





# A Heterophily-Based Polarization Measure for Multi-community Networks

Sreeja Nair<sup>1</sup> and Adriana Iamnitchi<sup>2</sup>

<sup>1</sup> University of South Florida, Tampa, USA  
sreejas@usf.edu

<sup>2</sup> Maastricht University, Maastricht, Netherlands  
a.iamnitchi@maastrichtuniversity.nl

**Abstract.** This work proposes a heterophily-based metric for quantifying polarization in social networks where multiple ideological, antagonistic communities coexist. This metric captures node-level polarization and is built on user's affinity towards other communities rather than their own. Node-level values can then be aggregated at the community, network, or sub-network level, providing a more detailed map of polarization. We tested our metric on the Polblogs network, White Helmets Twitter interaction network with two communities and the VoterFraud2020 domain network with five communities. We also tested our metric on dK-random graphs to verify that it results in low polarization scores, as expected. Finally, we compared our metric with two widely used polarization measures: Guerra's polarization index and RWC.

**Keywords:** Polarization · Communities · Heterophily

## 1 Introduction

Different polarization metrics have been proposed in the literature from several vantage points, including network topology [13,14,17], content semantics and sentiment [5,7]. Current network-based polarization measures [13,14,17] are tailored based on the assumption that a polarized network consists of two opposing communities. According to Esteban et al. [11], individuals can be grouped into multiple, antagonistic communities in a polarized society. Most efforts on measuring polarization assume that the polarized networks consist of exactly two antagonistic groups, and thus need to ignore the neutral nodes, or add them to one of the extreme groups. Our metrics address this limitation by acknowledging the existence of multiple communities.

This paper proposes a heterophily-based polarization metric called “cross-community affinity” that can be applied to networks with two or more communities with conflicting positions, goals, and viewpoints. We consider these communities are placed equi-distantly in a one-dimensional space. This assumption is supported by two facts: First, it allows us to compare with other metrics in the literature. Second, it reflects the datasets we use for our empirical evaluation.

The cross-community affinity of a node represents the node’s affinity to communities with a different ideology than its own. Our proposed metric measures the node-level value that can be aggregated to any higher level, such as the community level, the network level, or any sub-network level. With this approach, we can understand which nodes or communities contribute most to polarization, enabling a more detailed picture and the possibility of directing interventions to particular nodes.

The rest of the paper is structured as follows. Section 2 presents the relevant works in this area. Section 3 explains the metric we propose. Section 4 describes the datasets used in this study and reports the results of experiments performed. Section 5 summarizes the results and discusses the future work.

## 2 Polarization Metrics in the Literature

Measuring polarization using structural characteristics inferred from network representations of social or political systems is a common topic in the literature, along with two other approaches: survey-based approaches [9], which measure distributional properties of public opinion through surveys; and content-based approaches [4, 8, 21], which use NLP tools to identify opposing groups on the network.

Conover et al. [6] suggest that polarization has a significant impact on the structures of social networks because it results in the formation of two groups that are well connected within themselves but have few ties to one another. Guerra et al. [14] present a polarization metric that centers on investigating nodes that belong to the community boundary, which captures the concepts of antagonism and polarization. Another polarization metric, the Polarization index [17], measures how far apart two groups are in terms of ideology, assuming their populations are equal. Garimella et al. [13] established the Random Walk Controversy (RWC) metric, which uses the random walk to see how likely information is to stay inside or reach out to other groups. Salloum et al. [20] examine the polarization measures mentioned above via simulations and demonstrate that all of them produce high polarization scores even for random networks with density and degree distributions close to typical real-world networks.

However, these metrics are developed based on the assumption that the polarized network consists of exactly two communities. In this paper, we propose a heterophily-based polarization metric called cross-community affinity, which measures the affinity of a node to other clusters rather than its own.

## 3 Cross-community Affinity: A Heterophily-Based Polarization Metric

We propose a new polarization metric called cross-community affinity to serve two specific objectives. First, it should adapt to a variable number of ideological groups connected by different antagonizing forces. Second, we want this metric to be applicable to different granularity, from node to full network and other network-based groupings in between.

As in previous work [13, 14], the basis of this polarization metric is a node’s connectivity with groups other than its own. In order to capture that, we introduce a heterophily-based metric consistent with its definition [15] that captures how a node is connected to different groups via both direct and indirect links. We assume the polarization of a network is the inverse of the average cross-community affinity, that is:

$$\text{Polarization} = - \text{Avg. cross-community affinity}$$

In order to define the metric, we use the following intuition. First, a network can have multiple communities. We assume a constant between 0 and 1 represents an ideological distance between different groups. Intuitively, a connection with an ideologically opposite node should weigh differently than a connection with an ideologically similar node. In a political system, one could consider the difference between a far right—far left connection vs. a leaning right and center political positioning connection. To account for the ideological difference between communities, we define communities as being in a one-dimensional space and equally spaced apart. The datasets (described in Sect. 4.1) we looked at implicitly position themselves on the one-dimensional space. Specifically, our use of the VoterFraud2020 dataset labeled using Media Bias Fact Check considers political orientation in a uni-dimensional space. We are providing a weight factor to represent the ideological distance. For simplicity, we consider that the distance between consecutive communities is constant and equal to  $\frac{1}{|C|-1}$ , where  $C$  is the number of communities (as shown in Appendix A.4). This assumption can, of course, be relaxed in a scenario in which, for example, the ideological distance between extreme left and leaning left is smaller than between leaning left and center.

Second, we assume that both direct and indirect connections can have an impact on a node’s cross-community affinity. However, as well accepted in the literature [12], indirect connections have a much smaller impact on one’s beliefs than direct connections. It has been empirically observed by Friedkin [12] that people’s awareness of others’ actions is restricted to people who were either in direct contact or had at least one contact in common. Moreover, the impact of such connections is typically a function of the overall number of connections a node has: the more neighbors, the less the impact of any one neighbor may be. To implement this, we assume that the ideological difference between nodes from the same community is  $-1$ . This value was chosen such that a node’s affinity for its community reduces its cross-community affinity.

We thus define cross-community affinity (CCA) of a node  $i$  as the sum between the effects of its direct neighbors and indirect neighbors on its ideological openness:

$$CCA(i) = DNE(i) + \alpha \times INE(i) \quad (1)$$

where  $DNE(i)$  is the direct neighbor effect on node  $i$  and  $INE(i)$  the indirect neighbor effect on node  $i$ .  $\alpha$  is the impact factor of the indirect neighbor effect. For simplicity we consider  $\alpha = 1/h$ , where  $h$  is the number of social hops between node  $i$  and the given set of nodes (in this case  $h = 2$ ).

We consider the direct neighbor effect on node  $i$  as the sum of the relative impact of  $i$ 's direct neighbors as follows:

$$DNE(i) = \sum_{c \in C} w_{(s(i),c)} \times \frac{k_c(i)}{k(i)} \quad (2)$$

where  $C$  is the set of communities in the network,  $s(i)$  is the community to which node  $i$  belongs,  $w_{(s(i),c)}$  is the ideology based distance between  $i$ 's community and community  $c$ .  $k_c(i)$  denotes the number of neighbors of  $i$  in the community  $c$  and  $k(i)$  denotes the total number of neighbors of  $i$ . Similarly, we consider the indirect neighbor effect on  $i$  as the average of the relative effects of its 2-hop neighbors over all different communities.

$$INE(i) = \frac{1}{|C_{N(i)}|} \sum_{c \in C_{N(i)}} ANE_c(i) \quad (3)$$

where  $C_{N(i)}$  is the set of communities in the  $i$ 's neighborhood and  $|C_{N(i)}|$  is the number of communities in the  $i$ 's neighborhood.  $ANE_c$  represents the average neighbor effect of  $i$ 's immediate neighbors by examining neighbors' neighborhood. We calculate the individual neighbor effect of each neighbor of node  $i$  to determine how their neighbors are distributed throughout the communities. To determine the impact of neighbor  $j$  on node  $i$  we calculate neighbor effect(NE)  $j$  on  $i$  as follows:

$$NE(j, i) = \sum_{g \in C} w_{(s(i),g)} \times \frac{k_g(j)}{k(j) - 1} \quad (4)$$

where  $g$  is the community to which node  $j$ 's neighbors belong,  $w_{(s(i),g)}$  is the ideology distance between  $i$ 's community and community  $g$ ,  $k_g(j)$  represents the number of  $j$ 's neighbors in community  $g$  and  $k(j)$  is the total number of  $j$ 's neighbors, from which we exclude  $i$ .

$CCA(i)$  has a value ranging from  $-1.5$  to  $1.5$ . The  $CCA(i)$  is minimum ( $CCA(i) = -1.5$ ) if all nodes in the immediate and two-hop neighborhood belong to the same community as node  $i$ . If all neighbors up to two hops away are in the node's extreme opposite community, the cross-community affinity is maximum ( $CCA(i) = 1.5$ ). Cross-community affinity can thus be aggregated at different granularities, from node-specific to any grouping of nodes in the networks, whether connected or not by, for example, averaging the node-specific affinity. A node-specific cross-community affinity can tell whether the node contributes to the network polarization. The network-level polarization  $P$  can thus be obtained as the negative average cross-community affinity:

$$P = -\frac{1}{|N|} \sum_{i \in N} CCA(i) \quad (5)$$

where  $N$  is the set of nodes in the network. Appendix A.3 shows the different scenarios of a network and their respective CCA.

## 4 Empirical Evaluation

We evaluate our proposed metric on networks with different numbers of ideological groups. We use three datasets: Polblogs [3] and White Helmet Twitter interaction network [19] which each have two antagonistic communities, and the VoterFraud2020 domain network [18], with five communities.

### 4.1 Datasets

The Polblogs network [3] is a publicly available network of hyperlinks between political blogs about politics leading up to the 2004 United States presidential election. Each node in this network is labelled as either conservative (right) or liberal (left). Edges are the interaction between blogs such as citation, blogroll links etc. We consider the network as an undirected labelled network.

White Helmets Twitter dataset is the interaction network [19] based the tweets on White Helmets for a period from April 2018 to April 2019. Each node in this network is labelled as either pro-White Helmets or anti-White Helmets.

The VoterFraud2020 domain network [18] is derived from the publicly available VoterFraud2020 dataset [2], a Twitter dataset related to voter fraud claims about the US 2020 Presidential election. In this network, nodes are the web domains of URLs posted in tweets, and links connect domains that were tweeted by the same user. This network of websites is structurally divided into communities. Each node is labeled based on its media bias and credibility using publicly available source Media Bias Fact Check (MBFC) [1]. The labels are: right, right-center, center, left-center, and left. However, after this labeling strategy, 75.6% of the nodes remained unknown because they are not included in the MBFC database. To assign labels to the ‘unknown’ nodes, we relabelled them as the dominant label in the node’s direct neighborhood. That is, we started with unlabeled nodes with the largest proportion of labeled nodes in their one-hop neighborhood and labeled them as the majority. We recursively applied this methodology until all nodes were labeled. Edge distribution of this network is depicted in Appendix A.1 Table 1 shows the network properties of Polblogs, White Helmets twitter network and VoterFraud2020 domain network. Appendix A.2 depicts the visual representation of these datasets.

**Table 1.** Network properties of Polblogs, White Helmets twitter network and the VoterFraud2020 domain network.

	Nodes	Edges	Average degree	Assortativity	Clustering Coefficient
Polblogs	1224	16715	27.3	-0.22	0.32
White Helmets	9024	11641	2.6	-0.47	0.14
VoterFraud2020	2461	48189	39.2	-0.33	0.68

## 4.2 Cross-community Affinity in the Polblogs Network

Polblogs networks has two communities: conservative (right) and liberal (left). The edges connecting two communities are only 9%. 50.9% nodes (623 nodes) have connections to the opposite community. First, the ideology-based distances between these two communities are defined. As discussed in the preceding section, the ideology-based distance between the same communities is  $-1$ . In contrast, the connection to the most polar community gets the maximum weight of 1. In the Polblogs network, the weights between conservative-conservative and liberal-liberal are  $-1$  and the weight between conservative and liberal is 1.

We compute the average cross-community affinity value across each community and the entire network to determine cross-community affinity at the community and network levels. Using Eq. 5, the polarization score of conservative is 1.13, liberal is 1.0, and the network is 1.07. These values indicate that the communities and the whole network are polarized. One of the key benefits of having a metric that captures polarization at the node level is that, (Appendix A.2 Fig. 3a) we can determine which nodes contribute to the polarization.

Next, we evaluate how the metric works on a random graph. Our intuition is that randomizing the network should reduce polarization [20]. We generate a set of random networks using dK series [16]. dK-series generate random graphs that preserve desired prescribed properties of the original.  $0K$  ( $d = 0$ ) creates the Erdős-Rényi network with the same average node degree as the original graph.  $1K$  ( $d = 1$ ) creates the configuration model, fixing the degree sequence of the original graph. As compared to the polarization score of 1.07, the average polarization value for generated  $0K$  is 0.58 and  $1K$  is 0.02. These networks have lower polarization score than the original Polblogs network, which means that they are less polarized. This observation gives us confidence that measuring polarization using the methodology we proposed captures random behavior.

## 4.3 Cross-community Affinity in the White Helmets Twitter Interaction Network

We conducted a similar experiment on the White Helmets Twitter network. Around 73% of users are anti-White Helmets, and 27% are pro-White Helmets. The size of the communities is significantly different from the Polblogs, where it has an almost similar size for communities (52% and 48%). The connection between anti-White Helmets and pro-White Helmets users is 0.3%. Only 0.2% of users have interaction with the opposite community. As in the Polblogs experiment, we used  $-1$  for the ideology-based distance between the same communities and 1 for the opposing communities. The polarization score for the network and each community is 1.49. The score indicates that the network is highly polarized.

Next, we created a set of  $0K$  and  $1K$  graphs for the White Helmets Twitter dataset. The average polarization score for  $0K$  graphs is 0.94 and for  $1K$  graphs is 0.35. Consistent with the Polblogs results, the random graphs generated for White Helmets also yield lower polarization score indicating they are less polarized.

#### 4.4 Cross-community Affinity in the VoterFraud2020 Domain Network

Next the experiment is conducted on the VoterFraud2020 domain network with five communities. Given that there are five communities in the network, the distance between two adjacent communities is defined as  $1/(|C| - 1)$ , or  $1/4$  (Appendix A.4). The polarization scores computed using Eq. 5 for each community are: right: 0.88; right-center:  $-0.34$ ; center:  $-0.19$ ; left-center: 0.62; left: 0.05; and for the entire network: 0.61. The right-center and center communities are less polarized compared to other communities. That is because the right-center has a comparatively higher number of edges to the right community. Similarly, the center community contains more links to left-center and right. Even if the network-level polarization score shows that the network is polarized, the community-level polarization score reveals that two communities do not contribute to the polarization state of the network. Using only a network-level polarization score, it is impossible to determine how different communities contribute to polarization, thus obscuring information that might be useful in limiting damage or directing intervention.

We also created random networks via dK-distributions. A set of random graphs with same number of nodes and same average degree (0K) are generated. The average polarization score for 0K graphs is  $-0.06$  and 1K graphs is 0.02 compared to the original network’s score is 0.61. The polarization value dropped for the random networks even when the network’s degree sequence was preserved. More experiments on VoterFraud2020 domain network are shown in Appendix A.5

#### 4.5 Comparison with Existing Polarization Metrics

In this section, we compare our cross-community affinity metric with two widely used polarization metrics: Guerra’s polarization index (PI) [14] and random walk controversy score (RWC) [13]. The RWC score has been described as state-of-the-art [10, 22]. The range of polarization values for PI is  $-0.5$  to  $0.5$ , and RWC is  $-1$  to  $1$ . Our metric P ranges from  $-1.5$  to  $1.5$ . The higher the value, the higher the polarization. Table 2 shows the polarization value calculated using P, PI, and RWC. We can see that polarization values reduce consistently for Polblogs random networks. The PI value for White Helmets-0K increased compared to

**Table 2.** Comparison of polarization value calculated by P, PI, and RWC

	Polblogs			WhiteHelmets			VoterFraud2020		
	Original	0K	1K	Original	0K	1K	Original	0K	1K
P	1.07	0.58	$-0.00$	1.49	0.94	0.34	0.61	$-0.06$	0.02
PI	0.18	$-0.02$	$-0.32$	0.22	0.34	0.02	N/A	N/A	N/A
RWC	0.42	0.22	$-0.05$	0.92	0.45	$-0.04$	N/A	N/A	N/A

the original White Helmets dataset. This shows that the PI failed to capture the randomness of the network. P and RWC show a consistent drop in value, indicating that random networks show low or no polarization. The results also show that our metric works consistently as the current state-of-the-art metric, RWC. PI and RWC for VoterFraud2020 are N/A because of multiple communities.

According to Salloum et al. [20], RWC displays a severe problem related to hubs. RWC captures how likely a random user on either side is to be exposed to an authoritative user (higher degree node) from the opposing side. Even in a non-polarized network, a random network with one or more hubs can keep the random walker confined to its community, producing a high polarization value. CCA calculates a polarity score for each node separately. So, having one or more hubs will not affect our metric. Another issue with RWC is that we need to specify the parameter ‘k,’ which represents the number of authoritative users in each group. While doing experiments, we noticed that the same graph produces different polarization scores with a value ‘k’ change. So we need to be extra mindful while using RWC. Another limitation of RWC acknowledged by the author [13] is that it reports low controversy score for the Karate Club network with 34 nodes and 78 edges. The author mentions that the graph may be too small for random-walk-based measures to function correctly. According to the literature the RWC score for the Karate Club network is 0.11 whereas our polarization metric shows 1.02. Our polarization metric performs appropriately for networks with small size.

## 5 Summary

This paper proposes the cross-community affinity polarization metric as a new way to measure polarization. The cross-community affinity is a heterophily-based measure that captures the connectedness of nodes to groups other than their own. It has two specific goals. First, it adapts to a different number of ideological groups. Second, it applies to various levels of granularity, ranging from individual nodes to entire networks, as well as other network-based groups in between. The network-level polarization score can be obtained as the negative of average cross-community affinity. We evaluate our proposed metric on networks with multiple ideological groups. In addition, we compared them to randomized versions of our network datasets generated using dK distributions. The results show lower polarization values for the randomized networks. We also compared our metric with two widely used existing polarization metrics.

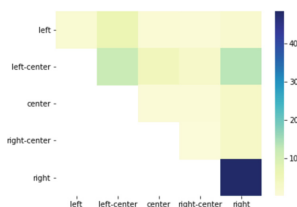
Our work has limits that merit mentioning. First, for simplicity we consider ideological difference to be a one-dimensional space. Second, our metric is now tailored to undirected unweighted networks. These are essential agenda items for future research. With a metric that captures polarization at the node level, it is possible to determine which nodes or communities contribute to the polarization. Assessing how distinct communities contribute to polarization is feasible, providing knowledge that may be valuable for limiting damage or directing intervention.



## Appendix A Additional Materials

### A.1 Edge Distribution of VoterFraud2020 Domain Network

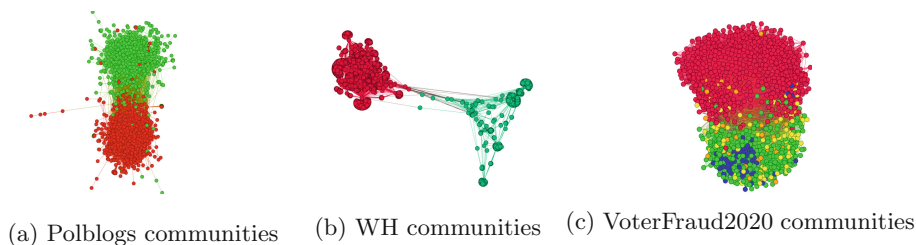
Figure 1 depicts the distribution of edges to communities after relabelling unknown based on the dominant label in the node’s direct neighborhood. The majority of the edges in the right community are to themselves. Left-center also has most of its edges to left-center and right.



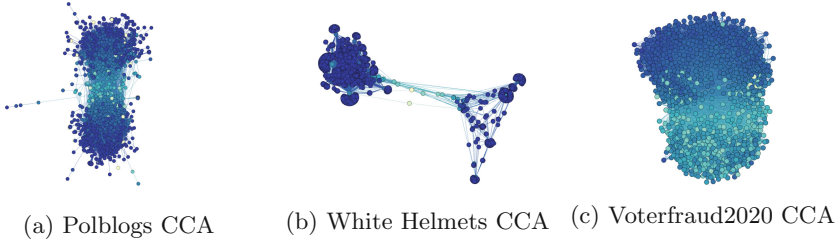
**Fig. 1.** Heatmap of the edge distribution between communities in the VoterFraud2020 domain network with five communities: left, left-center, center, right-center, right.

### A.2 Visual Representation of Datasets

**Polblogs.** Figure 2a shows the visualization of Polblogs community structure. Light green represents liberal and red represents conservative. Figure 3a displays the Polblogs network colored based on the cross-community affinity of each node. The greater the value, the lighter the shade. Different nodes have distinct hues, which demonstrates that their values vary. The figure is dominated by a darker hue, indicating that the majority of nodes have low cross-community affinity values, resulting in a polarized network.



**Fig. 2.** (a) The Polblogs interaction network: light green represents liberal nodes, and red represents conservative; (b) The White Helmets Twitter network. The colors red represents anti-White Helmets and green represents pro-White Helmets; (c) The Voter-Fraud2020 domain network, with colors reflects the node’s political orientation: red for right, orange for right-center, yellow for center, green for left-center and blue for left.



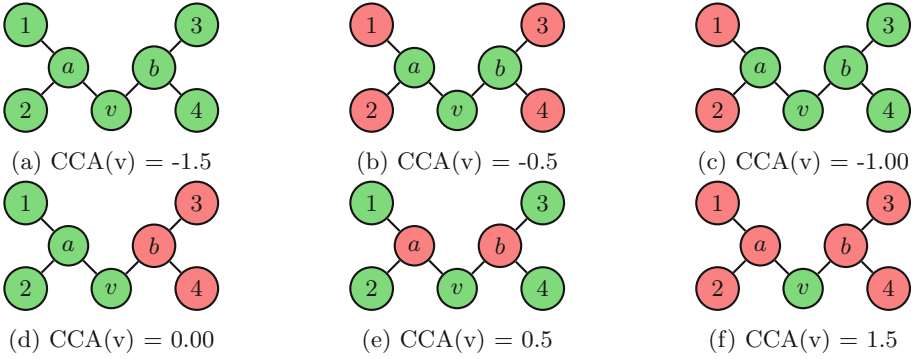
**Fig. 3.** (a) The Polblogs network colored based on cross-community affinity (b) The White Helmets Twitter colored based on cross-community affinity. (c) The Voter-Fraud2020 domain network colored based on cross-community affinity. The darker the color, the lower the value, thus higher polarization.

**White Helmets Twitter Interaction Network.** Figure 2b depicts the visual representation of communities in the WhiteHelmets interaction network. The colors red represents anti-White Helmets and green represents pro-White Helmets. Figure 3b displays the White Helmets network colored based on the cross-community affinity of each node. The greater the value, the lighter the shade. The figure is dominated by a darker hue, indicating a polarized network.

**VoterFraud2020 Domain Network.** The visual representation of communities in the network is shown in Fig. 2c. The color reflects the political orientation of the nodes, with red for right, orange for right-center, yellow for center, green for left-center, and blue for left. The right (47.4%) and the left-center (35.5%) constitute the majority of the network. The center has 7.4% nodes, the left has 6.1%, and the right-center has 3.6% nodes. Figure 3c shows the VoterFraud2020 domain network colored based on nodes' cross-community affinity value. Darker hue means low cross-community affinity. Overall, the graph shows darker shade indicating that the network is polarization.

### A.3 Scenarios of Network for CCA Calculation

Figure 4 depicts various scenarios of a network with seven nodes and two communities: red and green. For each scenario, the cross-community affinity for node  $v$  is provided. In scenario Fig. 4a all the immediate neighbors and two-hop neighbors of node  $v$  are members of the same community, indicating the absence of cross-community affinity. In this instance,  $CCA(v)$  has a value of -1.5. Similarly, in scenario Fig. 4f, all neighbors inside a two-hop neighborhood belong to the opposing community, resulting in a node with maximum cross-community affinity, where,  $CCA(v)$  equals 1.5.  $CCA(v) = 0$  in the Fig. 4d, because the neighborhood of node  $v$  is equally distributed among both communities.



**Fig. 4.** Different scenario of a network with 7 nodes and two communities: red and green. The cross-community affinity of node  $v$ ,  $CCA(v)$  is shown. (Color figure online)

**A.4 Ideological Distance for Five Communities**

Table 3 shows the distance between communities in a scenario in which we consider  $C = 5$ . Right and left communities are on the ends of the spectrum. As a result, the distance between them is 1.

**Table 3.** The ideological distance between 5 communities for VoterFraud2020 domain network

	Right	Right-center	Center	Left-center	Left
Right	-1	1/4	2/4	3/4	1
Right-center	1/4	-1	1/4	2/4	3/4
Center	2/4	1/4	-1	1/4	2/4
Left-center	3/4	2/4	1/4	-1	1/4
Left	1	3/4	2/4	1/4	-1

**A.5 Relabelling VoterFraud2020 Domain Network**

We randomized the labeling of the nodes of VoterFraud2020 domain network to see the effect on the polarization metric. We relabelled the network in three ways. In the first case, we randomly relabelled “unknown” without altering the community size of labeled nodes. We determined the number of “unknown” required by each community to maintain the community sizes. Then, “unknown” was arbitrarily assigned to each community. Even though we relabel “unknown” only, 75.6% of nodes in the network are “unknown”, making the network at least 75% random.

In the second case, we relabelled all the nodes in the network randomly but kept the number of nodes in each community the same as in the original

VoterFraud2020 domain network. In the third case, the five labels are equally distributed to the network, thus creating five equal-sized communities. These random labeling are performed 10 times, and the results provided are averaged over these outcomes. The polarization scores for these experiments are given in Table 4. The left-center community score in the experiment where “unknown” labels are randomly assigned shows that the community is polarized, indicating that this network is not totally random. The network had negative polarization values in all randomization experiments, indicating a lack of polarization.

**Table 4.** Community-level and network-level polarization score for VoterFraud2020 domain network with different labellings of unknown nodes.

	Original network	Random labelling of “unknown”	Random labelling keeping the original community sizes	Random labelling with equally divided community sizes
Right	0.88	-0.52	-0.44	-0.40
Right-center	-0.34	-0.37	-0.20	-0.18
Center	-0.19	-0.04	-0.09	-0.21
Left-center	0.62	0.11	-0.22	-0.26
Left	0.05	-0.38	-0.48	-0.48
Network	0.61	-0.20	-0.28	-0.31

## References

1. Media Bias/Fact Check News. <https://mediabiasfactcheck.com/>. Accessed 25 July 2022
2. Abilov, A., Hua, Y., Matatov, H., Amir, O., Naaman, M.: Voterfraud 2020: a multi-modal dataset of election fraud claims on Twitter. In: ICWSM (2021)
3. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 U.S. election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery, LinkKDD 2005, pp. 36–43. Association for Computing Machinery, New York (2005)
4. Belcastro, L., Cantini, R., Marozzo, F., Talia, D., Trunfio, P.: Learning political polarization on social media using neural networks. *IEEE Access* **8**, 47177–47187 (2020)
5. Borge-Holthoefer, J., Magdy, W., Darwish, K., Weber, I.: Content and network dynamics behind egyptian political polarization on Twitter. In: Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pp. 700–711 (2015)
6. Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on Twitter. In: Fifth International AAAI Conference on Weblogs and Social Media, January 2011
7. Darwish, K.: Quantifying polarization on twitter: the kavanaugh nomination. In: Weber, I., Darwish, K.M., Wagner, C., Zagheni, E., Nelson, L., Aref, S., Flöck, F. (eds.) *SocInfo 2019*. LNCS, vol. 11864, pp. 188–201. Springer, Cham (2019). [https://doi.org/10.1007/978-3-030-34971-4\\_13](https://doi.org/10.1007/978-3-030-34971-4_13)

8. Demszky, D., Garg, N., Voigt, R., Zou, J.Y., Gentzkow, M., Shapiro, J.M., Jurafsky, D.: Analyzing polarization in social media: Method and application to tweets on 21 mass shootings. In: NAACL (2019)
9. DiMaggio, P., Evans, J., Bryson, B.: Have American's social attitudes become more polarized? *Am. J. Sociol.* **102**(3), 690–755 (1996)
10. Emamgholizadeh, H., Nourizade, M., Tajbakhsh, M.S., Hashminezhad, M., Esfahani, F.N.: A framework for quantifying controversy of social network debates using attributed networks: biased random walk (brw). *Soc. Netw. Anal. Min.* **10**, 1–20 (2020)
11. Esteban, J., Ray, D.: On the Measurement of Polarization. *Econometrica* **62**(4), 819–851 (1994)
12. Friedkin, N.E.: Horizons of observability and limits of informal control in organizations. *Soc. Forces* **62**(1), 54–77 (1983)
13. Garimella, K., Morales, G., Gionis, A., Mathioudakis, M.: Quantifying controversy in social media. *ACM Trans. Soc. Comput.* **1**, May 2015
14. Guerra, P.H.C., Meira, W., Cardie, C., Kleinberg, R.D.: A measure of polarization on social media networks based on community boundaries. In: ICWSM (2013)
15. Lozares, C., Verd, J.M., Cruz, I., Barranco, O.: Homophily and heterophily in personal networks. from mutual acquaintance to relationship intensity. *Quality & Quantity* **48**, September 2014
16. Mahadevan, P., Krioukov, D., Fall, K., Vahdat, A.: Systematic topology analysis and generation using degree correlations. *ACM SIGCOMM Computer Communication Review* **36**, June 2006
17. Morales, A.J., Borondo, J., Losada, J.C., Benito, R.M.: Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: Interdisciplinary J. Non-linear Sci.* **25**(3), 033114 (2015)
18. Nair, S., Iamnitchi, A.: The polarized web of the voter fraud claims in the 2020 US presidential election. In: Workshop Proceedings of the 15th International AAAI Conference on Web and Social Media. International Workshop on Social Sensing (2021)
19. Nair, S., Ng, K., Iamnitchi, A., Skvoretz, J.: Diffusion of social conventions across polarized communities: an empirical study. *Social Network Analysis and Mining* **11**, December 2021
20. Salloum, A., Chen, T.H.Y., Kivelä, M.: Separating polarization from noise: comparison and normalization of structural polarization measures. In: Proceedings of the ACM on Human-Computer Interaction **6**(CSCW1), April 2022
21. Yang, M., Wen, X., Lin, Y.R., Deng, L.: Quantifying content polarization on Twitter. In: 2017 IEEE 3rd International Conference on Collaboration and Internet Computing (CIC), pp. 299–308 (2017)
22. Ortiz de Zarate, J., Di Giovanni, M., Feuerstein, E., Brambilla, M.: Measuring Controversy in Social Networks Through NLP, pp. 194–209. Springer International Publishing (09 2020)