Measuring and Classifying Network Polarization: Exploration of Structural Metrics

Sara Montese[1] and Marius Behret[2]

[1] Universitat Politecnica de Catalunya, Barcelona, Spain
s308482@studenti.polito.it
[2] Universitat Politecnica de Catalunya, Barcelona, Spain
mb98giwi@studserv.uni-leipzig.de

Keywords: Network Science · Polarization · Normalization.

1 Introduction

Political polarization, now escalating globally [12], has become a focal point of interest not only in political discourse but also in academic research. This phenomenon, characterized by the division of individuals into isolated groups with limited interaction and understanding, poses significant challenges to the cohesion and functionality of communities, societies, and democratic frameworks [16]. This growing divisiveness underscores the urgency of studying polarization. Defined by the Oxford Dictionary as the "division into two sharply contrasting groups or sets of beliefs or opinions" [7], polarization presents complex dynamics that are critical to understand for maintaining societal harmony and democratic integrity. The increasing attention towards this subject in various sectors highlights its relevance and the need for comprehensive strategies to address its multifaceted impacts.

1.1 Related work

In research, two main fields of polarization measurement can be identified, namely content-based and network-based procedures [24]. Content-based network research focuses more on the identification of opposing stances among groups through textual exploration like NLP-techniques. Often the focus there is on social media communication and communities. The second area of research, the network-based or structural polarization measures, are centered around the structural elements of polarization, deduced from network models of social or political frameworks [24]. Structural metrics often interpret the division between communities identified through structural analysis as indicative of polarization among groups within the system. Due to their ability to encompass theoretically-based elements of political polarization, and their relative cost-effectiveness compared to content-based and survey methodologies, these structural polarization metrics are emerging as appealing tools for practical applications such as this research project.

The common procedure for polarization measurement starts with the construction of the graph, followed by the detection of two distinct communities and finalized by an application of polarization measures on the separated graph. In many research approaches [24, 11, 1], the community detection is done with METIS [14], which enables the possibility of detecting specifically two communities. Examples of other graph partition algorithms, which were used in the context of polarization measurement, are Regularized Spectral Clustering and Modularity Optimization [24] or the bisection-based algorithm [26]. More recently, the FluidC algorithm [19] was applied in a few studies in order to detect two different communities [27, 2].

Today, multiple structural division measures exist, although some of them do not necessarily measure polarization. Homophily, for example, which is the inclination to form social connections with others who are similar to oneself, can result in segregation and polarization [17], but this tendency alone may not always be enough to cause such outcomes [6]. Even though they are very similar, the concepts segregation, homophily and polarization should be handled in distinctive ways [20]. Some of the scores which were directly used as polarization measures, include the Random Walk Controversy (RWC) [11], Adaptive Walk Controversy (ARWC), Betweenness Centrality Controversy (BCC) [11], Boundary Polarization (BP) [13], Dipole Polarization (DP) [18], E-I Index (EI) [15], Adaptive E-I Index (AEI) [5], and Modularity (MOD) [28]. The Segregation Matrix Index (SMI) [10], and Spatial Segregation Index (SSI) [8] were also categorized as scores that are able to measure polarization [20]. In [24], Salloum et al. underline the complexity in interpreting polarization scores on their own. They emphasize the challenge to determine the level of polarization in a network or to compare the polarization between networks using current metrics. This difficulty stems from the inability to assess whether a network is more polarized than a random one based solely on its score, without considering the network's specific features. To address this, the researchers propose normalizing polarization scores against a distribution from corresponding configuration models. This procedure significantly improved the accuracy in classifying 203 labeled networks [22], offering a more effective method for analyzing and comparing network polarization.

1.2 Motivation

In this research paper, we are motivated by the work of Salloum et al. [24], aiming to discern whether a network is polarized. To achieve this, we employ a range of polarization metrics including the EI, AEI, RWC, and Modularity as outlined in [24], along with the SMI as proposed in [20], and additionally considering their normalized versions. Our approach involves a comparative analysis to evaluate the impact of normalization and fundamental characteristics like the size and density of networks, also by measuring the polarization of randomized networks. For practical implementation, we utilize retweet networks for training as per [24], and test our findings on unclassified retweet networks detailed in [22]. This method allows us to explore the dynamics of polarization in retweet networks, providing insights into their structural aspects.

2 Results

In Figure 1, we conduct a comparison between the observed network data and randomized networks. Regarding RWC and modularity, the scores generated by random models encompass the observed score for most networks. The black bar, representing the observed score of a single network, is consistently covered by various colors, each corresponding to scores produced by different random models. This suggests that the number of links and the size of the networks (d = 0) already provide substantial predictive power for much of the observed score.

For negative EI, AEI, and BP, an increase in retained features corresponds to higher scores, bringing them closer to the original values. Retaining more features results in scores that closely resemble those of the original networks, indicating the significance of feature preservation.

The Segregation Matrix Index (SMI) stands out as an exception. In the ER model, the SMI is smaller than in the configuration model. This suggests that the ER model leads to less segregation or more integration between entities compared to the original network. The ER model, lacking preservation of the original network's degree distribution, is more likely to form edges between different groups than the configuration model, which does preserve the degree distribution. Consequently, the ER model exhibits greater homogeneity and less clustering than the configuration model.

2.1 Standardized and Denoised Polarization

Figure 2a shows the ROC curves and AUC values of the Logistic Regression Classifier, before score normalization. RWC and AEI exhibit the highest AUC values at 0.846, indicating exceptional effectiveness in class separation. With values close to 1, RWC and AEI are robust indicators for discerning polarization within the networks. Also EI, with an AUC of 0.845, remains a reliable score for classification. With a moderate AUC value of 0.794, SMI provides demonstrates as well its capability to distinguish between classes effectively. The AUC value of 0.731 indicates that BP is a less effective score for classification. Moreover, the high false positive rate of modularity suggests caution in relying on this measure as a standalone indicator for polarization as the risk of misclassification is non-negligible.

Figure 2b shows the ROC curves and AUC values of the Logistic Regression Classifier, after score normalization. As shows in the plot, the visual smoothness of the curve is compromised. A possible explanation is that there are too few threshold points. Despite the limited resolution, the AUC can still be calculated. As before score normalization, also in this case RWC, AEI and EI exhibit high AUC values, as opposed to modularity.

2.2 Experiments

Table 1 provides accuracy scores for Logistic Regression (LR) and Decision Tree (DT) classifiers, with comparisons between the original and normalized datasets.

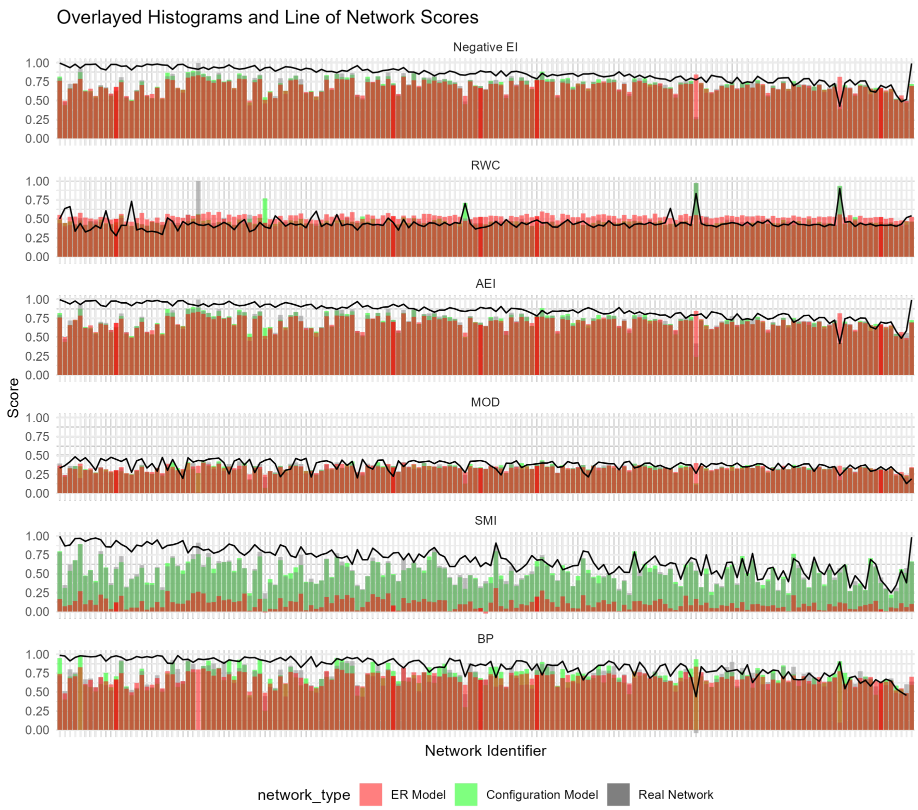
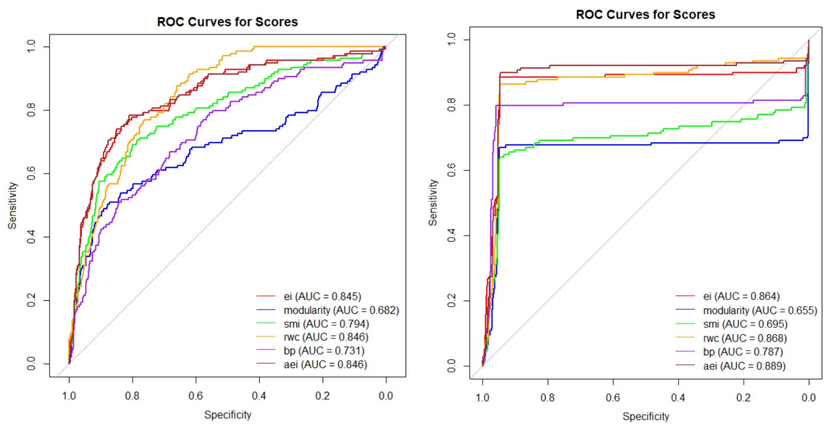


Fig. 1: Polarization scores for the observed networks and their randomized versions. Each bar corresponds to a score, and scores for a network and its randomized versions are on top of each other. An interpretation for the figure is that, the amount of color that is shown tells how much of the total bar height (the score value) is explained by the corresponding network feature. Note that in some cases, the randomized networks produce higher scores than the original network and in this case the black bar is fully covered by the colored bar(s). In this case we draw a black horizontal line on top of the colored bars indicating the height of the black bar.



(a) Before normalization (b) After normalization

Fig. 2: ROC curves and AUC values for the task of predicting manually curated labeling of polarized and non-polarized networks. The results shown (left) for the score values before the normalization and (right) after the normalization with standardized and denoised scores (See Section 2.1). The plots refer to the Logistic Regression Classifier.

In the original dataset, the performance between the Decision Tree and Logistic Regression classifiers are notable. The Decision Tree exhibits superior performance in the training set, surpassing Logistic Regression, whereas Logistic Regression demonstrates a slight advantage in the validation set. The normalization process introduces distinctive effects on the two classifiers. Logistic Regression shows a decrease in accuracy, suggesting potential challenges in adapting to the normalized data. Conversely, the Decision Tree undergoes a significant improvement after normalization, indicating its adaptability and robustness in handling normalized features. Decision Tree performs well in both the normalized training and validation sets, implying effective learning and generalization capabilities, and further emphasizing its suitability for the classification task on the normalized data.

	Original		Normalized	
Accuracy	LR	DT	LR	DT
Training	0.86	0.90	0.82	0.97
Validation	0.82	0.79	0.81	0.96

Table 1: Accuracy scores of classifiers before and after normalization.

Table 2 presents scores for the polarization metrics on external datasets, specifically focusing on networks related to #israel, #obama, and #occupywallstreet.

Network	|V|	|E|	-EI	RWC	AEI	MOD	SMI	BP
#israel	3698	4174	0.946	0.622	0.944	0.308	0.836	0.164
#israel $d = 0$	3140	4095	0.8	0.09	0.882	0.359	0.932	0.559
#israel $d = 1$	3210	3859	0.888	0.041	0.798	0.296	0.766	0.164
#obama	3212	3426	0.999	0.976	0.999	0.134	0.993	0.11
#obama $d = 0$	2646	3326	0.803	0.094	0.804	0.402	0.8	0.577
#obama $d = 1$	2512	2942	0.877	0.695	0.876	0.397	0.812	0.171
#occupywallstreet	3609	3933	0.191	0.364	0.221	0.07	-0.176	0.766
#occupywallstreet $d = 0$	3043	3840	0.821	0.066	0.818	0.352	0.778	0.556
#occupywallstreet $d = 1$	2961	3063	0.999	0.799	0.999	0.0	0	0

Table 2: Scores results on external datasets[21]. |V| and |E| refer to the network vertices and edges after the pre-processing step.

Table 3 presents classification results for external networks using Logistic Regression and Decision Tree classifiers, both before and after normalization. True labels were assigned based on manual verification of tweets. In the original dataset, both LR and DT perform well in predicting polarization or non-polarization. However, there are instances where LR predicts non-polarization while DT predicts polarization, suggesting divergent predictions after normalization. It is worth noting that the classifiers have poor performances on the configuration model.

	Original		Normalized		
Network	LR	DT	LR	DT	True label
#israel	1	1	0	1	1
#israel d=0	0	0	0	0	0
#israel d=1	1	1	0	1	0
#obama	1	1	0	1	1
#obama d=0	0	0	0	0	0
#obama d=1	1	1	0	1	0
#occupywallstreet	0	0	0	0	0
#occupywallstreet d=0	0	0	0	0	0
#occupywallstreet d=1	1	1	1	1	0

Table 3: Classification results of external networks. True labels were assigned based on manual verification of tweets.

3 Discussion

Interpreting polarization scores independently is challenging. Without contextualizing these scores with the network's specific attributes, discerning whether a network exhibits greater polarization than a random one becomes intricate. To face this challenge, we follow the approach of normalizing polarization scores by comparing them to a range of scores from corresponding configuration models. We confirm that this method notably enhances the effectiveness of classification of Decision Trees, that reach 97% of accuracy on the validation set. Logistic Regression classifier, on the other hand, shows slightly worse performances after normalization. In examining Logistic Regression results, AUC values reveal that RWC, AEI, and EI stand out as the most reliable scores for distinguishing polarization within the Twitter network. Meanwhile, Modularity, exhibiting the lowest AUC, indicates challenges in effectively classifying polarization using this metric. Regarding the limitations of our methodology, our research predominantly relies on datasets sourced from Twitter. Acknowledging this as a constraint, it is important to note that Twitter serves as a primary platform for online public discourse and is among the few sources with accessible data. This makes it a logical selection for our study. We manually chose and classified the retweet networks from [21] which might introduce bias. Furthermore, our approach depends heavily on graph partitioning, a critical stage for the effectiveness of our controversy detection methods. Recognizing that graph partitioning is a complex yet thoroughly researched problem, we utilize established techniques for this task. Nonetheless, developing a method that could eliminate the need for graph partitioning would be highly advantageous. The definition of what constitutes a controversial or polarized topic can be subjective, varying based on context, subject, and area of study. Our assessment relies on an intuitive judgment that a topic is controversial or polarized. While this may not always hold true, the alternative – manually labeling or surveying thousands of users – is less feasible. Thus, we believe this approach is a practical compromise for developing scalable methodologies.

4 Methodology

In this paper, our primary objective is to evaluate the effectiveness of commonly-used structural polarization measures in order to classify a network as polarized/non-polarized. To achieve this, we investigate the influence of six structural polarization scores through the utilization of null models based on random networks. These null models are designed to maintain specific structural properties while expressing maximum randomness. We employ two distinct null models:

- an Erdős-Rényi model [9], obtained by fixing the number of links,
- a configuration model [3] obtained by fixing the degree sequence.

The Erdős-Rényi network fixing the number of links is designated as $d = 0$, the configuration model fixing the degree sequence is $d = 1$. Our approach to measuring controversy is visualized in 3. We employ a pipeline with four stages, namely

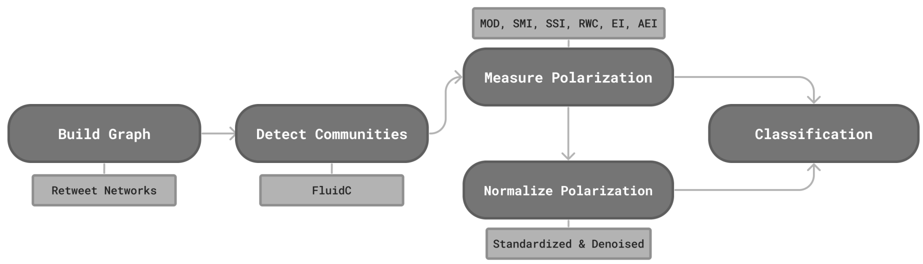


Fig. 3: Our polarization measurement pipeline

graph building, graph partitioning, controversy assessment and classification. A brief overview of each stage is presented below.

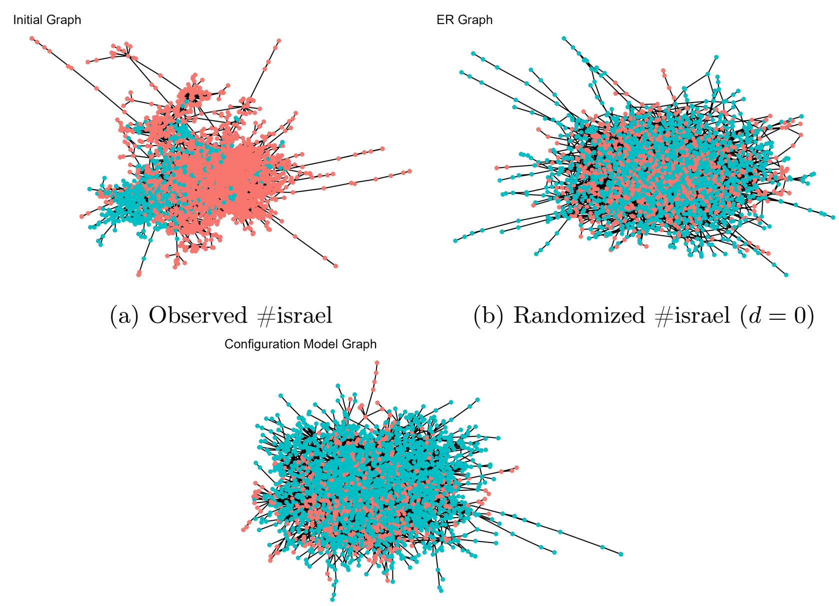
4.1 Graph Building

Our study incorporates real-world data obtained from Twitter. The dataset [23] includes 183 Twitter retweet networks collected during the 2019 Finnish Parliamentary Elections and labelled by hand as polarized/non-polarized. The first 150 networks are built around single hashtags, such as #translaw. The remaining 33 networks are constructed using a combination of hashtags focused on specific topics like climate change. Each network has a node set containing all users who posted an original tweet containing at least one hashtag related to the given topic and all users who retweeted at least one of these tweets. The nodes within the networks symbolize anonymized Twitter accounts, and the directed edges signify endorsements through retweets. Graphs with a node count exceeding 5000 were excluded, resulting in a total of 173 networks. In the network, self-loop edges arise when a user retweets their own tweet multiple times. Isolated nodes, representing tweets without any social connections, may introduce bias. Parallel edges can indicate various retweets of the same tweet or different tweets being retweeted by the same user. Consequently, we performed preprocessing steps, including obtaining the largest network component, eliminating self-loops, and removing multi-edges. 'Quote retweets' (retweets with added comments) were excluded from our analysis.

4.2 Community Detection

Our goal is to partition the network into two similarly-sized groups with minimal ties (retweets) between them. In the context of endorsement networks, this aims to identify two groups with the lowest level of agreement. We initially applied an ensemble of community detection algorithms (Louvain [4], Leiden[25], and Regularized Spectral Clustering [29]) and selected the one maximizing modularity. With this approach, restrictions on the balance between group size are required because there is the risk that the algorithm will lead to a trivial partition with a single degree-one node in one group and the rest of the nodes in the other group. While this satisfies the objective of minimizing inter-group ties to exactly one tie, it clearly does not capture our intentions. On the other hand, forcing the partitions to be perfectly balanced when real groups are not the same size will result in the partition to divide the larger group, which incorrectly inflates inter-group agreement relative to within-group agreement. To address this, we explored an alternative approach using a consensus matrix. Valid similarities, above a specified threshold, were considered, and a connected component search method was applied to assign each node to its cluster. This ensured that nodes frequently clustered together belonged to the same cluster, emphasizing inter-group differences. While effective, this approach is sensitive to noise and outliers, and the remaining connected component may differ from the initial clustering. For our objective of finding two groups with minimal agreement, the Fluid Communities[19] algorithm offers a straightforward and intuitive interpretation. Fluid, based on the propagation methodology, is efficient, scalable, and capable of identifying a variable number of communities in a network.

Figure 4 shows an example of application of the above steps.



(a) Observed #israel

(b) Randomized #israel ($d = 0$)

(c) Randomized #israel ($d = 1$)

Fig. 4: The network illustrates a Twitter conversation structure related to #israel. The visual representation includes randomized versions generated using ER and configuration model for the network.

4.3 Polarization Assessment

This stage takes the graph generated in the first stage, which is partitioned in the second stage, as input. It computes the values of multiple controversy measures that characterize the level of controversy surrounding the topic within the network. Essentially, a controversy measure aims to quantify the degree of separation between the two partitions. We evaluate six distinct measures for each network, and additional details are provided below.

Polarization Score	Domain	Parameters
Random Walk Controversy (RWC)	[-1, 1]	no. of influencers (k) per group = 10 no. of simulations = 50 no. of walks = 150
Modularity (MOD)	[-0.5, 1]	-
Boundary Polarization (BP)	[-0.5, 0.5]	-
E-I Index (EI)	[-1, 1]	-
Adaptive E-I Index (AEI)	[-1, 1]	-
Segregation Matrix Index (SMI)	[-1, 1]	-

Table 4: Polarization scores used, their domains and parameters

Certain metrics, such as SMI, assess polarization within individual groups, whereas others evaluate the polarization of the entire network. Our objective is to gauge the polarization of both groups simultaneously and modify the metrics accordingly, as detailed in the subsequent sections.

Random Walk Controversy (RWC) The RWC measures the likelihood of encountering authoritative content from opposing viewpoints in a network, based on the idea that more polarized groups are less likely to interact with influential members of other groups [11]. The methodology involves first the identification of influential nodes, which are the ones with the highest degrees. Then, a random walk is simulated, which begins from either side with the same probability and stops when it reaches an influencer. After multiple random walks, the RWC score is calculated as

$$P_{RWC} = p_{AA}p_{BB} - p_{AB}p_{BA}$$

Here, p_{AB} denotes the conditional probability for the random walk, given that it began in B and it stopped in B. The other formulas are computed in an akin way. The domain of the RWC is from -1 to 1, where high numbers indicate a high polarization, networks with $p_{RWC} = 0$ are considered not polarized, and a negative $p_{RWC} = 0$ indicates a likely contact with content produced by an influencer of the other community. We follow the proposal of [24] and set the number of influencers $k = 10$ for every network. To have a balance between computational resources and accuracy, we set the number of walks to 150 and the number of simulations to 50.

Boundary Polarization (BP) The Boundary Polarization measure is based on the premise that polarization is suggested by a low concentration of high-degree (influential) nodes at the community boundaries [13]. It posits that the more distant an influential user is from the boundary, the higher the level of antagonism within the network. For this measure, two sets of nodes are defined: boundary nodes and internal nodes. A node belongs to the boundary set if it links to at least one node of the opposite community and to another node that is not connected to any node of the opposite side. The internal nodes are those not in the boundary set. The measure is mathematically defined as follows:

$$P_{BP} = \frac{1}{|C|} \sum_{s \in C} \left(\frac{d_I(s)}{d_C(s) + d_I(s)} - 0.5 \right)$$

In this context, d_C represents the count of edges connecting a specific node s to nodes in community C, and d_I denotes the number of edges linking the same node s to nodes in community I. The calculation of the score takes into account the total number of boundary nodes in C for normalization. The resultant values of P_{BP} can vary from -0.5 to 0.5, where a value of 0.5 signifies the highest level of polarization. A network with little to no polarization is expected to have values approaching zero. Conversely, negative values suggest a propensity for the boundary nodes of community A to form connections more with the opposite community B rather than with their own community.

E-I Index (EI) The E-I Index, also referred to as the Krackhardt E/I Ratio, is a straightforward measure that calculates the relative density of internal connections within a community in comparison to its external connections and is particularly effective when analyzing two communities [15]. It is defined based on the ratio of the number of internal connections to the number of external connections.

$$P_{EI} = \frac{|C|}{|C|}$$

Here, the cut set $(s, t) \in E | s \in A, t \in B$ is C and its complement $C = E/C$ is C.

Adaptive EI Index (AEI) The Adaptive E-I Index is a refined version of the E-I Index that incorporates variations in community sizes by considering the density of links within each community [5]. This index aligns with the standard E-I Index in scenarios where the two communities being compared have an equal number of nodes. The index is formulated by adjusting the standard E-I Index to account for the ratio of actual to potential links within each community, as well as the ratio of observed links between the communities relative to all potential links. This adjustment makes the Adaptive E-I Index a more versatile tool for analyzing networks with communities of differing sizes. It can be defined as

$$P_{AEI} = \frac{\sigma_{AA} + \sigma_{BB} - (\sigma_{AB} + \sigma_{BA})}{\sigma_{AA} + \sigma_{BB} + (\sigma_{AB} + \sigma_{BA})}$$

.

While σ_{AA} (correspondingly σ_{BB} denotes the proportion of real and possible connections within the group A, σ_{AB} is the actual number of connections between the groups A and B. These two are divided by the amount of all possible connections.

Modularity (MOD) The MOD, a widely recognized metric for assessing discrepancies in social networks, evaluates how distinct the communities in a network are compared to those in a set of random graphs generated by a configuration model. It does this by using the standard modularity formula, which is commonly applied to assess the quality of community divisions within a network.

$$P_Q = \frac{1}{2|E|} \sum_{ij} \left(W_{ij} - \frac{k_i k_j}{2|E|} \right) \delta(c_i, c_j)$$

Here, $|E|$ represents the total number of edges in the network, W_{ij} is an element of the adjacency matrix (indicating the presence or absence of an edge between nodes i and j), k_i and k_j are the degrees of nodes i and j respectively, and $\delta(c_i, c_j)$ is the Kronecker delta function which equals 1 when nodes i and j belong to the same community and 0 otherwise.

[28]

Segregation Matrix Index (SMI) The SMI is a measure used to analyze the cohesiveness within social groups [10] and was not originally developed as a polarization metric. It's based on the ratio of inward to outward interactions within a group, reflecting how members of a group interact more with each other than with outsiders. SMI considers both the number and intensity of these interactions. A higher SMI value indicates a more cohesive group, characterized by a greater tendency for internal interaction and segregation from others. In our application, the final index is computed as the weighted average of the SMI for each group, with weights assigned based on group size.

4.4 Normalization

Here, we follow the proposal from [24], in order to remove the effect of specific network properties, such as degree distribution and network size, and fluctuations regarding the scores of the configuration models. The reduction of impact from the network properties is done by averaging the polarization score across several shuffled instances of the network using the configuration model (with $d = 1$). This average is then deducted from the actual observed polarization score. Specifically, for a given network G and a polarization score Φ, the normalized score is established as:

$$\hat{\Phi}(G) = \Phi(G) - \langle \Phi(G_{CM}) \rangle$$

$\Phi(G)$ denotes the polarization score of the real network, $\Phi(G_{CM})$ is the polarization score from the configuration model graphs and $\hat{\Phi}(G)$ is the *denoised polarization score*. To take into account the fluctuations of the configuration models, the denoised polarization score is divided by the standard deviation of the score value distribution:

$$\hat{\Phi}_z(G) = \frac{\Phi(G) - \langle \Phi(G_{CM}) \rangle}{\sqrt{\langle \Phi(G_{CM})^2 \rangle - \langle \Phi(G_{CM}) \rangle^2}}$$

where $\hat{\Phi}_z(G)$ is the so called *standardized and denoised polarization score.*

4.5 Classification

After computing scores for each network, we use two distinct classification approaches to determine the polarization status of a set of graphs based on these scores. The dataset, consisting of scores from 173 networks, is split into training (80%) and validation (20%) sets. In both classification methodologies, training is executed on the training set, and testing is carried out on the validation set. The initial classification method involves Logistic Regression. The second strategy involves training a Decision Tree classifier.

References

[1] Md Tanvir Al Amin, Charu Aggarwal, Shuochao Yao, Tarek Abdelzaher, and Lance Kaplan. "Unveiling polarization in social networks: A matrix factorization approach". In: *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*. IEEE. 2017, pp. 1–9.

[2] Faisal Alatawi, Paras Sheth, and Huan Liu. "Quantifying the Echo Chamber Effect: An Embedding Distance-based Approach". In: *arXiv preprint arXiv:2307.04668* (2023).

[3] K. Bailey, Daniel B. Fosdick, Daniel B. Larremore, Joel Nishimura, and Johan Ugander. "Configuring random graph models with fixed degree sequences". In: *SIAM Review* 60.2 (2018), pp. 315–355.

[4] Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 10 (2008), P10008.

[5] Ted Hsuan Yun Chen, Ali Salloum, Antti Gronow, Tuomas Ylä-Anttila, and Mikko Kivelä. "Polarization of climate politics results from partisan sorting: Evidence from Finnish Twittersphere". In: *Global Environmental Change* 71 (2021), p. 102348. ISSN: 0959-3780. DOI: https://doi.org/10.1016/j.gloenvcha.2021.102348. URL: https://www.sciencedirect.com/science/article/pii/S0959378021001278.

[6] Pranav Dandekar, Ashish Goel, and David T Lee. “Biased assimilation, homophily, and the dynamics of polarization”. In: *Proceedings of the National Academy of Sciences* 110.15 (2013), pp. 5791–5796.

[7] *Definition of Polarization.* https://www.oed.com/view/Entry/29442549. Accessed: 2024-01-17. Oxford Dictionaries, 2024.

[8] Federico Echenique and Roland G Fryer Jr. “A measure of segregation based on social interactions”. In: *The Quarterly Journal of Economics* 122.2 (2007), pp. 441–485.

[9] P. Erdős and A. Rényi. “On random graphs I”. In: *Publicationes Mathematicae* 6 (1959), pp. 290–297.

[10] Meir Fershtman. “Cohesive group detection in a social network by the segregation matrix index”. In: *Social Networks* 19.3 (1997), pp. 193–207.

[11] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. “Quantifying controversy on social media”. In: *ACM Transactions on Social Computing* 1.1 (2018), pp. 1–27.

[12] Noam Gidron, James Adams, and Will Horne. *American Affective Polarization in Comparative Perspective.* Elements in American Politics. Cambridge University Press, 2020. DOI: 10.1017/9781108914123.

[13] Pedro Guerra, Wagner Meira Jr, Claire Cardie, and Robert Kleinberg. “A measure of polarization on social media networks based on community boundaries”. In: *Proceedings of the international AAAI conference on web and social media.* Vol. 7. 1. 2013, pp. 215–224.

[14] George Karypis and Vipin Kumar. “A fast and high quality multilevel scheme for partitioning irregular graphs”. In: *SIAM Journal on scientific Computing* 20.1 (1998), pp. 359–392.

[15] David Krackhardt and Robert N. Stern. “Informal Networks and Organizational Crises: An Experimental Simulation”. In: *Social Psychology Quarterly* 51.2 (1988), pp. 123–140. ISSN: 01902725. URL: http://www.jstor.org/stable/2786835 (visited on 01/18/2024).

[16] Antonis Matakos, Evimaria Terzi, and Panayiotis Tsaparas. “Measuring and moderating opinion polarization in social networks”. In: *Data Mining and Knowledge Discovery* 31 (2017), pp. 1480–1505.

[17] Miller McPherson, Lynn Smith-Lovin, and James M Cook. “Birds of a Feather: Homophily in Social Networks”. In: *Annual Review of Sociology* 27.1 (2001), pp. 415–444. DOI: 10.1146/annurev.soc.27.1.415. eprint: https://doi.org/10.1146/annurev.soc.27.1.415. URL: https://doi.org/10.1146/annurev%20.soc.27.1.415.

[18] Alfredo Jose Morales, Javier Borondo, Juan Carlos Losada, and Rosa M Benito. “Measuring political polarization: Twitter shows the two sides of Venezuela”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.3 (2015).

[19] Ferran Parés, Dario Garcia Gasulla, Armand Vilalta, Jonatan Moreno, Eduard Ayguadé, Jesús Labarta, Ulises Cortés, and Toyotaro Suzumura. “Fluid communities: A competitive, scalable and diverse community detection algorithm”. In: *Complex Networks & Their Applications VI: Proceed-*

ings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications). Springer. 2018, pp. 229–240.

[20] Samantha C Phillips, Joshua Uyheng, and Kathleen M Carley. "A high-dimensional approach to measuring online polarization". In: *Journal of Computational Social Science* (2023), pp. 1–32.

[21] Ryan A. Rossi and Nesreen K. Ahmed. "The Network Data Repository with Interactive Graph Analytics and Visualization". In: *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*. 2015. URL: http://networkrepository.com.

[22] Ali Salloum, Ted Hsuan Yun Chen, and Mikko Kivelä. "Polarized and nonpolarized Twitter networks from the 2019 Finnish Parliamentary Elections". Version Version 2. In: (2023). DOI: 10.5281/zenodo.8434372. URL: https://doi.org/10.5281/zenodo.8434372.

[23] Ali Salloum, Ted Hsuan Yun Chen, and Mikko Kivelä. "Polarized and nonpolarized Twitter networks from the 2019 Finnish Parliamentary Elections". In: Zenodo, Oct. 2023. DOI: 10.5281/zenodo.8434372. URL: https://doi.org/10.5281/zenodo.8434372.

[24] Ali Salloum, Ted Hsuan Yun Chen, and Mikko Kivelä. "Separating polarization from noise: comparison and normalization of structural polarization measures". In: *Proceedings of the ACM on human-computer interaction* 6.CSCW1 (2022), pp. 1–33.

[25] Vincent A Traag, Ludo Waltman, and Nees Jan van Eck. "From Louvain to Leiden: Guaranteeing well-connected communities". In: *Scientific Reports* 9.1 (2019), p. 5233.

[26] Giacomo Villa, Gabriella Pasi, and Marco Viviani. "Echo chamber detection and analysis: a topology-and content-based approach in the COVID-19 scenario". In: *Social Network Analysis and Mining* 11.1 (2021), p. 78.

[27] Giacomo Villa, Gabriella Pasi, and Marco Viviani. "Echo chamber detection and analysis: a topology-and content-based approach in the COVID-19 scenario". In: *Social Network Analysis and Mining* 11.1 (2021), p. 78.

[28] Andrew Scott Waugh, Liuyi Pei, James H. Fowler, Peter J. Mucha, and Mason A. Porter. *Party Polarization in Congress: A Network Science Approach*. 2011. arXiv: 0907.3509 `[physics.soc-ph]`.

[29] Yilin Zhang and Karl Rohe. "Understanding regularized spectral clustering via graph conductance". In: *Advances in Neural Information Processing Systems* 31 (2018).