Influence Diffusion Dynamics and Influence Maximization in Social Networks with Friend and Foe Relationships*

Yanhua Li[†], Wei Chen[§], Yajun Wang[§] and Zhi-Li Zhang[†]
[†]Dept. of Computer Science & Engineering, Univ. of Minnesota, Twin Cities, Minneapolis, MN, US
[§] Microsoft Research Asia, Beijing, China
{yanhua,zhzhang}@cs.umn.edu,{weic,yajunw}@microsoft.com

ABSTRACT

Influence diffusion and influence maximization in large-scale online social networks (OSNs) have been extensively studied because of their impacts on enabling effective online viral marketing. Existing studies focus on social networks with only friendship relations, whereas the foe or enemy relations that commonly exist in many OSNs, e.g., Epinions and Slashdot, are completely ignored. In this paper, we make the first attempt to investigate the influence diffusion and influence maximization in OSNs with both friend and foe relations, which are modeled using positive and negative edges on signed networks. In particular, we extend the classic voter model to signed networks and analyze the dynamics of influence diffusion of two opposite opinions. We first provide systematic characterization of both short-term and long-term dynamics of influence diffusion in this model, and illustrate that the steady state behaviors of the dynamics depend on three types of graph structures, which we refer to as balanced graphs, anti-balanced graphs, and strictly unbalanced graphs. We then apply our results to solve the influence maximization problem and develop efficient algorithms to select initial seeds of one opinion that maximize either its short-term influence coverage or long-term steady state influence coverage. Extensive simulation results on both synthetic and real-world networks, such as Epinions and Slashdot, confirm our theoretical analysis on influence diffusion dynamics, and demonstrate that our influence maximization algorithms perform consistently better than other heuristic algorithms.

Categories and Subject Descriptors

E.1 [**Data**]: Data Structures; H.3.3 [**Information Systems**]: Information Storage and Retrieval—*Information Search and Retrieval*

General Terms

Theory, Algorithms, Design

Keywords

Signed social networks, voter model, influence maximization

*This study was done partly when Yanhua Li was a summer intern at Microsoft Research Asia.

WSDM'13, February 4–8, 2013, Rome, Italy.

1. INTRODUCTION

As the popularity of online social networks (OSNs) such as Facebook and Twitter continuously increases, OSNs have become an important platform for the dissemination of news, ideas, opinions, etc. The openness of the OSN platforms and the richness of contents and user interaction information enable intelligent online recommendation systems and viral marketing techniques. For example, if a company wants to promote a new product, it may identify a set of influential users in the online social network and provide them with free sample products. They hope that these influential users could influence their friends, and friends of friends in the network and so on, generating a large influence cascade so that many users adopt their product as a result of such word-of-mouth effect. The question is how to select the initial users given a limited budget on free samples, so as to influence the largest number of people to purchase the product through this "word-of-mouth" process. Similar situations could apply to the promotion of ideas and opinions, such as political candidates trying to find early supporters for their political proposals and agendas, government authorities or companies trying to win public support by finding and convincing an initial set of early adopters to their ideas.

The above problem is referred to as the *influence maximization* problem in the literature, which has been extensively studied in recent years [8–10, 15, 16, 19, 23, 34, 36]. In these studies, several influence diffusion models are proposed to formulate the underlying influence propagation processes, including linear threshold (LT) model, independent cascade (IC) model, voter model, etc. A number of approximation algorithms and scalable heuristics are designed under these models to solve the influence maximization problem.

However, all existing studies only look at networks with positive (i.e., friend, altruism, or trust) relationships, where in reality, relationships also include negative ones, such as foe, spite or distrust relationships. In Ebay, users develop trust and distrust in agents in the network; In online review and news forums, such as Epinions and Slashdot, readers approve or denounce reviews and articles of each other. Some recent studies [11, 21, 22] already look into the network structures with both positive and negative relationships. As a common sense exploited in many existing social influence studies [8–10, 15, 19], positive relationships carry the influence in a positive manner, i.e., you would *more likely* trust and adopt your friends' opinions. In contrast, we consider that negative relationships often carry influence in a reverse direction — if your foe chooses one opinion or votes for one candidate, you would *more likely* be influenced to do the opposite. This echoes the principles that "the friend of my enemy is my enemy" and "the enemy of my enemy is my friend". Structural balance theory has been developed based on these assumptions in social science (see Chapter 5 of [13] and the references therein). We acknowledge that in real social networks, people's reactions to the influence from their friends or foes could be complicated, i.e., one could take the opposite opinion of what her foe suggests for one situation or topic, but may adopt the suggestion from the same person for a different topic, because she trusts her foe's expertise in that particular topic. In this study, we consider the influence diffusion for a single topic, where one always takes the opposite opinion of what her foe suggests. This is our first attempt to model influence diffusion in signed networks, and such topic-dependent simplification is commonly employed in prior influence diffusion studies on unsigned networks [8–10, 15, 16, 19]. Our work aims at providing a mathematical analysis on the influence diffusion dynamic incorporated with negative relationship and applying our analysis to the algorithmic problem of influence maximization.

1.1 Our contributions

In this paper, we extend the classic voter model [12, 18] to incorporate negative relationships for modeling the diffusion of opinions in a social network. Given an unsigned directed graph (digraph), the basic voter model works as follows. At each step, every node in the graph randomly picks one of its *outgoing* neighbors and adopts the opinion of this neighbor. Thus, the voter model is suitable to interpret and model opinion diffusions where people's opinions may switch back and forth based on their interactions with other people in the network. To incorporate negative relationships, we consider signed digraphs in which every directed edge is either positive or negative, and we consider the diffusion of two opposite opinions, e.g., black and white colors. We extend the voter model to signed digraphs, such that at each step, every node randomly picks one of its outgoing neighbors, and if the edge to this neighbor is positive, the node adopts the neighbor's opinion, but if the edge is negative, the node adopts the opposite of the neighbor's opinion (Section 2).

We provide detailed mathematical analysis on the voter model dynamics for signed networks (Section 3). For short-term dynamics, we derive the exact formula for opinion distribution at each step. For long-term dynamics, we provide closed-form formulas for the steady state distribution of opinions. We show that the steady state distribution depends on the graph structure: we divide signed digraphs into three classes of graph structures — balanced graphs, anti-balanced graphs, and strictly unbalanced graphs, each of which leads to a different type of steady state distributions of opinions. While balanced and unbalanced graphs have been extensively studied by structural balance theory in social science [13], the anti-balanced graphs form a new class that has not been covered before, to the best of our knowledge. Moreover, our long-term dynamics not only cover strongly connected and aperiodic digraphs that most of such studies focus on, but also weakly connected and disconnected digraphs, making our study more comprehensive.

We then study the influence maximization problem under the voter model for signed digraphs (Section 4). The problem here is to select at most k initial white nodes while all others are black, so that either in short term or long term the expected number of white nodes is maximized. This corresponds to the scenario where one opinion is dominating the public and an alternative opinion (e.g. a competing political agenda, or a new innovation) tries to win over supporters as much as possible by selecting some initial seeds to influence on. We provide efficient algorithms that find optimal solutions for both short-term and long-term cases. In particular, for long-term influence maximization, our algorithm provides a comprehensive solution covering weakly connected and disconnected signed digraphs, with nontrivial computations on influence coverage of seed nodes.

Finally, we conduct extensive simulations on both real-world and synthetic networks to verify our analysis and to show the effectiveness of our influence maximization algorithm (Section 5). The simulation results demonstrate that our influence maximization algorithms perform consistently better than all other heuristic algorithms. To the best of our knowledge, we are the first to study influence diffusion and influence maximization in signed networks, and the first to apply the voter model to this case and provide efficient algorithms for influence maximization under voter model for signed networks.

Due to space constraints, some of the proofs and additional materials are omitted and delegated to our technical report [24].

1.2 Related work

In this subsection, we discuss the topics that are closely related to our problem, such as: (1) influence maximization and voter model, (2) signed networks, and (3) competitive influence diffusion.

Influence maximization and voter model. Influence maximization has been extensively studied in the literature. The initial work [19] proposes several influence diffusion models and provides the greedy approximation algorithm for influence maximization. More recent works [8–10, 15, 16, 23, 34] study efficient optimizations and scalable heuristics for the influence maximization problem. In particular, the voter model is proposed in [12, 18], and is suitable for modeling opinion diffusions in which people may switch opinions back and forth from time to time due to the interactions with other people in the network. Even-Dar and Shapira [15] study the influence maximization problem in the voter model on simple unsigned and undirected graphs, and they show that the best seeds for long-term influence maximization are simply the highest degree nodes. As a contrast, we show in this paper that seed selection for signed digraphs are more sophisticated, especially for weakly connected or disconnected signed digraphs. More voter model related research is conducted in physics domain, where the voter model, the zero-temperature Glauber dynamics for the Ising model, invasion process, and other related models of population dynamics belong to the class of models with two absorbing states and epidemic spreading dynamics [1, 30, 38]. However, none of these works study the influence diffusion and influence maximization of voter model under signed networks.

Signed networks. The signed networks with both positive and negative links have gained attentions recently [3, 20–22]. In [21, 22], the authors empirically study the structure of real-world social networks with negative relationships based on two social science theories, i.e., balance theory and status theory. Kunegis et al. [20] study the spectral properties of the signed undirected graphs, with applications in link predictions, spectral clustering, etc. Borgs et al. [3] proposes a generalized PageRank algorithm [35] for signed networks with application to online recommendations, where the distrust relations are considered as adversarial or arbitrary user behaviors, thus the outgoing relations of distrusted users are ignored while ranking nodes. Our algorithm can also be viewed as a node ranking algorithm that generalizes the PageRank algorithm, by treating distrust links as generating negative influence rather than ignoring distrusted users' opinions, and thus our ranking method is different from [3]. Overall, none of the above work studies influence diffusion and influence maximization in signed networks.

Competitive influence diffusion. A number of recent studies focus on competitive influence diffusion and maximization [2, 4, 6, 7, 17, 29], in which two or more competitive opinions or innovations are diffusing in the network. Although they consider two or more competitive or opposing influence diffusions, they are all on unsigned networks, different from our study here on diffusion with both positive and negative relationships.

2. VOTER MODEL ON SIGNED NETWORKS

We consider a weighted directed graph (digraph) $G = (V, E, A)$, where V is the set of vertices, E is the set of directed edges, and A is the weighted adjacency matrix with $A_{ij} \neq 0$ if and only if $(i, j) \in E$, with A_{ij} as the weight of edge (i, j). The voter model was first introduced for unsigned graphs, with nonnegative adjacency matrices A's. In this model, each node holds one of two opposite opinions, represented by black and white colors. Initially each node has either black or white color. At each step $t \geq 1$, every node i randomly picks one outgoing neighbor j with the probability proportional to the weight of (i, j), namely $A_{ij} / \sum_\ell A_{i\ell}$, and changes its color to j's color. The voter model also has a random walk interpretation. If a random walk starts from i and stops at node j at step t, then i's color at step t is j's color at step 0.

In this paper, we extend the voter model to signed digraphs, in which the adjacency matrix A may contain negative entries. A positive entry A_{ij} represents that i considers j as a friend or i trusts j, and a negative A_{ij} means that i considers j as a foe or i distrusts j. The absolute value $|A_{ij}|$ represents the strength of this trust or distrust relationship. The voter model is thus extended naturally such that one always takes the same opinion from his/her friend, and the opposite opinion of his/her foe. Technically, at each step $t \geq 1$, i randomly picks one outgoing neighbor j with probability $|A_{ij}| / \sum_\ell |A_{i\ell}|$, and if $A_{ij} > 0$ (edge (i, j) is positive) then i changes its color to j's color, but if $A_{ij} < 0$ (edge (i, j) is negative) then i changes its color to the opposite of j's color. The random walk interpretation can also be extended for signed networks: if the t-step random walk from i to j passes an even number of negative edges, then i's color at step t is the same as j's color at step 0; while if it passes an odd number of negative edges, then i's color at step t is the opposite of j's color at step 0.

Table 1: Notations and terminologies

$G = (V, E, A)$, $\bar{G} = (V, E, \bar{A})$	G is a signed digraph, with signed adjacency matrix A and $\bar{G}$ is the unsigned version of G, with adjacency matrix $\bar{A}$
A^+, A^-	A^+ (resp. A^-) is the non-negative adjacency matrix representing positive (resp. negative) edges of G, with $A = A^+ - A^-$ and $\bar{A} = A^+ + A^-$.
$\mathbf{1}$, π, x_0, x_t, x, x_e, x_o	Vector forms. All vectors are $\|V\|$-dimensional column vectors by default; $\mathbf{1}$ is all one vector, π is the stationary distribution of an ergodic digraph $\bar{G}$; x_0 (resp. x_t) is the white color distribution at the beginning (resp. at step t); x is the steady state white color distribution; x_e (resp. x_o) is the steady state white color distribution for even (resp. odd) steps.
d, d^+, d^-, D	d, d^+, and d^- are weighted out-degree vectors of G, where $d = \bar{A}\mathbf{1}$, $d^+ = A^+\mathbf{1}$, and $d^- = A^-\mathbf{1}$; $D = diag[d]$ is the diagonal degree matrix filled with entries of d.
$P, \bar{P}$	$P = D^{-1}A$ is the signed transition matrix of G and $\bar{P} = D^{-1}\bar{A}$ is the transition probability matrix of $\bar{G}$.
$v_Z, \hat{v}_S, \hat{v}_{Z,S_Z}$	Given a vector v, a node set $Z \subseteq V$, v_Z is the projection of v on Z. Given a partition $S, \bar{S}$ of V, $\hat{v}_S$ is signed such that $\hat{v}_S(i) = v(i)$ if $i \in S$, and $\hat{v}_S(i) = -v(i)$ if $i \notin S$. Given a partition $S_Z, \bar{S}_Z$ of Z, $\hat{v}_{Z,S_Z}$ is taking the projection of v on Z first, then negating the signs for entries in $\bar{S}_Z$.
$I, \hat{I}_S, B_Z$	I is the identity matrix. $\hat{I}_S = diag[\hat{\mathbf{1}}_S]$ is the signed identity matrix. B_Z is the projection of a matrix B to $Z \subseteq V$.

Given a signed digraph $G = (V, E, A)$, let $G^+ = (V, E^+, A^+)$ and $G^- = (V, E^-, A^-)$ denote the unsigned subgraphs consisting of all positive edges E^+ and all negative edges E^-, respectively, where A^+ and A^- are the corresponding non-negative adjacency matrices. Thus we have $A = A^+ - A^-$. Similar to unsigned digraphs, G is *aperiodic* if the greatest common divisor of the lengths of all cycles in G is 1, and G is *ergodic* if it is strongly connected and aperiodic. A *sink component* of a signed digraph is a strongly connected component that has no outgoing edges to any nodes outside the component. When studying the long-term dynamics of the voter model, we assume that all signed strongly connected components are ergodic. We first study the case of ergodic graphs, and then extend it to the more general case of weakly connected or disconnected graphs with ergodic sink components. Table 1 provides notations and terminologies used in the paper.

3. ANALYSIS OF VOTER MODEL DYNAMICS ON SIGNED DIGRAPHS

In this section, we study the short-term and long-term dynamics of the voter model on signed digraphs. In particular, we answer the following two questions.
(i) Short-term dynamics: Given an initial distribution of black and white nodes, what is the distribution of black and white nodes at step $t > 0$?
(ii) Convergence of voter model: Given an initial distribution of black and white nodes, would the distribution converge? If so, what is the steady state distribution of black and white nodes?

3.1 Short-term dynamics

To study voter model dynamics on signed digraphs, we first define the *signed transition matrix* as follows.

Definition 1 (Signed transition matrix). *Given a signed digraph $G = (V, E, A)$, we define the* signed transition matrix *of G as $P = D^{-1}A$, where $D = diag[d_i]$ is the diagonal matrix and $d_i = \sum_{j \in V} |A_{ij}|$ is the weighted out-degree of node i.*

Next proposition characterizes the dynamics of the voter model at each step using the *signed transition matrix*.

Proposition 1. *Let $G = (V, E, A)$ be a signed digraph and denote the initial white color distribution vector as x_0, i.e., $x_0(i)$ represents the probability that node i is white initially. Then, the white color distribution at step t, denoted by x_t can be computed as*

$$x_t = P^t x_0 + (\sum_{i=0}^{t-1} P^i) g^-, \tag{1}$$

where $g^- = D^{-1}A^-\mathbf{1}$, i.e. $g^-(i)$ is the weighted fraction of outgoing negative edges of node i.

PROOF. (Sketch) Based on the signed digraph voter model defined in Section 2, x_t can be iteratively computed as

$$x_t(i) = \sum_{j \in V} \frac{A^+_{ij}}{d_i} x_{t-1}(j) + \sum_{j \in V} \frac{A^-_{ij}}{d_i} (1 - x_{t-1}(j)). \tag{2}$$

The matrix form of eq.(2) yields Eq.(1). □

3.2 Convergence of signed transition matrix

Eq.(1) infers that the long-term dynamic, i.e., the vector x_t when t goes to infinity, depends critically on the limit of P^t and $\sum_{i=0}^{t-1} P^i$. We show below that the limiting behaviors of the two matrix sequences are fundamentally determined by the structural balance of signed digraph G, which connects to the social balance theory well studied in the social science literature (cf. [13]). We now define three types of signed digraphs based on their balance structures.

Definition 2 (Structural balance of signed digraphs). *Let $G = (V, E, A)$ be a signed digraph.*

1. **Balanced digraph**. *G is* balanced *if there exists a partition $S, \bar{S}$ of nodes in V, such that all edges within S and $\bar{S}$ are positive and all edges across S and $\bar{S}$ are negative.*
2. **Anti-balanced digraph**. *G is* anti-balanced *if there exists a partition $S, \bar{S}$ of nodes in V, such that all edges within S and $\bar{S}$ are negative and all edges across S and $\bar{S}$ are positive.*
3. **Strictly unbalanced digraph**. *G is* strictly unbalanced *if G is neither balanced nor anti-balanced.*

The balanced digraphs defined above correspond to the balanced graphs originally defined in social balance theory. It is known that a balanced graph can be equivalently defined by the condition that all circles in G without considering edge directions contain an even number of negative edges [13]. On the other hand, the concept of anti-balanced digraphs seems not appearing in the social balance theory. Note that balanced digraphs and anti-balanced digraphs are not mutually exclusive. For example, a four node circle with one pair of non-adjacent edges being positive and the other pair being negative is both balanced and anti-balanced. However, for studying long-term dynamics, we only need the above categorization for aperiodic digraphs, for which we show below that balanced digraphs and anti-balanced digraphs are mutually exclusive.

Proposition 2. *An aperiodic digraph G cannot be both balanced and anti-balanced.*

With the above proposition, we know that balanced graphs, anti-balanced graphs, and strictly unbalanced graphs indeed form a classification of aperiodic digraphs, where anti-balanced graphs and strictly unbalanced graphs together correspond to unbalanced graphs in the social balance theory. We identify anti-balanced graphs as a special category because it has a unique long-term dynamic behavior different from other graphs. An example of anti-balanced graphs is a graph with only negative edges. In general, anti-balanced graphs could be viewed as an extreme in which many hostility exist among individuals, e.g., networks formed by bidders in auctions [5, 33].

The next lemma characterizes the limiting behavior of P^t of ergodic signed digraphs with all three balance structures. Given a signed digraph $G = (V, E, A)$, let $\bar{G} = (V, E, \bar{A})$ corresponds to its unsigned version ($\bar{A}_{ij} = |A_{ij}|$ for all $i, j \in V$). When $\bar{G}$ is ergodic, a random walk on $\bar{G}$ has a unique stationary distribution, denoted as π. That is, $\pi^T = \pi^T \bar{P}$, where $\bar{P} = D^{-1}\bar{A}$ is the transition probability matrix for $\bar{G}$. Henceforth, we always use $S, \bar{S}$ to denote the corresponding partition for either balanced graphs or anti-balanced graphs.

Lemma 1. *Given an ergodic signed digraph $G = (V, E, A)$, let $\bar{G} = (V, E, \bar{A})$ be the* unsigned *digraph. When G is balanced or strictly unbalanced, P^t converges, and when G is anti-balanced, the odd and even subsequences of P^t converge to two opposite matrices, i.e.,*

$$\begin{aligned}
&\textit{Balanced } G: && \lim_{t\to\infty} P^t = \hat{\boldsymbol{1}}_S \hat{\pi}_S^T \\
&\textit{Strictly unbalanced } G: && \lim_{t\to\infty} P^t = \boldsymbol{0} \\
&\textit{Anti-balanced } G: && \lim_{t\to\infty} P^{2t} = \hat{\boldsymbol{1}}_S \hat{\pi}_S^T \\
& && \lim_{t\to\infty} P^{2t+1} = -\hat{\boldsymbol{1}}_S \hat{\pi}_S^T,
\end{aligned}$$

The above lemma clearly shows different convergence behaviors of P^t for three types of graphs. In particular, P^t of anti-balanced graphs exhibits a bounded oscillating behavior in long term.

Now, we consider a weakly connected signed digraph $G = (V, E, A)$ with one ergodic sink component G_Z with node set Z, which only has incoming edges from the rest of the signed digraph G_X with node set $X = V \setminus Z$. Then, the signed transition matrix P has the following block form.

$$P = \left[\begin{array}{c|c} P_X & P_Y \\ \hline \boldsymbol{0} & P_Z \end{array}\right], \tag{3}$$

where P_X and P_Z are the block matrices for components G_X and G_Z, and P_Y represents the one-way connections from G_X to G_Z. Then, the t-step transition matrix P^t can be expressed as

$$P^t = \left[\begin{array}{c|c} P_X^{(t)} & P_Y^{(t)} \\ \hline \boldsymbol{0} & P_Z^{(t)} \end{array}\right], \tag{4}$$

where $P_X^{(t)} = P_X^t$, $P_Z^{(t)} = P_Z^t$ and $P_Y^{(t)} = \sum_{i=0}^{t-1} P_X^i P_Y P_Z^{t-1-i}$. When G_Z is balanced or anti-balanced, we use $S_Z, \bar{S}_Z$ to denote the partition of Z defining its balance or anti-balance structure. Then, we denote column vectors

$$u_b = (I_X - P_X)^{-1} P_Y \hat{\mathbf{1}}_{Z,S_Z}, \tag{5}$$

$$\text{and } u_u = (I_X + P_X)^{-1} P_Y \hat{\mathbf{1}}_{Z,S_Z}. \tag{6}$$

The reason that $I_X - P_X$ is invertible is because $\lim_{t\to\infty} P_X^t = \mathbf{0}$, which is in turn because there is a path from any node i in G_X to nodes in Z (since Z is the single sink), and thus informally a random walk from i eventually reaches and then stays in G_Z. The same reason applies to $I_X + P_X$.

Let π_Z denote the stationary distribution of nodes in G_Z, and $\hat{\pi}_{Z,S_Z}$ is signed, with $\hat{\pi}_{Z,S_Z}(i) = \pi_Z(i)$ for $i \in S_Z$, and $\hat{\pi}_{Z,S_Z}(i) = -\pi_Z(i)$, otherwise. Lemma 2 discloses the convergence of P^t given various balance structures of G_Z.

Lemma 2. *Given the formulation in Eq.(4), we have*

$$\begin{aligned}
&\textit{Balanced } G_Z: && \lim_{t\to\infty} P^t = \left[\begin{array}{c|c} \boldsymbol{0} & u_b \hat{\pi}_{Z,S_Z}^T \\ \hline \boldsymbol{0} & \hat{\boldsymbol{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \end{array}\right] \\
&\textit{Strictly unbalanced } G_Z: && \lim_{t\to\infty} P^t = \boldsymbol{0} \\
&\textit{Anti-balanced } G_Z: && \lim_{t\to\infty} P^{2t} = \left[\begin{array}{c|c} \boldsymbol{0} & -u_u \hat{\pi}_{Z,S_Z}^T \\ \hline \boldsymbol{0} & \hat{\boldsymbol{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \end{array}\right] \\
& && \lim_{t\to\infty} P^{2t+1} = \left[\begin{array}{c|c} \boldsymbol{0} & u_u \hat{\pi}_{Z,S_Z}^T \\ \hline \boldsymbol{0} & -\hat{\boldsymbol{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \end{array}\right]
\end{aligned}$$

Weakly connected digraphs with multiple ergodic sinks or disconnected digraphs can be similarly analyzed.

3.3 Long-term dynamics

Based on the structural balance classification and the convergence of signed transition matrix discussed above, we are now ready to analyze the long-term dynamics of the voter model on signed digraphs. Formally, we are interested in characterizing x_t with $t \to \infty$, i.e.,

$$x = \lim_{t\to\infty} x_t = \lim_{t\to\infty} (P^t x_0 + (\sum_{i=0}^{t-1} P^i) g^-). \tag{7}$$

If the even and odd subsequences of x_t converge separately, we denote $x_e = \lim_{t\to\infty} x_{2t}$, $x_o = \lim_{t\to\infty} x_{2t+1}$.

In the following theorem, we first discuss the long-term dynamics of voter model on ergodic signed digraphs.

Theorem 1. *Let $G = (V, E, A)$ be an ergodic signed digraph, we have*

$$\text{Balanced } G: \qquad x = \hat{\mathbf{1}}_S \hat{\pi}_S^T (x_0 - \tfrac{1}{2}\mathbf{1}) + \tfrac{1}{2}\mathbf{1} \tag{8}$$

$$\text{Strictly unbalanced } G: \qquad x = \tfrac{1}{2}\mathbf{1} \tag{9}$$

$$\text{Anti-balanced } G: \qquad x_e = \hat{\mathbf{1}}_S \hat{\pi}_S^T (x_0 - \tfrac{1}{2}\mathbf{1}) + \tfrac{1}{2}\mathbf{1} \tag{10}$$

$$x_o = -\hat{\mathbf{1}}_S \hat{\pi}_S^T (x_0 - \tfrac{1}{2}\mathbf{1}) + \tfrac{1}{2}\mathbf{1} \tag{11}$$

Theorem 1 has several implications. First of all, for strictly unbalanced digraphs, each node has equal steady state probability of being black or white, and it is not determined by the initial distribution x_0. Secondly, anti-balanced digraphs has the same steady state distribution as the corresponding balanced graph for even steps, and for odd steps, the distribution oscillates to the opposite ($x_o = \mathbf{1} - x_e$).

For a balanced ergodic digraph G with partition $S, \bar{S}$, it is easy to check that it has the following two equilibrium states: in one state all nodes in S are white while all nodes in $\bar{S}$ are black; and in the other state all nodes in S are black while all nodes in $\bar{S}$ are white. We call these two states the *polarized states*. Using random walk interpretation, we show in the following theorem that with probability 1, the voter model dynamic converges to one of the above two equilibrium states.

Theorem 2. *Given an ergodic signed digraph $G = (V, E, A)$, if G is balanced with partition $S, \bar{S}$, the voter model dynamic converges to one of the polarized states with probability 1, and the probability of nodes in S being white is $\hat{\pi}_S^T(x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}$. Similarly, if G is anti-balanced, with probability 1 the voter model dynamic oscillates between the two polarized states eventually, and the probability of nodes in S being white at even steps is $\hat{\pi}_S^T(x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}$.*

Theorem 3 introduces the long-term dynamics of the weakly connected signed digraphs. We consider weakly connected G with a single sink ergodic component G_Z, and use the same notations as in Section 3.2.

Theorem 3. *Let $G = (V, E, A)$ be a weakly connected signed digraph with a single sink component G_Z and a non-sink component G_X. The long term white color distribution vector x is expressed in two parts:*

$$x^T = \lim_{t\to\infty} x_t^T = [x_{XY}^T, x_Z^T].$$

where x_Z is the limit of x_{tZ} on G_Z with initial distribution x_{0Z} and is given as in Theorem 1, and vector x_{XY} is given below with respect to the balance structure of G_Z:

$$\text{Balanced } G_Z: \qquad x_{XY} = \tfrac{1}{2}\mathbf{1}_X + u_b \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z)$$

$$\text{Strictly unbalanced } G_Z: \qquad x_{XY} = \tfrac{1}{2}\mathbf{1}_X$$

$$\text{Anti-balanced } G_Z\text{, even } t: \qquad x_{XY,e} = \tfrac{1}{2}\mathbf{1}_X - u_u \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z)$$

$$\text{Anti-balanced } G_Z\text{, odd } t: \qquad x_{XY,o} = \tfrac{1}{2}\mathbf{1}_X + u_u \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z),$$

where u_b and u_u are defined in Eq.(5) and Eq.(6).

Theorem 3 characterizes the long-term dynamics when the underlying graph is a weakly connected signed digraph with one ergodic sink component. We can see that the results for balanced and anti-balanced sink components are more complicated than the ergodic digraph case, since how non-sink components are connected to the sink subtly affects the final outcome of the steady state behavior. In steady state, while the sink component is still in one of the two polarized states as stated in Theorem 2, the non-sink components exhibit more complicated color distribution, for which we provide probability characterizations in Theorem 3. Our results can be readily extended to the case with more than one ergodic sink components and disconnected digraphs. When the network only contains positive directed edges, the voter model dynamics can be interpreted using digraph random walk theory [25–28].

4. INFLUENCE MAXIMIZATION

With the detailed analysis on voter model dynamics for signed digraphs, we are now ready to solve the influence maximization problem. Intuitively, we want to address the following question: *If only at most k nodes could be selected initially and be turned white while all other nodes are black, how should we choose seed nodes so as to maximize the expected number of white nodes in short term and in long term, respectively?*

4.1 Influence maximization problem

We consider two types of short-term influence objectives, one is the *instant influence*, which counts the total number of influenced nodes at a step $t > 0$; the other is the *average influence*, which takes the average number of influenced nodes within the first t steps. These two objectives have different implications and applications. For example, political campaigns try to convince voters who may change their minds back and forth, but only the voters' opinions on the voting day are counted, which matches the *instant influence*. On the other hand, a credit card company would like to have customers keep using its credit card service as much as possible, which is better interpreted by the *average influence*. When t is sufficiently large, it becomes the long-term objective, and long-term average influence coincides with long-term instant influence when the dynamic converges.

Formally, we define the *short-term instant influence* $f_t(x_0)$ and the *short-term average influence* $\bar{f}_t(x_0)$ as follows:

$$f_t(x_0) := \mathbf{1}^T x_t(x_0) \text{ and } \bar{f}_t(x_0) := \frac{\sum_{i=0}^t f_i(x_0)}{t+1}. \tag{12}$$

Moreover, we define *l*ong term influence as

$$f(x_0) := \lim_{t\to\infty} \frac{\sum_{i=0}^t f_i(x_0)}{t+1}. \tag{13}$$

Note that when the dynamic converges (e.g. ergodic balanced or ergodic strictly unbalanced graphs), $f(x_0) = \lim_{t\to\infty} f_t(x_0)$. For ergodic anti-balanced graphs (or sink components), it is essentially the average of even- and odd-step limit influence.

Given a set $W \subseteq V$, Let e_W be the vector in which $e_W(j) = 1$ if $j \in W$ and $e_W(j) = 0$ if $j \notin W$, which represents the initial seed distribution with only nodes in W as white seeds. Let e_i be the shorthand of $e_{\{i\}}$. Unlike unsigned graphs, if initially no white seeds are selected on a signed digraph G, i.e., $x_0 = \mathbf{0}$, the instant influence $f_t(\mathbf{0})$ at step t is in general non-zero, which is referred to as the *ground influence* of the graph G at t. The influence contribution of a seed set W does not count such ground influence, as shown in definition 3.

Definition 3 (Influence contribution). *The* instant influence contribution *of a seed set W to the t-th step instant influence objective, denoted by $c_t(W)$, is the difference between the instant influence at step t with only nodes in W selected as seeds and the ground influence at step t: $c_t(W) = f_t(e_W) - f_t(\mathbf{0})$. The* average influence contribution *$\bar{c}_t(W)$ and* long-term influence contribution *$c(W)$ are defined in the same way: $\bar{c}_t(W) = \bar{f}_t(e_W) - \bar{f}_t(\mathbf{0})$ and $c(W) = f(e_W) - f(\mathbf{0})$.*

We are now ready to formally define the influence maximization problem.

Definition 4 (Influence maximization). *The* influence maximization *problem for short-term instant influence is finding a seed set W of at most k seeds that maximizes W's instance influence contribution at step t, i.e., finding $W_t^* = \arg\max_{|W|\le k} c_t(W)$. Similarly, the problem for average influence and long-term influence is finding $\bar{W}_t^* = \arg\max_{|W|\le k} \bar{c}_t(W)$ and $W^* = \arg\max_{|W|\le k} c(W)$, respectively.*

We now provide some properties of influence contribution, which lead to the optimal seed selection rule. By Eq.(1), we have

$$c_t(W) = f_t(e_W) - f_t(\mathbf{0}) = \mathbf{1}^T x_t(e_W) - \mathbf{1}^T x_t(\mathbf{0}) = \mathbf{1}^T P^t e_W. \tag{14}$$

Let $c_t(i)$ be the shorthand of $c_t(\{i\})$, and let $c_t = [c_t(i)]$ denote the vector of influence contribution of individual nodes. Then $c_t^T = [c_t(i)]^T = \mathbf{1}^T P^t$. When $t \to \infty$, the long term influence contributions of individual nodes are obtained as a vector c:

$$c^T = \lim_{t\to\infty} \frac{\sum_{i=0}^{t} c_i^T}{t+1} = \lim_{t\to\infty} \frac{\mathbf{1}^T \sum_{i=0}^{t} P^i}{t+1}. \tag{15}$$

When P^t converges, we simply have $c^T = \mathbf{1}^T \lim_{t\to\infty} P^t$. (16)

Lemma 3 below discloses the important property that the influence contribution is a linear set function.

Lemma 3. *Given a white seed set W, $c_t(W) = \sum_{i\in W} c_t(i)$, $\bar{c}_t(W) = \sum_{i\in W} \bar{c}_t(i)$, and $c(W) = \sum_{i\in W} c(i)$.*

Given a vector v, let $n^+(v)$ denote the number of positive entries in v. By applying Lemma 3, we have the optimal seed selection rule for instant influence maximization as follows.

Optimal seed selection rule for instant influence maximization. *Given a signed digraph and a limited budget k, selecting top $\min\{k, n^+(c_t)\}$ seeds with the highest $c_t(i)$'s, $i \in V$, leads to the maximized instant influence at step $t > 0$.*

Note that the influence contributions of some nodes may be negative and these nodes should not be selected as white seeds, and thus the optimal solution may have less than k seeds. The rules for average influence maximization and long-term influence maximization are patterned in the same way. Therefore, the central task now becomes the computation of the influence contributions of individual nodes. Below, we will introduce our SVIM algorithm, for Signed Voter model Influence Maximization.

4.2 Short-term influence maximization

By applying Definition 3 and Lemma 3, we develop SVIM-S algorithm to solve the short-term instant and average influence maximization problem, as shown in Algorithm 1.

Algorithm 1 Short-term influence maximization SVIM-S

INPUT: Signed transition matrix P, short-term period t, budget k;
OUTPUT: White seed set W.
$c_t = \mathbf{1}$; $\bar{c}_t = \mathbf{1}$;
for $i = 1 : t$ **do**
    $c_t^T = c_t^T P$;(for instant influence maximization.)
    $\bar{c}_t = \bar{c}_t + c_t$; (for average influence maximization.)
$W =$ top $\min\{k, n^+(c_t)\}$ (resp. $\min\{k, n^+(\bar{c}_t)\}$) nodes with the highest $c_t(i)$ (resp. $\bar{c}_t(i)$) values, for instant (resp. average) influence maximization.

SVIM-S algorithm requires t vector-matrix multiplications, each of which takes $|E|$ times entry-wise multiplication operations. Hence the total time complexity of SVIM-S is $O(t \cdot |E|)$.

4.3 Long-term influence maximization

We now study the long-term influence contribution c and introduce the corresponding influence maximization algorithm SVIM-L. We will see that the computation of influence contribution c and seed selection schemes depends on the structural balance and connectedness of the graph. While seed selection for balanced ergodic digraphs still has intuitive explanations, the computation for weakly connected and disconnected digraphs is more involved and less intuitive.

4.3.1 Case of ergodic signed digraphs

When the signed digraph $G = (V, E, A)$ is ergodic, Lemma 4 below characterizes the long-term influence contributions of nodes, with respect to various balance structures.

Lemma 4. *Consider an ergodic signed digraph $G = (V, E, A)$. If G is* balanced*, with bipartition S and $\bar{S}$, the influence contribution vector $c = (|S| - |\bar{S}|)\hat{\pi}_S$. If G is* anti-balanced *or* strictly unbalanced*, $c = \mathbf{0}$.*

Based on Lemma 4, Algorithm 2 summarizes how to compute the long-term influence contribution c on ergodic signed digraphs.

Algorithm 2 $c = ergodic(G)$

INPUT: Signed transition matrix P.
OUTPUT: Long term influence contribution vector c
Detect the structure of ergodic signed digraph G;
if G is balanced, with bipartition S and $\bar{S}$ **then**
    Compute stationary distribution π of $\bar{P}$;
    $c = (|S| - |\bar{S}|)\hat{\pi}_S$;
else
    $c = \mathbf{0}$;

Lemma 4 suggests that for ergodic balanced digraphs, we should pick the larger component, e.g., S, if $|S| > |\bar{S}|$, and select the top $\min\{k, |S|\}$ nodes from S with the largest stationary distributions as white seeds. Selecting these nodes will make the probability of the larger component being white the largest.

4.3.2 Case of weakly connected signed digraphs

We first consider a weakly connected signed G which has a single ergodic sink component G_Z with only incoming edges from the remaining nodes $X = V \setminus Z$.

Lemma 5. *Consider a weakly connected digraph $G = (V, E, A)$ with a single ergodic sink component G_Z. If G_Z is* balanced*, with partition S_Z and $\bar{S_Z}$, the long term influence contribution vector $c^T = [c_X^T, c_Z^T]$, where $c_X = \mathbf{0}_X$ and $c_Z = (\mathbf{1}_X^T u_b + |S_Z| - |\bar{S}_Z|)\hat{\pi}_{Z,S_Z}$. If G is* anti-balanced *or* strictly unbalanced*, $c = \mathbf{0}$.*

Lemma 5 indicates that influence contribution of the balanced ergodic sink component is more complicated than that of the balanced ergodic digraph. This is because the sink component affects the colors of the non-sink component in a complicated way depending on how non-sink and sink components are connected. Therefore, the optimal seed selection depends on the calculation of the influence contributions of each sink node, and is not as intuitive as that for the ergodic digraph case.

More sink components. When there exist $m > 1$ ergodic sink components, i.e., $G_{Z1}, G_{Z2}, \cdots, G_{Zm}$, the rest of the graph G is considered as a single component G_X. Then the signed transition matrix P and P^t can be written as

$$P = \left[\begin{array}{c:c:c:c} P_X & P_{Y1} & \cdots & P_{Ym} \\ \hdashline \mathbf{0} & P_{Z1} & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & P_{Zm} \end{array}\right], P^t = \left[\begin{array}{c:c:c:c} P_X^t & P_{Y1}^{(t)} & \cdots & P_{Ym}^{(t)} \\ \hdashline \mathbf{0} & P_{Z1}^t & \mathbf{0} & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \hdashline \mathbf{0} & \mathbf{0} & \mathbf{0} & P_{Zm}^t \end{array}\right]$$

where $P_{Yi}^{(t)} = \sum_{j=0}^{t-1} P_X^j P_{Yi} P_{Zi}^{t-1-j}$. Hence, each sink ergodic component P_{Zi} along with P_X independently follows Lemma 5. Algorithm 3 below summarizes how to compute the node influence contributions of weakly connected signed digraphs. Note that by our assumption, we consider all sink components to be ergodic.

Algorithm 3 $c = weakly(G)$

INPUT: Signed transition matrix P.
OUTPUT: Influence contribution vector c.
Detect the structure of the weakly connected signed digraph G, and find its $m \geq 1$ signed ergodic sink components $G_{Z1}, \cdots, G_{Zm}$;
for $i = 1 : m$ **do**
  if G_{Zi} is balanced with partition $S_{Zi}, \bar{S}_{Zi}$ **then**
    Compute stationary distribution π_{Zi} of $\bar{P}_{Zi}$;
    $u_{bi} = (I_X - P_X)^{-1} P_{Yi} \hat{\mathbf{1}}_{Zi,S_{Zi}}$;
    $c_{Zi} = (\mathbf{1}_X^T u_{bi} + |S_{Zi}| - |\bar{S}_{Zi}|) \hat{\pi}_{Zi,S_{Zi}}^T$;
$c = [\mathbf{0}_X; c_{Z1}; \cdots; c_{Zm}]$

4.3.3 General case and SVIM-L algorithm

Given the above systematic analysis, we are now in a position to summarize and introduce our SVIM-L algorithm which solves the long-term voter model influence maximization problem for general aperiodic signed digraphs.

In general, a signed digraph consists $m \geq 1$ disconnected components, within each of which the node influence contribution follows Lemma 5. The long-term signed voter model influence maximization (SVIM-L) algorithm is constructed in Algorithm 4.

Algorithm 4 Long-term influence maximization SVIM-L

INPUT: Signed transition matrix P, budget k.
OUTPUT: White seed set W.
Detect the structure of a general aperiodic signed digraph G, and find the $m \geq 1$ disconnected components $G_1, \cdots, G_m$;
for $i = 1 : m$ **do**
  $c_{G_i} = weakly(G_i)$;
$c = [c_{G_1}; \cdots; c_{G_m}]$;
$W =$ top $\min\{k, n^+(c)\}$ nodes with the highest $c(i)$ values.

Complexity analysis. We consider $G = (V, E, A)$ to be weakly connected, since disconnected graph case can be treated independently for each connected component for the time complexity. SVIM-L algorithm consists of two parts. The first part extracts the connectivity and balance structure of the graph, which can be done using depth-first search with complexity $O(|E|)$. The second part uses Algorithm 3 to compute influence contributions of balanced ergodic sink components. The dominant computations are on the stationary distribution π_{Zi}'s and $(I_X - P_X)^{-1}$, which can be done by solving a linear equation system and matrix inverse in $O(|Z_i|^3)$ and $O(n_X^3)$, respectively, where $n_X = |X|$. Let b be the number of balanced sink components in G, n_Z be the number of nodes in the largest balanced sink component. Thus SVIM-L can be done in $O(bn_Z^3 + n_X^3)$ time. Alternatively, we can use iterative method for computing both π_{Zi}'s and $\mathbf{1}_X^T(I_X - P_X)^{-1}$, if the largest convergence time t_C of P_{Zi}^t's and P_X^t is small[1]. In this case, each iteration step involves vector-matrix multiplication and can be done in $O(m_B)$ time, where m_B is the number of edges of the induced subgraph G_B consisting of all nodes in the balanced sink components and X. Note that m_B and t_C are only related to subgraph G_B, which could be significantly smaller than G, and thus $O(t_C m_B)$ could be much smaller than the time of naive iterations on the entire graph. Overall SVIM-L can be done in $O(|E| + \min(bn_Z^3 + n_X^3, t_C m_B))$ time.

[1]Note that the convergence time of digraphs could be large [31,32], and even exponentially large at the worst case.

5. EVALUATION

In this section, we first use both synthetic datasets and real social network datasets to demonstrate the efficacy of our short-term and long-term seed selection schemes by comparing the performances with four baseline heuristics. Then, we evaluate how much the short-term and long-term influence can be improved by taking the edge signs into consideration.

5.1 Performance comparison with baseline heuristics

For different scenarios, we compare our SVIM-L and SVIM-S algorithms with four heuristics, i.e., (1) selecting seed nodes with the highest weighted outgoing degrees (denoted by $d^+ + d^-$ in the figures), (2) highest weighted outgoing positive degrees (denoted by d^+), (3) highest differences between weighted outgoing positive and negative degrees (denoted by $d^+ - d^-$), and (4) randomly selecting seed nodes (denoted by "Rand"), where in our evaluations, we run random seed selection 1000 times, and compare the average number of white nodes between our algorithm and other heuristics. Our evaluation results demonstrate that our seed selection scheme can increase up to 72% long-term influence, and 145% short-term influence over other heuristics.

5.1.1 Synthetic datasets

In this part, we generate synthetic datasets with different structures to validate our theoretical results.

Dataset generation model. We generate six types of signed digraphs, including balanced ergodic digraphs, anti-balanced ergodic digraphs, strictly unbalanced ergodic digraphs, weakly connected signed digraphs, disconnected signed digraphs with ergodic components, and disconnected signed digraph with weakly connected components (WCCs). All edges have unit weights. The following are graph configuration details.

We first create an unsigned ergodic digraph $\bar{G}$ with 9500 nodes, which has two ergodic components $\bar{G}_A$ and $\bar{G}_B$, with $[3000, 6500]$ nodes and $[3000, 6500] \times 8$ random directed edges, respectively. Moreover, there are 3000×8 random directed edges across $\bar{G}_A$ and $\bar{G}_B$. Ergodicity is checked through a simple connectivity and aperiodicity check. Given $\bar{G}$, a *balanced digraph* is obtained by assigning all edges within $\bar{G}_A$ and $\bar{G}_B$ with positive signs, and those across them with negative signs. Then, an *anti-balanced digraph* is generated by negating all edge signs of the balanced ergodic digraph. To generate a *strictly unbalanced digraph*, we randomly assign edge signs to all edges in $\bar{G}$ and make sure that there does not exist a balanced or anti-balanced bipartition.

Moreover, we generated a *disconnected signed digraph* and a weakly connected signed digraph for our study. We first generate 5 ergodic unsigned digraphs, $\bar{G}_1, \cdots, \bar{G}_5$ with $[500, 200, 800, 300, 2700]$ nodes and $[500, 200, 800, 300, 2700] \times 8$ edges, respectively. Then, we group $G_{23} = (G_2, G_3)$ and $G_{45} = (G_4, G_5)$ to form two ergodic balanced digraphs, and generate a strictly unbalanced ergodic digraph G_1 by randomly assigning signs to edges in $\bar{G}_1$. Three disconnected components G_1, G_{23}, G_{45} together form a disconnected signed digraph. To form a *weakly connected signed digraph*, we place in total 3000 random direct edges from G_1 to the balanced ergodic components G_{23} and G_{45}, where the nodes in subgraph G_1 only have outgoing edges to G_{23} and G_{45}. Moreover, we combine the above generated balanced ergodic digraph and the weakly connected signed digraph together forming a larger *disconnected signed digraph, with the weakly connected signed digraph as a component*.

Fig. 1-Fig. 6 present the evaluation results for one set of digraphs, where we observe that all digraphs we randomly generated exhibit consistent results. Our tests are conducted using Matlab on a standard PC server.

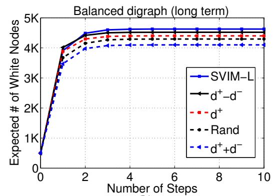

Figure 1: G is balanced

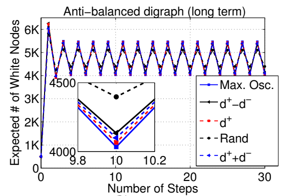

Figure 2: G is anti-balanced

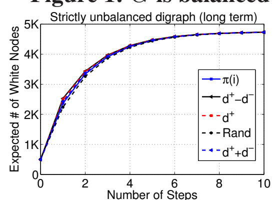

Figure 3: G is strictly unbalanced

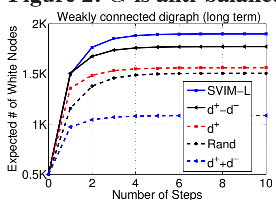

Figure 4: G is weakly connected

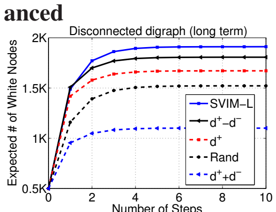

Figure 5: G is disconnected

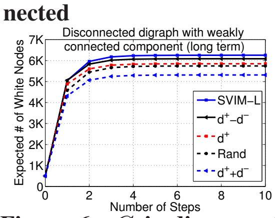

Figure 6: G is disconnected with WCC

Long-term influence maximization. In the evaluations, we set the influence budget as $k = 500$, and compare the average numbers of white nodes over steps between our algorithm and other heuristics. Fig. 1 shows that in the balanced ergodic digraph, SVIM-L algorithm achieves the highest long-term influence over other heuristics. When applying a heuristic seed selection scheme, denoted by $\mathtt{H}$, $f_t^{\mathtt{H}}$ represents the number of white nodes at step $t(\geq 1)$. Similarly, denote $f_t^{\mathtt{SVIM}}$ as the number of white nodes at step $t(\geq 1)$ for SVIM algorithm. We consider $\Delta f_t(\mathtt{SVIM}, \mathtt{H}) = (f_t^{\mathtt{SVIM}} - f_t^{\mathtt{H}})/f_t^{\mathtt{H}}$ as the influence increase of SVIM over the heuristic algorithm $\mathtt{H}$ at step t. The maximum influence increase is the maximum $\Delta f_t(\mathtt{SVIM}, \cdot)$ among all steps ($t \geq 1$) and all heuristics. Hence, in Fig. 1, we see that our SVIM-L algorithm outperforms all other heuristics. Especially, a maximum of 14% influence increase is observed for $t \geq 4$ with $4.68k$ and $4.1k$ white nodes for SVIM-L and random selection scheme, respectively. In the rest of this section, we will use the maximum influence increase as a metric to illustrate the efficacy of our SVIM algorithm. Fig. 2 shows the clear oscillating behavior on the anti-balanced ergodic digraph, and the average influence is the same for all algorithms. In fact, we also designed an algorithm to maximize the oscillation in this case, but due to space constraint we omit it in this paper. The inset shows that our algorithm (denoted as "Max. Osc.") indeed provides the largest oscillation. Fig. 3 shows the results in strictly unbalanced graph case, where the long-term influences of all algorithms converge to $4750 = |V|/2$, which matches Theorem 1. Fig. 4 and Fig. 5 show that SVIM-L algorithm performs the best, and it generates $5.6\% - 72\%$ long-term influence increases after the sixth step over other heuristics in the weakly connected signed digraph and the disconnected signed digraph. Fig. 6 shows that in a more general signed digraph, which consists of a weakly connected signed component and a balanced ergodic component, SVIM-L algorithm outperforms all other heuristics with up to 17% more long term influence, which occurs for $t \geq 4$. In general, we see that for weakly connected and disconnected digraphs, SVIM-L has larger winning margins over all other heuristics than the case of balanced ergodic digraphs (Fig. 4–6 vs. Fig.1). We attribute this to our accurate computation of influence contribution in the more involved weakly connected and disconnected digraph cases. Moreover, in all cases, the dynamics converge very fast, i.e., in only a few steps, which indicates that the convergence time of voter model on these random graphs are very small.

Table 2: Statistics of Epinions datasets

# of nodes	131580
# of edges	840799
# of positive edges	717129
# of negative edges	123670
# of nodes in largest SCC	41441
# of edges in largest SCC	693507
# of positive edges in largest SCC	614314
# of negative edges in largest SCC	79193
# of strongly connected components	88361

5.1.2 *Real datasets*

We conduct extensive simulations using real datasets, such as Epinions and Slashdot datasets, to validate our theoretical results and evaluate the performance of our SVIM algorithm.

Epinions Dataset. Epinions.com [14] is a consumer review online social site, where users can write reviews to various items and vote for or against other users. The signed digraph is formed with positive or negative directed edge (u, v) meaning that u trusts or distrusts v. The statistics are shown in Table 2. We compare our short-term SVIM-S algorithm with four heuristics, i.e., $d^+ + d^-$, d^+, $d^+ - d^-$ and random seed selection, on the entire Epinions digraph as well as the largest strongly connected component (SCC).

Our tests are conducted on both Epinions dataset and its largest strongly connected component (SCC), where the largest SCC is ergodic and strictly unbalanced. We first look at the comparison of instant influence maximization (*at* step t) among various seed selection schemes. Fig. 7-10 shows the expected maximum instant influence at each step by different methods. Note that since the initial seeds selected by SVIM-S algorithm hinge on t, the values on the curve of our selection scheme are associated with different optimal initial seed sets. On the other hand, the seed selections of other heuristics are independent to t, thus the corresponding curves represent the same initial seed sets. We choose the budget as 500 and 6000 in our evaluations, i.e., selecting at maximum 500 or 6000 initial white seeds. From Fig. 7-10, SVIM-S algorithm consistently performs better, and in some cases, e.g., Fig. 9, it generates $16\% - 145\%$ more influence than other heuristics at step 1.

Next we compare the seed selection schemes for maximizing the average influence *within* the first t steps. Fig. 11-14 show the expected maximum average influence within the first t steps by different methods. Again, the values on the curve of SVIM-S algorithm are associated with different initial seed sets. Fig. 11-14 show that with different budgets, i.e., 500 and 6000 seeds, SVIM-S algorithm performs better than all other heuristics, where in Fig. 13 a maximum of 64% more influence is achieved at $t = 8$. Moreover, in all these figures, we observe that our seed selection scheme results in the highest long-term influence over other heuristics.

Moreover, from Fig. 7-14, we observe that as t increases, the influences (i.e., the expected number of white nodes), for SVIM-S and all heuristics except for random seed selection schedule, increase for small t's, and then decrease and converge to the stationary state. In contrast, from Fig. 1-6, the influence increases monotonically with t. This happens because Epinions dataset (as

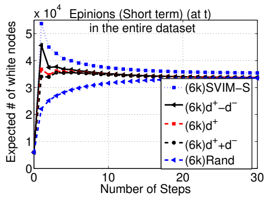

Figure 7: Instant influence in Epinions data with $k = 6$k

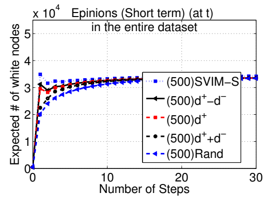

Figure 8: Instant influence in Epinions data with $k = 500$

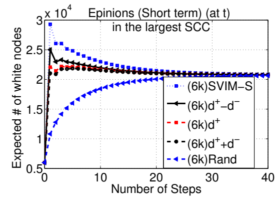

Figure 9: Instant influence in SCC with $k = 6$k

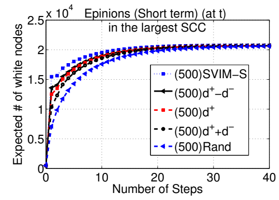

Figure 10: Instant influence in SCC with $k = 500$

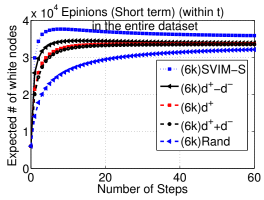

Figure 11: Average influence in Epinions data with $k = 6$k

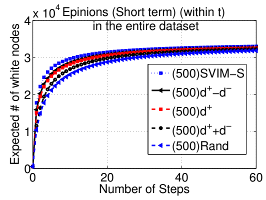

Figure 12: Average influence in Epinions data with $k = 500$

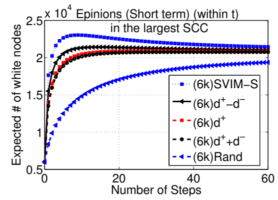

Figure 13: Average influence in SCC with $k = 6$k

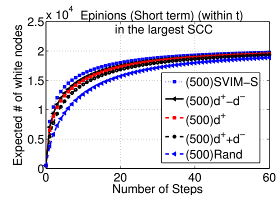

Figure 14: Average influence in SCC with $k = 500$

well as many real network datasets) has large portion (around 80%) of nodes in the non-sink components, where to maximize the long-term influence, only nodes in sink components should be selected, which governs the long-term influence dynamics of the whole graph, namely, sink nodes have higher long-term influence contributions. However, for short-term influence maximization, nodes with higher chances to influence more nodes in a few steps generally have large number of incoming links, which are able to influence a large number of nodes in either sink or non-sink components in a short period of time. Hence, in signed digraphs with large non-sink component, given a sufficiently large budget, the short-term influence can definitely outnumber the long-term influence. Our evaluations confirm this explanation. This interesting observation also leads to a problem that given a budget k, how to find the optimal time step t that generates the largest influence among all possible t's. We leaves this problem as our future work.

We also evaluate our SVIM-S algorithm on the entire slashdot dataset [22, 37] and its largest strongly connected component, where the results are delegated to our technical report [24] due to the limited space. In the simulations, similar results are obtained as that with Epinions dataset, where our SVIM-S algorithm performs the best among all methods tested, especially in the early steps.

Moreover, the convergence times for both real-world datasets are fast, in a few tens of steps, indicating good connectivity and fast mixing property of real-world networks. In summary, our evaluation results on both synthetic and real-world networks validate our theoretical results and demonstrate that our SVIM algorithms for both short term and long term are indeed the best, and often have significant winning margins.

5.2 The impacts of signed information

Unlike Epinions and Slashdot, many online social networks such as Twitter are simply represented by unsigned directed graphs, where friends and foe relationships are not explicitly represented on edges. Without edge signs, two types of information may be mis-represented or under-represented: (1) one may follow his foes for tracking purpose, but this link may be mis-interpreted as friend or trust relationship; and (2) one may not follow his foes publicly to avoid being noticed, but his foes may still generate negative influence to him. In this section, we investigate how much influence gain can be obtained by taking the edge signs into consideration, thus illustrate the significance of utilizing both friend and foe relationships in influence maximization.

Taking the synthetic networks and Epinions dataset (used in Sec 5.1) as examples, we apply our SVIM algorithm to compute the optimal initial seed sets in the original signed digraphs, and two types of "sign-missing" scenarios, i.e., the unsigned digraphs with only original positive edges (denoted by "Positive" graphs) and with all edges labeled by the same signs (denoted by "Sign ignored" graphs). Then, we examine the performances of those three initial seed sets in original signed digraphs.

Fig. 15-18 show the evaluation results, where the seed sets obtained by considering edge signs perform consistently better than those using unsigned graphs. In synthetic networks, we observed $5\% - 16\%$ more influence in balanced digraph for $t \geq 6$ (See Fig. 15), and $11.7\% - 58\%$ more influence in weakly connected digraph for $t \geq 6$ (See Fig. 16). Moreover, in Epinions dataset from Fig. 17-18, there is no impact on the long-term influence, since the underlying graphs are strictly unbalanced. However, in short term, the results demonstrate that taking edge signs into consideration always performs better, which generates at maximum of 38% and 21% more influence for the entire dataset (See Fig. 17) and the largest SCC (See Fig. 18), respectively. Both maximums occur at step 1. These results clearly demonstrate the necessity of utilizing sign information in influence maximization.

6. CONCLUSION

In this paper, we propose and study voter model dynamics on signed digraphs, and apply it to solve the influence maximization problem. We provide a rigorous mathematical analysis to completely characterize the short-term and long-term dynamics, and provide efficient algorithms to solve both short-term and long-term influence maximization problems. Simulation results on both synthetic and real-world graphs demonstrate that our influence maximization (SVIM) algorithms consistently outperform other heuristic algorithms.

There exist several open problems and future directions. One open problem is the convergence time of voter model dynamics on signed digraphs. For balanced and anti-balanced ergodic digraphs, our results show that their convergence times are the same as the corresponding unsigned digraphs. For strictly unbalanced ergodic digraphs and more general weakly connected signed digraphs, the problem is quite open. A future direction is to study influence diffusion in signed networks under other models, such as the voter model with a background color, the independent cascade model, and the linear threshold model.

7. ACKNOWLEDGEMENT

We would like to thank Christian Borgs and Jennifer T. Chayes for pointing out the relations between the signed digraph voter

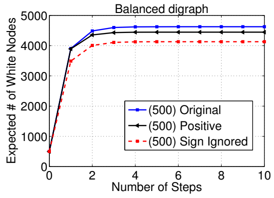

Figure 15: Synthetic balanced digraph

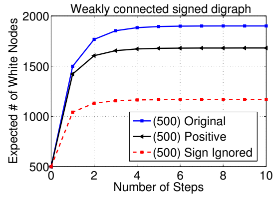

Figure 16: Synthetic weakly connected digraph

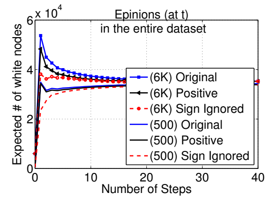

Figure 17: Epinions (the entire dataset)

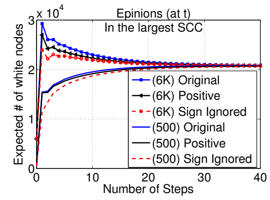

Figure 18: Epinions (the largest SCC)

model and concepts in physics, such as Ising model and Gauge transformations. We also thank Zhenming Liu for many useful discussions on this work.

8. REFERENCES

[1] M. Ángeles Serrano, K. Klemm, F. Vazquez, V. Eguíluz, and M. San Miguel. Conservation laws for voter-like models on random directed networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2009:P10024, 2009.

[2] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *WINE*, 2007.

[3] C. Borgs, J. Chayes, A. Kalai, A. Malekian, and M. Tennenholtz. A novel approach to propagating distrust. *Internet and Network Economics*, 2010.

[4] A. Borodin, Y. Filmus, and J. Oren. Threshold models for competitive influence in social networks. In *WINE*, 2010.

[5] F. Brandt, T. Sandholm, and Y. Shoham. Spiteful bidding in sealed-bid auctions. In *IJCAI*, 2007.

[6] C. Budak, D. Agrawal, and A. E. Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, 2011.

[7] W. Chen, A. Collins, R. Cummings, T. Ke, Z. Liu, D. Rincón, X. Sun, Y. Wang, W. Wei, and Y. Yuan. Influence maximization in social networks when negative opinions may emerge and propagate. In *SDM*, 2011.

[8] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, 2010.

[9] W. Chen, Y. Wang, and S. Yang. Efficient influence maximization in social networks. In *KDD*, 2009.

[10] W. Chen, Y. Yuan, and L. Zhang. Scalable influence maximization in social networks under the linear threshold model. In *ICDM*, 2010.

[11] K. Chiang, N. Natarajan, A. Tewari, and I. Dhillon. Exploiting longer cycles for link prediction in signed networks. 2011.

[12] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581, 1973.

[13] D. Easley and J. Kleinberg. *Networks, crowds, and markets: reasoning about a highly connected world*. Cambridge, 2010.

[14] Epinions. Dataset. *http://www.epinions.com/*.

[15] E. Even-Dar and A. Shapira. A note on maximizing the spread of influence in social networks. In *WINE*, 2007.

[16] A. Goyal, W. Lu, and L. V. S. Lakshmanan. Simpath: An efficient algorithm for influence maximization under the linear threshold model. In *ICDM*, 2011.

[17] X. He, G. Song, W. Chen, and Q. Jiang. Influence blocking maximization in social networks under the competitive linear threshold model. In *SDM*, 2012.

[18] R. Holley and T. Liggett. Ergodic theorems for weakly interacting infinite systems and the voter model. *The annals of probability*, 1975.

[19] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD*, 2003.

[20] J. Kunegis, S. Schmidt, A. Lommatzsch, J. Lerner, E. W. D. Luca, and S. Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM*, 2010.

[21] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Predicting positive and negative links in online social networks. In *WWW*, 2010.

[22] J. Leskovec, D. Huttenlocher, and J. Kleinberg. Signed networks in social media. In *CHI*. ACM, 2010.

[23] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In *KDD*, 2007.

[24] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. *arXiv:1111.4729 [cs.SI]*, Nov 2011.

[25] Y. Li and Z.-L. Zhang. Random walks on digraphs: A theoretical framework for estimating transmission costs in wireless routing. In *INFOCOM*, 2010.

[26] Y. Li and Z.-L. Zhang. Random walks on digraphs, the generalized digraph laplacian and the degree of asymmetry. *WAW*, 2010.

[27] Y. Li and Z.-L. Zhang. Digraph laplacian and the degree of asymmetry. *Internet Mathematics*, 8(4), 2012.

[28] Y. Li and Z.-L. Zhang. Random walks and green's function on digraphs: A framework for estimating wireless transmission costs. *IEEE/ACM Transactions on Networking*, PP(99):1–14, 2012.

[29] H. Ma, H. Yang, M. R. Lyu, and I. King. Mining social networks using heat diffusion processes for marketing candidates selection. In *CIKM*, 2008.

[30] N. Masuda and H. Ohtsuki. Evolutionary dynamics and fixation probabilities in directed networks. *New Journal of Physics*, 11:033012, 2009.

[31] A. Mohaisen, H. Tran, N. Hopper, and Y. Kim. On the mixing time of directed social graphs and security implications. In *ASIACCS*, 2012.

[32] A. Mohaisen, A. Yun, and Y. Kim. Measuring the mixing time of social graphs. In *IMC*, 2010.

[33] J. Morgan, K. Steiglitz, and G. Reis. The spite motive and equilibrium behavior in auctions. *The BE Journal of Economic Analysis & Policy*, 2(1):1102–1127, 2003.

[34] R. Narayanam and Y. Narahari. Determining the top-k nodes in social networks using the shapley value. In *AAMAS*, 2008.

[35] L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. *Technical report, Stanford University*, 1998.

[36] N. Pathak, A. Banerjee, and J. Srivastava. A generalized linear threshold model for multiple cascades. In *ICDM*, 2010.

[37] Slashdot. Dataset. *http://slashdot.org/*.

[38] V. Sood, T. Antal, and S. Redner. Voter models on heterogeneous networks. *Physical Review E*, 77(4):041121, 2008.