Voter Model on Signed Social Networks

Yanhua Li[†], Wei Chen[§], Yajun Wang[§] and Zhi-Li Zhang[†]

[†]Dept. of Computer Science & Engineering, Univ. of Minnesota, Twin Cities

[§] Microsoft Research Asia

{yanhua,zhzhang}@cs.umn.edu,{weic,yajunw}@microsoft.com

Abstract

Online social networks (OSNs) are becoming increasingly popular in recent years, which have generated great interest in studying the influence diffusion and influence maximization with applications to 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 the efficacy of our influence maximization algorithm over other heuristic algorithms.

Keywords: Signed networks, voter model, influence maximization, social networks

[*]A preliminary version of the results in this paper appeared in [28].

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, and 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, 17–19, 22, 23, 27, 37, 38]. 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, 25, 26] 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, 17, 22], 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 [15] 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, 17, 19, 22]. 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 [14, 21] 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 [15], 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 much better than 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.

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 [22] proposes several influence diffusion models and provides the greedy approximation algorithm for influence maximization. More recent works [8–10, 17, 19, 23, 27, 37] study efficient optimizations and scalable heuristics for the influence maximization problem. In

particular, the voter model is proposed in [14, 21], 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 [17] 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. Chung and Tsiatas [12] consider the voter models on hypergraphs as random walks on the associated weighted directed state graphs with various restrictions, such as "memoryless" or "partial memoryless" random walks, and provide bounds of the convergence time in various voter model settings. 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 [1,35,40]. 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 [3,24–26]. In [25,26], 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. [24] 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 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 considered as an influence ranking algorithm that generalizes the PageRank algorithm, but we treat distrust links as generating negative influence rather than ignoring distrusted users' opinions, and thus our ranking method is different from [3]. 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 [2,4,6,7,20,34], 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 for 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 $A_{ij} > 0$ (or edge (i, j) is positive) then i changes its color to j's color, but if $A_{ij} < 0$ (or 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.

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

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 the 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 $Z \subseteq V$.

ergodic sink components. Table 1 provides notations and terminologies used in the paper. Note that one basic fact we often use in studying long-term convergence behavior is: If matrix P satisfies $\lim_{t\to\infty} P^t = \mathbf{0}$, then $I - P$ is invertible and $(I - P)^{-1} = \lim_{t\to\infty} \sum_{i=0}^{t} P^i$.

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, and 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.*

The 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. Based on the signed digraph voter model defined in Section 2, x_t can be iteratively computed as

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

In matrix form, we have

$$x_t = D^{-1}Ax_{t-1} + D^{-1}A^-\mathbf{1} = Px_{t-1} + g^-, \tag{3}$$

which yields Eq.(1) by repeatedly applying Eq.(3). □

3.2 Convergence of signed transition matrix with relation to structural balance of signed digraphs

Eq.(1) infers that the long-term dynamics, 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 behavior of the two matrix sequences is 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. [15]). 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 [15]. On the other hand, the concept of anti-balanced digraphs seems not to appear 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.*

Proof. Suppose, for a contradiction, that an aperiodic digraph G is both balanced and anti-balanced. By the equivalent condition of balanced graphs, we know that all cycles of G have an even number of negative edges. Since an anti-balanced

graph will become balanced if we negate the signs of all its edges, we know that all cycles of G also have an even number of positive edges. Therefore, all cycles of G must have an even number of edges, which means their lengths have a common divisor 2, contradicting to the assumption that G is aperiodic. □

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.

Case of ergodic signed digraphs. Now, we discuss the limiting behavior of P^t of ergodic signed digraphs with three balance structures. A signed digraph $G = (V, E, A)$ is ergodic if and only if for any node i, there always exists a signed path to any other node in G and the common divisor of all cycle path lengths of i is 1. Here, a signed path R in a signed graph G is a sequence of nodes with the edges being directed from each node to the following one, where the length of the path, denoted as $|R|$, is the total number of directed edges in R. 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. Below, we first introduce Proposition 3 presenting that the balance structures of ergodic signed digraphs can be interpreted and distinguished in terms of the path lengths and path signs in G. As a result, Lemma 1 introduces the various limiting behaviors of P^t of ergodic signed digraphs with respect to three balance structures.

Proposition 3. *Let $G = (V, E, A)$ be an ergodic strictly unbalanced digraph. There exist two nodes i and j, and two directed paths from i to j with the same length but different signs.*

Proof. Given the following three statements, we prove `Statement 1` $\Rightarrow$ `Statement 2` $\Rightarrow$ `Statement 3`, which in turn proves this proposition, i.e., $\neg$`Statement 3` $\Rightarrow$ $\neg$`Statement 1`. We assume that G is a signed ergodic digraph.

Statement 1: For any two nodes i and j, all paths from i to j with the same length have same signs.

Statement 2: For any two nodes i and j, all paths from i to j with even length have same signs.

Statement 3: G is either balanced or anti-balanced.

(1) Proof by contradiction for `Statement 1` $\Rightarrow$ `Statement 2`. We assume that in G, there exist two even length paths R_{e1} and R_{e2} from i to j with different signs. Since G is ergodic, by Proposition 4 in Appendix A, there must exist a path, denoted by R_o, from j to i with odd length (no matter what sign it carries). Denote the length of these three paths as $|R_{e1}|$, $|R_{e2}|$ and $|R_o|$, respectively.

Then, $R_{c1} = R_{e1} + R_o$ forms a cycle at node i with odd length $|R_{e1}| + |R_o|$ and $R_{c2} = R_{e2} + R_o$ forms another cycle at i with odd length $|R_{e2}| + |R_o|$. Clearly, two cycles R_{c1} and R_{c2} carry different signs. Then, let $R'_{c1} = R_{c1}^{|R_{c2}|}$ denote a cycle of node i, by continuing R_{c1} for $|R_{c2}|$ times, which has the same sign with R_{c1} since $|R_{c2}|$ is odd. Similarly, we construct a cycle $R'_{c2} = R_{c2}^{|R_{c1}|}$ by continuing R_{c2} for $|R_{c1}|$ times, which has the same sign as R_{c2}. Thus R'_{c1} and R'_{c2} have the same length of $|R_{c1}||R_{c2}|$ but different signs, which contradicts to Statement 1.
(2) Proof for `Statement 2` $\Rightarrow$ `Statement 3`. By Proposition 4 in Appendix A, we know that between any two nodes there must exist even-length paths. By Statement 2, we partition V into S and $\bar{S}$, based on the signs of even length paths originated from a particular node $i \in V$. More specifically, S contains the nodes to which all even length paths from i have positive signs, and $\bar{S}$ contains the other set of nodes (note that i may not be in S).

We argue that (a) within S and $\bar{S}$, all edges have same signs; and (b) all edges between S and $\bar{S}$ have same signs. Since G contains both negative and positive edges, it must be either balanced or anti-balanced.

For (a), assume to the contrary that there exist two directed edges $R_{ab} = a \to b$ and $R_{cd} = c \to d$, which both reside in the same set, e.g., S with different signs. (The case for $\bar{S}$ is similar.)

We construct two even length paths from i to c and i to d as follows.

$$\begin{aligned} R_e(i,c) &= R_e(i,b) + R_e(b,c), \\ R_e(i,d) &= R_e(i,a) + R_{ab} + R_e(b,c) + R_{cd} \end{aligned}$$

where $R_e(x,y)$ represents the constructed even length path from node x to node y.

Since both $c, d \in S$, by construction, then $R_e(i,c)$ and $R_e(i,d)$ have same signs

$$sgn(R_e(i,c)) = sgn(R_e(i,d)). \tag{4}$$

On the other hand, since a and b are in the same group as c and d, $sgn(R_e(i,a)) =$

$sgn(R_e(i,b))$. Then, we have

$$sgn(R_e(i,c)) = sgn(R_e(i,b))sgn(R_e(b,c)), \tag{5}$$
$$\begin{aligned} sgn(R_e(i,d)) &= sgn(R_e(i,a))sgn(R_{ab})sgn(R_e(b,c))sgn(R_{cd}) \\ &= -sgn(R_e(i,b))sgn(R_e(b,c)). \end{aligned} \tag{6}$$

Eq.(6) comes from the assumption that R_{ab} and R_{cd} have different signs. Eq.(4) contradicts with Eq.(5) and Eq.(6).

For (b), assume that there exist two edges R_{ab} and R_{cd} with different signs between S and $\bar{S}$. Still consider the two even length paths $R_e(i,c)$ and $R_e(i,d)$ constructed before. Since c and d are not in the same side, $R_e(i,c)$ and $R_e(i,d)$ have opposite signs by the construction, i.e.,

$$sgn(R_e(i,c)) = -sgn(R_e(i,d)). \tag{7}$$

On the other hand, since a and b are in the different groups as well, $sgn(R_e(i,a)) = -sgn(R_e(i,b))$. Then, we have

$$sgn(R_e(i,c)) = sgn(R_e(i,b)) \cdot sgn(R_e(b,c)), \tag{8}$$
$$\begin{aligned} sgn(R_e(i,d)) &= sgn(R_e(i,a))sgn(R_{ab})sgn(R_e(b,c))sgn(R_{cd}) \\ &= sgn(R_e(i,b)) \cdot sgn(R_e(b,c)). \end{aligned} \tag{9}$$

However, Eq.(7) contradicts with Eq.(8) and Eq.(9). This completes the proof. □

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 π^T is the transpose of the stationary distribution vector π, and $\bar{P} = D^{-1}\bar{A}$ 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. We define the infinity norm of a matrix $M \in \mathbb{R}^{m\times m}$ as: $\|M\|_\infty := \max_{1\le i\le m} \sum_{j=1}^{m} |M_{ij}|$.

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, respectively.*

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

Proof. (1) When G is balanced, the signed transition matrix P can be written as $P = \hat{I}_S \bar{P} \hat{I}_S$. Since $\bar{G}$ is ergodic, we have $\lim_{t\to\infty} \bar{P}^t = \mathbf{1}\pi^T$. Thus,

$$\lim_{t\to\infty} P^t = \lim_{t\to\infty} (\hat{I}_S \bar{P} \hat{I}_S)^t = \hat{\mathbf{1}}_S \hat{\pi}_S^T,$$

where we use the simple facts $\hat{I}_S^2 = I$, $\hat{I}_S \mathbf{1} = \hat{\mathbf{1}}_S$, and $\pi^T \hat{I}_S = \hat{\pi}_S^T$.
(2) When G is anti-balanced, we have $P = -\hat{I}_S \bar{P} \hat{I}_S$. Thus,

$$\begin{aligned}
\lim_{t\to\infty} P^{2t} &= \lim_{t\to\infty} (-\hat{I}_S \bar{P} \hat{I}_S)^{2t} = \hat{\mathbf{1}}_S \hat{\pi}_S^T\\
\lim_{t\to\infty} P^{2t+1} &= \lim_{t\to\infty} (-\hat{I}_S \bar{P} \hat{I}_S)^{2t+1} = -\hat{\mathbf{1}}_S \hat{\pi}_S^T.
\end{aligned}$$

(3) ~~By~~ Proposition 3, ~~given a signed strictly unbalanced digraph G,~~ there ~~exist~~ a pair of nodes i and j, such that two paths R_1 and R_2 from i to j have the same length $\ell(i)$ and opposite signs. Consider a random walk from i. Let p_1 (resp. p_2) be the probability that the walk exactly follows R_1 (resp. R_2) in the first $\ell(i)$ steps. Let $R_{i,k}^{\ell(i)}$ be the set of all paths from i to k with length $\ell(i)$. Then, for a unit vector e_i with i-th entry equal to 1 and other entries as 0, we have $\|e_i^T P^{\ell(i)}\|_1$ as the ~~total~~ signed ~~probability~~ by randomly walking from node i to node j, which is bounded as follows:

$$\|e_i^T P^{\ell(i)}\|_1 = \sum_{k\in V} \left| \sum_{R\in R_{i,k}^{\ell(i)}} \mathrm{Prob}[R] sgn(R) \right| \le 1 - \min(p_1, p_2) = \rho_i.$$

For any node $i' \in V$, there must exist a path R' from $i' \to i$, due to the ergodicity of G, (where R' is empty when $i' = i$,) thus two paths $R_1' = R' + R_1$ and $R_2' = R' + R_2$ from i' to j have the same length, but opposite signs. With similar arguments as that for the node i, $\|e_{i'}^T P^{\ell(i')}\|_1 \le \rho_{i'}$ holds for any $i' \in V$. Let $\rho = \max_{i'\in V} \rho_{i'} < 1$ and $\ell = \max_{i'\in V} \ell(i')$, we conclude that $\|e_{i'}^T P^{\ell}\|_1 \le \rho$. We can express the identity matrix I as rows of e_i's, i.e., $I = [e_1^T; \cdots; e_{|V|}^T]$, which infers the following inequality:

$$\|P^\ell\|_\infty = \|IP^\ell\|_\infty = \max_{i'\in V} \|e_{i'}^T P^\ell\|_1 \le \max_{i'\in V} \rho_{i'} = \rho < 1.$$

Hence, by applying the fact that $\|AB\|_\infty \leq \|A\|_\infty \|B\|_\infty$ for any square matrices A and B, when $t \geq T = 2\ell$, the following inequality holds

$$\|P^t\|_\infty = \|P^{\frac{t}{\ell}\ell}\|_\infty \leq \rho^{\lfloor \frac{t}{\ell} \rfloor} \leq \rho^{\frac{t}{T}},$$

which infers $\lim_{t\to\infty} \|P^t\|_\infty = 0$, i.e., $\lim_{t\to\infty} P^t = \mathbf{0}$. □

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 the long term.

Case of weakly connected signed digraphs. 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.

$$P = \begin{bmatrix} P_X & P_Y \\ \mathbf{0} & P_Z \end{bmatrix}, \tag{10}$$

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

$$P^t = \begin{bmatrix} P_X^{(t)} & P_Y^{(t)} \\ \mathbf{0} & P_Z^{(t)} \end{bmatrix}, \tag{11}$$

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

$$u_b = (I_X - P_X)^{-1} P_Y \hat{\mathbf{1}}_{Z,S_Z}, \tag{12}$$

$$\text{and } u_u = (I_X + P_X)^{-1} P_Y \hat{\mathbf{1}}_{Z,S_Z}. \tag{13}$$

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 $I_X + P_X$. Lemma 2 provides the formal proof of the fact $\lim_{t\to\infty} P_X^t = \mathbf{0}$.

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 $\hat{\pi}_{Z,S_Z}(i) = -\pi_Z(i)$ for $i \in Z \setminus S_Z$. Lemma 2 describes the convergence of P^t given various balance structures of G_Z.

Lemma 2. *For weakly connected signed digraph $G = (V, E, A)$ with one ergodic sink components, with signed transition matrix given in Eq.(11), we have*

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

Proof. We discuss the convergence of P^t_X, P^t_Z, and $P^{(t)}_Y$ in Eq.(11).
(1) We first prove that P^t_X converges to $\mathbf{0}$, i.e., $\lim_{t\to\infty} P^t_X = \mathbf{0}$.

Since G_X does not contain sink components, any node $i \in X$ has a path to component G_Z. Let R_{iZ} be the shortest path from i to some node in Z, and $Prob[R_{iZ}]$ denote the probability that a random walk starting from i takes the path R_{iZ}. Hence we denote

$$p = \min_{i\in X} Prob[R_{iZ}], \text{ and } m = \max_{i\in X} |R_{iZ}|,$$

which implies that starting from any node $i \in X$, after m steps of random walk, there is at least probability p that it reaches component G_Z. Hence, we have $\|P^m_X\|_\infty \le (1-p) < 1$. Let $T = 2m$, then for any $t > T$, we have

$$\|P^t_X\|_\infty = \|P^{\frac{t}{m}m}_X\|_\infty \le (1-p)^{\lfloor \frac{t}{m} \rfloor} \le (1-p)^{\frac{t}{T}},$$

which implies $\lim_{t\to\infty} \|P^t_X\|_\infty = 0$, i.e., $\lim_{t\to\infty} P^t_X = \mathbf{0}$.
(2) For subgraph G_Z, Lemma 1 directly yields

$$\lim_{t\to\infty} P^t_Z = \begin{cases} \mathbf{0}, & \text{Strictly unbalanced } G_Z; \\ \mathbf{1}_{Z,S_Z}\pi^T_{Z,S_Z}, & \text{Balanced } G_Z; \\ \mathbf{1}_{Z,S_Z}\pi^T_{Z,S_Z}, & \text{Anti-balanced } G_Z, \text{ even } t; \\ -\mathbf{1}_{Z,S_Z}\pi^T_{Z,S_Z}, & \text{Anti-balanced } G_Z, \text{ odd } t. \end{cases} \tag{14}$$

(3) Below, we focus on proving the results on $\lim_{t\to\infty} P^{(t)}_Y$ using Proposition 6 in Appendix B.

When G_Z *is strictly unbalanced,* from Lemma 1 and (1) in this proof, $\lim_{t\to\infty} P_X^t = \mathbf{0}$ and $\lim_{t\to\infty} P_Z^t = \mathbf{0}$ hold, thus by Proposition 6 in Appendix B $\lim_{t\to\infty} P_Y^{(t)} = \mathbf{0}$.
When G_Z *is balanced,* Lemma 1 and Proposition 5 in Appendix A directly yield $(P_Z - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = P_Z^t - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T$ for any integer $t > 0$, and $\lim_{t\to\infty}(P_Z - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = \mathbf{0}$, thus

$$\begin{aligned}
&\lim_{t\to\infty} P_Y^{(t)} = \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z^{t-1-i} - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T) \\
&= \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^{t-1-i} + \lim_{t\to\infty} \sum_{i=0}^{t-2} P_X^i P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T \\
&= (I_X - P_X)^{-1} P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T = u_b \pi_{Z,S_Z}^T,
\end{aligned}$$

where the first term in the second line being $\mathbf{0}$ is due to Proposition 6 (ii) in Appendix B.
When G_Z *is anti-balanced,* applying Lemma 1 and Proposition 5 in Appendix A, we have for any integer $t > 0$, $(P_Z + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = P_Z^t - (-1)^t \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T$, and $\lim_{t\to\infty}(P_Z + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^t = \mathbf{0}$ hold true, thus

$$\begin{aligned}
&\lim_{t\to\infty} P_Y^{(t)} = \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z^{t-1-i} - (-1)^{t-1-i}(\mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T - \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)) \\
&= \lim_{t\to\infty} \sum_{i=0}^{t-1} P_X^i P_Y (P_Z + \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T)^{t-1-i} + \lim_{t\to\infty} \sum_{i=0}^{t-2} (-1)^{t-1-i} P_X^i P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T \\
&= (-1)^{t-1} \lim_{t\to\infty} \sum_{i=0}^{t-2} (-P_X)^i P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T = (-1)^{t-1}(I_X + P_X)^{-1} P_Y \mathbf{1}_{Z,S_Z}\pi_{Z,S_Z}^T \\
&= (-1)^{t-1} u_u \pi_{Z,S_Z}^T.
\end{aligned}$$

Hence, we have for anti-balanced G_Z: $\lim_{t\to\infty} P_Y^{(2t)} = -u_u \hat{\pi}_{Z,S_Z}^T$, and $\lim_{t\to\infty} P_Y^{(2t+1)} = u_u \hat{\pi}_{Z,S_Z}^T$. □

Multiple sink components and disconnected signed digraphs. 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 G_X. Then the signed transition matrix P and P^t can be written

as

$$P = \begin{bmatrix} P_X & P_{Y1} & \cdots & P_{Ym} \\ \mathbf{0} & P_{Z1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & P_{Zm} \end{bmatrix}, P^t = \begin{bmatrix} P_X^t & P_{Y1}^{(t)} & \cdots & P_{Ym}^{(t)} \\ \mathbf{0} & P_{Z1}^t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \ddots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & P_{Zm}^t \end{bmatrix}, \tag{15}$$

where $P_{Yi}^{(t)} = \sum_{j=0}^{t-1} P_X^j P_{Yi} P_{Zi}^{t-1-j}$, $1 \leq i \leq m$. Hence, each sink ergodic component P_{Zi} along with P_X and P_{Yi} independently follows Lemma 2. For disconnected signed digraph, with $m \geq 1$ ergodic or weakly connected components, each of which satisfies Lemma 1 or Lemma 2, respectively. For brevity, we omit the details here.

3.3 Long-term dynamics

Based on the structural balance classification and the convergence of signed transition matrix discussed above, we are ready now 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.,

$$x = \lim_{t\to\infty} x_t = \lim_{t\to\infty} (P^t x_0 + (\sum_{i=0}^{t-1} P^i) g^-). \tag{16}$$

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

Before presenting the results on long-term dynamics of voter model, we first introduce the following useful lemma connecting a signed digraph G with another graph G' where all edge signs in G are negated.

Lemma 3. *Given a signed digraph $G = (V, E, A)$, let $G' = (V, E, -A)$ be a signed digraph with all edge signs negated from G. Then, for any initial color distribution x_0, at any $2t$ steps ($t > 0$), the color distributions $x_{2t}(G)$ on G and $x_{2t}(G')$ on G' are identical.*

Proof. Let $P' = -P$ denote the signed transition matrix of G', and denote the vector $g^- = D^{-1}A^-\mathbf{1}$ and $g'^- = D^{-1}(-A)^-\mathbf{1} = D^{-1}A^+\mathbf{1}$. Thus $g'^- = \mathbf{1} - g^-$. By Eq.(1), after two steps, we have

$$\begin{aligned} x_2(G') = P'^2 x_0 + P' g'^- + g'^- &= P^2 x_0 - P(\mathbf{1} - g^-) + \mathbf{1} - g^- \\ &= P^2 x_0 + P g^- + g^- = x_2(G), \end{aligned}$$

where the last equality uses facts $\mathbf{1} = D^{-1}\bar{A}\mathbf{1}$ and $P = D^{-1}A$. Since the lemma holds for two steps, then clearly it holds for all even steps. □

Next theorem discusses the case of ergodic signed digraphs.

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

$$\textit{Balanced } G: \quad x = \hat{\mathbf{1}}_S\hat{\pi}_S^T(x_0 - \tfrac{1}{2}\mathbf{1}) + \tfrac{1}{2}\mathbf{1} \tag{17}$$

$$\textit{Strictly unbalanced } G: \quad x = \tfrac{1}{2}\mathbf{1} \tag{18}$$

$$\textit{Anti-balanced } G: \quad x_e = \hat{\mathbf{1}}_S\hat{\pi}_S^T(x_0 - \tfrac{1}{2}\mathbf{1}) + \tfrac{1}{2}\mathbf{1} \tag{19}$$

$$x_o = -\hat{\mathbf{1}}_S\hat{\pi}_S^T(x_0 - \tfrac{1}{2}\mathbf{1}) + \tfrac{1}{2}\mathbf{1} \tag{20}$$

Proof. We discuss the limit in Eq. (16) for three possible balance structures of G.
Balanced digraphs. From Lemma 1 and Proposition 5 in Appendix A, it is easy to prove $P^m - \hat{\mathbf{1}}_S\hat{\pi}_S^T = (P - \hat{\mathbf{1}}_S\hat{\pi}_S^T)^m$ for any integer $m > 0$, which yields the following result on the second part in Eq. (16).

$$\lim_{t\to\infty}\sum_{i=0}^{t-1} P^i g^- = (I - P + \hat{\mathbf{1}}_S\hat{\pi}_S^T)^{-1}g^- + \lim_{t\to\infty}\sum_{i=1}^{t-1}\mathbf{1}_S\hat{\pi}_S^T g^- \tag{21}$$

$$= (I - P + \hat{\mathbf{1}}_S\hat{\pi}_S^T)^{-1}g^- = \frac{1}{2}\mathbf{1} - \frac{1}{2}\hat{\mathbf{1}}_S\hat{\pi}_S^T\mathbf{1}, \tag{22}$$

where the last term of Eq.(21) is canceled out due to the digraph flow circulation law [13, 30], i.e.,

$$\hat{\pi}_S^T g^- = \hat{\pi}_S^T D^{-1}A^-\mathbf{1} = \sum_{i\in S}\pi(i)\sum_{j\in\bar{S}}\bar{P}_{ij} - \sum_{i\in\bar{S}}\pi(i)\sum_{j\in S}\bar{P}_{ij} = 0.$$

The last equality in Eq.(22) holds because

$$\frac{1}{2}(I - P + \hat{\mathbf{1}}_S\hat{\pi}_S^T)(\mathbf{1} - \hat{\mathbf{1}}_S\hat{\pi}_S^T\mathbf{1}) - g^- = 0.$$

Eq.(17) is obtained by combining Eq.(22) with Lemma 1.
Anti-balanced Digraphs. Lemma 3 directly yields Eq.(19). The odd step influence distribution sequence is obtained by

$$x_o = Px_e + g^- = -\mathbf{1}_S\hat{\pi}_S^T(x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}\mathbf{1}.$$

Strictly unbalanced digraphs. From Theorem 1, $\lim_{t\to\infty} P^t = \mathbf{0}$ holds and thus we have

$$\lim_{t\to\infty} \sum_{i=0}^{t-1} P^i g^- = (I-P)^{-1} g^- = (D-A)^{-1} A^- \mathbf{1} = \frac{1}{2}\mathbf{1}. \tag{23}$$

The last equality comes from the fact that $(D-A)\mathbf{1} = 2A^-\mathbf{1}$. □

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 have the same steady state distribution as the corresponding balanced graph for even steps, and for odd steps, the distribution oscillates to the opposite ($x_o = \mathbf{1} - x_e$). Moreover, Eq.(17) can also be intuitively explained from the random walk interpretation of the voter model. In particular, starting from node i, if we perform a random walk for an infinite number of steps, the probability that the random walk stops at j is given by the stationary distribution $\pi(j)$. For balanced graphs, if i and j are from the same component (either S or $\bar{S}$), then the random walk must pass an even number of negative edges, so i takes the same color as j; if i and j are from opposite components, then the walk passes an odd number of negative edges and i takes the opposite of j's color. Thus, the steady distribution of $i \in S$ being white is given by $\pi_S^T x_{0S} + \pi_{\bar{S}}^T(\mathbf{1}_{\bar{S}} - x_{0\bar{S}})$, and the case of $i \in \bar{S}$ is symmetric. Some algebra manipulations can lead us to Eq.(17).

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 $\hat{\pi}_S^T(x_0 - \frac{1}{2}\mathbf{1}) + \frac{1}{2}$.*

Proof. Consider a balanced ergodic digraph G with partition $S, \bar{S}$. By ergodicity, given any two nodes i and j, with probability 1 the random walks starting from i

and j will meet eventually. If i and j are both in S, when the two walks meet at some node u, they both pass either an even number of negative edges (if $u \in S$) or an odd number of negative edges (if $u \in \bar{S}$). Therefore, i and j must be in the same color with probability 1. If i and j are from different components S and $\bar{S}$, a similar argument shows that they will have the opposite color with probability 1. Therefore the final state is one of the two polarized states. The probability of nodes in S being white is simply given by Theorem 1, Eq.(17). The case of anti-balanced ergodic digraphs can be argued in a similar way. □

Theorem 3 below 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:

$$\begin{aligned}
\textit{Balanced } G_Z\textit{:} \quad & x_{XY} = \tfrac{1}{2}\mathbf{1}_X + u_b \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z) \\
\textit{Strictly unbalanced } G_Z\textit{:} \quad & x_{XY} = \tfrac{1}{2}\mathbf{1}_X \\
\textit{Anti-balanced } G_Z\textit{, even } t\textit{:} \quad & x_{XY,e} = \tfrac{1}{2}\mathbf{1}_X - u_u \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z) \\
\textit{Anti-balanced } G_Z\textit{, odd } t\textit{:} \quad & x_{XY,o} = \tfrac{1}{2}\mathbf{1}_X + u_u \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \tfrac{1}{2}\mathbf{1}_Z)\ ,
\end{aligned}$$

where u_b and u_u are defined in Eq.(12) and Eq.(13).

Proof. Let initial distribution $x_0^T = [x_{0X}^T, x_{0Z}^T]$ and $g^{-T} = [g_X^{-T}, g_Z^{-T}]$. When $t \to \infty$, Eq. (1) can be written as

$$x^T = \lim_{t\to\infty} (P^t x_0)^T = [x_{XY}^T, x_Z^T] = [x_X^T + x_Y^T, x_Z^T],$$

where $x_X = \lim_{t\to\infty}(P_X^t x_{0X} + \sum_{i=0}^{t-1} P_X^i g_X^-)$, $x_Y = \lim_{t\to\infty}(P_Y^{(t)} x_{0Z} + \sum_{i=0}^{t-1} P_Y^{(i)} g_Z^-)$, and $x_Z = \lim_{t\to\infty}(P_Z^t x_{0Z} + \sum_{i=0}^{t-1} P_Z^i g_Z^-)$.

From Lemma 2, $\lim_{t\to\infty} P_X^t = \mathbf{0}$, thus $x_X = (I_X - P_X)^{-1} g_X^-$ holds for any ergodic G_Z. Since G_Z is ergodic, x_Z follows Theorem 1. Below we will focus on deriving x_Y, where the first part of x_Y satisfies Lemma 2, i.e.,

$$\lim_{t\to\infty} P_Y^{(t)} x_{0Z} = \begin{cases} \mathbf{0} & G_Z \text{ is strictly unbalanced} \\ u_b \hat{\pi}_{Z,S_Z}^T x_{0Z} & G_Z \text{ is balanced} \\ -u_u \hat{\pi}_{Z,S_Z}^T x_{0Z} & G_Z \text{ is anti-balanced, even } t \\ u_u \hat{\pi}_{Z,S_Z}^T x_{0Z} & G_Z \text{ is anti-balanced, odd } t. \end{cases}$$

The second part of x_Y can be further written down as

$$\begin{aligned} &\lim_{m\to\infty} \sum_{t=1}^{m} P_Y^{(t)} g_Z^- = \lim_{m\to\infty} \sum_{t=0}^{m-1} \sum_{i=0}^{t} (P_X^{t-i} P_Y P_Z^i) g_Z^- \\ &= \lim_{m\to\infty} \sum_{t=0}^{m-1} \sum_{i=0}^{m-t} (P_X^t P_Y P_Z^i) g_Z^- = \sum_{t=0}^{\infty} (P_X^t P_Y \sum_{i=0}^{\infty} P_Z^i) g_Z^- \end{aligned} \tag{24}$$

Now we discuss Eq.(24) under different balance structures of G_Z.
(1) G_Z **is strictly unbalanced.** From Lemma 2, $\lim_{t\to\infty} P^t = \mathbf{0}$. Then by Eq.(23) we directly obtain that $x_{XY} = \frac{1}{2}\mathbf{1}_X$. Applying Eq.(23) to $\sum_{i=0}^{\infty} P_Z^i g_Z^-$ in Eq.(24), we have

$$\lim_{m\to\infty} \sum_{t=1}^{m} P_Y^{(t)} g_Z^- = \frac{1}{2}(I_X - P_X)^{-1} P_Y \mathbf{1}_Z.$$

Thus, we obtain the following equation:

$$x_{XY} = x_X + x_Y = (I_X - P_X)^{-1}(g_X^- + \frac{1}{2} P_Y \mathbf{1}_Z) = \frac{1}{2}\mathbf{1}_X.$$

(2) G_Z **is balanced.** Using Eq.(22), we have

$$\lim_{m\to\infty} \sum_{t=1}^{m} P_Y^{(t)} g_Z^- = \frac{1}{2}(I_X - P_X)^{-1} P_Y (\mathbf{1}_Z - \hat{\mathbf{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T \mathbf{1}_Z).$$

Hence, we have

$$\begin{aligned} x_{XY} &= (I_X - P_X)^{-1}(g_X^- + \frac{1}{2} P_Y \mathbf{1}_Z) + u_b \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \frac{1}{2}\mathbf{1}_Z) \\ &= \frac{1}{2}\mathbf{1}_X + u_b \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \frac{1}{2}\mathbf{1}_Z). \end{aligned} \tag{25}$$

(3) G_Z **is anti-balanced.** Using Lemma 3, we can negate the signs of all edges in G so that the sink becomes balanced. Hence, we know that at even steps in long term,

$$x_{XY,e} = \frac{1}{2}\mathbf{1}_X - u_u \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \frac{1}{2}\mathbf{1}_Z), \tag{26}$$

where Eq.(26) and Eq.(25) are identical in the sense that P_X's and P_Y's in Eq.(26) and Eq.(25) have opposite signs. Moreover, the odd step influence distribution sequence is obtained

$$x_{XY,o} = P_X x_{XY,e} + P_Y x_{Z,e} + g_X^- = \frac{1}{2}\mathbf{1}_X + u_u \hat{\pi}_{Z,S_Z}^T (x_{0Z} - \frac{1}{2}\mathbf{1}_Z). \tag{27}$$

□

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. Using Eq.(15), Theorem 1 and Theorem 3 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 [29, 31–33].

4 Influence maximization

With the detailed analysis on voter model dynamics for signed digraphs, we are ready now 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

Influence maximization objectives. 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:

$$f_t(x_0) := \mathbf{1}^T x_t(x_0) \text{ and } \bar{f}_t(x_0) := \frac{\sum_{i=0}^{t} f_i(x_0)}{t+1}. \tag{28}$$

Moreover, we define *l*ong term influence as

$$f(x_0) := \lim_{t\to\infty} \frac{\sum_{i=0}^{t} f_i(x_0)}{t+1}. \tag{29}$$

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* $W_t^* = \arg\max_{|W|\leq k} c_t(W)$. *Similarly, the problem for average influence and long-term influence is finding* $\bar{W}_t^* = \arg\max_{|W|\leq k} \bar{c}_t(W)$ *and* $W^* = \arg\max_{|W|\leq k} c(W)$*, respectively.*

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

$$c_t(W) = f_t(e_W) - f_t(\mathbf{0}) = \mathbf{1}^T x_t(e_W) - \mathbf{1}^T x_t(\mathbf{0}) = \mathbf{1}^T P^t e_W. \tag{30}$$

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:

$$c^T = \lim_{t\to\infty} \frac{\sum_{i=0}^t c_i^T}{t+1} = \lim_{t\to\infty} \frac{\mathbf{1}^T \sum_{i=0}^t P^i}{t+1}. \tag{31}$$

When P^t converges, we simply have

$$c^T = \mathbf{1}^T \lim_{t\to\infty} P^t. \tag{32}$$

Lemma 4 below discloses the important property that the influence contribution is a linear set function.

Lemma 4. *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* $c(W) = \sum_{i\in W} c(i)$.

Proof. From Eq.(30), we have

$$c_t(W) = \mathbf{1}^T P^t e_W = \mathbf{1}^T P^t \sum_{i\in W} e_i = \sum_{i\in W} \mathbf{1}^T P^t e_i = \sum_{i\in W} c_t(i).$$

The linearity of $\bar{c}_t$ and c can be derived from that of c_t. □

Given a vector v, let $n^+(v)$ denote the number of positive entries in v. By applying Lemma 4, 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 4, 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

1: **INPUT:** Signed transition matrix P, short-term period t, budget k;
2: **OUTPUT:** White seed set W.
3: $c_t = \mathbf{1}$; $\bar{c}_t = \mathbf{1}$;
4: **for** $i = 1 : t$ **do**
5: $\quad c_t^T = c_t^T P$;(for instant influence maximization.)
6: $\quad \bar{c}_t = \bar{c}_t + c_t$; (for average influence maximization.)
7: $W =$ 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 $O(t \cdot |E|)$. Note that when t is sufficiently large (e.g., $t > |V|^2$), the FOR loop in Algorithm 1 can be modified as follows to reduce the number of FOR Loops to $O(\log t)$, and the total complexity to $O(\log t|V||E|)$.

$P_0 = P$;
for $i = 1 : \lfloor \log t \rfloor$ **do**
$\quad P = P^2$;(for instant influence maximization.)

$c_t^T = c_t^T P$;(for instant influence maximization.)
$\bar{c}_t = \bar{c}_t + c_t$; (for average influence maximization.)
for $i = 2^{\lfloor \log t \rfloor} + 1 : t$ **do**
$c_t^T = c_t^T P_0$;(for instant influence maximization.)
$\bar{c}_t = \bar{c}_t + c_t$; (for average influence maximization.)

$P = P^2$ in the first loop is performed by matrix multiplication in $O(|E||V| + |E|) \approx O(|E||V|)$, instead of vector-matrix multiplication in $O(|E|)$. Thus, the overall computation complexity is $O(\log t|V||E|)$. Hence, when $t > \log t|V|$, the modified algorithm 1 works better; otherwise, the original algorithm1 is better. When $t > |V|^2$, we have $t > \log t|V|$ for any $|V| > 1$, thus the modified algorithm 1 is better.

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 5 below characterizes the long-term influence contributions of nodes, with respect to various balance structures.

Lemma 5. *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*, $c = \mathbf{0}$.*

Proof. (1) When G is balanced, by Lemma 1 and Eq.(32),

$$c^T = \mathbf{1}^T \lim_{t\to\infty} P^t = \mathbf{1}^T \hat{\mathbf{1}}_S \hat{\pi}_S^T = (|S| - |\bar{S}|)\hat{\pi}_S.$$

(2) When G is strictly unbalanced, again by Lemma 1 and Eq.(32), we have $c^T = \mathbf{1}^T \lim_{t\to\infty} P^t = \mathbf{0}$.

(3) When G is anti-balanced, by Lemma 1 and Eq.(31), we have

$$c^T = \mathbf{1}^T \frac{\lim_{t\to\infty} P^{2t} + \lim_{t\to\infty} P^{2t+1}}{2} = \mathbf{0}.$$

□

Based on Lemma 5, 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 5 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.

Theorem 1 indicates that given an anti-balanced digraph G, with bipartition S and $\bar{S}$, the long-term dynamic x_t oscillates on odd and even steps, and their long-term influence contribution is 0. Define the strength of the oscillation as the difference of the long-term influence between the even vs odd steps, i.e., $\Delta f_{o,e} = |f_o(x_0) - f_e(x_0)|/2$. We can maximize such strength of the oscillation of the voter model on an anti-balanced ergodic digraph by properly choosing the initial white seeds (See Remark 1.)

Remark 1. *In an anti-balanced ergodic digraph $G = (V, E, A)$ with the bipartition S and $\bar{S}$ and a budget k. Let W' (resp. W'') denote two initial seed sets, ~~where~~ $\min\{k, |S|\}$ (resp. $\min\{k, |\bar{S}|\}$) nodes~~,~~ with highest stationary distribution $\pi(i)$'s in S (resp. $\bar{S}$)~~, are selected.~~ Then, the optimal W^* that maximizes the strength of oscillation $\Delta f_{o,e}$ is*

$$W^* := \underset{W\in\{W',W''\}}{\operatorname{argmax}} |\hat{\pi}_S^T(e_W - \frac{1}{2}\mathbf{1})|. \tag{33}$$

Proof. From Theorem 1, when t becomes sufficiently large, the vector x oscillates at two vectors on odd and even steps, respectively. Rewrite the strength of the oscillation as

$$\Delta f_{o,e} = \frac{|f_o(x_0) - f_e(x_0)|}{2} = |\mathbf{1}^T \frac{x_o(x_0) - x_e(x_0)}{2}| = |\mathbf{1}^T \hat{\mathbf{1}}_S \hat{\pi}_S^T (x_0 - \frac{1}{2}\mathbf{1})|$$
$$= ||S| - |\bar{S}|| \cdot |\hat{\pi}_S^T (x_0 - \frac{1}{2}\mathbf{1})|.$$

Let W be the initial seed set, then the oscillation strength maximization is formulated as

$$\max_{|W|\leq k} ||S| - |\bar{S}|| \cdot |\hat{\pi}_S^T (e_W - \frac{1}{2}\mathbf{1})|$$
$$= ||S| - |\bar{S}|| \cdot \max\{\max_{|W|\leq k}\{\hat{\pi}_S^T e_W\} - \frac{1}{2}\hat{\pi}_S^T \mathbf{1}, \max_{|W|\leq k}\{-\hat{\pi}_S^T e_W\} + \frac{1}{2}\hat{\pi}_S^T \mathbf{1}\}, \quad (34)$$

which contains two sub-problems, i.e., $\max_{|W|\leq k}\{\hat{\pi}_S^T e_W\}$ and $\max_{|W|\leq k}\{-\hat{\pi}_S^T e_W\}$. The first maximization problem can be rewritten as

$$\max_{|W|\leq k}\{\hat{\pi}_S^T e_W\} = \max_{|W|\leq k}\big(\sum_{i\in S} \pi(i) e_W(i) - \sum_{j\in \bar{S}} \pi(i) e_W(j)\big). \quad (35)$$

Thus, let W' denote the optimal solution to the problem in Eq.(35), which is obtained by choosing $\min\{k, |S|\}$ seeds with highest $\pi(i)$'s from S. Similarly, choosing $\min\{k, |\bar{S}|\}$ nodes with the highest $\pi(i)$'s from $\bar{S}$ yields the optimal solution, denoted by W'', to the second maximization problem $\max_{|W|\leq k}\{-\hat{\pi}_S^T e_W\}$. The optimal W to the problem in eq.(34) that maximizes the oscillation strength is the one in $\{W', W''\}$, with higher $|\hat{\pi}_S^T (e_W - \frac{1}{2}\mathbf{1})|$, which completes the proof of eq.(33).

□

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 6. *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*, $c = \mathbf{0}$.*

Proof. (1)When G_Z is balanced, by Lemma 2, $c_X = \mathbf{0}_X$, and

$$c_Z^T = (\mathbf{1}_X^T u_b + \mathbf{1}_Z^T \hat{\mathbf{1}}_{Z,S_Z}) \hat{\pi}_{Z,S_Z}^T = (\mathbf{1}_X^T u_b + |S_Z| - |\bar{S}_Z|) \hat{\pi}_{Z,S_Z}^T.$$

(2) When G_Z is strictly unbalanced, $c^T = \mathbf{1}^T \lim_{t\to\infty} P^t = \mathbf{0}$

(3) When G_Z is anti-balanced, by Lemma 2 the limits of odd and even subsequences of P^t cancel out, thus $c = \mathbf{0}$. □

Lemma 6 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.

Theorem 3 shows that in a weakly connected signed digraph G, with single anti-balanced sink component G_Z, the long term influence $f(x_0)$ oscillates on odd and even steps, and the average is $|V|/2$, which is invariant to the initial seed selection. Similar to Remark 1, we can maximize the oscillation strength by properly selecting initial seeds, i.e.,

$$\begin{aligned} W^* &= \operatorname*{argmax}_{|W|\le k} |f_e(e_W) - f_o(e_W)|/2 \\ &= \operatorname*{argmax}_{|W|\le k} |(\mathbf{1}_X^T u_u \hat{\pi}_{Z,S_Z}^T + \mathbf{1}_Z^T \hat{\mathbf{1}}_{Z,S_Z} \hat{\pi}_{Z,S_Z}^T)(e_{WZ} - \frac{1}{2}\mathbf{1}_Z)| \\ &= |\mathbf{1}_X^T u_u + |S_Z| - |\bar{S}_Z|| \cdot \operatorname*{argmax}_{|W|\le k} |\hat{\pi}_{Z,S_Z}^T (e_{WZ} - \frac{1}{2}\mathbf{1}_Z)| \end{aligned} \tag{36}$$

where the maximization objective is independent from x_{0X}, thus oscillation strength maximization problem objective in Eq.(36) for G is identical to that in Remark 1. Hence, Remark 1 also applies here.

Using Eq.(15), Lemma 5 and Lemma 6 can be readily extended to the case with more than one ergodic sink components and disconnected digraphs. 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.

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 influ-

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 $\bar{P}_{Zi}$;
    $u_{bi} = (I_X - P_X)^{-1} P_{Yi} \hat{\mathbf{1}}_{Zi,S_{Zi}}$;
    $c_{Zi} = (\mathbf{1}_X^T u_{bi} + |S_{Zi}| - |\bar{S}_{Zi}|)\hat{\pi}^T_{Zi,S_{Zi}}$;
$c = [\mathbf{0}_X; c_{Z1}; \cdots; c_{Zm}]$

ence 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 6. 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 [41] 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. (Note that the convergence time of ergodic digraphs could be exponentially large in general, as illustrated by an example in Appendix C). 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.

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.

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

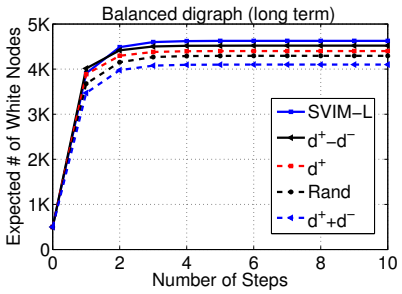

Figure 1: G is balanced

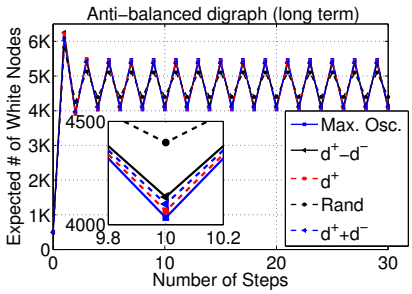

Figure 2: G is anti-balanced

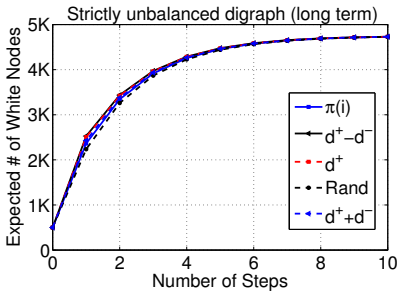

Figure 3: G is strictly unbalanced

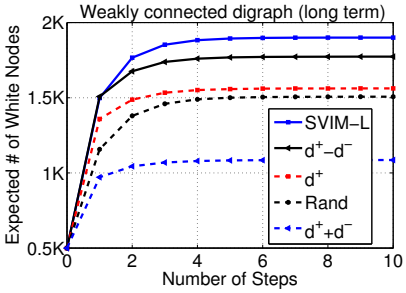

Figure 4: G is weakly connected

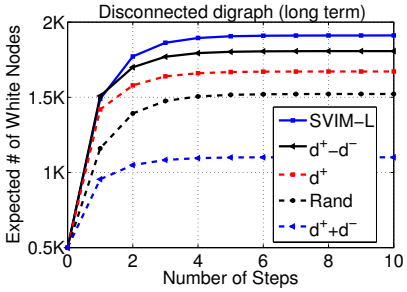

Figure 5: G is disconnected

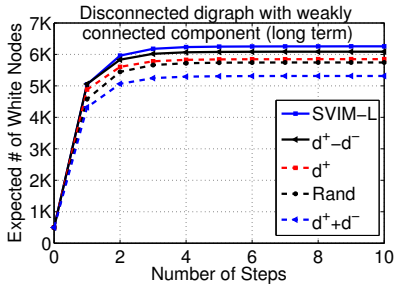

Figure 6: G is disconnected with WCC

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 maxi-

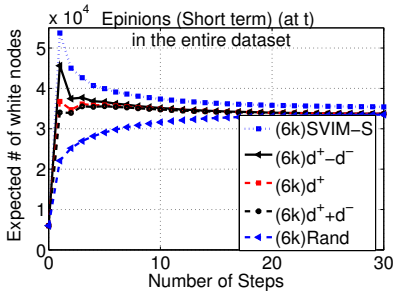

Figure 7: Instant influence in Epinions data with $k = 6$k

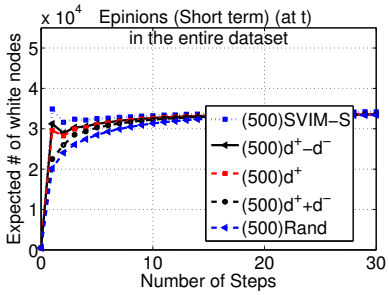

Figure 8: Instant influence in Epinions data with $k = 500$

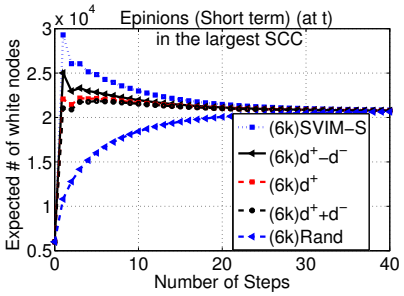

Figure 9: Instant influence in SCC with $k = 6$k

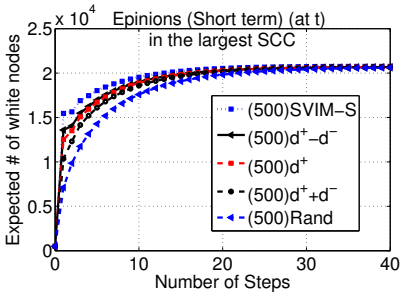

Figure 10: Instant influence in SCC with $k = 500$

mum 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. 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

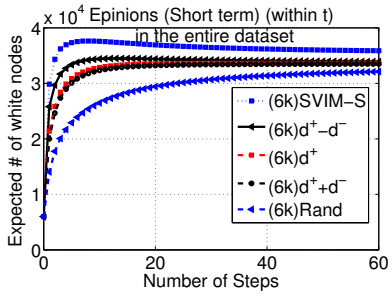

Figure 11: Average influence in Epinions data with $k = 6$k

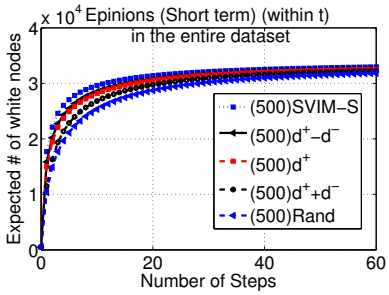

Figure 12: Average influence in Epinions data with $k = 500$

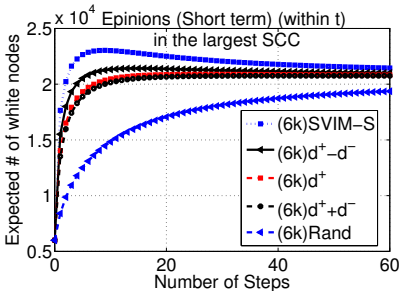

Figure 13: Average influence in SCC with $k = 6$k

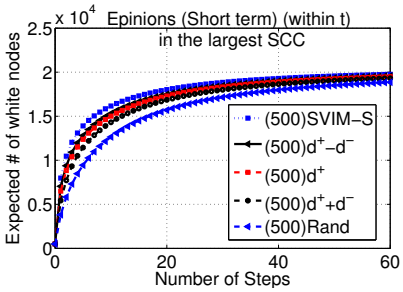

Figure 14: Average influence in SCC with $k = 500$

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 and Slashdot datasets

Statistics	Epinions	Slashdot
# of nodes	131580	77350
# of edges	840799	516575
# of positive edges	717129	396378
# of negative edges	123670	120197
# of nodes in largest SCC	41441	26996
# of edges in largest SCC	693507	337351
# of positive edges in largest SCC	614314	259891
# of negative edges in largest SCC	79193	77460
# of strongly connected components	88361	49209

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 [16] 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 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.

Slashdot Dataset. Slashdot.org [39] provides a discussion forum on various technology-related topics, where members can submit their stories, and comment on other members' stories. Its Slashdot Zoo feature allows members to tag each other as friends or foes, which in turn forms a signed online social network. The network was collected on 6-th November 2008 [26] and the statistics are shown in Table 2.

We evaluate instant influence and average influence of our SVIM-S algorithm on the entire slashdot dataset and its largest strongly connected component, respectively. Our results for $k = 6000$ are presented in Fig. 15-Fig. 18, which show that our SVIM-S algorithm performs the best among all methods tested, especially in the early steps. When changing the budget k, similar results were obtained, where we omitted them here for brevity.

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

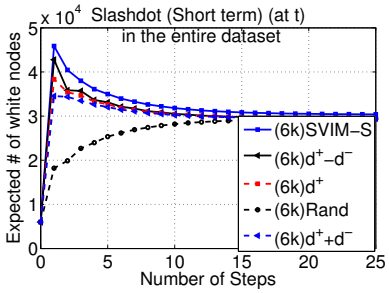

Figure 15: Instant influence in Slashdot data with $k = 6$k

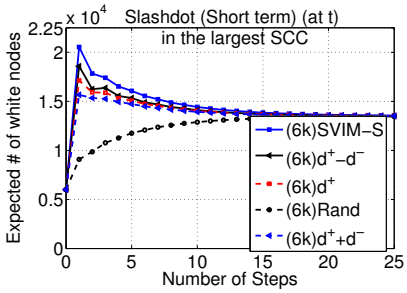

Figure 16: Instant influence in Slashdot SCC with $k = 6$k

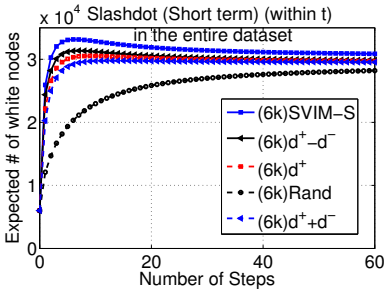

Figure 17: Average influence in Slashdot data with $k = 6$k

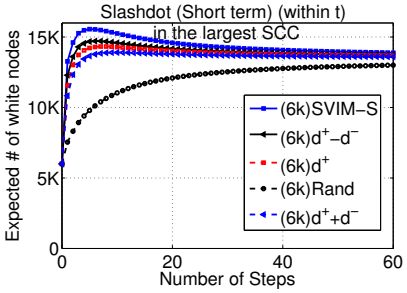

Figure 18: Average influence in Slashdot SCC with $k = 6$k

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. 19-22 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. 19), and $11.7\% - 58\%$ more influence in weakly connected

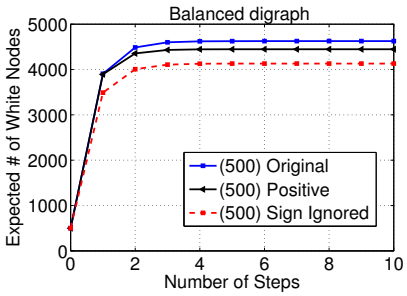

Figure 19: Synthetic balanced digraph

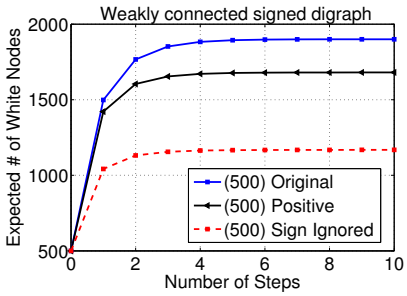

Figure 20: Synthetic weakly connected digraph

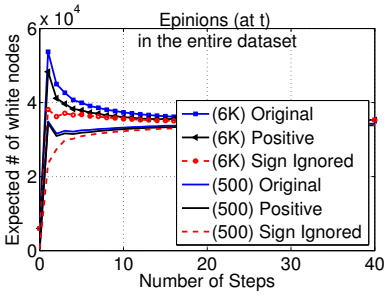

Figure 21: Epinions (the entire dataset) digraph

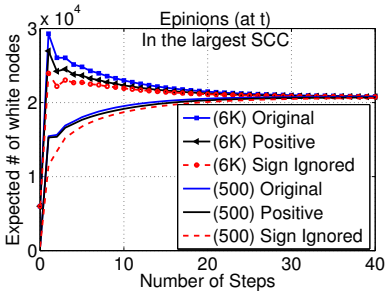

Figure 22: Epinions (the largest SCC)

for $t \geq 6$ (See Fig. 20). Moreover, in Epinions dataset from Fig. 21-22, 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. 21) and the largest SCC (See Fig. 22), 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 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. Extensive simulation results on both synthetic and real-world graphs demonstrate the efficacy of our signed voter model influence maximization (SVIM) algorithms. We also identify a class of anti-balanced

digraphs, which is not covered in the social balance theory before, and exhibits oscillating steady state behavior.

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 model and concepts in physics, such as Ising model and Gauge transformations. We also thank Zhenming Liu for many useful discussions on this work. This work was mostly done while the first author was working as full-time intern at Microsoft Research Asia.

This work was supported in part by the US NSF grant 0831734, CNS-1017647, the DTRA grant HDTRA1- 09-1-0050, and a DoD ARO MURI Award W911NF-12-1-0385.

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. In *WINE*, 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. In *CIKM*, 2011.

[12] F. Chung and A. Tsiatas. Hypergraph coloring games and voter models. *WAW '12: Proceedings of the 9th Workshop on Algorithms and Models for the Web Graph*, pages 1–16, 2012.

[13] F. R. K. Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9:1–19, sep 2005.

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

[15] D. Easley and J. Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge, 2010.

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

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

[18] A. Goyal, F. Bonchi, and L. V. S. Lakshmanan. A data-based approach to social influence maximization. *PVLDB*, 5(1):73–84, 2008.

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

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

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

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

[23] M. Kimura and K. Saito. Tractable models for information diffusion in social networks. In *PKDD*, 2006.

[24] 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.

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

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

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

[28] Y. Li, W. Chen, Y. Wang, and Z.-L. Zhang. Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships. In *WSDM '13: Proceedings of the 6th ACM International Conference on Web Search and Data Mining*. ACM, 2013.

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

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

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

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

[33] 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.

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

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

[36] 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.

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

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

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

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

[41] W. Stewart. Numerical methods for computing stationary distributions of finite irreducible markov chains. *Computational Probability*, 2000.

A Properties of ergodic digraphs

Proposition 4. *Let $G = (V, E, A)$ be an ergodic digraph. For any nodes $i, j \in V$, there exist two paths from i to j with even and odd length, respectively.*

Proof. Suppose, for a contradiction, that all paths from i to j have even lengths. This implies that all cycles passing through i must be even length, since otherwise we could follow node i's odd-length cycle followed by the even length path from i to j, making the entire path from i to j odd. Now we can consider any cycle C_r in G, not necessarily passing i. We claim that C_r must have even length. In fact, we can pick any node u on C_r, and construct a path from i to j with the following segments: R_1 from i to u, C_r, R_2 from u back to i, and R_3 from i to j. Since we know that $R_1 + R_2$ has even length and R_3 has even length, it must be the case that C_r has even length by our assumption. However, this means that all cycles in C has even lengths, contradicting to the aperiodicity of G.

The case of odd length paths can be proved in the same way. □

Proposition 5. *Let $\bar{G} = (V, E, \bar{A})$ be an ergodic unsigned digraph, with transition probability matrix $\bar{P}$ and stationary distribution vector π. $\bar{P}^t - \mathbf{1}\pi^T = (\bar{P} - \mathbf{1}\pi^T)^t$ holds for any integer $t > 0$.*

Proof. Using the facts that $\bar{P}\mathbf{1} = \mathbf{1}$ and $\pi^T\bar{P} = \pi^T$, it is easy to prove by induction that for any integer $t > 0$ $\bar{P}^t - \mathbf{1}\pi^T = (\bar{P} - \mathbf{1}\pi^T)^t$ holds. □

B Special matrix power series

Proposition 6. *Let $X \in \mathbb{R}^{m\times m}$, $Y \in \mathbb{R}^{m\times n}$ and $Z \in \mathbb{R}^{n\times n}$. If $\lim_{t\to\infty} X^t = \lim_{t\to\infty} Z^t = \mathbf{0}$, the following equalities hold:*

$$\textit{(i)}\ \lim_{t\to\infty} \sum_{i=0}^{t-1} X^i = (I - X)^{-1}, \tag{37}$$

$$\textit{(ii)}\ \lim_{t\to\infty} \sum_{i=0}^{t-1} X^i Y Z^{t-1-i} = 0, \tag{38}$$

Proof. (i) Let $\rho(X)$ be the spectral radius of matrix X, i.e., the largest absolute value of the eigenvalues of X. Notice that $\lim_{t\to\infty} X^t = \mathbf{0}$ if and only if $\rho(X) < 1$.

We first claim that, $I - X$ and $I - Z$ are invertible. Suppose $I - X$ is not invertible, there is a non-zero vector p such that $(I - X)p = \mathbf{0}$. Therefore, p is the eigenvector of X with eigenvalue 1, which contradicts $\lim_{t\to\infty} X^t = \mathbf{0}$. Same argument can be applied to $I - Z$. Hence, the left hand side of Eq.(37) equals to

$$\lim_{t\to\infty} \sum_{i=0}^{t} X^i = \lim_{t\to\infty} (I - X)^{-1}(I - X^{t+1}) = (I - X)^{-1}.$$

(ii) The max-norm of X is given by $\|X\|_{max} = \max_{i,j\le m}\{X_{ij}\}$. Let $X = Q_X \mathcal{J} Q_X^{-1}$ be the standard Jordan form of X, where Q_X is an invertible matrix. Denote $J = \mathbf{1}\mathbf{1}^T$ as the all-one matrix. Hence, we have

$$\begin{aligned}\|X^i\|_{max} &= \|Q_X \mathcal{J}^i Q_X^{-1}\|_{max} \le \|Q_X\|_{max}\|Q_X^{-1}\|_{max}\|J\mathcal{J}^i J\|_{max} \\ &\le \|Q_X\|_{max}\|Q_X^{-1}\|_{max} m^2 \|\mathcal{J}^i\|_{max}\end{aligned}$$

$\mathcal{J}^i$ is in form as

$$\mathcal{J}^i = \begin{bmatrix} \lambda_1^i & C_i^1\lambda_1^{i-1} & C_i^2\lambda_1^{i-2} & 0 & 0 \\ 0 & \lambda_1^i & C_i^1\lambda_1^{i-1} & 0 & 0 \\ 0 & 0 & \lambda_1^i & 0 & 0 \\ 0 & 0 & 0 & \lambda_{m_0}^i & C_i^1\lambda_{m_0}^{i-1} \\ 0 & 0 & 0 & 0 & \lambda_{m_0}^i \end{bmatrix}, \tag{39}$$

where $C_i^\ell = \frac{i!\ell!}{(i-\ell)!} \le i^m$ and each non-zero entry in $\mathcal{J}^i$ can be expressed as $C_\ell^i \lambda_k^{i-\ell}$, $1 \le k \le m_0$, $1 \le \ell \le \ell_0(k)$, with m_0 as the number of different eigenvalues of X and $\ell_0(k)$ as the multiplicity of the k-th eigenvalue of X. Hence, the absolute value of each non-zero entry in $\mathcal{J}^i$ is upper bounded as $|C_i^\ell \lambda_k^{i-\ell}| \le i^m \rho(X)^{i-m}$, which implies that

$$\|X^i\|_{max} \le \|Q_X\|_{max}\|Q_X^{-1}\|_{max} m^2 i^m \rho(X)^{i-m}$$

Let $\rho = \max(\rho(X), \rho(Z))$, we have

$$\begin{aligned}&\lim_{t\to\infty} \|\sum_{i=0}^{t-1} X^i Y Z^{t-1-i}\|_{max} \le \lim_{t\to\infty} tmn\|X^i\|_{max}\|Y\|_{max}\|Z^{t-1-i}\|_{max} \\ &\le \lim_{t\to\infty} tmnT_{max}(m^2 t^m \rho^{i-m})(n^2 t^n \rho^{t-i-1-n}) \le \lim_{t\to\infty} m^3 n^3 T_{max} t^{m+n+1}\rho^{t-1-n-m} = 0\end{aligned}$$

where $T_{max} = \|Y\|_{max}\|Q_X\|_{max}\|Q_X^{-1}\|_{max}\|Q_Z\|_{max}\|Q_Z^{-1}\|_{max}$.

□

C Illustration of exponential convergence time of P^t on ergodic digraph.

Given an unsigned ergodic digraph $\bar{G} = (V, E, \bar{A})$, with transition probability matrix $\bar{P}$, it has fixed stationary distribution π, i.e., $\pi^T = \pi^T \bar{P}$.

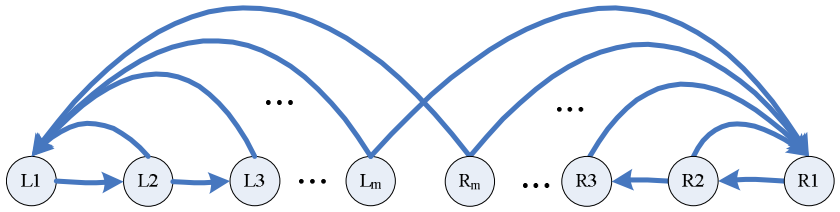

Figure 23: An example digraph with exponential convergence time. All edges are with unit weights.

The convergence time (or mixing time) of a random walk Markov chain on G is the time until the Markov chain is "close" to its stationary distribution π. To be precise, for an initial distribution x_0, let $x_t^T = x_0^T \bar{P}^t$ be the distribution at step t. The variation distance mixing time is defined as the smallest t such that for any subset $W \subseteq V$,

$$|(x_t^T - \pi^T)e_W| \leq \frac{1}{4},$$

where e_W is the vector such that $e_W(i) = 1$ if $i \in W$, and $e_W(i) = 0$ if $i \in V \setminus W$.

The convergence time is said to be exponentially large if there exists x_0 such that the convergence time of the random walk starting from x_0 is $2^{\Omega(n)}$, where $n = |V|$. Lemma 7 below illustrates that the convergence time of random walk on ergodic digraphs could be exponentially large.

Lemma 7. *There exist ergodic digraphs, such that the convergence time of the random walks on these digraphs are exponentially large.*

Proof. We prove this by construction. Fig. 23 shows an example digraph G, with $|V| = 2m$ nodes. On the left hand side, there are $m \geq 3$ nodes $L_1, L_2, \cdots, L_m$ connected by $m-1$ directed edges from L_1 to L_m, and every node L_i with $i > 1$ has a directed connection to the leftmost node L_1. The right hand side nodes have symmetric connections as the left hand side. Moreover, node L_m and R_m also have one more connection to R_1 and L_1, respectively, which connect two components together. It is clear that the graph is strongly connected and aperiodic (there exist cycles of length 2 and 3), and thus ergodic.

Let $x_t(L_i)$ denote the probability that the random walk is at node L_i at step t, and $x(L_i)$ be its stationary distribution. Similarly define $x_t(R_i)$ and $x(R_i)$ for node R_i. The graph is symmetric, thus we have $x(L_i) = x(R_i)$ for $1 \leq i \leq m$. Let $x(L_1) = x(R_1) = \rho/4$, we have $x(L_i) = x(R_i) = \rho/2^i$ for $i = 2, 3, \ldots, m$. Then, by solving $\sum_{i=1}^{m}(x(L_i) + x(R_i)) = 1$, we obtain $\rho = \frac{2^{m-1}}{3 \cdot 2^{m-2}-1}$. It is easy to

verify that indeed the obtained x is the stationary distribution of the random walks on the digraph.

Then, we consider the initial distribution as $x_0 = [1, 0, 0, \ldots, 0]$, and the subset $W = \{R_1, \cdots, R_m\}$ including all m nodes on the right-hand side. Let $x_t(W) = x_t^T \cdot e_W$ denote the total probability that the random walk is in some node in W at step t. The only edge from the left half to the right half is the edge from L_m to R_1. Thus all additions to $x_{t+1}(W)$ from $x_t(W)$ comes from this edge, namely $x_{t+1}(W) - x_t(W) \leq x_t(L_m)/2$. We now bound $x_t(L_m)$. For $t \leq m-1$, we know that $x_t(L_m) = 0$. For $t \geq m$, we have

$$x_t(L_m) = x_{t-1}(L_{m-1})/2 = x_{t-2}(L_{m-2})/2^2 = \cdots = x_{t-m+2}(L_2)/2^{m-2} \leq 1/2^{m-2}.$$

Hence, we have

$$x_t(W) = \sum_{i=1}^{t} (x_i(W) - x_{i-1}(W)) \leq t \cdot x_t(L_m)/2 \leq t/2^{m-1}.$$

Therefore, the smallest t that satisfies $|(x_t^T - \pi^T)e_W| = |x_t(W) - 1/2| \leq 1/4$ is such that $x_t(W) \geq 1/4$, which implies that $t/2^{m-1} \geq 1/4$ and $t \geq 2^{m-3}$. This completes the proof. $\square$