Minimizing Polarization and Disagreement in the Friedkin–Johnsen Model with Unknown Innate Opinions

Federico Cinus[1,2], **Atsushi Miyauchi**[2], **Yuko Kuroki**[2] and **Francesco Bonchi**[2,3]
[1]Sapienza University, Rome, Italy
[2]CENTAI Institute, Turin, Italy
[3]Eurecat, Barcelona, Spain
{name.surname}@centai.eu

Abstract

The bulk of the literature on opinion optimization in social networks adopts the Friedkin–Johnsen (FJ) opinion dynamics model, in which the innate opinions of all nodes are known: this is an unrealistic assumption. In this paper, we study opinion optimization under the FJ model without the full knowledge of innate opinions. Specifically, we borrow from the literature a series of objective functions, aimed at minimizing polarization and/or disagreement, and we tackle the budgeted optimization problem, where we can query the innate opinions of only a limited number of nodes. Given the complexity of our problem, we propose a framework based on three steps: (1) select the limited number of nodes we query, (2) reconstruct the innate opinions of all nodes based on those queried, and (3) optimize the objective function with the reconstructed opinions. For each step of the framework, we present and systematically evaluate several effective strategies. A key contribution of our work is a rigorous error propagation analysis that quantifies how reconstruction errors in innate opinions impact the quality of the final solutions. Our experiments on various synthetic and real-world datasets show that we can effectively minimize polarization and disagreement even if we have quite limited information about innate opinions.

1 Introduction

The synergetic effect of natural *homophily* and the algorithms employed by social media platforms, e.g., *who-to-follow* recommender systems and *"feed"* content rankers, largely constitutes the information diet of social media users, aligning it with their own opinions. This, together with the natural tendency to *confirmation bias* [Del Vicario *et al.*, 2017], leads to the so-called *"echo-chamber"* effect [Quattrociocchi *et al.*, 2016; Cinus *et al.*, 2022], where individuals with similar mindsets reciprocally reinforce their pre-existing beliefs, which in turn leads to *polarization* [Nikolov *et al.*, 2015; Pariser, 2011]. The rising awareness of the societal risks of extreme polarization driven by social media has spurred a great deal of research on algorithmic interventions aimed at mitigating these harmful effects [Hartman *et al.*, 2022; Garimella *et al.*, 2018; Aslay *et al.*, 2018; Garimella *et al.*, 2017; Tu *et al.*, 2020]. While the bulk of this literature focuses on a static setting, a growing body of work takes into consideration the dynamic nature of the underlying opinion-formation process [Musco *et al.*, 2018; Cinus *et al.*, 2023; Chen *et al.*, 2018; Zhu *et al.*, 2021; Xu *et al.*, 2021], in particular adopting the widely used *Friedkin–Johnsen (FJ)* opinion-dynamics model [Friedkin and Johnsen, 1990].

Background and Related Work. In the FJ model, social media users are depicted as nodes in a network and social ties are represented by edges. Each individual has an *innate opinion*, which may differ from their *expressed opinion* on social media, due to various factors, such as social pressure or fear of judgement. The model operates through an iterative process where users adjust their expressed opinions by taking a weighted average of their own innate opinion and the expressed opinions of their connected peers. It is well known that the equilibrium state of expressed opinions has an analytic form based on the Laplacian of the network and the innate opinions [Musco *et al.*, 2018]. Due to its linear algebraic nature, the model has inspired several optimization problems involving susceptibility [Abebe *et al.*, 2021; Marumo *et al.*, 2021], stubbornness [Xu *et al.*, 2022], exposure timelines [Zhou *et al.*, 2024], and adversarial attacks [Tu *et al.*, 2023]. Additionally, the model has led to some generalizations such as randomized interactions [Fotakis *et al.*, 2016], dynamic social pressure [Ferraioli and Ventre, 2017], and discrete opinion settings [Chierichetti *et al.*, 2013; Auletta *et al.*, 2016].

In the literature on optimization problems under the FJ model, the seminal work of [Musco *et al.*, 2018] introduced the problem of minimizing the sum of *polarization* and *disagreement* by intervening the weights of edges, and designed a polynomial-time approximation scheme based on the convexity of the objective function. Specifically, they investigated two problems: the first problem requires to find a weighted undirected graph that optimizes the objective function, given a total edge-weight budget and without considering a specific input network, while the second aims to optimize users' innate opinions. Following up on this work, several modeling and intervention strategies have been explored [Chen *et al.*, 2018; Abebe *et al.*, 2021; Zhu *et al.*, 2021; Cinus *et al.*, 2023; Zhou *et al.*, 2024; Xu *et al.*, 2021].

Most relevant to our work, [Cinus *et al.*, 2023] recently extended the problem of [Musco *et al.*, 2018] to general directed networks, where the intervention is on the weights of out-going edges of each node, i.e., *rebalancing* the relative importance of the accounts that the user follows, so as to calibrate the frequency with which the contents produced by various accounts are shown in the *social feed* of the user.

Our Contributions. All of this body of work on opinion optimization under the FJ model *assumes that the innate opinions of the nodes are all known and given as input*. However, in practical scenarios, obtaining such information is a task inherently imprecise and expensive. For example, analyzing user opinions on a controversial topic (e.g., COVID-19 vaccination or Brexit) on a social media platform would require either large-scale surveys or extensive behavioral analysis (e.g., posts, reposts, and likes on platforms like 𝕏). Furthermore, even in scenarios in which one is able to reconstruct all the opinions, how the inherent error in such opinion reconstruction influences the performance in the opinion optimization task, has not been addressed in the literature.

In this paper, to fill this gap, we consider a series of opinion optimization problems *without the full knowledge of innate opinions*, and tackling the budgeted optimization problems, where we can query the innate opinions of only a limited number of nodes. As objective functions to be minimized, following [Chen *et al.*, 2018; Musco *et al.*, 2018; Zhu *et al.*, 2021; Cinus *et al.*, 2023], we consider polarization, disagreement, and the sum of the two in both directed and undirected networks, for a total of six different objectives. As an intervention mechanism, following [Cinus *et al.*, 2023], we consider re-weighting the relative importance of the accounts that each user follows.

A crucial step our solution is to effectively reconstruct the innate opinions of nodes that we did not query. [Neumann *et al.*, 2024] studies opinion estimation in the FJ model. Their estimation approach is not directly applicable to our set of opinion optimization problems, which require gradient descent methods [Musco *et al.*, 2018; Cinus *et al.*, 2023]. A (non-trivial) adaptation of their algorithm would require numerous evaluations of both the objective values and the gradients across different solutions, resulting in a significant computational cost. Finally, [Neumann *et al.*, 2024] assumes to have access to an oracle for the expressed opinions (which we do not have) and consider only undirected networks, while we study the optimization problems also on directed graphs.

Roadmap. In §2, we provide the necessary background and introduce the six objective functions we consider. In §3, we propose a pipeline that integrates innate opinion reconstruction with opinion optimization and theoretically characterize the objectives in terms of convexity and Lipschitz continuity, deriving their solvability and an upper bound on optimization error based on opinion reconstruction error. In §4, we present methods for selecting nodes to query their innate opinions and reconstructing the opinions of unqueried nodes. For node selection, we use heuristics based on centrality measures. For opinion reconstruction, motivated by the strong *homophily* of innate opinions in real-world networks, we apply strategies including label propagation [Zhu and Ghahramani, 2002], graph neural networks [Kipf and Welling, 2017], and graph signal processing [Lorenzo *et al.*, 2018]. Finally, in §5, we present experiments on synthetic and real-world datasets with up to 1.6 million edges. Our key finding is that opinions can be effectively optimized even with limited information about innate opinions, using appropriate combinations of node selection strategies and opinion reconstruction methods.

2 Problem Definition

In this section, we first revisit the *Friedkin–Johnsen (FJ) model* [Friedkin and Johnsen, 1990], followed by six objective functions from the literature, i.e., polarization, disagreement, and their combination, in directed networks [Cinus *et al.*, 2023] and undirected networks [Musco *et al.*, 2018; Chen *et al.*, 2018; Xu *et al.*, 2021]. Finally, we define the problem to be addressed in this paper.

Although the FJ model is typically presented in the undirected case, we here focus on the more general and interesting case of directed graphs, following the treatment of [Cinus *et al.*, 2023]. We thus consider an edge-weighted *directed* graph $G = (V, E)$, with $n = |V|$ nodes and $m = |E|$ edges, where each node $i \in V$ corresponds to a user, and each directed edge $(i, j) \in E$ indicates that i *"follows"* j or, in other words, that j can influence the opinion of i. For each edge $(i, j) \in E$, the edge weight a_{ij} quantifies the strength of influence that user j exerts on user i, for example, based on how frequently content produced by j appears in the social feed of i. We assume that $a_{ij} > 0$ if $(i, j) \in E$ and $a_{ij} = 0$ if $(i, j) \notin E$ and we represent all the weights as a matrix $\mathbf{A}$, i.e., $\mathbf{A}[i, j] = a_{ij}$. In the FJ model, each node $i \in V$ has an *innate opinion* s_i about one topic, which may differ from their *expressed opinion* z_i on social media about the same topic, due to various factors, such as social pressure or fear of judgement. The sets of innate and expressed opinions, for all nodes in the network, are represented by vectors $\mathbf{s} \in \mathbb{R}^n$ and $\mathbf{z} \in \mathbb{R}^n$, respectively. The nodes update their expressed opinions, based on the expressed opinions of their neighbors and their own innate opinions. Specifically, for each node $i \in V$, its expressed opinion z_i at time $t + 1$ is given by the average of the expressed opinions of its neighbors at time t and its own innate opinion, weighted by the strength of their influence. If we denote by $\mathbf{D}^{\text{out}}$ the diagonal matrix whose i-th diagonal entry is the weighted out-degree of node i, i.e., $\mathbf{D}^{\text{out}}[i, i] = \sum_{j \in V} \mathbf{A}[i, j]$, and by $\mathbf{z}^{(t)}$ the vector of the expressed opinions at time t, the opinion-update rule can be written in matrix notation as

$$\mathbf{z}^{(t+1)} = (\mathbf{D}^{\text{out}} + \mathbf{I})^{-1}(\mathbf{A}\mathbf{z}^{(t)} + \mathbf{s}). \quad (1)$$

By iterating Eq. (1) and using the matrix convergence theorems [Burden *et al.*, 2015, Theorem 7.17 and Lemma 7.18], we can find the equilibrium of the system, where the opinions of all nodes have converged to a steady state. Specifically, the equilibrium is given by

$$\mathbf{z}^* = (\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}, \quad (2)$$

where $\mathbf{L} = \mathbf{D}^{\text{out}} - \mathbf{A}$ is the Laplacian matrix of G [Cinus *et al.*, 2023]. Eq. (2) shows that the equilibrium opinions depend only on the innate opinions and the structure of the social network.

Following the literature [Cinus *et al.*, 2023], we assume that the adjacency matrix $\mathbf{A}$ of the directed graph G is row-stochastic, i.e., $\mathbf{A}\mathbf{1} = \mathbf{1}$, where $\mathbf{1}$ denotes the all-ones vector. This assumption allows for a straightforward interpretation that the total amount of influence each node receives sums to 1. In this case, the Laplacian $\mathbf{L}$ is given by $\mathbf{L} = \mathbf{D}^{\text{out}} - \mathbf{A} = \mathbf{I} - \mathbf{A}$. Thus, the equilibrium opinion is written as

$$\mathbf{z}^* = (2\mathbf{I} - \mathbf{A})^{-1}\mathbf{s}.$$

Finally, as in the literature [Musco *et al.*, 2018; Cinus *et al.*, 2023], we assume that all opinions are mean-centered, i.e., $\sum_{u \in V} z_u^* = 0$.

We are now ready to introduce the six objective functions.

Definition 1 (Polarization for directed graphs (P-DIR)). *The polarization for the equilibrium opinion vector* $\mathbf{z}^*$ *is defined to be the deviation of the opinions of nodes from the average opinion [Musco* et al.*, 2018]. As the opinions are mean-centered, the polarization at the equilibrium is defined as* $\sum_{u \in V} {z_u^*}^2 = {\mathbf{z}^*}^\top \mathbf{z}^*$*. Following [Cinus* et al.*, 2023], the objective function of P-Dir is given by*

$$f(\mathbf{s}, \mathbf{L}{=}\mathbf{I}{-}\mathbf{A}) := \mathbf{s}^\top (2\mathbf{I} - \mathbf{A})^{-\top} (2\mathbf{I} - \mathbf{A})^{-1}\mathbf{s}. \quad (3)$$

Definition 2 (Disagreement for directed graphs (D-DIR)). *Following [Cinus* et al.*, 2023], the disagreement for the equilibrium opinion vector* $\mathbf{z}^*$ *is defined to be the degree of difference of neighboring opinions on G given by* $\sum_{(u,v) \in E} a_{uv}(z_u^* - z_v^*)^2 = \frac{1}{2}{\mathbf{z}^*}^\top (\mathbf{I} + \mathbf{D}^{\text{in}} - 2\mathbf{A})\mathbf{z}^*$*, where* $\mathbf{D}^{\text{in}}$ *is the in-degree counterpart of* $\mathbf{D}^{\text{out}}$*. Then, the objective function of D-Dir is*

$$f(\mathbf{s}, \mathbf{L}{=}\mathbf{I}{-}\mathbf{A}) := \frac{1}{2}\mathbf{s}^\top (2\mathbf{I}{-}\mathbf{A})^{-\top} (\mathbf{I}{+}\mathbf{D}^{\text{in}}{-}2\mathbf{A})(2\mathbf{I}{-}\mathbf{A})^{-1}\mathbf{s}. \quad (4)$$

Definition 3 (Polarization plus Disagreement for directed graphs (PD-DIR)). *We define polarization plus disagreement as the sum of polarization* (3) *and disagreement* (4)*, as in [Cinus* et al.*, 2023]:*

$$f(\mathbf{s}, \mathbf{L}{=}\mathbf{I}{-}\mathbf{A}) := \frac{1}{2}\mathbf{s}^\top (2\mathbf{I}{-}\mathbf{A})^{-\top} (2\mathbf{I}{+}\mathbf{D}^{\text{in}}{-}2\mathbf{A})(2\mathbf{I}{-}\mathbf{A})^{-1}\mathbf{s}. \quad (5)$$

The Undirected Case. Next, we consider the undirected version of the three objective functions above. Undirected graphs are useful to model social networks in which each link represents a bidirectional "friendship" relation. The formalization used so far applies straightforwardly to the undirected case by considering any undirected edge $\{u, v\}$ as the two directed edges (u, v) and (v, u). In the undirected case, we can use the symmetry of $\mathbf{A}$ and $\mathbf{L}$ to simplify the notation. Following [Musco *et al.*, 2018], the objective functions for undirected graphs are then defined as follows:

Definition 4 (Polarization for undirected graphs (P-UNDIR)).

$$f(\mathbf{s}, \mathbf{L}) := \mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-2}\mathbf{s}. \quad (6)$$

Definition 5 (Disagreement for undirected graphs (D-UNDIR)).

$$f(\mathbf{s}, \mathbf{L}) := \mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-1}\mathbf{L}(\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}. \quad (7)$$

Definition 6 (Polarization plus Disagreement for undirected graphs (PD-UNDIR)).

$$f(\mathbf{s}, \mathbf{L}) := \mathbf{s}^\top (\mathbf{I} + \mathbf{L})^{-1}\mathbf{s}. \quad (8)$$

2.1 Our Problem

Given a graph G and influence weights $\mathbf{A}$ along its edges, our goal is to adjust the edge weights $\mathbf{A}$ so as to minimize one of the six objective functions above. As discussed in Introduction, our intervention corresponds to re-weighting the relative importance of the accounts that each user follows, so as to calibrate the frequency with which the contents produced by various accounts are shown to the user. Earlier works have studied similar tasks on undirected [Chen *et al.*, 2018; Musco *et al.*, 2018] and directed graphs [Cinus *et al.*, 2023]. However, the problem we consider is much more complex, as we assume that *we have no prior knowledge of the innate opinions* $\mathbf{s}$, and instead, we are given a *budget* $b \in \mathbb{Z}_{>0}$ that represents the number of nodes we can query their innate opinions. We assume that if we query the innate opinion s_v for $v \in V$, we can obtain the exact value of s_v.

Following [Cinus *et al.*, 2023], we restrict the feasible set of solutions to adjacency matrices where the set of edges is a subset of the edges in the input graph and the out-degree of each node is preserved. By doing so, we are asked to use only pre-existing links and preserve the total engagement of each user in the social network. Formally, given the adjacency matrix $\mathbf{A}$ for a directed graph G, we define the convex set of feasible solutions as follows:

$$\mathcal{C}(\mathbf{A}) = \{\mathbf{X} \in \mathbb{R}_{\geq 0}^{n \times n} \mid \mathbf{X}\mathbf{1} = \mathbf{A}\mathbf{1}, \mathbf{A}[i,j] = 0 \implies \mathbf{X}[i,j] = 0\}.$$

In the case of undirected graphs, the feasible set is

$$\mathcal{C}(\mathbf{L}) = \{\mathbf{X} \in \mathcal{L}^n \mid \text{Tr}(\mathbf{X}) = \text{Tr}(\mathbf{L}),\ \mathbf{L}[i,j] = 0 \Rightarrow \mathbf{X}[i,j] = 0\}$$

where $\mathcal{L}^n$ is the set of Laplacians for graphs with n nodes. Here we adopt the approach of [Musco *et al.*, 2018], where the weighted degree of the nodes is not necessarily preserved.

We are now ready to define the problem.

Problem 1. *We are given an edge-weighted directed (resp. undirected) graph* $G = (V, E)$ *with the unknown innate opinion vector* $\mathbf{s} \in \mathbb{R}^n$*. We are also given a budget* $b \in \mathbb{Z}_{>0}$*. The goal is to find a new adjacency matrix* $\mathbf{A}^* \in \mathcal{C}(\mathbf{A})$ *(resp. Laplacian* $\mathbf{L}^* \in \mathcal{C}(\mathbf{L})$*) that minimizes the specified objective function among Eqs.* (3)–(5) *(resp. Eps.* (6)–(8)*) defined for the unknown innate opinion vector* $\mathbf{s}$*, under the assumption that we can query the innate opinions of at most* b *nodes.*

3 Characterization

Given the complexity of our problem, we propose a framework based on three steps (which we will detail in Section 4): (1) select b nodes and observe their innate opinions, (2) reconstruct the innate opinions of all nodes, using the b observed opinions and the network structure, and (3) optimize the objective function with the reconstructed opinions.

In this section, to characterize the importance of steps (1) and (2), we quantify how the reconstruction error of innate opinions affects the final quality of solutions. The main result of our analysis is presented in Theorem 1, revealing how the reconstruction error together with the Lipschitz constant of the objective function bound the error of optimization. To complete the analysis, we then derive the Lipschitz constant of each objective function with respect to the opinion vectors.

All proofs are deferred to the Supplementary Material[1].

[1] https://github.com/FedericoCinus/Query-MinPD

Objective	Convex	Gradient w.r.t. $\mathbf{s}$	K
P-DIR	✗	$2(2\mathbf{I}-\mathbf{A})^{-\top}(2\mathbf{I}-\mathbf{A})^{-1}\mathbf{s}$	2
D-DIR	✗	$(2\mathbf{I}-\mathbf{A})^{-\top}(\mathbf{I}+\mathbf{D}^{\text{in}}-2\mathbf{A})(2\mathbf{I}-\mathbf{A})^{-1}\mathbf{s}$	$1+\Delta(G)$
PD-DIR	✗	$(2\mathbf{I}-\mathbf{A})^{-\top}(2\mathbf{I}+\mathbf{D}^{\text{in}}-2\mathbf{A})(2\mathbf{I}-\mathbf{A})^{-1}\mathbf{s}$	$2+\Delta(G)$
P-UNDIR	✗	$2(\mathbf{I}+\mathbf{L})^{-2}\mathbf{s}$	2
D-UNDIR	✗	$2(\mathbf{I}+\mathbf{L})^{-1}\mathbf{L}(\mathbf{I}+\mathbf{L})^{-1}\mathbf{s}$	$2\Delta(G)$
PD-UNDIR	✔	$2(\mathbf{I}+\mathbf{L})^{-1}\mathbf{s}$	2

Table 1: Summary of the six objectives functions, convexity w.r.t. the Laplacian matrix $\mathbf{L}$, gradients w.r.t. the opinion vector $\mathbf{s}$, and their corresponding Lipschitz constants K.

3.1 Hardness

First, we discuss the hardness of the node selection problem for opinion reconstruction and the non-convexity of most objectives. These results build a taxonomy of objective functions related to the FJ model, as summarized in Table 1.

Node Selection for Opinion Reconstruction. In general, selecting an optimal subset of nodes to query, which minimizes the error of some estimate, is a well-known problem in optimal design [Pukelsheim, 2006]. This involves choosing a subset of b sample locations to recover the unknown parameter $\mathbf{s}$. When experiments are selected integrally, as in our problem, standard criteria such as A-optimal design, D-optimal design, and E-optimal design are known to be NP-hard, even in the simplest cases like linear experiments [Madan *et al.*, 2019]. This implies the potential difficulty of node selection in our problem.

Non-Convexity of Objective Functions. We next state the non-convexity of our objectives, except for PD-UNDIR being convex, related to the solvability of the problems.

Proposition 1. *The objectives* (3)–(7) *are not matrix-convex.*

Proposition 2. *The objective* (8) *is matrix-convex.*

3.2 Error Analysis

Let $\mathbf{s}$ and $\hat{\mathbf{s}}$ be the true innate opinions and the reconstructed innate opinions, respectively. We define the *reconstruction error* of $\hat{\mathbf{s}}$ to be $\|\mathbf{s}-\hat{\mathbf{s}}\|$. In general, a function f on $\mathbb{R}^n$ is said to be *K-Lipschitz continuous* if there exists a constant K such that for all $\mathbf{x}, \mathbf{y}$ in its domain, $|f(\mathbf{x}) - f(\mathbf{y})| \leq K\|\mathbf{x}-\mathbf{y}\|$, where the parameter K is called a *Lipschitz constant*. The following theorem shows how the reconstruction error together with the Lipschitz constant of the objective function bound the error of optimization.

Theorem 1. *Let* $\mathbf{L}^*$ *and* $\hat{\mathbf{L}}$ *be optimal solutions, which minimize* $f(\mathbf{s}, \mathbf{L})$ *and* $f(\hat{\mathbf{s}}, \mathbf{L})$ *in Eqs.* (3)–(8)*, respectively, over the feasible set defined. Suppose that* f *is* K*-Lipschitz continuous with respect to the first argument. Then, it holds that* $f(\mathbf{s}, \hat{\mathbf{L}}) - f(\mathbf{s}, \mathbf{L}^*) \leq 2K\|\mathbf{s}-\hat{\mathbf{s}}\|$.

This implies the following multiplicative approximation:

Corollary 1. *Under the assumptions of Theorem 1 together with* $f(\mathbf{s}, \mathbf{L}^*) \neq 0$*, it holds that*

$$\frac{f(\mathbf{s}, \hat{\mathbf{L}})}{f(\mathbf{s}, \mathbf{L}^*)} \leq 1 + \frac{2K\|\mathbf{s}-\hat{\mathbf{s}}\|}{f(\mathbf{s}, \mathbf{L}^*)}.$$

The right-hand-side represents an approximation ratio of $\hat{\mathbf{L}}$ with respect to the true innate opinions $\mathbf{s}$. From Proposition 1, we know that most of the objectives are non-convex, and for those objectives, no exact algorithm is known in the literature. However, for the objective PD-UNDIR, we can apply Corollary 1, due to the convexity of the objective (Proposition 2).

Finally, we provide Lipschitz constants for each objective function in Eqs. (3)–(8). Along with Theorem 1 and Corollary 1, this establishes an upper bound on the optimization error associated with the reconstruction error.

Proposition 3. *The objective functions in Eqs.* (3)–(8) *are Lipschitz continuous on the space* $\mathbb{R}^n$ *with the following Lipschitz constants: for* P-DIR*,* $K = 2$*; for* D-DIR*,* $K = 1 + \Delta(G)$*; for* PD-DIR*,* $K = 2 + \Delta(G)$*; for* P-UNDIR*,* $K = 2$*; for* D-UNDIR*,* $K = 2\Delta(G)$*; and for* PD-UNDIR*,* $K = 2$*; where* $\Delta(G)$ *is the maximum (in-)degree of* G *(directed or undirected).*

The proof proceeds as follows: First, we derive the gradient of the objective function with respect to the opinion vector to express the Lipschitz constant in its infinitesimal form. Next, we relate the Lipschitz constant to the spectral norm of the gradient matrix, which can then be bounded.

4 Methods

Here, we outline the three steps of our proposed framework, with detailed explanations, pseudocodes, and time complexity analyses provided in the Supplementary Material.

4.1 Node Selection Strategies

To mitigate computational bottlenecks, we employ three heuristic approaches based on centrality measures, which select the top-b nodes with respect to the following: *Degree Centrality*, which represents the sum of a node's in-degree and out-degree; *Closeness Centrality* [Bavelas, 1950], which is inversely proportional to the total shortest path distances from a node to all others; and *PageRank* [Page *et al.*, 1999] with a damping factor of 0.85.

As a *baseline*, we also include a random uniform strategy for node selection, which selects b nodes uniformly at random without leveraging any structural properties of the network.

4.2 Opinion Reconstruction Methods

Based on the innate opinions of the b selected nodes, we aim to reconstruct the innate opinions of the remaining nodes as accurately as possible. To this end, we consider three types of reconstruction methods: Label Propagation-based, GNN-based, and Graph Signal Processing-based algorithms.

Label Propagation (LP). We extended LP [Zhu and Ghahramani, 2002] to handle continuous values by initializing all node values to zero, setting selected nodes to their true values, and iteratively updating the remaining nodes to the average of their neighbors over a fixed number of iterations.

Graph Neural Networks (GNN). We use a GCN [Kipf and Welling, 2017] to reconstruct unknown opinions by propagating known opinions from selected nodes, initializing node features with these values (or zero if unavailable), and training with an MSE loss.

Dataset	**Statistics**		
	n	m	Opinion Distribution
Referendum	2,479	154,831	
Brexit	7,281	530,607	
VaxNoVax	11,632	1,599,220	
directed/moreno-highschool	70	366	
directed/wiki-talk-ht	82	154	
directed/moreno-innovation	108	510	
directed/moreno-oz	216	2,667	
directed/librec-filmtrust-trust	425	1,363	
directed/dnc-temporalGraph	949	4,029	
directed/librec-ciaodvd-trust	1,309	27,239	
directed/moreno-health	2,298	11,999	
undirected/ucidata-zachary	34	78	
undirected/moreno-beach	43	336	
undirected/moreno-train	64	243	
undirected/out.mit	96	2,539	
undirected/dimacs10-football	115	613	

Table 2: Statistics of our datasets. We plot the distribution of the standardized innate opinions (i.e., average opinion is zero). The first three datasets contain direct follow networks on $\mathbb{X}$ and real opinions. The other networks obtained from KONECT [Kunegis, 2013] are associated with opinions sampled from Gaussian distributions.

Graph Signal Processing (GSP). We use GSP to reconstruct opinions, assuming the graph signal $f : V \to \mathbb{R}$ (opinions in our case) is bandlimited and can be expressed as a linear combination of a limited number of the Laplacian eigenvectors. Using mild assumptions for perfect recovery and noise from [Lorenzo *et al.*, 2018], we apply the best linear unbiased estimator (BLUE) [Winer *et al.*, 1971] to infer opinions from sampled nodes.

4.3 Optimization

In directed graphs, the objective functions (3)–(5) are non-convex, and no approximation algorithm is known in the literature. To address these computational challenges, we employ a constrained gradient-descent approach as in [Cinus *et al.*, 2023]. In the case of undirected graphs, we formulated the problem with PD-UNDIR as a semidefinite programming (SDP) with CVX as in [Musco *et al.*, 2018]. For the other objectives and constraint, we apply the projection steps in [Cinus *et al.*, 2023]. We compute a local minimum, using the reconstructed opinions as input.

5 Experimental Evaluation

In this section, we asses our framework and the impact of the reconstruction error on solution quality, compared to the ground-truth opinions. We present experiments on 16 networks with up to 1.6 million edges, considering both real opinions and synthetic opinions generated with varying distributions and polarization levels. Dataset statistics are provided in Table 2. Additional experiments on opinion distribution variations, runtime analysis, sensitivity studies, and baselines comparison are provided in the Supplementary Material.

Evaluation. We evaluate the accuracy of an algorithm as follows. Let $\mathbf{L}_{\mathrm{ALG}}$ and $\mathbf{L}^*_{\mathrm{ALG}}$ be the solutions obtained by the algorithm with the reconstructed innate opinions and the true innate opinions, respectively. Note that we can only compute $\mathbf{L}_{\mathrm{ALG}}$ when the true innate opinions are unavailable, but $\mathbf{L}^*_{\mathrm{ALG}}$ is also computed in experiments for evaluation. Then the quality of solution is measured by $\frac{f(\mathbf{s},\mathbf{L}_{\mathrm{ALG}})}{f(\mathbf{s},\mathbf{L}^*_{\mathrm{ALG}})}$, which we refer to as the *multiplicative error*. As the denominator is (even beyond) the best possible achievable by the algorithm, the multiplicative error can be interpreted as a measure of how far the solution's quality deviates from this benchmark. Therefore, the smaller the multiplicative error is, the better the solution's quality.

Reproducibility. Experimental settings are in the Supplementary Material. Our code is available at https://github.com/FedericoCinus/Query-MinPD.

5.1 Results

Real-world Datasets (Directed Graphs). Results are presented in Table 3. We use the three real-world datasets from $\mathbb{X}$, as shown in Table 2, and validate performance across the three objectives for directed graphs in Table 1. Our framework tests three reconstruction methodologies – LP, GNN, and GSP, from Section 4, using reconstructed opinions from $b = 0.2|V|$ selected nodes, with this choice validated in Figure 2. Nodes for opinion reconstruction were selected based on Degree Centrality. Other selections of node sizes and strategies are tested in subsequent experiments.

The results show that LP, despite being the most computationally efficient, consistently achieves the lowest errors across all objectives and datasets, with multiplicative errors ranging from 1.16 to 2.08. For the P-DIR objective, LP reduces the error by up to 1.13 compared to GSP, even though this strategy has been optimized for undirected graphs. For the D-DIR objective, the maximum difference between the multiplicative errors reduces to 0.65, but it is even greater than that in PD-DIR (with a maximum difference of 0.42). The general trend suggests that the D-DIR objective appears to be the most challenging to optimize, with a maximum multiplicative error exceeding 2. This could be due to a dependence between network size and the error bound, as indicated by Proposition 3, where the function f is shown to be Lipschitz continuous with a constant related to $\Delta(G)$.

Semi-Synthetic Datasets (Directed Graphs). Results are presented in Table 4. We consider the 8 real-world directed networks in Table 2 and a polarized opinion distribution that reflects community structures.

In general, our framework yields multiplicative errors below 2. The GNN reconstruction methodology consistently achieves multiplicative errors below this value. Nevertheless, LP, while the fastest method, provides the lowest multiplicative errors, except for "ciaodvd-trust" and "wiki talk ht" networks, where it consistently shows higher multiplicative errors across all three objectives. P-DIR, at this scale, consistently exhibits the highest errors across all methods and networks. The maximum multiplicative errors for the least performing methods in this objective function reach up to 2.74.

Objective	Rec Method / Network	Multiplicative Error GNN	GSP	LP
P-DIR	Referendum	1.33	2.35	**1.22**
	Brexit	2.09	2.32	**1.76**
	VaxNoVax	1.85	2.36	**1.41**
D-DIR	Referendum	1.99	2.07	**1.42**
	Brexit	2.66	2.53	**2.08**
	VaxNoVax	2.16	1.96	**1.44**
PD-DIR	Referendum	1.19	1.58	**1.16**
	Brexit	1.31	1.33	**1.19**
	VaxNoVax	1.33	1.39	**1.16**

Table 3: Multiplicative errors for the three objectives (D-DIR, P-DIR, PD-DIR) for 3 real-world directed graphs with different sizes (n). Opinions are derived from the average stance of tweets users retweeted.

Similar trends are observed with uniformly distributed opinions (Table 8 in the Supplementary Material): LP outperforms the other methods except for "wiki talk ht" and "dnc-temporal" networks.

Semi-Synthetic Datasets (Undirected Graphs). Results are presented in Table 5. We consider the 5 real-world undirected networks in Table 2 with a polarized opinion distribution that reflects community structures. Network sizes are limited to approximately 100 nodes to avoid the computational bottleneck inherent in the SDP approach for finding an optimal solution. We test the performance of our framework in minimization of PD-UNDIR. This problem is well-studied in the literature and includes a standard projection step onto the set of SD matrices [Musco *et al.*, 2018]. For P-UNDIR and D-UNDIR, no projection step is known in the literature.

Multiplicative errors show greater variability across different graphs in undirected settings compared to directed ones. This would be because the current optimization problem (PD-UNDIR minimization) allows the algorithm to find a global optimum, resulting in a relatively small value of the denominator in the multiplicative error calculation.

Bounds on the multiplicative error are presented as averages, following Corollary 1 and Proposition 3. These bounds are sensitive to the numerical value of the global minimum of the objective. When the minimum is quite small, it can lead to very large bounds that are not practically useful. This occurred in 2 out of 5 instances in our experiments, specifically with the "beach" and "mit" networks. As a result, the bounds are larger than the actual error, indicating that, in practice, the problem is less challenging than theoretically predicted, and a tighter bound likely exists at this graph size scale. The bounds on the optimization error are proportional to the reconstruction errors, meaning that better performance is closely linked to improved reconstruction accuracy. For instance, the GNN strategy consistently achieves lower multiplicative errors (up to 1.7 times lower) compared to the GSP method. The LP method is comparable to GNN, except for "football" network, where it outperforms the others with a multiplicative error below 2.

Effect of Node Selection Strategies. The results are presented in Table 6. We consider real-world datasets to compare different node selection strategies to select $b = 0.2|V|$ nodes for reconstruction with the LP method.

Objective	Rec Method / Network	Multiplicative Error GNN	GSP	LP
P-DIR	highschool	1.87 ± 0.40	1.97 ± 0.30	$\mathbf{1.68 \pm 0.28}$
	wiki talk ht	1.28 ± 0.18	$\mathbf{1.25 \pm 0.13}$	1.34 ± 0.27
	innovation	1.98 ± 0.22	2.06 ± 0.25	$\mathbf{1.80 \pm 0.21}$
	oz	1.83 ± 0.17	2.74 ± 0.22	$\mathbf{1.78 \pm 0.17}$
	film-trust	1.28 ± 0.06	1.47 ± 0.09	$\mathbf{1.27 \pm 0.06}$
	dnc-temporal	1.30 ± 0.14	1.36 ± 0.08	$\mathbf{1.25 \pm 0.14}$
	ciaodvd-trust	$\mathbf{1.56 \pm 0.07}$	2.50 ± 0.14	2.54 ± 0.14
	health	1.75 ± 0.07	1.92 ± 0.07	$\mathbf{1.57 \pm 0.06}$
D-DIR	highschool	1.48 ± 0.20	1.48 ± 0.17	$\mathbf{1.42 \pm 0.15}$
	wiki talk ht	1.24 ± 0.18	$\mathbf{1.21 \pm 0.14}$	1.34 ± 0.17
	innovation	1.69 ± 0.17	1.71 ± 0.15	$\mathbf{1.51 \pm 0.15}$
	oz	1.57 ± 0.15	1.93 ± 0.12	$\mathbf{1.47 \pm 0.14}$
	film-trust	1.23 ± 0.05	1.24 ± 0.06	$\mathbf{1.22 \pm 0.05}$
	dnc-temporal	1.30 ± 0.13	$\mathbf{1.25 \pm 0.11}$	1.35 ± 0.11
	ciaodvd-trust	$\mathbf{1.45 \pm 0.04}$	1.53 ± 0.06	1.52 ± 0.03
	health	1.45 ± 0.06	1.50 ± 0.05	$\mathbf{1.36 \pm 0.04}$
PD-DIR	highschool	1.43 ± 0.15	1.42 ± 0.12	$\mathbf{1.32 \pm 0.09}$
	wiki talk ht	1.23 ± 0.18	$\mathbf{1.20 \pm 0.12}$	1.28 ± 0.23
	innovation	1.43 ± 0.11	1.39 ± 0.12	$\mathbf{1.35 \pm 0.10}$
	oz	$\mathbf{1.39 \pm 0.09}$	1.59 ± 0.09	1.40 ± 0.09
	film-trust	1.16 ± 0.04	1.24 ± 0.05	$\mathbf{1.15 \pm 0.03}$
	dnc-temporal	1.19 ± 0.12	1.22 ± 0.08	$\mathbf{1.16 \pm 0.12}$
	ciaodvd-trust	$\mathbf{1.19 \pm 0.05}$	1.46 ± 0.07	1.56 ± 0.05
	health	1.37 ± 0.02	1.37 ± 0.02	$\mathbf{1.29 \pm 0.02}$

Table 4: Average multiplicative errors for the three objectives (D-DIR, P-DIR, PD-DIR) for 8 real-world directed graphs with different sizes ($|V|$). Opinions are Gaussian distributed around a mean corresponding to one of the assigned communities. This opinions are reconstructed with $b = 0.20|V|$ sampled nodes.

Rec Method / Network	Multiplicative Error (Bound) GNN	GSP	LP
zachary	$\mathbf{1.69 \pm 0.29}$ (7)	2.40 ± 0.36 (9)	1.70 ± 0.24 (8)
beach	$\mathbf{2.22 \pm 0.16}$ (67)	4.03 ± 0.70 (94)	2.27 ± 0.20 (70)
train	$\mathbf{1.46 \pm 0.13}$ (4)	2.69 ± 0.41 (7)	1.82 ± 0.26 (6)
mit	$\mathbf{1.07 \pm 0.11}$ (9,540)	$\mathbf{1.07 \pm 0.11}$ (12,067)	1.09 ± 0.15 (9,071)
football	2.11 ± 0.53 (3)	2.57 ± 0.72 (4)	$\mathbf{1.93 \pm 0.44}$ (3)

Table 5: Average multiplicative errors in PD-UNDIR minimization in undirected graphs.

On average, PageRank proves to be the most effective strategy for selecting nodes, while Degree Centrality shows consistently strong performance compared to the random strategy. For P-DIR, using Degree Centrality and PageRank can reduce the multiplicative error by up to 0.2 and 0.24, respectively, compared to random selection. For D-DIR, using Degree Centrality can reduce the multiplicative error by up to 0.3 compared to random selection; 0.33 for PageRank. For PD-DIR, using Degree Centrality can reduce the multiplicative error by up to 0.12 compared to random selection; 0.15 for PageRank. Closeness Centrality yields comparable results, although it performs less consistently, particularly on the "Referendum" network. Results for real directed networks with synthetic opinions are presented in Table 10 in the Supplementary Material. These results are consistent, showing the superiority of Degree Centrality and PageRank, except for "ciaodvd trust" network where the random strategy outperforms the others. It is worth noting that selecting nodes uniformly at random tends to cover diverse parts of the network, and given the fact that the innate opinions in real-world networks have a strong homophily, the random strategy

Objective	Sel Method / Network	Multiplicative Error / Closeness centrality	Degree	PageRank	Random
P-DIR	Referendum	1.44	1.22	**1.21**	1.27
	Brexit	**1.74**	1.76	1.75	1.87
	VaxNoVax	**1.37**	1.41	**1.37**	1.61
D-DIR	Referendum	1.57	**1.42**	1.44	1.52
	Brexit	2.06	2.08	**2.04**	2.35
	VaxNoVax	**1.41**	1.44	**1.41**	1.74
PD-DIR	Referendum	1.23	**1.16**	1.17	1.22
	Brexit	**1.18**	1.19	**1.18**	1.28
	VaxNoVax	1.14	1.16	**1.13**	1.28

Table 6: Multiplicative errors for the three objectives (P-DIR, D-DIR, PD-DIR) for different node selection strategies in real-world datasets.

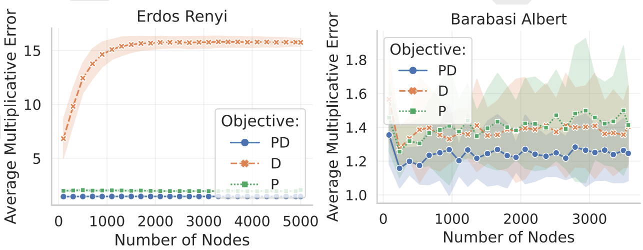

Figure 1: Average multiplicative errors vs. number of nodes in (left) Erdős Rényi graph with $p = 0.25$, and polarized distribution of opinions; (right) Barabási Albert graph with $m = 5$, and polarized distribution of opinions.

would be reasonable and sometimes better than sophisticated ones. Despite the strong performance of PageRank, Degree Centrality strikes a better balance among quality, consistency, and computational efficiency. As such, we use Degree Centrality as the default strategy in subsequent experiments.

Effect of Network Size. The results are depicted in Figure 1. We consider two synthetic directed networks with sizes ranging from 100 to 5,000 nodes. We create a polarized opinion distribution reflecting communities and test the capabilities of our framework using the LP strategy. We measure the performance in minimizing the three objectives, P-DIR, D-DIR, and PD-DIR, with respect to the number of nodes. We use Degree Centrality to select 20% of nodes in each instance for opinion reconstruction.

Except for D-DIR, multiplicative errors are generally constant, with higher volatility observed in the Barabási-Albert graph. At this range of network sizes, D-DIR shows an increasing error with respect to the number of nodes, followed by a plateau, suggesting that non-constant error is introduced by some network structures. These observations align with our upper bounds and underscore the importance of characterizing the inter-dependence between networks and the objectives. This is the first necessary step toward understanding the minimization of such known objectives in an unknown opinion setting, necessitating further analysis and experiments.

Effect of the number of selected nodes. The results are depicted in Figure 2. We consider the "Referendum" dataset to compare different node selection sizes, ranging from 100 up to the size of the network (2,479 nodes).

As expected, the multiplicative error in each objective decreases as the number of selected nodes increases, but dif-

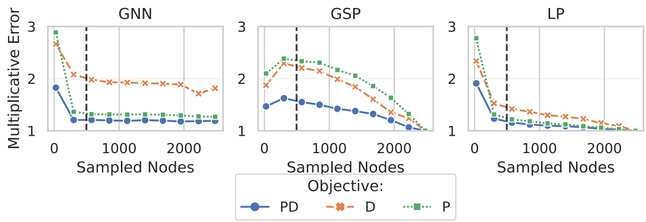

Figure 2: Multiplicative error vs. number of sampled nodes in the Referedum dataset.

ferent patterns emerge. D-DIR shows the slowest rate of decrease compared to the other objectives across reconstruction strategies. In particular, the GNN strategy exhibits a plateau in the error curve, while LP displays a monotonically decreasing behavior. This is why LP has been chosen as the main reconstruction strategy in other experiments, in addition to its computational efficiency. The significant drop in multiplicative errors occurs between 15–20% of the node size. The 20% threshold, indicated in black, represents the selected node size used in all other experiments. In GSP, the number of selected nodes is linked to the number of frequencies, which is always smaller than the number of selected nodes. A sensitivity analysis is presented in Figure 3 in the Supplementary Material.

6 Conclusions

This paper contributes to the literature on opinion optimization in social networks under the Friedkin–Johnsen (FJ) model, which assumes innate opinions are fully known. We address the novel problem of opinion optimization under a budget constraint, where the goal is to minimize polarization and/or disagreement by querying a limited number of nodes for their innate opinions. To tackle this, we propose a framework integrating node selection, opinion reconstruction, and optimization, systematically evaluating alternative strategies for each component and identifying the most effective approaches. Our results demonstrate the framework's practicality, achieving multiplicative errors consistently below 2 and as low as 1.1. Additionally, our error propagation analysis quantifies how reconstruction errors in innate opinions impact the quality of final solutions, offering guidelines for researchers and practitioners, particularly for objectives involving disagreement, whose bounds scale with network size.

Although our experiments scale to networks with up to 1.6 million edges, larger real-world networks require more scalable methods. Techniques like GraphSAGE [Hamilton *et al.*, 2017] could address this challenge, while active learning strategies adapted to graph structures may improve node selection efficiency. Robustness in heterophilic networks, where dissimilar nodes connect, remains an open challenge, and developing heuristics for such cases is a key avenue for future work. From a theoretical perspective, tighter bounds leveraging network structures that constrain Laplacian eigenvalues could provide stronger guarantees for optimization performance. Finally, real-world deployment must account for platform-specific constraints, which can be addressed by adapting the solution space accordingly.

Ethics Statement

Societal Impact. Our work proposes a framework for opinion optimization under the Friedkin–Johnsen model to mitigate polarization while addressing the challenge of incomplete information. By leveraging the interplay between network structure, innate opinions, and opinion dynamics, this research contributes to efforts aimed at designing interventions that reduce polarization and promote healthier discourse in online environments.

Ethical Aspects. Although this work is grounded in theoretical models and analysis, it carries significant ethical implications. Our experiments are conducted using anonymized datasets to ensure that no personally identifiable information is used or exposed. Nonetheless, methods for inferring opinions inherently carry risks, such as enabling the targeting of individuals based on their inferred opinions. Furthermore, the framework and analysis presented here could, in principle, be repurposed to maximize polarization rather than mitigating it, simply by changing the sign of the objective function.

We acknowledge these risks and emphasize that our work aims to advance ethical and privacy-conscious machine learning. By prioritizing the minimization of polarization and promoting responsible approaches to opinion dynamics, we seek to contribute to societal challenges constructively. Balancing these potential harms, we hope our research sets an example for the development of fairer, more transparent, and ethically grounded interventions in opinion optimization.

Acknowledgments

The work of Yuko Kuroki is supported by Japan Science and Technology Agency (JST) Strategic Basic Research Programs PRESTO "R&D Process Innovation by AI and Robotics: Technical Foundations and Practical Applications" grant number JPMJPR24T2. The authors would like to thank the anonymous reviewers for their insightful comments and useful feedback.

References

[Abebe *et al.*, 2021] Rediet Abebe, T-H Hubert Chan, Jon Kleinberg, Zhibin Liang, David Parkes, Mauro Sozio, and Charalampos E. Tsourakakis. Opinion dynamics optimization by varying susceptibility to persuasion via non-convex local search. *ACM Transactions on Knowledge Discovery from Data*, 16(2):1–34, 2021.

[Aslay *et al.*, 2018] Çigdem Aslay, Antonis Matakos, Esther Galbrun, and Aristides Gionis. Maximizing the diversity of exposure in a social network. In *Proceedings of the 2018 IEEE International Conference on Data Mining*, pages 863–868, 2018.

[Auletta *et al.*, 2016] Vincenzo Auletta, Ioannis Caragiannis, Diodato Ferraioli, Clemente Galdi, and Giuseppe Persiano. Generalized discrete preference games. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 53–59, 2016.

[Bavelas, 1950] Alex Bavelas. Communication patterns in task-oriented groups. *The Journal of the Acoustical Society of America*, 22(6):725–730, 1950.

[Burden *et al.*, 2015] Richard L. Burden, J. Douglas Faires, and Annette M. Burden. *Numerical Analysis*. Cengage Learning, 2015.

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

[Chierichetti *et al.*, 2013] Flavio Chierichetti, Jon Kleinberg, and Sigal Oren. On discrete preferences and coordination. In *Proceedings of the 14th ACM Conference on Electronic Commerce*, pages 233–250, 2013.

[Cinus *et al.*, 2022] Federico Cinus, Marco Minici, Corrado Monti, and Francesco Bonchi. The effect of people recommenders on echo chambers and polarization. In *Proceedings of the 16th International AAAI Conference on Web and Social Media*, pages 90–101, 2022.

[Cinus *et al.*, 2023] Federico Cinus, Aristides Gionis, and Francesco Bonchi. Rebalancing social feed to minimize polarization and disagreement. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*, pages 369–378, 2023.

[Del Vicario *et al.*, 2017] Michela Del Vicario, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. Modeling confirmation bias and polarization. *Scientific Reports*, 7(1):40391, 2017.

[Ferraioli and Ventre, 2017] Diodato Ferraioli and Carmine Ventre. Social pressure in opinion games. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 3661–3667, 2017.

[Fotakis *et al.*, 2016] Dimitris Fotakis, Dimitris Palyvos-Giannas, and Stratis Skoulakis. Opinion dynamics with local interactions. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pages 279–285, 2016.

[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.

[Garimella *et al.*, 2017] Kiran Garimella, Aristides Gionis, Nikos Parotsidis, and Nikolaj Tatti. Balancing information exposure in social networks. *Proceedings of Advances in Neural Information Processing Systems*, pages 4663–4671, 2017.

[Garimella *et al.*, 2018] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Reducing controversy by connecting opposing views. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, pages 5249–5253, 2018.

[Hamilton *et al.*, 2017] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. *Proceedings of Advances in Neural Information Processing Systems*, pages 1024–1034, 2017.

[Hartman *et al.*, 2022] Rachel Hartman, Will Blakey, Jake Womick, Chris Bail, Eli J. Finkel, Hahrie Han, John Sarrouf, Juliana Schroeder, Paschal Sheeran, Jay J. Van Bavel, Robb Willer, and Kurt Gray. Interventions to reduce partisan animosity. *Nature Human Behaviour*, 6(9):1194–1205, 2022.

[Kernighan and Lin, 1970] Brian W. Kernighan and Shen Lin. An efficient heuristic procedure for partitioning graphs. *The Bell System Technical Journal*, 49(2):291–307, 1970.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Proceedings of the 5th International Conference on Learning Representations*, 2017.

[Kunegis, 2013] Jerome Kunegis. Konect: The koblenz network collection. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 1343–1350, 2013.

[Lorenzo *et al.*, 2018] PaoloDi Lorenzo, Sergio Barbarossa, and Paolo Banelli. Sampling and recovery of graph signals. In *Cooperative and Graph Signal Processing: Principles and Applications*, pages 261–282. 2018.

[Madan *et al.*, 2019] Vivek Madan, Mohit Singh, Uthaipon Tantipongpipat, and Weijun Xie. Combinatorial algorithms for optimal design. In *Proceedings of the 32nd Conference on Learning Theory*, pages 2210–2258, 2019.

[Marumo *et al.*, 2021] Naoki Marumo, Atsushi Miyauchi, Akiko Takeda, and Akira Tanaka. A projected gradient method for opinion optimization with limited changes of susceptibility to persuasion. In *Proceedings of the 30th ACM International Conference on Information and Knowledge Management*, pages 1274–1283, 2021.

[Minici *et al.*, 2022] Marco Minici, Federico Cinus, Corrado Monti, Francesco Bonchi, and Giuseppe Manco. Cascade-based echo chamber detection. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pages 1511–1520, 2022.

[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 2018*, pages 369–378, 2018.

[Neumann *et al.*, 2024] Stefan Neumann, Yinhao Dong, and Pan Peng. Sublinear-time opinion estimation in the Friedkin–Johnsen model. In *Proceedings of The ACM Web Conference 2024*, pages 2563–2571, 2024.

[Nikolov *et al.*, 2015] Dimitar Nikolov, Diego F.M. Oliveira, Alessandro Flammini, and Filippo Menczer. Measuring online social bubbles. *PeerJ Computer Science*, 1:e38, 2015.

[O'Donoghue *et al.*, 2016] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068, 2016.

[O'Donoghue, 2021] Brendan O'Donoghue. Operator splitting for a homogeneous embedding of the linear complementarity problem. *SIAM Journal on Optimization*, 31:1999–2023, 2021.

[Page *et al.*, 1999] Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the web. Technical report, Stanford Infolab, 1999.

[Pariser, 2011] Eli Pariser. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin UK, 2011.

[Petersen and Pedersen, 2008] Kaare B. Petersen and Michael S. Pedersen. The matrix cookbook. *Technical University of Denmark*, 7(15):510, 2008.

[Pukelsheim, 2006] Friedrich Pukelsheim. *Optimal Design of Experiments*. Society for Industrial and Applied Mathematics, 2006.

[Quattrociocchi *et al.*, 2016] Walter Quattrociocchi, Antonio Scala, and Cass R. Sunstein. Echo chambers on Facebook. *Harvard Public Law Working Paper Forthcoming*, 2016. Available at SSRN 2795110.

[Tu *et al.*, 2020] Sijing Tu, Çigdem Aslay, and Aristides Gionis. Co-exposure maximization in online social networks. In *Proceedings of Advances in Neural Information Processing Systems*, 2020.

[Tu *et al.*, 2023] Sijing Tu, Stefan Neumann, and Aristides Gionis. Adversaries with limited information in the Friedkin–Johnsen model. In *Proceedings of the 29th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2201–2210, 2023.

[Winer *et al.*, 1971] Ben J. Winer, Donald R. Brown, and Kenneth M. Michels. *Statistical Principles in Experimental Design*, volume 2. McGraw-Hill New York, 1971.

[Xu *et al.*, 2021] Wanyue Xu, Qi Bao, and Zhongzhi Zhang. Fast evaluation for relevant quantities of opinion dynamics. In *Proceedings of The ACM Web Conference 2021*, pages 2037–2045, 2021.

[Xu *et al.*, 2022] Wanyue Xu, Liwang Zhu, Jiale Guan, Zuobai Zhang, and Zhongzhi Zhang. Effects of stubbornness on opinion dynamics. In *Proceedings of the 31st ACM International Conference on Information and Knowledge Management*, pages 2321–2330, 2022.

[Zhou *et al.*, 2024] Tianyi Zhou, Stefan Neumann, Kiran Garimella, and Aristides Gionis. Modeling the impact of timeline algorithms on opinion dynamics using low-rank updates. In *Proceedings of The ACM Web Conference 2024*, pages 2694–2702, 2024.

[Zhu and Ghahramani, 2002] Xiaojin Zhu and Zoubin Ghahramani. Learning from labeled and unlabeled data with label propagation. Technical report, CMU CALD tech report CMU-CALD-02-107, 2002.

[Zhu *et al.*, 2021] Liwang Zhu, Qi Bao, and Zhongzhi Zhang. Minimizing polarization and disagreement in social networks via link recommendation. In *Proceedings of Advances in Neural Information Processing Systems*, pages 2072–2084, 2021.