Counteracting Filter Bubbles with Homophily-Aware Link Recommendations

Robert Warton[1], Chris Volny[1], and Kevin S. Xu[1,2](✉)

[1] Department of Electrical Engineering and Computer Science, University of Toledo, Toledo, OH 43606, USA
robert.warton@rockets.utoledo.edu, cvolny@gmail.com, ksx2@case.edu
[2] Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH 44106, USA

Abstract. With the prevalence of interaction on social media, data compiled from these networks are perfect for analyzing social trends. One such trend that this paper aims to address is political homophily. Evidence of political homophily is well researched and indicates that people have a strong tendency to interact with others with similar political ideologies. Additionally, as links naturally form in a social network, either through recommendations or indirect interaction, new links are very likely to reinforce communities. This serves to make social media more insulated and ultimately more polarizing. We aim to address this problem by providing link recommendations that will reduce network homophily. We propose several variants of common neighbor-based link prediction algorithms that aim to recommend links to users who are similar but also would decrease homophily. We demonstrate that acceptance of these recommendations can indeed reduce the homophily of the network, whereas acceptance of link recommendations from a standard common neighbors algorithm does not.

Keywords: Polarization · Political homophily · Link prediction · Common neighbors · Modularity reduction

1 Introduction

The Internet, particularly social media platforms, enables unprecedented opportunities for individual expression and the global exchange of thought. As social platforms vie for users and their engagement, significant gains have been made in positive link suggestions and curated content feeds. Activist and author Eli Pariser describes an emergent phenomenon that individualized content feeds and search results effectively filter conflicting viewpoints and promote isolated intellectual and information bubbles [15]. Filter bubbles can have negative impacts

R. Warton and C. Volny—Contributed equally to this work.

R. Thomson et al. (Eds.): SBP-BRiMS 2022, LNCS 13558, pp. 155–164, 2022.
https://doi.org/10.1007/978-3-031-17114-7_15

on a community by excluding potential links outside the bubble while promoting isolation within. This also leads to disrupting the spread of information through a social network. One instance of this isolation of particular interest is the divisive political climate leading up to the 2020 United States presidential election.

Many approaches for link recommendation in online social networks (OSNs), e.g. friend or follow suggestions, are functions of reciprocity (this person follows you), common neighbors (friends of friends), preferential attachment (popular page), or a composite of these strategies (your friends follow this page). Consequently, these strategies often yield recommendations that promote filter bubbles. In particular, methods based on common neighbors can give recommendations that are likely to be accepted; however, the nature of recommendations based on a person's direct neighbors are unlikely to give recommendations outside of a network's isolated communities. These recommendation systems are likely not inherently maligned and do fulfill a desirable outcome of using OSNs: to connect with like-minded peers or information sources. However, simply selecting the best match in similarity often encourages filter bubbles due to homophily—the tendency for a person to form links with similar people.

A direct approach to this problem of recommendations that encourage filter bubbles would be to give recommendations that seek to directly break up communities. This is relatively easy in principle; for instance, if a network is categorized by many small communities, accepting random link recommendations would result in successfully breaking up the communities. However, this is entirely counterintuitive to the purpose standard social network link recommendations serve in the first place. Random recommendations would be unlikely to be accepted and would be seen as poor recommendations; therefore, this is not a good solution to filter bubbles. It would instead be desirable to generate recommendations that are likely to be accepted while simultaneously breaking up the insulated communities within the network.

In this paper, we consider a Twitter network that exhibits political homophily, with nodes labeled as left, right, or neutral depending on their political affiliation. Our main contributions are as follows:

- We define a *polarization index* for a node that indicates the tendency of a node to form edges with other nodes with the same or different affiliation.
- We propose 3 homophily-aware common neighbors-based scores with the goal of providing relevant link recommendations that can reduce homophily.
- We demonstrate that our proposed homophily-aware link recommendation scores can indeed produce relevant link recommendations that, if accepted, would reduce political polarization on a Twitter network.

2 Related Work

McPherson et al.'s work in homophily in social networks [9] lays out the sociological phenomenon from which the filter bubble emerges: people's attraction to similarity in their social network. Moeller and Helberger survey the domains, scopes, and causes of filter bubble works and describes a contrast in polarization concerns between the US (political) and European (news audiences) [11]. Geschke et al. [5] proposes a three-leveled filtering process; individual, social, and technological; and uses an Agent-Based Model to study the contributions of each filter layer to content filtering and possible strategies to remediate filtering. Various works have studied the effects of content diversity over time [13] and strategies to counteract bubbles in recommendation systems [14]. Garimella et al. have studied quantifying controversy in social networks and proposed a strategy to visualize, inform, and counter polarization on controversial issues [3,4].

Mehrabi et al.'s survey on bias and fairness in machine learning [10] provides an excellent introduction into the problem domain Pariser describes in his book [15]: algorithms having unforeseen consequences in promoting bias in a system. Of the types of bias described, the problem domain of political homophily can most aptly be categorized as historical and emergent biases: networks of high assortative mixing will yield recommendations that reinforce that trend.

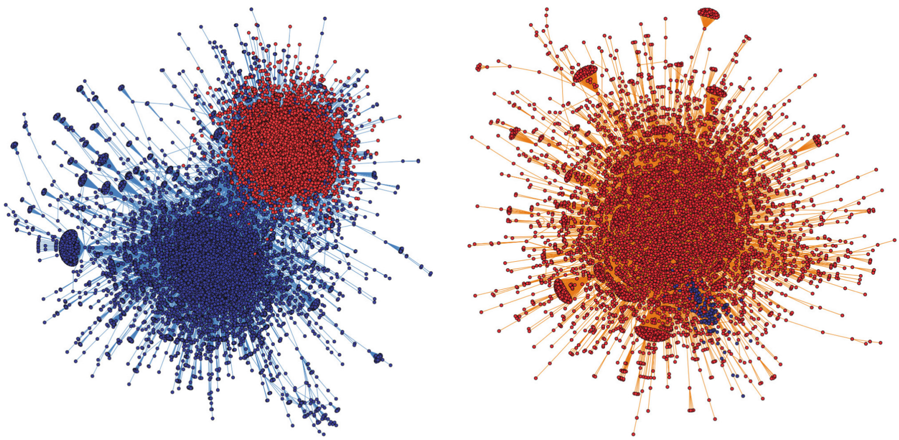
Masrour et al.'s work in fairness-aware network link prediction [8] discusses a novel approach to link prediction while maintaining fairness to a protected attribute such as gender. They utilize the modularity [12] with respect to the protected attribute as a measure of homophily and provide a model for adversarial learning to train "fair" link predictions.

3 Data Description

We consider the data collected by Conover et al. [2] from Twitter, a social network with high political community homogeneity still featuring cross-community connections. This is a dataset of mentions and retweets recorded during the Fall 2010 United States midterm election via the Twitter "gardenhose" stream into three networks: retweets, mentions, and the combination of those two.

In the retweet network, retweet events recorded a directed edge from the subscriber who retweeted to the original poster. The retweet network showed remarkably high polarization characterized as the vast majority of retweet edges occurring between politically aligned nodes. The mentions network, however, shows more lax community insulation and shows evidence that user-to-user communication tended to publish across political lines more often than with retweeting. This aligns with the findings of Majmundar et al.'s "why we retweet scale" [7] categories of to "show approval" and "showing support" users retweeting content are intentionally expanding the audience of a tweet to their own followers (implying support for the views or opinions of that entry).

Nodes within these networks were labeled as left, right, or neutral based on political affiliation using hashtag co-occurrence with the seed tags `#p2` (progressive 2.0) and `#tcot` (top conservative on Twitter): both popular and polarized hashtags for the election cycle studied. When other hashtags co-occurred with these, they were studied and, if confirmed to be politically aligned, were included in their node categorization effort. The resultant networks published are the largest connected components of each graph with the aforementioned node labels applied; this is the dataset used in this paper. A visualization of the retweet network with nodes colored by political affiliation is shown in Fig. 1.



(a) Retweet network (b) Mentions network

Fig. 1. Force-directed layouts of (a) retweet and (b) mentions networks. Red nodes denote right-leaning users, and blue nodes denote left-leaning users. Community structure is strong in the retweet network and weak in the mentions network. Figure credit: [2].

Table 1. Network properties after conversion to undirected network

Network type	Nodes	Edges	Avg. degree	Modularity
Retweet	18,470	48,053	5.20	0.475
Mentions	7,175	11,681	3.26	0.109
Combined	22,405	59,926	5.35	0.417

This dataset serves as a reasonable size of network labeled by a divisive topic. This clearly demonstrates the tendency to form isolated communities, and is a good target for trying to avoid the filter bubble phenomenon. Encouraging fairness in the flow of information is beneficial to a divided political climate where each party is likely to have an agenda.

4 Methodology

We convert the 3 networks from directed to undirected based on the norm of the Twitter platform: both users are aware of these interactions, and therefore, it can be argued this constitutes a reciprocated network connection of peer awareness. Some basic summary statistics of the networks are shown in Table 1, where the modularity is computed with respect to the political affiliation as a measure of assortative mixing [12]. Each of the three political affiliations (left, right, and neutral) is treated as its own group when computing modularity. The high modularity of the retweet network suggests that there is a lot of assortative mixing, i.e. users mostly retweet from other users with the same political affiliation.

Many of the polarizing qualities of the retweet network carried over into the combined network; albeit, dampened by cross community connections of the mentions network. We use the combined network as the main target of further analysis because it draws from both retweets and mentions and averages out some of the assortative and disassortative features of both networks.

4.1 Polarization Index

A common approach to measure homophily at the level of an individual ego node is the External-Internal (EI) homophily index [6]. The EI index is given by (1).

$$\mathrm{EI}(u) = \frac{N_{\text{external}} - N_{\text{internal}}}{N_{\text{external}} + N_{\text{internal}}} \in [-1, 1], \tag{1}$$

where N_{external} and N_{internal} denote number of edges the ego node forms to external and internal nodes, respectively. Here, a node is considered to be internal to another node if it has the same political affiliation, and external otherwise. This gives a metric that indicates if a node has a strong tendency to form links with nodes of the same political affiliation (-1) or the opposite affiliation (1). Nodes close to 0 have little tendency to prioritize links based on political affiliation. Using the EI index, neutrals would normally be considered as external to both the left and right.

However, for our purposes since the aim is to split up the more polarized groups, it is beneficial to treat neutrals as neither external nor internal to any nodes. This approach was chosen so that neutrals are considered favorably. This is a modification of the EI homophily index, which we will refer to as the *polarization index (PI)* (2).

$$\mathrm{PI}(u) = \frac{N_{\text{opposite}} - N_{\text{same}}}{N_{\text{opposite}} + N_{\text{same}} + N_{\text{neutral}}} \in [-1, 1] \tag{2}$$

The distribution of polarization indices in the combined network is shown in Fig. 2. We also make use of the absolute value of the polarization index, which can be thought of as a measure of a node's general tendency to form links based on political affiliation.

4.2 Homophily-Aware Common Neighbors

We propose 3 homophily-aware based link recommendation scores that attempt to extend common neighbor-based recommendations in a different way using political alignment.

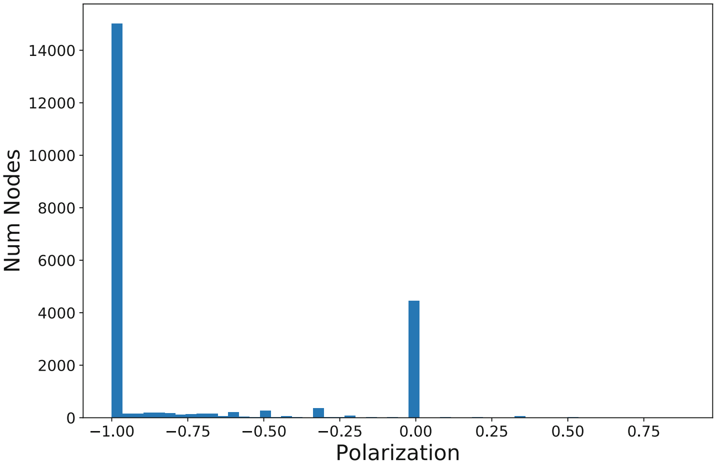


Fig. 2. Polarization index of nodes in the combined network. Most nodes have a polarization index close to −1, indicating strong political polarization.

Common Neighbors of Opposite Alignment: The first is nearly identical to common neighbors, but only considering nodes of opposite political alignment to recommend. For neutral ego nodes, we consider nodes of either left or right alignments. Recommending common neighbors of opposite alignment is a direct way of breaking up the two polarized communities while keeping the score rooted in common neighbors.

Ego Score: Our second score directly uses the polarization from a node's ego network, so we refer to it as the *ego score*. Inspired by Adamic and Adar's approach of inversely weighting common neighbors by degree to give higher weight to low-degree nodes [1], we consider weighting common neighbors with respect to polarization index. We weight common neighbors with polarization index close to 0 more strongly because they are more neutral and weight more insulated common neighbors (with polarization index close to −1 or 1) less, as shown in (3):

$$\text{Ego Score}(u, v) = \sum_{w \in \text{CN}(u,v)} \big[1 - |\text{PI}(w)|\big], \tag{3}$$

where $\text{CN}(u, v)$ denotes the set of common neighbors between nodes u and v. This favors recommendations with a high number of common neighbors, but also offers the added effect that the polarization index of the common neighbors is a major determining factor in the recommendation. Additionally, there is the benefit that the recommendations are not forced to be of opposite alignment.

Adjacent Score: The last score directly includes the number of common neighbors, but weights it by the difference of the absolute polarization of each of the nodes. The absolute value of the polarization index is in $[0, 1]$ and can be thought of as the level of insulation, regardless of whether the node prefers to connect with internal or external nodes. Nodes close to 1 will have a strong preference to connect to one type of node (either internal or external), while nodes close to 0 do not have a strong preference. We call this score the *adjacent score*, as it only includes polarization of each of the two nodes. The adjacent score shown in (4) will have a strong preference to connect nodes of high and low absolute polarization, while still factoring in the number of common neighbors $|\mathrm{CN}(u, v)|$ between u and v:

$$\text{Adjacent Score}(u, v) = |\mathrm{CN}(u, v)| \, \big||\mathrm{PI}(u)| - |\mathrm{PI}(v)|\big|. \tag{4}$$

Each of these scores gives insight into how the inclusion of homophily can affect recommendations but incorporates it in a different manner.

5 Experiments

We use the homophily-aware common neighbor scores proposed in Sect. 4.2 to generate a ranked list of recommendations for each node u. We then accept the best recommendation for node u (call this node v) by adding the edge (u, v) to the network. However, edge (u, v) may have already been added if node u was the best recommendation for node v, and node v was considered before node u.

We consider two different approaches to handle these duplicate recommendations. The first is to *allow duplicate recommendations* by not adding any new edge for node u. This results in less total edges added to the network. The second approach is to *avoid duplicate recommendations* by choosing the second best recommendation (call this node w) by adding (u, w) if the edge (u, v) for the best recommendation already exists. (If an edge with the second best recommendation already exists, then we choose the third best, and so on.)

By accepting the best recommendation for each node, we can get a direct comparison of how each of the methods is changing the network[1].

5.1 Evaluation Metrics

An effective homophily-aware link recommendation should have two qualities:

1. *Link recommendation quality:* The user should be likely to accept the recommendation.
2. *Polarization reduction:* If accepted, the polarization of the network should decrease.

The polarization reduction can be measured by comparing the polarization of the network before and after accepting the recommendation. We compute the *modularity change* to be the difference in modularity from the original network to the new network after accepting the best recommendation for each node. We

[1] Code to reproduce experiments is available at https://github.com/IdeasLabUT/Homophily-Aware-Link-Recommendation.

use the modularity change as the evaluation metric for change in polarization due to recommendation.

To measure the link recommendation quality, one would likely have to deploy the homophily-aware link recommendation algorithm onto an online social network and then compare acceptance rates of recommendations compared to a baseline recommendation algorithm that does not consider political homophily, e.g. common neighbors (CN). In the absence of such a deployment, we consider an alternative approach. We assume that CN-based recommendations will give good recommendations (i.e. likely to be accepted). With this in mind, we can investigate the similarity between the recommendations given by our scores with the potential CN recommendations.

We consider 3 evaluation metrics for these similarities. First, for each ego node, we examine the rank of the top recommendation in the ranked list of CN recommendations. A lower rank indicates that the recommendation is better matched with CN recommendations, with the best possible rank of 1 indicating that our recommendation aligns with CN. We consider the *mean rank* over all ego nodes in the network. Since the mean rank may be prone to outliers due to a small number of recommendations having very high ranks, we also consider the *median rank* over all ego nodes.

Additionally, for each ego node, we compute the number of CNs that our top recommendation has with the ego node. We divide by the maximum number of CNs that any node may have with the ego node to get a proportion of maximum CNs. A higher proportion indicates better agreement with CN recommendations, with the best possible proportion of 1 indicating that our recommendation has the same number of CNs as the recommendation by the CN score. We consider the *mean proportion of maximum CNs* over all ego nodes in the network.

All 3 of these metrics give us an idea of how well matched our top recommendation is with CN, which we use as a proxy for how likely our top recommendation is to be accepted.

6 Results

The results of our experiments are shown in Table 2. First, notice that accepting the top recommendation from the baseline common neighbors (CN) score barely changes the modularity of the network—in other words, it is not increasing polarization but also not decreasing polarization. On the other hand, each of the 3 homophily-aware scores reduces modularity, with the highest reduction achieved by simply recommending the node of the opposite political alignment with the most CNs, denoted by CN (opposite). Thus, all the homophily-aware link recommendations do indeed decrease the political polarization of the network, although the decrease is very small for the Ego Score (3).

Next, we consider the similarity of the recommendations to CN recommendations. CN (opposite) generally achieves the worst link recommendation quality of the 3 homophily-aware scores, with the lowest mean proportion of maximum CNs. The Ego Score (3) achieves the best link recommendation quality, which is not much lower than that achieved by the baseline CN score.

Table 2. Link recommendation results assuming top recommendation is accepted (with different ways of handling duplicate top recommendations as described in Sect. 5). Median rank, mean rank, and mean proportion of maximum CNs measure link recommendation quality. Modularity change measures polarization reduction. Best result for each metric is shown in bold.

(a) Allowing duplicate recommendations				
Evaluation Metric	Common Neighbors	CN (opposite)	Ego Score (3)	Adjacent Score (4)
Modularity change	−0.008	**−0.173**	−0.017	−0.082
Median rank	**1**	6	**1**	6
Mean rank	**1**	17.38	8.93	19.42
Mean proportion of maximum CNs	**1**	0.831	0.946	0.906
# of added edges	21,868	19,076	21,816	**21,909**
(b) Avoiding duplicate recommendations				
Evaluation Metric	Common Neighbors	CN (opposite)	Ego Score (3)	Adjacent Score (4)
Modularity change	−0.009	**−0.175**	−0.018	−0.084
Median rank	**1**	6	**1**	6
Mean rank	**1.02**	17.38	9.28	19.28
Mean proportion of maximum CNs	**0.999**	0.831	0.944	0.906
# of added edges	22,261	19,322	22,254	**22,356**

Between the 3 homophily-aware recommendation scores, we can observe different trade-offs. CN (opposite) is forced to recommend a node on the opposite end of the political spectrum, even when no such node with many CNs exist, so accepting its recommendation will always reduce modularity. For about 3,000 ego nodes, no such node with *any* CNs exist, so it cannot even make a recommendation! This is why the number of added edges for CN (opposite) is about 3,000 lower than for the other methods. Thus, the evaluation metrics for link recommendation quality for CN (opposite) may even be too favorable, since it avoids making recommendations for ego nodes that don't have any CNs with any nodes of the opposite political affiliation. Of the 3 homophily-aware scores, the Adjacent Score seems to offer the best balance between the recommendation quality and polarization reduction.

7 Conclusion

This paper provides evidence and methods for how the use of homophily-aware link recommendation could decrease political polarization on social media. We proposed 3 approaches to provide relevant link recommendations while also potentially reducing homophily, if accepted. These approaches are all based on common neighbors. The main limitation in our study is the evaluation of link recommendation quality, which we performed by comparing our recommendations to those of a common neighbors score. To better evaluate the recommendation quality and impact of our homophily-aware approaches, they would need to be deployed onto an online social network to evaluate how they change acceptance rates of recommendations.

Acknowledgements. We thank the anonymous reviewers for their suggestions to improve this paper. This material is based upon work supported by the National Science Foundation grants IIS-1755824 and IIS-2047955.

References

1. Adamic, L.A., Adar, E.: Friends and neighbors on the Web. Soc. Networks **25**, 211–230 (2003)
2. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on Twitter. In: Proceedings of the 5th International AAAI Conference on Web and Social Media, pp. 89–96 (2011)
3. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M.: Mary, mary, quite contrary: Exposing Twitter users to contrarian news. In: Proceedings of the 26th International Conference on World Wide Web Companion, pp. 201–205 (2017)
4. Garimella, K., Morales, De Francisci Morales, G., Gionis, A., Mathioudakis, M.: Quantifying controversy on social media. ACM Trans. Soc. Comput. **1**(1), 1–27 (2018)
5. Geschke, D., Lorenz, J., Holtz, P.: The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers. Br. J. Soc. Psychol. **58**(1), 129–149 (2019)
6. Krackhardt, D., Stern, R.N.: Informal networks and organizational crises: an experimental simulation. Soc. Psychol. Quart. **51**(2), 123–140 (1988)
7. Majmundar, A., Allem, J.P., Boley Cruz, T., Unger, J.B.: The why we retweet scale. PLoS ONE **13**(10), e0206076 (2018)
8. Masrour, F., Wilson, T., Yan, H., Tan, P.N., Esfahanian, A.: Bursting the filter bubble: Fairness-aware network link prediction. In: Proceedings of the 34th AAAI Conference on Artificial Intelligence, pp. 841–848 (2020)
9. McPherson, M., Smith-Lovin, L., Cook, J.M.: Birds of a feather: homophily in social networks. Ann. Rev. Sociol. **27**(1), 415–444 (2001)
10. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Comput. Surv. **54**(6), 1–35 (2021)
11. Moeller, J., Helberger, N.: Beyond the filter bubble: Concepts, myths, evidence and issues for future debates. Tech. rep., University of Amsterdam (2018). https://dare.uva.nl/search?identifier=478edb9e-8296-4a84-9631-c7360d593610
12. Newman, M.: Networks. Oxford University Press, 2nd edn. (2018)
13. Nguyen, T.T., Hui, P.M., Harper, F.M., Terveen, L., Konstan, J.A.: Exploring the filter bubble: the effect of using recommender systems on content diversity. In: Proceedings of the 23rd International Conference on World Wide Web, pp. 677–686 (2014)
14. Oku, K., Hattori, F.: User evaluation of fusion-based approach for serendipity-oriented recommender system. In: Proceedings of the Workshop on Recommendation Utility Evaluation: Beyond RMSE, pp. 39–44 (2012)
15. Pariser, E.: The filter bubble: What the Internet is hiding from you. Penguin UK (2011)