Opinion Maximization in Social Networks: An Inverse Optimization Perspective

Yilu Liu[1,2], **Bo Xue**[1,2], **Yiming Yao**[1,2] and **Qingfu Zhang**[1,2,*]
[1]Department of Computer Science, City University of Hong Kong, Hong Kong SAR, China
[2]City University of Hong Kong Shenzhen Research Institute, Shenzhen, China
{yilu.liu, boxue4-c, yimingyao3-c}@my.cityu.edu.hk, qingfu.zhang@cityu.edu.hk

Abstract

Given a social network $\mathcal{G}$ with n nodes and m edges, the opinion maximization (OM) problem aims at identifying k ($k \ll n$) opinion leaders to maximize their opinion propagation in $\mathcal{G}$. Despite its significance and prevalence, existing OM methods often struggle to strike a satisfactory balance between theoretical performance and practical efficiency. This paper provides an inverse optimization perspective for OM by reformulating it as an opinion minimization (OMin) problem, whose objective function holds succinct expression and desired properties. Then, we introduce a bounded estimation scheme (BES) that can estimate the objective function of OMin with bounded error in $\widetilde{O}(kn+m)$ expected time, thereby making the optimization for OMin more efficient. To further enhance the optimization efficiency, a decomposition-based decremental estimation algorithm (DDEA) with a near $(1-1/e)$ approximation ratio in $\widetilde{O}(kmn)$ expected time is specifically designed for OMin. Extensive experiments on real-world social networks of varying scales demonstrate the effectiveness of BES and the superiority of DDEA.

1 Introduction

In the current digital age, social networks have become crucial media for us to express and exchange opinions. To maximize the dissemination of opinions, many opinion publishers tend to select a number of opinion leaders (OLs). For instance, governments may choose certain institutions to release news or policies [Howard *et al.*, 2000], while advertisers may select celebrities or bloggers to promote products [Casaló *et al.*, 2020]. Consequently, how to appropriately select OLs has emerged as an important research topic, commonly referred to as the opinion maximization (OM) problem [Gionis *et al.*, 2013].

The initial OM problem was proposed by [Gionis *et al.*, 2013]. They assumed that opinion dynamics on a social network $\mathcal{G}$ with n nodes and m edges follow the classical Friedkin-Johnsen (FJ) model [Friedkin and Johnsen, 1990], and studied how to select k ($k \ll n$) OLs holding stubborn opinions of "1" (the maximum opinion value) to maximize the overall opinion (the sum of opinions of all nodes) in $\mathcal{G}$. To address this problem, they demonstrated that the objective function of the OM problem satisfies monotonicity and modularity, and then employed the classical greedy algorithm (GA) [Nemhauser *et al.*, 1978] with an approximation ratio of $(1-1/e)$ along with several simple heuristic algorithms to solve it. Subsequently, numerous heuristic and metaheuristic algorithms have been further developed to tackle this problem, such as GA variants [Liu *et al.*, 2018; Abebe *et al.*, 2018], centrality-based methods [He *et al.*, 2023; Hunter and Zaman, 2022], and metaheuristic algorithms [Liu *et al.*, 2025b; Wan *et al.*, 2025]. Additionally, extended OM problems under specific scenarios (e.g., behavioral constraints [Nayak *et al.*, 2019; Hudson and Khamfroush, 2020], signed networks [Xu *et al.*, 2021; He *et al.*, 2021], and multi-party opinion publishers [Zhou and Zhang, 2023; He *et al.*, 2025; Saha *et al.*, 2025]) have also been recently studied.

This paper focuses on the original OM problem proposed by [Gionis *et al.*, 2013]. Despite the progress achieved, existing methods typically struggle to achieve a satisfactory balance between theoretical performance and practical efficiency. Specifically, while GA and its variants can often provide theoretical guarantees, the multiple rounds of objective incremental evaluation they require often limit their scalability. Centrality-based heuristic algorithms, as they generally do not involve the evaluation of specific objectives, exhibit favorable scalability, but their practical performance is highly unstable due to the lack of theoretical support. Metaheuristic algorithms, by adopting various search strategies, often achieve better practical performance, but their complex algorithmic structures and numerous objective evaluations similarly constrain their scalability. Considering these facts, we aim to design an approach that integrates both theoretical performance and practical efficiency.

Upon careful analysis of the OM formulation, we find that its optimization challenges mainly stem from the complicated representation and high computational complexity of its objective function. Therefore, we first reformulate the OM problem into an opinion minimization (OMin) problem, whose objective function can be more succinctly expressed while still retaining monotonicity and modularity. Since the objective function of OMin is still compu-

*Corresponding Author

tationally expensive, we then introduce a bounded estimation scheme (BES) that can estimate it with bounded error in nearly linear expected time, allowing iterative algorithms (e.g., GA, metaheuristic algorithms) to address OMin more efficiently. Next, to further enhance the efficiency for solving OMin, a decomposition-based decremental estimation algorithm (DDEA) is specifically designed. By structurally decomposing the objective increments within the GA framework, DDEA can achieve bounded estimation for each decomposed component, thereby significantly reducing computational complexity without substantially compromising the theoretical performance. We conduct extensive experiments on real-world social networks of varying scales. The results demonstrate the effectiveness of BES and the superiority of DDEA over several state-of-the-art OM approaches.

In summary, this paper makes the following contributions.

- We provide an inverse optimization perspective on the OM problem by reformulating it as the OMin problem. We show that the objective function of OMin could be expressed in a more succinct form while still preserving monotonicity and modularity.
- We propose BES to efficiently calculate the objective function of OMin, which could provide bounded error in $\widetilde{O}(kn+m)$ expected time, allowing iterative algorithms to address OMin in a more efficient manner.
- We specifically design DDEA to address OMin. We show that DDEA could achieve a near $(1-1/e)$ approximation ratio for OMin in $\widetilde{O}(kmn)$ expected time.
- Extensive experiments on real-world social networks of varying scales demonstrate the effectiveness of BES and the superiority of DDEA.

2 Related Work

Influence maximization in social networks. The proposal of the OM problem was mainly inspired by the influence maximization (IM) problem. Unlike the OM problem, the IM problem typically employs diffusion models (e.g., independent cascade (IC) model, linear threshold (LT) model) to describe the influence diffusion process of seed nodes (similar to OLs in the OM problem), with the objective of maximizing the number of nodes activated by the seed nodes. The IM problem was first introduced by [Kempe *et al.*, 2003]. After showing the modularity of the objective function under the IC and LT models, they also adopted GA as the optimization method. Given that the evaluation of the objective function involves expensive Monte Carlo simulations, several approximation methods have been proposed and widely used, such as expected diffusion value [Jiang *et al.*, 2011] and two-hop approximation [Lee and Chung, 2014]. Building upon these approximation methods, numerous heuristic and metaheuristic approaches have also been proposed to solve the IM problem, such as centrality-based measures [Guo *et al.*, 2025; Ullah *et al.*, 2024; Engsig *et al.*, 2024], memetic algorithms [Gong *et al.*, 2016], and swarm-intelligence algorithms [Wang *et al.*, 2019]. Furthermore, IM problems under various conditional constraints (e.g., competitiveness [Hong *et al.*, 2020], fairness [Rahmattalabi *et al.*, 2021], robustness [Wang *et al.*, 2023]) have been extensively studied in recent years.

Opinion optimization in social networks. In addition to the OM problem studied in this paper, numerous opinion optimization problems based on the FJ model have been proposed in recent years. [Matakos *et al.*, 2017] defined a polarization index to measure the extent to which opinions on a social network deviate from the mean, and similarly investigated how to select k OLs to minimize this polarization index. Subsequently, [Musco *et al.*, 2018] unified this polarization index with the disagreement index previously proposed by [Dandekar *et al.*, 2013] into a polarization-disagreement index, and studied how to modify the Laplacian matrix of the social network to optimize it. Given the significant practical implications of the polarization-disagreement index, many other optimization approaches have been further introduced to optimize it, such as recommending new edges [Zhu *et al.*, 2021; Wang and Kleinberg, 2023], adjusting edge weights [Cinus *et al.*, 2023], shifting the overall internal opinions [Chen and Rácz, 2021; Tu *et al.*, 2023], or increasing the overall opinion stubbornness [Shirzadi and Zehmakan, 2025]. Furthermore, some other opinion optimization problems have also garnered considerable attention, such as competitiveness maximization [Liu *et al.*, 2023] and leader-follower disagreement minimization [Liu *et al.*, 2026].

3 Preliminaries

Social networks. Following most OM studies, a social network is denoted by a connected, undirected, and weighted graph $\mathcal{G} = (V, E)$, where $V = \{v_1, \ldots, v_n\}$ is the node set and $E = \{e_1, \ldots, e_m\}$ is the edge set. The adjacency matrix is $\boldsymbol{A} = (a_{ij})_{n\times n}$, where $a_{ij} = a_{ji} = w \in (0, 1]$ if an edge of normalized weight w connects v_i and v_j, and $a_{ij} = a_{ji} = 0$ otherwise. The diagonal degree matrix is $\boldsymbol{D} = diag(d_1, \ldots, d_n)$, with $d_i = \sum_{j=1}^{n} a_{ij}$ representing the degree of node v_i. From these, the Laplacian matrix is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, and the forest matrix as $\boldsymbol{F} = (\boldsymbol{I} + \boldsymbol{L})^{-1} = (f_{ij})_{n\times n}$, where $\boldsymbol{I}$ is an identity matrix of appropriate size. The matrix $\boldsymbol{F}$ is doubly stochastic, satisfying $f_{ij} \geq 0$, $\boldsymbol{1}^\top \boldsymbol{F} = \boldsymbol{1}^\top$, and $\boldsymbol{F}\boldsymbol{1} = \boldsymbol{1}$.

Opinion maximization. In the FJ model [Friedkin and Johnsen, 1990], each node v_i holds an internal (innate) opinion $s_i \in [0, 1]$ and an external (expressed) opinion $z_i \in [0, 1]$, where s_i remains unchanged and z_i will be updated by:

$$z_i\,(t+1) = \frac{s_i + \sum_{j=1}^{n} a_{ij} z_j(t)}{1 + \sum_{j=1}^{n} a_{ij}}, \tag{1}$$

where $z_i(t)$ is the state of z_i at time step t. As t approaches infinity, $z_i(t)$ will reach its equilibrium state z_i. (1) shows that z_i is determined by s_i and the neighboring external opinions of v_i. Thus, if there are some OLs holding stubborn opinions, the equilibrium external opinions of all nodes will be influenced to varying degrees. Thereby, based on the FJ model, the OM problem is defined as below.

Problem 1 (OM [Gionis *et al.*, 2013]). *Given a social network $\mathcal{G} = (V, E)$ and a positive integer $k \ll |V|$, the aim* *of OM is to identify an OL set $L \subseteq V$ ($|L| = k$, $z_i \equiv 1$ for $v_i \in L$) to maximize the overall opinion in $\mathcal{G}$:*

$$\mathcal{F}(L) = \sum_{i=1}^{n} z_i. \tag{2}$$

4 Problem Reformulation

Optimization challenges of opinion maximization. From Problem 1, it can be found that the optimization challenges of OM mainly stem from the following two aspects:

- *The complex representation of the objective function.* To compute (2), we must first obtain the z_i value for each node. Let $\boldsymbol{z}_l = [z_1, \ldots, z_k]$ and $\boldsymbol{z}_c = [z_1, \ldots, z_{n-k}]$ be the equilibrium external opinion vectors of OLs and common nodes, respectively. Similarly, we can write $\boldsymbol{A}$ and $\boldsymbol{D}$ as the following block form:

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{l,l} & \boldsymbol{A}_{l,c} \\ \boldsymbol{A}_{c,l} & \boldsymbol{A}_{c,c} \end{pmatrix}, \boldsymbol{D} = \begin{pmatrix} \boldsymbol{D}_{l,l} & \boldsymbol{D}_{l,c} \\ \boldsymbol{D}_{c,l} & \boldsymbol{D}_{c,c} \end{pmatrix}. \tag{3}$$

  Since OLs hold the stubborn opinion 1, we have

$$\boldsymbol{z}_l = \boldsymbol{1}. \tag{4}$$

  Then, combining (1) and (4), $\boldsymbol{z}_c$ satisfies

$$\boldsymbol{z}_c = \frac{\boldsymbol{s}_c + \boldsymbol{A}_{c,l}\boldsymbol{z}_l + \boldsymbol{A}_{c,c}\boldsymbol{z}_c}{\boldsymbol{I} + \boldsymbol{D}_{c,c}}, \tag{5}$$

  which shows that

$$\boldsymbol{z}_c = (\boldsymbol{I} + \boldsymbol{L}_{c,c})^{-1}(\boldsymbol{A}_{c,l}\boldsymbol{z}_l + \boldsymbol{s}_c), \tag{6}$$

  where $\boldsymbol{L}_{c,c} = \boldsymbol{D}_{c,c} - \boldsymbol{A}_{c,c}$. Thus, (2) can be denoted as

$$\begin{aligned}\mathcal{F}(L) &= \sum_{i=1}^{n} z_i = \boldsymbol{1}^\top \boldsymbol{z}_l + \boldsymbol{1}^\top \boldsymbol{z}_c \\ &= \boldsymbol{1}^\top \boldsymbol{z}_l + \boldsymbol{1}^\top (\boldsymbol{I} + \boldsymbol{L}_{c,c})^{-1}(\boldsymbol{A}_{c,l}\boldsymbol{z}_l + \boldsymbol{s}_c),\end{aligned} \tag{7}$$

  which is very complicated, as it involves four different dynamic components: $\boldsymbol{L}_{c,c}$, $\boldsymbol{A}_{c,l}$, $\boldsymbol{z}_l$, and $\boldsymbol{s}_c$.

- *The high computational complexity of the objective function.* From (7), we can find that the calculation of $\mathcal{F}(L)$ involves inverting an $(n-k) \times (n-k)$ matrix, which generally requires $O(n^3)$ time complexity.

These two optimization challenges significantly limit the scalability of iterative algorithms (e.g., GA, metaheuristic algorithms). Therefore, we next address these challenges through problem reformulation.

Formulation of opinion minimization. As illustrated in Problem 1, the OM problem essentially models the stance of opinion publisher as the maximum value "1", and then selects k OLs to maximize the overall opinion (i.e., to maximize the propagation effect of opinion "1"). Therefore, we can naturally adopt an alternative yet equivalent modeling approach: model the stance of opinion publisher as the minimum value "0", and select k OLs to minimize the overall opinion (i.e., to maximize the propagation effect of opinion "0"). Under this modeling framework, $\boldsymbol{z}_l = \boldsymbol{0}$, which causes both $\boldsymbol{1}^\top \boldsymbol{z}_l$ and $\boldsymbol{A}_{c,l}\boldsymbol{z}_l$ in (7) to equal $\boldsymbol{0}$, thereby significantly simplifying the computation of (7). To further unify the two remaining dynamic components (i.e., $\boldsymbol{L}_{c,c}$ and $\boldsymbol{s}_c$), we will next proceed to simplify the representation of the FJ model after selecting OLs with stubborn opinion "0".

Let $\boldsymbol{z} = [z_1, \ldots, z_n]$ be the external opinion vector of all nodes. By introducing an auxiliary diagonal matrix $\boldsymbol{X}_k = diag(x_1, \ldots, x_n)$, where $x_i = 0$ if $v_i \in L$ and $x_i = 1$ otherwise, the FJ model after selecting OLs with stubborn opinion "0" can be denoted by

$$\boldsymbol{z}(t+1) = \frac{\boldsymbol{X}_k \boldsymbol{s} + \boldsymbol{X}_k \boldsymbol{A} \boldsymbol{z}(t)}{\boldsymbol{I} + \boldsymbol{D}}, \tag{8}$$

Since $\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_k \boldsymbol{A}$ is still an M-matrix, The equilibrium state of $\boldsymbol{z}(t)$ in (8) is

$$\boldsymbol{z}_k = \boldsymbol{Y}_k \boldsymbol{s}_k, \tag{9}$$

where $\boldsymbol{Y}_k = (\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_k \boldsymbol{A})^{-1}$ and $\boldsymbol{s}_k = \boldsymbol{X}_k \boldsymbol{s}$. Thus, the overall opinion can be denoted as

$$\mathcal{H}(L) = \boldsymbol{1}^\top \boldsymbol{z}_k. \tag{10}$$

It can be found that $\mathcal{H}(L)$ only has one dynamic component $\boldsymbol{X}_k$, which is more succinct than $\mathcal{F}(L)$. Thereby, taking $\mathcal{H}(L)$ as the objective function, we reformulate Problem 1 as the following OMin problem.

Problem 2 (OMin)**.** *Given a social network $\mathcal{G} = (V, E)$ and a positive integer $k \ll |V|$, the aim of OMin is to identify an OL set $L \subseteq V$ ($|L| = k$, $z_i \equiv 0$ for $v_i \in L$) such that $\mathcal{H}(L)$ is minimized.*

Properties of opinion minimization. [Gionis *et al.*, 2013] demonstrated the monotonicity and modularity of $\mathcal{F}(L)$ using the random walk theory. Considering that $\mathcal{H}(L)$ possesses a more succinct expression, we will next mathematically demonstrate that $\mathcal{H}(L)$ still satisfies monotonicity and modularity.

Theorem 1 (Monotonicity of $\mathcal{H}(L)$)**.** *$\mathcal{H}(L)$ is monotonically decreasing. In other words, for any two OL sets $U \subseteq W \subseteq V$, $\mathcal{H}(U) \geq \mathcal{H}(W)$ holds.*

Proof. For brevity, we use u to denote $|U|$, and $\boldsymbol{x} \succeq \boldsymbol{y}$ to denote that $\boldsymbol{x} - \boldsymbol{y}$ is non-negative. Then, by (9), we have

$$\begin{aligned}&(\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_u \boldsymbol{A})\boldsymbol{z}_u = \boldsymbol{s}_u, \\ &(\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_u \boldsymbol{A} + (\boldsymbol{X}_u - \boldsymbol{X}_w)\boldsymbol{A})\boldsymbol{z}_w = \boldsymbol{s}_w.\end{aligned} \tag{11}$$

Subtracting the two equations in (11) yields

$$\boldsymbol{z}_u - \boldsymbol{z}_w = \boldsymbol{Y}_u \left(\boldsymbol{s}_u - \boldsymbol{s}_w + (\boldsymbol{X}_u - \boldsymbol{X}_w)\boldsymbol{A}\boldsymbol{z}_w\right). \tag{12}$$

As $\boldsymbol{F} = (\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{A})^{-1} \succeq \boldsymbol{0}$ is doubly stochastic, we have $\boldsymbol{Y}_u = (\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_u \boldsymbol{A})^{-1} \succeq \boldsymbol{0}$. Then, as $U \subseteq W$, $\boldsymbol{s}_u \succeq \boldsymbol{s}_w$ and $\boldsymbol{X}_u \succeq \boldsymbol{X}_w$ hold, showing that $\boldsymbol{z}_u \succeq \boldsymbol{z}_w$. Therefore, we can derive $\boldsymbol{1}^\top(\boldsymbol{z}_u - \boldsymbol{z}_w) \geq 0$, which demonstrates that Theorem 1 holds true. □

Theorem 2 (Modularity of $\mathcal{H}(L)$)**.** *$\mathcal{H}(L)$ is supermodular. In other words, for any two OL sets $U \subseteq W \subseteq V$, $\mathcal{H}(U) - \mathcal{H}(U \cup \{v^*\}) \geq \mathcal{H}(W) - \mathcal{H}(W \cup \{v^*\})$ holds for any node $v^* \in V \backslash W$.*

Proof. For brevity, we use U^* to denote $U \cup \{v^*\}$, and $\boldsymbol{y}_u$ to denote $(\boldsymbol{s}_u - \boldsymbol{s}_{u^*} + (\boldsymbol{X}_u - \boldsymbol{X}_{u^*})\boldsymbol{A}\boldsymbol{z}_{u^*})$. Similarly, by (12), we have

$$\begin{aligned}\boldsymbol{z}_u - \boldsymbol{z}_{u^*} &= \boldsymbol{Y}_u \boldsymbol{y}_u,\\ \boldsymbol{z}_w - \boldsymbol{z}_{w^*} &= \boldsymbol{Y}_w(\boldsymbol{y}_u - (\boldsymbol{y}_u - \boldsymbol{y}_w)).\end{aligned} \tag{13}$$

Since $\boldsymbol{s}_u - \boldsymbol{s}_{u^*} = \boldsymbol{s}_w - \boldsymbol{s}_{w^*}$, $\boldsymbol{X}_u - \boldsymbol{X}_{u^*} = \boldsymbol{X}_w - \boldsymbol{X}_{w^*}$, and $\boldsymbol{z}_{u^*} \succeq \boldsymbol{z}_{w^*}$, we have $\boldsymbol{y}_u \succeq \boldsymbol{y}_w$. Thus, subtracting the two equations in (13) yields

$$\begin{aligned}&(\boldsymbol{z}_u - \boldsymbol{z}_{u^*}) - (\boldsymbol{z}_w - \boldsymbol{z}_{w^*}) \succeq (\boldsymbol{Y}_u - \boldsymbol{Y}_w)\, \boldsymbol{y}_u\\ &= \boldsymbol{Y}_u(-\boldsymbol{Y}_u^{-1} + \boldsymbol{Y}_w^{-1})\boldsymbol{Y}_w \boldsymbol{y}_u\\ &= \boldsymbol{Y}_u((\boldsymbol{X}_u - \boldsymbol{X}_w)\boldsymbol{A})\boldsymbol{Y}_w \boldsymbol{y}_u.\end{aligned} \tag{14}$$

As demonstrated in the proof of Theorem 1, $\boldsymbol{Y}_u \succeq \boldsymbol{0}$, $\boldsymbol{Y}_w \succeq \boldsymbol{0}$, and $\boldsymbol{y}_u \succeq \boldsymbol{0}$ hold, showing that $(\boldsymbol{z}_u - \boldsymbol{z}_{u^*}) \succeq (\boldsymbol{z}_w - \boldsymbol{z}_{w^*})$. Therefore, we can derive $\boldsymbol{1}^\top(\boldsymbol{z}_u - \boldsymbol{z}_{u^*}) \geq \boldsymbol{1}^\top(\boldsymbol{z}_w - \boldsymbol{z}_{w^*})$, which demonstrates that Theorem 2 holds true. □

Despite these desired properties of $\mathcal{H}(L)$, its computational complexity is still bounded by $O(n^3)$, making iterative optimization methods (e.g., GA, metaheuristic algorithms) difficult to be applied to optimize it in large-scale networks. Therefore, we propose BES to achieve bounded estimation of $\mathcal{H}(L)$, elaborated in the next section.

5 Bounded Estimation Scheme

It can be found that the general $O(n^3)$ computational complexity of $\mathcal{H}(L)$ mainly comes from inverting $\boldsymbol{Y}_k^{-1} = \boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_k\boldsymbol{A}$. To accelerate this critical step, we consider using the following solver for symmetric and diagonally dominant matrix (SDDM) [Spielman and Teng, 2014].

Lemma 1 (SDDM Solver [Spielman and Teng, 2014]). *Given an $n \times n$ SDDM $\boldsymbol{B}$ with $2m$ non-zero off-diagonal entries and an accuracy parameter $\delta > 0$, there is a solver $\boldsymbol{u} = Solve(\boldsymbol{B}, \boldsymbol{v}, \delta)$ satisfying $||\boldsymbol{u} - \boldsymbol{B}^\dagger\boldsymbol{v}||_{\boldsymbol{B}} \leq \delta||\boldsymbol{B}^\dagger\boldsymbol{v}||_{\boldsymbol{B}}$, where $||\boldsymbol{v}||_{\boldsymbol{B}} = \sqrt{\boldsymbol{v}^\top\boldsymbol{B}\boldsymbol{v}}$ and $\boldsymbol{B}^\dagger$ is the pseudo-inverse of $\boldsymbol{B}$. The expected time of the solver is $\widetilde{O}(m)$, where $\widetilde{O}(\cdot)$ suppresses the logarithmic factors.*

Lemma 1 establishes that the product $\boldsymbol{B}^{-1}\boldsymbol{v}$ can, with the aid of an SDDM solver, be consistently approximated by $\boldsymbol{u}$ within controllable accuracy in nearly linear time. It is thus reasonable to contemplate estimating $\boldsymbol{z}_k = \boldsymbol{Y}_k\boldsymbol{s}_k$ in an analogous manner. Nevertheless, despite the diagonal dominance of $\boldsymbol{Y}_k^{-1}$, its asymmetric nature prevents the direct application of this estimation approach. To address this obstacle, we subsequently construct an SDDM transformation for $\boldsymbol{Y}_k^{-1}$.

First, we analyze the properties of $\boldsymbol{Y}_1^{-1}$, as it serves as the basic case for our purpose.

Lemma 2. *When $L = \{v_i\}$, $\boldsymbol{Y}_1^{-1}$ satisfies*

- $\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_i = (1 + d_i)^{-1}$.
- *For any $j \neq i$, $\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_j = 0$.*

Proof. Let $\boldsymbol{a}_i = \boldsymbol{A}\boldsymbol{e}_i$. Then, $\boldsymbol{Y}_1$ can be computed by

$$\boldsymbol{Y}_1 = (\boldsymbol{F}^{-1} + \boldsymbol{e}_i\boldsymbol{a}_i^\top)^{-1}, \tag{15}$$

According to the Sherman-Morrison formula [Sherman and Morrison, 1950], we have

$$\boldsymbol{Y}_1 = \boldsymbol{F} - \frac{\boldsymbol{F}\boldsymbol{e}_i\boldsymbol{a}_i^\top\boldsymbol{F}}{1 + \boldsymbol{a}_i^\top\boldsymbol{F}\boldsymbol{e}_i}. \tag{16}$$

Since $(\boldsymbol{I} + \boldsymbol{D} - \boldsymbol{A})\boldsymbol{F} = \boldsymbol{I}$, $\boldsymbol{A}\boldsymbol{F} = (\boldsymbol{I} + \boldsymbol{D})\boldsymbol{F} - \boldsymbol{I}$ holds. Thereby, we have

$$\begin{aligned}\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_i &= \boldsymbol{e}_i^\top\boldsymbol{F}\boldsymbol{e}_i - \boldsymbol{e}_i^\top\frac{\boldsymbol{F}\boldsymbol{e}_i\boldsymbol{a}_i^\top\boldsymbol{F}}{1 + \boldsymbol{a}_i^\top\boldsymbol{F}\boldsymbol{e}_i}\boldsymbol{e}_i^\top\\ &= \boldsymbol{e}_i^\top\boldsymbol{F}\boldsymbol{e}_i - \frac{(\boldsymbol{e}_i^\top\boldsymbol{F}\boldsymbol{e}_i)(\boldsymbol{e}_i^\top\boldsymbol{A}\boldsymbol{F}\boldsymbol{e}_i)}{1 + (\boldsymbol{e}_i^\top\boldsymbol{A}\boldsymbol{F}\boldsymbol{e}_i)}\\ &= f_{ii} - \frac{f_{ii}(f_{ii}(1 + d_i) - 1)}{1 + f_{ii}(1 + d_i) - 1} = \frac{1}{1 + d_i}.\end{aligned} \tag{17}$$

Similarly, we have

$$\begin{aligned}\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_j &= \boldsymbol{e}_i^\top\boldsymbol{F}\boldsymbol{e}_j - \boldsymbol{e}_i^\top\frac{\boldsymbol{F}\boldsymbol{e}_i\boldsymbol{a}_i^\top\boldsymbol{F}}{1 + \boldsymbol{a}_i^\top\boldsymbol{F}\boldsymbol{e}_i}\boldsymbol{e}_j^\top\\ &= \boldsymbol{e}_i^\top\boldsymbol{F}\boldsymbol{e}_j - \frac{(\boldsymbol{e}_i^\top\boldsymbol{F}\boldsymbol{e}_i)(\boldsymbol{e}_i^\top\boldsymbol{A}\boldsymbol{F}\boldsymbol{e}_j)}{1 + (\boldsymbol{e}_i^\top\boldsymbol{A}\boldsymbol{F}\boldsymbol{e}_i)}\\ &= f_{ij} - \frac{f_{ii}f_{ij}(1 + d_i)}{1 + f_{ii}(1 + d_i) - 1} = 0.\end{aligned} \tag{18}$$

Hence, Lemma 2 holds true. □

Lemma 3. *There is an SDDM transformation for $\boldsymbol{Y}_1^{-1}$, $\boldsymbol{T}_1^{-1} = \boldsymbol{Y}_1^{-1} + \boldsymbol{a}_i\boldsymbol{e}_i^\top$, satisfying*

- *For any $j \neq i$, $\boldsymbol{T}_1\boldsymbol{e}_j = \boldsymbol{Y}_1\boldsymbol{e}_j$.*
- $\boldsymbol{T}_1\boldsymbol{s}_1 = \boldsymbol{Y}_1\boldsymbol{s}_1$.

Proof. According to the Sherman-Morrison formula and the results in Lemma 2, we have

$$\begin{aligned}\boldsymbol{T}_1 &= \boldsymbol{Y}_1 - \frac{\boldsymbol{Y}_1\boldsymbol{a}_i\boldsymbol{e}_i^\top\boldsymbol{Y}_1}{1 + \boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{a}_i}\\ &= \boldsymbol{Y}_1 - \frac{\boldsymbol{Y}_1\boldsymbol{a}_i\boldsymbol{e}_i^\top\boldsymbol{Y}_1}{1 + \sum_j a_{ij}\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_j}\\ &= \boldsymbol{Y}_1 - \boldsymbol{Y}_1\boldsymbol{a}_i\boldsymbol{e}_i^\top\boldsymbol{Y}_1.\end{aligned} \tag{19}$$

Therefore, we have

$$\begin{aligned}\boldsymbol{T}_1\boldsymbol{e}_j &= \boldsymbol{Y}_1\boldsymbol{e}_j - \boldsymbol{Y}_1\boldsymbol{a}_i\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_j\\ &= \boldsymbol{Y}_1\boldsymbol{e}_j - \boldsymbol{Y}_1\boldsymbol{a}_i(\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_j)\\ &= \boldsymbol{Y}_1\boldsymbol{e}_j,\end{aligned} \tag{20}$$

which yields

$$\boldsymbol{T}_1\boldsymbol{s}_1 = \sum_j s_j\boldsymbol{T}_1\boldsymbol{e}_j = \sum_j s_j\boldsymbol{Y}_1\boldsymbol{e}_j = \boldsymbol{Y}_1\boldsymbol{s}_1. \tag{21}$$

Hence, Lemma 3 holds true. □

Figure 1 uses an example social network to illustrate Lemmas 2 and 3. When $L = \{v_4\}$, $\boldsymbol{e}_i^\top\boldsymbol{Y}_1\boldsymbol{e}_i = (1 + d_4)^{-1} = 0.5$ and the off-diagonal elements in the fourth row of $\boldsymbol{Y}_1$ are all 0. Using the SDDM transformation, all elements except those in the fourth column will remain unchanged, which results in $\boldsymbol{Y}_1\boldsymbol{s}_1 = \boldsymbol{T}_1\boldsymbol{s}_1$. Based on the theoretical results in Lemmas 2 and 3, we can use induction to obtain the following SDDM transformation of $\boldsymbol{Y}_k^{-1}$.

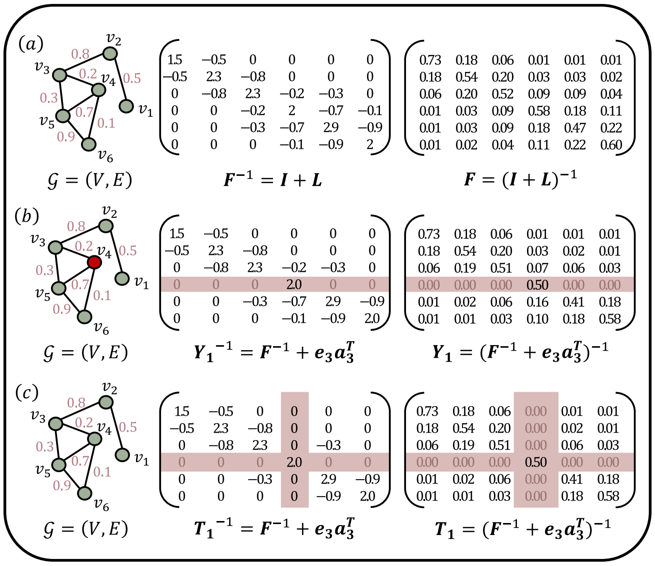

Figure 1: Illustration of Lemmas 2 and 3. (a) An example social network and its forest matrix. (b) $\boldsymbol{Y}_1^{-1}$ and $\boldsymbol{Y}_1$ when $L = \{v_4\}$. (c) $\boldsymbol{T}_1^{-1}$ and $\boldsymbol{T}_1$ when $L = \{v_4\}$.

Theorem 3 (SDDM Transformation of $\boldsymbol{Y}_k^{-1}$)**.** *There is an SDDM transformation of $\boldsymbol{Y}_k^{-1}$, i.e., $\boldsymbol{T}_k^{-1} = \boldsymbol{Y}_k^{-1} + \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{X}_k)$, satisfying $\boldsymbol{Y}_k \boldsymbol{s}_k = \boldsymbol{T}_k \boldsymbol{s}_k$.*

Theorem 3 demonstrates that transforming $\boldsymbol{Y}_k^{-1}$ into $\boldsymbol{T}_k^{-1}$ allows $\boldsymbol{z}_k$ to be equivalently computed as $\boldsymbol{T}_k \boldsymbol{s}_k$. Based on this result, $\boldsymbol{z}_k$ can be estimated via the SDDM solver presented in Lemma 1 (i.e., $\tilde{\boldsymbol{z}}_k = \text{Solve}(\boldsymbol{T}_k^{-1}, \boldsymbol{s}_k, \delta)$). Next, we derive the bounds for this estimation.

Theorem 4 (Bounded Estimation of $\mathcal{H}(L)$)**.** *For social network $\mathcal{G} = (V, E)$ with internal opinion vector $\boldsymbol{s}$, OL set L, and error parameter $\epsilon \in (0, 1/2]$, if the accuracy parameter δ satisfies $0 < \delta \leq \epsilon/n$, then the following relation holds:*

$$|\tilde{\mathcal{H}}(L) - \mathcal{H}(L)| \leq \epsilon, \tag{22}$$

where $\tilde{\mathcal{H}}(L) = \boldsymbol{1}^\top \tilde{\boldsymbol{z}}_k$ and $\tilde{\boldsymbol{z}}_k = \text{Solve}(\boldsymbol{T}_k^{-1}, \boldsymbol{s}_k, \delta)$.

Proof. It can be found that

$$\begin{aligned}|\boldsymbol{1}^\top \tilde{\boldsymbol{z}}_k - \boldsymbol{1}^\top \boldsymbol{z}_k| &= |\boldsymbol{1}^\top \tilde{\boldsymbol{z}}_k - \boldsymbol{1}^\top \boldsymbol{T}_k \boldsymbol{s}_k| \\ &\leq \|\boldsymbol{1}\| \|\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k\| = \sqrt{n} \|\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k\|. \end{aligned} \tag{23}$$

Since $\boldsymbol{T}_k^{-1} = \boldsymbol{I} + \boldsymbol{D} - \boldsymbol{X}_k \boldsymbol{A} + \boldsymbol{A}(\boldsymbol{I} - \boldsymbol{X}_k)$, we can derive

$$\begin{aligned}\|\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k\|_{\boldsymbol{T}_k^{-1}} &\geq \sqrt{(\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k)^\top \boldsymbol{I} (\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k)} \\ &= \|\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k\|. \end{aligned} \tag{24}$$

Then, we have

$$\begin{aligned}&\|\tilde{\boldsymbol{z}}_k - \boldsymbol{T}_k \boldsymbol{s}_k\|_{\boldsymbol{T}_k^{-1}} \leq \delta \|\boldsymbol{T}_k \boldsymbol{s}_k\|_{\boldsymbol{T}_k^{-1}} \\ &\leq \delta \sqrt{\boldsymbol{s}_k^\top \boldsymbol{T}_k \boldsymbol{T}_k^{-1} \boldsymbol{T}_k \boldsymbol{s}_k} = \delta \sqrt{\boldsymbol{s}_k^\top \boldsymbol{T}_k \boldsymbol{s}_k} \\ &\leq \delta \sqrt{\boldsymbol{1}^\top \boldsymbol{F} \boldsymbol{1}} = \delta \sqrt{n} \end{aligned} \tag{25}$$

Algorithm 1 BES

Input: social network: $\mathcal{G} = (V, E)$; OL set: L; internal opinion vector $\boldsymbol{s}$; error parameter ϵ;
Output: OL set: $\boldsymbol{z}_k$;

```
s_k ← s, T_k^{-1} ← I + L;
for i in L do
    s_k[i] ← 0;
    for j = 1 to n do
        if j = i then continue;
        T_k^{-1}[i, j] ← 0, T_k^{-1}[j, i] ← 0;
    end for
end for
z̃_k ← Solve(T_k^{-1}, s_k, ε/n);
H̃(L) ← 1^⊤ z̃_k;
return H̃(L).
```

Combing (23), (24), and (25), we have

$$|\boldsymbol{1}^\top \tilde{\boldsymbol{z}}_k - \boldsymbol{1}^\top \boldsymbol{z}_k| \leq \delta n \leq \epsilon. \tag{26}$$

showing that Theorem 4 holds. □

Based on Theorem 4, BES is presented in Algorithm 1. With the assistance of BES, the computational complexity of evaluating $\mathcal{H}(L)$ can be reduced from $O(n^3)$ to $\widetilde{O}(kn + m)$. This reduction significantly enhances the efficiency of various iterative algorithms that rely on objective evaluation (e.g., GA, metaheuristics). We will further demonstrate this in the experimental section.

6 Decomposition-Based Decremental Estimation Algorithm

Due to the monotonicity and modularity of the objective function in the OM problem, [Gionis *et al.*, 2013] employed GA with $(1 - 1/e)$ approximation ratio for optimization. Given that OMin also satisfies monotonicity and modularity, GA remains the preferred choice for providing theoretical performance guarantees. The GA for the OMin problem is outlined in Algorithm 2. As can be seen, the kn objective evaluations required by GA result in a complexity as high as $O(kn^4)$. Although our proposed BES can effectively reduce the computational complexity of objective evaluation, the resulting complexity of $\widetilde{O}(kn(kn + m))$ still limits algorithmic scalability. Therefore, we introduce DDEA to further enhance optimization efficiency without significantly compromising theoretical performance guarantees.

The main idea of DDEA is to perform a decomposed estimation of the marginal decrease in line 3 of Algorithm 2. Let $\Delta\mathcal{H}_i^{t+1} = \mathcal{H}(L^t) - \mathcal{H}(L^t \cup \{v_i\})$ be the marginal decrease of adding node $v_i \in V \backslash L^t$ to L^t, it can be calculated by

$$\Delta\mathcal{H}_i^{t+1} = \boldsymbol{1}^\top \boldsymbol{T}_t \boldsymbol{s}_t - \boldsymbol{1}^\top \boldsymbol{T}_{t+1} \boldsymbol{s}_{t+1}. \tag{27}$$

The relation of $\boldsymbol{T}_t$ and $\boldsymbol{T}_{t+1}$ can be expressed by

$$\boldsymbol{C}_{t+1} = (\boldsymbol{T}_t^{-1} + \boldsymbol{e}_i \boldsymbol{a}_i^\top)^{-1} = \boldsymbol{T}_t - \frac{\boldsymbol{T}_t \boldsymbol{e}_i \boldsymbol{a}_i^\top \boldsymbol{T}_t}{1 + \boldsymbol{a}_i^\top \boldsymbol{T}_t \boldsymbol{e}_i}, \tag{28}$$

$$\boldsymbol{T}_{t+1} = (\boldsymbol{C}_{t+1}^{-1} + \boldsymbol{a}_i \boldsymbol{e}_i^\top)^{-1} = \boldsymbol{C}_{t+1} - \frac{\boldsymbol{C}_{t+1} \boldsymbol{a}_i \boldsymbol{e}_i^\top \boldsymbol{C}_{t+1}}{1 + \boldsymbol{e}_i^\top \boldsymbol{C}_{t+1} \boldsymbol{a}_i},$$

Algorithm 2 GA for OMin

Input: social network: $\mathcal{G} = (V, E)$; number of OLs: k; internal opinion vector $\boldsymbol{s}$;
Output: solution set: L^k;

$L^0 \leftarrow \emptyset$;
for t **from** 0 **to** $k-1$ **do**
  $v_i \leftarrow \mathrm{argmax}_{v_i \in V \setminus L^t}(\mathcal{H}(L^t) - \mathcal{H}(L^t \cup \{v_i\}))$;
  $L^{t+1} \leftarrow L^t \cup \{v_i\}$;
end for
return L^k.

where $\boldsymbol{C}_{t+1}$ is an auxiliary matrix. Based on the theoretical results in Lemmas 2 and 3, it can be found that

$$\boldsymbol{T}_{t+1}\boldsymbol{s}_{t+1} = \boldsymbol{C}_{t+1}\boldsymbol{s}_{t+1}. \tag{29}$$

Combining (27), (28), (29), and $\boldsymbol{s}_{t+1} = \boldsymbol{s}_t - s_i\boldsymbol{e}_i$, $\Delta\mathcal{H}_i^{t+1}$ can be decomposed into a polynomial that mainly consists of four components:

$$\Delta\mathcal{H}_i^{t+1} = s_i c_1(1 - c_3c_4) + c_1c_2c_4, \tag{30}$$

where $c_1 = \boldsymbol{1}^\top\boldsymbol{T}_t\boldsymbol{e}_i$, $c_2 = \boldsymbol{a}_i^\top\boldsymbol{T}_t\boldsymbol{s}_t$, $c_3 = \boldsymbol{a}_i^\top\boldsymbol{T}_t\boldsymbol{e}_i$, and $c_4 = (1+\boldsymbol{a}_i^\top\boldsymbol{T}_t\boldsymbol{e}_i)^{-1}$. Since the structure of these four components are very similar to that of $\mathcal{H}(L) = \boldsymbol{1}^\top\boldsymbol{Y}_k\boldsymbol{s}_k$, we can use the SDDM solver again to achieve bounded estimation for these four components, and then for $\Delta\mathcal{H}_i^{t+1}$.

Theorem 5 (Bounded Estimation of $\Delta\mathcal{H}_i^{t+1}$)**.** *For social network $\mathcal{G} = (V, E)$ with internal opinion vector $\boldsymbol{s}_t$, SDDM $\boldsymbol{T}_t^{-1}$, OL set L^t, node $v_i \in V \setminus L^t$, and error parameter $\epsilon \in (0, 1/2]$, we define quantities $\boldsymbol{p} = \mathit{Solve}(\boldsymbol{T}_t^{-1}, \boldsymbol{e}_i, \delta)$, $\boldsymbol{q} = \mathit{Solve}(\boldsymbol{T}_t^{-1}, \boldsymbol{s}_t, \delta)$, $\tilde{c}_1 = \boldsymbol{1}^\top\boldsymbol{p}$, $\tilde{c}_2 = \boldsymbol{a}_i^\top\boldsymbol{q}$, $\tilde{c}_3 = \boldsymbol{a}_i^\top\boldsymbol{p}$, and $\tilde{c}_4 = (1+\boldsymbol{a}_i^\top\boldsymbol{p})^{-1}$. If the accuracy parameter δ satisfies $0 < \delta \le \epsilon/(36n^2)$, then the following relation holds:*

$$|\Delta\tilde{\mathcal{H}}_i^{t+1} - \Delta\mathcal{H}_i^{t+1}| \le \epsilon, \tag{31}$$

where $\Delta\tilde{\mathcal{H}}_i^{t+1} = s_i\tilde{c}_1(1-\tilde{c}_3\tilde{c}_4) + \tilde{c}_1\tilde{c}_2\tilde{c}_4$.

The proof idea is similar to that of Theorem 4. With Theorem 5, the algorithmic framework of DDEA is presented in Algorithms 3 and 4. Since Algorithm 3 only involves $\widetilde{O}(mn)$ expected time, the time complexity of DDEA is $\widetilde{O}(kmn)$. Then, as the error of estimating $\Delta\mathcal{H}_i^{t+1}$ is bounded by ϵ, we can follow the proof idea of the original GA to demonstrate the theoretical performance guarantee of DDEA.

Theorem 6 (Theoretical Performance Guarantee of DDEA)**.** *The output L^k of DDEA satisfies*

$$\mathcal{H}(\emptyset) - \mathcal{H}(L^k) \ge (1 - \frac{1}{e})(\mathcal{H}(\emptyset) - \mathcal{H}(L^o) - 2k\epsilon), \tag{32}$$

where L^o is the optimal OL set for OMin.

7 Experiments

Datasets and baselines. The largest connected components of eight common real-world social networks of different scales are selected as our test datasets, and their basic information are shown in the first three columns of Table 1.

Algorithm 3 Index($\cdot$)

Input: social network: $\mathcal{G} = (V, E)$; OL set L^t; internal opinion vector $\boldsymbol{s}_t$; SDD matrix: $\boldsymbol{T}_t^{-1}$; error parameter: ϵ;
Output: index of the node offering maximal approximate marginal decrease on $\mathcal{H}(L^t)$: i;

$\tilde{\boldsymbol{\Delta}} \leftarrow \boldsymbol{0}$, $\delta \leftarrow \epsilon/(36n^2)$;
for i **from** 1 **to** n **do**
  if $v_i \in L^t$ **then** continue;
  $\boldsymbol{p} \leftarrow \mathrm{Solve}(\boldsymbol{T}_t^{-1}, \boldsymbol{e}_i, \delta)$, $\boldsymbol{q} \leftarrow \mathrm{Solve}(\boldsymbol{T}_t^{-1}, \boldsymbol{s}_t, \delta)$;
  $\tilde{c}_1 \leftarrow \boldsymbol{1}^\top\boldsymbol{p}$, $\tilde{c}_2 \leftarrow \boldsymbol{a}_i^\top\boldsymbol{q}$,
  $\tilde{c}_3 \leftarrow \boldsymbol{a}_i^\top\boldsymbol{p}$, $\tilde{c}_4 \leftarrow (1+\boldsymbol{a}_i^\top\boldsymbol{p})^{-1}$;
  $\tilde{\boldsymbol{\Delta}}_i \leftarrow s_i\tilde{c}_1(1-\tilde{c}_3\tilde{c}_4) + \tilde{c}_1\tilde{c}_2\tilde{c}_4$;
end for
$i \leftarrow \mathrm{argmax}_i\, \tilde{\boldsymbol{\Delta}}_i$;
return i.

Algorithm 4 DDEA

Input: social network: $\mathcal{G} = (V, E)$; number of OLs: k; internal opinion vector $\boldsymbol{s}$; error parameter ϵ;
Output: OL set: L^k;

$L^0 \leftarrow \emptyset$, $\boldsymbol{s}_0 \leftarrow \boldsymbol{s}$, $\boldsymbol{T}_0^{-1} \leftarrow \boldsymbol{I} + \boldsymbol{L}$;
for t **from** 0 **to** $k-1$ **do**
  Get the index of the node offering maximal approximate marginal decrease on $\mathcal{H}(L^t)$:
  $i \leftarrow \mathrm{Index}(\mathcal{G}, L^t, \boldsymbol{s}_t, \boldsymbol{T}_t^{-1}, \epsilon)$;
  $L^{t+1} \leftarrow L^t \cup \{v_i\}$, $\boldsymbol{s}_{t+1} \leftarrow \boldsymbol{s}_t - s_i\boldsymbol{e}_i$,
  $\boldsymbol{T}_{t+1}^{-1} \leftarrow \boldsymbol{T}_t^{-1} + \boldsymbol{e}_i\boldsymbol{a}_i^\top + \boldsymbol{a}_i\boldsymbol{e}_i^\top$;
end for
return L^k.

All these datasets are available in public network libraries including KONECT [Kunegis, 2013] and Network Repository [Rossi and Ahmed, 2015].

To assess BES and DDEA, we select GA, a state-of-the-art metaheuristic OM method MACOM [Liu *et al.*, 2025b], and three state-of-the-art heuristic methods for identifying influential nodes (Improved Local and Global Centrality (ILGC) [Guo *et al.*, 2025], Electric Potential Centrality (EPC) [Ullah *et al.*, 2024], and DomiRank Centrality (DRC) [Engsig *et al.*, 2024]) as our baselines. The parameters of these baselines strictly follow their original publication specifications.

Basic settings. Following many related studies, the internal opinions of social members are randomly sampled from [0,1]. To ensure the theoretical performances of DDEA, we set ϵ to $1/200k$ (i.e., $2k\epsilon = 10^{-2}$). Since the test datasets have different scales, we use $\bar{\mathcal{F}} = \mathcal{F}(L)/n$ and $\bar{\mathcal{H}} = \mathcal{H}(L)/n$ to evaluate algorithmic effectiveness, and the average runtime across 10 independent runs are used to evaluate algorithmic efficiency [Liu *et al.*, 2022; Liu *et al.*, 2025a].

All experiments are conducted on a PC with a 3.0 GHz Intel Core i7 processor, and all algorithms are implemented by the Julia language[1]. Besides, the source code of the SDDM solver is publicly available[2].

[1]https://julialang.org/
[2]https://github.com/danspielman/Laplacians.jl

$\mathcal{G}$	n	m	Runtime/s(↓)	
			EC	BES
$\mathcal{G}_1$. EmailUniv	1,133	5,451	0.09	0.02
$\mathcal{G}_2$. Hamsterster	2,000	16,097	0.45	0.03
$\mathcal{G}_3$. GrQc	4,158	13,422	2.55	0.03
$\mathcal{G}_4$. PagesComp	14,113	52,126	74.22	0.06
$\mathcal{G}_5$. GemsecRO	41,773	125,826	2102.19	0.12
$\mathcal{G}_6$. Douban	154,908	327,162	–	0.25
$\mathcal{G}_7$. Delicious	536,108	1,365,961	–	1.22
$\mathcal{G}_8$. YoutubeSnap	1,134,890	2,987,624	–	2.93

Table 1: Runtime of Exact Computation (EC) and BES for a random OL set with $k = 100$ on all test datasets, where the symbol "–" indicates results that could not be obtained within 10000 seconds, and the numbers with a gray background indicate better results.

	Method	$\bar{\mathcal{F}}(\uparrow)$	$\bar{\mathcal{H}}(\downarrow)$	Runtime/s(↓)
$\mathcal{G}_1$	GA	0.7795	0.2341	4338.35
	GA+BES	0.7795	0.2341	324.09
	MACOM	0.7785	0.2363	1998.61
	MACOM+BES	0.7787	0.2362	37.94
$\mathcal{G}_2$	GA	–	–	–
	GA+BES	0.7984	0.2003	1595.92
	MACOM	0.7977	0.2022	7024.30
	MACOM+BES	0.7975	0.2025	93.43

Table 2: $\bar{\mathcal{F}}$, $\bar{\mathcal{H}}$, and Runtime of GA, MACOM, BES-assisted GA (GA+BES), and BES-assisted MACOM (MACOM+BES) for addressing the OMin problem with $k = 50$ on $\mathcal{G}_1$ and $\mathcal{G}_2$.

Effectiveness of BES. We first demonstrate the effectiveness of BES. As shown in Table 1, when evaluating a random OL set with $k = 100$, BES significantly reduces the runtime while exhibiting excellent scalability. For example, on $\mathcal{G}_8$, which contains more than one million nodes and edges, its runtime does not exceed three seconds. In contrast, the runtime of EC on $\mathcal{G}_5$ with about 42 thousands nodes and 126 thousands edges has already exceeded half an hour. To further illustrate that BES can improve the efficiency of iterative algorithms, we next conduct experimental analyses using GA and MACOM as examples. As shown in Table 2, with the assistance of BES, the performance of GA and MACOM is not significantly affected, yet their runtime is notably reduced. Moreover, compared with GA, BES brings a more pronounced efficiency improvement to MACOM, which is mainly because MACOM requires fewer evaluations of the objective function. In summary, the experimental results align closely with our theoretical analysis, strongly demonstrating the superiority of BES.

Effectiveness of DDEA. We next demonstrate the efficiency of DDEA. As shown in Figure 2 and Figure 3, on $\mathcal{G}_3$, DDEA and GA+BES achieve identical optimal performance across all settings of k, while the runtime of DDEA is two orders of magnitude lower than that of GA+BES. This is mainly

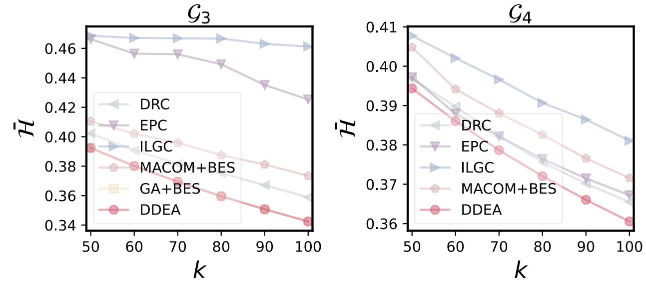

Figure 2: $\bar{\mathcal{H}}$ of DDEA, GA+BES, MACOM+BES, ILGC, EPC, and DRC vs. k on $\mathcal{G}_3$ and $\mathcal{G}_4$.

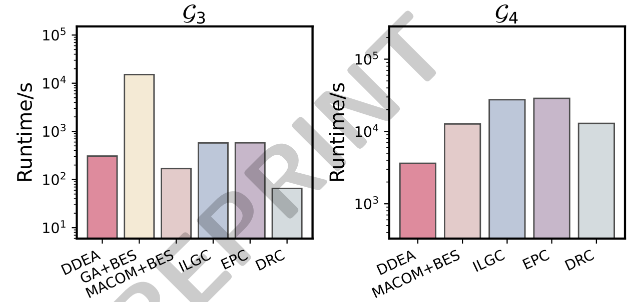

Figure 3: Runtime of DDEA, GA+BES, MACOM+BES, ILGC, EPC, and DRC when $k = 100$ on $\mathcal{G}_3$ and $\mathcal{G}_4$.

because DDEA and GA+BES share similar greedy principles and error estimation strategies, yet the $O(kmn)$ time complexity of DDEA significantly reduces its runtime. When applied to the larger-scale $\mathcal{G}_4$, DDEA not only maintains optimal performance but also achieves the shortest runtime. This is largely because MACOM, ILGC, and EPC all involve the computation of graph-related distance matrices, while DRC requires calculating the smallest eigenvalue of the adjacency matrix, leading to their higher time complexities. In summary, the experimental results strongly confirm that DDEA possesses not only theoretical superiority but also outstanding practical efficiency.

8 Conclusion

This paper has provided an inverse optimization perspective on the OM problem by reformulating it as the OMin problem. By showing that the objective function of OMin could be expressed in a succinct form while still preserving monotonicity and modularity, we have first proposed BES to achieve efficient calculation for its objective function, which could provide bounded error in $\widetilde{O}(kn + m)$ expected time. Then, we have specifically designed DDEA with a near $(1 - 1/e)$ approximation ratio for OMin in $\widetilde{O}(kmn)$ expected time. Extensive experimental results have demonstrated that BES could significantly enhance the efficiency of iterative algorithms to address OMin, while DDEA has not only theoretical advantage but also practical efficiency superiority. We believe this work could inspire more efforts on opinion optimization in social networks.

Acknowledgments

This work was supported in part by the Research Grants Council of the Hong Kong SAR, China, under Grant CityU11215622, and in part by the Natural Science Foundation of China under Grant 62276223.

References

[Abebe *et al.*, 2018] Rediet Abebe, Jon Kleinberg, David Parkes, and Charalampos E Tsourakakis. Opinion dynamics with varying susceptibility to persuasion. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1089–1098, 2018.

[Casaló *et al.*, 2020] Luis V Casaló, Carlos Flavián, and Sergio Ibáñez-Sánchez. Influencers on instagram: Antecedents and consequences of opinion leadership. *Journal of Business Research*, 117:510–519, 2020.

[Chen and Rácz, 2021] Mayee F Chen and Miklós Z Rácz. An adversarial model of network disruption: Maximizing disagreement and polarization in social networks. *IEEE Transactions on Network Science and Engineering*, 9(2):728–739, 2021.

[Cinus *et al.*, 2023] Federico Cinus, Aristides Gionis, and Francesco Bonchi. Rebalancing social feed to minimize polarization and disagreement. In *Proceedings of the ACM International Conference on Information and Knowledge Management (CIKM)*, pages 369–378, 2023.

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

[Engsig *et al.*, 2024] Marcus Engsig, Alejandro Tejedor, Yamir Moreno, Efi Foufoula-Georgiou, and Chaouki Kasmi. Domirank centrality reveals structural fragility of complex networks via node dominance. *Nature Communications*, 15(1):56, 2024.

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

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

[Gong *et al.*, 2016] Maoguo Gong, Chao Song, Chao Duan, Lijia Ma, and Bo Shen. An efficient memetic algorithm for influence maximization in social networks. *IEEE Computational Intelligence Magazine*, 11(3):22–33, 2016.

[Guo *et al.*, 2025] Qiang Guo, Meng-Meng Ye, and Jian-Guo Liu. Identifying important nodes based on local and global information. *Expert Systems with Applications*, page 129192, 2025.

[He *et al.*, 2021] Qiang He, Lihong Sun, Xingwei Wang, Zhenkun Wang, Min Huang, Bo Yi, Yuantian Wang, and Lianbo Ma. Positive opinion maximization in signed social networks. *Information Sciences*, 558:34–49, 2021.

[He *et al.*, 2023] Qiang He, Hui Fang, Jie Zhang, and Xingwei Wang. Dynamic opinion maximization in social networks. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):350–361, 2023.

[He *et al.*, 2025] Qiang He, Li Zhang, Hui Fang, Xingwei Wang, Lianbo Ma, Keping Yu, and Jie Zhang. Multistage competitive opinion maximization with Q-learning-based method in social networks. *IEEE Transactions on Neural Networks and Learning Systems*, 36(4):7158–7168, 2025.

[Hong *et al.*, 2020] Wenjing Hong, Chao Qian, and Ke Tang. Efficient minimum cost seed selection with theoretical guarantees for competitive influence maximization. *IEEE Transactions on Cybernetics*, 51(12):6091–6104, 2020.

[Howard *et al.*, 2000] Kim Ammann Howard, Todd Rogers, Beth Howard-Pitney, June A Flora, Gregory J Norman, and Kurt M Ribisl. Opinion leaders' support for tobacco control policies and participation in tobacco control activities. *American Journal of Public Health*, 90(8):1283, 2000.

[Hudson and Khamfroush, 2020] Nathaniel Hudson and Hana Khamfroush. Behavioral information diffusion for opinion maximization in online social networks. *IEEE Transactions on Network Science and Engineering*, 8(2):1259–1268, 2020.

[Hunter and Zaman, 2022] David Scott Hunter and Tauhid Zaman. Optimizing opinions with stubborn agents. *Operations Research*, 70(4):2119–2137, 2022.

[Jiang *et al.*, 2011] Qingye Jiang, Guojie Song, Cong Gao, Yu Wang, Wenjun Si, and Kunqing Xie. Simulated annealing based influence maximization in social networks. In *Proceedings of the AAAI conference on Artificial Intelligence (AAAI)*, pages 127–132, 2011.

[Kempe *et al.*, 2003] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 137–146, 2003.

[Kunegis, 2013] Jérôme Kunegis. KONECT: The koblenz network collection. In *Proceedings of the Web Conference (WWW)*, pages 1343–1350, 2013.

[Lee and Chung, 2014] Jong-Ryul Lee and Chin-Wan Chung. A fast approximation for influence maximization in large social networks. In *Proceedings of the ACM Web Conference (WWW)*, pages 1157–1162, 2014.

[Liu *et al.*, 2018] Xinyue Liu, Xiangnan Kong, and Philip S Yu. Active opinion maximization in social networks. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1840–1849, 2018.

[Liu *et al.*, 2022] Yilu Liu, Jing Liu, and Xiangyi Teng. Single-particle optimization for network embedding preserving both local and global information. *Swarm and Evolutionary Computation*, 71:101069, 2022.

[Liu *et al.*, 2023] Yilu Liu, Jing Liu, and Kai Wu. Cost-effective competition on social networks: a multi-objective optimization perspective. *Information Sciences*, 620:31–46, 2023.

[Liu *et al.*, 2025a] Yilu Liu, Xi Lin, Liang Zhao, and Qingfu Zhang. Few for many: Towards efficient and flexible many-objective optimization. *IEEE Transactions on Evolutionary Computation*, 2025.

[Liu *et al.*, 2025b] Yilu Liu, Qingfu Zhang, and Zhenkun Wang. Community opinion maximization in social networks. *IEEE Transactions on Evolutionary Computation*, 29(5):1760–1773, 2025.

[Liu *et al.*, 2026] Yilu Liu, Xi Lin, and Qingfu Zhang. Leader-follower disagreement minimization in social networks. *IEEE Transactions on Evolutionary Computation*, 30(2):810–822, 2026.

[Matakos *et al.*, 2017] Antonis Matakos, Evimaria Terzi, and Panayiotis Tsaparas. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery*, 31(5):1480–1505, 2017.

[Musco *et al.*, 2018] Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. Minimizing polarization and disagreement in social networks. In *Proceedings of the ACM Web Conference (WWW)*, pages 369–378, 2018.

[Nayak *et al.*, 2019] Anuj Nayak, Seyyedali Hosseinalipour, and Huaiyu Dai. Smart information spreading for opinion maximization in social networks. In *Proceedings of the IEEE International Conference on Computer Communications (INFOCOM)*, pages 2251–2259, 2019.

[Nemhauser *et al.*, 1978] George L Nemhauser, Laurence A Wolsey, and Marshall L Fisher. An analysis of approximations for maximizing submodular set functions - I. *Mathematical Programming*, 14:265–294, 1978.

[Rahmattalabi *et al.*, 2021] Aida Rahmattalabi, Shahin Jabbari, Himabindu Lakkaraju, Phebe Vayanos, Max Izenberg, Ryan Brown, Eric Rice, and Milind Tambe. Fair influence maximization: A welfare optimization approach. In *Proceedings of the AAAI conference on Artificial Intelligence (AAAI)*, pages 11630–11638, 2021.

[Rossi and Ahmed, 2015] Ryan A. Rossi and Nesreen K. Ahmed. The network data repository with interactive graph analytics and visualization. In *Proceedings of the Annual AAAI Conference on Artificial Intelligence (AAAI)*, 2015.

[Saha *et al.*, 2025] Arkaprava Saha, Xiangyu Ke, Arijit Khan, and Laks VS Lakshmanan. Beyond influence: Voting theory for opinion maximization. *The VLDB Journal*, 34(5):57, 2025.

[Sherman and Morrison, 1950] Jack Sherman and Winifred J Morrison. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. *The Annals of Mathematical Statistics*, 21(1):124–127, 1950.

[Shirzadi and Zehmakan, 2025] Mohammad Shirzadi and Ahad N Zehmakan. Do stubborn users always cause more polarization and disagreement? a mathematical study. In *Proceedings of the ACM International Conference on Web Search and Data Mining (WSDM)*, pages 309–317, 2025.

[Spielman and Teng, 2014] Daniel A Spielman and Shanghua Teng. Nearly linear time algorithms for preconditioning and solving symmetric, diagonally dominant linear systems. *SIAM Journal on Matrix Analysis and Applications*, 35(3):835–885, 2014.

[Tu *et al.*, 2023] Sijing Tu, Stefan Neumann, and Aristides Gionis. Adversaries with limited information in the friedkin-johnsen model. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pages 2201–2210, 2023.

[Ullah *et al.*, 2024] Aman Ullah, Salah Ud Din, Nasrullah Khan, Cobbinah B Mawuli, and Junming Shao. Towards investigating influencers in complex social networks using electric potential concept from a centrality perspective. *Information Fusion*, 109:102439, 2024.

[Wan *et al.*, 2025] Rong Wan, Fengfeng Wei, and Weineng Chen. GA-ComFit: A genetic algorithm with community fitness sharing niching for multimodal opinion maximization. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2025.

[Wang and Kleinberg, 2023] Yanbang Wang and Jon Kleinberg. On the relationship between relevance and conflict in online social link recommendations. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

[Wang *et al.*, 2019] Liang Wang, Zhiwen Yu, Fei Xiong, Dingqi Yang, Shirui Pan, and Zheng Yan. Influence spread in geo-social networks: A multiobjective optimization perspective. *IEEE Transactions on Cybernetics*, 51(5):2663–2675, 2019.

[Wang *et al.*, 2023] Shuai Wang, Yaochu Jin, and Ming Cai. Enhancing the robustness of networks against multiple damage models using a multifactorial evolutionary algorithm. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(7):4176–4188, 2023.

[Xu *et al.*, 2021] P Xu, W Hu, J Wu, and W Liu. Opinion maximization in social trust networks. In *Proceedings of the International Joint Conferences on Artificial Intelligence (IJCAI)*, pages 1251–1257, 2021.

[Zhou and Zhang, 2023] Xiaotian Zhou and Zhongzhi Zhang. Opinion maximization in social networks via leader selection. In *Proceedings of the ACM Web Conference (WWW)*, pages 133–142, 2023.

[Zhu *et al.*, 2021] Liwang Zhu, Qi Bao, and Zhongzhi Zhang. Minimizing polarization and disagreement in social networks via link recommendation. In *Proceedings of the Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2021.