Social Influence Computation and Maximization in Signed Networks with Competing Cascades

Abstract—**Often in marketing, political campaigns and social media, two competing products or opinions propagate over a social network. Studying social influence in such competing cascades scenarios enables building effective strategies for maximizing the propagation of one process by targeting the most "influential" nodes in the network. The majority of prior work however, focuses on unsigned networks where individuals adopt the opinion of their neighbors with certain probability. In real life, relationships between individuals can be positive (e.g., friend-of relationship) or negative (e.g. connection between "foes"). According to social theory, people tend to have similar opinions to their friends but opposite of their foes. In this work, we study the problem of competing cascades on signed networks, which has been relatively unexplored. Particularly, we study the progressive propagation of two competing cascades in a signed network under the Independent Cascade Model, and provide an approximate analytical solution to compute the probability of infection of a node at any given time. We leverage our analytical solution to the problem of competing cascades in signed networks to develop a heuristic for the influence maximization problem. Unlike prior work, we allow the seed-set to be initialized with populations of both cascades with the end goal of maximizing the spread of one cascade. We validate our approach on several large-scale real-world and synthetic networks. Our experiments demonstrate that our influence maximization heuristic significantly outperforms stat-of-the-art methods, even more so when the network is dominated by distrust relationships.**

Keywords*—*analytical framework; diffusion models; social influence; social networks

I. Introduction

Online Social Networks have become a prevalent platform for the diffusion of ideas and the dissemination of news, a medium for political debates and opinion sharing, as well as the main channel for product marketing and innovation spreading. In deciding whether to adopt an innovation, a political idea, or a product, people are frequently influenced, either explicitly or implicitly, by their social contacts, aggregate social behavior, and external factors, or a combination of the above [1], [2]. Several influence diffusion models have been proposed in the literature to formulate the underlying influence propagation process [3], [4], [1], [5], [2]. Even though, current models of influence spread enable the study of complex and realistic scenarios, most models assume that the spreading process takes place on an unsigned network. In reality however, the polarity of relationships might not be always positive [6], [7], [8]. In online social networks such as Slashdot and Epinions for example, relationships might have a positive (e.g., represent "friends" or "trust") or negative (e.g., model "foe", "spite" or "distruct" relationships) connotation. Intuitively, positive relationships carry influence in a positive way, whereas negative edges carry influence in a reverse direction (i.e., one is more likely to follow a friend's choice, yet do the opposite of a foe).

People's reactions to social influence can be further complicated when multiple competing processes unfold over the network (e.g., multiple companies with similar products vie for sales using competing viral marketing campaigns) [9], [10], [11]. Intuitively, if a company wants to market a new product for it to be adopted by a large fraction of the network, it may try to identify and initially target a small set of "influential" members of the network on the premise of "viral marketing" [12]. The problem of influence maximization has been extensively studied for two widely used diffusion models, i.e., Independent Cascade (IC) and Linear Threshold (LT) and their extensions [3], [13]. As finding the optimal seed-set for influence maximization has been shown to be NP-hard [3], greedy algorithms and heuristics have been proposed to solve the problem [3], [14], [15], [16]. However, in the real-world, it is rarely the case that only a single company promotes a single product at any given time [9], [10], [11]. But then, how should few initial nodes be chosen for starting the process so that the expected total influence in a given signed network is maximized under a model of influence spread $\mathcal{M}$ in the presence of competing cascades? Once central aspect of this problem is the estimation of expected influence spread $\sigma(S)$, given the seed set, which is typically done using numerous simulations. Even for a single cascade diffusing based on Independent Cascade Model, it has been shown that exact computation of the $\sigma(S)$ is #P-hard [14]. Estimation of the influence spread through analytical computation can reduce computation by avoiding expensive simulations.

To fill the gap of influence computation and maximization in signed networks with competing diffusion processes, we propose a novel signed network influence maximization (SiNiMax) problem. The purpose of SiNiMax is to find a small set of seeds with maximum positive (similarly negative) influence in a signed social network. Unlike the few recent studies on influence maximization on signed networks [17], [7], [8], SiNiMax enables seeds to be either positive or negative; this facilitates diversification of the seed-set portfolio, taking advantage of both positive and negative relationships at the same time. Our framework enables the study of the spreading dynamics of two concurrent yet interdependent contagion processes over a signed network. Specifically, we extend the unified model of influence [2] to signed networks for competing cascades and study the dynamics of influence diffusion for two opposite opinions, which are modeled as positive and negative, and are spread over positive and negative edges on a signed network. We first characterize analytically the contagion phenomena of two competing cascades and compute the temporal evolution of influence in a signed network. We show how our closed-form expression can be used to efficiently study the unfolding dynamics of opposite opinions in a signed network without requiring extensive simulations. We then apply our model to solve the influence maximization problem and develop efficient algorithms to select initial seeds of either opinion that maximize influence coverage. We use both synthetic and real-world large-scale networks, such as Epinions and Slashdot, to confirm our theoretical analysis on competing influence diffusion dynamics over signed networks, and demonstrate that our influence maximization algorithm outperforms other heuristics.

Our work solves a fundamental problem and paves the way towards a more comprehensive description of the dynamics of competing processes over signed networks with applications to numerous applications and domains, including but not limited to competing cascades in online social networks and interacting diseases in viral epidemiology. The rest of this paper is organized as follows. Section II develops the analytical solution for computing infection probabilities for competing cascades in signed networks. In Section III we define the influence maximization problem that we propose to solve. We build on our solutions to develop a heuristic in Section IV. We describe our experiments in Section V. The related work has been discussed in Section VI and the conclusions from our work have been drawn in Section VII.

II. Unified Model of Competing Cascades in Signed Networks

We consider a weighted, directed, and signed graph $G = (V, E, W)$, where V is the set of nodes, and E is the set of directed edges. Edges represent influence; edge $(u, v) \in E$ if node u influences v. W is a matrix whose element w_{uv} denotes the signed weight of an edge (u, v) in the graph. Entries in matrix W are non-negative when the network is unsigned, but may contain both positive and negative entries when graph G is signed. Particularly, a positive entry w_{uv} may represent friendship or trust, whereas a negative value would be indicative of a foe or distrust relationship (i.e., node u distrusts node v. The absolute value $|w_{uv}|$ denotes the strength of influence.

To extend the present understanding of multiple contagions as they simultaneously spread through a signed network, we consider the case of two influence diffusion processes that spread in discrete time steps according to some propagation model $\mathcal{M}$. For simplicity, we describe the diffusion process of competing cascades in a signed network for the standard Independent Cascade model (ICM) [3], with the note that our results can be easily extended to other influence models.

In our modeling, two cascades spread over the network according to $\mathcal{M}$. We use two colors, black and white, to differentiate between them. A node can therefore be either susceptible, black or white. Initially, all nodes are susceptible, i.e., have not been exposed to any of the two cascades. We study the problem of progressive diffusion, according to which nodes that become colored (infected by either black or white) cannot become susceptible again. Additionally, we assume that once a node is colored it cannot change color in the future. At every time step t, with probability $p_{v,u}$, each node v that was infected at $t-1$ attempts to infect its outgoing neighbors u with its own color. A susceptible node on which multiple influence attempts were made, randomly selects one of such attempts and changes from susceptible to colored.

We first describe the spreading dynamics of two concurrent yet interdependent contagion processes over an unsigned network. We extend the Unified Model (UM) [2], a generalized analytical model of influence in networks, that incorporates both pairwise and collective influence dynamics into the diffusion mechanism for accurate calculation of the probability of infection at any time t. According to UM, the probability of node u being infected under ICM at or before time t, $B_{u,t}$ is given by $B_{u,t} = 1 - (1 - r_{u,t})(1 - B_{u,t-1})$. After some algebraic manipulation, the previous equation can be written as $B_{u,t} = B_{u,t-1} + r_{u,t}(1 - B_{u,t-1})$, where $r_u(t)$ denotes collective influence. According to [2] collective influence can be used to model local neighborhood effects, aggregate social behavior, and external factors, or a combination of the above. However, as shown in [2], it is possible (in case of ICM) to aggregate the effect of multiple pairwise infection attempts into the collective influence term. We use terms $B^+_{u,t}$ and $B^-_{u,t}$ to denote the probability of infection of node u with black and white respectively at or before time t. Similarly, $r^+_{u,t}$ and $r^-_{u,t}$ represent collective influence due to black and white influence respectively. From the perspective of an initially susceptible node, the probability of being colored black at time t can be formalized as:

$$P(u \text{ colored black at time } t) = P(u \text{ susceptible before } t, \text{ collective black influence succeds and white fails}) \quad (1)$$

However, the probability of node u being colored black at time t can be calculated as $A^+_{u,t} = B^+_{u,t} - B^-_{u,t-1}$. Therefore, Equation 1 becomes

$$A^+_{u,t} = r^+_{u,t}(1 - r^-_{u,t})(1 - B^+_{u,t} - B^-_{u,t-1}), \quad (2)$$

where

$$B^+_{u,t} = B^+_{u,t-1} + r^+_{u,t}(1 - r^-_{u,t})(1 - B^+_{u,t-1} - B^-_{u,t-1}), \quad (3)$$

$$B^-_{u,t} = B^-_{u,t-1} + r^-_{u,t}(1 - r^+_{u,t})(1 - B^+_{u,t-1} - B^-_{u,t-1}). \quad (4)$$

The collective influence probabilities for each competing cascade can be computed by separately due to the independence between the two influence processes in a single time step (i.e., the event of an attempt of a black infection on a node is independent of an attempt of white infection). Therefore extending the idea of collective influence of ICM from [2]to multiple cascades

$$r^+_{u,t} = 1 - \prod_{v \to u} (1 - p^+_{v,u} A^+_{v,t-1}) \quad (5)$$

and

$$r^-_{u,t} = 1 - \prod_{v \to u} (1 - p^-_{v,u} A^-_{v,t-1}) \quad (6)$$

where $p^+_{v,u}$ and $p^-_{v,u}$ represent the probabilities of node v exerting influence on u when v is colored black or white repectively. Equations 2, 3, 4, 5, and 6 constitute our novel analytical solution for computing infection probability of any node at any time t for competing cascades that propagate based on the ICM model on an unsigned network.

We naturally extend influence propagation model $\mathcal{M}$ (in this case ICM) for signed networks based on the social principles "the friend of my enemy is my enemy" and "the enemy of my enemy is my friend" [18]. Specifically, the influence is flipped when traversing a negative edge (u, v) between nodes u and v. Intuitively, if node u is colored black and is successful in infecting v, then v will become infected with white. However, from the perspective of v, u's attempt to influence v with black through a negative edge is equivalent to u trying to pass along white infection to v through an unsigned edge. Therefore, flipping the infection (from black to white or vice-versa) of the incoming neighbor which has a negative link and removing the sign from the link is equivalent to the diffusion process of signed network. Thus, competing cascades in a signed network can be reduced into an equivalent problem of competing cascades in an unsigned network. Particularly, Equations 3 and 4 can be used to calculate the infection probabilities in a signed network. For the calculation to be valid, the formulas for collective influence need to be modified. Formally,

$$r^{+}_{u,t} = 1 - \prod_{v \xrightarrow{+} u} (1 - p^{+}_{v,u} A^{+}_{v,t-1}) \prod_{v \xrightarrow{-} u} (1 - p^{-}_{v,u} A^{-}_{v,t-1}) \quad (7)$$

and

$$r^{-}_{u,t} = 1 - \prod_{v \xrightarrow{+} u} (1 - p^{-}_{v,u} A^{-}_{v,t-1}) \prod_{v \xrightarrow{-} u} (1 - p^{+}_{v,u} A^{+}_{v,t-1}) \,. \quad (8)$$

III. Influence Maximization in Signed Networks

We redefine the problem of Influence Maximization for signed networks as follows:

Definition 1 (Signed Network Influence Maximization): Given a diffusion model M of competing cascades $C = \{\text{black, white}\}$ running on a signed graph $G(V, E, w)$ possibly weighted, and an integer m, find $S \subseteq V \times C$ such that $|S| = m$, and $\forall (u, c) \in S$, infecting u with c at $t = 0$ maximizes the expected spread of the black infection denoted by $\sigma^{M,+}(S)$.

Note that since 'black' and 'white' are merely symbols representing the two cascades, Influence Maximization with 'black' infection is equivalent to the one with 'white' infection, and so an algorithm to maximize $\sigma^{M,+}(S)$ would also be able to maximize $\sigma^{M,-}(S)$. The most prominent difference of this problem from the traditional Influence Maximization problem [3] is that in the seed set S, along with the choice of nodes, we also need to decide the infection state black or white. Due to the sign on the links, it is possible that initializing a node with white color can lead to more black infections. Next we show the NP-hardness of the problem.

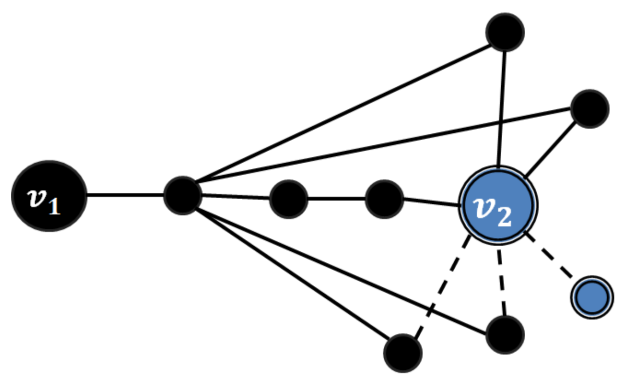
Theorem 1: Signed Network Influence Maximization problem is NP-hard.

Proof: Consider an instance of the traditional Influence Maximization problem where we need to find a seed-set of nodes infecting which would create maximum number of expected infections at steady state under Independent Cascade Model with parameter p.

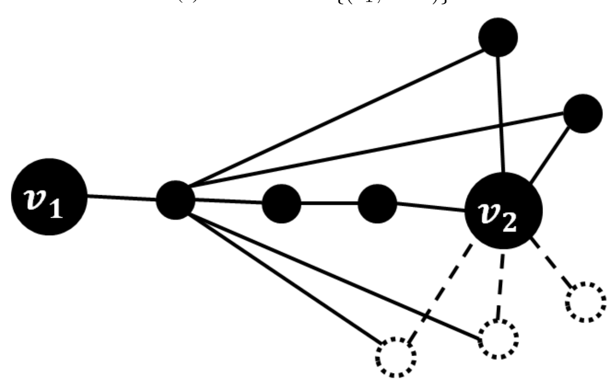
Suppose in the SiNIMax we set all the links in the graph to positive and $p^{-}_{v,u} = 0, p^{+}_{v,u} = p, \forall (v, u) \in E$. Now we proceed to solve SiNiMax to maximize $\sigma^{ICM,+}(S)$. Clearly, $\forall (u, c) \in S$, $c =$ black because 'white' cannot propagate and if $(u, \text{white}) \in S$, then replacing that element with $(u, +)$ can only increase $\sigma^{ICM,+}(S)$. Notice that this is equivalent to the traditional ICM as there is only one type of infection that propagates in the network. Therefore, a solution to SiNIMax would provide a solution to the traditional Influence Maximization problem which is known to be NP-hard.

Hence, SiNiMax is NP-hard. ■

Our formulation SiNiMax is similar to PRIM [8] with the only difference that we allow the addition of a node with initial coloring of 'white' for maximization of $\sigma^{ICM,+}(S)$. We point out one major difference of competing ICM over a signed network when compared to other formulations of ICM, in the following theorem.



(a) Seed set $S = \{(v_1, \text{black})\}$



(b) Seed set $S = \{(v_1, \text{black}), (v_2, \text{black})\}$

Fig. 1. Spread of 'black' and 'white' infections with (a) $S = (v_1, \text{black})$ and (b) $S = (v_1, \text{black})$. Solid line represents a positive link and dashed line represents a negative link. Blue nodes represent ambiguous infection.

Theorem 2: The function $\sigma^{ICM,+}(S)$ is not monotonic.

Proof: We disprove the monotonicity of $\sigma^{ICM,+}(S)$ by a counter-example. Consider the graph in Figure 1, where solid line represents a positive link and dashed line represents a negative link. Suppose $p_{v,u} = 1$ for all links. Starting with seed-set $\{(v_1, \text{black})\}$ (Figure 1(a)) we find that 8 nodes end up being colored 'black'. Two nodes have ambiguous infection.

Their state is dependent on which infection state, 'black' or 'white', node v_2 decides to adopt. In any case these two nodes must have opposite infections due to a negative link between them (v_2 is the only neighbor of the other node and so its infection must come through v_2). Therefore, $\sigma^{ICM,+}(S) = 9$. Inclusion of v_2 in the seed-set with 'black' color (Figure 1(b)) creates 7 'black' and 3 'white' infections. Therefore $\sigma^{ICM,+}(\{(v_1, \text{black}), (v_2, \text{black})\}) < \sigma^{ICM,+}(\{(v_1, \text{black})\})$. Infecting v_2 with 'white' in the seed-set creates at most 6 'black' infections. Thus, $\sigma^{ICM,+}(\{(v_1, \text{black}), (v_2, \text{white})\}) < \sigma^{ICM,+}(\{(v_1, \text{black})\})$, disproving the monotonicty. ■

IV. Seed-set Selection Heuristic

Based on the analytical solution obtained in II, we develop a heuristic for selecting the seed-set S for maximization of $\sigma^{M,+}(S)$. Our approach is an incremental one, where we start with an empty set and include (u, c) at time $t-1$ if infecting u with c at $t-1$ would create most number of new 'black' infections (expected value) at time t. The traditional greedy approach, on the other hand, would be to chooses (u, c) which maximizes $\sigma^{M,+}(S \cup (u, c)) - \sigma^{M,+}(S)$. We refrain from using the traditional greedy approach because it would require calculation of total spread (instead of immediate spread as in our heuristic) adding more computational requirements. Also, due to the lack of monotonicity the $(1-1/e)$-approximation [3] is no longer guaranteed. Owing to the low value of p, it follows that the effect of infection of a node decays quickly with distance. Therefore, the node creating the most new number of infections is expected to have high contribution to $\sigma^{M,+}(S)$.

A. OSSUMS

Let $B^+_{u,t}(k, +)$ represent the probability of node u being colored 'black' at time t if node k is infected with 'black' at time $t-1$, if it was not already infected. This is done by setting $B^+_{k,t-1}$ to $1-B^-_{k,t-1}$ and calculating its impact on node u. Similarly, $B^+_{u,t}(k, -)$ represent the probability of 'black' infection of node u at time t if node k is infected with 'white' at time $t-1$.

$$B^+_{u,t}(k,+) = \begin{cases} r^+_{u,t}(k,c)(1 - r^-_{u,t}(k,c)) \\ (1 - B^+_{u,t-1} - B^-_{u,t-1}) & \text{if } u \neq k \\ 1 - B^-_{u,t-1} & \text{if } u = k \end{cases} \quad (9)$$

Let $r^{(k,+)}_{u,t}$ and $r^{(k,-)}_{u,t}$ be the effective collective influence due to the infection of k with 'black' and 'white' respectively.

Our heuristic, termed Online Seed-set Selection using Unified Model on Signed Networks (OSSUMS), is based on selecting a node k and infecting it with $c \in \{\text{black, white}\}$

$$(k, c) = \arg\max_{(k,c)} \sum_u \left(B^+_{u,t}(k, c) - B^+_{u,t}\right) \quad (10)$$

Now,

$$\begin{aligned} &B^+_{u,t}(k,c) - B^+_{u,t} \\ &\quad = \left[r^+_{u,t}(k,c)(1 - r^-_{u,t}(k,c)) - r^-_{u,t}(k,c)(1 - r^+_{u,t}(k,c))\right] \\ &\quad (1 - B^+_{u,t-1} - B^-_{u,t-1}) \end{aligned} \quad (11)$$

To evaluate $\left(B^+_{u,t}(k, c) - B^+_{u,t}\right)$ for $u \neq k$, we need to consider four cases:

1) Node k is infected by 'black' and the link from k to u is positive:
Since k is infected with 'black' and $k \xrightarrow{+} u$, it affects only the 'black' collective influence. Therefore

$$r^-_{u,t}(k, +) = r^-_{u,t}. \quad (12)$$

And,

$$\begin{aligned} r^+_{u,t}(k,+) &= 1 - (1 - p_{k,u}(1 - B^+_{k,t-2} - B^-_{k,t-1})) \\ &\prod_{j \xrightarrow{+} u, j \neq k} (1 - p_{j,u} A^+_{j,t-1}) \prod_{j \xrightarrow{-} u,} (1 - p_{j,u} A^-_{j,t-1}) \\ &= 1 - \frac{1 - p_{k,u}(1 - B^+_{k,t-2} - B^-_{k,t-1})}{1 - p_{k,u} A^+_{k,t-1}} \\ &\prod_{j \xrightarrow{+} u} (1 - p_{j,u} A^+_{j,t-1}) \prod_{j \xrightarrow{-} u,} (1 - p_{j,u} A^-_{j,t-1}) \\ &= 1 - \frac{1 - p_{k,u}(1 - B^+_{k,t-2} - B^-_{k,t-1})}{1 - p_{k,u} A^+_{k,t-1}} (1 - r^+_{u,t}) \end{aligned} \quad (13)$$

Using Equations 11, 12 and 13, for this case

$$\begin{aligned} &\sum_{k \xrightarrow{+} u} (B^+_{u,t}(k,+) - B^+_{u,t}) \\ &= \sum_{k \xrightarrow{+} u} \frac{p_{k,u}(1 - B^+_{k,t-1} - B^-_{k,t-1})}{1 - p_{k,u} A^+_{k,t-1}} \\ &(1 - B^+_{u,t-1})(1 - B^-_{u,t-1})(1 - r^+_{u,t} - r^-_{u,t}) \end{aligned} \quad (14)$$

2) Node k is infected with 'black' and the link from k to u is negative: In this case, since this action does not affect the collective influence for 'black' infection on u,

$$r^+_{u,t}(k, +) = r^+_{u,t}. \quad (15)$$

And

$$\begin{aligned} r^-_{u,t}(k,+) &= 1 - (1 - p_{k,u}(1 - B^+_{k,t-2} - B^-_{k,t-1})) \\ &\prod_{j \xrightarrow{+} u} (1 - p_{j,u} A^-_{j,t-1}) \prod_{j \xrightarrow{-} u, j \neq k} (1 - p_{j,u} A^+_{j,t-1}) \\ &= 1 - \frac{1 - p_{k,u}(1 - B^+_{k,t-2} - B^-_{k,t-1})}{1 - p_{k,u} A^+_{k,t-1}} \\ &\prod_{j \xrightarrow{+} u} (1 - p_{j,u} A^-_{j,t-1}) \prod_{j \xrightarrow{-} u,} (1 - p_{j,u} A^+_{j,t-1}) \\ &= 1 - \frac{1 - p_{k,u}(1 - B^+_{k,t-2} - B^-_{k,t-1})}{1 - p_{k,u} A^+_{k,t-1}} (1 - r^-_{u,t}) \end{aligned} \quad (16)$$

Using Equations 11, 15 and 16, for this case

$$\sum_{k \xrightarrow{-} u} \left(B^+_{u,t}(k,+) - B^+_{u,t}\right) = \sum_{k \xrightarrow{-} u} \frac{-p_{k,u}(B^+_{k,t-1} + B^-_{k,t-1})}{1 - p_{k,u}A^+_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1}) r^+_{u,t}(1 - r^-_{u,t}) \tag{17}$$

3) Node k is infected with 'white' and the link from k to u is positive: Similar to Case 2, it can be shown that

$$\sum_{k \xrightarrow{+} u} \left(B^+_{u,t}(k,-) - B^+_{u,t}\right) = \sum_{k \xrightarrow{+} u} \frac{-p_{k,u}(B^+_{k,t-1} + B^-_{k,t-1})}{1 - p_{k,u}A^-_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1}) r^+_{u,t}(1 - r^-_{u,t}) \tag{18}$$

4) Node k is infected with 'white' and the link from k to u is negative: Similar to Case 1, it can be shown that

$$\sum_{k \xrightarrow{-} u} \left(B^+_{u,t}(k,-) - B^+_{u,t}\right) = \sum_{k \xrightarrow{-} u} \frac{p_{k,u}(1 - B^+_{k,t-1} - B^-_{k,t-1})}{1 - p_{k,u}A^-_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1})(1 - r^+_{u,t})(1 - r^-_{u,t}) \tag{19}$$

Finally, a node k is added to the seed-set with color $c \in$ {black, white} which maximizes the following:

$$\max_k \max\{\sum_u \left(B^+_{u,t}(k,+) - B^+_{u,t}\right), \sum_u \left(B^+_{u,t}(k,-) - B^+_{u,t}\right)\} \tag{20}$$

where the objective function is computed by combining Equations 1, 17, 18 and 19 which produces:

$$\begin{aligned}\sum_u (B^+_{u,t}(k,+) - B^+_{u,t}) &= \sum_{k \xrightarrow{+} u} \Bigg(\frac{p_{k,u}(1 - B^+_{k,t-1} - B^-_{k,t-1})}{1 - p_{k,u}A^+_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1})(1 - r^+_{u,t})(1 - r^-_{u,t}) \Bigg) \\ &- \sum_{k \xrightarrow{-} u} \Bigg(\frac{p_{k,u}(B^+_{k,t-1} + B^-_{k,t-1})}{1 - p_{k,u}A^+_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1}) r^+_{u,t}(1 - r^-_{u,t}) \Bigg)\end{aligned} \tag{21}$$

TABLE I. SUMMARY OF THE DATASETS

	Epinions	Slashdot
# nodes	131828	82144
# edges	841372	549202
# positive edges	717667	425072
# negative edges	123705	124130

And

$$\begin{aligned}\sum_u (B^+_{u,t}(k,-) - B^+_{u,t}) &= \sum_{k \xrightarrow{-} u} \Bigg(\frac{p_{k,u}(1 - B^+_{k,t-1} - B^-_{k,t-1})}{1 - p_{k,u}A^-_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1})(1 - r^+_{u,t})(1 - r^-_{u,t}) \Bigg) \\ &- \sum_{k \xrightarrow{+} u} \Bigg(\frac{p_{k,u}(B^+_{k,t-1} + B^-_{k,t-1})}{1 - p_{k,u}A^-_{k,t-1}} (1 - B^+_{u,t-1} - B^-_{u,t-1}) r^+_{u,t}(1 - r^-_{u,t}) \Bigg)\end{aligned} \tag{22}$$

The OSSUMS heuristic is summarized in Algorithm 1.

Algorithm 1 Online Seed-set Selection using Unified Model on Signed Network (OSSUMS)

```
function OSSUM(G, m)
    S ← ∅
    for t = 1 → m do
        (k, c) = argmax_(k,c) Σ_u (B+_{u,t}(k, c) − B+_{u,t})
                         ▷ Computed using Equations 21 and 22
        S ← S ∪ (k, c)
    end for
    return S
end function
```

B. Complexity Analysis

Computing $r^+_{u,t}$ and $r^-_{u,t}$ require $O(indegree(u))$ computations. Doing this for all nodes u requires $O(\sum_u (1 + indegree(u))) = O(|V| + |E|)$ operations. Once these values are calculated, Equation 21 and 22 are to be evaluated for each node k, which takes $O(\sum_k (1 + outdegree(k))) = O(|E| + |V|)$ computations. This is to be repeated for selection of each seed-set. Therefore, the time complexity of finding m nodes for seed-set S for ICM using OSSUMS is $O(m(|V| + |E|))$.

V. EXPERIMENTS

The experiments were conducted on two datasets: Epinions and Slashdot [6]. Both are trust relations among users in the corresponding social media. There link from u to v is positive if u trusts v (u considers v a friend) and negative if u distrusts v (u considers v a foe). The influence graph is constructed by flipping the direction of these edges. If u trusts v, then there is a positive link from v to u, and if u distrusts v, then there is a negative link from v to u. The datasets are summarized in Table V.

We compared our heuristic with the following heuristics

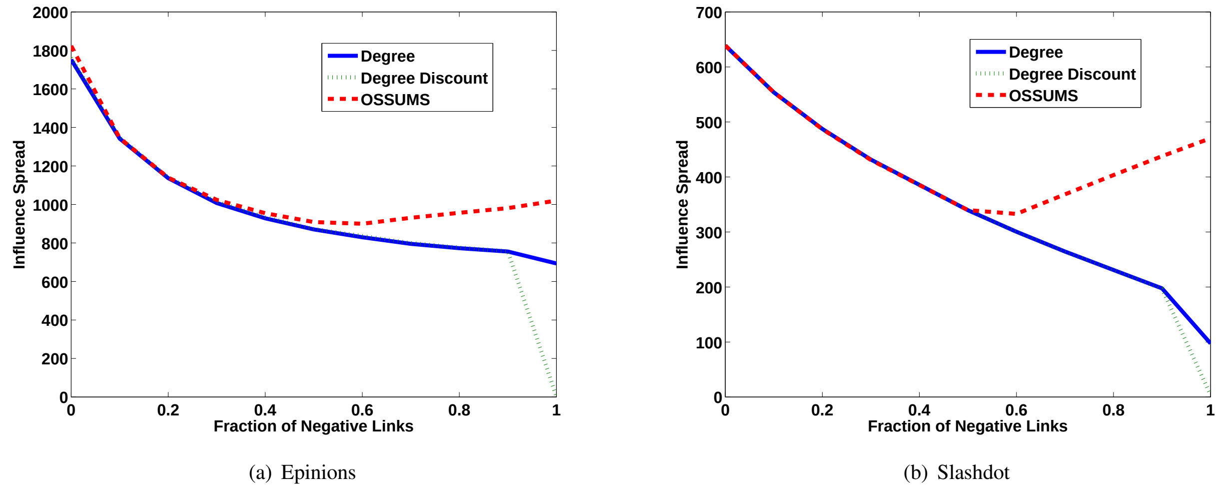


Fig. 2. Influence spread achieved by the heuristics on graphs with varying fraction of negative links.

- Degree: Choose the nodes with maximum positive degree and color them 'black'.
- Degree Discount [15]: A heuristic designed for traditional ICM, which performs a form of weighted discount based on the parameter p. This heuristic is applied after removing all negative edges, and the selected nodes are colored 'black'.

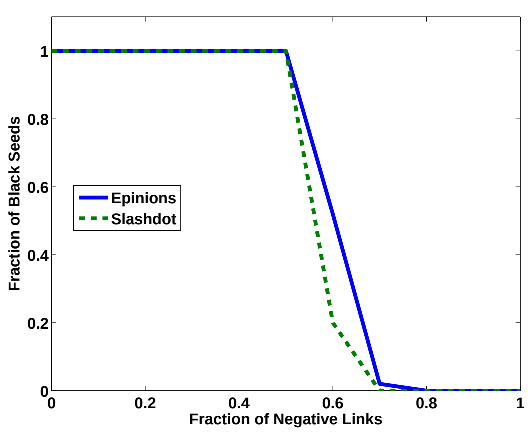


Fig. 3. Fraction of nodes in the seed-set included with 'black' color by OSSUMS. The fraction drops very quickly when majority of links in the network are negative.

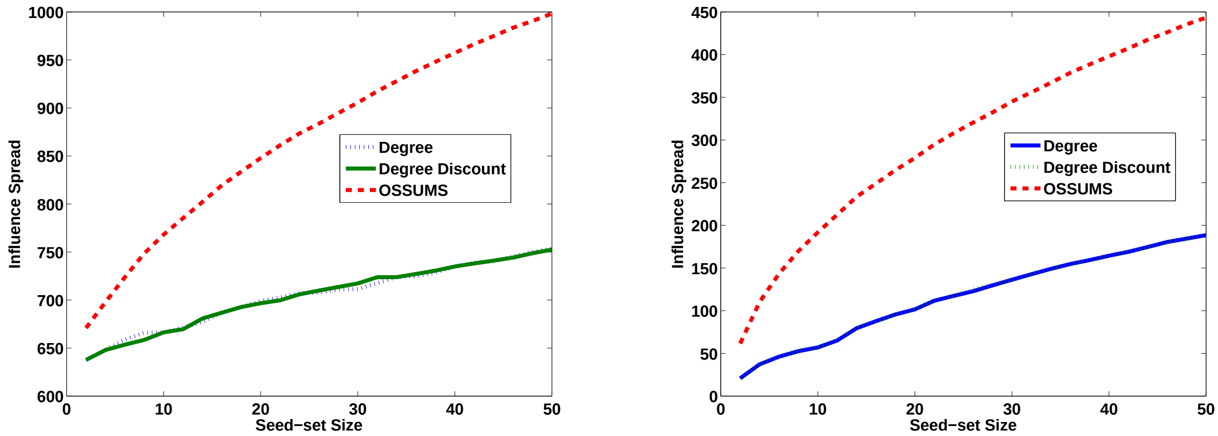
To study the effect of negative links on the influence spread achieved by the heuristics, first we ignored the actual signs of the links and randomly assign the signs. The selection was done at random with probability of negative link ranging from 0 to 1, resulting in a range of graphs with fraction of negative links $0, 0.1, 0.2, \ldots, 1$. Figure 2 shows the result. Observe that in both cases, the spread achieved by the heuristics are almost equal when the fraction of negative links is low. However, when the graph has more negative links OSSUMS significantly outperform other heuristics. Also, note that least spread is achieved by OSSUMS when the fraction of negative links is around 0.5, which suggests that achieving high influence spread is more difficult when there are almost equal number of positive and negative links compared to when all the links are negative.

When more links are negative in a network, i.e., there are more distrust relations between the individuals, it becomes important to include 'white' colored nodes in the seed set for maximization of 'black' infection. This is demonstrated in Figure V, which shows the fraction of nodes in the seed set included with 'black' color by OSSUMS. The fraction drops very quickly when almost half of the links have negative signs.When the fraction of negative links reaches 0.7, almost all the nodes in the seed set are included with 'white' color.

We also studied the influence spread achieved by the heuristics with varying size of the seed set. No significant difference between them was observed on the original graphs. This is due to the fact that the number of negative links in both graph are less compared to the number of positive links (14.70% and 22.60% for Epinions and Slashdot, respectively). It follows from the results of our previous experiments (Figure 2) that the heuristics do not differ significantly when the fraction of negative links is less than 0.5. Therefore we flipped the sign of all the edges in the original graph so that there are 14.70% and 22.60% positive links in Epinions and Slashdot, respectively. The results of influence spreads achieved in these 'flipped' datasets are shown in Figure V. Note that OSSUMS outperforms the other two heuristics by a huge margin in both cases. Again, this is consistent with our earlier claim that significant advantage is observed with OSSUMS when the number of negative links dominate the positive links.

VI. Related Work

Information dissemination has been thoroughly studied on unsigned networks [3], [14], [15], [16]. Among the models that have been proposed Independent Cascade Model [3] has been studied extensively. Computing the exact expectation of influence spread has been shown to be #P-hard [14]. Typically thousands of Monte Carlo simulations are run to estimate the influence spread. An approximate analytical solution was proposed [2] in the form of a Unified Model that covers computation of expected influence for several models including



(a) Epinions

(b) Slashdot

Fig. 4. Influence spread achieved by varying size of the seed-set by the heuristics in the datasets after flipping the signs of edges.

ICM. However, the solution is specific to unsigned graph with a single infection. We extend the formula for ICM in signed networks with two competing infections.

Diffusion of multiple cascades has been the focus of [11], [19]. [20] studied the diffusion of multiple cascades and their interactions. Instead we study the spread of competing infections, where a node can be infected by only one of the infections prevalent in the network. Competing cascades have been studied in [9] from game theoretic perspective for maximizing the expected diffusion of an opinion against a competing one. [7] proposed influence maximization on a voter model on a social network with positive and negative links. They find the optimal seed-set for influence maximization on signed networks where opinions are flipped when flowing through a negative link. We assume a similar modeling of flipping infections over negative links, however, we show that is NP-hard to find the optimal seed set in our model. A similar model on unsigned network was proposed in [17], where opinions propagate according to ICM, and positive opinions get flipped randomly with certain probability. Unlike their model, the expected influence spread is not monotonic, making the influence maximization more difficult. Our model is same as IC-P [8], however our influence maximization of one infection we allow the inclusion of the opposite infection in the seed-set. Inclusion of opposite infection is important when the majority of links in the network are negative as demonstrated in our experiments.

VII. Conclusion

We studied the propagation of competing cascades in signed networks according to an extension of Independent Cascade Model where infections are flipped when propagated over negative links. We extended the Unified Model [2] to competing cascades in signed networks, that provided an approximate analytical solution to the problem of calculating the probability of infection of either of the two competing cascades for any node at any time t. We then defined SiNiMax, a novel signed network influence maximization problem for competing cascades. We proposed a heuristic, Online Seed-set Selection using Unified Model for Signed networks (OSSUMS), for this problem that diversifies the seed-set portfolio, taking advantage of both positive and negative relationships.

We validated our approach through experiments on real-world large-scale signed networks. We also quantified the effect of density of negative links on influence maximization. We demonstrated that no significant difference is observed among the performance of the heuristics when the majority of links is positive. However, we demonstrated that when the majority of links is negative, our heuristic significantly outperforms state-of-the-art heuristics for influence maximization. This result was a direct outcome of OSSUMS ability to incorporate seeds from both cascades into the influence maximization problem. We in fact observed that more seeds of the opposing cascade were selected for maximizing the cascade of interest when negative links are more prevalent in the network. As future work, we plan to extend our approach to multiple cascades and other influence models on signed networks.

Acknowledgment

This work is supported by Chevron U.S.A. Inc. under the joint project, Center for Interactive Smart Oilfield Technologies (CiSoft), at the University of Southern California.

References

[1] C. Chelmis and V. K. Prasanna, “The role of organization hierarchy in technology adoption at the workplace,” in *Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. ACM, 2013, pp. 8–15.

[2] A. Srivastava, C. Chelmis, and V. K. Prasanna, “Influence in social networks: A unified model?” in *Advances in Social Networks Analysis and Mining (ASONAM), 2014 IEEE/ACM International Conference on*. IEEE, 2014, pp. 451–454.

[3] D. Kempe, J. Kleinberg, and É. Tardos, “Maximizing the spread of influence through a social network,” in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2003, pp. 137–146.

[4] C. Budak, D. Agrawal, and A. El Abbadi, “Diffusion of information in social networks: Is it all local?” in *ICDM*, 2012, pp. 121–130.

[5] C. Chelmis, A. Srivastava, and V. K. Prasanna, “Computational models of technology adoption at the workplace,” *Social Network Analysis and Mining*, vol. 4, no. 1, pp. 1–18, 2014.

[6] J. Leskovec, D. Huttenlocher, and J. Kleinberg, "Signed networks in social media," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2010, pp. 1361–1370.

[7] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang, "Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships," in *Proceedings of the sixth ACM international conference on Web search and data mining*. ACM, 2013, pp. 657–666.

[8] D. Li, Z.-M. Xu, N. Chakraborty, A. Gupta, K. Sycara, and S. Li, "Polarity related influence maximization in signed social networks," *PloS one*, vol. 9, no. 7, p. e102199, 2014.

[9] S. Bharathi, D. Kempe, and M. Salek, "Competitive influence maximization in social networks," in *Internet and Network Economics*. Springer, 2007, pp. 306–311.

[10] A. Borodin, Y. Filmus, and J. Oren, "Threshold models for competitive influence in social networks," in *Internet and network economics*. Springer, 2010, pp. 539–550.

[11] X. He, G. Song, W. Chen, and Q. Jiang, "Influence blocking maximization in social networks under the competitive linear threshold model." in *SDM*. SIAM, 2012, pp. 463–474.

[12] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2001, pp. 57–66.

[13] W. Chen, W. Lu, and N. Zhang, "Time-critical influence maximization in social networks with time-delayed diffusion process," *arXiv preprint arXiv:1204.3074*, 2012.

[14] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2010, pp. 1029–1038.

[15] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2009, pp. 199–208.

[16] A. Goyal, W. Lu, and L. V. Lakshmanan, "Simpath: An efficient algorithm for influence maximization under the linear threshold model," in *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 211–220.

[17] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincon, X. Sun, Y. Wang, W. Wei, and Y. Yuan, "Influence maximization in social networks when negative opinions may emerge and propagate." in *SDM*, vol. 11. SIAM, 2011, pp. 379–390.

[18] D. Easley and J. Kleinberg, *Networks, crowds, and markets: Reasoning about a highly connected world*. Cambridge University Press, 2010.

[19] N. Pathak, A. Banerjee, and J. Srivastava, "A generalized linear threshold model for multiple cascades," in *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010, pp. 965–970.

[20] S. A. Myers and J. Leskovec, "Clash of the contagions: Cooperation and competition in information diffusion." in *ICDM*, vol. 12. Citeseer, 2012, pp. 539–548.