Factors in Recommending Contrarian Content on Social Media

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

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

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

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

ABSTRACT

Polarization is a troubling phenomenon that can lead to societal divisions and hurt the democratic process. It is therefore important to develop methods to reduce it.

We propose an algorithmic solution to the problem of reducing polarization. The core idea is to expose users to content that challenges their point of view, with the hope broadening their perspective, and thus reduce their polarity. Our method takes into account several aspects of the problem, such as the estimated polarity of the user, the probability of accepting the recommendation, the polarity of the content, and popularity of the content being recommended.

We evaluate our recommendations via a large-scale user study on Twitter users that were actively involved in the discussion of the US elections results. Results shows that, in most cases, the factors taken into account in the recommendation affect the users as expected, and thus capture the essential features of the problem.

ACM Reference format:
Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. Factors in Recommending Contrarian Content on Social Media. In *Proceedings of WebSci '17, Troy, NY, USA, June 25-28, 2017,* 4 pages.
DOI: http://dx.doi.org/10.1145/3091478.3091515

1 INTRODUCTION

Polarization around controversial issues is a well-studied phenomenon in the social sciences [11].Social media have arguably amplified polarization, thanks to the scale of discussions and their publicity [7]. This paper studies how to reduce polarization on social media by recommending *contrarian* content, i.e., content that expresses a point-of-view opposing the one held by the target user. In particular, we examine which features might be used to develop such a content recommender system.

WebSci '17, Troy, NY, USA

978-1-4503-4896-6/17/06…$15.00
DOI: http://dx.doi.org/10.1145/3091478.3091515

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

Due to phenomena such as homophily, confirmation bias, and selective exposure, social media often create echo chambers [5, 8]. These chambers, in turn, cultivate isolation and misunderstanding in society [18], and deepen its polarization.

A potential solution to this problem is to encourage users to consider points of view different from their own. Thus, in this paper, we study methods to recommend content items (e.g., news articles, opinion pieces, blog posts) that express a contrarian point of view, while at the same time being appealing to the target user.

In particular, given metrics that measure the polarization of users and items (such as those proposed in recent research [3]), our goal is to recommend an item that nudges the user towards the opposite polarity. That is, we seek to propose content produced by a user v to another user u, thus informing u of a different viewpoint, and hoping that u will endorse v.

Clearly, some content is more likely to be endorsed than other. For instance, people in the "center" might be easier to convince than people on the two extreme ends of the political spectrum [13]. We take this issue into account by modeling the *acceptance probability* for a recommendation as a separate component of the model.

We blend these factors, together with other signals such as topic and popularity, to create a ranked list of recommendations. Our solution employs a well-known weighted rank-aggregation algorithm at its core [17].

We evaluate our proposal by running an online user study with Twitter users. We focus on the recent 2016 US presidential elections, and generate recommendations for the thousands of users involved in this highly-polarizing controversial discussion. The results of the study show that the two main factors used in the recommendation, the polarity and the acceptance probability models, are predictive of the responses of the users.

In summary, we make the following contributions:

- We study the problem of bridging echo chambers algorithmically, in a language- and domain-agnostic way. Previous studies that address this problem focus mostly on understanding *how* to recommend content to an ideologically opposite side, while we

focus on *which* contrarian content to recommend. We believe that the two approaches complement each other in bringing us closer to bursting filter bubbles.

- We build on top of results from recent user studies [14, 15, 19] on how users prefer to consume content from opposing views, and formulate the task as a content-recommendation problem based on an endorsement graph, while also taking into account the acceptance probability of a recommendation.
- We evaluate the proposed solution via a user study on Twitter users, and demonstrate the validity of the main factors involved in the recommendation.

2 RELATED WORK

Although the Web was envisioned as a place of open discussions on a wide range of topics, many people tend to restrict themselves to viewing and sharing information that conforms with their beliefs. A wide body of recent studies has explored [1, 2] and quantified [3] the notions of "filter bubble" and "echo chambers".

Munson et al. [15] created a browser widget that measures the bias of users based on the news articles they read. Their study shows that users are willing to slightly change views once they are shown their biases. Graells-Garrido et al. [9] 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. Graells-Garrido et al. [10] propose to find 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. Park et al. [16] propose methods for presenting multiple aspects of news to reduce bias.

Most relevant to this work is the recent study about the problem of reducing the overall polarization of a controversial topic in a network [6]. The study tries to find the best edges that can be added to an endorsement graph so that the polarization score of the network is reduced. In this paper, we focus on reducing the polarization of an individual user (local objective), instead of the entire network (global objective).

There have also been a number of demos and systems: Wall Street Journal's *Blue feed-Red feed*[1] raises awareness about the extent to which viewpoints on a matter can differ, by showing side-by-side articles expressing very liberal and very conservative viewpoints; *Politecho*[2] displays how polarizing the content on a user's news feed is when compared to their friends'; *Escape your bubble*[3] is a browser extension to add hand-curated content from the opposite side in Facebook; automated bots have been created to respond to posts containing harassment or fake news,[4] with an attempt to de-polarize the discussion and educate users. Moreover, new social media platforms have been proposed that aim to be designed in such a way to encourage discussions and debates, such as the Filterburst project,[5] Rbutr,[6] where users can post rebuttals of other urls, and a wikipedia for debates.[7]

The proposed method differs from existing ones in many ways. First, our approach is completely algorithmic, unlike most demos listed above, which involve manual curation. Second, as discussed above, it builds on top of existing research and incorporates key findings of previous work.

3 PRELIMINARIES

A topic of discussion is identified as the set of tweets that satisfy a text query – e.g., all tweets that contain a specific hashtag. We represent a topic with an *endorsement graph* $G(V, E)$, where vertices V represent users and edges E represent *endorsements*.

It has been shown that an endorsement graph captures well the extent to which a topic is controversial [3]. In particular, the endorsement graph of a controversial topic has a *multimodal clustered structure*, where each cluster of vertices represents one viewpoint on the topic. As we focus on two-sided controversies, we identify the two sides of a controversial topic by employing a *graph-partitioning* algorithm, which partitions the graph into *two* subgraphs. In this work, we specifically focus on recommending content in the form of news items, such as articles, blog posts, and opinion pieces. The item pool for the recommendation comprises all the links shared by the active users during the observation window.

User polarization score. We use a recently-proposed methodology to define the polarization score for each user in the graph [4]. The score is based on the expected hitting time of a random walk that starts from the user under consideration and ends on a high-degree vertex on either side. Typically, in a retweet graph, high-degree vertices on each side are indicators of authoritative content generators. We denote the set of the k highest degree vertices on each side by X^+ and Y^+. Intuitively, a vertex is assigned a score of higher absolute value (closer to +1 or −1), if, compared to other vertices in the graph, it takes a very different time to reach a high-degree vertex on either side (X^+ or Y^+) (in terms of information flow). Specifically, for each vertex $u \in V$ in the graph, we consider a random walk that starts at u, and estimate the expected number of steps, l_u^X before the random walk reaches any high-degree vertex in X^+. Considering the distribution of values of l_u^X across all vertices $u \in V$, we define $\rho^X(u)$ as the fraction of vertices $v \in V$ with $l_v^X < l_u^X$. We define $\rho^Y(u)$ similarly. Obviously, we have $\rho^X(u), \rho^Y(u) \in [0, 1)$. The polarization score of a user is then defined as

$$\rho(u) = \rho^X(u) - \rho^Y(u) \quad \in (-1, 1). \tag{1}$$

Following this definition, a vertex that is close to high-degree vertices X^+, compared to most other vertices, will have $\rho^X(u) \approx 1$; on the other hand, if the same vertex is far from high-degree vertices Y^+, it will have $\rho^Y(u) \approx 0$; leading to a polarization score $\rho(u) \approx 1 - 0 = 1$. The opposite is true for vertices that are far from X^+ but close to Y^+; leading to a polarization score $\rho(u) \approx -1$.

Item polarization score. Once we have obtained polarization scores for users in the graph, it is straightforward to derive a similar score for content items shared by these users. Specifically, we define

[1] http://graphics.wsj.com/blue-feed-red-feed/
[2] http://politecho.org/
[3] https://www.escapeyourbubble.com/
[4] http://wpo.st/4kVR2, https://goo.gl/Xl6x9t
[5] http://www.filterburst.com/
[6] http://rbutr.com/
[7] http://www.debatepedia.org/en/index.php/Welcome_to_Debatepedia%21

the polarization score of an item i as the average of the polarization scores of the set of users who have shared i, denoted by U_i:

$$\rho(i) = \frac{1}{|U_i|} \sum_{u \in U_i} \rho(u) \quad \in (-1, 1). \tag{2}$$

Acceptance probability. Not all recommendations are agreeable, especially if they do not conform to the user's beliefs. To reduce these effects, we define an acceptance probability, which quantifies the degree to which a user is likely to endorse the recommended content. We use the item and user polarization scores defined above to estimate the likelihood that a target user u endorses (i.e., retweets) the recommended item i. We build an acceptance model by adapting a similar one based on the feature of user polarization [6]. High absolute values of user polarization (close to -1 or $+1$) indicate that the user belongs clearly to one side of the controversy, while middle-range values (close to 0) indicate that the user is in the middle of the two sides. It was shown that users from either side accept content from different sides with different probabilities, and these probabilities can be inferred from the graph structure [6]. For example, a user with polarization close to -1 is more likely to endorse a user with a negative polarization than a user with polarization $+1$. This intuition directly translates to endorsing items, and therefore can be used for our recommendation problem.

Based on this intuition, we define the acceptance probability $p(u, i)$ of a user u endorsing item i as

$$p(u, i) = {}^{N_e(\rho(u),\, \rho(i))}/_{N_x(\rho(u),\, \rho(i))}, \tag{3}$$

where $N_e(\rho(u), \rho(i))$ and $N_x(\rho(u), \rho(i))$ are the number of times a user with polarity $\rho(u)$ has endorsed or was exposed to (respectively) content of polarity $\rho(i)$. In practice, the polarity scores are bucketed to smooth the probabilities.

4 RECOMMENDATION FACTORS

This section describes the factors used to generate recommendations. Though our main focus is to connect users with content that expresses a contrarian point of view, we also want to maximize the chances of such a recommendation being endorsed by the user. We take into account several factors: reduction in polarization score of the target user; exclusivity of the candidate items (polarity of the items); acceptance probability of recommendation based on polarization scores; topic diversity; popularity/quality of the candidate item. Next, we describe these factors in more detail.

Reduction of user polarization score. The maximum reduction of user polarization score is achieved by putting the user in contact with an authoritative source from the opposing side. Leveraging this idea, we build a list of items L_1 by considering items shared by high degree nodes on the opposite side of the target user, and ranking them by the potential decrease in user polarization score.

Exclusivity on either side. We consider items that are almost exclusively shared by one of the sides. Specifically, we denote by n_i^X and n_i^Y the number of users who shared each item i on side X and Y, respectively. For each side, we generate a list L_2 ranked by the ratio of shares ${}^{n_i^X}/_{n_i^Y}$ (for side X) and ${}^{n_i^Y}/_{n_i^X}$ (for side Y).

Acceptance probability. For a given user u, all items sorted in decreasing order of acceptance probability $p(u, i)$ make up list L_3.

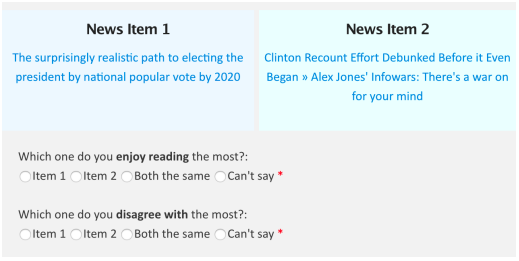

Figure 1: Screenshot of the interface shown for a user with a high polarity on the political left (Democrat).

Topic diversity. We want to ensure that the recommendations are topically diverse. To achieve this, for each user, we compute a vector t_u that contains the topics extracted from the tweets written and the items shared by the user. Similarly, we extract a vector of topics t_i for each item. Topics are defined as *named entity*, and we extract them using the tool tagme.[8] Given a user vector t_u, we compute the cosine similarity with all item vectors t_i, and rank items in increasing order of cosine similarity (list L_4).

Popularity on either side. Finally, we take into account the popularity of the recommended items, so that users receive content that is popular and, likely, of good quality. For each item, we compute a popularity score as the maximum number of retweets obtained by a tweet that contains this item. We produce list L_5 of items in decreasing popularity score.

Rank Aggregation. Given the 5 ranked lists discussed above, we use a weighted rank-aggregation scheme to generate the final recommendations. The intuition behind using rank aggregation is that items that are highly ranked in many lists, are also highly ranked in the output list. In particular, we use a weighted rank-aggregation technique proposed by Pihur et al. [17], whose goal is to minimize the objective function

$$\phi(\delta) = \sum_{i=1}^{5} w_i d(\delta, L_i), \tag{4}$$

where δ is the optimal ranked output list, d is any distance function (we use the Spearman footrule distance), and w_i are the importance weights of each list. We can set the weights to generate highly contrarian recommendations (by giving large weights to L_1 and L_2) or recommendations that are likely to be accepted (by giving large weight to L_3).

5 EVALUATION

Dataset. We collect all tweets containing the hashtag `#USelections`, used in discussions about the US presidential elections during Nov 9–12, 2016. From the 6.2 M tweets collected, we build an endorsement graph with 6764 nodes (users) and 9896 edges (retweets). To filter out noise, the graph contains an edge between two users only if at least 5 retweets between the two users occur. We partition the graph to obtain the two sides by using METIS [12]. For recommendation items (urls), we consider items that have been shared at least 5 times in our dataset. The final pool contains 10 210 candidate items, which include news articles, blog posts, opinion pieces, etc.

[8] https://services.d4science.org/web/tagme

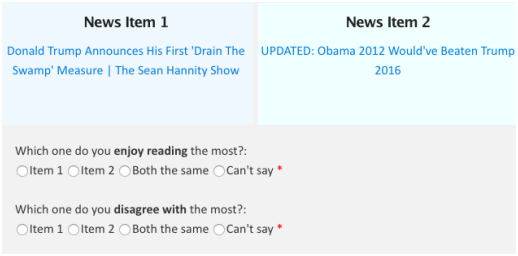

Figure 2: Screenshot of the interface shown for a user with a high polarity on the political right (Republican).

Table 1: Results from the user study.

Main factor	Item1 (Acceptance)	Item2 (Contrarian)	Both the same	Can't say
Enjoy	51	19	8	15
Disagree	22	57	7	7

User study. We run an online user study involving all 6764 users in the dataset with the aim of evaluating how users percieve the two main conflicting factors proposed, i.e. the contrarian features (L_1, L_2) and acceptance features (L_3). For each user in the study, we generate two recommended items that are personalized based on their Twitter activity: one item is highly contrarian, while the other is more likely to be accepted, according to our model. In more detail, by using the methodology described above, we compute two recommendations for each user: in the first one we give a high weight (60%) to contrarian features (L_1 and L_2), while in the second one we give high weight (60%) to acceptance probability (L_3). We distribute the remaining 40% equally among other features.

The main research questions we investigate are: (*i*) is a high acceptance probability factor predictive of content with higher acceptance? and (*ii*) are contrarian factors predictive of more disagreement with the user? To simplify the task for the user, we set up the user study as a relative comparison between the two recommendations, rather than asking for absolute judgments. Since the two recommendations are generated completely independently, we assume that they do not influence the users decision making process in choosing one over the other.

We create a web form[9] with two recommended items, customized for each user, with the item weighted by the acceptance features shown on the left and contrarian features on the right. Figures 1 and 2 show two instances of the web form. Looking at Figure 1, given the left-leaning political affiliation of the user, the recommendation on the left side (News item 1) looks more agreeable than the recommendation on the right side (News item 2). The opposite is true for Figure 2, which targets a right-leaning user.

We contacted users on Twitter with the following private message: "@username We are scientists studying social media. Would u like to help science by participating in a survey? http://bit.ly/XXXXX'', and waited for two weeks for them to respond. In total, we sent around 6700 messages and received 93 valid responses after removing duplicates (1.4% response rate).

Our expectation is that users enjoy reading the item with high acceptance probability, and disagree with the contrarian item. The results, summarized in Table 1, confirm our expectations. Indeed, most users enjoy reading the item with high acceptance, and disagree with the contrarian item. Specifically, 44 out of the 93 users (47%) reported that at the same time they enjoy the first item, and disagree with the second. For a few users (n=7), we were able to generate enjoyable recommendations that they disagreed with. While this was not the goal of the specific user study, it is indeed our ultimate goal, and thus these results are highly encouraging.

[9] http://bit.ly/2jOQBxP

Acknowledgements. This work has been supported by the Academy of Finland project "Nestor" (286211) and the EC H2020 RIA project "SoBigData" (654024).

6 REFERENCES

[1] Lada A Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 US election: divided they blog. In *LinkKDD*. 36–43.

[2] Michael Conover, Jacob Ratkiewicz, Matthew R Francisco, Bruno Gonçalves, Filippo Menczer, and Alessandro Flammini. 2011. Political Polarization on Twitter. In *ICWSM*.

[3] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying Controversy in Social Media. In *WSDM*. 33–42.

[4] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2016. Quantifying Controversy in Social Media. *arXiv preprint arXiv:1507.05224* (2016).

[5] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. The Ebb and Flow of Controversial Debates on Social Media. In *ICWSM*.

[6] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2017. Reducing Controversy by Connecting Opposing Views. In *WSDM*. 81–90.

[7] Kiran Garimella and Ingmar Weber. 2017. A Long-Term Analysis of Polarization on Twitter. In *ICWSM*.

[8] 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.

[9] Eduardo Graells-Garrido, Mounia Lalmas, and Daniele Quercia. 2013. Data portraits: Connecting people of opposing views. *arXiv preprint arXiv:1311.4658* (2013).

[10] Eduardo Graells-Garrido, Mounia Lalmas, and Daniele Quercia. 2014. People of opposing views can share common interests. In *WWW Companion*. 281–282.

[11] Daniel J Isenberg. 1986. Group polarization: A critical review and meta-analysis. *Journal of personality and social psychology* (1986).

[12] George Karypis and Vipin Kumar. 1995. METIS - Unstructured Graph Partitioning and Sparse Matrix Ordering System. (1995).

[13] Q Vera Liao and Wai-Tat Fu. 2014. Can you hear me now?: mitigating the echo chamber effect by source position indicators. In *CSCW*. 184–196.

[14] Q Vera Liao and Wai-Tat Fu. 2014. Expert voices in echo chambers: effects of source expertise indicators on exposure to diverse opinions. In *CHI*. 2745–2754.

[15] Sean A Munson and others. 2013. Encouraging Reading of Diverse Political Viewpoints with a Browser Widget.. In *ICWSM*.

[16] Souneil Park and others. 2009. NewsCube: delivering multiple aspects of news to mitigate media bias. In *CHI*. 443–452.

[17] Vasyl Pihur and others. 2009. RankAggreg, an R package for weighted rank aggregation. *BMC bioinformatics* 10, 1 (2009), 1.

[18] Cass R Sunstein. 2009. *Republic. com 2.0.* Princeton University Press.

[19] VG Vydiswaran, ChengXiang Zhai, Dan Roth, and Peter Pirolli. 2015. Overcoming bias to learn about controversial topics. *JASIST* (2015).