at most $O(s \cdot n^\beta) = O(n^{\beta+1})$.

Set $\beta = \alpha + 2$. The total number of vertices in the reduced graph is $N = O(n^{\alpha+\beta} + |E| \cdot n^\alpha) = O(n^{2\alpha+2})$. Thus we get that if there is a k-cover for G then the optimal I_A yields $\sigma(I_A) = \Omega(N)$, whereas any I_A that does not correspond to a k-cover yields $\sigma(I_A) = O(N^{(\alpha+3)/(2\alpha+2)})$. Hence, any algorithm that gives an approximation ratio of $o(N^{1-(\alpha+3)/(2\alpha+2)})$ can solve the NP-complete vertex cover problem. Therefore the approximation ratio of any poly-time algorithm is $\Omega(N^{1/2-\epsilon})$, for all $\epsilon > 0$, unless $P = NP$.