

Minimizing Polarization and Disagreement in Social Networks

Cameron Musco
Massachusetts Institute of
Technology (MIT)
cnmusco@mit.edu

Christopher Musco
Massachusetts Institute of
Technology (MIT)
cpmusco@mit.edu

Charalampos E. Tsourakakis
Boston University (BU)
ctsourak@bu.edu

ABSTRACT

The rise of social media and online social networks has been a disruptive force in society. Opinions are increasingly shaped by interactions on online social media, and social phenomena including disagreement and polarization are now tightly woven into everyday life. In this work we initiate the study of the following question:

Given n agents, each with its own initial opinion that reflects its core value on a topic, and an opinion dynamics model, what is the structure of a social network that minimizes *disagreement* and *controversy* simultaneously?

This question is central to recommender systems: should a recommender system prefer a link suggestion between two online users with similar mindsets in order to keep disagreement low, or between two users with different opinions in order to expose each to the others viewpoint of the world, and decrease overall levels of polarization and controversy? Such decisions have an important global effect on society [48]. Our contributions include a mathematical formalization of this question as an optimization problem and an exact, time-efficient algorithm. We also prove that there always exists a graph with $O(n/\epsilon^2)$ edges that is a $(1 + \epsilon)$ approximation to the optimum. Our formulation is an instance of optimization over *graph topologies*, see also [6, 11, 45]. Furthermore, for a given graph, we show how to optimize the same objective over the agents' innate opinions in polynomial time. Finally, we perform an empirical study of our proposed methods on synthetic and real-world data that verify their value as mining tools to better understand the trade-off between of disagreement and polarization. We find that there is a lot of space to reduce both controversy and disagreement in real-world networks; for instance, on a Reddit network where users exchange comments on politics, our methods achieve a reduction in controversy and disagreement of the order 6.2×10^4 .

CCS CONCEPTS

• **Mathematics of computing** → **Graph algorithms**;

KEYWORDS

Social Networks; Opinion Dynamics; Polarization; Optimization

ACM Reference Format:

Cameron Musco, Christopher Musco, and Charalampos E. Tsourakakis. 2018. Minimizing Polarization and Disagreement in Social Networks. In *WWW*

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW 2018, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5639-8/18/04.

<https://doi.org/https://doi.org/10.1145/3178876.3186103>

2018: *The 2018 Web Conference, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 10 pages. <https://doi.org/https://doi.org/10.1145/3178876.3186103>

1 INTRODUCTION

The enormous popularity of social media and online social networks has led to fundamental changes in how humans share and form opinions. Several events have stirred fierce debates online, such as the internal memo of James Damore at Google that leaked online and showed how society and users of social media become strongly polarized around controversial issues [12]. Social phenomena such as disagreement and polarization that have existed in human societies for millenia, are now taking place in an online virtual world, with a huge impact on society. Furthermore, opinions are increasingly shaped by social media. For this reason Facebook's recommender systems were recently accused of influencing voters by propagating fake news and Russian ads during the U.S. national elections [46], thus stirring additional controversy about the role of tech giants in society. Google has faced similar accusations about influencing the outcome of elections [41].

From a business perspective, recommender systems aim to maximize revenue – some in the short-term, others in the long-term – by optimizing mathematical objectives related to click-through rate and user engagement. However, given the increasing power of tech-giants like Facebook and Google, a better understanding of the effect of recommended links on society is required [47].

The following facts motivate our work. Valdis Krebs analyzed purchasing trends on Amazon. He found that during the 2008 presidential elections, people who already supported Barack Obama tended to be the same people buying books that painted him in a positive light. Similarly, people who disliked him, bought books that painted him in a negative light [29, Chapter 1]. This bias is known as *confirmation bias* [30], and lies at the root of the spread of conspiracy theories and fake news. Put simply, most humans avoid challenging their opinions. Therefore, a recommender system trained on real-data, whose goal is to maximize revenue and increase user engagement, may naturally end up creating “echo-chambers”. Equivalently, the recommended links minimize the *disagreement* that the user experiences. In the context of social networks, connections between users with similar mindsets are preferred over connections between users with different mindsets.

On the other hand, minimizing disagreement leads to greater *polarization*. To see intuitively why this is the case, consider a topic with two prevalent opinions such as supporting either Democrat or Republican politics. As users connect to users with similar mindsets, two clusters with strong and extreme opinions are formed, leading to greater polarization between the two groups [5]. This polarization has harmful effects on society; for example, reaching political

consensus becomes much harder. This trade-off between *disagreement* and *polarization* motivates our work. We are interested in understanding the structure of a social network that minimizes both disagreement and polarization. We introduce and study two key problems, summarized below. For both problems, we use as our underlying opinion dynamics model the Friedkin-Johnsen model [21] that includes both disagreement and consensus, as it associates each node with an innate opinion and an expressed opinion. For details see Section 2.

• **Minimizing Polarization and Disagreement Over Graphs.**

We initiate the study of the following important question in understanding the global effect of recommended connections in social networks on society. We formalize this problem mathematically, and provide a polynomial time algorithm by proving that our optimization formulation is convex. Our formulation is an instance of optimization over *graph topologies*. For other such optimization problems, see [6, 11, 45].

PROBLEM 1. *Given n agents, each with its own initial opinion that reflects its core value on a topic, and an opinion dynamics model, what is the structure of a connected social network with a given total edge weight that minimizes polarization and disagreement simultaneously?*

We additionally prove that there always exists a graph with at most $O(n/\epsilon^2)$ edges that achieves a $(1 + \epsilon)$ -approximation to the above problem. That is, disagreement and polarization can be minimized by very sparsely connected networks.

• **Minimizing Polarization and Disagreement Over Opinions.**

We also study the following optimization problem. Here, the social network is given and we wish to modify the agents' innate opinions to minimize polarization and disagreement at equilibrium. This problem aims to understand the effect of targeted ads or recommendations designed to influence innate opinions.

PROBLEM 2. *Given a network G on n agents, each with its own initial opinion, an opinion dynamics model, and a budget $\alpha > 0$, how should we change the initial opinions using total opinion mass at most α in order to minimize polarization and disagreement?*

• **Experiments.** We evaluate our methods on synthetic and real data. Our findings indicate that our graph optimization methods result in significantly lower levels of polarization and disagreement on Twitter and Reddit, and confirm that our proposed methods can be used as mining tools to better understand the effects of link recommendation. We also observe experimentally that existing graph topologies are far away from optimizing polarization and disagreement. For instance, our proposed method shows that we can obtain a 6.2×10^4 -fold reduction in polarization and disagreement by optimizing the graph topology of our Reddit dataset.

Roadmap. The paper is organized as follows: Section 2 presents related work. Section 3 presents our algorithmic contributions. Section 4 presents experimental findings on synthetic and real-world data. Finally, Section 5 concludes the paper.

Notation. We use the following notation throughout our paper. Let $G(V, E, w)$ be a weighted connected undirected graph, with $V = [n]$ and $|E| = m$. Let $N(i)$ be the neighborhood of node i and $d(i) = \sum_{j \in N(i)} w_{ij}$ be its degree. Let A be the adjacency matrix, and $L \stackrel{\text{def}}{=} D - A$ be the combinatorial Laplacian. Here $D = \text{diag}(d(1), \dots, d(n))$ is a diagonal matrix with the degrees in its diagonal. Each agent $i \in V$ has an innate opinion $s_i \in [0, 1]$. Let $s = (s_1, \dots, s_n) \in [0, 1]^n$ be the vector of innate opinions. Finally, let $\vec{1}, \vec{0}$ be the all-ones and all-zeros vectors respectively.

2 RELATED WORK

To the best of our knowledge, we are the first to formally define and study the problems suggested in Section 1. In the following we review related work.

Modeling opinion dynamics. Opinion dynamics has been a topic of intense study by political scientists, economists, sociologists, physicists, control systems scientists, and computer scientists, among others. Opinion dynamics model social learning processes. Such models have important applications in understanding political voting (e.g., [1, 8]), viral marketing (e.g., [14]), and various other phenomena that take place in social media (e.g., [40]). There exist *discrete* and *continuous* models. In the former, agents have only two possible opinions, 0 or 1. In the latter, opinions span a continuous range; popular choices in the literature are intervals $[0, 1]$, and $[-1, 1]$. The voter model is a popular discrete model that was originally described by Clifford and Sudbury [10] in the context of a spatial conflict between animals for territory. In general, discrete models apply various interaction mechanisms that may update an agent's opinion. Such mechanisms include randomly adopting the opinion of a connected neighbor or applying a local majority rule [18].

DeGroot introduced a continuous opinion dynamics model in his seminal work on consensus formation [16]. A set of n individuals in society start with initial opinions on a subject. Individual opinions are updated using the average of the neighborhood of a fixed social network. Friedkin and Johnsen [22] extended the DeGroot model to include both disagreement and consensus by mixing each individual's *innate belief* with some weight into the averaging process. We discuss this model in greater detail below. Other continuous models are the Hegselmann-Krause model [28] and the Deffuant-Weisbuch model [49]. Das, Gollapudi, Munagala use a model which combines the continuous and discrete approach to fit opinion data better [14]. Bhawalkar, Gollapudi, Munagala study the scenario where the network and the agents' opinions co-evolve [?].

It is worth mentioning that many aspects of opinion formation have been incorporated into existing models such as opinion leaders, external influences and many others. We refer the interested reader to the excellent survey by Mossel and Tamuz and references therein for more related work on opinion dynamics models [34].

Friedkin-Johnsen opinion dynamics. Each node maintains a persistent internal (or innate) opinion s_i , that remains constant. The node updates its expressed opinion z_i through repeated averaging. More precisely, if $w_{ij} \geq 0$ is the weight on edge $(i, j) \in E$ and $N(i)$ denotes the neighborhood of node i , then in one time step agent i updates its opinion to be the average

$$z_i = \frac{s_i + \sum_{j \in N(i)} w_{ij} z_j}{1 + \sum_{j \in N(i)} w_{ij}}.$$

Let z^* be the equilibrium of this process. The value z_i^* at equilibrium is the *expressed opinion* of node i . It is well-known that the equilibrium z^* is the solution to a linear system of equations:

$$z^* = (I + L)^{-1} s \quad \text{[Friedkin-Johnsen Equilibrium]} \quad (1)$$

Here L is the combinatorial Laplacian of the connection graph G . Notice that $(I + L)$ is always invertible since it is positive definite (in fact, all of its eigenvalues are greater than or equal to 1). The Friedkin-Johnsen model has been used by Bindel et al. as a model for understanding the price of anarchy in society when individuals selfishly update their opinions in order to minimize the stress they experience [4]. Here, stress consists of two terms, the stress due to the fact that node i at equilibrium may express a different opinion than its innate one, and the difference of its expressed opinion and the expressed opinions of its neighbors. Formally, the stress of a node i is defined as $(z_i^* - s_i)^2 + \sum_{j \in N(i)} w_{ij} (z_i^* - z_j^*)^2$. Gionis, Terzi, and Tsaparas use this model to suggest a variation of influence maximization [27].

Polarization. Munson et al. have created a browser widget that measures the bias of users based on the news articles they read, and suggests articles of the opposing view to reduce polarization [35]. Liao and Fu provide a tool that makes users aware of the extremity of their opinion, and how well justified it is given their expertise on the topic [33]. Dandekar et al. define a polarizing opinion formation process that increases a disagreement index at the end of the process, show limitations of existing linear models to capture extreme polarization, and suggest a non-linear model [13]. Flaxman et al. study political “echo-chambers” by examining browsing histories and show several interesting empirical findings with respect to how social networks affect humans’ opinions [20].

Closest to our work lies that of Matakos et al. [?]. In their work, they focus on the ℓ_2 norm of the equilibrium vector z^* under the Friedkin-Johnsen model, which is close to our measure of polarization, introduced in Section 3.1, except that they do not mean center the opinion vector. Given a parameter k , they consider the problem of choosing a set of k nodes such that if their innate opinions are set to zero, the proposed polarization index will be minimized. This problem is NP-hard, and focuses just on polarization, without considering disagreement. Close to our work at a conceptual level also lies the work of Garimella, Morales, Gionis, and Mathioudakis [24], which focuses on Twitter data. Garimella et al. create graphs from Twitter data, propose a new definition of polarization, and detect topics that cause intense debates based on this measure. In follow-up work, they consider the problem of making link recommendations to reduce polarization [25, 26].

Finally, an important aspect of using opinion dynamics with real data is opinion mining. For example, how do we map tweets to opinions, i.e., real numbers? A lot of work in this area has been done by the natural language processing (NLP) community; see the

recent survey by Nakov for more details on sentiment analysis and Twitter opinion mining [36].

Optimizing over graph topologies. Problem 1 (see Section 1) aims to optimize an objective over graph topologies. We review some instances of related graph topology optimization problems, most of which lie outside the graph mining literature, which has focused on edge recommendation rather than global structural results.

In 2009, Daitch, Kelner, and Spielman suggested learning graphs that fit vector points well in order to leverage graph mining techniques for machine learning [11]. They propose using the objective $\|LX\|_F^2$ where L is the graph Laplacian and X is a data matrix representing n points in d dimensions [11]. Their objective minimizes the reconstruction error over all points, where the reconstruction error of a point is the ℓ_2 norm of the difference of the point and the weighted average of its neighbors. Other groups have considered related approaches; see for instance [17, 31].

The central problem in phylogenetics is to learn an evolutionary tree that explains the observed existence of n species. A popular family of techniques are distance-based; find an evolutionary tree (whose leaves correspond to the species) which most closely matches a set of $\binom{n}{2}$ pairwise genetic distances [19]. In machine learning, a wide variety of techniques (e.g., Chow-Liu) have been introduced to learn graphical models from data via optimization problems over graph topologies [32]. Boyd et al. have studied the problem of finding the fastest mixing Markov chain on a graph [6, 45], casting it as an optimization problem over all possible weighted subgraphs of a given graph.

3 PROPOSED METHOD

3.1 Polarization and Disagreement

Following (1), let $z^* = (I + L)^{-1} s$ be the equilibrium vector of opinions according to the Friedkin-Johnsen model, for a social network $G(V, E, w)$, and innate opinions $s : V \rightarrow [0, 1]$.

Disagreement. We define the disagreement $d(u, v)$ of edge (u, v) as the squared difference between the opinions of u, v at equilibrium: $d(u, v) \stackrel{\text{def}}{=} w_{uv} (z_u^* - z_v^*)^2$. We define total disagreement $D_{G,s}$ as:

$$D_{G,s} \stackrel{\text{def}}{=} \sum_{(u,v) \in E} d(u,v). \quad \text{[Disagreement]} \quad (2)$$

Polarization. Intuitively, polarization should measure how opinions at equilibrium deviate from the average. There are many ways to quantify this. We choose the standard definition of variance, i.e., the second moment of the opinions. Specifically, let \bar{z} be the mean-centered equilibrium vector:

$$\bar{z} = z^* - \frac{z^{*T} \mathbf{1}}{n} \mathbf{1}.$$

Then the polarization $P_{G,s}$ is defined to be:

$$P_{G,s} \stackrel{\text{def}}{=} \sum_{u \in V} \bar{z}_u^2 = \bar{z}^T \bar{z} \quad \text{[Polarization]} \quad (3)$$

We now introduce the Polarization-Disagreement index, i.e., the objective we care about.

$$\mathcal{I}_{G,s} \stackrel{\text{def}}{=} P_{G,s} + D_{G,s} \quad [\text{Polarization-Disagreement index}] \quad (4)$$

Two useful propositions and an observation follow.

PROPOSITION 3.1. *The disagreement $D_{G,s}$ satisfies the equation:*

$$D_{G,s} = \sum_{(u,v) \in E} w_{uv} (\bar{z}_u - \bar{z}_v)^2.$$

PROOF. Let $\mu \stackrel{\text{def}}{=} \frac{z^{*T} \mathbf{1}}{n}$ be the mean of z^* , so the i -th coordinate of \bar{z} is $\bar{z}_i = z_i^* - \mu$. Now, observe that $w_{uv} (\bar{z}_u - \bar{z}_v)^2 = w_{uv} (z_u^* - \mu - z_v^* + \mu)^2 = d(u,v)$. The result follows by summing over all edges. \square

Observation 1. *The disagreement $D_{G,s}$ is a quadratic form. Specifically, $D_{G,s} = z^{*T} L z^*$, and by Proposition 3.1, $D_{G,s} = \bar{z}^T L \bar{z}$.*

PROPOSITION 3.2. *Let $\bar{s} = s - \frac{s^T \mathbf{1}}{n} \mathbf{1}$ be the mean-centered innate opinion vector. Then, $\bar{z} = (I + L)^{-1} \bar{s}$.*

PROOF. For any graph, $L \mathbf{1} = 0$. Therefore $(I + L) \mathbf{1} = \mathbf{1}$, or equivalently $\mathbf{1} = (I + L)^{-1} \mathbf{1}$. This implies that $z^{*T} \mathbf{1} = s^T (I + L)^{-1} \mathbf{1} = s^T \mathbf{1}$. By these facts, we obtain that

$$\begin{aligned} \bar{z} &= z^* - \frac{z^{*T} \mathbf{1}}{n} \mathbf{1} = (I + L)^{-1} s - \frac{z^{*T} \mathbf{1}}{n} \mathbf{1} = (I + L)^{-1} s - \frac{s^T \mathbf{1}}{n} \mathbf{1} \\ &= (I + L)^{-1} \left(s - \frac{\mathbf{1}^T s}{n} \mathbf{1} \right) = (I + L)^{-1} \bar{s}. \end{aligned} \quad \square$$

Proposition 3.2 provides an alternative way of computing \bar{z} . We can either find it in the obvious way, i.e., find z^* and then center it around zero. Alternatively, we can first center s around zero, and then obtain \bar{z} as the Friedkin-Johnsen equilibrium when the innate opinion vector is \bar{s} .

Trade-off between polarization and disagreement. Before proceeding into any technical details, we show a simple example that illustrates the trade-off between polarization and disagreement. Suppose there are three agents, two of which have opinion 0, and one has opinion 1 on a certain topic, i.e., $s = [0, 0, 1]$. We wish to recommend one link with weight 1 between these three agents. The recommendation that agrees with human confirmation bias, and therefore does not cause any dissatisfaction to the three agents, is the edge between nodes 1 and 2. The equilibrium opinion vector is the same as s , i.e., $z^* = [0, 0, 1]$. The total disagreement is 0, and the polarization is equal to $(-1/3)^2 + (-1/3)^2 + (2/3)^2 = 0.667$. An alternative choice is to recommend the edge between nodes 1, 3. The equilibrium now is $[1/3, 0, 2/3]$. The total polarization is equal to $0^2 + (-1/3)^2 + (1/3)^2 = 0.222$. The total disagreement is $(2/3 - 1/3)^2 = 0.111$. Therefore, the second recommendation resulted in a better outcome with respect to the sum of polarization and disagreement. By symmetry, edge (2, 3) has the same effect as edge (1, 3). The results are summarized in Table 1.

Recommended link	$P_{G,s}$	$D_{G,s}$	$\mathcal{I}_{G,s}$
(1, 2)	0.667	0	0.667
(1, 3)	0.111	0.222	0.333
(2, 3)	0.111	0.222	0.333

Table 1: Trade-off between polarization and disagreement for three agents with innate opinions $s_1 = s_2 = 0, s_3 = 1$. For details, see Section 3.1.

3.2 Optimizing over Graph Topologies

Using the definitions of Section 3.1, we can now formulate Problem 1 from Section 1 mathematically. The objective is to minimize the sum of two terms, polarization and disagreement. Here, \mathcal{L} is the set of valid combinatorial Laplacians of connected graphs. Observe that the trace of the Laplacian is equal to twice the total edge weight of the corresponding graph.

$$\begin{aligned} \text{Problem 1} \quad & \min_{L \in \mathbb{R}^{n \times n}} \quad \bar{z}^T \bar{z} + \bar{z}^T L \bar{z} \\ & \text{subject to} \quad L \in \mathcal{L} \\ & \quad \quad \quad \text{Tr}(L) = 2m \end{aligned} \quad (5)$$

Note that by (3) and Observation 1, $\bar{z}^T \bar{z} + \bar{z}^T L \bar{z} = D_{G,s} + P_{G,s}$, where G is the weighted graph corresponding to the Laplacian L . Thus (5) is equivalent to minimizing the polarization-disagreement index $\mathcal{I}_{G,s}$ over all graphs G with total edge weight m .

THEOREM 3.3. *The objective $\bar{z}^T \bar{z} + \bar{z}^T L \bar{z}$ is a convex function of the edge weights in the graph G corresponding to the Laplacian L .*

PROOF. Using Proposition 3.2 we rewrite the objective as:

$$\begin{aligned} \bar{z}^T \bar{z} + \bar{z}^T L \bar{z} &= \bar{s}^T (I + L)^{-1} (I + L)^{-1} \bar{s} + \bar{s}^T (I + L)^{-1} L (I + L)^{-1} \bar{s} \\ &= \bar{s}^T (I + L)^{-1} (I + L) (I + L)^{-1} \bar{s} = \bar{s}^T (I + L)^{-1} \bar{s}. \end{aligned}$$

It is known that the function $f(L) = (I + L)^{-1}$ is matrix-convex when $L \in \mathcal{L}$ and hence positive semidefinite [?]. That is, for any $\lambda \in (0, 1)$,

$$\lambda Z_1^{-1} + (1 - \lambda) Z_2^{-1} \geq (\lambda Z_1 + (1 - \lambda) Z_2)^{-1}.$$

This gives that $x^T (I + L)^{-1} x$ is convex for all vectors $x \in \mathbb{R}^n$. \square

Additionally, the set of Laplacians \mathcal{L} which we optimize over is convex (standard fact).

Claim 1. *The set $\mathcal{L} \stackrel{\text{def}}{=} \{L^{n \times n} : L \text{ Laplacian, Tr}(L) = 2m\}$ is convex.*

By Theorem 3.3 and Claim 1 we obtain that Problem 1 is solvable in polynomial time. One may use gradient descent or second order methods, see [7]. In the following, we show how to compute the gradient in a closed form in order to perform gradient descent more efficiently (compared to relying on numerical approximations of it).

Gradient. Let $N = \binom{n}{2}$, and e_i be the i -th standard basis vector in \mathbb{R}^N . We can write $L = B \text{diag}(w) B^T$ where $B \in \mathbb{R}^{n \times N}$ is the sign oriented incidence matrix and $w \in \mathbb{R}^N$ is the vector of edge weights for the graph G corresponding to L . Let b_i be the i -th column of B , $i = 1, \dots, N$. Observe that the graph Laplacian corresponding to

the weight vector $w + \epsilon e_i$, is $L + \epsilon b_i b_i^T$. Hence, if we perturb the i -th coordinate of any weight vector by ϵ , i.e., to $w + \epsilon e_i$ we obtain the Laplacian $L + \epsilon b_i b_i^T$. The Sherman-Morrison formula for the matrix pseudo-inverse yields:

$$(I + L + \epsilon b_i b_i^T)^+ = (I + L)^{-1} - \epsilon \frac{(I + L)^{-1} b_i b_i^T (I + L)^{-1}}{1 + \epsilon b_i^T (I + L)^{-1} b_i},$$

and hence, thinking of $(I + L)^{-1}$ as a matrix-valued function of w ,

$$\begin{aligned} \frac{\partial (I + L)^{-1}}{\partial w_i} &= \\ \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left[(I + L)^{-1} - \epsilon \frac{(I + L)^{-1} b_i b_i^T (I + L)^{-1}}{1 + \epsilon b_i^T (I + L)^{-1} b_i} - (I + L)^{-1} \right] &= \\ - (I + L)^{-1} b_i b_i^T (I + L)^{-1}. & \end{aligned}$$

Therefore, by linearity, and the fact that the objective $\bar{z}^T \bar{z} + \bar{z}^T L \bar{z} = \bar{s}^T (I + L)^{-1} \bar{s}$ as shown in Theorem 3.3,

$$\frac{\partial \bar{s}^T (I + L)^{-1} \bar{s}}{\partial w_i} = -\bar{s}^T (I + L)^{-1} b_i b_i^T (I + L)^{-1} \bar{s}.$$

Non-convexity. Perhaps surprisingly, a slightly more general form of our objective, where one of the two terms is multiplied by any factor $\rho \geq 0$ (i.e., polarization and disagreement are weighted differently), is not convex!

THEOREM 3.4. *Let $\rho \bar{z}^T \bar{z} + \bar{z}^T L \bar{z}$, $\rho \geq 0$ be our objective. For $\rho = 0$, the objective is a non-convex function of the edge weights.*

PROOF. By Propositions 3.1 and 3.2 we obtain that

$$z^{*T} L z^* = \bar{z}^T L \bar{z} = \bar{s}^T (I + L)^{-1} L (I + L)^{-1} \bar{s}.$$

To prove that this is not convex, it suffices to prove that $f(L) = (I + L)^{-1} L (I + L)^{-1}$ is non-convex. Let $L_1, L_2 \in \mathcal{L}$. Assume for the sake of contradiction that $f(L) = (I + L)^{-1} L (I + L)^{-1}$ is convex. Then, the following follows by convexity for any $\lambda \in (0, 1)$,

$$\begin{aligned} \lambda f(L_1) + (1 - \lambda) f(L_2) &\geq f(\lambda L_1 + (1 - \lambda) L_2) \rightarrow \\ \lambda (I + L_1)^{-1} L_1 (I + L_1)^{-1} + (1 - \lambda) (I + L_2)^{-1} L_2 (I + L_2)^{-1} &\geq \\ (I + \lambda L_1 + (1 - \lambda) L_2)^{-1} (\lambda L_1 + (1 - \lambda) L_2) (I + \lambda L_1 + (1 - \lambda) L_2)^{-1}. & \end{aligned}$$

Set $\lambda = 0.5$, and consider the following two Laplacian matrices corresponding to two paths on 3 nodes.

$$L_1 = \begin{bmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{bmatrix}, L_2 = \begin{bmatrix} 2 & -1 & -1 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

It is easy to verify numerically that the minimum eigenvalue of $\lambda (I + L_1)^{-1} L_1 (I + L_1)^{-1} + (1 - \lambda) (I + L_2)^{-1} L_2 (I + L_2)^{-1} - (I + \lambda L_1 + (1 - \lambda) L_2)^{-1} (\lambda L_1 + (1 - \lambda) L_2) (I + \lambda L_1 + (1 - \lambda) L_2)^{-1}$ is negative. This contradicts the convexity assumption of $f(L)$. \square

We note that it is easy to construct more counterexamples for other ρ values (details omitted).

Finally, we prove that there always exists a graph with $O(\frac{n}{\epsilon^2})$ edges that lies within a multiplicative $(1 \pm \epsilon + O(\epsilon^2))$ factor of

the optimal – that is, approximately minimizing the polarization-disagreement index $\mathcal{I}_{G,s}$ does not require a dense graph G . Our proof relies on the seminal work on spectral sparsification [3, 42–44]. The next theorem proves a sub-optimal result in terms of the number of edges in G , but uses the sparsification algorithm due to Spielman and Srivastava based effective resistances, which has the advantages of being fast and easy to implement [42].

THEOREM 3.5. *There always exists a graph G with $O(n \log n / \epsilon^2)$ edges that achieves polarization-disagreement index $\mathcal{I}_{G,s}$ within a multiplicative $(1 + \epsilon + O(\epsilon^2))$ factor of optimal for Problem 1.*

PROOF. Let L be the Laplacian that minimizes the sum of polarization and disagreement subject to total edge weight constraint m , i.e, the solution to (5). By applying the Spielman-Srivastava algorithm to L with error parameter ϵ' , we obtain a spectral sparsifier whose combinatorial Laplacian \tilde{L} satisfies for any $x \in \mathbb{R}^n$:

$$(1 - \epsilon') x^T L x \leq x^T \tilde{L} x \leq (1 + \epsilon') x^T L x. \quad (6)$$

This in turn implies,

$$\begin{aligned} (1 - \epsilon') x^T (L + I) x &\leq x^T (\tilde{L} + I) x \leq (1 + \epsilon') x^T (L + I) x \rightarrow \\ \frac{1}{1 + \epsilon'} x^T (L + I)^{-1} x &\leq x^T (\tilde{L} + I)^{-1} x \leq \frac{1}{1 - \epsilon'} x^T (L + I)^{-1} x. \end{aligned}$$

By setting $x = s$, $OPT \stackrel{\text{def}}{=} s^T (L + I)^{-1} s$, and using the Taylor expansion for the fraction $\frac{1}{1 - \epsilon'} = 1 + \epsilon' + O(\epsilon'^2)$ when ϵ' is small, we obtain that $s^T (\tilde{L} + I)^{-1} s \leq (1 + \epsilon' + O(\epsilon'^2)) \cdot OPT$.

Note that we may not have $\text{Tr}(\tilde{L}) = 2m$. However, by (6), $\text{Tr}(\tilde{L}) \leq (1 + \epsilon') \text{Tr}(L) = (1 + \epsilon') 2m$. Thus we can scale \tilde{L} to have trace $\text{Tr}(\tilde{L}) = 2m$ and will still have $s^T (\tilde{L} + I)^{-1} s \leq \frac{1}{1 - \epsilon'} \cdot (1 + \epsilon' + O(\epsilon'^2)) \cdot OPT \leq (1 + \epsilon + O(\epsilon^2)) \cdot OPT$ if we set $\epsilon' = \epsilon/2$. Finally, notice that $s^T (\tilde{L} + I)^{-1} s \geq s^T (L + I)^{-1} s = OPT$ since L is the optimizer of (5). \square

As mentioned, we use the Spielman-Srivastava sparsification algorithm [42] in our proof since we use it in our experiments. However, a follow-up result due to Batson-Spielman-Srivastava [3] reduces the number of edges to linear in the number of nodes. Using the same mathematical arguments, but invoking [3] instead of [42] we obtain the following corollary of Theorem 3.5.

COROLLARY 3.6. *There always exists a graph with $O(n/\epsilon^2)$ edges that achieves polarization-disagreement index $\mathcal{I}_{G,s}$ within a multiplicative $(1 + \epsilon + O(\epsilon^2))$ factor of the optimal for Problem 1.*

3.3 Optimizing over Innate Opinions

We next give a mathematical formulation of Problem 2. We show how to simplify this formulation to obtain a convex optimization program (specifically, an SDP) which can be solved in polynomial time using standard algorithms [7].

Equation (7) provides a straight-forward way to model our problem. We wish to minimize the sum of polarization and disagreement subject to the structure implied by the dynamics, and a total budget α on the total change of innate opinions. The variable we optimize over is ds , the change in opinions. We restrict this change to decrease the innate opinions, i.e., $ds \leq \vec{0}$.

$$\begin{aligned}
& \min_{ds \in \mathbb{R}^n} && \vec{z}^T \bar{z} + \vec{z}^T L \bar{z} \\
& \text{subject to} && \vec{z}^* = (I + L)^{-1}(s + ds) \\
& && \bar{z} = \vec{z}^* - \frac{\vec{1}^T \vec{z}^*}{n} \vec{1} \\
& && \vec{1}^T ds \geq -\alpha \\
& && ds \leq \vec{0} \\
& && s + ds \geq \vec{0}
\end{aligned} \tag{7}$$

Our main result is the following proposition.

PROPOSITION 3.7. *The formulation of Problem 2 in (7) is solvable in polynomial time.*

PROOF. We can simplify the objective of (7) using our analysis from Section 3.2. Specifically, our problem is equivalent to

$$\begin{aligned}
& \text{minimize} && (s + ds)^T (I + L)^{-1}(s + ds) \\
& \text{subject to} && ds \leq \vec{0} \\
& && \vec{1}^T ds \geq -\alpha \\
& && s + ds \geq \vec{0}
\end{aligned} \tag{8}$$

By expanding the above objective, we observe that it is a standard quadratic form $x^T Qx + 2b^T x + c$ where $Q \stackrel{\text{def}}{=} (I + L)^{-1}$ is symmetric positive semidefinite, $b = (I + L)^{-1}s$ and $c = s^T (I + L)^{-1}s$ (note that b, c are fixed with respect to the variable ds). This objective is convex, and the set of constraints form a convex set, proving the Proposition. \square

We remark that the above convex formulation (7) can accommodate other types constraints, such as restricting ds to be in a specific range of values, or allowing positive changes in the innate opinions. In our experimental section, we use the basic formulation without additional constraints.

4 EXPERIMENTAL RESULTS

4.1 Experimental Setup

Datasets. We use two datasets collected by De et al. [15]. Specifically, De et al. collect text produced by users, and interactions between them [15]. The text is mapped into opinions using NLP tools [39], and the interactions are used to create a network between users. There are many preprocessing details, which can be found in [15]. Here, we provide a brief description of both datasets.

Twitter is a network with $n = 548$ nodes, and $m = 3,638$ edges. Edges correspond to user interactions. All nodes' opinions are available to us. This dataset was collected with the aim of analyzing the debate on Twitter about the Delhi legislative assembly elections of 2013. This was an irresolute event with three major parties winning roughly equal shares of the vote. The hashtags used to collect tweets were #BJP, #APP, #Congress, #Polls2013, tweeted over the period December 9th to 15th.

Reddit is a network with $n = 556$ and $m = 8,969$ edges. There is an edge between two users if there exist two subreddits (other than politics) that both users posted in during the given time period. The topic of interest is politics.

Aside from the above real-world datasets, we use a wide variety of synthetic data. Since most real-world networks have a skewed degree distributions, we focus on power law random networks generated using the Norros-Reittu model [38], an important random

	Proposed method					
	ER(0.5)	PL(2)	PL(2.5)	PL(3)	L^*	L^* -sparsified
$s \sim PL(1.5)$	14.38	16.10	22.06	53.05	11.60	11.60
$s \sim PL(2)$	25.98	45.16	72.11	107.23	19.24	19.27
$s \sim PL(2.5)$	94.87	103.62	121.21	166.38	85.55	85.56

Table 2: Polarization-disagreement minimization for random power law opinions vectors. Average polarization-disagreement indices $\bar{I}_{G,s}$ over 5 experiments for Erdős-Rényi, random power law graphs, and the optimal graphs L^* and \tilde{L}^* -sparsified output by our solutions to Problem 1. Rows correspond to generating innate opinions s according to a power law distribution with slopes 1.5, 2, and 2.5 respectively. For details see Section 4.2.

graph model that produces networks that mimic real-world networks in some respects [2, 23]. We also use skewed distributions for generating opinions. We use the *randht.m* file by Aaron Clauset [9] to generate opinions according to a power law with a given slope. We normalize opinions to the range $[0, 1]$ by dividing by the maximum observed value (i.e., there is always a node with opinion 1).

Machine specs. All experiments were run on a laptop with 1.7 GHz Intel Core i7 processor and 8GB of main memory.

Code. Our code was written in Matlab. Our code is publicly available at <https://github.com/tsourolampis/polarization-disagreement>.

4.2 Experimental Findings

Synthetic experiments. Table 2 shows our results on learning optimal graph topologies (i.e. solving Problem 1) for various innate opinion vectors averaged over 5 experiments. Each row corresponds to sampling a vector of 100 innate opinions according to the power law distribution with slopes 1.5, 2, and 2.5 respectively. Column L^* shows the objective value achieved by our optimization algorithm, and the last column shows the \tilde{L}^* -sparsified the objective value after we sparsify L^* using the Spielman-Srivastava algorithm (see Theorem 3.5). We observe a negligible loss, but achieve a dramatic reduction in the number of edges. The optimal solution L^* in all experiments corresponds to a weighted clique, i.e., there are 4950 non-zero edges. The sparsified versions of the optimal solution have on average 761, 697, and 700 edges for the three different distributions on s respectively.

For comparison, in Table 2 we also show the polarization-disagreement index for several randomly generated graphs. The first column shows the results obtained by a random binomial graph, i.e., an Erdős-Rényi graph. We use $p = 0.5$ as the underlying edge probability. We also generate power law networks according to the Norros-Reittu model with slopes 2, 2.5, and 3 respectively. In general, we observe that random binomial graphs achieve a close-to-optimal performance, in contrast to power law networks. Understanding this observation using formal random graph theory [23] is an interesting direction for future research. We also observe empirically that the polarization-disagreement index increases as the slope of the random network increases from 2 to 3.

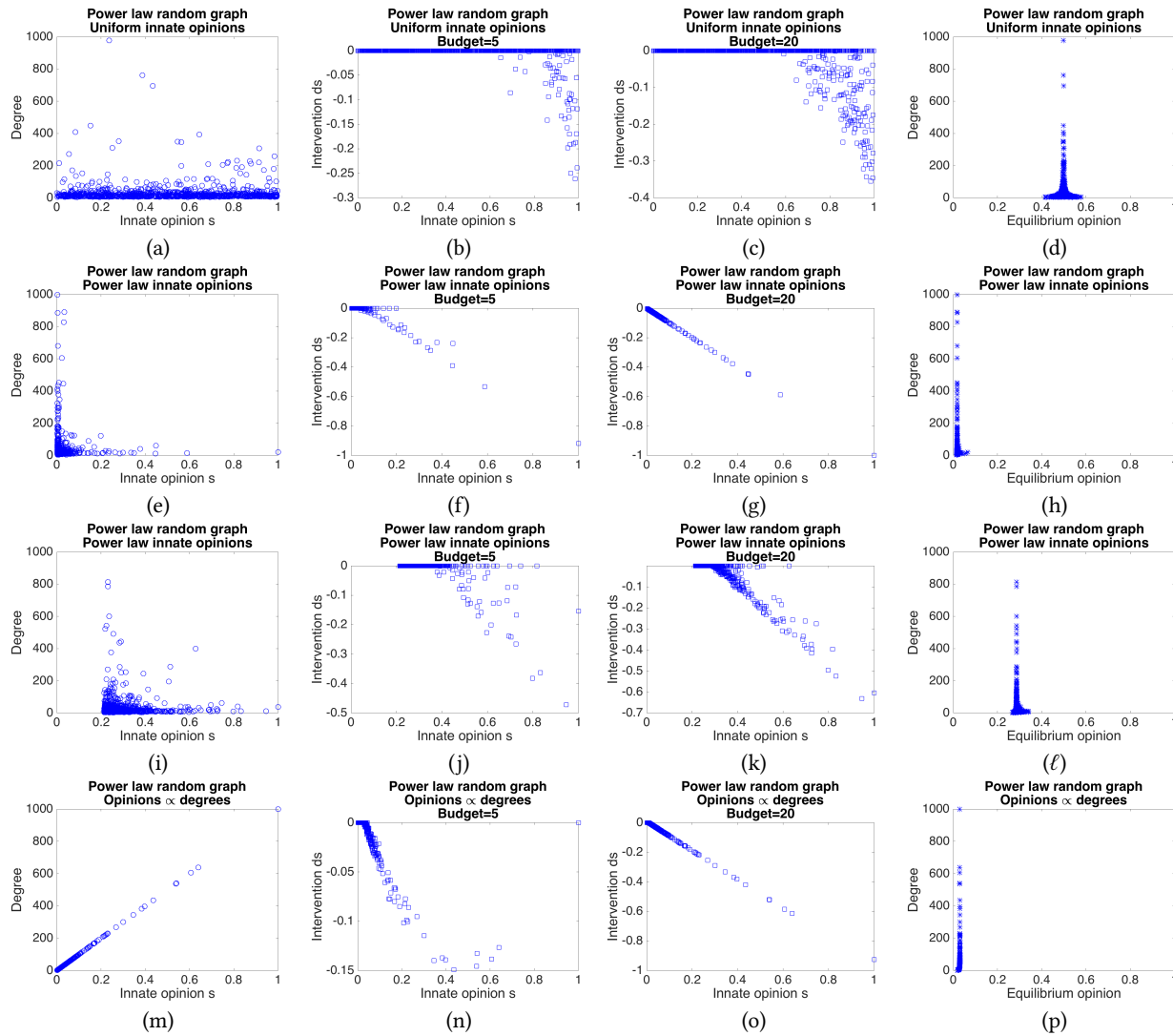


Figure 1: Optimizing over opinion interventions for varying innate opinions distributions on random power law graph topologies generated according to the Norros-Reittu model [38]. For details, see Section 4.2.

Figure 1 shows our experimental findings when we optimize over innate opinions. We generate power-law random graphs using the Norros-Reittu model with slope 2, a slope that lies close to the slopes of several real-world networks [37]. Similarly we generate random initial opinion vectors according to various distributions. The first row (Figures 1(a),(b),(c),(d)) corresponds to uniform innate opinions. The second row (Figures 1(e),(f),(g),(h)) to power law opinions with slope 2, and the third row (Figures 1(i),(j),(k),(l)) to slope 5. For the fourth row (Figures 1(m),(n),(o),(p)) we set the opinion of each node v deterministically to $\frac{deg(v)}{\sum_u deg(u)}$. This setting allows us to clearly see if the algorithm chooses to change the innate opinions of only the highest degree nodes, or if the graph topology forces the algorithm to choose other nodes (which is actually the case). The first column (Figures 1(a),(e),(i),(m)) plots the degree of

each node versus its innate opinion. Figures 1(b),(f),(j),(n), and Figures 1(c),(g),(k),(o) plot the optimal change ds for each node versus its innate opinion for budgets $\alpha = 5$, and $\alpha = 20$ respectively. We observe that the optimization algorithm tends to prefer to reduce high innate opinions. However we can also observe that graph topology plays a key role, as in certain cases it changes opinions of nodes with lower innate opinions even if it could choose higher innate opinions. This is obvious in all figures, but most striking in Figure 1(n); the algorithm does not modify the node with innate opinion 1 when the budget is 5, but only modifies other nodes. Similar remarks can be made for the rest of the figures. This fact is more pronounced when $\alpha = 5$ than when $\alpha = 20$. since as the budget α increases, the algorithm can always reduce the values of the highest innate opinions, caring less and less about the graph topology. Finally, Figures 1(d),(h), (l), (p) plot the opinion vector at equilibrium

after our algorithm’s intervention (i.e, $z = (I + L)^{-1}(s + ds)$), when the budget is 20. We observe that in all cases opinions get closer, but the actual value that they concentrate around depends on the distribution. For instance, in Figure 1(d) we observe concentration around 0.5 but in Figure 1(h) around 0. This should not be very surprising; the average of the uniform opinion distribution is 0.5, whereas when s is distributed as a power law distribution, most of the values initially are close to 0. The results presented in Figure 1 are representative observations across many experiments.

	Twitter		Reddit
$\mathcal{I}_{Twitter,s}$	199.84	$\mathcal{I}_{Reddit,s}$	138.02
# Edges	3,638	# Edges	8,969
$\mathcal{I}_{G^*,s}$	30.113	$\mathcal{I}_{G^*,s}$	0.0022
# Edges	120,000	# Edges	103,050
$\mathcal{I}_{\hat{G}^*,s}$	30.114	$\mathcal{I}_{\hat{G}^*,s}$	0.0022
# Edges	3,455	# Edges	7,521

Table 3: Optimizing over graph topologies. First row, shows the objective, and the number of edges for the Twitter and Reddit networks. Second row shows the objective and the number of edges for the optimal solution. Third row shows the objective and the number of edges for the sparsified optimal solution. For details, see Section 4.2.

Real-world experiments. Table 3 summarizes our experimental results for the Twitter and Reddit datasets. Given the Twitter network among the users, and their innate opinions, the polarization-disagreement index $\mathcal{I}_{G,s}$ is equal to 199.84. By optimizing over graph topologies, we find a graph G^* with the same total weight 8,969 as the original graph that reduces the polarization-disagreement index to 30.113. However, this optimal graph is dense; the total weight is spread over 120,000 edges. By sparsifying using effective resistances according to Spielman-Srivastava, we obtain a sparse graph using just 3,455 edges. The polarization-disagreement index is equal to 30.114. Again, we observe that there is a negligible change in the objective value (on the order of 10^{-3}), but there is a drastic reduction in the number of edges.

The results for Reddit are impressive; the polarization-disagreement index reduces from 138.02 to 0.0022. Both the original graph and the optimal graph have the same total weight. The weight of the optimal graph is spread over 103,050 edges, but by sparsifying we get a graph with 7,521 edges with polarization-disagreement index 0.0022; the difference from the optimal value is $< 10^{-5}$.

We checked the structure of our optimal graphs, suspecting that two clusters would emerge. Specifically, we centered the innate vector s around the mean, and we obtained two sets of nodes: those with innate opinions below the mean, and those with innate opinions above the mean. Intuitively, small disagreement graphs should have a strong community structure on these two sets of nodes, and small polarization graphs should have a large max cut (edges across the two groups). The optimal graphs do not exhibit such a clear two-block structure, but look closer to being random. Combined with our finding on random binomial graphs from the synthetic experiments, it is an interesting question whether random

binomial graphs are indeed close to the optimal solution for any initial opinion vector.

Runtime-wise, the optimization took a couple of hours for each network; observe that the total number of variables grows as $O(n^2)$. It is an open question to come up with an efficient primal-dual procedure that takes advantage of specific features of our convex programs and Theorem 3.5 to improve the overall runtime.

We also considered the problem of optimizing the vector of innate opinions using different budget values. We show how the equilibria opinions vectors look before and after intervention. Figures 2(b)(c) show four opinion vectors permuted according to the sorting permutation of s for Twitter and Reddit respectively: the vector of innate opinions s that is sorted, the original equilibrium vector, and two optimized equilibria vectors, when the budget is 5, and 20 respectively. We observe that nodes’ opinions converge closer and closer to the average, achieving significantly lower objective values as the budget increases.

5 CONCLUSION

Humans tend to prefer links that minimize disagreement due to the well-known confirmation bias, among other reasons. However, this may lead to polarized communities. In this work we introduce the notion of the *polarization-disagreement index* to understand better this trade-off between disagreement and polarization in online social networks and online social media. To the best of our knowledge, we provide the first formulation for finding an optimal topology which minimizes the sum of polarization and disagreement under a popular opinion formation model. We prove interesting facts about our objective, including the fact that there is always a graph with $O(n/\epsilon^2)$ edges that provides a $(1 + \epsilon)$ approximation to the optimal objective. We also provide efficient procedures to optimize the objective. We complement our results on optimizing graph topologies, by considering a version of the problem where we optimize over innate opinions for a fixed network. Our proposed tools provide useful information about the importance of certain edges and nodes with respect to opinion formation. We provide an extensive empirical study of our results using synthetic data and two real-world datasets (Twitter, and Reddit).

Our work opens many interesting questions. We used the popular Friedkin-Johnsen opinion formation model, which takes into account both consensus and disagreement in the underlying opinion update process. The same questions we asked here can be also asked using other opinion formation models [34]. Our experimental results strongly indicate that Erdős-Rényi graphs are not far away from achieving the optimal polarization-disagreement index $\mathcal{I}_{G,s}$. Can we prove that Erdős-Rényi graphs are always a good approximation to the optimal solution? If yes, can we prove a similar result for expanders? Do links that cross communities always improve the polarization-disagreement index, or not? Another interesting research direction is understanding how well we can approximate the non-convex objective of Theorem 3.4, where a weighting is added to the polarization-disagreement index. Furthermore, we focused on disagreement and polarization, two key phenomena in human societies. Can we develop a similar framework for optimizing other mathematical objectives for social good?

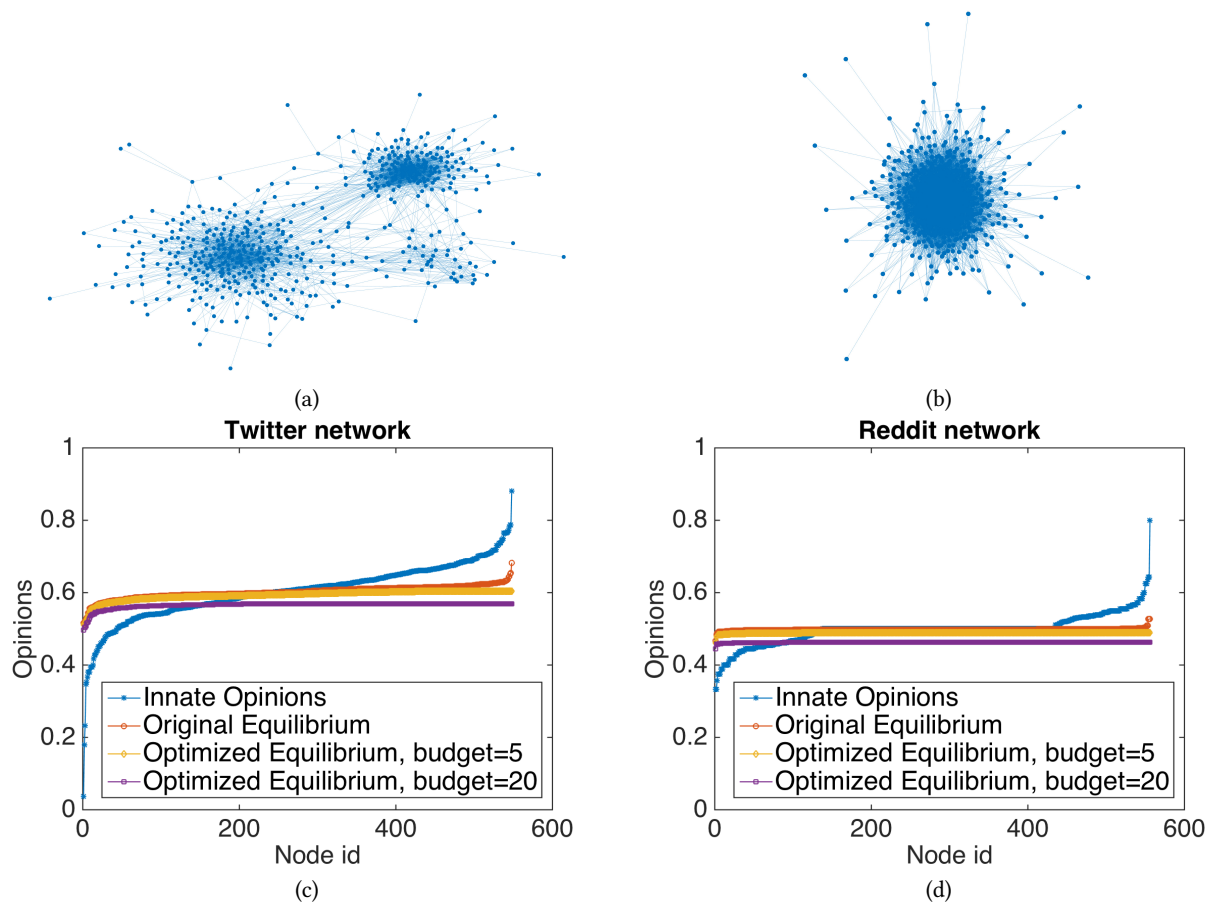


Figure 2: (a) Twitter network. (b) Reddit network. (c),(d) Opinions per node for Twitter and Reddit respectively. Plots shows the innate opinions (*), the opinions at equilibrium with no innate opinion intervention (o), and the opinions at equilibrium using an intervention with budget 5 (diamonds), and 20 (squares) respectively.

REFERENCES

- [1] D. Acemoglu and A. Ozdaglar. Opinion dynamics and learning in social networks. *Dynamic Games and Applications*, 1(1):3–49, 2011.
- [2] W. Aiello, F. Chung, and L. Lu. A random graph model for massive graphs. In *Proceedings of the Thirty-second Annual ACM Symposium on Theory of Computing*, pages 171–180. Acm, 2000.
- [3] J. Batson, D. A. Spielman, and N. Srivastava. Twice-Ramanujan sparsifiers. *SIAM Journal on Computing*, 41(6):1704–1721, 2012.
- [4] D. Bindel, J. Kleinberg, and S. Oren. How bad is forming your own opinion? *Games and Economic Behavior*, 92:248–265, 2015.
- [5] L. Boxell, M. Gentzkow, and J. M. Shapiro. Is the internet causing political polarization? Evidence from demographics. Technical report, National Bureau of Economic Research, 2017.
- [6] S. Boyd, P. Diaconis, and L. Xiao. Fastest mixing Markov chain on a graph. *SIAM review*, 46(4):667–689, 2004.
- [7] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [8] D. Braha and M. A. de Aguiar. Voting contagion: Modeling and analysis of a century of US presidential elections. *PLoS One*, 12(5):e0177970, 2017.
- [9] A. Clauset, C. R. Shalizi, and M. E. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009.
- [10] P. Clifford and A. Sudbury. A model for spatial conflict. *Biometrika*, 60(3):581–588, 1973.
- [11] S. I. Daitch, J. A. Kelner, and D. A. Spielman. Fitting a graph to vector data. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 201–208. ACM, 2009.
- [12] J. Damore. Google’s ideological echo chamber, July 2017. <https://tinyurl.com/yde86v9e>.
- [13] P. Dandekar, A. Goel, and D. T. Lee. Biased assimilation, homophily, and the dynamics of polarization. *Proceedings of the National Academy of Sciences*, 110(15):5791–5796, 2013.
- [14] A. Das, S. Gollapudi, and K. Munagala. Modeling opinion dynamics in social networks. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining, WSDM ’14*, pages 403–412, 2014.
- [15] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, and S. Chakrabarti. Learning a linear influence model from transient opinion dynamics. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 401–410. ACM, 2014.
- [16] M. H. DeGroot. Reaching a consensus. *Journal of the American Statistical Association*, 69(345):118–121, 1974.
- [17] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst. Learning Laplacian matrix in smooth graph signal representations. *IEEE Transactions on Signal Processing*, 64(23):6160–6173, 2016.
- [18] R. Durrett. *Lecture notes on particle systems and percolation*. Brooks/Cole Pub Co, 1988.
- [19] J. Felsenstein. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39(4):783–791, 1985.
- [20] S. Flaxman, S. Goel, and J. M. Rao. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, 80(S1):298–320, 2016.
- [21] N. E. Friedkin and E. C. Johnsen. Social influence and opinions. *Journal of Mathematical Sociology*, 15(3-4):193–206, 1990.
- [22] N. E. Friedkin and E. C. Johnsen. Social positions in influence networks. *Social Networks*, 19(3):209–222, 1997.
- [23] A. Frieze and M. Karoński. *Introduction to random graphs*. Cambridge University Press, 2015.

- [24] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy in social media. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 33–42. ACM, 2016.
- [25] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Reducing controversy by connecting opposing views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 81–90. ACM, 2017.
- [26] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis. Exposing Twitter users to contrarian news. *arXiv preprint arXiv:1703.10934*, 2017.
- [27] A. Gionis, E. Terzi, and P. Tsaparas. Opinion maximization in social networks. In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 387–395. SIAM, 2013.
- [28] R. Hegselmann, U. Krause, et al. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.
- [29] C. Kadushin. *Understanding social networks: Theories, concepts, and findings*. OUP USA, 2012.
- [30] D. Kahneman. *Thinking, fast and slow*. Macmillan, 2011.
- [31] V. Kalofolias. How to learn a graph from smooth signals. In *The 19th International Conference on Artificial Intelligence and Statistics (AISTATS 2016)*. Journal of Machine Learning Research (JMLR), 2016.
- [32] D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.
- [33] Q. V. Liao and W.-T. Fu. Can you hear me now?: Mitigating the echo chamber effect by source position indicators. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 184–196. ACM, 2014.
- [34] E. Mossel, O. Tamuz, et al. Opinion exchange dynamics. *Probability Surveys*, 14:155–204, 2017.
- [35] S. A. Munson, S. Y. Lee, and P. Resnick. Encouraging reading of diverse political viewpoints with a browser widget. In *ICWSM*, 2013.
- [36] P. Nakov. Semantic sentiment analysis of Twitter data. *arXiv preprint arXiv:1710.01492*, 2017.
- [37] M. E. Newman. Power laws, Pareto distributions and Zipf’s law. *Contemporary Physics*, 46(5):323–351, 2005.
- [38] I. Norros and H. Reittu. On a conditionally Poissonian graph process. *Advances in Applied Probability*, 38(1):59–75, 2006.
- [39] J. W. Pennebaker, M. E. Francis, and R. J. Booth. Linguistic inquiry and word count: LIWC 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.
- [40] W. Quattrociocchi, G. Caldarelli, and A. Scala. Opinion dynamics on interacting networks: media competition and social influence. *Scientific Reports*, 4:4938, 2014.
- [41] D. Shultz. Could Google influence the presidential election?. October 2016.
- [42] D. A. Spielman and N. Srivastava. Graph sparsification by effective resistances. *SIAM Journal on Computing*, 40(6):1913–1926, 2011.
- [43] D. A. Spielman and S.-H. Teng. Spectral sparsification of graphs. *SIAM Journal on Computing*, 40(4):981–1025, 2011.
- [44] D. A. Spielman and S.-H. 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.
- [45] J. Sun, S. Boyd, L. Xiao, and P. Diaconis. The fastest mixing Markov process on a graph and a connection to a maximum variance unfolding problem. *SIAM review*, 48(4):681–699, 2006.
- [46] N. Times. Facebook’s Russia-linked ads came in many disguises, Oct. 2017. <https://www.nytimes.com/2017/10/02/technology/facebook-russia-ads-.html>.
- [47] N. Times. Tech giants, once seen as saviors, are now viewed as threats, October 2017.
- [48] C. B. Williams, G. J. Gulati, et al. Social networks in political campaigns: Facebook and the 2006 midterm elections. In *annual meeting of the American Political Science Association*, volume 1, pages 49–62, 2007.
- [49] J. Zhang and Y. Hong. Opinion evolution analysis for short-range and long-range Deffuant–Weisbuch models. *Physica A: Statistical Mechanics and its Applications*, 392(21):5289–5297, 2013.