Reducing Controversy by Connecting Opposing Views

Kiran Garimella
Aalto University
Helsinki, Finland
kiran.garimella@aalto.fi

Aristides Gionis
Aalto University
Helsinki, Finland
aristides.gionis@aalto.fi

Gianmarco De Francisci Morales
Qatar Computing Research Institute
Doha, Qatar
gdfm@acm.org

Michael Mathioudakis
HIIT, Aalto University
Helsinki, Finland
michael.mathioudakis@hiit.fi

ABSTRACT

Society is often polarized by controversial issues that split the population into groups with opposing views. When such issues emerge on social media, we often observe the creation of 'echo chambers', i.e., situations where like-minded people reinforce each other's opinion, but do not get exposed to the views of the opposing side. In this paper we study algorithmic techniques for bridging these chambers, and thus reduce controversy. Specifically, we represent the discussion on a controversial issue with an *endorsement graph*, and cast our problem as an *edge-recommendation problem* on this graph. The goal of the recommendation is to reduce the *controversy score* of the graph, which is measured by a recently-developed metric based on random walks. At the same time, we take into account the *acceptance probability* of the recommended edge, which represents how likely the edge is to materialize in the endorsement graph.

We propose a simple model based on a recently-developed user-level controversy score, that is competitive with state-of-the-art link-prediction algorithms. Our goal then becomes finding the edges that produce the largest reduction in the controversy score, in expectation. To solve this problem, we propose an efficient algorithm that considers only a fraction of all the possible combinations of edges. Experimental results show that our algorithm is more efficient than a simple greedy heuristic, while producing comparable score reduction. Finally, a comparison with other state-of-the-art edge-addition algorithms shows that this problem is fundamentally different from what has been studied in the literature.

1. INTRODUCTION

Polarization around controversial issues is a well studied phenomenon in the social sciences [19, 36]. Social media have arguably eased the emergence of such issues, thanks to the scale of discussions and the publicity they foster. This paper studies how to reduce the polarization in controversial issues on social media by creating bridges across opposing sides.

We focus on controversial issues that create discussions online. Usually, these discussions involve a fair share of "retweeting" or "sharing" opinions of authoritative figures that the user agrees with. Therefore, it is natural to model the discussion as an *endorsement graph*: a vertex v represents a user, and a directed edge (u, v) represents the fact that user u endorses the opinion of user v.

Given this modus operandi, and the existence of confirmation bias, homophily, selective exposure, and related social phenomena in human activities, the existence of echo chambers online is not surprising [13, 8]. The existence of these chambers is a hindrance to the democratic process and to the functioning of society at large, as they cultivate isolation and misunderstanding across sections of the society.

A solution to this problem is to create *bridges* that connect people of opposing views. By putting different parts of the endorsement graph in contact, we hope to reduce the polarization of the discussion the graph represents.

We operationalize this concept by leveraging recent advances in quantifying online controversy [11]. In particular, given a metric that measures how controversial an issue discussed on social media is, our goal is to find a small number of edges, called bridges, that minimize this measure. That is, we seek to propose (content produced by) a user v to another user u, hoping that u endorses v by spreading their opinion. This action would create a new edge (a bridge) in the endorsement graph, thus reducing the controversy score of the graph itself.

Clearly, some bridges are more likely to materialize than others. For instance, people in the 'center' might be easier to convince than people on the two extreme ends of the political spectrum [22]. We take this issue into account by modeling an *acceptance probability* for a bridge as a separate component of the model. This component can be implemented by any generic link-prediction algorithm that gives a probability of materialization to each non-existing edge. However, we propose a simple model based on a recently developed user-level controversy score [12] which nicely captures the dynamics and properties of the endorsement graph. Therefore, we seek bridges that minimize the *expected* controversy score, according to their acceptance probabilities.

The core of this paper is an algorithm to solve the aforementioned problem. We show that a brute-force approach is not only unfeasible, as it requires one to evaluate a combinatorial number of candidates, but also unnecessary. Moreover, our algorithm needs to consider far fewer than the $\mathcal{O}(n^2)$ possible edges (where n is the number of vertices in the graph) needed by a simple greedy heuristic.

Experimental results show that our algorithm is able to minimize the controversy score of a graph efficiently and as effectively as the greedy algorithm. In addition, they show that previously-proposed methods for edge addition that optimize for different objective functions are not applicable to the problem at hand.

In summary, our contributions are the following:

- We study the problem of bridging echo chambers algorithmically, in a language and domain agnostic way for the first time. Previous studies that try to address this problem focus mostly on understanding *how* to recommend content to an ideologically opposite side, while our focus is on *who* to recommend contrarian content to. We believe that the two approaches complement each other in bringing us closer to bursting the filter bubble.

- We build on top of results from recent user studies [28, 23, 39] on how users prefer to consume content from opposing views, and formulate the task as an edge-recommendation problem in an endorsement graph, while also taking into account the acceptance probability of a recommendation.

- We provide a method to estimate the acceptance probability of a recommendation that fits well in this setting.

- We propose an efficient algorithm to solve the problem, which considers fewer candidates than a greedy baseline.

- We extensively evaluate the proposed algorithm on real-world data, and demonstrate that it outperforms several sensible baselines.

2. RELATED WORK

Making recommendations to decrease polarization. The Web offers the opportunity to easily access any kind of information. Nevertheless, several studies have observed that, when offered choice, users prefer to be exposed to agreeable and like-minded content. For instance, Liao et al. [21] report that "*even when opposing views were presented side-to-side, people would still preferentially select information that reinforced their existing attitudes.*" This selective-exposure phenomenon has led to increased fragmentation and polarization online. A wide body of recent studies have studied [2, 7, 25] and quantified [3, 11, 17, 26] this divide.

Given the ill-fated consequences of polarization on society [31, 37], it is well-worth investigating whether online polarization and filter bubbles can be avoided. One simple way to achieve this is to "nudge" individuals towards being exposed to opposing view-points, an idea that has motivated several pieces of work in the literature.

Liao and Fu [22, 23] attempt to limit the *echo chamber effect* by making users aware of other users' stance on a given issue, the extremity of their position, and their expertise. Their results show that participants who seek to acquire more accurate information about an issue are exposed to a wider range of views, and agree more with users who express moderately-mixed positions on the issue.

Vydiswaran et al. [39] perform a user study aimed to understand ways to best present information about controversial issues to users so as to persuade them. Their main relevant findings reveal that factors such as showing the credibility of a source, or the expertise of a user, increases the chances of other users believing in the content. In a similar spirit, Munson et al. [28] create a browser widget that measures and displays the bias of users based on the news articles they read. Their study concludes that showing users their bias nudges them to read articles of opposing views.

Graells-Garrido et al. [15] show that mere display of contrarian content has negative emotional effect. To overcome this effect, they propose a visual interface for making recommendations from a diverse pool of users, where diversity is with respect to user stances on a topic. In contrast, Munson et al. [27] show that not all users value diversity and that the way of presenting information (e.g., highlighting vs. ranking) makes a difference in the way users perceive information. In a different direction, Graells-Garrido et al. [16] propose to find "intermediary topics" (i.e., topics that may be of interest to both sides) by constructing a *topic graph*. They define intermediary topics to be those topics that have high betweenness centrality and topic diversity.

Based on the papers discussed above, we make the following observations:

(a) Although several studies have been proposed to solve the problem of decreasing polarization, there is a lack of an algorithmic approach that works in a domain- and language-independent manner. Instead, the approaches listed above are mostly based on user studies or hand-crafted datasets. To our knowledge, this paper is the first to offer such an algorithmic approach.

(b) Additionally, the studies discussed above focus mostly on understanding *how* to recommend content to an ideologically opposite side. Instead, the approach presented in this paper deals with the problem of finding *who* to recommend contrarian content to. Combining the two approaches can bring us a step closer to bursting the filter bubble.

(c) The studies discussed above suggest that (i) it is possible to nudge people by recommending content from an opposing side [28], (ii) extreme recommendations might not work [16], (iii) people "in the middle" are easier to convince [22], (iv) expert users and hubs are often less biased and can play a role in convincing others [23, 39]

In the design of our algorithm we explicitly take into account these considerations (i)–(iv).

Adding edges to modify the graph structure. In addition to the work on explicitly reducing polarization in social media, there are several papers which aim to make a graph more cohesive by adding edges, where cohesiveness is quantified using graph-theoretic properties such as shortest paths [32, 30], closeness centrality [33], diameter [9], eccentricity [34], communicability [4, 5], synchronizability [40], and natural connectivity [6].

The paper conceptually closest to ours is the one by Tong et al. [38], which aims to add and remove edges in a graph to reduce the dissemination of content (e.g., viruses). The proposed approach maximizes the largest eigenvalue, which determines the epidemic threshold, and thus the properties of information dissemination in networks.

The similarity of the above-mentioned approaches to our paper is limited to the fact that the goal is to modify a graph by edge additions. However, our proposed approach and objective function is predominantly different from those found in other works.

3. PRELIMINARIES AND PROBLEM DEFINITION

To ensure an algorithmic approach to identifying controversial issues and selecting which edges to recommend in order to reduce controversy in a social network, we need to rely on a measure of controversy. As reviewed in Section 2, there are several measures for quantifying controversy in social media [2, 3, 7, 11, 25, 26]. In this paper, we adopt the controversy measure proposed by Garimella et al. [11], as it is the most recent work and it was shown to work reliably in multiple domains; in contrast, other measures focus on a single topic (usually politics) or require domain-specific knowledge. We revise the proposed measure and modify its formulation to adapt it to our current problem. The adopted controversy measure consists of the following steps [11]:

(i) Given a topic t for which we want to quantify its controversy level, we create an *endorsement graph* $G = (V, E)$. This graph represents users who have generated content relevant to t. For instance, if t is specified by a hashtag, the vertices of the endorsement graph are the set of all users who have used this hashtag. The edges of the endorsement graph are defined by the *retweets* among the users, in order to capture user-to-user endorsement.

(ii) The vertices of the endorsement graph $G = (V, E)$ are partitioned into two disjoint sets X and Y, i.e., $X \cup Y = V$ and $X \cap Y = \varnothing$. The partitioning is based on the graph structure and it is obtained using a graph-partitioning algorithm. The intuition is that, for controversial topics, the partitions X and Y are well separated and correspond to the opposing sides of the controversy.

(iii) The last step of computing the controversy measure relies on a *random walk*. In particular, the measure, called *random-walk controversy* (RWC) score, is defined as the difference of the probability that a random walk starting on one side of the partition will stay on the same side and the probability that the random walk will cross to the other side. This measure is computed via two personalized PageRank computations, where the probability of restart is set to a random vertex on each side, and the final probability is taken by considering the stationary distribution of only the high-degree vertices.

In more detail, let P be the column-stochastic transition probability matrix for the random walk, and let X^* and Y^* be the sets of the k_1, k_2 highest in-degree vertices of the two partitions X and Y, respectively. Let r_X be the personalized PageRank vector for the random walk starting in X with restart vector $e_X = \text{uniform}(X)$ and restart probability $(1 - \alpha) \in [0, 1]$, and similarly for r_Y.

Let P_X and P_Y be the transition matrices corresponding to the two random walks starting from the corresponding side. Note that if there are no dangling vertices in the graph then $P_X = P_Y = P$. In the case of dangling vertices, following standard practice, the matrices P_X and P_Y are defined so that the transition probabilities from the dangling vertices are equal to the restart vectors e_X and e_Y, respectively. The personalized PageRank for the two random walks (starting in X and starting in Y) is given by equations:

$$r_X = \alpha P_X r_X + (1 - \alpha) e_X$$
$$r_Y = \alpha P_Y r_Y + (1 - \alpha) e_Y. \tag{1}$$

Let c_X be a vector of size n having value 1 in the coordinates that correspond to the high-degree vertices X^* and 0 else-where, and similarly define c_Y. The random-walk controversy score $\text{RWC}(G, X, Y)$ is defined as:

$$\text{RWC}(G, X, Y) = (c_X{}^\mathsf{T} r_X + c_Y{}^\mathsf{T} r_Y) - (c_Y{}^\mathsf{T} r_X + c_X{}^\mathsf{T} r_Y)$$
$$= (c_X - c_Y)^\mathsf{T}(r_X - r_Y). \tag{2}$$

By using Equations (1), Equation (2) can be written as:

$$\text{RWC}(G, X, Y) =$$
$$(1 - \alpha)(c_X - c_Y)^\mathsf{T}((I - \alpha P_X)^{-1} e_X - (I - \alpha P_Y)^{-1} e_Y),$$

or

$$\text{RWC}(G, X, Y) = (1 - \alpha)(c_X - c_Y)^\mathsf{T}(M_X^{-1} e_X - M_Y^{-1} e_Y), \tag{3}$$

for $M_X = (I - \alpha P_X)$ and $M_Y = (I - \alpha P_Y)$.

Given the controversy measure $\text{RWC}(G)$, the problem we consider in this paper can be formulated as follows.

PROBLEM 1 (k-EDGEADDITION). *Given a graph $G(V, E)$ whose vertices are partitioned into two disjoint sets X and Y ($X \cup Y = V$ and $X \cap Y = \varnothing$), and an integer k, find a set of k edges $E' \subseteq V \times V \setminus E$ to add to G and obtain a new graph $G' = (V, E \cup E')$, so that the controversy score $\text{RWC}(G', X, Y)$ is minimized.*

Note that the two partitions X and Y are considered fixed and part of the input. We also consider the high-degree vertices on which the score depends the same in G and G'.

4. ALGORITHMS

A brute-force approach to solve the problem needs to consider all $\mathcal{O}(\binom{n^2}{k})$ combinations of k possible edges to add. A more efficient greedy heuristic would select k edges in k steps, and at each step evaluate the improvement in the value of RWC given by any of the remaining $\mathcal{O}(n^2)$ edges. Even for the greedy approach, though, the number of possible edges to consider is prohibitively large in real settings. Since computation of the controversy score is an expensive operation, we would like to invoke the function as few times as possible. That is, we aim to consider far fewer candidate edges — ideally sub-linear in real-world settings.

At a high level, the algorithm we propose works as follows. It considers only the edges between the high-degree vertices of each side. For each such edge, it computes the reduction in the RWC score obtained when that edge is added to the original graph. It then selects the k edges that lead to the lowest score when added to the graph individually.

Exemplary case

To motivate the proposed algorithm, we study an exemplary case. We use this case to justify our choice to add edges which connect high-degree vertices across the two sides.

Consider a hypothetical directed graph shown in Figure 1. The graph consists of two disjoint stars, each comprised of n vertices. Intuitively, each star represents one side of the controversy. The center of each star is the highest degree vertex of each side. Following the definition of Problem 1 for $k = 1$, we ask which directed edge we should add in order to minimize the controversy score RWC of the entire graph.

Without loss of generality, we consider the following four cases of edges: (i) from a to c, (ii) from a to d, (iii) from b to c, (iv) from b to d. Among these four edges, the first one,
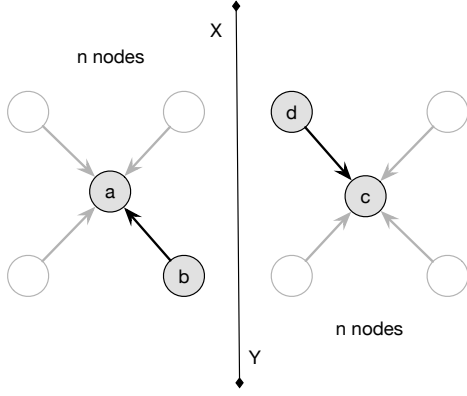
Figure 1: Exemplary case of a graph that consists of two disjoint star-like graphs, each of size n. We wish to add one directed edge so as to minimize the resulting RWC score.

from a to c, connects the two centers of the two stars. We can analytically formulate the RWC score we obtain when each of these edges is added to the original graph, denoted with $s_{a \to c}$, $s_{a \to d}$, $s_{b \to c}$, $s_{b \to d}$, respectively. The respective RWC scores are given by the following formulas (details omitted due to lack of space):

$$s_{a \to c} = \frac{(-\alpha^2 + \alpha) \cdot n + (\alpha - 1)^2}{(\alpha^2 + \alpha + 1) \cdot n - \alpha^2} + \frac{\alpha \cdot n - \alpha + 1}{(\alpha + 1) \cdot n - \alpha}$$

$$s_{a \to d} = \frac{(-\alpha^3 + \alpha) \cdot n + \alpha^3 - \alpha^2 - \alpha + 1}{(\alpha^3 + \alpha^2 + \alpha + 1) \cdot n - \alpha^3} + \frac{\alpha \cdot n - \alpha + 1}{(\alpha + 1) \cdot n - \alpha}$$

$$s_{b \to c} = \frac{2\alpha \cdot n - 3\alpha + 2}{(\alpha + 1) \cdot n - \alpha}$$

$$s_{b \to d} = \frac{\alpha \cdot n - \alpha + 1}{(\alpha + 1) \cdot n - \alpha} + \frac{2\alpha \cdot n - 3\alpha - \alpha^2 + 2}{2(\alpha + 1) \cdot n + \alpha^2 - 2\alpha}$$

THEOREM 1. *For $n \to \infty$, $\alpha \in [0, 1]$, we have*

$$s_{a \to c} \leq s_{a \to d}, s_{b \to c}, s_{b \to d}.$$

PROOF. We have

$$s_{a \to c} \underset{n \to \infty}{\to} \frac{-\alpha^2 + \alpha}{\alpha^2 + \alpha + 1} + \frac{\alpha}{\alpha + 1}$$

$$s_{a \to d} \underset{n \to \infty}{\to} \frac{-\alpha^3 + \alpha}{\alpha^3 + \alpha^2 + \alpha + 1} + \frac{\alpha}{\alpha + 1}$$

$$s_{b \to c} \underset{n \to \infty}{\to} \frac{2\alpha}{\alpha + 1}$$

$$s_{b \to d} \underset{n \to \infty}{\to} \frac{\alpha}{\alpha + 1} + \frac{2\alpha}{2(\alpha + 1)} = \frac{2\alpha}{\alpha + 1}$$

and the inequalitites follow trivially. □

Therefore, the edge from vertex a to vertex c is the one that leads to the minimum score. Theorem 1 provides the optimal edge for a special case. Even though real graphs do not match this case exactly, they often have a structure that resembles star-graphs in certain ways: a small number of

highly popular vertices receive incoming edges (endorsements) from a large number of other vertices. We can think that, in a controversial setting, there are thought leaders and followers. Most activity in the endorsement graph happens when spreading the voice of these leaders across their side. This creates a polarized structure which resembles a union of stars on each side of the controversy. The theorem suggests intuitively that edges between high-degree vertices of either side are good candidates to add to obtain a low RWC score.

The exemplary case described above motivates us to consider edges between high-degree vertices from either side. The algorithm for selecting the edges to be added is shown as Algorithm 1. Its running time is $\mathcal{O}(k_1 \cdot k_2)$, where k_1, k_2 are the number of high-degree vertices chosen in X and Y respectively.

Algorithm 1: Algorithm for k-EDGEADDITION

Input: Graph G, number of edges to add, k; k_1, k_2 high degree vertices in X, Y respectively
Output: List of k edges that minimize the objective function, RWC

1 Initialize: Out ← *empty list* ;
2 **for** $i = 1{:}k_1$ **do**
3   vertex u = X[i];
4   **for** $j = 1{:}k_2$ **do**
5   vertex v = Y[j];
6   Compute δRWC$_{u \to v}$, the decrease in RWC if the edge (u, v) is added;
7   Append δRWC$_{u \to v}$ to Out;
8   Compute δRWC$_{v \to u}$, the decrease in RWC if the edge (v, u) is added;
9   Append δRWC$_{v \to u}$ to Out;
10 sorted ← sort(Out) by δRWC by decreasing order ;
11 **return** top k from sorted;

4.1 Incorporating Acceptance Probabilities

Problem 1 seeks the edges that lead to the lowest RWC score *if added* to the graph. In a recommendation setting, however, the selected edges do not always materialize (e.g., the recommendation might be rejected by the user). In such a setting, it is more appropriate to consider edges that minimize the RWC score *in expectation*, under a probabilistic model $\mathbb{A}$ that provides the probability that a set of edges are accepted once recommended. This consideration leads us to the following formulation of our problem.

PROBLEM 2 (k-EDGEADDITIONEXPECTATION). *Given a graph $G = (V, E)$ whose vertices are partitioned into two disjoint sets X and Y ($X \cup Y = V$ and $X \cap Y = \varnothing$), and an integer k, find a set of k edges $E' \subseteq V \times V \setminus E$ to add to G and obtain a new graph $G' = (V, E \cup E')$, so that the expected controversy score $E_A[\text{RWC}(G', X, Y)]$ is minimized under acceptance model $\mathbb{A}$.*

We build such an acceptance model $\mathbb{A}$ on the feature of *user polarity* proposed by Garimella et al. [12]. Intuitively, this polarity score of a user, which takes values in the interval $[-1, 1]$, captures how much the user belongs to either side of the controversy. High absolute values (close to -1 or 1) indicate that the user clearly belongs to one side of the

controversy, while central values (close to 0) indicate that the user is in the middle of the two sides. We employ user polarity as a feature for our acceptance model because, intuitively, we expect users from each side to accept content from different sides with different probabilities, and we assume these probabilities are encoded in, and can be learned from, the graph structure itself. For example, a user with polarity close to -1 is more likely to endorse a user with a negative polarity than a user with polarity $+1$.

Technically, the polarity score R_u of user u is defined as follows. Let l_u^X and l_u^Y be the expected time a random walk needs to hit the high degree vertices of side X and Y, respectively, starting from vertex u. Moreover, let $\rho^X(u) \in [0, 1]$ and $\rho^Y(u) \in [0, 1]$ be the fraction of other vertices u' for which $l_{u'}^X < l_u^X$ and $l_{u'}^Y < l_u^Y$, respectively. The polarity of user u is then defined as

$$R_u = \rho^X(u) - \rho^Y(u) \in [-1, 1]. \tag{4}$$

Now let u and v be two users with polarity R_u and R_v, respectively. Moreover, assume that u is not connected to v in the current instantiation of the graph. Let $p(u, v)$ be the probability that u accepts a recommendation to connect to v. We estimate $p(u, v)$ from training data. Given a dataset of user interactions, we estimate $p(u, v)$ as the fraction

$$N_{endorsed}(R_u, R_v)/N_{exposed}(R_u, R_v)$$

where $N_{exposed}(R_u, R_v)$ and $N_{endorsed}(R_u, R_v)$ are the number of times a user with polarity R_v was *exposed to* or *endorsed* (respectively) content generated by a user of polarity R_u. $N_{exposed}(R_u, R_v)$ is computed by assuming that if v follows u, v is exposed to all content generated by u. In practice, the polarity scores are bucketed to smooth the probabilities. An experimental evaluation in Section 6.2 shows that polarity scores learned this way predict the existence of an edge across datasets with good accuracy.

For a recommended edge (u, v) from vertex u to vertex v, with acceptance probability $p(u, v)$ and RWC decrease $\delta\text{RWC}_{u \to v}$, the *expected decrease* in RWC when the edge is recommended individually is

$$E(u, v) = p(u, v) \cdot \delta\text{RWC}_{u \to v}.$$

Algorithm 1 can be efficiently extended to target the expected RWC decrease by using Fagin's algorithm [10]. Specifically, we take as input two ranked lists of edges (u, v), one ranked by decreasing $\delta\text{RWC}_{u \to v}$ (as currently produced in the course of Algorithm 1) and another one ranked by decreasing probability of acceptance $p(u, v)$. Fagin's algorithm parses the two lists in parallel to find the edges that optimize the expected decrease $E(u, v)$. We refer the interested reader to the original work for details [10].

5. INCREMENTAL COMPUTATION OF RWC

The RWCscore, as defined in Section 3 can be computed via personalized PageRank, which is usually implemented by power iterations. However, since we are only interested in computing the incremental change in RWC after adding an edge, we propose a new way to efficiently compute it.

Consider the transition probability matrix P. After the addition of one (directed) edge from vertex a to vertex b, only one column of P is affected: the column that corresponds to the origin vertex (a) of the directed edge. Let q be the out degree of a. Specifically, before the addition of the edge, the ath column of the matrix has the following form.

$$P^T = \begin{bmatrix} & & & \dots & & & & \\ & & & \dots & & & & \\ \frac{1}{q} & \frac{1}{q} & \dots & \frac{1}{q} & 0 & 0 & \dots & 0 \\ & & & \dots & & & & \\ & & & \dots & & & & \end{bmatrix} \tag{5}$$

After adding the new outgoing edge from a, the transition probability matrix has the following form,

$$P^{'T} = \begin{bmatrix} & & & \dots & & & & \\ & & & \dots & & & & \\ \frac{1}{q+1} & \frac{1}{q+1} & \dots & \frac{1}{q+1} & \frac{1}{q+1} & 0 & \dots & 0 \\ & & & \dots & & & & \\ & & & \dots & & & & \end{bmatrix} \tag{6}$$

with an additional $\frac{1}{q+1}$ at the bth index, and all other columns of the matrix are unchanged.

Define $u^\mathsf{T} = [0\ 0\ 0\ 0\ \dots\ 1\ 0\ 0\ \dots\ 0]$ (the ath vector of the standard basis of $\mathbb{R}^n$). Similarly, define v^T as a column vector with a 1 at the bth position and 0 elsewhere.

Define z^T: (i) If the outgoing vertex a is not a dangling vertex, as $\frac{1}{q+1}\left[\frac{1}{q}\ \frac{1}{q}\ \frac{1}{q}\ \frac{1}{q}\ \frac{1}{q}\ \dots\ \frac{1}{q}\ -1\ \dots\ 0\ 0\ 0\right]$ (i.e., $\frac{1}{q(q+1)}$ at all non zero neighbor indices, and $\frac{-1}{q+1}$ at the index of the incoming vertex), we can also say that z_x/z_y is the column vector in P_x/P_y corresponding to the outgoing vertex, multiplied by $\frac{1}{q+1}$ and a -1 at the index of the incoming vertex; and, (ii) if the outgoing vertex is a dangling vertex, as $e_x - v$ or $e_y - v$, depending on the side.

The updated transition probability matrix P' is given by:

$$P' = P - zu^\mathsf{T}. \tag{7}$$

Let $M_x = I - \alpha P_x$ and $M_x' = I - \alpha P_x'$. Expanding the formula for M_x', we get

$$M_x' = I - \alpha P_x' = I - \alpha P_x + \alpha z_x u^\mathsf{T} = M_x + \alpha z_x u^\mathsf{T}. \tag{8}$$

Similarly for M_y', $M_y' = M_y + \alpha z_y u^\mathsf{T}$. As we can see, for any single edge addition, RWCcan be re-computed by using only additional vectors that depends on the vertex that is affected. Moreover, the inverse of M_x' (needed in Equation (3)) can be computed efficiently by using the Sherman-Morrison formula.

LEMMA 1 (SHERMAN-MORRISON FORMULA [14]). *Let* $\mathbf{M}$ *be a square* $n \times n$ *invertible matrix and* $\mathbf{M}^{-1}$ *its inverse. Moreover, let* $\mathbf{a}$ *and* $\mathbf{b}$ *be any two column vectors of size* n. *Then, the following equation holds*

$$(\mathbf{M} + \mathbf{a}\mathbf{b}^T)^{-1} = \mathbf{M}^{-1} - \mathbf{M}^{-1}\mathbf{a}\mathbf{b}^T\mathbf{M}^{-1}/(1 + \mathbf{b}^T\mathbf{M}^{-1}\mathbf{a}).$$

Now, from Equation (3), the updated RWC, RWC' is, $\text{RWC}' = (1 - \alpha)(c_x - c_y)^\mathsf{T}(M_x'^{-1}e_x - M_y'^{-1}e_y)$, and the update in RWC can be written as

$$
\begin{aligned}
\delta(\text{RWC}) &= \text{RWC}' - \text{RWC} \\
&= (1 - \alpha)(c_x - c_y)^\mathsf{T}\big((M_x'^{-1}e_x - M_x^{-1}e_x) \\
&\qquad\qquad + (M_y^{-1}e_y - M_y'^{-1}e_y)\big) \\
&= (1 - \alpha)(c_x - c_y)^\mathsf{T}\Big(-\big(\frac{\alpha M_x^{-1}z_x u^\mathsf{T} M_x^{-1}}{1 + \alpha u^\mathsf{T} M_x^{-1}z_x}\big)e_x \\
&\qquad\qquad + \big(\frac{\alpha M_y^{-1}z_y u^\mathsf{T} M_y^{-1}}{1 + \alpha u^\mathsf{T} M_y^{-1}z_y}\big)e_y\Big).
\end{aligned} \tag{9}
$$

Table 1: Datasets statistics: hashtag used to collect dataset, number of tweets, size of retweet graph.

Dataset	# Tweets	Retweet graph	
		$|V|$	$|E|$
#beefban	84 543	1610	1978
#nemtsov	183 477	6546	10 172
#netanyahuspeech	254 623	9434	14 476
#russia_march	118 629	2134	2951
#indiasdaughter	167 704	3659	4323
#baltimoreriots	218 157	3902	4505
#indiana	116 379	2467	3143
#ukraine	287 438	5495	9452
obamacare	123 320	3132	3241
guncontrol	117 679	2633	2672

Table 2: Algorithms explored for link prediction.

Algorithm	Summary	AUC
Vertex polarity	Link recommendation based on vertex polarity	0.79
Adamic-Adar [1]	Link prediction based on number of common neighbors	0.60
Reliability [41]	Block stochastic model	0.66
RAI [35]	Using community detection to improve link prediction	0.60
SLIM [29]	Collaborative filtering recommendation	0.71
FISM [20]	Content-based recommendation	0.66

In light of Equation (9), the costly inverse computation need not be performed in each iteration to compute the updated RWC score. When a new edge is added to the graph, we just compute the vectors z_x, z_y, and u, and use Equation (9) to directly compute the incremental change in RWC, instead of computing the new RWC and taking the difference. The matrix multiplication $M_*^{-1} z_* u^{\mathsf{T}} M_*^{-1}$ can be computed efficiently by grouping the matrices as $(M_*^{-1} z_*)(u^{\mathsf{T}} M_*^{-1})$. As we see in Section 6.6, this approach provides an order of magnitude speed up in the runtime of our algorithm.

6. EXPERIMENTS

In this section, we provide an evaluation of the two algorithms proposed in Section 4. We use the acronym ROV (_recommend opposing view_) to refer to Algorithm 1, and ROV-AP (_recommend opposing view - with acceptance probability_) to refer to its variation that also considers edge acceptance probabilities.

6.1 Datasets

We use Twitter datasets on known controversial issues. The datasets have also been used in previous studies [11, 24]. Dataset statistics are shown in Table 1. Eight of the datasets consist of tweets collected by tracking single hashtags over a small period of time. The remaining two datasets (_obamacare_, _guncontrol_) consist of tweets collected via the Twitter streaming API[1] by tracking the corresponding keywords for two years. We process the datasets and construct _retweet graphs_. We remark that even though all our datasets are from Twitter, our work can be applied on any graph with a clustered structure separating the sides of a controversy.

6.2 Comparison with other link prediction and recommendation systems

Let us first evaluate the choice of using vertex polarity scores to predict edge acceptance (Section 4.1). To perform this evaluation we compare our approach to other state-of-the-art link-prediction algorithms, which are listed in Table 2.

Following Section 4.1, to estimate acceptance probabilities as a function of user polarity, we first bucket the user polarity scores into 10 equally sized buckets, from -1 to +1. Then, we estimate acceptance probabilities $p(u, v)$ separately for each bucket combination of users u and v. We train a model and cross-validate across all datasets. The median AUC is 0.79,

which indicates that endorsement graphs across different datasets have similar edge-formation criteria.

We compare our approach with existing link-recommendation methods. The implementations are obtained from Librec [18]. Table 2 reports the results. As we can see, our approach, which uses vertex polarity scores for predicting links, works as well as the best link-recommendation algorithm. Note that the objective here is not to propose yet another link-recommendation algorithm, nor to claim that our method works better than other approaches in general. Rather, our objective is to validate the use of vertex polarities to create a model for edge-acceptance probabilities.

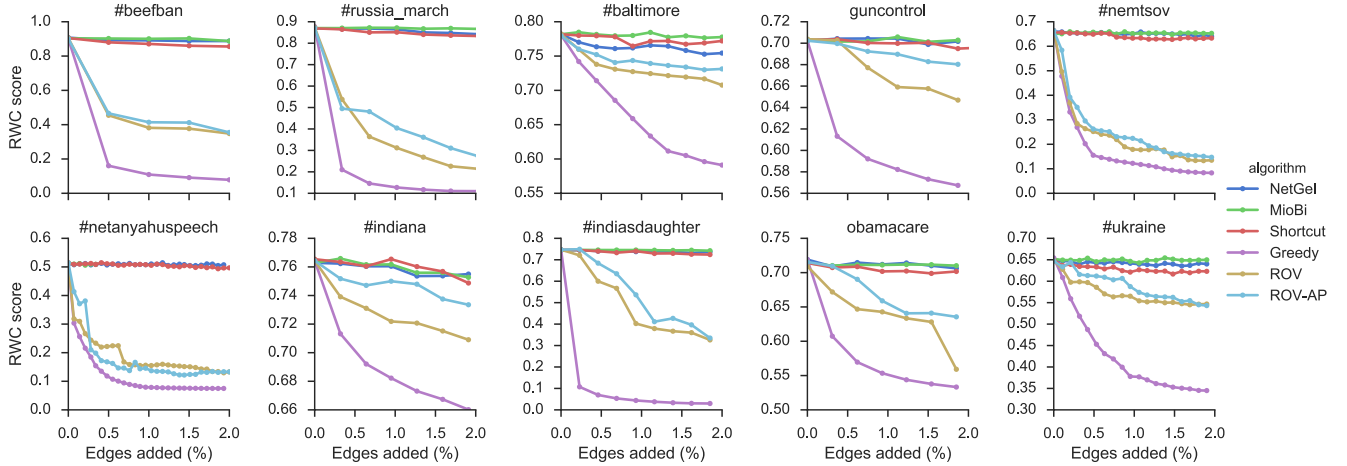### 6.3 Comparison with other related approaches

As mentioned earlier, this paper is the first tp addresses the problem of adding edges to reduce controversy. However, there exist other methods that consider adding edges to improve other structural graph properties. In this section, we compare our approach with three such recent methods: (_i_) NetGel [38], which maximizes the largest eigenvalue; (_ii_) MioBi [6], which maximizes the average eigenvalue; and (_iii_) Shortcut [32], which minimizes the average shortest path. We also experiment with the simple greedy version of our approach, which does not use the heuristic proposed in Section 4, but considers all possible edges.

The results are shown in Figure 2. As expected, the greedy brute-force algorithm performs the best. Our algorithm, ROV, which considers only a small fraction of possible edges, performs quite well, and in some cases, is on par with the greedy. The version of our algorithm with edge acceptance probabilities, ROV-AP, comes next. It is worth noting that even though the choice of edges for ROV-AP is based on a different criterion, the performance of the algorithm in terms of the RWC score is not impacted much. On the other hand, as we will see in Section 6.5, using edge acceptance probabilities improves significantly the real world applicability of our approach.

The other methods (NetGet, MioBi and Shortcut) do not perform particularly well. This is expected, as those methods are not designed to optimize our objective function. Overall, our results demonstrate the need for a specialized method to reduce controversy.

6.4 Edge-addition strategies

Let us now evaluate different edge-addition strategies. The goal is to test the hypothesis that adding edges among high-degree vertices on the two sides of the controversy gives the highest decrease in polarity score. For each of the 10 datasets, we generate a list of random high-degree vertices and non-high-degree vertices on each side. We then generate a list of

---

[1] https://dev.twitter.com/streaming/public

Figure 2: Comparison of the proposed methods (ROV and ROV-AP) with related approaches (NetGel, MioBi, Shortcut) for 2% of the total edges added. The Greedy algorithm considers all possible edges.
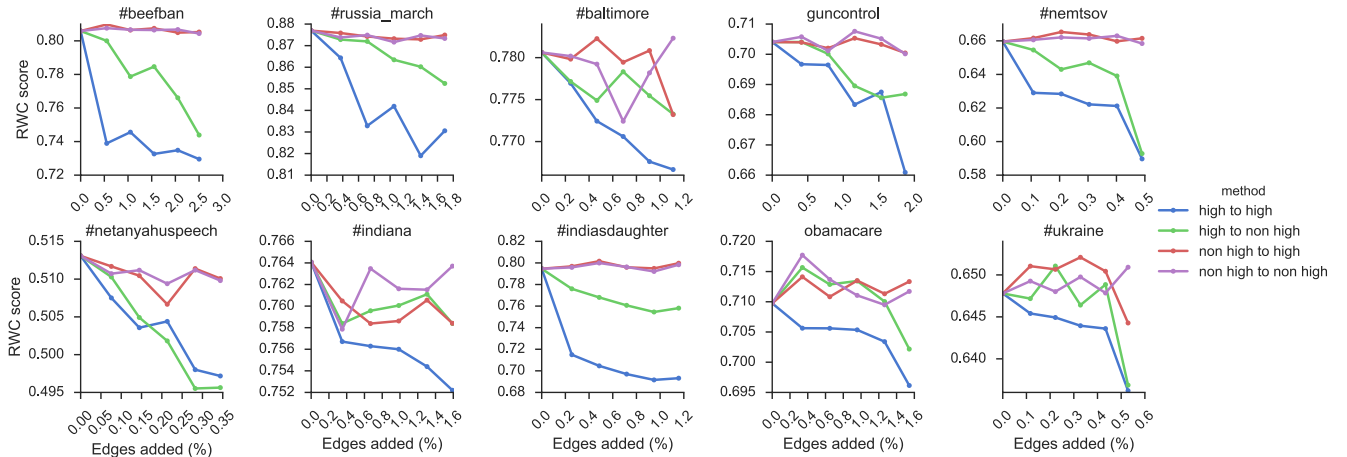
Figure 3: Comparison of different edge-addition strategies after the addition of 50 edges.

50 edges, drawn at random from the sampled vertices, and corresponding to the 4 possible combinations (high/non-high to high/non-high edges). Figure 3 shows the results of these simulations. We see that, despite the fact that high-degree vertices are selected at random, connecting such vertices gives the highest decrease in polarity score (blue line).

6.5 Case study

In order to provide qualitative evidence on the functioning of our algorithms on real-world datasets, we conduct a case study on three datasets. The datasets are chosen for the ease of the interpretation of the results, since they represent topics of wider interest (compared to *beefban*, for example, which is specific to India).

The results of the case study are summarized in Table 3. We can verify that the recommendations we obtain are meaningful and agree with our intuition for the proposed methods. The most important observation is that when comparing ROV and ROV-AP we see a clear difference in the type of edges recommended. For example, for *obamacare*, ROV recommends edges from *mittromney* to *barackobama*, and from *barackobama* to *paulryanvp* (2012 republican vice president nominee). Even though these edges indeed connect opposing

sides, they might be hard to materialize in the real world. This issue is mitigated by ROV-AP, which recommends edges between less popular users, yet connects opposing viewpoints. Examples include the edge (*csgv*, *dloesch*) for guncontrol, which connects a pro-gun-control organization to a conservative radio host, or the edge (*farhankvirk*, *pamelageller*), which connects an islamist blogger with a user who wants to "Stop the Islamization of America."[2]

Additionally, we provide a quantitative comparison of the output of the two algorithms, ROV and ROV-AP, by extracting several statistics regarding the recommended edges. In particular we consider: (*i*) *Total number of followers*. We compute the median number of followers from all edges suggested by ROV and ROV-AP. A high value indicates that the users are more central. (*ii*) *Overlap of tweet content*, For each edge we compute the Jaccard similarity of the text of the tweets of the two users. We aggregate these values for each dataset, by taking the median among all edges. A higher

---

[2]Note that since some of the data is from 2012-13, some accounts may have been deleted/moved (e.g., *paulryanvp*, *truthteam2012*). Also, some accounts may have changed stance in these years. Interested readers can use the Internet Archive Wayback Machine to have a look at past profiles.
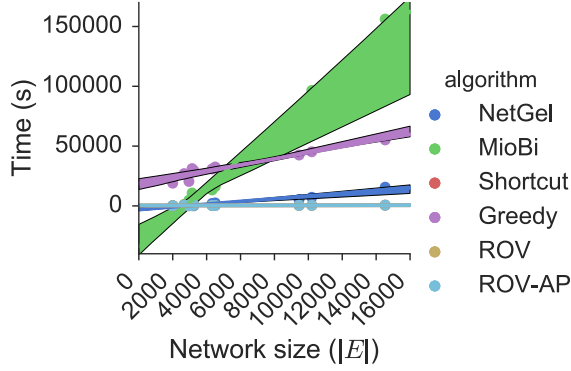
Figure 4: Running time of the proposed algorithms and competitors. ROV and ROV-AP **almost overlap.**

Figure 5: Relative Speed-up produced by using our proposed method in Section 5

value indicates that there is higher similarity between the tweet texts of the two users recommended by the algorithm. (*iii*) *Fraction of common retweets.* For each recommended edge (x, y), we obtain all other users who retweeted users x and y, and compute the Jaccard similarity of the two sets. As before, we aggregate for each dataset, by taking the median among all edges. A higher value indicates that there is a higher agreement in endorsement for users x, y on the topic.

The results are presented in Table 4. We observe that the results agree with our intuition. For example, ROV-AP produces edges with a lower number of followers (not extremely popular users), who have more common retweets, and a higher overlap in terms of tweet content.

6.6 Running time

Finally, we measure the performance of our algorithms in terms of running time. Figure 4 shows that both our algorithms ROV and ROV-AP are fast in comparison to other approaches. Greedy and MioBi are the slowest overall.

Moreover, Figure 5 shows the improvement in running time due to the incremental computation of Section 5. We observe that there is almost an order of magnitude improvement for all the datasets (from 2x – 60x). The density of the graph is indicated by the density of the grey lines in the plot. In general, the speedup is larger for denser graphs.

7. CONCLUSIONS

We considered the problem of bridging opposing views on social media by recommending relevant content to certain users (edges in the endorsement graph). Our work builds on recent studies of controversy in social media and uses a random walk-based score as a measure of controversy. We first proposed a simple, yet efficient, algorithm to bridge opposing sides. Furthermore, inspired by recent user studies on how users prefer to consume content from opposing views, we improved the algorithm to take into account the probability of a recommendation being accepted. Finally, we also proposed a way to incrementally compute the random-walk score by using matrix operations, which typically gives more than an order of magnitude improvement in runtime. We evaluated our algorithms on a wide range of real-world datasets in Twitter, and showed that our methods outperform other baselines.
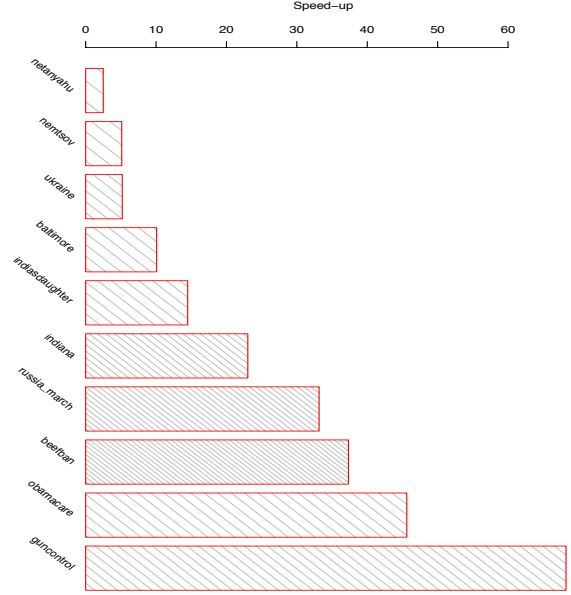
Future work. Our approach relies on a random walk-based optimization function [11]. Although this measure has been proven to be effective it has a few drawbacks. In particular, the measure is applicable to controversies having two sides. One way to overcome this restriction is to assume the presence of multiple clusters, and define the measure accordingly. In the future, we plan to experiment with this generalization of our method, as well as, investigate the edge-recommendation problem for other objective functions.

As mentioned in Section 2, previous work deals mostly with the problem of *how* to connect opposing sides, while our work provides methods for selecting *who* to recommend. Another interesting direction is studying the problem of *what* to recommend.

8. REFERENCES

[1] L. A. Adamic and E. Adar. Friends and neighbors on the web. *Social networks*, 25(3):211–230, 2003.

[2] L. A. Adamic and N. Glance. The political blogosphere and the 2004 us election: divided they blog. In *LinkKDD*, pages 36–43, 2005.

[3] L. Akoglu. Quantifying political polarity based on bipartite opinion networks. In *ICWSM*, 2014.

[4] F. Arrigo and M. Benzi. Edge modification criteria for enhancing the communicability of digraphs. *arXiv preprint arXiv:1508.01056*, 2015.

[5] F. Arrigo and M. Benzi. Updating and downdating techniques for optimizing network communicability. *SIAM Journal on Scientific Computing*, 38(1):B25–B49, 2016.

[6] H. Chan, L. Akoglu, and H. Tong. Make it or break it: Manipulating robustness in large networks. In *SDM*, pages 325–333. SIAM, 2014.

Table 3: Twitter handles of the top edges picked by our algorithms for different datasets.

	obamacare		guncontrol		#netanyahuspeech	
	vertex1	vertex2	vertex1	vertex2	vertex1	vertex2
ROV	mittromney	barackobama	ghostpanther	barackobama	maxblumenthal	netanyahu
	realdonaldtrump	truthteam2012	mmflint	robdelaney	bipartisanism	lindasuhler
	barackobama	drudge_report	miafarrow	chuckwoolery	harryslaststand	rednationrising
	barackobama	paulryanvp	realalexjones	barackobama	lindasuhler	marwanbishara
	michelebachmann	barackobama	goldiehawn	jedediahbila	thebaxterbean	worldnetdaily
ROV-AP	kksheld	ezraklein	chuckwoolery	csgv	farhankvirk	pamelageller
	lolgop	romneyresponse	liamkfisher	miafarrow	medeabenjamin	annebayefsky
	irritatedwoman	motherjones	csgv	dloesch	2afight	sttbs73
	hcan	romneyresponse	jonlovett	spreadbutter	rednationrising	palsjustice
	klsouth	dennisdmz	drmartyfox	huffpostpol	jvplive	chucknellis

Table 4: Quantitative comparison of recommendations from ROV and ROV-AP. * indicates that the result is statistically significant with $p < 0.1$, and ** with $p < 0.001$. **Significance is tested using Welch's t-test for inequality of means.**

	ROV	ROV-AP
NumFollowers	50729	36160*
ContentOverlap	0.054	0.073**
CommonRetweets	0.029	0.063**

[7] M. Conover, J. Ratkiewicz, M. Francisco, B. Gonçalves, F. Menczer, and A. Flammini. Political Polarization on Twitter. In *ICWSM*, 2011.

[8] M. Del Vicario, A. Bessi, F. Zollo, F. Petroni, A. Scala, G. Caldarelli, H. E. Stanley, and W. Quattrociocchi. Echo chambers in the age of misinformation. *arXiv preprint arXiv:1509.00189*, 2015.

[9] E. D. Demaine and M. Zadimoghaddam. Minimizing the diameter of a network using shortcut edges. In *Algorithm Theory-SWAT 2010*, pages 420–431. Springer, 2010.

[10] R. Fagin, A. Lotem, and M. Naor. Optimal aggregation algorithms for middleware. *Journal of computer and system sciences*, 66(4):614–656, 2003.

[11] 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*, WSDM '16, pages 33–42, New York, NY, USA, 2016. ACM. ISBN 978-1-4503-3716-8. . URL http://doi.acm.org/10.1145/2835776.2835792.

[12] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy in social media. *arXiv preprint arXiv:1507.05224*, 2016.

[13] R. K. Garrett. Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication*, 14(2):265–285, 2009.

[14] G. H. Golub and C. F. Van Loan. *Matrix computations*, volume 3. JHU Press, 2012.

[15] E. Graells-Garrido, M. Lalmas, and D. Quercia. Data portraits: Connecting people of opposing views. *arXiv preprint arXiv:1311.4658*, 2013.

[16] E. Graells-Garrido, M. Lalmas, and D. Quercia. People of opposing views can share common interests. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 281–282. International World Wide Web Conferences Steering Committee, 2014.

[17] P. H. C. Guerra, W. Meira Jr, C. Cardie, and R. Kleinberg. A measure of polarization on social media networks based on community boundaries. In *ICWSM*, 2013.

[18] G. Guo, J. Zhang, Z. Sun, and N. Yorke-Smith. Librec: A java library for recommender systems. In *Proceedings of the 23rd International Conference on User Modeling,*

Adaptation and Personalization, 2015.

[19] D. J. Isenberg. Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, 50(6):1141, 1986.

[20] S. Kabbur, X. Ning, and G. Karypis. Fism: factored item similarity models for top-n recommender systems. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 659–667. ACM, 2013.

[21] Q. V. Liao and W.-T. Fu. Beyond the filter bubble: interactive effects of perceived threat and topic involvement on selective exposure to information. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2359–2368. ACM, 2013.

[22] 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.

[23] Q. V. Liao and W.-T. Fu. Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2745–2754. ACM, 2014.

[24] H. Lu, J. Caverlee, and W. Niu. Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, pages 213–222. ACM, 2015.

[25] Y. Mejova, A. X. Zhang, N. Diakopoulos, and C. Castillo. Controversy and sentiment in online news. *Symposium on Computation + Journalism*, 2014.

[26] A. Morales, J. Borondo, J. Losada, and R. Benito. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos*, 25(3), 2015.

[27] S. A. Munson and P. Resnick. Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1457–1466. ACM, 2010.

[28] S. A. Munson, S. Y. Lee, and P. Resnick. Encouraging reading of diverse political viewpoints with a browser widget. In *ICWSM*, 2013.

[29] X. Ning and G. Karypis. Slim: Sparse linear methods for top-n recommender systems. In *2011 IEEE 11th International Conference on Data Mining*, pages 497–506. IEEE, 2011.

[30] M. Papagelis, F. Bonchi, and A. Gionis. Suggesting ghost edges for a smaller world. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, pages 2305–2308. ACM, 2011.

[31] E. Pariser. *The filter bubble: What the Internet is hiding from you.* Penguin UK, 2011.

[32] N. Parotisidis, E. Pitoura, and P. Tsaparas. Selecting shortcuts for a smaller world. In *SIAM International Conference on Data Mining (SDM)*, pages 28–36. SIAM,

2015.

[33] N. Parotsidis, E. Pitoura, and P. Tsaparas. Centrality-aware link recommendations. In *Proceedings of the Ninth ACM International Conference on Web Search and Data Mining*, pages 503–512. ACM, 2016.

[34] S. Perumal, P. Basu, and Z. Guan. Minimizing eccentricity in composite networks via constrained edge additions. In *Military Communications Conference, MILCOM 2013-2013 IEEE*, pages 1894–1899. IEEE, 2013.

[35] S. Soundarajan and J. Hopcroft. Using community information to improve the precision of link prediction methods. In *Proceedings of the 21st International Conference on World Wide Web*, pages 607–608. ACM, 2012.

[36] C. R. Sunstein. The law of group polarization. *Journal of political philosophy*, 10(2):175–195, 2002.

[37] C. R. Sunstein. *Republic. com 2.0.* Princeton University Press, 2009.

[38] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 245–254. ACM, 2012.

[39] V. Vydiswaran, C. Zhai, D. Roth, and P. Pirolli. Overcoming bias to learn about controversial topics. *Journal of the Association for Information Science and Technology*, 2015.

[40] A. Zeng, L. Lü, and T. Zhou. Manipulating directed networks for better synchronization. *New Journal of Physics*, 14(8):083006, 2012.

[41] T. Zhou, L. Lü, and Y.-C. Zhang. Predicting missing links via local information. *The European Physical Journal B*, 71 (4):623–630, 2009.