# Fast Evaluation for Relevant Quantities of Opinion Dynamics

Wanyue Xu
Fudan University
Shanghai, China
xuwy@fudan.edu.cn

Qi Bao
Fudan University
Shanghai, China
20110240002@fudan.edu.cn

Zhongzhi Zhang*
Fudan University
Shanghai, China
zhangzz@fudan.edu.cn

## ABSTRACT

One of the main subjects in the field of social networks is to quantify conflict, disagreement, controversy, and polarization, and some quantitative indicators have been developed to quantify these concepts. However, direct computation of these indicators involves the operations of matrix inversion and multiplication, which make it computationally infeasible for large-scale graphs with millions of nodes. In this paper, by reducing the problem of computing relevant quantities to evaluating $\ell_2$ norms of some vectors, we present a nearly linear time algorithm to estimate all these quantities. Our algorithm is based on the Laplacian solvers, and has a proved theoretical guarantee of error for each quantity. We execute extensive numerical experiments on a variety of real networks, which demonstrate that our approximation algorithm is efficient and effective, scalable to large graphs having millions of nodes.

## CCS CONCEPTS

• **Human-centered computing** → **Social network analysis**; • **Networks** → **Network dynamics**; • **Information systems** → **Social networks**.

## KEYWORDS

Opinion dynamics, social network, multi-agent system, polarization, disagreement, conflict, controversy, Laplacian solver

## 1 INTRODUCTION

Online social networks and social media are increasingly becoming an important part of our lives, which have led to a fundamental change of ways people share and shape opinions [1, 13, 23]. Particularly, the enormous popularity of social media and online social networks produces diverse social phenomena, such as polarization, disagreement, conflict, and controversy, which have been a hot

subject of study in different disciplines, especially social science. In fact, some phenomena, for example, disagreement and polarization, have taken place in human societies for millenia, but now they are more apparent in an online virtual world.

In addition to the identification of aforementioned social phenomena, the issue of how to quantify these phenomena has received increasing amounts of attention. Thus far, various measures have been developed to quantify these phenomena, such as disagreement [10, 31], polarization [10, 29, 31], conflict [7], and controversy [7]. Most of these measures are based on the Friedkin-Johnsen (FJ) social-opinion dynamics model [17], which is an important extension of the DeGroot's opinion model [12]. Although the expressions of these quantitative metrics seem very concise, rigorous determination for them in large-scale graphs is a computational challenge, since it involves matrix inversion and multiplication.

In this paper, we address the problem of fast calculation for the aforementioned quantitative measures related to opinion dynamics modelled by the well-established FJ model, by exploiting the connection [7, 31] between forest matrix [5, 6] and these key quantities. To this end, we first represent these quantities in terms of the $\ell_2$ norm of some vectors. We then provide an algorithm to approximate these quantities in nearly linear time with the number of edges. Our algorithm has a proved error guarantee. Extensive experiments on many real network datasets indicate that our algorithm is efficient and effective, which is scalable to large graphs with millions of nodes.

**Related work.** The focus of this paper is to propose a fast algorithm approximately evaluating the quantitative metrics of some key social phenomena. We use the FJ model as our underlying opinion dynamics model. Below, we review some work that is closely related to ours.

It is well known, the FJ model is a significant extension of the DeGroot model for opinion dynamics, where opinion many be the understanding or position of individuals on a certain popular topic or subject. The DeGroot model is an iterative averaging model, with each individual having only one opinion [12]. At each time step, any individual updates its opinion as a weighted average of its neighbors. For the DeGroot model on a connected graph, it will reach consensus [2]. Under the formalism of the DeGroot model or its variants, many consensus protocols have been proposed or studied [32], especially in the literature of systems [14, 15, 22, 28, 36] and cybernetics literature [34, 41–43].

Although the FJ model is an extension of the DeGroot model, the former is significantly different from the latter. In the DeGroot model, each node has only one opinion, while the FJ model associates each node with two opinions: internal opinion and expressed opinion. Since its establishment, the FJ model has attracted much attention. A sufficient condition for stability of the FJ model was obtained in [35], and the equilibrium expressed opinion was derived

in [3, 11]. Some interpretations of the FJ model were provided in [3] and [19]. And some optimization problems based on the FJ model were also introduced, such as opinion maximization [20]. Moreover, further extensions of the FJ model were suggested and studied in recent papers [23, 38]. For example, some multidimensional extensions have been presented for the FJ model [18, 33].

Except for the properties, interpretations, and extension of the FJ model, some measures for disagreement [31], polarization [29, 31], conflict [7], and controversy [7] for this popular model have also been developed and studied. These quantitative metrics provide deep insight into understanding social phenomena. However, exact computation for these key measures is difficult and even impossible for large graphs, since it takes cube running time. In this paper, we give a computationally cheaper approach for approximating these quantities.

## 2 PRELIMINARY

In this section, we briefly introduce some basic concepts about undirected weighted graphs, Laplacian matrix, spanning (rooted) forest, forest matrix, FJ opinion dynamics model, its relevant quantities and their connections with the forest matrix.

### 2.1 Graph and Laplacian Matrix

Let $\mathcal{G} = (V, E, w)$ be a connected undirected weighted simple graph with $n$ nodes and $m$ edges, where $V = \{v_1, v_2, \cdots, v_n\}$ is the node set, $E = \{e_1, e_2, \cdots, e_m\}$ is the edge set, and $w : E \to \mathbb{R}_+$ is the edge weight function, with the weight of an edge $e$ denoted by $w_e$. Let $w_{\max}$ and $w_{\min}$ denote, respectively, the maximum and minimum weight among all edges in $E$. A graph is a tree if it is connected but has no cycles. We consider a graph with only an isolated node as a tree. A forest is a particular graph that is a disjoint union of trees. Thus, a forest may be connected or disconnected. In the sequel, we interchangeably use $v_i$ and $i$ to represent node $v_i$ if incurring no confusion.

The connections of graph $\mathcal{G}$ are encoded in its extended adjacency matrix $A = (w_{ij})_{n \times n}$, with the element $w_{ij}$ at row $i$ and column $j$ representing the strength of connection between nodes $i$ and $j$. If nodes $i$ and $j$ are adjacent by an edge $e$ with weight $w_e$, then $w_{ij} = w_{ji} = w_e$; $w_{ij} = w_{ji} = 0$ otherwise. Let $\Theta_i$ be the set of neighbours of node $i$. Then the weighted degree of a node $i$ is $d_i = \sum_{j=1}^{n} w_{ij} = \sum_{j \in \Theta_i} w_{ij}$. The weighted diagonal degree matrix of $\mathcal{G}$ is defined as $D = \text{diag}(d_1, d_2, \ldots, d_n)$, and the Laplacian matrix of $\mathcal{G}$ is defined to be $L = D - A$.

An alternative construction of $L$ is to use the incidence matrix $B \in \mathbb{R}^{|E| \times |V|}$, which is an $m \times n$ signed edge-node incidence matrix. The entry $b_{ev}$, $e \in E$ and $v \in V$, of $B$ is defined as follows: $b_{ev} = 1$ if node $v$ is the head of edge $e$, $b_{ev} = -1$ if node $v$ is the tail of edge $e$, and $b_{ev} = 0$ otherwise. Let $e_i$ denote the $i$-th standard basis vector. For an edge $e \in E$ with two end nodes $i$ and $j$, the row vector of $B$ corresponding to $e$ can be written as $b_{ij} \triangleq b_e = e_i - e_j$. Let $W = \text{diag}(w_1, w_2, \ldots, w_m)$ be an $m \times m$ diagonal matrix with the $e$-th diagonal entry being the weight of edge $w_e$. Then the Laplacian matrix $L$ of $\mathcal{G}$ can also be represented as $L = B^\top W B$. Moreover, $L$ can be written as the sum of product of block matrices as $L = \sum_{e \in E} w_e b_e b_e^\top$, which indicates that $L$ is a symmetric and positive semidefinite matrix.

The positive semidefiniteness of Laplacian matrix $L$ implies that all its eigenvalues are non-negative. Moreover, for a connected graph $\mathcal{G}$, its Laplacian matrix $L$ has a unique zero eigenvalue. Let $\mathbf{1}$ denote the $n$-dimensional column vector with all entries being ones, i.e. $\mathbf{1} = (1, 1 \cdots, 1)^\top$, which is an eigenvector of $L$ associated with eigenvalue 0. That is, $L\mathbf{1} = \mathbf{0}$, where $\mathbf{0}$ is the zero vector. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{n-1} \geq \lambda_n = 0$ be the $n$ eigenvalues of $L$, and let $u_i$ be the orthogonal eigenvector corresponding to $\lambda_i$. Then, $L$ has an eigendecomposition of form $L = U \Lambda U^\top = \sum_{i=1}^{n-1} \lambda_i u_i u_i^\top$ where $\Lambda = \text{diag}(\lambda_1, \lambda_2, .., \lambda_{n-1}, 0)$ and $u_i$ is the $i$-th column of matrix $U$. Let $\lambda_{\max}$ and $\lambda_{\min}$ be, respectively, the maximum and nonzero minimum eigenvalue of $L$. Then, $\lambda_{\max} = \lambda_1 \leq n\, w_{\max}$ [40], and $\lambda_{\min} = \lambda_{n-1} \geq w_{\min}/n^2$ [27].

### 2.2 Spanning Forests and Forest Matrix

For a graph $\mathcal{G} = (V, E, w)$, a subgraph $\mathcal{H}$ is a graph whose sets of nodes and edges are subsets of $V$ and $E$, respectively. If $\mathcal{H}$ and $\mathcal{G}$ have the same node set $V$, we call $\mathcal{H}$ a spanning subgraph of $\mathcal{G}$. A spanning forest on $\mathcal{G}$ is a spanning subgraph of $\mathcal{G}$ that is a forest. A spanning rooted forest of $\mathcal{G}$ is a spanning forest of $\mathcal{G}$, where each tree has a node marked as its root. For a subgraph $\mathcal{H}$ of graph $\mathcal{G}$, the product of the weights of all edges in $\mathcal{H}$ is referred to as the weight of $\mathcal{H}$, denoted as $\varepsilon(\mathcal{H})$. If $\mathcal{H}$ has no edges, its weight is set to be 1. For any nonempty set $S$ of subgraphs, we define its weight $\varepsilon(S)$ as $\varepsilon(S) = \sum_{\mathcal{H} \in S} \varepsilon(\mathcal{H})$. If $S$ is empty, we set its weight to be zero [5, 6].

Suppose that $\Gamma$ is the set of all spanning forests of graph $\mathcal{G}$ and $\Gamma_{ij}$ is the set of those spanning forests of $\mathcal{G}$ with nodes $v_i$ and $v_j$ in the same tree rooted at node $v_i$. Based on the above notions associated with spanning rooted forests, we can define the forest matrix $\Omega = \Omega(\mathcal{G})$ of graph $\mathcal{G}$ [4, 21]. Let $I$ be the identity matrix. Then the forest matrix is defined as $\Omega = (I + L)^{-1} = (\omega_{ij})_{n \times n}$, where the entry $\omega_{ij} = \varepsilon(\Gamma_{ij})/\varepsilon(\Gamma)$ [5, 6]. For an arbitrary pair of nodes $v_i$ and $v_j$ in graph $\mathcal{G}$, $\omega_{ij} \geq 0$ with equality if and only if $\mathcal{G}$ is disconnected. Moreover, $\omega_{ij} = 0$ if and only if there is no path between $v_i$ and $v_j$ [30].

If every edge in $\mathcal{G}$ has unit weight, then $\varepsilon(\Gamma)$ is equal to the total number of spanning rooted forests of $\mathcal{G}$, and $\varepsilon(\Gamma_{ij})$ equals the number of spanning rooted forests of $\mathcal{G}$, where nodes $v_i$ and $v_j$ are in the same tree rooted at $v_i$. For example, in the 5-node path graph $\mathcal{P}_5$, there are exactly 55 spanning rooted forests, among which there are 13 forests where $v_2$ belongs to a tree rooted at $v_1$. Figure 1 illustrates all the 55 spanning rooted forests in $\mathcal{P}_5$, where the 13 spanning rooted forests with green background are those, for each of which $v_1$ and $v_2$ belong to the same tree with $v_1$ being the root. According to Fig. 1, the forest matrix for graph $\mathcal{P}_5$ is

$$\Omega = \frac{1}{55} \begin{pmatrix} 34 & 13 & 5 & 2 & 1 \\ 13 & 26 & 10 & 4 & 2 \\ 5 & 10 & 25 & 10 & 5 \\ 2 & 4 & 10 & 26 & 13 \\ 1 & 2 & 5 & 13 & 34 \end{pmatrix}.$$

The forest matrix $\Omega$ is a symmetric and positive definite matrix, the eigendecomposition of which can be written as $\Omega = U \tilde{\Lambda} U^\top = \sum_{i=1}^{n} \frac{1}{\lambda_i + 1} u_i u_i^\top$ where $\tilde{\Lambda}$ is a diagonal matrix given by $\tilde{\Lambda} = \text{diag}(\frac{1}{1 + \lambda_1},$
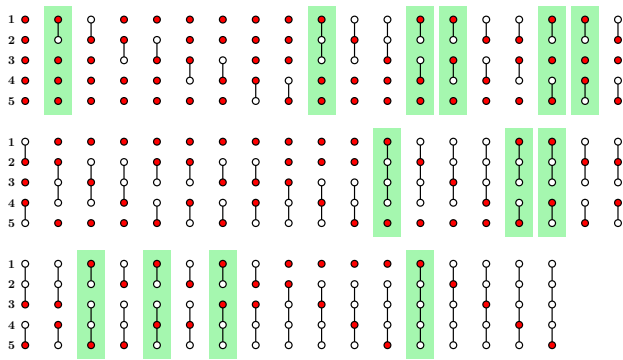
**Figure 1: All the spanning rooted forests in the path graph $\mathcal{P}_5$ with 5 nodes. Those with green background are forests with $v_2$ being in a tree rooted at $v_1$. The full nodes denote root nodes.**

$\cdots, \frac{1}{1+\lambda_{n-1}}, \frac{1}{1+\lambda_n})$ with $\frac{1}{1+\lambda_1} \leq \frac{1}{1+\lambda_2} \leq \cdots \leq \frac{1}{1+\lambda_n} = 1$. It has been shown that [5, 6] $\Omega$ is a doubly stochastic matrix satisfying $\Omega \mathbf{1} = \mathbf{1}$ and $\mathbf{1}^\top \Omega = \mathbf{1}^\top$. For any connected graph, $\omega_{ij} > 0$. Moreover, $\sum_{j=1}^n \omega_{ij} = 1$ for $i = 1, 2, \ldots, n$, and $\sum_{i=1}^n \omega_{ij} = 1$ or $j = 1, 2, \ldots, n$.

The forest matrix $\Omega$ is related to various practical applications [16, 24, 39]. For example, its entry $\omega_{ij}$ can be used to gauge the *proximity* between nodes $v_i$ and $v_j$: the less the value of $\omega_{ij}$, the "farther" $v_i$ from $v_j$ [5]. Furthermore, since $\Omega$ is doubly stochastic, $\omega_{ij}$ can be explained as the fraction of the connectivity of $v_i$ and $v_j$ in the total connectivity of $v_i$ (or $v_j$) with all nodes [6].

Particularly, forest matrix is in fact the fundamental matrix [29] of the FJ opinion dynamics model [17]. Various important quantities of the FJ model can be expressed in terms of the linear combination of the entries for forest matrix or quadric forms of forest matrix or its variant matrices [3, 7, 31]. In the sequel, we will show that by using the properties of forest matrix, one can provide a fast algorithm evaluating relevant quantities for the FJ opinion dynamics model.

# 3 FRIEDKIN-JOHNSEN OPINION DYNAMICS MODEL AND ITS RELEVANT QUANTITIES

This section is denoted to brief introduction to the FJ model of opinion formation, as well as the definitions and measures for conflict, disagreement, polarization, and controversy, relying on this popular model. Particularly, we give an explanation and some properties of equilibrium expressed opinions of the FJ model, using the forest matrix.

## 3.1 Friedkin-Johnsen Model

As one of the first opinion dynamics models, the FJ model [17] is an extension of the DeGroot's opinion model [12]. In the DeGroot model, every node has only one opinion that is updated as the weighted average of its neighbors. Different from the DeGroot model, in the FJ model, each node $i \in V$ has two different kinds of opinions: one is the internal (or innate) opinion $s_i$, the other is the expressed opinion $z_i$. The internal opinion $s_i$ is assumed to remain constant, private to node $i$, while the expressed opinion $z_i$

evolves as a weighted average of its corresponding internal opinion $s_i$ and the expressed opinions of $i$'s neighbors. More precisely, the updating rule of $z_i$ is

$$z_i = \frac{s_i + \sum_{j \in \Theta_i} w_{ij} z_j}{1 + \sum_{j \in \Theta_i} w_{ij}}. \tag{1}$$

Note that in the above updating process, we make a common assumption in the literature that the weight of internal opinion is unit. On the other hand, as popular choice in the literature, we assume that for all $i \in V$, its internal opinion $s_i$ is in the interval $[0, 1]$. Let $s = (s_1, s_2, \ldots, s_n)^\top$. Then the expressed opinions updated by iterative process (1) converge to a unique equilibrium opinion vector. Let $z = (z_1, z_2, \ldots, z_n)^\top$, with the value $z_i$ being the expressed opinion of node $i$ at equilibrium. It was shown [3] that the equilibrium expressed opinion vector $z$ is the solution to a linear system of equations:

$$z = (I + L)^{-1} s. \tag{2}$$

Equation (2) shows that the equilibrium expressed opinion for every node is determined by the forest matrix $\Omega = (I + L)^{-1}$, with $z_i$ given by $z_i = \sum_{j=1}^n \omega_{ij} s_j$, for each $i \in V$. Concretely, for all $i = 1, 2 \ldots, n$, $z_i$ is a weighted average of internal opinions of all nodes, with the weight of internal opinion $s_j$ being $\omega_{ij}$, where $j = 1, 2 \ldots, n$. Considering that $\Omega$ is a doubly stochastic and $s_i \in [0, 1]$ for all $i = 1, 2 \ldots, n$, we have that for every $i$, $z_i \in [0, 1]$. Moreover, $\sum_{i=1}^n s_i = \sum_{i=1}^n z_i$, which means that the total expressed opinion is equal to the total internal opinion, although the equilibrium expressed opinion for a single node may be different from its internal opinion. This conservation law is independent of the network structure. In this sense, we provide a novel interpretation and some properties of equilibrium expressed opinion vector $z$ according to the forest matrix. Our interpretation is different from previous ones, which are in terms of game theory [3] and electrical networks [19], respectively.

## 3.2 Measures for Conflict, Disagreement, Polarization, and Controversy

In the FJ model, the opinions of nodes often do not reach consensus, leading to conflict, disagreement, polarization, and controversy, which are common phenomena in social networks and have been the subject of many recent works. Below we survey some quantitative measures of these phenomena based on the FJ opinion formation model.

As known to us all, in the FJ model, individuals differ in their internal opinions and expressed opinions. The extent of this difference can be measured by internal conflict defined as follows [7].

DEFINITION 3.1. *For a graph $\mathcal{G} = (V, E, w)$, its internal conflict $C_I(\mathcal{G})$ is the sum of squares of the differences between internal and expressed opinions over all nodes:*

$$C_I(\mathcal{G}) = \sum_{i \in V} (z_i - s_i)^2. \tag{3}$$

DEFINITION 3.2. *[10, 31] For a graph $\mathcal{G} = (V, E, w)$, its disagreement $D(\mathcal{G})$ is defined by*

$$D(\mathcal{G}) = \sum_{(i,j) \in E} w_{ij} (z_i - z_j)^2. \tag{4}$$

Note that the disagreement $D(\mathcal{G})$ is called external conflict of graph $\mathcal{G}$ in [7]. We also note that the FJ model has been used to understand the price of anarchy in society when individuals selfishly update their opinions with an aim to minimize the stress they experience [3]. The stress of a node $i$ is defined as $(z_i - s_i)^2 + \sum_{j \in \Theta_i} w_{ij}(z_i - z_j)^2$, while the sum of the stress for all nodes is $C_{\mathrm{I}}(\mathcal{G}) + D(\mathcal{G})$, which is exactly the sum of internal conflict and external conflict defined above.

If the equilibrium expressed opinions have an increased divergence, we say that opinion formation dynamics are polarizing. Intuitively, polarization should measure how equilibrium expressed opinions deviate from their average. There are many ways to quantify polarization. We here choose the metric proposed in [31] to measure polarization.

DEFINITION 3.3. *For a graph $\mathcal{G} = (V, E, w)$, let $\bar{z}$ be the mean-centered equilibrium vector given by $\bar{z} = z - \frac{z^\top 1}{n} 1$. Then the polarization $P(\mathcal{G})$ is defined to be:*

$$P(\mathcal{G}) = \sum_{i \in V} \bar{z}_i^2 = \bar{z}^\top \bar{z}. \tag{5}$$

In addition to $P(\mathcal{G})$, the polarization can also be measured by the controversy that quantifies how much the expressed opinion varies across the individuals in the whole graph $\mathcal{G}$.

DEFINITION 3.4. *For a graph $\mathcal{G} = (V, E, w)$, the controversy $C(\mathcal{G})$ is the sum of the squares of the equilibrium expressed opinions:*

$$C(\mathcal{G}) = \sum_{i \in V} z_i^2 = z^\top z. \tag{6}$$

In [29], the quantity $C(\mathcal{G})/n$ is introduced as the polarization index.

There can be a tradeoff between disagreement and controversy [31]. The sum of the disagreement and controversy is called disagreement-controversy index, which is named polarization-disagreement index in [31].

DEFINITION 3.5. *For a graph $\mathcal{G} = (V, E, w)$, the disagreement-controversy index $I_{\mathrm{dc}}(\mathcal{G})$ is the sum of the disagreement $D(\mathcal{G})$ and controversy $C(\mathcal{G})$ :*

$$I_{\mathrm{dc}}(\mathcal{G}) = D(\mathcal{G}) + C(\mathcal{G}). \tag{7}$$

It is easy to verify [7, 31] that the disagreement-controversy index $I_{\mathrm{dc}}(\mathcal{G})$ is equal to the inner product between internal opinion vector $s$ and expressed opinion vector $z$, that is, $I_{\mathrm{dc}}(\mathcal{G}) = \sum_{i=1}^n s_i z_i$.

Let $\bar{s} = s - \frac{1^\top s}{n} 1$. The above-mentioned quantities can be expressed in matrix-vector notation as stated in Proposition 1 [7, 31].

PROPOSITION 1. *For a graph $\mathcal{G} = (V, E, w)$, the internal conflict $C_{\mathrm{I}}(\mathcal{G})$, disagreement $D(\mathcal{G})$, polarization $P(\mathcal{G})$, controversy $C(\mathcal{G})$, and disagreement-controversy index $I_{\mathrm{dc}}(\mathcal{G})$ can be conveniently expressed in terms of quadratic forms as:*

$$C_{\mathrm{I}}(\mathcal{G}) = z^\top L^2 z = s^\top (I + L)^{-1} L^2 (I + L)^{-1} s, \tag{8}$$

$$D(\mathcal{G}) = z^\top L z = s^\top (I + L)^{-1} L (I + L)^{-1} s, \tag{9}$$

$$P(\mathcal{G}) = \bar{z}^\top \bar{z} = \bar{s}^\top (I + L)^{-1} (I + L)^{-1} \bar{s}, \tag{10}$$

$$C(\mathcal{G}) = z^\top z = s^\top (I + L)^{-1} (I + L)^{-1} s, \tag{11}$$

$$I_{\mathrm{dc}}(\mathcal{G}) = s^\top z = s^\top (I + L)^{-1} s. \tag{12}$$

Notice that matrices $L$ and $(I + L)^{-1}$ have identical eigenspaces, implying that they commute, that is, $L(I + L)^{-1} = (I + L)^{-1} L$. Thus, we have

$$C_{\mathrm{I}}(\mathcal{G}) = \bar{s}^\top (I + L)^{-1} L^2 (I + L)^{-1} \bar{s},$$

$$D(\mathcal{G}) = \bar{s}^\top (I + L)^{-1} L (I + L)^{-1} \bar{s},$$

due to $L1 = 0$.

After expressing the quantities concerned, in what follows we will provide a fast algorithm evaluating these quantities.

## 4 FAST APPROXIMATION ALGORITHM FOR CONFLICT, DISAGREEMENT, POLARIZATION, AND CONTROVERSY

As shown in Proposition 1, for internal conflict $C_{\mathrm{I}}(\mathcal{G})$, disagreement $D(\mathcal{G})$, polarization $P(\mathcal{G})$, controversy $C(\mathcal{G})$, and disagreement-controversy index $I_{\mathrm{dc}}(\mathcal{G})$, exactly computing them needs to invert matrix $L + I$, which takes $O(n^3)$ time. This is computationally impractical for large graphs.

In this section, we develop a fast algorithm for approximately evaluating those interesting quantities in nearly linear time with respect of the number of edges in $\mathcal{G}$. To achieve this goal, we first reduce the problem for evaluating the above quantities to computing the $\ell_2$ norm of different vectors. Then, we estimate the $\ell_2$ norm by applying linear system solvers [9, 26] in order to significantly reduce the computational complexity.

According to Proposition 1, we can explicitly represent the concerned quantities in $\ell_2$ norm of vectors as stated in Lemma 4.1.

LEMMA 4.1. *For a graph $\mathcal{G} = (V, E, w)$, the internal conflict $C_{\mathrm{I}}(\mathcal{G})$, disagreement $D(\mathcal{G})$, polarization $P(\mathcal{G})$, controversy $C(\mathcal{G})$, and disagreement-controversy index $I_{\mathrm{dc}}(\mathcal{G})$ can be expressed, respectively, in terms of $\ell_2$ norm as:*

$$C_{\mathrm{I}}(\mathcal{G}) = z^\top L^2 z = \|L(I + L)^{-1} s\|^2, \tag{13}$$

$$D(\mathcal{G}) = z^\top L z = \|W^{1/2} B (I + L)^{-1} s\|^2, \tag{14}$$

$$P(\mathcal{G}) = \bar{z}^\top \bar{z} = \|(I + L)^{-1} \bar{s}\|^2, \tag{15}$$

$$C(\mathcal{G}) = z^\top z = \|(I + L)^{-1} s\|^2, \tag{16}$$

$$I_{\mathrm{dc}}(\mathcal{G}) = \|W^{1/2} B (I + L)^{-1} s\|^2 + \|(I + L)^{-1} s\|^2, \tag{17}$$

*where $W^{1/2}$ is a diagonal matrix defined as $W^{1/2} = \mathrm{diag}(\sqrt{w_1}, \sqrt{w_2}, \sqrt{w_3}, \ldots, \sqrt{w_m})$.*

Having reduced the computation of the relevant quantities to evaluating $\ell_2$ norms of some vectors in $\mathbb{R}^n$ or $\mathbb{R}^m$, we continue to compute the $\ell_2$ norms. However, directly calculating the $\ell_2$ norms does not help to reduce the computational cost, since it still requires inverting matrix $I + L$. In order to reduce computational time, we resort to the efficient linear system solvers [9], which avoids the inverse operation by solving a system of equations [26].

LEMMA 4.2. *There is a nearly linear time solver $y = \text{SOLVE}(T, x, \delta)$, which takes an $n \times n$ positive semi-definite matrix $T$ with $m$ non-zero entries, a column vector $x$, and an accuracy parameter $\delta$, and returns a column vector $y$ satisfying $\|y - T^\dagger x\|_T \le \delta \|T^\dagger x\|_T$, where $\|v\|_T = \sqrt{v^\top T v}$ and $T^\dagger$ is the pseudo-inverse of $T$. The expected time for performing this solver is $O\left(m \log^3 n \log\left(\frac{1}{\delta}\right)\right)$.*

Lemma 4.2 can significantly reduce the computational time for evaluating those quantities in the form of $(I + L)x$ with an ideal approximation guarantee. For example, as will be shown below, $p = \textsc{Solve}(I + L, s, \delta)$ is a good approximation of the expressed opinion vector $z$.

We next use Lemma 4.2 to obtain approximations for the quantities concerned. Prior to this, we introduce some notations and their properties. Let $a \geq 0$ and $b \geq 0$ be two nonnegative scalars. We say $a$ is an $\epsilon$-approximation ($0 \leq \epsilon \leq 1/2$) of $b$ if $(1 - \epsilon)a \leq b \leq (1 + \epsilon)a$, denoted by $a \approx_\epsilon b$. The $\epsilon$-approximation has the following basic property: For nonnegative scalars $a$, $b$, $c$, and $d$, if $a \approx_\epsilon b$ and $c \approx_\epsilon d$, then $a + c \approx_\epsilon b + d$. For two matrices $X$ and $Y$, we write $X \preceq Y$ if $Y - X$ is positive semidefinite, that is, $x^\top X x \leq x^\top Y x$ holds for every real vector $x$. Then, we have $I \preceq I + L$, $L \preceq I + L$, $I + L \preceq w_{\max}(n + 1)I$, and $\frac{1}{n\, w_{\max}} L \preceq I$. These relations are useful to prove the following lemmas, which we will apply to obtain $\epsilon$-approximation for the quantities we care about.

LEMMA 4.3. *Given an undirected weighted graph $G = (V, E, w)$ with each edge weight in the interval $[w_{\min}, w_{\max}]$, the Laplacian matrix $L$, and a parameter $\epsilon \in (0, \frac{1}{2})$, let $x$ be an arbitrary vector, and let $y = \textsc{Solve}(I + L, x, \delta)$, where*

$$\delta \leq \frac{\epsilon}{3\sqrt{w_{\max}(n + 1)}}.$$

*Then, the following relation holds:*

$$(1 - \epsilon)\|(I + L)^{-1}x\|^2 \leq \|y\|^2 \leq (1 + \epsilon)\|(I + L)^{-1}x\|^2.$$

**Proof.** According to Lemma 4.2, we have

$$\|y - (I + L)^{-1}x\|^2_{I+L} \leq \delta^2\|(I + L)^{-1}x\|^2_{I+L}.$$

The term on the left-hand side (lhs) can be bounded as

$$\|y - (I + L)^{-1}x\|^2_{I+L} \geq \|y - (I + L)^{-1}x\|^2$$
$$\geq \left| \|y\| - \|(I + L)^{-1}x\| \right|^2,$$

while the term on the right-hand side (rhs) can be bounded as

$$\|(I + L)^{-1}x\|^2_{I+L} \leq w_{\max}(n + 1)\|(I + L)^{-1}x\|^2.$$

Combining the above-obtained relations, we have

$$\left| \|y\| - \|(I + L)^{-1}x\| \right|^2$$
$$\leq \delta^2 w_{\max}(n + 1)\|(I + L)^{-1}x\|^2,$$

which implies

$$\frac{\left| \|y\| - \|(I + L)^{-1}x\| \right|}{\|(I + L)^{-1}x\|} \leq \sqrt{\delta^2 w_{\max}(n + 1)} \leq \frac{\epsilon}{3}$$

and

$$(1 - \frac{\epsilon}{3})^2\|(I + L)^{-1}x\|^2 \leq \|y\|^2$$
$$\leq (1 + \frac{\epsilon}{3})^2\|(I + L)^{-1}x\|^2.$$

Considering $0 < \epsilon < \frac{1}{2}$, we get

$$(1 - \epsilon)\|(I + L)^{-1}x\|^2 \leq \|y\|^2$$
$$\leq (1 + \epsilon)\|(I + L)^{-1}x\|^2,$$

which completes the proof. □

LEMMA 4.4. *Given an undirected weighted graph $G = (V, E, w)$ with each edge weight in the interval $[w_{\min}, w_{\max}]$, the incident matrix $B$, diagonal edge weight matrix $W$, Laplacian matrix $L$, and a parameter $\epsilon \in (0, \frac{1}{2})$, let $s = (s_1, s_2, \ldots, s_n)^\top$ be the internal opinion vector with each $s_i \in [0, 1]$ for $i = 1, 2, \ldots, n$, $\bar{s} = s - \frac{1^\top s}{n} 1 = (\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_n)^\top$ be the mean-centered internal opinion vector, and let $q = \textsc{Solve}(I + L, \bar{s}, \delta)$, where*

$$\delta \leq \frac{\epsilon\|\bar{s}\|}{3w_{\max}n(n + 1)}\sqrt{\frac{w_{\min}}{w_{\max}n(n + 1)}}.$$

*Then, the following relation holds:*

$$(1 - \epsilon)\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2$$
$$\leq \|W^{1/2}Bq\|^2 \leq (1 + \epsilon)\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2.$$

**Proof.** By Lemma 4.2, we have

$$\|q - (I + L)^{-1}\bar{s}\|^2_{I+L} \leq \delta^2\|(I + L)^{-1}\bar{s}\|^2_{I+L}.$$

The lhs can be bounded as

$$\|q - (I + L)^{-1}\bar{s}\|^2_{I+L}$$
$$\geq \|q - (I + L)^{-1}\bar{s}\|^2_L$$
$$= \|W^{1/2}Bq - W^{1/2}B(I + L)^{-1}\bar{s}\|^2$$
$$\geq \left| \|W^{1/2}Bq\| - \|W^{1/2}B(I + L)^{-1}\bar{s}\| \right|^2,$$

while rhs is bounded as

$$\|(I + L)^{-1}\bar{s}\|^2_{I+L} \leq w_{\max}(n + 1)\|(I + L)^{-1}\bar{s}\|^2$$
$$\leq n(n + 1)w_{\max},$$

where $\bar{s}_i \leq 1$, $i = 1, 2, \ldots, n$, is used. Combining the above-obtained results gives

$$\left| \|W^{1/2}Bq\| - \|W^{1/2}B(I + L)^{-1}\bar{s}\| \right|^2$$
$$\leq \delta^2 n(n + 1)w_{\max}.$$

On the other hand, since $\bar{s}^\top 1 = 0$ due to $\bar{s} = s - \frac{1^\top s}{n} 1$,

$$\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2 \geq \frac{w_{\min}}{n^2(n + 1)^2 w_{\max}^2}\|\bar{s}\|^2.$$

Thus, one has

$$\frac{\left| \|W^{1/2}Bq\| - \|W^{1/2}B(I + L)^{-1}\bar{s}\| \right|}{\|W^{1/2}B(I + L)^{-1}\bar{s}\|}$$
$$\leq \sqrt{\frac{\delta^2 n^3(n + 1)^3 w_{\max}^3}{w_{\min}\|\bar{s}\|^2}} \leq \frac{\epsilon}{3}.$$

In other words,

$$(1 - \frac{\epsilon}{3})^2\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2$$
$$\leq \|W^{1/2}Bq\|^2 \leq (1 + \frac{\epsilon}{3})^2\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2.$$

Because $0 < \epsilon < \frac{1}{2}$, we obtain

$$(1 - \epsilon)\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2$$
$$\leq \|W^{1/2}Bq\|^2 \leq (1 + \epsilon)\|W^{1/2}B(I + L)^{-1}\bar{s}\|^2,$$

which completes the proof. □

LEMMA 4.5. *Given an undirected weighted graph $\mathcal{G} = (V, E, w)$ with each edge weight in the interval $[w_{\min}, w_{\max}]$, Laplacian matrix $L$, and a parameter $\epsilon \in (0, \frac{1}{2})$, let $s = (s_1, s_2, \ldots, s_n)^\top$ be the internal opinion vector with each $s_i \in [0, 1]$ for $i = 1, 2, \ldots, n$, $\bar{s} = s - \frac{1^\top s}{n} 1 = (\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_n)^\top$ be the mean-centered internal opinion vector, $q = $ SOLVE$(I + L, \bar{s}, \delta)$, where*

$$\delta \leq \frac{\epsilon w_{\min} \|\bar{s}\|}{3 w_{\max}^2 n^3 (n + 1) \sqrt{n + 1}}.$$

*Then, the following relation holds:*

$$(1 - \epsilon)\|L(I + L)^{-1}\bar{s}\|^2 \leq \|Lq\|^2$$
$$\leq (1 + \epsilon)\|L(I + L)^{-1}\bar{s}\|^2.$$

**Proof.** Making use of Lemma 4.2, we obtain

$$\|q - (I + L)^{-1}\bar{s}\|_{I+L}^2 \leq \delta^2 \|(I + L)^{-1}\bar{s}\|_{I+L}^2.$$

The term on the lhs can be bounded as

$$\|q - (I + L)^{-1}\bar{s}\|_{I+L}^2$$
$$\geq \|q - (I + L)^{-1}\bar{s}\|^2 \geq \frac{1}{n w_{\max}} \|Lq - L(I + L)^{-1}\bar{s}\|^2$$
$$\geq \frac{1}{n w_{\max}} \left| \|Lq\| - \|L(I + L)^{-1}\bar{s}\| \right|^2.$$

According to the proof of Lemma 4.4,

$$\|(I + L)^{-1}\bar{s}\|_{I+L}^2 \leq n(n + 1)w_{\max}.$$

Thus, we have

$$\left| \|Lq\| - \|L(I + L)^{-1}\bar{s}\| \right|^2 \leq \delta^2 n^2 (n + 1) w_{\max}^2.$$

On the other hand,

$$\|L(I + L)^{-1}\bar{s}\|^2 \geq \frac{w_{\min}^2}{n^4(n + 1)^2 w_{\max}^2} \|\bar{s}\|^2.$$

Combining the above relations leads to

$$\frac{\left| \|Lq\| - \|L(I + L)^{-1}\bar{s}\| \right|}{\|L(I + L)^{-1}\bar{s}\|} \leq \sqrt{\frac{\delta^2 n^6 (n + 1)^3 w_{\max}^4}{w_{\min}^2 \|\bar{s}\|^2}} \leq \frac{\epsilon}{3},$$

which can be recast as

$$(1 - \frac{\epsilon}{3})^2 \|L(I + L)^{-1}\bar{s}\|^2 \leq \|Lq\|^2$$
$$\leq (1 + \frac{\epsilon}{3})^2 \|L(I + L)^{-1}\bar{s}\|^2.$$

Using the condition $0 < \epsilon < \frac{1}{2}$, one obtains

$$(1 - \epsilon)\|L(I + L)^{-1}\bar{s}\|^2 \leq \|L\tilde{z}\|^2$$
$$\leq (1 + \epsilon)\|L(I + L)^{-1}\bar{s}\|^2,$$

which completes the proof. □

Based on the Lemmas 4.2, 4.3, 4.4, and 4.5, we propose a fast and efficient algorithm APPROXIM to approximate the internal conflict $C_I(\mathcal{G})$, disagreement $D(\mathcal{G})$, polarization $P(\mathcal{G})$, controversy $C(\mathcal{G})$, and disagreement-controversy index $I_{dc}(\mathcal{G})$ for any undirected weighted graph $\mathcal{G}$. In Algorithm 1, we present the pseudocode of APPROXIM, where $\delta$ is less than or equal to the parameters $\delta$'s in Lemmas 4.3, 4.4, and 4.5.

Our approximation algorithm APPROXIM is both accurate and efficient, as summarized in Theorem 4.1.

---

**Algorithm 1:** APPROXIM $(\mathcal{G}, s, \epsilon)$

**Input** : $\mathcal{G}$: a graph with edge weight in $[w_{\min}, w_{\max}]$
$s$: initial opinion vector
$\epsilon$: the error parameter in $(0, \frac{1}{2})$

**Output** : $\{\tilde{C}_I(\mathcal{G}), \tilde{D}(\mathcal{G}), \tilde{P}(\mathcal{G}), \tilde{C}(\mathcal{G}), \tilde{I}_{dc}(\mathcal{G})\}$

1 $\delta = \frac{\epsilon w_{\min} \|\bar{s}\|}{3 w_{\max}^2 n^3 (n+1) \sqrt{n+1}}$

2 $\bar{s} = s - \frac{1^\top s}{n} 1$

3 $\tilde{z} = $ SOLVE$(I + L, s, \delta)$

4 $q = $ SOLVE$(I + L, \bar{s}, \delta)$

5 $\tilde{C}_I(\mathcal{G}) = \|L\tilde{z}\|^2$

6 $\tilde{D}(\mathcal{G}) = \|W^{1/2}Bq\|^2$

7 $\tilde{P}(\mathcal{G}) = \|q\|^2$

8 $\tilde{C}(\mathcal{G}) = \|\tilde{z}\|^2$

9 $\tilde{I}_{dc}(\mathcal{G}) = \tilde{D}(\mathcal{G}) + \tilde{C}(\mathcal{G})$

10 **return** $\{\tilde{C}_I(\mathcal{G}), \tilde{D}(\mathcal{G}), \tilde{P}(\mathcal{G}), \tilde{C}(\mathcal{G}), \tilde{I}_{dc}(\mathcal{G})\}$

---

THEOREM 4.1. *Given an undirected weighted graph $\mathcal{G}$ with $n$ nodes and $m$ edges, an error parameter $\epsilon \in (0, \frac{1}{2})$, and the internal opinion vector $s$, the algorithm APPROXIM$(\mathcal{G}, s, \epsilon)$ runs in expected time $O\left(m \log^4 n \log\left(\frac{r}{\epsilon}\right)\right)$ where $r = \frac{w_{\max}}{w_{\min}}$, and returns the $\epsilon$-approximation $\tilde{C}_I(\mathcal{G}), \tilde{D}(\mathcal{G}), \tilde{P}(\mathcal{G}), \tilde{C}(\mathcal{G}), \tilde{I}_{dc}(\mathcal{G})$ for the internal conflict $C_I(\mathcal{G})$, disagreement $D(\mathcal{G})$, polarization $P(\mathcal{G})$, controversy $C(\mathcal{G})$, and disagreement-controversy index $I_{dc}(\mathcal{G})$, satisfying $\tilde{C}_I(\mathcal{G}) \approx_\epsilon C_I(\mathcal{G}), \tilde{D}(\mathcal{G}) \approx_\epsilon D(\mathcal{G}), \tilde{P}(\mathcal{G}) \approx_\epsilon P(\mathcal{G}), \tilde{C}(\mathcal{G}) \approx_\epsilon C(\mathcal{G}), and \tilde{I}_{dc}(\mathcal{G}) \approx_\epsilon I_{dc}(\mathcal{G})$.*

## 5 EXPERIMENTS

In this section, we assess the efficiency and accuracy of our approximation algorithm APPROXIM. To this end, we implement this algorithm on various real networks and compare the running time and accuracy of APPROXIM with those corresponding to the exact algorithm, called EXACT. For the EXACT, it computes relevant quantities by directly inverting matrix $I + L$, performing product of related matrices, and then calculating corresponding $\ell_2$ norms.

**Environment.** Our extensive experiments were run on a Linux box with 4-core 4.2GHz Intel i7-7700K CPU and 32GB of main memory. Our code for both the approximation algorithm APPROXIM and exact algorithm EXACT was written in *Julia v1.5.1*. The solver SOLVE we use is based on the technique in [26], the *Julia language* implementation for which is open and accessible on the website [1].

**Datasets.** All the real-world networks we consider are publicly available in the Koblenz Network Collection [25] and Network Repository [37]. The first three columns of Table 1 are related information of networks, including the network name, the number of nodes, and the number of edges. For a network with $n$ nodes and $m$ edges, we use $n'$ and $m'$ to denote, respectively, the number of nodes and edges in its largest connected component. All our experiments were conducted on the largest connected components. The smallest network consists of 4991 nodes, while the largest network has more than one million. In Table 1, the networks listed

---

[1]http://danspielman.github.io/Laplacians.jl/latest/

**Table 1: Statistics of real-world networks used in our experiments and comparison of running time (seconds, $s$) between Exact and Approxim for three internal distributions (uniform distribution, exponential distribution, and power-law distribution) with error parameter $\epsilon = 10^{-6}$.**

| Network | $n'$ | $m'$ | Running time ($s$) for algorithms Exact and Approxim | | | | | |
| | | | Uniform distribution | | Exponential distribution | | Power-law distribution | |
| | | | Exact | Approxim | Exact | Approxim | Exact | Approxim |
|---|---|---|---|---|---|---|---|---|
| Erdös992 | 4,991 | 7,428 | 2.24 | 2.66 | 2.29 | 2.65 | 2.51 | 2.72 |
| Advogato | 5,054 | 43,015 | 2.75 | 2.80 | 2.43 | 2.71 | 2.51 | 2.77 |
| PagesGovernment | 7,057 | 89,429 | 7.52 | 2.70 | 9.72 | 2.75 | 7.82 | 2.59 |
| WikiElec | 7,066 | 100,727 | 7.59 | 2.69 | 6.68 | 2.69 | 7.34 | 2.63 |
| HepPh | 11,204 | 117,619 | 28.14 | 2.74 | 27.67 | 2.61 | 27.58 | 2.62 |
| Anybeat | 12,645 | 49,132 | 40.59 | 2.65 | 40.87 | 2.74 | 40.21 | 2.73 |
| PagesCompany | 14,113 | 52,126 | 56.02 | 2.78 | 55.61 | 2.96 | 55.67 | 2.75 |
| AstroPh | 17,903 | 196,972 | 115.50 | 2.91 | 117.64 | 2.87 | 117.59 | 2.82 |
| CondMat | 21,363 | 91,286 | 204.15 | 2.83 | 215.57 | 3.04 | 200.75 | 2.59 |
| Gplus | 23,613 | 39,182 | 279.01 | 2.82 | 281.00 | 2.96 | 274.52 | 2.75 |
| GemsecRO | 41,773 | 125,826 | 1585.75 | 3.52 | 1621.83 | 3.17 | 1587.00 | 3.27 |
| GemsecHU | 47,538 | 222,887 | 2410.30 | 6.52 | 2395.84 | 7.24 | 2430.26 | 6.95 |
| PagesArtist | 50,515 | 819,090 | 2895.73 | 10.06 | 2976.45 | 7.50 | 3017.09 | 8.07 |
| Brightkite | 56,739 | 212,945 | 4203.67 | 8.66 | 4312.25 | 6.74 | 4072.74 | 8.53 |
| Themarker | 69,317 | 1,644,794 | — | 5.02 | — | 5.04 | — | 5.09 |
| Slashdot | 70,068 | 358,647 | — | 3.31 | — | 3.36 | — | 3.34 |
| BlogCatalog | 88,784 | 2,093,195 | — | 5.54 | — | 5.50 | — | 5.53 |
| WikiTalk | 92,117 | 360,767 | — | 3.21 | — | 3.28 | — | 3.29 |
| Buzznet | 101,163 | 2,763,066 | — | 6.39 | — | 6.33 | — | 6.31 |
| LiveMocha | 104,103 | 2,193,083 | — | 6.46 | — | 6.44 | — | 6.46 |
| Douban | 154,908 | 327,162 | — | 3.30 | — | 3.35 | — | 3.37 |
| Gowalla | 196,591 | 950,327 | — | 4.43 | — | 4.30 | — | 4.29 |
| Academia | 200,167 | 1,022,440 | — | 4.51 | — | 4.44 | — | 4.49 |
| GooglePlus | 201,949 | 1,133,956 | — | 4.33 | — | 4.33 | — | 4.30 |
| Citeseer | 227,320 | 814,134 | — | 4.04 | — | 3.87 | — | 4.02 |
| MathSciNet | 332,689 | 820,644 | — | 4.32 | — | 4.37 | — | 4.27 |
| TwitterFollows | 404,719 | 713,319 | — | 3.91 | — | 3.94 | — | 3.86 |
| Flickr | 513,969 | 3,190,452 | — | 8.21 | — | 8.22 | — | 8.07 |
| Delicious | 536,108 | 1,365,961 | — | 5.48 | — | 5.39 | — | 5.30 |
| IMDB | 896,305 | 3,782,447 | — | 11.10 | — | 11.22 | — | 10.75 |
| YoutubeSnap | 1,134,890 | 2,987,624 | — | 9.00 | — | 8.89 | — | 8.98 |
| Lastfm | 1,191,805 | 4,519,330 | — | 11.79 | — | 11.79 | — | 12.12 |
| Pokec | 1,632,803 | 22,301,964 | — | 75.91 | — | 76.50 | — | 76.14 |
| Flixster | 2,523,386 | 7,918,801 | — | 19.46 | — | 19.67 | — | 19.56 |
| LiveJournal | 4,033,137 | 27,933,062 | — | 80.60 | — | 80.15 | — | 79.40 |

**Table 2: Relative error for estimated $\tilde{C}_I(\mathcal{G})$, $\tilde{D}(\mathcal{G})$, $\tilde{P}(\mathcal{G})$, $\tilde{C}(\mathcal{G})$ for three internal distributions with input parameter $\epsilon = 10^{-6}$.**

| Network $\mathcal{G}$ | Relative error of four estimated quantities for three internal opinion distributions ($\times 10^{-7}$) | | | | | | | | | | | |
| | Uniform distribution | | | | Exponential distribution | | | | Power-law distribution | | | |
| | $\tilde{C}_I(\mathcal{G})$ | $\tilde{D}(\mathcal{G})$ | $\tilde{P}(\mathcal{G})$ | $\tilde{I}_{dc}(\mathcal{G})$ | $\tilde{C}_I(\mathcal{G})$ | $\tilde{D}(\mathcal{G})$ | $\tilde{P}(\mathcal{G})$ | $\tilde{I}_{dc}(\mathcal{G})$ | $\tilde{C}_I(\mathcal{G})$ | $\tilde{D}(\mathcal{G})$ | $P(\mathcal{G})$ | $\tilde{I}_{dc}(\mathcal{G})$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Erdös992 | 1.1002 | 0.0018 | 0.0018 | 0.1379 | 6.5276 | 0.0077 | 0.0078 | 0.0903 | 1.3976 | 0.0010 | 0.0009 | 0.0372 |
| Advogato | 2.6516 | 0.0058 | 0.0077 | 0.0900 | 2.0039 | 0.0138 | 0.0186 | 0.0827 | 0.0975 | 0.0022 | 0.0025 | 0.0015 |
| PagesGovernment | 0.6307 | 0.0214 | 0.0544 | 0.0432 | 5.1476 | 0.0068 | 0.0165 | 0.0643 | 0.0926 | 0.0020 | 0.0076 | 0.2168 |
| WikiElec | 1.0029 | 0.0021 | 0.0030 | 0.0826 | 4.9890 | 0.0002 | 0.0003 | 0.0582 | 0.1049 | 0.0074 | 0.0220 | 0.0516 |
| HepPh | 0.5886 | 0.0009 | 0.0015 | 0.2008 | 0.6633 | 0.0039 | 0.0063 | 0.0605 | 2.5963 | 0.0051 | 0.0098 | 0.0380 |
| Anybeat | 0.1904 | 0.0152 | 0.0185 | 0.0366 | 0.6107 | 0.0064 | 0.0077 | 0.1517 | 0.0256 | 0.0044 | 0.0067 | 0.0532 |
| PagesCompany | 0.6721 | 0.0090 | 0.0129 | 0.0038 | 0.7049 | 0.0046 | 0.0069 | 0.0061 | 0.8048 | 0.0099 | 0.0136 | 0.2874 |
| AstroPh | 0.8058 | 0.0051 | 0.0109 | 0.1292 | 2.9936 | 0.0020 | 0.0041 | 0.0873 | 0.0143 | 0.0012 | 0.0120 | 0.0122 |
| CondMat | 2.0614 | 0.0051 | 0.0088 | 0.1528 | 2.1314 | 0.0027 | 0.0049 | 0.0670 | 0.2099 | 0.0056 | 0.0155 | 0.0666 |
| Gplus | 1.6386 | 0.0012 | 0.0014 | 0.2219 | 0.4878 | 0.0001 | 0.0001 | 0.0198 | 0.1471 | 0.0005 | 0.0006 | 0.0756 |
| GemsecRO | 4.5966 | 0.0230 | 0.0358 | 0.3691 | 2.1684 | 0.0069 | 0.0109 | 0.1042 | 2.0774 | 0.0257 | 0.0507 | 0.4375 |
| GemsecHU | 0.4021 | 0.0342 | 0.0826 | 0.2018 | 4.6511 | 0.0153 | 0.0369 | 0.1320 | 0.8842 | 0.0003 | 0.0004 | 0.0831 |
| PagesArtist | 1.9994 | 0.0093 | 0.0228 | 0.4941 | 1.3243 | 0.0077 | 0.0194 | 0.4454 | 0.0419 | 0.0026 | 0.0125 | 0.0766 |
| Brightkite | 1.2047 | 0.0022 | 0.0026 | 0.1436 | 0.5782 | 0.0190 | 0.0221 | 0.1564 | 0.5104 | 0.0421 | 0.0324 | 0.0027 |

in an increasing order of the number of nodes in their largest connected components.

**Internal opinion distributions.** In our experiments, the internal opinions are generated according to three different distributions: uniform distribution, exponential distribution, and power-law distribution, with the latter two distributions generated according by the randht.py file in [8]. For the uniform distribution, the opinion $s_i$ of node $i$ is distributed uniformly in the range of $[0, 1]$. For the exponential distribution, we choose the probability density $e^{x_{\min}} e^{-x}$ to generate $n'$ positive numbers $x$ with minimum value $x_{\min} > 0$. Then, dividing each $x$ by the maximum observed value, we normalize these $n'$ numbers to the range $[0, 1]$ as the internal opinions of nodes. Note that there is always a node with internal opinion 1 due to the normalization operation. Similarly, for the power-law distribution, we use the probability density $(\alpha - 1)x_{\min}^{\alpha-1} x^{-\alpha}$ with $\alpha = 2.5$ to generate $n'$ positive numbers, and normalize them to interval $[0, 1]$ as the internal opinions.

**Efficiency.** Table 1 reports the running time of the two algorithms Approxim and Exact on different networks we consider in the experiments. We note that we cannot compute those related quantities using the algorithm Exact for the last 7 networks, due to the high memory and time cost, but we can run the algorithm Approxim. In all our experiments, we set the error parameter $\epsilon$ equal to $10^{-6}$. For each of the three internal opinion distributions in different networks, we record the running times of Approxim and Exact. Table 1 shows that for all considered networks, the running time for Approxim is less than that for Exact. For moderately large networks with more than ten thousand nodes, Exact is much slower than Approxim. For example, for the GemsecRO network, the running time for Approxim is less than one hundredth of that for Exact. Finally, for large graphs particularly those with over 150 thousand nodes, our approximation algorithm Approxim shows a very obvious efficiency advantage, since Exact fails to run for these networks.

**Accuracy.** Except for the high-efficiency, our approximation algorithm Approxim has also high-accuracy in practice. To demonstrate this, we assess the accuracy of algorithm Approxim in Table 2. For each of the three distributions of internal opinions, we compare the approximate results of Approxim with the exact result of Exact for all networks examined but the last seven ones in Table 2. For each quantity $\rho$ concerned, we use the mean relative error $\sigma = |\rho - \tilde{\rho}|/\rho$ of $\tilde{\rho}$ obtained by Approxim as an estimation of $\rho$. In Table 2, we present the mean relative errors of the four estimated quantities, internal conflict $\tilde{C}_{\mathrm{I}}(\mathcal{G})$, disagreement $\tilde{D}(\mathcal{G})$, polarization $\tilde{P}(\mathcal{G})$, and controversy $\tilde{C}(\mathcal{G})$, for different real networks with input error parameter $\epsilon = 10^{-6}$. The results show that for all quantities concerned of all networks examined, the actual relative errors are ignorable, with all errors less than $10^{-7}$, much smaller than the proved theoretical guarantee.

## 6 CONCLUSION

Conflict, disagreement, controversy, and polarization in social networks have received considerable recent attention, and some indices have been proposed to quantify these concepts. However, direct computation of these quantities is computationally challenging for large networks, because it involves inverting matrices. In this paper, we addressed the problem of efficiently computing these quantities in an undirected unweighted graph. To this end, we developed an approximation algorithm that is based on seminal work about Laplacian solver for solving linear system of equations, which can avoid the operation of matrix inversion. The algorithm has an almost linear computation complexity with respect to the number of edges in the graph, and simultaneously possesses a theoretical guarantee for accuracy. We performed extensive experiments on a diverse set of real network datasets to demonstrate that our presented approximation algorithm works both efficiently and effectively, especially for large-scale networks with millions of nodes.

## REFERENCES

[1] Brian DO Anderson and Mengbin Ye. 2019. Recent advances in the modelling and analysis of opinion dynamics on influence networks. *International Journal of Automation and Computing* 16, 2 (2019), 129–149.

[2] Roger L Berger. 1981. A necessary and sufficient condition for reaching a consensus using DeGroot's method. *J. Amer. Statist. Assoc.* 76, 374 (1981), 415–418.

[3] David Bindel, Jon Kleinberg, and Sigal Oren. 2015. How bad is forming your own opinion? *Games and Economic Behavior* 92 (2015), 248–265.

[4] Pavel Chebotarev. 2008. Spanning forests and the golden ratio. *Discrete Applied Mathematics* 156, 5 (2008), 813–821.

[5] P. Yu Chebotarev and E. V. Shamis. 1997. The matrix-forest theorem and measuring relations in small social groups. *Automation and Remote Control* 58, 9 (1997), 1505–1514.

[6] P. Yu Chebotarev and E. V. Shamis. 1998. On Proximity Measures for Graph Vertices. *Automation and Remote Control* 59, 10 (1998), 1443–1459.

[7] Xi Chen, Jefrey Lijffijt, and Tijl De Bie. 2018. Quantifying and minimizing risk of conflict in social networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1197–1205.

[8] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. 2009. Power-law distributions in empirical data. *SIAM Rev.* 51, 4 (2009), 661–703.

[9] Michael B Cohen, Rasmus Kyng, Gary L Miller, Jakub W Pachocki, Richard Peng, Anup B Rao, and Shen Chen Xu. 2014. Solving SDD linear systems in nearly $m \log^{1/2} n$ time. In *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*. ACM, 343–352.

[10] Pranav Dandekar, Ashish Goel, and David T Lee. 2013. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences* 110, 15 (2013), 5791–5796.

[11] Abhimanyu Das, Sreenivas Gollapudi, Rina Panigrahy, and Mahyar Salek. 2013. Debiasing social wisdom. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 500–508.

[12] Morris H DeGroot. 1974. Reaching a consensus. *J. Amer. Statist. Assoc.* 69, 345 (1974), 118–121.

[13] Yucheng Dong, Min Zhan, Gang Kou, Zhaogang Ding, and Haiming Liang. 2018. A survey on the fusion process in opinion dynamics. *Information Fusion* 43 (2018), 57–65.

[14] Haibo Du, Guanghui Wen, Guanrong Chen, Jinde Cao, and Fuad E Alsaadi. 2017. A distributed finite-time consensus algorithm for higher-order leaderless and leader-following multiagent systems. *IEEE Transactions on Systems, Man and Cybernetics: Systems* 47, 7 (2017), 1625–1634.

[15] Sheng-Li Du, Weiguo Xia, Wei Ren, Xi-Ming Sun, and Wei Wang. 2018. Observer-based consensus for multiagent systems under stochastic sampling mechanism. *IEEE Trans. Syst., Man, Cybern. Syst.* 48, 12 (2018), 2328–2338.

[16] Francois Fouss, Alain Pirotte, J-M Renders, and Marco Saerens. 2007. Random-walk computation of similarities between nodes of a graph with application to collaborative recommendation. *IEEE Transactions on Knowledge and Data Engineering* 19, 3 (2007), 355–369.

[17] Noah E Friedkin and Eugene C Johnsen. 1990. Social influence and opinions. *Journal of Mathematical Sociology* 15, 3-4 (1990), 193–206.

[18] Noah E Friedkin, Anton V Proskurnikov, Roberto Tempo, and Sergey E Parsegov. 2016. Network science on belief system dynamics under logic constraints. *Science* 354, 6310 (2016), 321–326.

[19] Javad Ghaderi and Rayadurgam Srikant. 2014. Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate. *Automatica* 50, 12 (2014), 3209–3215.

[20] Aristides Gionis, Evimaria Terzi, and Panayiotis Tsaparas. 2013. Opinion maximization in social networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 387–395.

[21] VE Golender, VV Drboglav, and AB Rosenblit. 1981. Graph potentials method and its application for chemical information processing. *Journal of Chemical Information and Computer Sciences* 21, 4 (1981), 196–204.

[22] Huifen Hong, Wenwu Yu, Guanghui Wen, and Xinghuo Yu. 2017. Distributed robust fixed-time consensus for nonlinear and disturbed multiagent systems. *IEEE Transactions on Systems, Man and Cybernetics: Systems* 47, 7 (2017), 1464–1473.

[23] Peng Jia, Anahita MirTabatabaei, Noah E Friedkin, and Francesco Bullo. 2015. Opinion dynamics and the evolution of social power in influence networks. *SIAM Rev.* 57, 3 (2015), 367–397.

[24] Yujia Jin, Qi Bao, and Zhongzhi Zhang. 2019. Forest distance closeness centrality in disconnected graphs. In *Proceedings of 2019 IEEE International Conference on Data Mining*. IEEE, 339–348.

[25] Jérôme Kunegis. 2013. KONECT: The Koblenz Network Collection. In *Proceedings of the 22nd International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee.

[26] Rasmus Kyng and Sushant Sachdeva. 2016. Approximate Gaussian elimination for Laplacians - fast, sparse, and simple. In *Proceedings of IEEE 57th Annual Symposium on Foundations of Computer Science*. IEEE, 573–582.

[27] Huan Li and Aaron Schild. 2018. Spectral Subspace Sparsification. In *Proceedings of 2018 IEEE 59th Annual Symposium on Foundations of Computer Science*. IEEE, 385–396.

[28] Xue Lin and Yuanshi Zheng. 2017. Finite-time consensus of switched multiagent systems. *IEEE Transactions on Systems, Man and Cybernetics: Systems* 47, 7 (2017), 1535–1545.

[29] Antonis Matakos, Evimaria Terzi, and Panayiotis Tsaparas. 2017. Measuring and Moderating Opinion Polarization in Social Networks. *Data Mining and Knowledge Discovery* 31, 5 (2017), 1480–1505.

[30] Russell Merris. 1997. Doubly stochastic graph matrices. *Publikacije Elektrotehničkog fakulteta. Serija Matematika* 8 (1997), 64–71.

[31] Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. 2018. Minimizing polarization and disagreement in social networks. In *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, 369–378.

[32] R. Olfati-Saber, J. A. Fax, and R. M Murray. 2007. Consensus and cooperation in networked multi-agent systems. *Proc. IEEE* 95, 1 (2007), 215–233.

[33] Sergey E Parsegov, Anton V Proskurnikov, Roberto Tempo, and Noah E Friedkin. 2016. Novel multidimensional models of opinion dynamics in social networks. *IEEE Trans. Automat. Control* 62, 5 (2016), 2270–2285.

[34] Yi Qi, Zhongzhi Zhang, Yuhao Yi, and Huan Li. 2019. Consensus in Self-Similar Hierarchical Graphs and Sierpiński Graphs: Convergence Speed, Delay Robustness, and Coherence. *IEEE Transactions on Cybernetics* 49, 2 (2019), 592–603.

[35] Chiara Ravazzi, Paolo Frasca, Roberto Tempo, and Hideaki Ishii. 2015. Ergodic Randomized Algorithms and Dynamics Over Networks. *IEEE Transactions on Control of Network Systems* 1, 2 (2015), 78–87.

[36] Hamed Rezaee and Farzaneh Abdollahi. 2019. Stationary consensus control of a class of high-order uncertain nonlinear agents with communication delays. *IEEE Transactions on Systems, Man and Cybernetics: Systems* 49, 6 (2019), 1285–1290.

[37] Ryan Rossi and Nesreen Ahmed. 2015. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. AAAI, 4292–4293.

[38] Justin Semonsen, Christopher Griffin, Anna Squicciarini, and Sarah Rajtmajer. 2019. Opinion dynamics in the presence of increasing agreement pressure. *IEEE Transactions on Cybernetics* 49, 4 (2019), 1270–1278.

[39] Mathieu Senelle, Silvia Garcia-Diez, Amin Mantrach, Masashi Shimbo, Marco Saerens, and François Fouss. 2014. The sum-over-forests density index: Identifying dense regions in a graph. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36, 6 (2014), 1268–1274.

[40] Daniel A. Spielman and Nikhil Srivastava. 2011. Graph Sparsification by Effective Resistances. *SIAM Journal of Computing* 40, 6 (2011), 1913–1926.

[41] Yuxin Wei and Guo-Ping Liu. 2017. Consensus tracking of heterogeneous discrete-time networked multiagent systems based on the networked predictive control scheme. *IEEE Transactions on Cybernetics* 47, 8 (2017), 2173–2184.

[42] Xiaotai Wu, Yang Tang, Jinde Cao, and Wenbing Zhang. 2016. Distributed consensus of stochastic delayed multi-agent systems under asynchronous switching. *IEEE Transactions on Cybernetics* 46, 8 (2016), 1817–1827.

[43] Yuhao Yi, Zhongzhi Zhang, and Stacy Patterson. 2020. Scale-free loopy structure is resistant to noise in consensus dynamics in power-law graphs. *IEEE Transactions on Cybernetics* 50, 1 (2020), 190–200.