Competitive Influence Maximization in Social Networks

Shishir Bharathi David Kempe Mahyar Salek

Department of Computer Science, University of Southern California

Abstract. Social networks often serve as a medium for the diffusion of ideas or innovations. An individual's decision whether to adopt a product or innovation will be highly dependent on the choices made by the individual's peers or neighbors in the social network. In this work, we study the game of innovation diffusion with multiple competing innovations such as when multiple companies market competing products using viral marketing. Our first contribution is a natural and mathematically tractable model for the diffusion of multiple innovations in a network. We give a $(1-1/e)$ approximation algorithm for computing the best response to an opponent's strategy, and prove that the "price of competition" of this game is at most 2. We also discuss "first mover" strategies which try to maximize the expected diffusion against perfect competition. Finally, we give an FPTAS for the problem of maximizing the influence of a single player when the underlying graph is a tree.

1 Introduction

Social networks are graphs of individuals and their relationships, such as friendships, collaborations, or advice seeking relationships. In deciding whether to adopt an innovation (such as a political idea or product), individuals will frequently be influenced, explicitly or implicitly, by their social contacts. In order to effectively employ *viral marketing* [1, 2], i.e., marketing via "word-of-mouth" recommendations, it is thus essential for companies to identify "opinion leaders" to target, in the hopes that influencing them will lead to a large cascade of further recommendations. More formally, the *influence maximization problem* is the following: Given a probabilistic model for influence, determine a set A of k individuals yielding the largest expected cascade.[1] The formalization of influence maximization as an optimization problem is due to Domingos and Richardson [1], who modeled influence by an arbitrary Markov random field, and gave heuristics for maximization. The first provable approximation guarantees are given in [3–5].

In this paper, we extend past work by focusing on the case when multiple innovations are competing within a social network. This scenario will frequently arise in the real world: multiple companies with comparable products will vie for

[1] More realistic models considering different marketing actions which affect multiple individuals can usually be reduced to the problem as described here (see [3]).

sales with competing word-of-mouth cascades; similarly, many innovations face active opposition also spreading by word of mouth. We propose a natural generalization of the *independent cascade model* [2] to multiple competing influences (see Section 2 for details). Our model extends Hotelling's model of competition [6], and is related to competitive facility location and Voronoi games [7, 8]. We first study second-mover strategies and equilibria of the resulting activation game and show that:

Theorem 1. *The last agent i to commit to a set S_i for initial activation can efficiently find a $(1 - 1/\mathrm{e})$ approximation to the optimal S_i.*

Theorem 2. *The price of competition of the game (resulting from lack of coordination among the agents) is at most a factor of 2.*

We analyze good first-mover strategies for the two-player game in Section 4, and give exact algorithms for simple special cases. Finally, we give an FPTAS for maximizing the influence of a single player on bidirected trees, even when the edges in opposite directions have different probabilities. [3, 4] proved that the general version of influence maximization is NP-complete, and we conjecture it is so even for arborescences directed into a root.

2 Models and Preliminaries

The social network is represented as a directed graph $G = (V, E)$. Following the independent cascade model [2, 3], each edge $e = (u, v) \in E$ has an activation probability p_e. Each node can be either *inactive* or *active*; in the latter case, it is assigned a color denoting the influence for which it is active (intuitively, the product that the node has adopted). To avoid tie-breaking for simultaneous activation attempts by multiple neighbors, we augment the model by a notion of *activation time* for each activation attempt. When node u becomes active at time t, it attempts to activate each currently inactive neighbor v. If the activation attempt from u on v succeeds, v will become active, of the same color as u, at time $t+T_{uv}$, where the T_{uv} are independent and exponentially distributed continuous random variables. Subsequently, v will try to activate inactive neighbors, and so forth. Thus, a node always has the color of the first neighbor succeeding in activating it.

In the influence maximization game, each of b players selects a set S_i of at most k_i nodes. A node selected by multiple players will take the color of one of the players uniformly at random. Then, with S_i being active for influence i, the process unfolds as described above until no new activations occur. Letting $T_1, \ldots, T_b$ be the active sets at that point, the goal of each player i is to maximize $\mathrm{E}[|T_i|]$. Player i is indifferent between strategies S_i and S'_i if their expected gain is the same. In particular, the game is thus not a zero-sum game. Simple examples show that in general, this game has no pure strategy Nash Equilibria; however, it does have mixed-strategy Nash Equilibria.

3 Best Response Strategies

In order to gain a better understanding of the influence maximization game, we first focus on best response strategies for players.

Lemma 1. *Suppose that the strategies $S_j, j \neq i$ for other players are fixed. Then, player i's payoff $\mathrm{E}\left[|T_i| \mid S_1, \ldots, S_b\right]$ from the strategy S_i is a monotone and submodular function of S_i.*

Proof. We obtain a deterministic equivalent of the activation process by choosing independently if each edge $e = (u, v)$ will constitute a successful activation attempt by u on v (a biased coin flip with probability p_e), as well as the activation time T_e, beforehand. Then, we consider running the (now deterministic) activation process using these outcomes and delays.

If node u has color j, and activates node v successfully, we color the edge (u, v) with color j. A path P is called a color-j path if all its edges have color j. Then, a node u ends up colored with color j iff there is a color-j path from some node in S_j to u.

Conditioned on any outcome of all random choices as well as all $S_j, j \neq i$, the set of nodes reachable along color-i paths from S_i is the union of all nodes reachable from any one node of S_i. Thus, if $S_i \subseteq S'_i$, the set of nodes reachable from $S_i + v$, but not from S_i, is a superset of those reachable from $S'_i + v$, but not from S'_i (by monotonicity). Thus, given fixed outcomes of all random choices and $S_j, j \neq i$, the number of nodes reachable from S_i is a monotone and submodular function of S_i. Being a non-negative linear combination of submodular functions (with coefficients equal to the probabilities of the outcomes of the random choices), the objective function of player i is thus also monotone and submodular. ■

The above lemma implies that for the last player to commit to a strategy, the greedy algorithm of iteratively adding a node with largest marginal gain is within a factor $(1 - 1/\mathrm{e})$ of the best response (see [3]), thus proving Theorem 1. Second, as the expected total number of active nodes at the end is also a monotone submodular function of $S := \bigcup_j S_j$, the game meets the requirements of a *valid utility system* as defined by Vetta [10]. We can apply Theorem 3.4 of [10] to obtain that the expected total number of nodes activated in any Nash Equilibrium is at least half the number activated by the best solution with a single player controlling all of the $\sum_i b_i$ initial activations. This proves Theorem 2.

4 First Mover Strategies

We now consider first mover strategies in a duopoly, with 2 players called "red" and "blue". The following variant of the competitive influence maximization problem is motivated by its similarity both to the case of multiple disjoint directed lines (discussed briefly below) and to a fair division problem: Given n lines of lengths $\ell_1, \ldots, \ell_n$, the red player first gets to make any k cuts, creating

$k+n$ *pieces* whose lengths sum up to the original lengths. The blue player picks the k largest segments ("blue pieces") and the red player gets the next-largest $\min(n,k)$ segments ("red pieces").

Assume for now that we know a "cutoff point" c such that all blue pieces have size at least c, and all red pieces have size at most c. Let $F(i,r,b,c)$ denote the maximum total size of r red pieces in the i^{th} line and $G(i,r,b,c)$ the maximum total size of r red pieces over the first i lines. Then, we obtain the following recurrence relation, which turns into a dynamic program in the standard way.

$$G(i,r,b,c) = \max_{r'=0\ldots r} \max_{b'=0\ldots b} F(i,r',b'(r'),c) + G(i-1,r-r',b-b'(r'),c)$$

The main issue is then to reduce the number of candidates for the cutoff point c to a strongly polynomial number. The following lemma shows that we only need to try out nk candidate values $\ell_i/j, i=1,\ldots,n, j=1,\ldots,k$ for c (retaining the best solution found by the dynamic program), making the algorithm strongly polynomial.

Lemma 2. *The optimal solution cuts each line segment into equal-sized pieces.*

Proof. First, we can remove unused line segments from the problem instance. Second, partially used line segments can be converted to completely used line segments by adding the unused part to an existing blue segment (if it exists) or to an existing red segment (if no blue piece exists). The latter may entice the blue player to take a red piece. But this frees up a formerly blue piece (of size at least c) to be picked up by the red player.

W.l.o.g., all pieces of the same color on a line segment are of the same size. If the optimal solution contains an unevenly cut line with red and blue pieces, we increase the sizes of all red pieces and decrease the sizes of all blue pieces until the line is cut evenly. As before, the red player's gain cannot be reduced by the blue player switching to a different piece, because any new piece the red player may obtain after the blue player switches will have size at least c. ■

The above algorithm can be extended to deal with directed lines and even outdirected arborescences. In the former case, the slight difference is that the "leftmost" piece of any line is not available to the red player. These extensions are deferred to the full version due to space constraints.

5 Influence Maximization on Bidirected Trees

While the single-player influence maximization problem is APX-hard in general [3], special cases of graph structures are more amenable to approximation. Here, we will give an FPTAS for the influence maximization problem for bi-directed trees. (This FPTAS can be extended to bounded treewidth graphs with a significant increase in complexity.) Given a target ε, we will give a $1-\varepsilon$ approximation based on a combination of dynamic programming and rounding of probabilities.

For the subtree rooted at node v, let $G(v,k,q^+,q^-)$ denote the expected number of nodes that will be activated by an optimum strategy, provided that (1) v is activated by its parent with probability at most q^-, and (2) v has to be activated by its subtree with probability at least q^+.

Let v be a node of degree d with children $v_1,\ldots v_d$. Then, for the respective subproblems, we can choose arbitrary $k_1,\ldots,k_d$, $q_1^+,\ldots,q_d^+$, $q_1^-,\ldots,q_d^-$, such that (1) $\sum_i k_i = k$, $q^+ \leq 1-\prod_i(1-q_i^+ p_{v_i,v})$ and $q_i^- \leq p_{v,v_i}(1-(1-q^-)\prod_{j\neq i}(1-q_j^+ p_{v_j,v}))$ if v is selected, or (2) $\sum_i k_i = k-1$, $q^+ \leq 1$ and $q_i^- \leq p_{v,v_i}$ if v is not selected. If (1) or (2) is satisfied, we call the values *consistent*. For consistent values, the optimum can be characterized as:

$$G(v,k,q^+,q^-) = \max_{(k_i),(q_i^+),(q_i^-)} \sum_{i=1}^{m} G(v_i,k_i,q_i^+,q_i^-) + 1 - (1-q^+)(1-q^-). \quad (1)$$

As discussed above, the maximum is over both the case that v is selected, and that it is not. It can be computed via a nested dynamic program over the values of i. In this form, $G(v,k,q^+,q^-)$ may have to be calculated for exponentially many values of q^+ and q^-. To deal with this problem, we define $\delta = \varepsilon/n^3$, and compute (and store) the values $G(v,k,q^+,q^-)$ only for q^+ and q^- which are multiples of δ between 0 and 1. The number of computed entries is then polynomial in n and $1/\varepsilon$. Let $G'(v,k,q^+,q^-)$ denote the gain obtained by the best consistent solution to the rounding version of the dynamic program, and $\lfloor q \rfloor_\delta$ the value of q rounded down to the nearest multiple of δ. Then, for the rounding version, we have

Theorem 3. *For all v,k,q^+,q^-, there exists a value $r^+ \leq q^+$ with $q^+ - r^+ \leq \delta|T_v|$, such that $G(v,k,q^+,q^-) - G'(v,k,r^+,\lfloor q^- \rfloor_\delta) \leq \delta|T_v|^3$, where $|T_v|$ is the number of nodes in the subtree rooted at v.*

Applying the theorem at the root of the tree, we obtain that the rounding dynamic program will find a solution differing from the optimum by at most an additive $\delta n^3 \leq \varepsilon \leq \varepsilon \cdot \mathrm{OPT}$, proving that the algorithm is an FPTAS.

Proof. We will prove the theorem by induction on the tree structure. It clearly holds for all leaves, by choosing $r^+ = \lfloor q^+ \rfloor_\delta$. Let v be an internal node of degree d, with children $v_1,\ldots,v_d$. Let $(k_i),(q_i^+),(q_i^-)$ be the arguments for the optimum subproblems of $G(v,k,q^+,q^-)$. By induction hypothesis, applied to each of the subtrees, there are values $r_i^+ \leq q_i^+$ with $q_i^+ - r_i^+ \leq \delta|T_{v_i}|$, such that $G(v_i,k_i,q_i^+,q_i^-) - G'(v_i,k_i,r_i^+,\lfloor q_i^- \rfloor_\delta) \leq \delta|T_{v_i}|^3$.

Define $r^+ := \lfloor 1-\prod_i(1-r_i^+ p_{v_i,v}) \rfloor_\delta$, (or $r^+ = 1$, if the optimum solution included node v). By definition, r^+ is consistent with the r_i^+. Using Lemma 4 below and the inductive guarantee on the r_i^+ values, we obtain directly that $q^+ - r^+ \leq \delta|T_v|$ (where we used the fact that $\sum_i |T_{v_i}| + 1 = |T_v|$). Next, we define $r_i^- = \lfloor p_{v,v_i}(1-(1-\lfloor q^- \rfloor_\delta)\prod_{j\neq i}(1-r_j^+ p_{v_j,v})) \rfloor_\delta$ for all i. Again, the r_i^- are consistent by definition, and by using the inductively guaranteed bounds on $q_j^+ - r_j^+$ as well as Lemma 4, we obtain that $q_i^- - r_i^- \leq \delta(|T_v|+1)$ for all i.

Now, applying Lemma 3, we obtain that $G'(v_i, k_i, r_i^+, q_i^-) - G'(v_i, k_i, r_i^+, r_i^-) \leq \delta|T_{v_i}|(|T_v| + 1)$, for all i. In other words, because the input values to the subproblems did not need to be perturbed significantly to make the r_i^- consistent, the value of the rounding dynamic program cannot have changed too much. Combining these bounds with the inductive assumption for each subproblem, we have:

$$\begin{aligned} G'(v, k, r^+, \lfloor q^- \rfloor_\delta) &\geq \textstyle\sum_i G'(v_i, k_i, r_i^+, r_i^-) + 1 - (1 - r^+)(1 - \lfloor q^- \rfloor_\delta) \\ &\geq \textstyle\sum_i (G'(v_i, k_i, r_i^+, \lfloor q_i^- \rfloor_\delta) - \delta|T_{v_i}|(|T_v| + 1)) \\ &\quad +1 - (1 - q^+ + \delta|T_v|)(1 - q^- + \delta) \\ &\overset{IH}{\geq} \textstyle\sum_i (G(v_i, k_i, q_i^+, q_i^-) - \delta(|T_{v_i}|^3 + (|T_v| + 1)|T_{v_i}|)) \\ &\quad +1 - (1 - q^+)(1 - q^-) - \delta(1 + |T_v|) \\ &\geq G(v, k, q^+, q^-) - \delta|T_v|^3. \end{aligned}$$ ■

The following two lemmas are proved by induction; their proofs are deferred to the full version due to space constraints.

Lemma 3. *If* $r^- \leq q^-$, *then* $G'(v, k, q^+, q^-) - G'(v, k, q^+, r^-) \leq |T_v|(q^- - r^-)$.

Lemma 4. *For any* $a_1, \ldots, a_n$ *and* $b_1, \ldots, b_n$,

$$\textstyle\prod_{i=1}^n a_i - \prod_{i=1}^n b_i = \sum_{i=1}^n (a_i - b_i) \cdot \prod_{j=1}^{i-1} a_j \cdot \prod_{j=i+1}^n b_j$$

Acknowledgments We would like to thank Bobby Kleinberg and Ranjit Raveendran for useful discussions, and anonymous reviewers for helpful feedback on previous versions of this paper.

References

1. Domingos, P., Richardson, M.: Mining the network value of customers. In: Proc. 7th KDD. (2001) 57–66
2. Goldenberg, J., Libai, B., Muller, E.: Talk of the network: A complex systems look at the underlying process of word-of-mouth. Marketing Letters **12** (2001) 211–223
3. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence in a social network. In: Proc. 9th KDD. (2003) 137–146
4. Kempe, D., Kleinberg, J., Tardos, E.: Influential nodes in a diffusion model for social networks. In: Proc. 32nd ICALP. (2005)
5. Mossel, E., Roch, S.: On the submodularity of influence in social networks. In: Proc. 38th ACM STOC. (2007)
6. Hotelling, H.: Stability in competition. The Economic Journal **39** (1929) 41–57
7. Ahn, H., Cheng, S., Cheong, O., Golin, M., van Oostrom, R.: Competitive facility location along a highway. In: Computing and Combinatorics : 7th Annual International Conference, COCOON 2001, Guilin, China. (2001) 237–246
8. Cheong, O., Har-Peled, S., Linial, N., Matoušek, J.: The one-round voronoi game. Discrete And Computational Geometry **31** (2004) 125–138
9. Nemhauser, G., Wolsey, L., Fisher, M.: An analysis of the approximations for maximizing submodular set functions. Mathematical Programming **14** (1978) 265–294
10. Vetta, A.: Nash equlibria in competitive societies with applications to facility location, traffic routing and auctions. In: Proc. 43rd IEEE FOCS. (2002) 416–425