

# Quantifying and Minimizing Risk of Conflict in Social Networks

Xi Chen

Dept. of Electronics and Information  
Systems, IDLab, Ghent University  
xi.chen@ugent.be

Jefrey Lijffijt

Dept. of Electronics and Information  
Systems, IDLab, Ghent University  
jefrey.lijffijt@ugent.be

Tijl De Bie

Dept. of Electronics and Information  
Systems, IDLab, Ghent University  
tijl.debie@ugent.be

## ABSTRACT

Controversy, disagreement, conflict, polarization and opinion divergence in social networks have been the subject of much recent research. In particular, researchers have addressed the question of how such concepts can be quantified given people's prior opinions, and how they can be optimized by influencing the opinion of a small number of people or by editing the network's connectivity.

Here, rather than optimizing such concepts given a specific set of prior opinions, we study whether they can be optimized in the average case and in the worst case over all sets of prior opinions. In particular, we derive the worst-case and average-case conflict risk of networks, and we propose algorithms for optimizing these.

For some measures of conflict, these are non-convex optimization problems with many local minima. We provide a theoretical and empirical analysis of the nature of some of these local minima, and show how they are related to existing organizational structures.

Empirical results show how a small number of edits quickly decreases its conflict risk, both average-case and worst-case. Furthermore, it shows that minimizing average-case conflict risk often does not reduce worst-case conflict risk. Minimizing worst-case conflict risk on the other hand, while computationally more challenging, is generally effective at minimizing both worst-case as well as average-case conflict risk.

## KEYWORDS

Social networks; conflict; controversy; disagreement measures

### ACM Reference Format:

Xi Chen, Jeffrey Lijffijt, and Tijl De Bie. 2018. Quantifying and Minimizing Risk of Conflict in Social Networks. In *KDD 2018: 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom*. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3219819.3220074>

## 1 INTRODUCTION AND MOTIVATION

The study of how opinions form through social interactions with others with potentially differing opinions has long been studied in the social sciences (see e.g. [11, 14]). Today, online social networks offer unprecedented access to both social interactions and publicly expressed opinions on controversial matters. This now allows one to quantitatively study differences of opinions on a large scale, as well as to moderate them through targeted interventions. This

newfound ability offers new opportunities for conflict prevention and mitigation, as well as for more effective marketing campaigns.<sup>1</sup>

**Background.** Much prior research has focused on opinions on political matters [1, 9, 27]. However, recent work has often studied the problem in a more generic manner (independent of the topic of controversy) [21, 25, 30]. The identification of controversial issues has been studied using tools from sentiment analysis [7, 26], as well as by relying on the structure of the social network and the distribution of opinions across it [1, 3, 8, 9]. Besides identifying or quantifying controversy or conflict, the question of how it can be influenced has received increasing amounts of attention [17, 25, 28]. Strategies that have been considered include editing the graph (or even designing it from scratch), and attempting to alter the opinions of a small number of individuals [6, 16, 19, 25, 28].

Most of these results are based on the opinion formation model by Friedkin and Johnsen [14], which extended DeGroot's model of opinion averaging [11]. In Friedkin and Johnsen's model, individuals are assumed to hold an ('a priori') *internal opinion*, while they may *express* an opinion that may differ from it but that is more socially acceptable (i.e. more similar to their friends' opinions). To model this, it is assumed that individuals are connected to each other in a social network, and that individuals' expressed opinion is a weighted average of their own internal opinion and their neighbors' expressed opinions, with weights representing the strength of the connections in the network.

**Shortcomings in the state-of-the-art.** An important problem with Friedkin and Johnsen's model is that, while external opinions are hard to measure, access to internal opinions is near-impossible in practice. Another shortcoming of the dominant line of research attempting to reduce conflict by editing the social network is that it tends to focus on a single or a given set of controversial topics. Yet, different issues do not generally correspond to different social networks, such that editing a social network to minimize conflict on one issue may actually increase conflict on another.

**Contributions in this paper.** In this paper, we depart from the existing literature in focusing on *risk of conflict*, rather than on conflict around one particular issue. In this way, we overcome both shortcomings of prior work discussed above. We still rely on Friedkin and Johnsen's model of opinion formation to quantify the risk of networks to conflict (which we discuss in detail in Sec. 2). However, the proposed quantifications are independent of any particular set of internal (or external) opinions, depending purely on the topology of the network. In this way, we bypass the problem that quantifying internal opinions is beyond reach in practice. Moreover, attempting to reduce the *risk* of conflict, leads to more robust network editing

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

KDD 2018, August 19–23, 2018, London, United Kingdom

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-5552-0/18/08.

<https://doi.org/10.1145/3219819.3220074>

<sup>1</sup>It also creates risks: it could allow oppressive governments to design more effective propaganda, or hostile actors to incite conflict rather than prevent it. These risks are an additional reason for these matters to be studied by the scientific community.

strategies than reducing conflict for one particular assignment of internal opinions.

More specifically, we propose two measures of conflict risk: the *worst-case conflict risk (WCR)* and the *average-case conflict risk (ACR)*, respectively quantifying the amount of conflict in the *worst-case*, and *on average*, over all possible internal opinions. Subsequently, we demonstrate how both WCR and ACR can be minimized by locally editing the network. We do this for a number of pre-existing measures of conflict and disagreement discussed in Sec. 3, most notably the *internal conflict* (the extent to which individuals are torn by expressing an opinion that differs from their internal opinion), *external conflict* (the extent to which neighboring individuals express different opinions), and *controversy* (the overall variation in expressed opinion). A side-result in this paper is an equality relating these different conflict measures, leading to what we refer as a *conservation law of conflict*: the sum of the internal conflict, twice the external conflict, and controversy is a constant.

In Sec. 4 we propose two types of algorithms (one coordinate descent, and one conditional gradient descent) to locally edit the social network to reduce the WCR and ACR for a number of these measures of conflict. Empirical results are provided in Sec. 5, evaluating the effectiveness of the proposed algorithms at reducing risk of conflict, providing additional insight into the local minima of the measures, and discussing conflict risk in random network models.

We end with related work in Sec. 6 and conclusions in Sec. 7.

**Notation.** Let  $G = (V, E, w)$  be an undirected positive-weighted network with  $V = \{1, \dots, n\}$  the set of nodes,  $E \subseteq V \times V$  the set of  $m = |E|$  edges (with  $(i, j) \in E$  iff  $(j, i) \in E$ ), and  $w$  a weight function mapping an edge  $e \in E$  onto its weight  $w(e) > 0$ . We denote with  $A$  the (symmetric) adjacency matrix (with zero diagonal), defined by  $a_{ij} = w(i, j)$  iff  $(i, j) \in E$  and  $a_{ij} = 0$  otherwise. With  $N(i)$  we denote the set of neighboring nodes of node  $i$ :  $N(i) \triangleq \{j \in V \mid (i, j) \in E\}$ . Let  $\mathbf{1}$  denote the vector of ones of appropriate size. Furthermore, let  $\mathbf{d} \triangleq A\mathbf{1}$  denote the vector containing the weighted degrees of all nodes, and  $D \triangleq \text{diag}(\mathbf{d})$  the diagonal degree matrix. Then the Laplacian matrix is defined as  $L \triangleq D - A$ .

## 2 OPINION FORMATION MODELS

Here we briefly discuss the models of opinion formation on social networks, as formalized above, related to the present paper.

**The dynamic model.** According to DeGroot’s model [11], people’s opinions are updated gradually through repeated communication. In the model, every person  $i \in V$  has an opinion  $s_i(t)$  at time  $t$ , and it is influenced by its direct neighbors so as to evolve into a different opinion  $s_i(t+1)$  in the next time step. More precisely, their opinion is updated as the weighted sum of their own opinion (with weight  $w_{ii}$ ) and those of the neighbors (with weight  $w_{ij}$  for neighbor  $j$ ). Given a weighted graph  $G = (V, E, w)$ , and the opinions  $s_i(t)$  of the nodes at time  $t$ , the updating rule is defined as:

$$s_i(t+1) = \frac{w_{ii}s_i(t) + \sum_{j \in N(i)} w_{ij}s_j(t)}{w_{ii} + \sum_{j \in N(i)} w_{ij}} \quad (1)$$

This model formalizes opinion formation as a repeated averaging process of one’s opinion with one’s neighbors.

**The static model.** In 1990, Friedkin and Johnsen extended the model by DeGroot to have two different kinds of opinions [14]:

**Table 1: Measures for conflict in undirected networks**

Name	$\mathbf{z}$	$\mathbf{s}$
internal conflict: $ic$	$\mathbf{z}^T L^2 \mathbf{z}$	$\mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} L^2 (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$
external conflict: $ec$	$\mathbf{z}^T L \mathbf{z}$	$\mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} L (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$
controversy: $c$	$\mathbf{z}^T \mathbf{z}$	$\mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-2} \mathbf{s}$
resistance: $r$	$\mathbf{z}^T \mathbf{s}$	$\mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$

an internal opinion  $s_i$  and an expressed opinion  $z_i$ . The internal opinions of every person are assumed fixed, while the expressed opinions are influenced by the node’s own internal opinion as well the expressed opinions of the neighbors, as follows:

$$z_i = \frac{w_{ii}s_i + \sum_{j \in N(i)} w_{ij}z_j}{w_{ii} + \sum_{j \in N(i)} w_{ij}}. \quad (2)$$

Expressed in matrix-vector notation, and with  $w_{ii} = 1$  (a common assumption in the literature that we also make in this paper), this equation is solved by (3) below at equilibrium [6]:

$$\mathbf{z} = (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}. \quad (3)$$

In this model, the internal opinion  $s_i$  of node  $i$  is considered a constant, and private to each individual, while the expressed opinion  $z_i$  is public, and a compromise between the internal opinion of node  $i$  and the expressed opinion of node  $i$ ’s neighbors.

**REMARK 1.** In this paper, we will generally assume that the internal opinions are mean-centered. Note that in that case, also  $\mathbf{z}$  will be mean-centered. As opinions are arguably relative, this assumption should not incur any loss of generality. Rather on the contrary: some measures of opinions are affected by the mean of  $\mathbf{s}$  (as we will point out later), which is arguably undesirable, such that assuming  $\mathbf{s}$  has zero mean enhances the usability of the proposed measures.

## 3 CONFLICT AND CONFLICT RISK

In this paper, we rely on Friedkin and Johnsen’s model of opinion formation and discuss a number of (previously known) measures of conflict in terms of the internal opinions  $\mathbf{s}$  and expressed opinions  $\mathbf{z} = (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$ . Note that we will often use the term *conflict* in a more generic manner in this paper, to signify conflict, controversy, or disagreement more generally.

In Sec. 3.1 we survey the measures of conflict and discuss how they can be computed using matrix-vector operations. Section 3.2 introduces an intriguing though intuitive connection between some of these measures. Finally, in Sec. 3.3 we discuss how the risk of conflict, as quantified by the proposed measures, can be formulated, both in the worst case (WCR) and in the average-case (ACR).

### 3.1 Conflict Measures

Table 1 provides an overview of the proposed measures, which we will discuss in greater detail below.

**Internal Conflict  $ic$ .** The internal conflict measure is designed to quantify the extent to which individuals’ internal and expressed opinions differ.

**Definition 3.1.** The internal conflict  $ic$  is the sum of squares of the differences between individual internal and expressed opinions:

$$ic = \sum_i (z_i - s_i)^2.$$

The following proposition provides a convenient matrix-vector expression for it. The proof is elementary and omitted for brevity.

**PROPOSITION 3.2.**  $ic = \mathbf{z}^T \mathbf{L}^2 \mathbf{z} = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} \mathbf{L}^2 (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$ .

**External Conflict  $ec$ .** Arguably the most relevant measure in practice, the external conflict measure quantifies the extent to which the expressed opinions of neighbors are in disagreement with each other. Formally:

**Definition 3.3.** The external conflict  $ec$  is the weighted sum of squares of the pairwise differences between the expressed opinions of neighbors in the network:

$$ec = \sum_{(i,j) \in E} w_{ij} (z_i - z_j)^2.$$

Again, it can be expressed conveniently in matrix-vector form:

**PROPOSITION 3.4.**  $ec = \mathbf{z}^T \mathbf{L} \mathbf{z} = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} \mathbf{L} (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$ .

The proposed measure of external conflict is closely related to the so-called Network Disagreement Index (NDI) in [10], except that there are two different opinions in our work: it is equal to the NDI evaluated on the external opinions.

**Controversy  $c$ .** Given the expressed opinions, the controversy does not depend on the network structure, and simply quantifies how much the opinion varies across the individuals in the network:

**Definition 3.5.** The controversy  $c$  is the sum of the squares of the expressed opinions:

$$c = \sum_i z_i^2.$$

Again, this can be trivially expressed in matrix-vector form:

**PROPOSITION 3.6.**  $c = \mathbf{z}^T \mathbf{z} = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-2} \mathbf{s}$ .

The controversy  $c$  is equivalent with the polarization index proposed by Matakos et al. [25], although they normalized the measure by  $n$ , the number of nodes in the network. For zero mean  $\mathbf{s}$  (and hence zero mean  $\mathbf{z}$ ), as we assume in this paper, the controversy is also equivalent to the Global Disagreement Index [10], defined as:

$$\gamma(\mathbf{x}) := \sum_{i < j} (x_i - x_j)^2 \quad (4)$$

More specifically, the GDI is a constant factor  $n$  times larger than the controversy.

**Resistance  $r$ .** The final measure we wish to discuss is the *resistance*.<sup>2</sup>

<sup>2</sup> Its suggested name stems from its mathematical form, which is closely related to the effective resistance in graphs [12]:  $R_{ij} = (\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{L}^+ (\mathbf{e}_i - \mathbf{e}_j)$ , thus it is called resistance. In a graph, the effective resistance between two nodes  $i$  and  $j$  is:  $(\mathbf{e}_i - \mathbf{e}_j)^T \mathbf{L}^+ (\mathbf{e}_i - \mathbf{e}_j)$ .  $\mathbf{e}_i$  has one at position  $i$  and zeros elsewhere. If  $\mathbf{s} = \mathbf{e}_i - \mathbf{e}_j$  where only the opinions of the two nodes count,

$$r = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s} = (\mathbf{e}_i - \mathbf{e}_j)^T (\mathbf{L} + \mathbf{I})^{-1} (\mathbf{e}_i - \mathbf{e}_j).$$

**Definition 3.7.** The resistance  $r$  is the inner product between expressed and internal opinion vectors:

$$r = \sum_i s_i z_i.$$

It can again be expressed in matrix-vector notation:

**PROPOSITION 3.8.**  $r = \mathbf{s}^T \mathbf{z} = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s}$ .

The resistance was in fact introduced earlier by Musco et al. [28] (where it was given no name). It was introduced there as the sum of the controversy and external conflict:

**PROPOSITION 3.9.** *Resistance is the sum of external conflict and controversy:  $r = ec + c$ .*

Their work included an algorithm for optimizing the network to reduce conflict given a specified internal opinion vector  $\mathbf{s}$ , and took advantage of the fact that resistance is matrix-convex in  $\mathbf{L}$ .

**Summary.** Thus, each of the measures can be written in the form

$$* = \mathbf{s}^T \mathbf{M}_* \mathbf{s},$$

where  $*$  is one of  $ic$ ,  $ec$ ,  $c$ , or  $r$ , and  $\mathbf{M}_{ic} = (\mathbf{L} + \mathbf{I})^{-1} \mathbf{L}^2 (\mathbf{L} + \mathbf{I})^{-1}$ ,  $\mathbf{M}_{ec} = (\mathbf{L} + \mathbf{I})^{-1} \mathbf{L} (\mathbf{L} + \mathbf{I})^{-1}$ ,  $\mathbf{M}_c = (\mathbf{L} + \mathbf{I})^{-2}$ , and  $\mathbf{M}_r = (\mathbf{L} + \mathbf{I})^{-1}$ .

We note in passing that the matrices  $\mathbf{L}$  and  $(\mathbf{L} + \mathbf{I})$  obviously have the same eigenspaces, such that they commute – i.e. the factors in the expressions for  $\mathbf{M}_*$  can be freely rearranged.

## 3.2 A conservation law of conflict

In this section, we state an identity that implies that the different measures of conflict act like communicating vessels: reducing one implies that another one must be increased.

**THEOREM 3.10 (CONSERVATION LAW OF CONFLICT).** *Given a network and an internal opinion vector  $\mathbf{s}$ , then the sum of  $ic$ ,  $2ec$ , and  $c$  is a constant equal to  $\mathbf{s}^T \mathbf{s}$ :*

$$ic + 2ec + c = \mathbf{s}^T \mathbf{s}.$$

**PROOF.**  $ic + 2ec + c = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} (\mathbf{L}^2 + 2\mathbf{L} + \mathbf{I}) (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s} = \mathbf{s}^T (\mathbf{L} + \mathbf{I})^{-1} (\mathbf{L} + \mathbf{I})^2 (\mathbf{L} + \mathbf{I})^{-1} \mathbf{s} = \mathbf{s}^T \mathbf{s}$ .  $\square$

Note that the constant  $\mathbf{s}^T \mathbf{s}$  could be regarded as *the internal controversy*: the inherent controversy on a particular topic. The conservation law essentially states that in a social network, this inherent controversy is divided over external conflict, internal conflict, and a remaining amount of controversy. The relative proportions of each of these measures of conflict depend on the structure of the network in relation to the internal opinion vector  $\mathbf{s}$ .

To understand this better, let  $\mathbf{L} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^T$  be the eigenvalue decomposition of  $\mathbf{L}$ . It is known from graph signal processing [31] that eigenvectors corresponding to small eigenvalues are slowly varying over the graph (i.e. the  $i$ 'th and  $j$ 'th entries of such an eigenvector tend to be similar if  $i$  and  $j$  are nearby in the graph), while the large eigenvalues correspond to eigenvectors that fluctuate rapidly over the graph. The eigenvalue decompositions of the diagonal matrices with eigenvalues  $\mathbf{M}_*$  are then given by:

$$\mathbf{M}_{ic} = \mathbf{U} \mathbf{\Lambda}^2 (\mathbf{\Lambda} + \mathbf{I})^{-2} \mathbf{U}^T, \mathbf{M}_{ec} = \mathbf{U} \mathbf{\Lambda} (\mathbf{\Lambda} + \mathbf{I})^{-2} \mathbf{U}^T, \mathbf{M}_c = \mathbf{U} (\mathbf{\Lambda} + \mathbf{I})^{-2} \mathbf{U}^T.$$

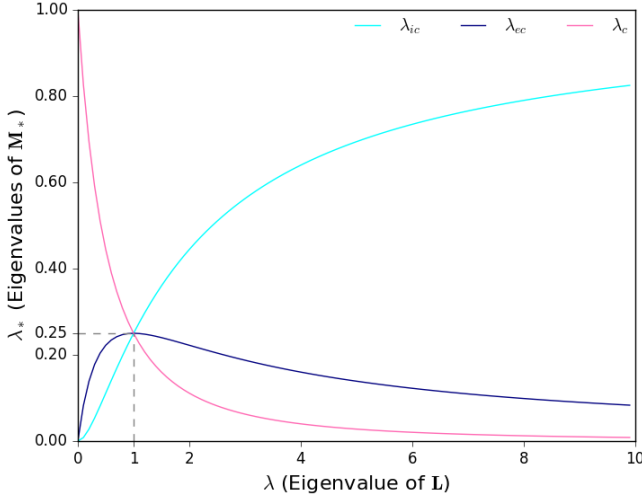


Figure 1: Eigenvalues in the Conservation Law.

In other words, any eigenvalue  $\lambda$  of the Laplacian  $\mathbf{L}$  yields a corresponding eigenvalue of the  $\mathbf{M}_*$  matrices as follows:

$$\lambda_{ic} = \frac{\lambda^2}{(\lambda + 1)^2}, \quad \lambda_{ec} = \frac{\lambda}{(\lambda + 1)^2}, \quad \lambda_c = \frac{1}{(\lambda + 1)^2}.$$

These eigenvalues are plotted as a function of the eigenvalue  $\lambda$  of the Laplacian in Fig. 1.<sup>3</sup>

Note that  $\lambda_c$  increases with  $\lambda$ ,  $\lambda_{ic}$  decreases with  $\lambda$ , and  $\lambda_{ec}$  first increases to reach a maximum value of 0.25 at  $\lambda = 1$  after which it decreases again.

For a fixed 2-norm of the internal opinion vector  $\mathbf{s}$ , the measure of conflict with  $\mathbf{M}_*$  is larger the more it is aligned with the eigenvectors corresponding to the largest eigenvalues of  $\mathbf{M}_*$ . Figure 1 shows that this differs for the different measures.

For  $\mathbf{s}$  aligning with the smoothest eigenvectors of the Laplacian (i.e. those corresponding to small eigenvalues  $\lambda$  of the Laplacian), the internal and external conflicts are small, but the controversy is large as internal opinions remain unmoderated by dissenting neighbors. This situation would arise when the graph contains different (nearly) disconnected communities, and within each community the internal opinion is constant, while between the communities the internal opinions differ. As  $\mathbf{s}$  becomes more aligned with less smooth eigenvectors (i.e. with larger eigenvalues), the external conflict starts to increase because conflicts between neighbors are starting to arise. For the same reason, the internal conflict starts to increase, and the controversy starts to decrease. The external conflict reaches its maximum when  $\mathbf{s}$  is aligned with eigenvectors of  $\mathbf{L}$  with eigenvalue  $\lambda \approx 1$ . As  $\lambda$  keeps increasing, meaning  $\mathbf{s}$  aligns with more high-frequency eigenvectors, the moderating effect of neighbors starts to become more important, resulting in a decrease of external conflict as well as the controversy. Essentially, the conflict is increasingly internalized in a network where neighbors often have different internal opinions.

<sup>3</sup> Note that the conservation law is reflected in this figure in the following equality, as can be visually verified from Fig. 1:  $\lambda_{ic} + 2\lambda_{ec} + \lambda_c = 1$ .

### 3.3 Conflict Risk of a Network

The measures from Sec. 3.1 quantify the various types of conflict given an internal opinion vector  $\mathbf{s}$ . Prior work (see Sec. 6) has focused on tweaking the network or the opinions of a selection of individuals to reduce such measures. Often, however, the internal opinions are not accessible. More fundamentally, one might wish to minimize conflict on more than one, including yet unknown controversial issues. We therefore argue that it is more sensible to engineer a network so as to reduce the *risk* of conflict, rather than the conflict for one specific internal opinion vector  $\mathbf{s}$ . We propose two ways of quantifying risk of conflict, discussed in turn below.

**Average-case Conflict Risk (ACR).** The ACR is defined as the expected conflict, where the expectation is taken w.r.t. the internal opinions. To evaluate this, a probabilistic model for the internal opinions is needed, and we propose to use the uniform distribution over all vectors from  $\{-1, 1\}^n$ , such that  $E[\mathbf{s}\mathbf{s}^T] = \mathbf{I}$ . Thus:

$$\begin{aligned} \text{ACR}_* &= E[\mathbf{s}^T \mathbf{M}_* \mathbf{s}] = E[\text{Tr}(\mathbf{s}\mathbf{s}^T \mathbf{M}_*)] = \text{Tr}(E[\mathbf{s}\mathbf{s}^T] \mathbf{M}_*), \\ &= \text{Tr}(\mathbf{M}_*). \end{aligned}$$

**Worst-case Conflict Risk (WCR).** This is an alternative (and more robust) measure, defined as the maximum conflict over all possible internal opinion vectors  $\mathbf{s} \in \{-1, 1\}^n$ :

$$\text{WCR}_* = \max_{\mathbf{s} \in \{-1, 1\}^n} \mathbf{s}^T \mathbf{M}_* \mathbf{s}.$$

Note that  $\mathbf{M}_* \geq \mathbf{0}$  for all measures, such that this is an instance of Boolean Quadratic Maximization (BQM) problem [24, 29]. While this problem is NP-hard, it can be approximated by solving the following semidefinite programming (SDP) relaxation of the problem (here,  $\Sigma$  is a symmetric real-valued matrix):

$$\begin{aligned} u_* &= \max_{\Sigma} \text{Tr}(\Sigma \mathbf{M}_*), \\ \text{s.t. } \Sigma &\geq \mathbf{0}, \\ \text{diag}(\Sigma) &= \mathbf{1}. \end{aligned}$$

Nesterov [29] proved that this strategy achieves a  $\frac{2}{\pi}$  approximation:

$$\frac{2}{\pi} u_* \leq \text{WCR}_* \leq u_*.$$

To derive an estimate for the worst-case  $\mathbf{s} \in \{-1, 1\}^n$  from  $\Sigma$ , Goemans and Williamson's randomized rounding strategy [20] can be used: Let  $\Sigma = \mathbf{C}\mathbf{C}^T$  be a Cholesky decomposition of  $\Sigma$ , and let  $\mathbf{x} \in \mathbb{R}^n$  be a randomly sampled vector from some rotation-invariant distribution. Then, for  $\mathbf{s} = \text{sign}(\mathbf{C}\mathbf{x})$ , it holds that (where the expectation is over the random vector  $\mathbf{x}$ ):

$$\frac{2}{\pi} \text{WCR}_* \leq E[\mathbf{s}^T \mathbf{M}_* \mathbf{s}] \leq \text{WCR}_*.$$

I.e., the estimated worst-case opinion vector achieves a conflict that is not smaller than  $\frac{2}{\pi}$  the actual worst-case conflict.

SDPs can be solved in polynomial time:  $O(n^{4.5})$ . While this is still a high complexity, in practice such SDPs can be solved without further optimizations for thousands of nodes on commodity machines, and results for the Maximum Cut problem suggest that scaling is possible much beyond that (to millions of nodes) by exploiting tight approximations, further relaxations, or dedicated optimization approaches [5, 32].

## 4 MINIMIZING THE CONFLICT RISK

### 4.1 Algorithms

Here we discuss how the ACR and WCR can be optimized by adding or deleting edges in the network. Note that only the resistance is known to be convex, such that we should not hope for convergence to a global optimum. Yet, we argue that the question of convexity is purely academic here: in practice, graph edits can typically be made only in small amounts, either because of budget constraints, or because of practical considerations. For example, a company may wish to increase its productivity by organizing a team-building event or reorganizing office space so as to create new conflict-risk reducing connections, but such operations are costly and cannot in practice redesign the complete network structure. Thus, what we should be interested in is a fast decrease of the ACR or WCR given the number of edges added or deleted, rather than eventual convergence to a possible local minimum – let alone a global one.

The edits we consider are edge additions or deletions, or more precisely the increase or decrease of edge weights as long as they remain in the range  $[0, 1]$ . We keep them within this range because it makes no sense to talk about a negative edge strength in social networks, and there is a bound on the strength of connections. Our algorithms can easily be adapted to handle different bounds.

Below, we discuss two algorithmic approaches to this end: one is a conditional gradient method, and suggests a number of edge additions or deletions simultaneously. The other is a coordinate descent method, and suggests adding or deleting just a single edge.

**The optimization problems.** Let  $A_0$  be the initial adjacency matrix, and  $A$  the optimized adjacency matrix with corresponding matrix  $M_*$ . With  $\|\cdot\|_1$  the entry-wise one-norm, the optimization problems for ACR and WCR are thus:

$$\begin{aligned} \text{ACR: } & \min_A \text{Tr}(M_*^T A), \\ & \text{s.t. } 0 \leq A \leq 1, \quad \text{and} \quad \|A - A_0\|_1 \leq 2k. \end{aligned}$$

$$\begin{aligned} \text{WCR: } & \min_A \max_{s \in \{-1, 1\}^n} s^T M_* s, \\ & \text{s.t. } 0 \leq A \leq 1, \quad \text{and} \quad \|A - A_0\|_1 \leq 2k, \end{aligned}$$

where  $k$  is a bound on the sum of absolute values of weight changes (the factor 2 stems from the fact that  $A$  is symmetric). The entry-wise one-norm on  $A - A_0$  ensures this difference tends to be sparse, such that only few edge weights tend to be updated at the minimum.

For the WCR, this problem is complicated by the inner maximization. We handle this optimization problem by alternating optimization: before each conditional gradient or coordinate descent step, we solve the inner maximization as detailed in the previous section, and then assume  $s$  to be fixed. We found however, that robustness of this strategy can be increased by using not a single  $s$ , but a small set of  $\ell$  vectors  $s$  all obtained by randomized rounding. More specifically, written in terms of  $S \in \{-1, 1\}^{n \times \ell}$  containing these different  $s$  vectors as its columns, we solve:

$$\begin{aligned} \text{Robust WCR: } & \min_A \text{Tr}(S^T M_* S), \\ & \text{s.t. } 0 \leq A \leq 1, \quad \text{and} \quad \|A - A_0\|_1 \leq 2k. \end{aligned}$$

Thus, rather than minimizing the risk of conflict for one given worst-case opinion vector, the average over a set of approximately worst-case opinion vectors is minimized. The added robustness

of this strategy stems from the fact that different approximately worst-case opinion vectors can be similarly bad, such that editing the graph to reduce risk for one can increase risk for another. In this case, the alternating minimization would fail. Minimizing the risk averaged over a set approximately worst-case opinion vectors thus increases robustness. Note that for  $S = I$ , the WCR reduces to the ACR. Thus, it suffices to discuss the optimization of the WCR in what follows. Both conditional gradient and coordinate descent first compute the gradient of the ACR and WCR. The gradients for the different measures are summarized in Table 2.

**Conditional gradient descent [13, 23].** The conditional gradient method seeks a step  $\Delta$  most aligned with the gradient, while respecting the constraints after taking a finite step along that direction. More specifically, this step direction is found by solving:

$$\begin{aligned} \min_{\Delta} \text{Tr} \left( \frac{\partial \text{Tr}(S^T M_* S)}{\partial L} \cdot (\text{diag}(\Delta I) - \Delta) \right), \\ \text{s.t. } 0 \leq A + \Delta \leq 1, \quad \text{and} \quad \|\Delta\|_1 \leq 2k', \end{aligned}$$

where  $k' \ll k$  limits the step size. Here, the objective computes the inner product between the gradient with respect to  $L$  and  $\text{diag}(\Delta I) - \Delta$ , as changing  $A$  by adding  $\Delta$  amounts to a step of  $\text{diag}(\Delta I) - \Delta$  on the Laplacian. Note again that these constraints induce sparsity in the solution vector. The experiments indeed confirmed that often  $\Delta$  contains exactly  $2k'$  1's or -1's.

**Coordinate descent.** The coordinate descent method first computes the gradient with respect to the (symmetric) adjacency matrix from the gradient with respect to the Laplacian (as listed in Table 2):

$$\frac{\partial \text{Tr}(S^T M_* S)}{\partial a_{ij}} = \frac{\partial \text{Tr}(S^T M_* S)}{\partial l_{ii}} + \frac{\partial \text{Tr}(S^T M_* S)}{\partial l_{jj}} - 2 \frac{\partial \text{Tr}(S^T M_* S)}{\partial l_{ij}}.$$

Positive  $\frac{\partial \text{Tr}(S^T M_* S)}{\partial a_{ij}}$  means that reducing  $a_{ij} > 0$  will reduce the objective. Conversely, negative  $\frac{\partial \text{Tr}(S^T M_* S)}{\partial a_{ij}}$  means that increasing  $a_{ij} < 1$  will reduce the objective. Thus, the algorithm takes the  $\frac{\partial \text{Tr}(S^T M_* S)}{\partial a_{ij}}$  with largest absolute value for which either  $a_{ij} > 0$  and  $\frac{\partial \text{Tr}(S^T M_* S)}{\partial a_{ij}} > 0$ , or for which  $a_{ij} < 1$  and  $\frac{\partial \text{Tr}(S^T M_* S)}{\partial a_{ij}} < 0$ . In the former case, the algorithm sets  $a_{ij} = a_{ji} = 0$ , and in the latter it sets  $a_{ij} = a_{ji} = 1$ .

**Conditional gradient versus coordinate descent.** The coordinate descent method is computationally obviously easier, but convergence may be slower than with the conditional gradient method. They are compared with each other in the empirical results section.

### 4.2 Local optima of the ACR for different risk measures

As pointed out, only the resistance is known to be convex, such that the ACR and WCR are prone to local minima. Relying on the gradients in Table 2, we can prove the following proposition.

**PROPOSITION 4.1.** *The complete graph forms local minimum for the ACR of conflict measures  $ec$ ,  $c$ , and  $r$ .*

**Table 2: Middle Matrices and Gradients**

*	$\mathbf{M}_*$	ACR: $\frac{\partial \text{Tr}(\mathbf{M}_*)}{\partial \mathbf{L}}$	WCR: $\frac{\partial \text{Tr}(\mathbf{S}^T \mathbf{M}_* \mathbf{S})}{\partial \mathbf{L}}$
$ic$	$(\mathbf{L} + \mathbf{I})^{-2} \mathbf{L}^2$	$2(\mathbf{L} + \mathbf{I})^{-2} - 2(\mathbf{L} + \mathbf{I})^{-3}$	$\mathbf{L}(\mathbf{L} + \mathbf{I})^{-2} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-1} + (\mathbf{L} + \mathbf{I})^{-1} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-2} \mathbf{L}$
$ec$	$(\mathbf{L} + \mathbf{I})^{-2} \mathbf{L}$	$-(\mathbf{L} + \mathbf{I})^{-2} + 2(\mathbf{L} + \mathbf{I})^{-3}$	$(\mathbf{L} + \mathbf{I})^{-2} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-2} - \mathbf{L}(\mathbf{L} + \mathbf{I})^{-2} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-2} \mathbf{L}$
$c$	$(\mathbf{L} + \mathbf{I})^{-2}$	$-2(\mathbf{L} + \mathbf{I})^{-3}$	$-(\mathbf{L} + \mathbf{I})^{-1} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-2} - (\mathbf{L} + \mathbf{I})^{-2} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-1}$
$r$	$(\mathbf{L} + \mathbf{I})^{-1}$	$-(\mathbf{L} + \mathbf{I})^{-2}$	$-(\mathbf{L} + \mathbf{I})^{-1} \mathbf{S} \mathbf{S}^T (\mathbf{L} + \mathbf{I})^{-1}$

**Table 3: Gradient matrix elements for size  $n$  complete graph**

Matrix	Diagonal	Off-diagonal
$-(\mathbf{L} + \mathbf{I})^{-2}$	$-\frac{n+3}{(n+1)^2}$	$-\frac{n+2}{(n+1)^2}$
$-2(\mathbf{L} + \mathbf{I})^{-3}$	$-2\frac{n^2+3n+4}{(n+1)^3}$	$-2\frac{n^2+3n+3}{(n+1)^3}$
$-(\mathbf{L} + \mathbf{I})^{-2} + 2(\mathbf{L} + \mathbf{I})^{-3}$	$\frac{n^2+2n+5}{(n+1)^3}$	$\frac{n^2+3n+4}{(n+1)^3}$

**Table 4: Risks for complete graph of size  $n$**

*	$ic$	$ec$	$c$	$r$
$\text{Tr}(\mathbf{M}_*)$	$\frac{n^2(n-1)}{(n+1)^2}$	$\frac{n(n-1)}{(n+1)^2}$	$\frac{n(n+3)}{(n+1)^2}$	$\frac{2n}{n+1}$

**PROOF.** The adjacency matrix of a size  $n$  complete graph consists of 0 on the diagonal and 1 elsewhere, thus the corresponding Laplacian matrix has  $n - 1$  on the diagonal and  $-1$  elsewhere. In Table 3, the elements in the corresponding ACR gradients with respect to the Laplacian are shown. We will show from these that no feasible step can be found that improves the objectives for a complete graph.

Indeed, for a complete graph (with all weights equal to 1), edge weights can only be decreased. However, decreasing the weight of the edges increases the objective: for a step of  $-\delta$  on  $w_{ij}$ , the external conflict is increased by  $2\frac{n-1}{(n+1)^3}\delta$ , the controversy by  $\frac{4}{(n+1)^3}\delta$  and the resistance by  $\frac{2}{(n+1)^2}\delta$ . For  $n > 1$  these changes are strictly positive, such that the ACR would be increased after decreasing any  $w_{ij}$  by  $1 \geq \delta > 0$ .  $\square$

**Derivative results.** A number of results immediately follow from this proposition. Recall that resistance is convex on  $\mathbf{L}$  [28], so this local minimum is a global one. Furthermore, note that from the conservation law, it follows directly that the gradient of  $ic + 2ec + c$  is equal to 0. Thus, it is trivial to show that for  $ic$ , a complete graph is a local maximum of the ACR. Finally, for a complete graph of size  $n$  (i.e., the number of node is  $n$ ,  $n > 1$ ), the values of ACR for different conflict measures are given in Table 4. Using this table, it can be shown that larger complete graph has smaller conflict risks than two smaller complete graphs with the same total number of nodes. For complete graphs of size  $n_1$ ,  $n_2$ , and  $n_1 + n_2$  ( $n_1, n_2 \geq 3$ ,  $n_1, n_2 \in \mathbb{Z}$ ),

$$\text{Tr}[\mathbf{M}_{ec}(n_1)] + \text{Tr}[\mathbf{M}_{ec}(n_2)] > \text{Tr}[\mathbf{M}_{ec}(n_1 + n_2)].$$

(As long as  $n_1 n_2 - n_1 - n_2 - 3 \geq 0$ , the above inequality holds, which can be proved using  $\text{Tr}(\mathbf{M}_{ec})$  in Table 4.)

**Table 5: Dataset summary statistics.**

Network	Karate	Facebook	ER	BA	WS
Nodes	34	4039	$n$	$n$	$n$
Edges	78	88234	$m$	$m_1$	$\frac{nK}{2}$
Avg degree	4.5882	43.6910	$\frac{2m}{n}$	$\frac{2m_1}{n}$	$K$

**Table 6: ACR for random networks of size  $n = 1000$ ,  $m \approx 5000$ .**

ACR	ic	ec	c	r
ER	796.6	94.1	15.3	109.3
BA	759.3	109.7	21.2	131.0
WS	804.2	91.2	13.3	104.5

We also showed empirically that for the  $ec$  a set of disconnected components are optimal where each component is either a clique, a sufficiently long chain, or a tree where each leaf node is separated by at least two edges from a bifurcation node (see Sec. 5 for details).

## 5 EMPIRICAL EVALUATION

### 5.1 Datasets

We use real social networks as well as synthetic data shown in Table 5. The real-world datasets we use are the Karate network with 34 nodes and a Facebook network consisting of 4039 users. The Karate network is a social network of friendships between 34 members of a Karate club [33]. The Facebook network contains friend circles and was collected through the Facebook app surveys [22].

The synthetic data includes three random network models: Erdős-Rényi (ER) random networks with binomial degree distribution; Barabási-Albert (BA) random networks with power-law degree distribution; and Watts-Strogatz (WS) small world random networks.

### 5.2 Experimental Findings

We investigate the following questions: (1) What types of networks have the highest risks for what types of conflict measures; (2) What are the local minima of the ACR for the various measures; (3) For the external conflict: how do the actual conflict, ACR, and WCR evolve as the ACR or WCR is being minimized; (4) How do the coordinate and conditional gradient descent methods compare for the external conflict. Due to space constraints, some results are summarized, and details are deferred to a forthcoming extended report. <sup>4</sup>

**5.2.1 Conflict risk for different measures in random networks.** We investigated how the ACR for different conflict measures compare to each other across ER, BA, and WS models. We generated random networks of very similar sizes and densities according to these models, and we compared their ACR for different conflict measures.

Across a wide range of graph densities, the WS network is consistently the most high-risk for  $ic$ , while the BA network is consistently the most high-risk for  $ec$ . For  $c$  and  $r$  the most high-risk network depends on the density, although usually the BA or ER networks carry the highest risk. Table 6 gives an example.

These findings can be interpreted in terms of the properties of the random network models. In the WS network, the  $ic$  is probably

<sup>4</sup>All code is available at <https://bitbucket.org/ghentdatascience/conflict-risk-public>

high due to the short path lengths and high clustering coefficient, which causes opinions to be strongly moderated. In the BA network, the existence of high-degree *hubs* along with a fat tail of small-degree nodes may cause considerable *ec* between these hubs (which are strongly moderated) and their surrounding nodes (which are moderated only by very few nodes).

**5.2.2 Empirical study of the local optima of ACR with different conflict measures.** We used the coordinate and conditional gradient descent methods to optimize the ACR (i.e.,  $\text{Tr}(\mathbf{M}_*)$ ) until convergence, to investigate the structure of the network at the local minima. The following findings complement and corroborate the theoretical analysis of the local minima from the previous section.

**Internal Conflict** In our experiments, after convergence the network always contains no edges. As in that case internal and expressed opinions coincide, the *ic* is then equal to zero, this is obviously the global minimum.

**External Conflict** In our experiments, the local minima always contained sets of disconnected subgraphs that are cliques, trees, and chains, and sometimes cliques with a chain attached to one of its nodes. Yet, the particular local minimum found differs for different initial graphs, and also slightly for the different algorithms and choices of  $k'$ .

**Controversy** The local minimum found is always the completely connected graph. While this problem is not known to be convex, we conjecture that it has only one local minimum.

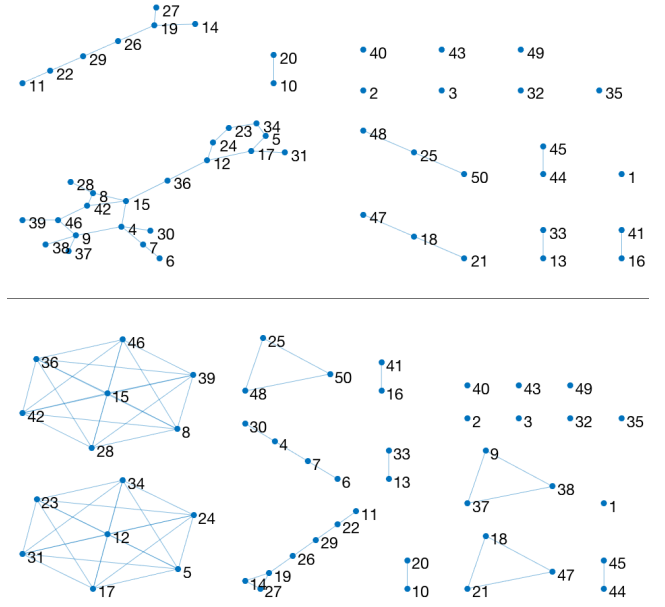
**Resistance** We know from theory that this ACR minimization for resistance is a convex problem. Thus, the minimum found is always the global minimum, namely the complete graph.

Clearly the *ec*, which is arguably the most relevant among the conflict measures in practice, also exhibits the most complex behavior. One example of how the network changes when minimizing the *ec* is shown in Fig. 2, where the bottom network is the local minimum for the network on the top. Typical adjustments during both the coordinate descent and the conditional gradient algorithm are: a chain of three nodes always forms a triangle (see node 25, 48, 50); two nodes at the same end of a chain/tree will always be connected (see node 14, 27); connections that are not strong enough will break (see node 36 between node 12 and 15).

**REMARK 2.** Interestingly, the structures at the local optima of the ACR for *ec* seem to correspond with common management structures in companies: a flat organization corresponds to a clique, while a hierarchical organization corresponds to a tree. Management practice may well have evolved this way in part because it minimizes conflict.

In the sequel, for conciseness we focus on the *ec* alone, as this is arguably the most useful and most interesting measure.

**5.2.3 Effectiveness of minimizing ACR versus WCR for *ec*.** Here we investigate the effectiveness of both ACR and WCR. In particular, we investigated on one ER network and the Karate network how the ACR, WCR, and the conflict for three different internal opinion vectors, evolved over consecutive iterations. The three fixed opinion vectors include a random vector  $\mathbf{s}_1$ , and two vectors found as  $\text{sign}(v)$  where  $v$  is an eigenvector of the Laplacian: the 10th smallest (i.e. low-frequency on the graph,  $\mathbf{s}_2$ ) as well as the  $n - 10$ th (i.e. high-frequency on the graph,  $\mathbf{s}_3$ ).



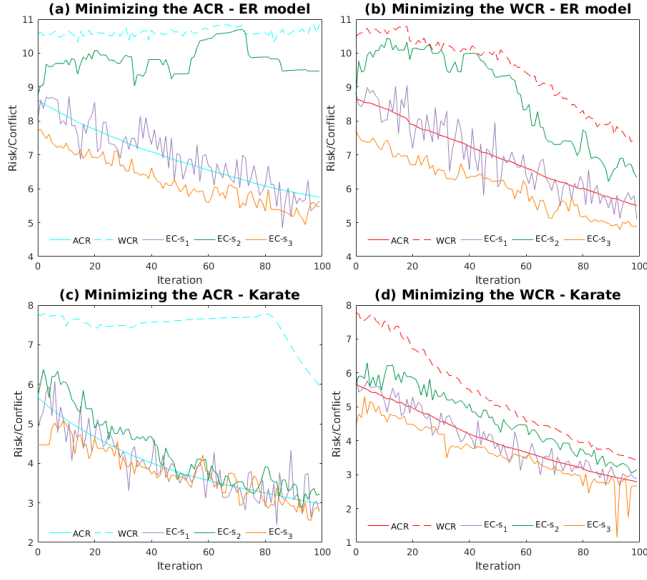
**Figure 2: Optimization of the ACR of *ec* on an ER network ( $n = 50, p = 0.03$ ) with gradient descent ( $k' = 2$ ).**

Figure 3 shows that the optimization for ACR will not necessarily improve the WCR, and also does not improve the *ec* for the low-frequency vector  $\mathbf{s}_2$ , while the optimization over the WCR always decreases also the ACR and the risks for all three given opinion vectors. The fact that the WCR is an upper bound for the ACR as well as for the conflict for any given internal opinion vector probably explains this. Yet, it is remarkable that minimizing the more robust measure WCR does not seem to reduce much the rate at which also the ACR reduces.

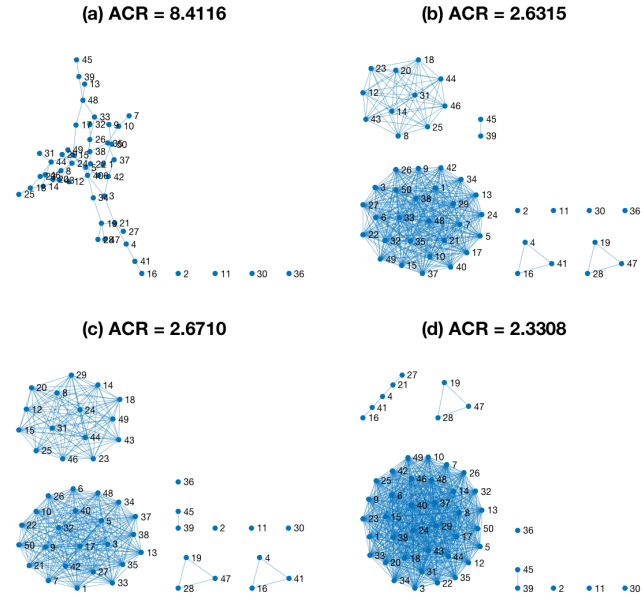
**5.2.4 How does the performance of conditional gradient descent compare to that of coordinate descent?** The following experiment illustrates our observation that conditional gradient descent typically converges to a better local minimum than coordinate descent. This may be because conditional gradient descent can make larger steps at each iteration, thus allowing it to escape bad local minima more easily. Figure 4 shows an example of their different performances, which is consistent with our theoretical conclusion in Sec. 4 about local optima structures, i.e., larger complete graphs contains less external conflict ACR than smaller ones adding to the same size.

**5.2.5 Real-world networks.** For space reasons, a detailed analysis of real-world networks is postponed to an extended report, but we summarize the main findings here. The ACR for the Karate network is minimized by forming a complete network for *ec*, *c* and *r*, and the network without edge for *ic*. Connections within the ‘friend circles’ in the Facebook network are found to be strengthened when minimizing the *ec* ACR, while those between circles are gradually deleted: the originally connected network is ultimately divided into several connected components as the optimization continues. It takes 3 to 5 seconds for one coordinate step on Facebook dataset at





**Figure 3: The ACR, WCR, and conflict for the three described internal opinion vectors over consecutive iterations. (a), (b) are based on an ER model ( $n = 50, m = 60$ ) with gradient descent  $k' = 1$ ; (c), (d) on Karate with coordinate descent.**



**Figure 4: Optimal results using the two algorithms. (a) is the original graph; (b) is the result of coordinate descent; (c) is the result of gradient descent with  $k' = 5$  at each iteration; (d) is the result of gradient descent with  $k' = 25$ .**

the beginning and the time increases as edges are added, which is acceptable in practice.

## 6 RELATED WORK

Social network analysis research almost invariably relies on data from online social media and microblogging sites. In particular Twitter [9, 15, 21, 27] is often the scene of controversial debates. Notable studies are Conover et al., who performed research on the retweet and mention networks from Twitter, and differentiated between the two mechanisms [9], and Garimella et al., who used conversation graphs obtained from twitter to quantify controversy for any topic [15]. While online social media expose the users to various kinds of opinions, the effects of ‘filter bubbles’ and ‘echo chamber’ have attracted increasing amounts of attention in recent years [4, 18]: when people only get information that corroborates their own opinions and communicate only with like-minded people, there is a risk that society will be increasingly fragmented and polarized, although there is an ongoing debate about this issue [18].

Research about polarization and controversy has so far mostly focused on political issues. Morales et al. studied the emergence of political polarization and quantified its effects by a polarization index [27]. Akoglu quantified the political polarity of individuals and political issues by doing classification and ranking tasks [1]. It defines a node classification task on edge-signed (+/-) bipartite opinion network, then predicts latent political classes of people and opinion subjects and ranks people and issues.

Opinion formation models are not always used; some prior work focuses on the underlying structure of the social network, or assumes there are only two groups for ‘pro’ and ‘contra’. Coletto et al. used only local patterns of user interactions (motifs) [8]. Guerra et al. focused on the nodes in the community boundaries [21]. Random Walk Controversy (RWC) scores are used to quantify controversy in [15] as the difference between the properties of a random walk ending in different opinion partitions. Amin et al. studied the problem of identifying and separating polarization using a matrix factorization based gradient descent algorithm [3].

Different measures have been proposed for quantifying polarization or controversy. Modularity is regarded as a traditional measure for polarization [30], but Guerra et al. argue that it is not a good measure since non-polarized networks may also be divided into modular communities in [21]. Then they proposed their novel polarization metric  $P$  based on boundary nodes and found that polarized networks tend to have low concentration on high-degree nodes in the boundary between two communities. The Social Network Distance (SND) is a distance measure that quantifies the likelihood of evolution of one snapshot of a social network into another snapshot under a chosen opinion dynamic model in [2]. To quantify controversy in social networks in any topic domain, a three-step pipeline is proposed in [15]. It was found that the RWC outperformed many other controversy measures, including the betweenness, embedding, boundary connectivity, and dipole moment.

A major and increasingly important focus of research is whether polarization and controversy can be *engineered*, e.g. by editing the graph or affecting opinions of a selected set of individuals. In [17], the edge-recommendation problem is studied based on the endorsement graph, with the goal to reduce the controversy score (namely the RWC), and the acceptable probability of the recommended edge is taken into account. The addition of edges is discussed in [6] in order to reduce the social cost, namely the lack of agreement in



the network, and it is argued as intuitive because the exposure to opposite opinions can reduce disagreement. The expressed opinion vector  $\mathbf{z}$  above is obtained at the Nash Equilibrium in the social game of opinion formation [6]. Moreover, they firstly studied the problem of moderating people's opinions to reduce the polarization. Based on the same opinion formation dynamics, the promotion problem called the **CAMPAIGN** was studied in [19]. It aimed to promote a product by setting the expressed opinions of  $k$  nodes to 1 such that the overall opinions  $g(\mathbf{z})$  over the network can be maximized. The expressed opinion  $z_i$  represents the affection of node  $i$  for the product, and it lies in the range from 0 to 1. This work provides a good example of shifting from the problem of measuring opinion differences to the area of influence maximization.

## 7 CONCLUSIONS AND FURTHER WORK

Research into the formation of conflict, disagreement, and related concepts was until recently the subject of the social sciences only. Today however, the fact that opinion formation takes place increasingly on online social platforms creates new possibilities to address related issues from a computer science perspective, building on models of opinion formation from the social sciences. Specifically, it creates the potential to quantify, mitigate, and reduce conflict and disagreement. Prior research on this topic has focused on a single issue of controversy, and the reduction of conflict on this issue, in particular by manipulating the structure of the network.

In this paper we included a small survey of existing measures, and identified an insightful identity between them that amounts to a conservation law of conflict. However, we also argued that reducing one of these measures of conflict for a single issue is problematic, reducing conflict on a single issue may increase it for another. Indeed, in practice a network is not tied to a single issue, and even when it is, the individual opinions may be hard to gauge. To resolve this, we take a novel perspective on this problem, focusing on identifying a limited number of edges to add or remove in the network so as to reduce the *risk* of conflict, both on average and in the worst-case over all possible opinions. We have demonstrated the usefulness of these characterizations of conflict risk, studied their behavior in a range of networks, developed effective algorithms for optimizing them, and confirmed that their minimization minimizes actual risk on some random opinion assignments.

In further work, we plan to investigate further the theoretical properties of these measures, in particular of the worst-case risk. Additionally, we plan to improve our implementations and investigate other algorithmic improvements for enhanced scalability.

**Acknowledgements.** This work was supported by the ERC under the EU's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agr. no. 615517, FWO (project no. G091017N, G0F9816N), the EU's Horizon 2020 research and innovation programme and the FWO under the Marie Skłodowska-Curie Grant Agr. no. 665501.

## REFERENCES

- [1] Leman Akoglu. 2014. Quantifying political polarity based on bipartite opinion networks. In *Proc. of ICWSM*.
- [2] Victor Amelkin, Petko Bogdanov, and Ambuj K Singh. 2017. A distance measure for the analysis of polar opinion dynamics in social networks. In *Proc. of ICDE*. 159–162.
- [3] Md Tanvir Al Amin, Charu Aggarwal, Shuochao Yao, Tarek Abdelzaher, and Lance Kaplan. 2017. Unveiling polarization in social networks: A matrix factorization approach. In *Proc. of INFOCOM*. 1–9.
- [4] Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
- [5] Tijl De Bie and Nello Cristianini. 2006. Fast SDP relaxations of graph cut clustering, transduction, and other combinatorial problems. *Journal of Machine Learning Research* 7 (2006), 1409–1436.
- [6] David Bindel, Jon Kleinberg, and Sigal Oren. 2015. How bad is forming your own opinion? *Games and Economic Behavior* 92 (2015), 248–265.
- [7] Yoonjung Choi, Yuchul Jung, and Sung-Hyon Myaeng. 2010. Identifying controversial issues and their sub-topics in news articles. *Intelligence and Security Informatics* (2010), 140–153.
- [8] Mauro Coletto, Kiran Garimella, Aristides Gionis, and Claudio Lucchese. 2017. A motif-based approach for identifying controversy. arXiv:1703.05053. (2017).
- [9] Michael Conover, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political polarization on twitter. In *Proc. of ICWSM*. 89–96.
- [10] Pranav Dandekar, Ashish Goel, and David T Lee. 2013. Biased assimilation, homophily, and the dynamics of polarization. *PNAS* 110, 15 (2013), 5791–5796.
- [11] Morris H DeGroot. 1974. Reaching a consensus. *JASA* 69, 345 (1974), 118–121.
- [12] W Ellens, FM Spieksma, P Van Mieghem, A Jamakovic, and RE Kooij. 2011. Effective graph resistance. *Linear algebra and its applications* 435, 10 (2011), 2491–2506.
- [13] Marguerite Frank and Philip Wolfe. 1956. An algorithm for quadratic programming. *Naval Research Logistics* 3, 1-2 (1956), 95–110.
- [14] Noah E Friedkin and Eugene C Johnsen. 1990. Social influence and opinions. *Journal of Mathematical Sociology* 15, 3-4 (1990), 193–206.
- [15] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying controversy in social media. In *Proc. of ICWSM*. 33–42.
- [16] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. Reducing controversy by connecting opposing views. In *Proc. of ICWSM*. 81–90.
- [17] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Balancing opposing views to reduce controversy. arXiv:1611.00172. (2016).
- [18] R Kelly Garrett. 2009. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication* 14, 2 (2009), 265–285.
- [19] Aristides Gionis, Evimaria Terzi, and Panayiotis Tsaparas. 2013. Opinion maximization in social networks. In *Proc. of SDM*. 387–395.
- [20] Michel X Goemans and David P Williamson. 1995. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *JACM* 42, 6 (1995), 1115–1145.
- [21] Pedro Henrique Calais Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. 2013. A measure of polarization on social media networks based on community boundaries. In *Proc. of ICWSM*. 215–224.
- [22] Jure Leskovec and Julian J McAuley. 2012. Learning to discover social circles in ego networks. In *Proc. of NIPS*. 539–547.
- [23] Evgeny S Levitin and Boris T Polyak. 1966. Constrained minimization methods. *USSR Computational mathematics and mathematical physics* 6, 5 (1966), 1–50.
- [24] Zhi-Quan Luo, Wing-Kin Ma, Anthony Man-Cho So, Yinyu Ye, and Shuzhong Zhang. 2010. Semidefinite relaxation of quadratic optimization problems. *IEEE Signal Processing Magazine* 27, 3 (2010), 20–34.
- [25] Antonis Matakos, Evimaria Terzi, and Panayiotis Tsaparas. 2017. Measuring and moderating opinion polarization in social networks. *DMKD* 31, 5 (2017), 1480–1505.
- [26] Yelena Mejova, Amy X Zhang, Nicholas Diakopoulos, and Carlos Castillo. 2014. Controversy and sentiment in online news. arXiv:1409.8152. (2014).
- [27] AJ Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25, 3 (2015), 033114.
- [28] Cameron Musco, Christopher Musco, and Charalampos E Tsourakakis. 2017. Minimizing polarization and disagreement in social networks. arXiv:1712.09948. (2017).
- [29] Yu Nesterov. 1998. Semidefinite relaxation and nonconvex quadratic optimization. *Optimization methods and software* 9, 1-3 (1998), 141–160.
- [30] Mark EJ Newman. 2006. Modularity and community structure in networks. *PNAS* 103, 23 (2006), 8577–8582.
- [31] David I Shuman, Narang Sunil K, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. 2013. The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine* 30, 3 (2013), 83–98.
- [32] Po-Wei Wang and J Zico Kolter. 2016. The mixing method for maxcut-SDP problem. In *NIPS LHDS Workshop*.
- [33] Wayne W Zachary. 1977. An information flow model for conflict and fission in small groups. *Journal of Anthropological Research* 33, 4 (1977), 452–473.