5^{th} International Conference on Computational Social Science IC^2S^2
July 17-20, 2019, University of Amsterdam, The Netherlands

Reducing Polarization by Connecting Opposing Views

Kiran Garimella[⋆], Gianmarco De Francisci Morales[†], Aristides Gionis[◇] and Michael Mathioudakis‡
⋆ MIT, USA; † ISI Foundation, Italy; ◇ Aalto University, Finland; ‡ University of Helsinki, Finland

Keywords: polarization, controversy, social media, recommendation, network design

Discussions on polarizing issues such as abortion or gun control on social media are often associated with the creation of "echo chambers", where people hear opinions of like-minded people, but not ones from the opposing side. In our work, we study algorithmic techniques to bridge echo chambers, and thus reduce polarization. To achieve this goal, we prompt users to view, and possibly endorse, content that was posted online by social media users of an opposing side — and thus create a connection between such users. For creating effective connections, we face the following trade-off. On the one hand, it is desirable to create connections between highly polarized users of opposing sides, as this would successfully bridge the echo chamber. On the other hand, such connections are not likely to materialize in practice, as highly polarized users avoid endorsing views of the opposing side. The algorithms we develop navigate this trade-off to achieve the highest possible returns in expectation.

Measuring polarization. The task of connecting opposing views requires defining a measure for polarization. Our measure of choice is based on the notion of *endorsement graph* [2], a graph where nodes represent users who participate in a given discussion, and directed edges represent endorsement between users. In the case of Twitter, for example, the edges of such a graph represent *retweets* between users (Figure 1).

Once the endorsement graph is built for a given discussion, we quantify the polarization represented in the graph in two steps. First, we obtain a partitioning of the graph into two sets of nodes (e.g., see red and blue nodes of Figure 1). Intuitively, if the discussion is polarized, each partition corresponds to one side of the discussion. Second, we measure the separation of the two partitions, by using a random walk-based measure called *random-walk controversy* (RWC) score [2]. The higher the magnitude of separation of the two partitions according to RWC, the higher the polarization represented in the endorsement graph.

User polarization. Besides having a global measure of polarization for a discussion, we are also interested in having a measure that captures how polarized an individual user is with respect to the same discussion. Since RWC is our global measure of choice, we have adapted it to define a similar, random walk-based polarization measure iRWC for individual users [3]. Intuitively, iRWC captures how much closer a user is to one side compared to the other in terms of random walk length. It takes values between -1 and 1, and a high absolute value indicates that the user is polarized towards one of the two sides.

Connecting opposing views. Our task is to recommend possible connections between users, so that polarization as measured by RWC is minimized. In the case of Twitter, for example, this would correspond to a recommendation for a user u to retweet content posted by another user v. If the recommendation were to be accepted, it would lead to the inclusion of a directed edge (u, v) in the endorsement graph G. Note that minimization of RWC is considered *in expectation*, as connections between different pairs of users have different probability to materialize. Given

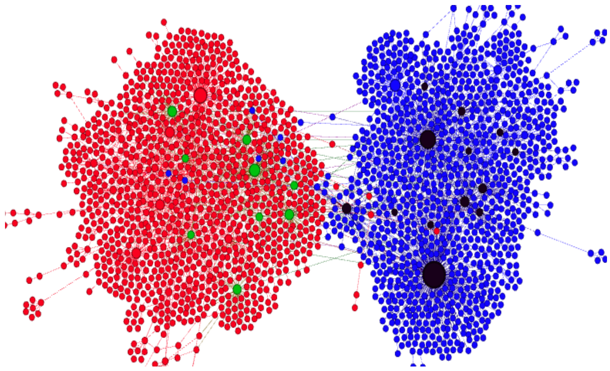

Figure 1: An example retweet graph for a political discussion in Russia related to the killing of opposition leader Boris Nemtsov. The green and black dots indicate nodes picked by our algorithm ROV-AP for recommendation.

the probability that a connection between any two pairs of users be accepted, we consider the following problem: *recommend k possible connections, so that the expected* RWC *score of the resulting endorsement graph is minimized.*

Acceptance probabilities. To obtain an estimate of the acceptance probability p for a recommended connection between users u and v, we use historical data to build a simple binomial model parametrized by the iRWC polarities of users u and v. In the case of Twitter, we essentially estimate p as the fraction of times that a user with polarity iRWC(u) has retweeted content posted by a user of polarity iRWC(v), out of all the times they have been exposed to content of such polarity. Our empirical findings confirm that indeed users of more distant polarities have lower probability to endorse each other's content.

Algorithm. One could employ a natural greedy algorithm to select the k connections to recommend. However, its running time is impractical for moderately-sized datasets, as it would require $\mathcal{O}(kn^2)$ evaluations of RWC, where n is the number of users that participate in a discussion (and are represented as nodes in the endorsement graph). Instead, we show that one can effectively reduce the running time of the greedy approach by considering only connections between nodes of sufficiently large degree. We refer to the resulting algorithm as ROV-AP (recommending opposing views with acceptance probability).

Empirical findings. We evaluate the proposed algorithm on real datasets and compare it with state-of-the-art edge-addition techniques [1, 4, 5]. The results show that it performs better both in terms of result quality and running time. By manual inspection of the recommendations produced by ROV-AP for Twitter, we find that indeed the algorithm recommends connections between user accounts that combine moderate popularity with moderate polarity. Figure 1 contains examples of recommendations generated by ROV-AP.

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

[2] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy in social media. In *WSDM '16*, pages 33–42, 2016.

[3] K. Garimella, G. De Francisci Morales, A. Gionis, and M. Mathioudakis. Quantifying controversy on social media. *Transactions on Social Computing*, 2018.

[4] 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.

[5] H. Tong, B. A. Prakash, T. Eliassi-Rad, M. Faloutsos, and C. Faloutsos. Gelling, and melting, large graphs by edge manipulation. In *CIKM '12*, pages 245–254, 2012.