Reducing Controversy by Connecting Opposing Views*

Kiran Garimella[1], Gianmarco De Francisci Morales[2], Aristides Gionis[3], Michael Mathioudakis[4]
[1] EPFL, Switzerland
[2] Qatar Computing Research Institute, Qatar
[3] Aalto University, Finland
[4] University of Helsinki, Finland
kiran.garimella@epfl.ch, gdfm@acm.org, aristides.gionis@aalto.fi, michael.mathioudakis@helsinki.fi

Abstract

Controversial issues often 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 discussions as *graphs*, and cast our objective as an *edge-recommendation problem*. The goal of the recommendation is to reduce the *controversy score* of the graph, measured by a recently-developed metric based on random walks. At the same time, we take into account the *acceptance probability* of the recommended edges, which represent the probability that the recommended edges materialize in the graph.

1 Introduction

Polarization around controversial issues is a well-studied phenomenon in social sciences [Isenberg, 1986; Sunstein, 2002]. Social media have arguably eased the emergence of such issues, thanks to the publicity they foster. This paper studies how to reduce 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 "retweets" or "shared" opinions among users. 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.

It has been observed that online discussions form echo chambers [Garrett, 2009; Vicario *et al.*, 2015; Garimella *et al.*, 2018b], where net-citizens are not informed about opposing views. The phenomenon is amplified by behavioral traits such as confirmation bias, homophily, selective exposure, and related social phenomena. These chambers are a hindrance to the democratic process 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.

To make our objective concrete we utilize metrics for quantifying online controversy [Garimella *et al.*, 2016b; Garimella *et al.*, 2018a]. In particular, given a metric that measures how controversial an issue is, we aim to find a small number of edges, called bridges, which minimize this measure. That is, we seek to propose (content produced by) a user v to another user u. This action would create a new edge (a bridge) in the endorsement graph, thus reducing the controversy score of the discussion graph [Garimella *et al.*, 2016a].

Clearly, some bridges are more likely to materialize than others. For instance, in politics, people in the "center" might be easier to persuade than people on the two extreme ends of the political spectrum [Liao and Fu, 2014]. We take this issue into account by modeling the *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. Therefore, we seek bridges that minimize the *expected* controversy score, according to their acceptance probabilities.

Our main contribution 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) 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.

2 Preliminaries and Problem Definition

For selecting which edges to recommend in order to reduce controversy we need to rely on a measure of controversy.

*This is an abridged version of a homonymous paper that received the best student paper award in ACM WSDM 2017.

In this paper, we adopt a measure proposed in our previous work [Garimella *et al.*, 2018a], as 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 [Garimella *et al.*, 2018a]:

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

(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 is obtained via 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.

Given the controversy measure $\text{RWC}(G, X, Y)$, the problem we consider 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' (the size k of the recommendation is negligible).

3 Algorithm

A brute-force approach to solve the problem needs to consider all k-set combinations of non-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. However, even for the greedy approach the number of possible edges to consider is prohibitively large in real settings. Since computing the controversy score is expensive, we would like to invoke the function as few times as possible. That is, we aim to consider far fewer candidate edges.

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. It is possible to show via a simple example that the best edges for reducing our controversy measure are between high-degree vertices from opposite sides (for details visit the full version [Garimella *et al.*, 2017]). The algorithm is shown in Algorithm 1. Its running time is $\mathcal{O}(k_1 k_2)$, where k_1 and 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

Initialize: Out $\leftarrow$ *empty list* ;
for i = *1:*k_1 **do**
    vertex u = X[i];
    for j = *1:*k_2 **do**
        vertex v = Y[j];
        Compute $\delta\text{RWC}_{u \to v}$, the decrease in RWC if the edge (u, v) is added;
        Append $\delta\text{RWC}_{u \to v}$ to Out;
        Compute $\delta\text{RWC}_{v \to u}$, the decrease in RWC if the edge (v, u) is added;
        Append $\delta\text{RWC}_{v \to u}$ to Out;
sorted $\leftarrow$ sort(Out) by δRWC by decreasing order ;
return top k from sorted;

3.1 Adding 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 $E_{\mathbb{A}}[\text{RWC}(G', X, Y)]$ 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. [Garimella *et al.*, 2018a]. 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)

Dataset	# Tweets	Retweet graph	
		$\lvert V\rvert$	$\lvert E\rvert$
#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 1: Datasets statistics: hashtag used to collect dataset, number of tweets, size of retweet graph.

indicate that the user does not hold a strong opinion. 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 a recommended edge (u, v) from vertex u to vertex v, with acceptance probability $p(u, v)$ and RWC decrease $\delta\mathrm{RWC}_{u\to v}$, the *expected decrease* in RWC when the edge is recommended individually is

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

4 Experiments

In this section, we provide an evaluation of the two algorithms proposed in Section 3. 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.

4.1 Datasets

We use Twitter datasets on known controversial issues. The datasets have also been used in previous studies [Garimella *et al.*, 2018a; Lu *et al.*, 2015]. 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.

4.2 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 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 1 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).

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

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

Table 3: Quantitative comparison of recommendations from ROV and ROV-AP (median values). An asterisk * indicates that the result is statistically significant with $p < 0.1$, and two asterisks ** with $p < 0.001$. Significance is tested using Welch's t-test for inequality of means.

4.3 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 2. 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 `barackbobama`, 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 value indicates that there is higher similarity between the tweet texts of the two

[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.

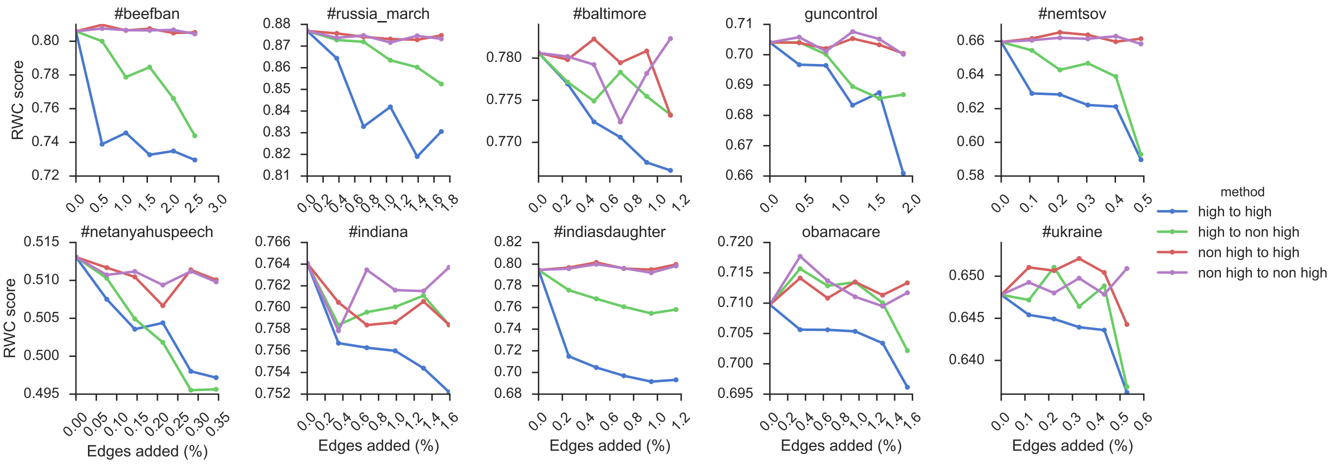

Figure 1: Comparison of different edge-addition strategies after the addition of 50 edges.

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

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.

Results are presented in Table 3. 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.

5 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. We evaluated our algorithms on a wide range of real-world datasets in Twitter, and showed that our methods outperform other baselines.

Acknowledgements

This work was supported by the Academy of Finland projects "Nestor" (286211), "Agra" (313927), and "AIDA" (317085), and the EC H2020 RIA project "SoBigData" (654024).

References

[Garimella *et al.*, 2016a] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Exploring Controversy in Twitter. In *Proceedings of the 19th Conference on Computer Supported Cooperative Work and Social Computing*, CSCW, pages 33–36. ACM, 2016.

[Garimella *et al.*, 2016b] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy in social media. In *Proceedings of the 9th ACM International Conference on Web Search and Data Mining*, WSDM, pages 33–42. ACM, 2016.

[Garimella *et al.*, 2017] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Reducing controversy by connecting opposing views. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, WSDM, pages 81–90. ACM, 2017.

[Garimella *et al.*, 2018a] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Quantifying controversy on social media. In *ACM Transactions on Social Computing*, volume 1, pages 1–27, 2018.

[Garimella *et al.*, 2018b] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, pages 913–922, 2018.

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

[Isenberg, 1986] Daniel Isenberg. Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology*, 50(6):1141–1142, 1986.

[Liao and Fu, 2014] Q-Vera Liao and Wai-Tat 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.

[Lu *et al.*, 2015] Haokai Lu, James Caverlee, and Wei 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.

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

[Vicario *et al.*, 2015] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. Echo chambers in the age of misinformation. *CoRR*, abs/1509.00189:1–7, 2015.