

Maximizing Contrasting Opinions in Signed Social Networks

Kaivalya Rawal* †

Arijit Khan*

*Nanyang Technological University, Singapore

†BITS-Pilani, India

Abstract—The classic influence maximization problem finds a limited number of influential seed users in a social network such that the expected number of influenced users in the network, following an influence cascade model, is maximized. The problem has been studied in different settings, with further generalization of the graph structure, e.g., edge weights and polarities, target user categories, etc. In this paper, we introduce a unique influence diffusion scenario involving a population that split into two distinct groups, with opposing views. We aim at finding the top- k influential seed nodes so to simultaneously maximize the adoption of two distinct, antithetical opinions in the two groups, respectively. Efficiently finding such influential users is essential in a wide range of applications such as increasing voter engagement and turnout, steering public debates and discussions on societal issues with contentious opinions. We formulate this novel problem with the voter model to simulate opinion diffusion and dynamics, and then design a linear-time and exact algorithm COSiNeMax, while also investigating the long-term opinion characteristics in the network. Our experiments with several real-world datasets demonstrate the effectiveness and efficiency of the proposed algorithm, compared to various baselines.

I. INTRODUCTION

A central characteristic of social networks is that it facilitates rapid dissemination of information among large groups of individuals [6]. Online social networks, such as *Facebook*, *Twitter*, *LinkedIn*, *Flickr*, and *Digg* are used for spreading ideas and messages. Users' behaviors and opinions are highly affected by their friends in social networks, which is defined as the *social influence*. Motivated by various real-world applications, e.g., viral marketing [9], social and political campaigning [8], social influence studies have attracted extensive research attention. The classic influence maximization problem [15], [9] identifies the top- k seed users in a social network such that the expected number of influenced users in the network, starting from those seeds and following an influence diffusion model, is maximized. The budget k on the seed set size usually depends on how many initial users the campaigner can directly influence by advertisements, re-tweets from “bots”, free samples and discounted prices.

In reality, societies are complex systems, and polarize into groups of individuals with dramatically opposite perspectives. This phenomenon is also evident in online social networks based on political affiliations, religious views, controversial topics, personal biases and preferences [12]. Therefore, each campaign is generally launched and promoted with certain target audience in mind, e.g., all Republican voters, people who prefer jazz over metal music, or Android over iPhones,

etc. Often, online campaigns have limited budgets and cannot afford to directly reach to all members of their target population. In such scenarios, it is desirable to minimize the number of seed users as permitted by the budget, while still maximizing the spread of the campaign in the target audience.

Furthermore, due to the existence of subgroups with differing views, relationships between social network users also include negative ones, such as foe, spite, and distrust relations. Indeed, *signed social networks* containing both positive and negative relationships are ubiquitous [21]. For example, in the explicit category, users can directly tag the polarity (positive or negative) to the relation between two users, e.g., *Epinions*, *Slashdot*, *Ebay*, and other online review and news forums. In the implicit category, the relationship polarities can be mined from the interaction data between users, such as, in *Twitter* a user u may support some users whom she follows (positive) and be against the others (negative). Following common sense and past literature on signed networks (including the *structural balance theory*) [4], [17], [18], we assume that positive relations carry the influence in a positive manner, that is, a user would more likely trust and adopt her friends' opinions. On the other hand, negative relations tend to carry influence in a reverse direction, i.e., if a user's foe chooses some opinion, the user would more likely be influenced to select the opposite one. Ignoring such relationship polarities between users and treating signed social networks as unsigned ones would result in over-estimation of positive influence spread, thereby leading to lower-quality solutions. *Social influence can be further complicated when competing campaigns are simultaneously spread over a signed social network. Therefore, influence and opinion dynamics in a signed social network is a critical problem that, unfortunately, remains pretty much open.*

In this work, we investigate a *novel* influence diffusion problem: **COSiNe** (**C**ontrasting **O**pinions **M**aximization in a **S**igned **S**ocial **N**etwork). We find a limited influential seed nodes which maximize the adoption of two distinct, antithetical opinions in two non-overlapping user groups with opposing views. The objective behind such influence maximization is to create awareness in a population by improving the quality of the debate on naturally contentious issues.

• **Applications.** An ideal application of our problem would be to increase awareness about infrequently discussed issues that are nonetheless controversial (such as capital punishment, nuclear energy, or affirmative action) — in a target population that naturally splits into two distinct ideological groups (such as democrats and republicans); in a forum that extensively debates topics and proposes mutually agreeable solutions based on compromise, diversity, and inclusion (such as the United States Senate or House of Representatives). Contrary to initial

This work was done when the first author was interning at Nanyang Technological University, Singapore.

expectations, polarization of opinions and increased conflict can often be beneficial [14], [20], [11], [1], as discussed in the following.

The benefit of conflicting opinions among collaborators has been clearly observed in Wikipedia. Controversial articles such as those on the Syrian Civil War, Israel/Palestine, or George W. Bush attract a higher number of edits. Higher polarization in the contributing community is associated with higher article quality for a broad range of articles – from politics to science and social issues [20]. Increased diversity is often correlated also with greater business performance. Similarly, disagreements amongst co-workers have been found to improve the decision making capabilities at the organisation level. Thus, encouraging different opinions about the same topic can be leveraged to improve the productivity of the organisation [11]. When dealt with correctly, such differences in thought and opinions are a force for good.

Lastly, we illustrate an example from the world of politics that is most similar to our “ideal” application scenario. Unlike the American presidential system, in countries based upon the Westminster parliamentary system, there is an appointed head of government, different from the head of the state, and an appointed head of opposition. This balance between the government and the opposition is considered integral to the success of a functioning democracy in diverse countries such as in Britain and in India [1]. An equivalent analysis was made for the political system in the United States of America in 1950 by the American Political Science Association [14] which recommended a stronger two party system in order to strengthen the democratic process. Both these analyses point to the importance of opposition in political discourse, and go on to show that policies being enacted and implemented benefit from engagement, and even opposition. Meaningful discourse and spirited debate requires people who inherently hold opposing beliefs on a given issue, and thus maximizing opposing influences can be beneficial for a legislative body from the point of view of the general population.

• **Challenges and contributions.** Contrasting opinions maximization, as required in our problem setting, is a non-trivial one. First, one must employ an influence cascade model that has properties different from those for commercial, *one-time product purchasing* based marketing strategies. For example, people’s opinions change over time; thus, activation based models, such as independent cascade (IC) and linear threshold (LT) models [15] are less appropriate in political contexts. Second, in reality a signed social network might not be perfectly balanced [18], that is, there may not exist a partition V_1, V_2 of the node set V , such that all edges with V_1 and V_2 are positive and all edges across V_1 and V_2 are negative. Such a network does not follow the social balance theory, and adds more complexity to the social influence cascade.

In this work, we employ the *voter model* [7], [13], [10], [18] to characterize influence diffusion in the two population groups of a social network. We define our model such that opposite influences, when applied on the same user, cancel

each other, leading to a decay in the influence strength on any given user. Our model does not mandate that a user’s choice be frozen upon one-time activation, explicitly allowing the user to switch opinions at later times. Moreover, voter model, being a stochastic one (it has a random walk based interpretation, which will be introduced in § II), can deal with signed networks that are not perfectly balanced. We then define our novel COSiNe problem (contrasting opinions maximization), and design an efficient, exact solution.

The main contributions of this paper are as follows.

- We study the novel problem (COSiNe) of finding the top- k seed nodes that maximize the adoption of two distinct, antithetical opinions in two given non-overlapping sets of target users, respectively, in a signed social network. We adapt the voter model to formulate our problem in §II.
- We design a linear-time, exact solution (COSiNeMax) for our problem. We demonstrate its correctness in §III.
- We further characterize two different long-term opinion dynamics in a signed social network under extreme scenarios, and investigate how our proposed method, COSiNeMax finds the seed nodes intelligently under such extreme situations (§IV).
- We conduct a thorough experimental evaluation with several real-world signed social networks to demonstrate the effectiveness and efficiency of our algorithm, compared to various baseline methods (§V).

II. PRELIMINARIES

We model a social network as a signed, directed graph with edge weights: $G = (V, E, \mathbf{A})$, where V is the set of nodes (users), $E \subseteq V \times V$ is the set of directed edges (links, connections, follower/followee relations, etc.), and \mathbf{A} is the weighted adjacency matrix with $A_{ij} \neq 0$ when the edge $(i, j) \in E$, with A_{ij} being the weight of the edge (i, j) . The weight A_{ij} represents the strength of j ’s influence on i . Moreover, as we consider a signed graph, the adjacency matrix \mathbf{A} may contain negative entries. A positive entry A_{ij} indicates a positive relation, i.e., i considers j as a friend or i trusts j , whereas a negative entry A_{ij} denotes a negative relation, that is, i considers j as a foe, or i distrusts j . The absolute value $|A_{ij}|$ represents the strength of this positive or negative relation — the higher, the stronger. We further denote by \mathbf{A}^+ and \mathbf{A}^- the (unsigned) matrices with only positive and negative entries of \mathbf{A} , respectively. Thus, $\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^-$.

A. Information Diffusion Model

The voter model was first introduced in [13], [7] to investigate territorial conflicts between two species and more abstractly, the properties of infinite systems of stochastic processes. It was then studied for maximizing influence in unsigned networks [10] and over signed networks [18]. We update the model from prior attempts in order to more naturally simulate the spread of *two contrasting ideas*, O_1 and O_2 , *simultaneously* in the same network.

We associate with each node a floating point value C in the range $[-1, 1]$, that probabilistically determines the node’s adopted idea O_1 or O_2 . The diffusion happens at discrete time

steps, and the C value at every node can change with each time step. The opinion or idea adopted by node i at time step t is represented by $C_t(i)$: $C_t(i) \rightarrow 1$ implies that the user is likely to adopt the idea O_1 at time step t , whereas $C_t(i) \rightarrow -1$ denotes that the user is likely to adopt the idea O_2 at time step t . In particular, the probability of node i adopting idea O_1 at time t is defined as $p(O_1) = \frac{1+C_t(i)}{2}$, and the probability of i adopting idea O_2 at time t is $p(O_2) = \frac{1-C_t(i)}{2}$. The two probabilities are defined so that they always sum up to one. In our voter model, each node starts uninfluenced in the beginning, i.e., $C_t = 0$ at time $t = 0$, except those nodes being influenced as seed nodes for ideas O_1 or O_2 by the campaigner. For seed nodes, $C_0 = 1$ and $C_0 = -1$, respectively.

At every time step t , each node $i \in V$ adopts the idea of its outgoing neighbour $j \in V$ with probability $p = \frac{|A_{ij}|}{\sum_l |A_{il}|}$ if $A_{ij} > 0$, and adopts the opposite idea if $A_{ij} < 0$. Formally, $C_t(i)$

$$\begin{aligned} &= \sum_{j \in V} \left(\frac{A_{ij}^+}{\sum_{l \in V} |A_{il}|} C_{t-1}(j) \right) - \sum_{j \in V} \left(\frac{A_{ij}^-}{\sum_{l \in V} |A_{il}|} C_{t-1}(j) \right) \\ &= \sum_{j \in V} \frac{A_{ij}^+ - A_{ij}^-}{\sum_{l \in V} |A_{il}|} C_{t-1}(j) \end{aligned} \quad (1)$$

There is also an alternative, *random walk* interpretation of this voter model [18]. In this interpretation, we consider a walk across the graph that starts at an arbitrary node u . At each time step, from the current node i , an outgoing edge $i \rightarrow j$ is chosen with probability $p = \frac{|A_{ij}|}{\sum_l |A_{il}|}$ for the random walk. This walk is deemed to terminate at time t on some node v . Then, according to the voter model, $C_t(u) = C_0(v)$ if the path $u \rightarrow \dots \rightarrow v$ has an even number of negative edges (a *positive path*), and $C_t(u) = -C_0(v)$ if the path has an odd number of negative edges (a *negative path*).

By defining the voter model this way, opposite influences on a particular node tend to “cancel” out. The voter model also allows the opinion of a user to flip between two contrasting ideas, based on her neighbors’ influences. Thus, our voter model is different from one-time, activation-based influence propagation models (e.g., independent cascade (IC) and linear threshold (LT) models [15]), and we employ it to study opinion diffusion and formation in online signed social networks.

B. Problem Statement

Two non-overlapping groups V_1 and V_2 among the social network users are given as an input to our problem, such that, $V_1 \cap V_2 = \phi$ and $V_1 \cup V_2 \subseteq V$. The campaigner aims at influencing all nodes in V_1 with the idea O_1 , and all nodes in V_2 with the idea O_2 . Clearly, the users outside both the groups V_1 and V_2 have no business value to the campaigner.

We define an *opinion vector* \mathbf{C}_t , according to the opinions of all the nodes in our network at any specific time t . Thus, for a network with $|V| = n$ nodes:

$$\mathbf{C}_t = \begin{bmatrix} C_t(0) \\ C_t(1) \\ \vdots \\ C_t(n-1) \end{bmatrix} \quad (2)$$

The voter model can be described in matrix form in terms of the opinion vector and a *transition matrix* $\mathbf{P} = \mathbf{D}^{-1}\mathbf{A}$. Here, \mathbf{D} is a diagonal matrix that consists of all entries of $(\mathbf{A}^+ + \mathbf{A}^-) \cdot \mathbf{1}$ in its diagonal. From Equation 1, we get:

$$\begin{aligned} C_t(i) &= \sum_{j \in V} \frac{A_{ij}^+ - A_{ij}^-}{\sum_{l \in V} |A_{il}|} C_{t-1}(j) \\ \implies C_t(i) &= \sum_{j \in V} \frac{A_{ij}}{\sum_{l \in V} |A_{il}|} C_{t-1}(j) \\ \implies \mathbf{C}_t &= \mathbf{D}^{-1} \mathbf{A} \mathbf{C}_{t-1} = \mathbf{P} \mathbf{C}_{t-1} = \mathbf{P}^t \mathbf{C}_0 \end{aligned} \quad (3)$$

Similar to the opinion vector, we define a partition vector $\boldsymbol{\rho}$ to describe two target populations V_1 and V_2 . We define element ρ_i in this vector, for each node $i \in V$, as below:

$$\rho_i = \begin{cases} +1 & \dots & \text{if } i \in V_1 \\ -1 & \dots & \text{if } i \in V_2 \\ 0 & \dots & \text{if } i \in V \wedge i \notin (V_1 \cup V_2) \end{cases} \quad (4)$$

The effectiveness ϵ_t of the advertising campaign across both target populations can now be measured by using the scalar product formula $\epsilon_t = \boldsymbol{\rho}^T \cdot \mathbf{C}_t$. This promotes opinion O_1 in partition V_1 and opinion O_2 in partition V_2 , while also penalising the reverse situation, that is, O_1 in V_2 and O_2 in V_1 . The formulation correctly ignores the opinions of the nodes that do not belong in either V_1 or V_2 , that the campaigner is agnostic towards. It is worth noting that ϵ_t is a function of three parameters. (1) Future time step t : input to the problem, (2) $\boldsymbol{\rho}$: which defines two non-overlapping target groups and is provided as an input to the problem, and (3) C_0 : the seed set that needs to be determined.

We consider budget k on the number of seed nodes, which is an input parameter. We are now ready to define our problem.

Problem 1. [COSiNe] Given a signed, directed graph with edge weights: $G = (V, E, \mathbf{A})$, a future time step $t > 0$, $\boldsymbol{\rho}$ vector which defines two non-overlapping target groups V_1, V_2 for two contrasting ideas O_1 and O_2 , respectively, and a budget k on the total number of seed nodes, find the top- k seed nodes, together with their advertisement types (between O_1 and O_2), such that the effectiveness $\epsilon_t = \boldsymbol{\rho}^T \cdot \mathbf{C}_t$ of the campaign is maximized.

III. ALGORITHM: SHORT-TERM OPINIONS MAXIMIZATION

In this section, we design an efficient and *exact* algorithm for the COSiNe problem and with a given, finite time step $t > 0$. We refer to this as “short-term” since t could be small and we do not look for characteristics of the opinion dynamics as $t \rightarrow \infty$. The long-term case will be discussed in § IV.

Our strategy for finding the most influential seed nodes is as follows. We compute the amount of influence of each node on the rest of the network at time t . It turns out that, according to our voter model, selecting the top- k individually most influential nodes as the seed nodes is equivalent to the set of k nodes with the highest influence. The correctness of our algorithm is proved in § III-A.

Our complete algorithm, COSiNeMax is given in Algorithm 1. To find the individual influence power $\epsilon(i)$ of each

node $i \in V$, we simulate random walks in the reverse direction of the actual influence diffusion (Lines 1-14). The number of walks terminating at a specific node can thus be used as a measure of the node's ability to influence other nodes, based on our voter model. We next select the top- k nodes having the maximum absolute influence power individually as the seed set (Lines 15-37). Furthermore, for a seed node j , if $\epsilon(j)$ is positive, it is influenced with idea O_1 ; otherwise the seed node is influenced with O_2 (Lines 29-33).

A. Proof of Correctness

We prove the correctness of Algorithm 1 in two steps. First, we show that the aggregate of the individual influence of k nodes is identical to the influence strength of the set consisting of the same k nodes together (Theorem 1). Second, we demonstrate that the seed set formed by the top- k nodes as selected by Algorithm 1 is indeed the best seed set given inputs G , t , k , and ρ (Theorem 2).

Theorem 1. *Let $\epsilon_t = \rho^T \cdot \mathbf{C}_t$ be the total influence of a seed set Ω consisting of k nodes. We denote by $\epsilon_t(i)$ the individual influence of a node $i \in \Omega$. Then, $\epsilon_t = \sum_{i \in \Omega} \epsilon_t(i)$.*

Proof. We denote by Ω the seed set with k nodes. The subset of seed nodes influenced by the idea O_1 is denoted as Ω^+ , whereas the subset of seed nodes influenced by the idea O_2 is denoted as Ω^- . Clearly, $\Omega_1 \cap \Omega_2 = \phi$ and $\Omega_1 \cup \Omega_2 = \Omega$. Let ϵ_t be the total influence by the seed set Ω , whereas we represent by $\epsilon_i(t)$ the individual influence when the seed set consists of the single node $i \in \Omega$.

Consider three vectors e_1 , e_2 , and e_i , each having dimensionality $|V|$. They represent various subsets of Ω : e_i consists of $|V| - 1$ zeros, with only the i -th element being ± 1 (depending on whether i has been influenced with idea O_1 or O_2 , respectively), representing the singleton set $\{i\}$. Analogously, e_1 consists of $+1$ corresponding to all nodes in the set Ω_1 , and e_2 consists of -1 for all nodes in the set Ω_2 . The rest of the elements in e_1 and e_2 are zeros. Formally,

$$e_1(j) = \begin{cases} 0 & \text{if } j \notin \Omega_1 \\ +1 & \text{if } j \in \Omega_1 \end{cases} \quad e_2(j) = \begin{cases} 0 & \text{if } j \notin \Omega_2 \\ -1 & \text{if } j \in \Omega_2 \end{cases}$$

$$e_i(j) = \begin{cases} 0 & \text{if } j \neq i \\ +1 & \text{if } j = i, j \in \Omega_1 \\ -1 & \text{if } j = i, j \in \Omega_2 \end{cases} \quad (5)$$

Thus, $e = e_1 + e_2$ is the vector denoting the seed set $\Omega = \Omega_1 \cup \Omega_2$. Next, we derive the following.

$$\begin{aligned} \epsilon &= \rho^T \cdot \mathbf{C}_t = \rho^T \cdot (\mathbf{P}^t e) \quad \triangleright \text{Following Equation 3} \\ &= \rho^T \cdot \mathbf{P}^t (e_1 + e_2) = \rho^T \cdot \mathbf{P}^t (\sum_{i \in \Omega_1} (e_i) + \sum_{i \in \Omega_2} (e_i)) \\ &= \sum_{i \in \Omega} (\rho^T \mathbf{P}^t e_i) = \sum_{i \in \Omega} (\rho^T \mathbf{C}_t(i)) \quad \triangleright \text{By Equation 3} \\ &= \sum_{i \in \Omega} \epsilon_i \end{aligned} \quad (6)$$

Hence, the theorem.

Theorem 2. *The seed set Ω , consisting of the top- k individually most influential nodes as selected by Algorithm 1, is the optimal seed set having size k .*

Proof. Notice that Algorithm 1 selects the top- k individually most influential nodes into the seed set Ω . Therefore, the

Algorithm 1 COSiNeMax: Maximize Contrasting Opinions

Require: Signed graph $G = (V, E, \mathbf{A})$; time step $t > 0$; ρ vector to define two non-overlapping target groups V_1, V_2 for two contrasting ideas O_1, O_2 , respectively; budget k

Ensure: Set Ω of top- k nodes, with their advertisement types (between O_1 and O_2), that maximizes $\epsilon_t = \rho^T \cdot \mathbf{C}_t$

```

1:  $\mathbf{P} = \mathbf{D}^{-1} \mathbf{A}$  ▷ Transition Matrix of  $G$ 
2:  $\epsilon \leftarrow [0, 0, 0, \dots, 0]$  ▷ Initialise row vector of size  $|V|$ 
3: for  $i \leftarrow 1, |V|$  do
4:   if  $i \in V_1$  then
5:      $\epsilon[i] \leftarrow +1$ 
6:   else if  $i \in V_2$  then
7:      $\epsilon[i] \leftarrow -1$ 
8:   else
9:      $\epsilon[i] \leftarrow 0$ 
10:  end if
11: end for
12: for  $i \leftarrow 1, t$  do
13:    $\epsilon = \epsilon \cdot \mathbf{P}$ 
14: end for
15:  $\Omega \leftarrow \Phi$  ▷  $\Omega$  is a set of tuples  $\langle i \in V, \tau(i) \rangle$ 
16:  $\tau(i)$  denotes the individual influence of node  $i$ 
17: for  $j \leftarrow 1, |V|$  do
18:   if  $\text{size}(\Omega) \leq k$  then
19:      $\text{insert}(\Omega, \langle j, |\epsilon[j]| \rangle)$ 
20:     if  $\epsilon(j) > 0$  then
21:        $\text{Opinion}(j) \leftarrow O_1$ 
22:     else
23:        $\text{Opinion}(j) \leftarrow O_2$ 
24:     end if
25:   else
26:      $\langle i, \tau(i) \rangle \leftarrow \min(\Omega)$  ▷ min is based on  $\tau()$  values
27:     if  $|\epsilon[j]| > \tau(i)$  then
28:        $\text{remove}(\Omega, \langle i, \tau(i) \rangle)$ 
29:        $\text{insert}(\Omega, \langle j, |\epsilon[j]| \rangle)$ 
30:       if  $\epsilon(j) > 0$  then
31:          $\text{Opinion}(j) \leftarrow O_1$ 
32:       else
33:          $\text{Opinion}(j) \leftarrow O_2$ 
34:       end if
35:     end if
36:   end if
37: end for
38: return  $\Omega, \text{Opinion}(i : i \in \Omega)$  ▷ Optimal seed nodes, with their advertisement types between  $O_1$  and  $O_2$ 

```

following holds: $\epsilon_j \geq \epsilon_i$ for all nodes $i, j \in V$, such that $j \in \Omega$ and $i \notin \Omega$.

We demonstrate that for any other seed set Ω' , such that $\Omega' \neq \Omega$, $|\Omega'| = |\Omega|$ cannot have more influence than that of Ω . Let us define $\omega' = \Omega' \setminus \Omega$, $\omega = \Omega \setminus \Omega'$, and $o = \Omega' \cap \Omega$. Note that since the size of both Ω and Ω' is k , $|\omega'| = |\omega|$.

We prove by contradiction: Following Theorem 1, and if possible, we assume that $\sum_{i \in \Omega'} \epsilon_i > \sum_{j \in \Omega} \epsilon_j$. Then, we get:

$$\begin{aligned} &\sum_{i \in \Omega'} \epsilon_i > \sum_{j \in \Omega} \epsilon_j \\ \implies &\sum_{i \in \omega' \cup o} \epsilon_i > \sum_{j \in \omega \cup o} \epsilon_j \\ \implies &\sum_{i \in \omega'} \epsilon_i + \sum_{i \in o} \epsilon_i > \sum_{j \in \omega} \epsilon_j + \sum_{j \in o} \epsilon_j \quad (7) \\ \implies &\sum_{i \in \omega'} \epsilon_i > \sum_{j \in \omega} \epsilon_j \\ \implies &\exists (i \in \omega', j \in \omega) \quad \text{such that} \quad \epsilon_i > \epsilon_j \\ \implies &\exists (i \notin \Omega, j \in \Omega) \quad \text{such that} \quad \epsilon_i > \epsilon_j \end{aligned}$$

This contradicts that Algorithm 1 selects the top- k individually most influential nodes into the seed set Ω . Hence, the theorem.

B. Time Complexity

The slowest step in our method is random walk simulation (Lines 12-14), that requires $\mathcal{O}(|E|t)$ time. Thus, the time complexity of our algorithm is $\mathcal{O}(|E|t)$, which is linear in the size of the input graph. For details, we refer to [19].

IV. LONG-TERM OPINIONS FORMULATION

We now turn our attention to the long-term scenario, that is, opinion dynamics as $t \rightarrow \infty$. In particular, we consider two extreme scenarios with respect to the two non-overlapping groups V_1 and V_2 in the signed social network. For simplicity, in this section we shall assume that $V_1 \cup V_2 = V$ and the graph is strongly connected.

• **Socially balanced partitions:** With respect to partitions V_1, V_2 , all intra-partition edges are positive, and all inter-partition edges are negative. • **Socially anti-balanced partitions:** With respect to partitions V_1, V_2 , all intra-partition edges are negative, and all inter-partition edges are positive.

Remarks. First, even though most real-world datasets do not exactly fall under the above two categories, a real-world network could resemble one of them. For example, we observe that the *Tagged* dataset that we use in our experiments, has more than three times as many positive inter-partition edges than all other kinds of edges combined, thereby making these partitions close to socially anti-balanced partitions. By analyzing the long-term opinion dynamics for the two categories, we demonstrate how intelligently our algorithm finds the seed nodes even under such extreme situations. Second, we employ our algorithm, COSiNeMax in all scenarios, as its optimality has been proved in §III-A irrespective of future time step t , graph structures, and node partitions.

For ease of discussion, we define a *signed path* in a signed, directed social network as a sequence of nodes with the edges being directed from each node to the following one. The *length of the path* is the total number of directed edges in it. The *sign of a path* is positive if there is an even number of negative edges along the path; otherwise the sign of a path is negative.

A. Socially Balanced Partitions

Recall that the campaigner’s objective is as follows: At time step t , all nodes in V_1 will adopt opinion O_1 , and nodes in V_2 will adopt opinion O_2 . We next show that *if the input partitions are socially balanced, then by following our algorithm, at $t \rightarrow \infty$, indeed nodes in V_1 will adopt opinion O_1 and nodes in V_2 will adopt O_2 .* To prove this, it is easy to verify that all paths that begin and end in the same partition have positive signs (due to even number of negative edges on those paths). Analogously, all paths that begin in one partition and end in the other partition must have negative signs because of odd number of negative edges on them. This has two implications.

First, COSiNeMax will select all seed nodes of O_1 only from the users in V_1 , and all seeds for O_2 only from V_2 . This is because in Lines 4-7 of Algorithm 1, all nodes in V_1 starts as positive, and in partition V_2 all nodes starts as negative (at $t = 0$). Now, repeated multiplications with the transition matrix \mathbf{P} (Lines 12-14) can be considered as a union of random walks. Therefore, at any arbitrary future time step t , all nodes in V_1 would remain positive, because all random walks starting at V_1 and also ending at V_1 must consist of only positive paths. Similarly, at any arbitrary future time step t , all nodes in V_2 would remain negative. Now, in Lines 29-33, the seed nodes are influenced based on their final sign, that is, if positive then

influenced with opinion O_1 , and otherwise with opinion O_2 . This concludes that the seed nodes for O_1 will only be selected from group V_1 , and those for O_2 will be picked only from V_2 .

Second, *for socially balanced partitions, if all seeds of O_1 are from V_1 , and all seeds for O_2 are from V_2 , then at $t \rightarrow \infty$, nodes in V_1 will adopt opinion O_1 and nodes in V_2 will adopt O_2 .* This holds because each path from any seed in V_1 to some other node in V_1 will always be a positive path, thereby carrying the same opinion as that of the seed (i.e., O_1), whereas every path from a seed in V_2 to some other node in V_1 will be a negative path, thereby carrying the opposite opinion to that of the seed (i.e., also O_1).

B. Socially Anti-balanced Partitions

We show that *if all seeds of O_1 are from V_1 , all seeds for O_2 are from V_2 , and when $t \rightarrow \infty$, then anti-balanced partitions switch opinions between O_1 and O_2 at even and odd time steps, respectively. (1) Even time steps.* For even time steps, we consider paths of even lengths. Among such paths, all paths that begin and end in the same partition have positive signs (due to even number of negative edges), and all paths that begin and end in different partitions have negative signs (due to odd number of negative edges). Hence, this is identical to the situation in socially balanced partitions, and similar results hold. In other words, (1) COSiNeMax will select all seed nodes of O_1 only from the users in V_1 , and all seeds for O_2 only from V_2 . (2) For socially anti-balanced partitions, if all seeds of O_1 are from V_1 , and all seeds for O_2 are from V_2 , then at $t \rightarrow \infty$, with t being even, nodes in V_1 will adopt opinion O_1 and nodes in V_2 will adopt O_2 .

(2) *Odd time steps.* For odd time steps (with $t \rightarrow \infty$), one can follow similar reasoning to show that the opposite case arises. We now consider paths of odd lengths. Among such paths, all paths that end in the same partition as they began have negative signs (due to odd number of negative edges), and all paths that end in the opposite partition as they began have positive signs (due to even number of negative edges). This results in swapping of opinions for the two partitions, relative to the ones in an even time step.

Notice that COSiNeMax intelligently selects seed nodes: When the objective is to maximize the adoption of O_1 at V_1 and O_2 at V_2 in an odd time step, in anti-balanced partitions as $t \rightarrow \infty$, COSiNeMax will select all seed nodes of O_1 only from the users in V_2 , and all seeds for O_2 only from V_1 .

V. EXPERIMENTAL RESULTS

We show empirical results to demonstrate effectiveness and efficiency of our solution, and compare it with three baselines. We analyze sensitivity of COSiNeMax by varying several parameters, e.g., number of seed and targets, time steps.

A. Environment Setup

Our code (available at: github.com/COSiNeMax/COSiNeMax) is implemented in Python, using sparse matrix operations from the *scipy* library, and the experiments were performed on a single core of a 16GB, 1.8GHz, Intel i7-8550U processor. Each experimental result is averaged over 10 runs.

TABLE I: Dataset characteristics

Dataset	#Nodes	#Edges	#Positive Edges	#Negative Edges
<i>Epinions</i>	132 585	701 926	605 854 (86%)	96 072 (14%)
<i>GitHub</i>	44 914	44 100 700	26 185 530 (59%)	17 915 170 (41%)
<i>Tagged</i>	5 607 448	546 799 071	443 895 613 (81%)	102 903 458 (19%)

1) *Datasets*: We summarize our datasets in Table I, and more details about them are given in our extended version [19]. **(1) *Epinions***. This social network dataset is extracted from the product review website *epinions.com*, where users may trust or distrust others. It is a signed and directed network: A user trusting another is represented with an edge of weight +1, and distrusting another is denoted by weight -1. **(2) *GitHub***. The dataset (blog.github.com/2009-07-29-the-2009-github-contest) is extracted from an anonymized dataset of user-repository interactions on *github.com*, utilising information about users “watching” other’s repositories. We connect any two users in the network with a bidirectional edge if they watch the same repository, with edge weight inversely proportional to the number of watchers for that repository. The sign of this edge is positive if both nodes view more single-language repositories (or, both view more multi-language repositories), and negative otherwise. **(3) *Tagged***. Our largest real-life dataset is collected from the online social network *tagged.com*. The nodes are partitioned into V_1 and V_2 using anonymized gender metadata.

2) *Competing Methods*: We compare the proposed COSiNeMax method (Algorithm 1) with three baselines. **(1) *Random***. Uniformly at random selection of k seed nodes. **(2) *Degree***. The top- k nodes with the highest out-degrees. **(3) *Individual InfMax***. In this baseline approach, we follow the *voter model* over signed networks [18], however we consider each target set separately. That is, we first compute the top- $\lfloor k/2 \rfloor$ seed nodes so to maximize the spread of the idea O_1 in the target partition V_1 . Next, we find another top- $\lfloor k/2 \rfloor$ seed nodes that maximize the spread of the idea O_2 within the target set V_2 . Therefore, by comparing with the *Individual Influence Maximization* approach as described above, we demonstrate the improvements due to our algorithm COSiNeMax, which returns the top- k optimal seed nodes considering the spread of two contrasting ideas O_1 and O_2 simultaneously.

For each baseline, at $t = 0$ we target a seed node i with idea O_1 if $i \in V_1$, and with O_2 if $i \in V_2$.

3) *Parameters Setup*: We vary the number of seed nodes from 1% to 5%, and by default consider all nodes in the networks as targets. Sensitivity analysis experiments by varying the number of seeds and targets are given in [19]. We also conduct experiments for both short-term and long-term with 30 and 500 time steps, respectively.

4) *Evaluation Metrics*: We employ two metrics for the effectiveness measure.

Expected number of correctly influenced nodes. We compute the number of nodes influenced by idea O_1 in target partition V_1 , and by O_2 in target partition V_2 . Recall that the probability of node i adopting idea O_1 at time t is defined as $p(O_1) = \frac{1+C_t(i)}{2}$, and the probability of i adopting idea O_2 at time t is $p(O_2) = \frac{1-C_t(i)}{2}$. Here, $C_t(i) \in [-1, 1]$ is computed following Equation 3.

Moreover, we disregard weakly influenced nodes, i.e., node $i \in V_1$ when its $p(O_1)$ is less than a predefined threshold (0.5), and $i \in V_2$ when its $p(O_2)$ is less than a predefined threshold (0.5). Such a user is likely to be undecided between two opposite opinions on a specific issue. Formally, we report:

Expected number of correctly influenced nodes

$$= \sum_{i \in V_1, C_t(i) > 0} \left(\frac{1 + C_t(i)}{2} \right) + \sum_{i \in V_2, C_t(i) < 0} \left(\frac{1 - C_t(i)}{2} \right)$$

Influence percentage w.r.t. all targets as seeds. We also measure *campaign effectiveness constrained by a limited number of seeds*, with respect to the hypothetical scenario when all target nodes can be employed as seeds. We recall that in § II, the effectiveness of the campaign was formulated as $\rho^T \cdot \mathbf{C}_t$. This promotes opinion O_1 in partition V_1 and opinion O_2 in partition V_2 , while penalising the reverse situation, that is, O_1 in V_2 and O_2 in V_1 .

To better compare the aforementioned campaign effectiveness of each baseline and our proposed algorithm, we compare it to the case when all target nodes are assigned as seed nodes. At time step $t = 0$, the seeds are influenced with the respective idea of the target partition that they belong to. According to the voter model, opposite influences on the same node cancel each other out, thus there could be a decay with time in the magnitude of influence. Let us denote by T_t the campaign effectiveness at time step t in this scenario (i.e., when all target nodes were seed nodes at $t = 0$).

Finally, we report $(\frac{\rho^T \cdot \mathbf{C}_t}{T_t} \times 100)\%$ as the influence percentage w.r.t. all target nodes used as seed nodes.

B. Effectiveness Results

We present effectiveness results on three networks (Figure 1). We find that our designed COSiNeMax achieves higher expected number of influenced nodes than all three baselines. Notice that *Epinions* (Figure 1(a)) shows some reduction in the expected number of correctly influenced nodes with larger time steps till it saturates. Such reduction is not observed in *GitHub* and *Tagged*. This is due to higher sparsity of *Epinions*, with the presence of many separated components, each consisting of a few nodes. In such a sparse network, random walks from seed nodes initially influence a large number of nodes. However, this influence is unable to sustain at later time steps due to sparsity of the graph. In other words, the sparsity of the network prevents long random walks from returning to the same nodes, thereby reducing the influence over time.

When we compare the influence percentage (w.r.t. all targets as seeds) of each algorithm, COSiNeMax also outperforms all baselines (Figure 2). However, the peak value obtained in each dataset is different, with *Epinions* having the highest at 120%, *GitHub* having 55%, and *Tagged* at 40%. The sparsity of *Epinions* dissipates the total influence T_t very rapidly, reducing it by almost 75 % in the first time step itself. This quick decrease in influence is prevented with COSiNeMax by selecting the seed nodes more intelligently, thus achieving the peak value at higher than 100%.

The oscillatory plots of the baselines in *Tagged* (Figures 1(c), 2(c)) can be explained based on graph structure and

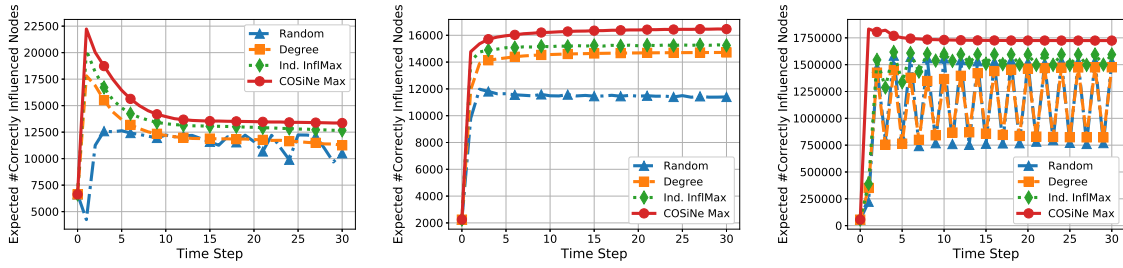


Fig. 1: Expected number of correctly influenced users for different time steps. Seeds are selected according to various algorithms.

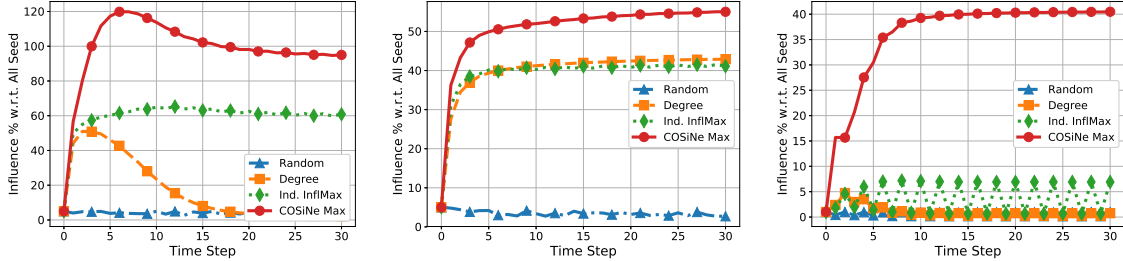


Fig. 2: Influence percentage w.r.t. ‘‘All Seed’’ for different time steps. Seeds are selected according to various algorithms. ‘‘All Seed’’ denotes the case when all target nodes are used as seeds, and influenced by the respective idea at $t = 0$ (this metric is defined in § V-A4).

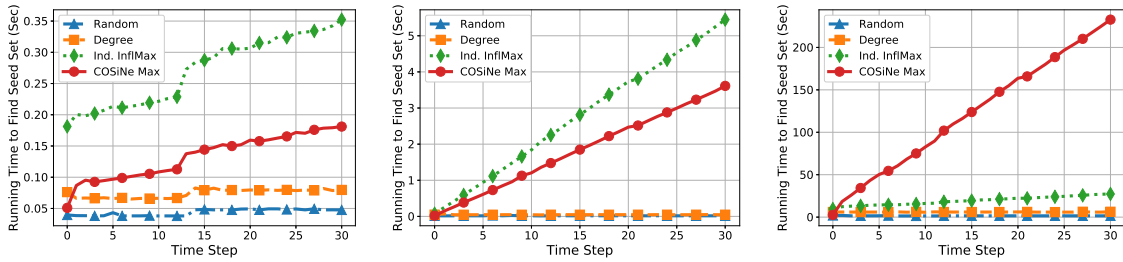


Fig. 3: Running time to find seed nodes according to various algorithms.

node partitions. *Tagged* has more than three times as many positive inter-partition edges than all other kinds of edges combined, thereby making these partitions close to socially anti-balanced partitions. Thus, if the seed nodes in the two partitions are not targeted by O_1 or O_2 intelligently, as it is done in case of baselines (see § V-A2), such oscillatory behaviour in influence spread arises. This is similar to the oscillatory behaviour discussed in § IV due to socially anti-balanced graph partitions. COSiNeMax is able to circumvent this problem by targeting all seed nodes in V_1 as O_1 when maximizing influence for even time steps, and as O_2 when maximizing influence for odd time steps.

C. Efficiency Results

We compare running time to find seed nodes by all algorithms in Figure 3. While time taken increases almost linearly with time steps for both COSiNeMax and Individual InfMax, it is evident that both Random and Degree are faster, and their seed set finding times are independent of input time step.

In case of Individual InfMax, the seed nodes are computed in two stages: once for opinion O_1 in the target set V_1 , and then for opinion O_2 in the target set V_2 . However, COSiNeMax holistically identifies all seed nodes in the entire graph. This explains why COSiNeMax is faster than Individual InfMax

over two smaller graphs. On the other hand, COSiNeMax requires more time than Individual InfMax over *Tagged*, which is a larger dataset and the complexity of performing random walks over entire graph dominates seed set finding time.

D. Longer-term Dynamics

We also study longer-term dynamics, with time steps from 0 to 500 (Figure 4). We find that all algorithms, except the Random baseline, achieves saturation over time, with no further variation in influence. The expected number of correctly influenced nodes and the influence percentage (w.r.t. all targets as seeds) in this saturated state are both higher for our COSiNeMax than the baselines.

VI. RELATED WORK

The classic influence maximization problem finds a limited number of seed users that generate the largest expected influence cascade in a social network. Influence maximization in the presence of a negative campaign was investigated in [2], which assumes that the later campaign has prior knowledge of rival side’s initial seed nodes. Bordin et. al. [3] analyzed the similar problem under the LT model; while [5] attempts at preventing the spread of an existing negative campaign in the network. However, as competitive new products from rival companies are often launched around the same time,

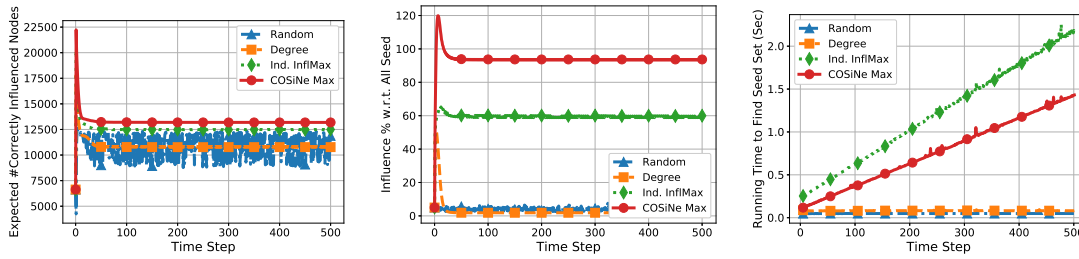


Fig. 4: Results on long-term opinions formation, *Epinions*, #seeds=1% of all users.

[16] considered influence maximization in the presence of multiple competing campaigners, who promote their products in a social network around the same time.

Our work is fundamentally different from prior literature. **First**, they generally consider activation based models (e.g., IC and LT) suitable for *one-time* product purchase. In contrast, our voter model allows users to switch opinions at later times based on their neighbors’ opinions. Thus, voter model is more suitable to study opinion diffusion and formation in online social networks. **Second**, although earlier works consider multiple competitive campaigns, different from our study they do not consider diffusion with both positive and negative edges in a *signed* social network. **Third**, due to the inherent complexity of IC, LT models and their variants, the problems investigated in those works are generally NP-hard and also #P-hard, while the voter model can solve our problem *exactly* in linear time.

With the prevalence of signed social networks, recent works investigated the problem of finding the seed set that maximizes positive influence, which is also known as positive influence maximization. [17] studied positive influence maximization under different extensions of IC and LT models. Li et al. [18] explored similar problem in a signed social network with voter model. Unlike ours, they do not aim at maximizing two contrasting opinions in two non-overlapping target regions. Moreover, in [18] all seed nodes can be influenced by only one type of idea, that is, for positive influence maximization, all seeds will be influenced by the positive idea. However, as demonstrated in our experiments, maximizing each influence *separately* (i.e., Individual InfMax) results in a sub-optimal solution compared to ours (i.e., COSiNeMax): We return optimal seed nodes considering the spread of two contrasting ideas *simultaneously*.

Due to lack of space, we discuss in [19] more related work on signed networks and polarization in social networks.

VII. CONCLUSIONS

We formulated and investigated the novel problem of contrasting opinions maximization in two distinct target groups, respectively, over a signed social network. Motivated by scenarios such as increasing voter engagement and turnout, steering public debates and discussions on societal issues with contentious opinions, we adapted the voter model to effectively study influence diffusion. We efficiently solved this problem, and designed an exact algorithm. We then empirically compared this algorithm with several baselines on three real-

world signed network datasets. Our analysis reveals that the proposed algorithm, COSiNeMax finds the seed set with the highest expected number of influenced nodes, and has the highest relative total influence. This behaviour is demonstrated over all datasets and for different variations of time steps, seed set budget, and target population size parameters. In future, it would be interesting to consider adaptive seeding, as opposed to one-time seeding, for even more effective short-term opinions maximization in a signed, social network.

REFERENCES

- [1] A. Beteille, “Democracy and It’s Institutions”, Oxford University Press, Chapter Government and Opposition, 2012.
- [2] S. Bharathi, D. Kempe, and M. Salek, “Competitive Influence Maximization in Social Networks”, WINE, 2007.
- [3] A. Borodin, Y. Filmus, and J. Oren, “Threshold Models for Competitive Influence in Social Networks”, WINE, 2010.
- [4] D. Cartwright and F. Harary, “Structural Balance: A Generalization of Heider’s Theory”, Psychological Review, vol. 63(5), 1956, pp. 277–293.
- [5] W. Chen et al., “Influence Maximization in Social Networks when Negative Opinions May Emerge and Propagate”. SDM, 2011.
- [6] W. Chen, L. V. S. Lakshmanan, and C. Castillo, “Information and Influence Propagation in Social Networks”, Morgan & Claypool, 2013.
- [7] P. Clifford and A. Sudbury, “A Model for Spatial Conflict”, Biometrika, vol. 60(3), 1973, pp. 581–588.
- [8] B. A. Conway, K. Kenski, and D. Wang, “The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary”, Journal of Computer-Mediated Communication, vol. 20(4), 2015, pp. 363–380.
- [9] P. Domingos and M. Richardson, “Mining the Network Value Customers”, KDD, 2001.
- [10] E. Even-Dar and A. Shapira, “A Note on Maximizing the Spread of Influence in Social Networks”, Internet & Network Econ”, 2007.
- [11] K. Ferrazzi, “The Benefits of Conflict at Work”, 2014, <http://fortune.com/2014/03/11/the-benefits-of-conflict-at-work>.
- [12] V. R. K. Garimella and I. Weber, “A Long-Term Analysis of Polarization on Twitter”, ICWSM, 2017.
- [13] R. A. Holley and T. M. Liggett, “Ergodic Theorems for Weakly Interacting Infinite Systems and the Voter Model”, Ann. Probab., vol. 3(4), 1975, pp. 643–663.
- [14] A. Schlesinger JR, “Toward A More Responsible Two-Party System: A Report”, American Political Science Association, 1950.
- [15] D. Kempe, J. Kleinberg, and E. Tardos, “Maximizing the Spread of Influence through Social Network”, KDD, 2003.
- [16] A. Khan, B. Zehnder, and D. Kossmann, “Revenue Maximization by Viral Marketing: A Social Network Host’s Perspective”, ICDE, 2016.
- [17] 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(7), 2014, pp. 1–12.
- [18] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang, “Influence Diffusion Dynamics and Influence Maximization in Social Networks with Friend and Foe Relationships”, WSDM, 2013.
- [19] K. Rawal and A. Khan, “Maximizing Contrasting Opinions in Signed Social Networks” (extended version), arXiv:1910.12017, 2019.
- [20] F. Shi, M. Teplitskiy, E. Duede, and J. A. Evans, “The Wisdom of Polarized Crowds”, Human Behaviour, 2019.
- [21] J. Tang, Y. Chang, C. Aggarwal, and et al., “A Survey of Signed Network Mining in Social Media”, ACM Comput. Surv., vol. 49(3), 2016.