

RESEARCH ARTICLE

# Negative ties highlight hidden extremes in social media polarization

E. Candellone<sup>1,2,\*</sup> , S. A. Babul<sup>3,\*</sup> , Ö. Togay<sup>1</sup> , A. Bovet<sup>4,5</sup>  and J. Garcia-Bernardo<sup>1,2,6</sup> 

<sup>1</sup>Department of Methodology and Statistics, Utrecht University, Utrecht, Netherlands, <sup>2</sup>Centre for Complex Systems Studies, Utrecht University, Utrecht, Netherlands, <sup>3</sup>Mathematical Institute, University of Oxford, Oxford, UK, <sup>4</sup>Department of Mathematical Modeling and Machine Learning, University of Zurich, Zurich, Switzerland, <sup>5</sup>Digital Society Initiative, University of Zurich, Zurich, Switzerland, and <sup>6</sup>ODISSEI SoDa Team, Utrecht University, Utrecht, Netherlands

\*E. Candellone and S. A. Babul are joint first authors.

**Corresponding author:** Elena Candellone; Email: [candellone.elena@gmail.com](mailto:candellone.elena@gmail.com)

## Abstract

Human interactions in the online world comprise a combination of positive and negative exchanges. These diverse interactions can be captured using signed network representations, where edges take positive or negative weights to indicate the sentiment of the interaction between individuals. Signed networks offer valuable insights into online political polarization by capturing antagonistic interactions and ideological divides on social media platforms. This study analyzes polarization on Menéame, a Spanish social media platform that facilitates engagement with news stories through comments and voting. Using a dual-method approach—Signed Hamiltonian Eigenvector Embedding for Proximity for signed networks and Correspondence Analysis for unsigned networks—we investigate how including negative ties enhances the understanding of structural polarization levels across different conversation topics on the platform. While the unsigned Menéame networks effectively delineates ideological communities, only by incorporating negative ties can we identify ideologically extreme users who engage in antagonistic behaviors: without them, the most extreme users remain indistinguishable from their less confrontational ideological peers.

**Keywords:** polarization; signed networks; social media

## 1. Introduction

Online social networks have changed how people interact with news content, stay informed about current events, and form opinions on related topics (Marchi (2012)). This new mechanism for communication and information spread plays a key role in facilitating an increased polarization around controversial issues and amplifying political divisions and conflicts (Adamic and Glance (2005); Barberá (2020); Conover et al., (2011); Falkenberg et al., (2022); Flaminio et al., (2023); Garimella et al., (2017); Hohmann et al., (2023); Törnberg (2022)). As a consequence, political polarization is increasing not only within small, active partisan groups (ZP Neal (2020)) but also spreading more widely among the general population (A Abramowitz and Saunders (2005); AI Abramowitz and Saunders (2008); ZP Neal (2020)).

In political science literature, a distinction is made between ideological, affective, and structural (or interactional) polarization (Adamic and Glance (2005); Barberá et al., (2015); Bramson et al., (2017); DiMaggio et al., (1996); Esau et al., (2024); Hohmann et al., (2023); Lelkes (2016); ZP Neal (2020); Salloum et al., (2022); Yarchi et al., (2020)). Ideological polarization refers to the widening gap between the political beliefs of different groups; affective polarization captures the emotional hostility and negative attitudes between political factions; and structural polarization examines the division of social interactions into homogeneous groups with minimal cross-group

engagement. Given the recent increase in data available on online interactions between social media users, structural polarization (Falkenberg et al., (2023); Salloum et al., (2022); Yarchi et al., (2020)) has become a rich area of study over the past few decades. Prior research has shown that social media platforms may amplify polarization by promoting the alignment of differences, facilitating the formation of echo chambers (clusters of like-minded users) and the rapid diffusion of biased information (Barberá (2020); Cinelli et al., (2021); Del Vicario et al., (2016); Ferraz de Arruda et al., (2022)). Ideological polarization can also be recovered from the network of interactions between social media users, under the assumption that users are more likely to interact with other users (or content) with whom they share a similar ideological stance (Barberá et al., (2015); Ribeiro et al., (2017)). The ideal points model, used to measure both voting behaviors (Clinton et al., (2004); Enelow and Hinich (1984); Poole (2005); Poole and Rosenthal (1985); Yu and Rodríguez (2021)) and ideology (Barberá (2015); Moody and Mucha (2013); Waugh et al., (2009)), is based on the same assumption: the probability of voting (positively or negatively) on some content depends on the latent ideological difference between the individual and the content. The model has been applied across various contexts, from a legislator voting on a piece of legislation to a user voting on a social media post.

Despite extensive research, studies of online polarization remain constrained by limited data. Social media platforms primarily provide information on positive interactions (e.g., likes or retweets on Twitter/X) or neutral interactions (e.g., mentions on Twitter/X), while negative interactions (e.g., downvotes on Reddit) are either unrecorded or only accessible to researchers at an aggregated level. As a result, most prior studies analyze structural polarization by constructing unsigned networks from online interactions, failing to distinguish between interactions that are positive or negative in nature. Drawing from *Emotional Information* theory, which claims that negative sentiment may be a strong indicator of negative links between individuals, we interpret observed negative links as signs of discord or tension (Beigi et al., (2020)). This distinction is crucial, as online interactions can be incredibly diverse, representing sentiments ranging from support to hostility. Negative interactions significantly impact offline social networks and individual outcomes (Offer (2021)). Signed network representations, which assign positive or negative weights to edges, offer a powerful tool for capturing this complexity.

Previous studies of structural polarization in online media using signed networks have primarily focused on three platforms: Epinions (Richardson et al., (2003)) and Slashdot (Leskovec et al., (2009)), where users explicitly label each other as friends or foes, and Wikipedia, where users cast votes in administrator elections (Leskovec et al., (2010)). More recently, signed network representations have been constructed from Reddit and Twitter data, by inferring the interaction sign (positive or negative) using sentiment analysis applied to comments and posts (Keuchenius et al., (2021); Pougué-Biyong (2023); Pougué-Biyong et al., (2021)). Signed networks allow for a refined analysis of polarization by identifying communities with internal coherence and cross-group antagonism (Cartwright and Harary (1956); Davis (1967); Harary (1953); Heider (1946)). Several methods exist to partition the signed graph into these factions, quantify graph-level polarization, or extract ideological information from the network structure (Aref and Neal (2021); Aref et al., (2019); Babul and Lambiotte (2024); Doreian and Mrvar (2009); Huang et al., (2021); Kirkley et al., (2019); Traag and Bruggeman (2009)). These studies suggest that negative ties provide a deeper understanding of structural polarization by revealing hidden antagonisms or patterns that may not be apparent in networks only composed of positive interactions (Aref and Neal (2020); Babul and Lambiotte (2024); Doreian and Mrvar (2009); Keuchenius et al., (2021); Traag and Bruggeman (2009)).

Our work seeks to understand the value of including negative ties through a case study on Menéame, a social media platform with naturally occurring signed signals. Menéame<sup>1</sup> is a popular Spanish social media platform that, with a similar structure to Reddit, primarily functions as a news aggregator. On Menéame, users can post links to news stories; the post appears in Menéame's news feed with the hyperlink, information about the user who posted the story, and a brief

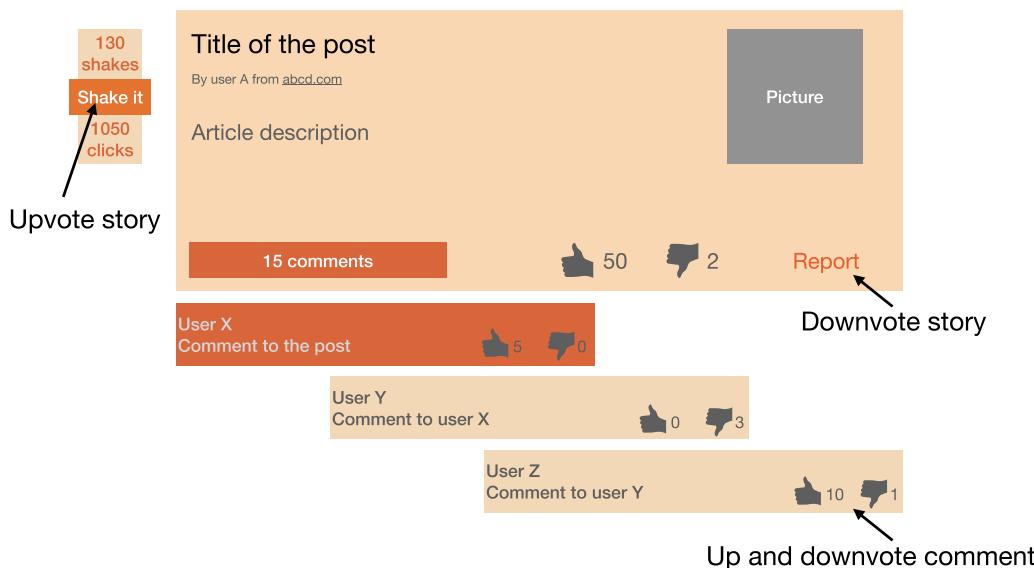
description of the story. These stories can then be voted (upvoted or downvoted) or commented on by other users in the ecosystem (Silva (2008)). The articles posted cover a wide range of topics, from sports to local and international politics. Given how users interact on the Menéame platform, we can naturally extract a signed network representation of the user base interactions, where the signed signals can be obtained directly from the up- and downvotes that users can leave on the articles and comments posted by other users. While Menéame has been studied before from various perspectives (Aragón et al., (2017); Gómez et al., (2013); Kaltenbrunner et al., (2011)), to the best of our knowledge, the social network of Menéame has never been studied in a way that exploits the natural signed representation of the social network ecosystem and studies polarization in such a context. Unlike other small community datasets, Menéame's data directly captures the sign of interaction dynamics at the comment level (rather than at the user level or inferred sentiment from text). In addition, the platform is designed so that the main newsfeed appears the same for all users, and there is no personalized recommendation algorithm; thus, the presence of structural polarization is due only to user preferences.

Our paper contributes to the literature in two ways; the first is that we have collected and made available a dataset from Menéame.<sup>2</sup> Our second contribution to the study of online polarization lies in our assessment of the value of including negative interactions in measuring polarization at the individual level. To do this, we study the patterns of structural polarization by constructing two types of networks from the Menéame data: a signed network representation that includes both positive and negative interactions, and an unsigned network that includes only positive interactions. Prior research on unsigned social media networks has identified structural position at the user level, revealing limited (but existing) communication between opposing groups (Barberá (2020); Barberá et al., (2015)). Moreover, studies using signed networks primarily seek to identify communities with a high in-group agreement and out-group antagonism (Doreian and Mrvar (2009); Esmailian and Jalili (2015); Keuchenius et al., (2021); Talaga et al., (2023); Traag and Bruggeman (2009)). More recently, methods have been developed to give each user a score that describes how polarized or extreme they are within the context of the network (Babul and Lambiotte (2024)). To quantify the value of negative ties to the study of polarization, we take a dual approach, leveraging methods designed for both signed and unsigned networks. We use the Signed Hamiltonian Eigenvector Embedding for Proximity (SHEEP) method (Babul and Lambiotte (2024)) for the signed network, and Correspondence Analysis (CA) (Greenacre (2017)) for the unsigned network. Both SHEEP and CA produce lower-dimensional embeddings of network data using spectral techniques and matrix decomposition, respectively, and have been used for latent ideology analysis (Babul and Lambiotte (2024); Barberá et al., (2015); Falkenberg et al., (2022); Falkenberg et al., (2023); Flamino et al., (2023); Peralta et al., (2024)). By evaluating the two methods side-by-side, we illuminate the insights gained by incorporating negative ties into network models of political polarization.

Our study addresses one key research question:

*What do negative interactions reveal about polarization that positive interactions cannot?*

We find that Menéame users can be grouped into two main ideological factions that exhibit structural polarization. The polarization between factions is much more pronounced in discussions around controversial topics (e.g., the Russia-Ukraine war) compared to discussions of general politics. The two methods we use—SHEEP and CA—largely agree in identifying ideological groups, and we verify the ideological groups we detect against an independent ideological measure. CA is perhaps even more effective at identifying disparate ideological groups within the unsigned network. However, negative ties reveal critical patterns, particularly at the extremes, that remain hidden when only positive interactions are analyzed. For instance, in the network of users voting on comments related to the Russia-Ukraine war, the extreme users identified by SHEEP are found to be those who upvote stories from the Russian-state-controlled news outlet *RT (Russia Today)*, while CA fails to distinguish these users from general left-wing users. More broadly, only SHEEP is capable of detecting extreme users who engage in high levels of antagonism. Our findings suggest



**Figure 1. Menéame platform.** Schematic representation of one of the stores in the platform. Users can upvote and downvote stories, and upvote and downvote comments within the story. Downvoting stories is possible only for registered users through the “Report” button, while upvoting stories is allowed to everyone. Only registered users can vote for comments. Comments with many positive votes appear on the platform highlighted in orange.

that the signed and unsigned networks offer complementary insights and that combining both methodologies enhances our understanding of polarization.

The paper is organized as follows. The section “Methods” describes the data collection process and the construction of the networks and introduces the two techniques we use to analyze polarization: SHEEP for signed networks and CA for unsigned networks. The section “Results and Discussion” presents our findings, comparing the insights gained from the signed and unsigned networks, focusing on topic-specific polarization and the role of the negative interactions. Finally, in the conclusion, we summarize our main contributions, discuss the implications of our findings, and outline potential directions for future research.

## 2. Methods

### 2.1. Dataset – Menéame social media platform

Menéame is a Spanish news aggregator platform created in 2005 that aims to enhance community participation in information and news diffusion. Users can post and interact with stories, i.e., posts containing a hyperlink to websites such as news outlet articles or social media posts, information about the user who posted the story, and a short description (Figure 1). The platform is divided into several sections. We focus on the main page (with the most popular stories) and the queue (where new stories are listed). Users can upvote, downvote, or comment on stories. They can also comment on or vote (up or down) on other comments. Stories that receive the most positive engagement and little negative engagement appear at the top of the feed. While the platform guidelines state that users should use negative votes to report spam<sup>3</sup> and to help remove content that goes against the platform guidelines, such uses are rare and result in account removal by moderators. In practice, negative votes are used primarily to express disapproval.

We collected all the stories, comments, and votes from the period spanning from December 1st, 2022, to August 8th, 2023, comprising 47,887 stories. In aggregate, our data contains 1,869,190

votes on those stories, 704,636 comments, and 3,113,863 votes on the comments. We found 11,156 unique users voting on stories, while 8,604 users were voting on other users' comments.

In this paper, we create two types of un-directed signed networks: a *user-to-user* network, using the votes on comments; and a *user-to-news outlet* network, using the votes on stories. We create these two networks for different topics, which we extract using the short summaries of the articles. The *user-to-news outlet* network is used to validate our approach since we can compare the ideological position of the news outlet generated by CA and SHEEP to external sources. The *user-to-user* network is our main dataset to study polarization in the platform. Although the format of the data and the platform could naturally produce a directed network representation, we make the modeling choice to remove the directions from the graph to focus on the impact of the signed and weighted edges. Empirically, we found that voting behavior between users tends to be symmetrical, as illustrated in Tables A1 and A2. Details on how we construct the two types of networks can be found in Subsection 2.3.

## 2.2. Topic modeling

Topic modeling consists of classifying texts into a finite number of categories (topics) (Blei and Lafferty (2009)). It can be supervised if there are existing labels or unsupervised, as in our case, where no topics are indicated in the data. Given the absence of a “ground truth” to compare the results of the algorithms, we employ two different algorithms: BERTopic (Grootendorst (2022)) and hierarchical Stochastic Block Model (hSBM) (Gerlach et al., (2018)). We then compare the topics obtained by the two methods, finding robust topics in a restricted subset of stories. In the following, we describe the two algorithms and the comparison technique we implemented.

### 2.2.1. BERTopic

The *BERTopic* algorithm, developed by Grootendorst (Grootendorst (2022)), presents a modular text-embedding-based approach for identifying and extracting topics from a given textual dataset based on the BERT (Bidirectional Encoder Representations from Transformers) language model (Devlin et al., (2019)). The algorithm involves the initial creation of text embeddings, followed by a dimensionality reduction and clustering process applied to these embeddings to form topics. Finally, each topic is associated with keywords using a variation of the Term Frequency-Inverse Document Frequency (TF-IDF). This last step facilitates the interpretation of the topics found with the algorithm. For a thorough description of the algorithm steps and the parameters chosen, please refer to Appendix A.2.

### 2.2.2. hSBM topic model

The TM-hSBM, proposed by Gerlach et al. (Gerlach et al., (2018)), is an application of community detection methods for topic inference on a text corpus. In contrast to BERTopic, which is based on embedding sentences into vectors, this approach involves creating a bipartite network with two groups of nodes: words and documents (e.g., comments or stories). Each word is connected to a document if it appears in that document. The method then entails applying a Bayesian hSBM inference (Peixoto (2014)), a method to detect communities in networks, to the bipartite network. In this case, it groups words into topics if they have a unique connectivity pattern—i.e., if they appear and are absent in similar documents (Gerlach et al., (2018)). Our detailed procedure is described in Appendix A.3.

### 2.2.3. Combining the results of BERTopic and hSBM

We apply both BERTopic and TM-hSBM to the story descriptions' text corpus. We found 180 topics using BERTopic containing 38,883 stories and 16,633 stories labeled as outliers. Table A3 in the Appendix shows the topic names and the number of stories per topic. We performed

**Table 1.** Number of stories, votes, upvotes, and downvotes per macro-topic in the dataset

Topic	Number of stories	Number of votes	Number of upvotes	Number of downvotes
Broad Politics	7,427	411,591	378,528	33,063
Russia-Ukraine War	2,394	77,828	64,240	13,588
Public services	1,621	85,154	81,637	3,517
Crime	1,537	73,465	69,074	4,391
Climate Change	1,458	41,847	39,661	2,186
Cryptocurrencies/tech	814	26,533	24,414	2,119
Inflation	763	26,724	25,316	1,408

outlier reduction techniques based on cosine similarities between embeddings. We also reduced the number of topics to 30.

TM-hSBM provides topics at different hierarchical levels. At the highest granularity level, it finds 114 topics, which are combined into 20, 4, and 1 topics, respectively. Table A4 in the Appendix shows the topic keywords for each level. Given the results obtained by the two methods independently, we compared them to obtain more robust topics. We computed the accuracy classification score and found that BERTopic with 30 reduced topics is the most similar to the second level of TM-hSBM.

We analyzed the most representative keywords for each topic identified using BERTopic based on their TF-IDF weights. For instance, Topic 1 included keywords such as “*ukraine, russia, war, russian (male), russian (female), putin, russian (plural), nato, military*,” which we manually labeled as the “*Russia-Ukraine war*” topic. We then assessed the overlap of each topic with the hSBM results.

For example, *hSBM Topic 1* (keywords: “*ukraine, russian (male), war, russia, ukrainian (male), trump, putin, militar, nato, invasion*”) and *hSBM Topic 16* (keywords: “*nuclear, chinese (male), ukraine, russian (male), militar, fire, russia, wagner, china, american*”) were grouped under the same macro-topic *BERTopic Topic 1* as they had an overlap of 52 and 46% respectively. In cases where multiple BERTopic topics aligned with the same hSBM topic, we merged the BERTopic topics. For instance, the *hSBM Topic 6* (“*rent, healthcare, hospital, union, labor, doctor, strike*”) was matched with the *BERTopic Topics 11* (“*health, strike, doctors, hospitals*”), 14 (“*workers, labor, work, employment*”), and 18 (“*education, students, teachers, schools*”) and manually labeled as “*Public Services*” (see Figure A2 to observe the overlap).

The combination of BERTopic and hSBM resulted in 7 macro-topics: *Broad Politics, Russia-Ukraine War, Public Services, Crime, Climate Change, Cryptocurrencies/tech, Inflation*. We then kept the stories classified in that topic by *both* algorithms and validated our approach by manually labeling a random sample of the comments. Out of 100 comments on which both algorithms agreed on the topic, we agreed with the classification in 93 instances and disagreed in 7. Out of 100 comments on which algorithms disagreed, we agreed with BERTopic in 34 cases, with hSBM in 26 cases, with neither algorithm in 39 cases, and with both algorithms in 1 case. Table 1 shows the number of stories for each macro-topic. While only a small fraction of the initial corpus is preserved, this refined categorization enables us to understand whether the importance of negative ties to assess polarization is topic-specific. In this paper, we focus on the two largest topics: *Broad Politics* and *Russia-Ukraine war*.

### 2.3. Creating networks from data

In the previous section, we described how we used the text corpus composed of short textual descriptions of each news story to divide these texts into macro-topics. Here, we seek to explore

the interactions between users and stories. Consequently, we construct two networks from our dataset: a network of *user-to-user* interactions and a network of interactions between *users and news outlets*.

In the first case, we consider a network  $G = (U, E)$ , where  $U$  is the set of users, and  $E$  is the set of edges. The network can be represented with an adjacency matrix  $A$  of dimension  $N \times N$  ( $N$  is the number of users), where each entry  $A_{i,j}$  is the sum of the signed votes of the user  $i$  to the comments of the user  $j$  and the signed votes of the user  $j$  to the comments of the user  $i$ . As a result, the network we construct is undirected, and the adjacency matrix  $A$  is symmetric. If there is no interaction between the two users, or if there are the same number of positive and negative interactions,  $A_{i,j} = 0$ .<sup>4</sup> The weight is bounded between  $[-n_{ij}, +n_{ij}]$ , where  $n_{ij}$  is the number of interactions between  $i$  and  $j$ . Specifically,  $A_{i,j} = -n_{ij}$  in case there are only negative votes between the two users, and  $A_{i,j} = +n_{ij}$  if there are only positive votes. We only consider “active” users who cast more than 10 votes during the period studied. This is a one-mode (unipartite) network, as all the nodes are of the same type (users).

In the second case, we consider a two-mode (bipartite) network  $B = (V, E')$ , where  $V$  is the set of nodes, and  $E'$  is the set of edges. In this case, nodes are of two types, users and news outlets, that form two disjoint sets, which we label  $U$  and  $O$  to represent users and news outlets. The edges connect nodes from one subset to the other only, i.e.,  $E' \subseteq U \times O$ . We can represent this network with an adjacency matrix that has the shape

$$A = \begin{pmatrix} 0 & I \\ I^T & 0 \end{pmatrix},$$

where  $I$  is the incidence matrix that has shape  $|U| \times |O|$ , and each entry  $I_{k,l}$  is the sum of the signed votes of user  $k$  on the stories from the news outlet  $l$ . Note that a user can vote on many different stories from a given news outlet, and this information is aggregated in our network. Similarly to the common practice in latent space models (Poole and Rosenthal (1985))—where the vote (positive/negative) is modeled as depending on the difference between latent ideological positioning—we removed the stories that only received positive votes. We found this step to be fundamental in quantifying the impact of controversial stories, and it reflects the fact that stories with only positive votes do not polarize the discussion.

#### 2.4. SHEEP embedding

SHEEP, developed by Babul and Lambiotte (Babul and Lambiotte (2024)), is a spectral embedding method capable of representing proximal information of nodes, using both positive and negative interactions. SHEEP is based on the minimization of the repelling Laplacian (Shi et al., (2019)), defined as

$$L_r = D^+ - A^+ - D^- - A^-, \quad (1)$$

where  $D^+$  (resp.,  $D^-$ ) and  $A^+$  (resp.,  $A^-$ ) are the degree and adjacency matrix of the positive (resp., negative) part of the network. Babul and Lambiotte (Babul and Lambiotte (2024)) proved the equivalence between the spectrum of the repelling Laplacian and the Hamiltonian in one dimension, where positive edges are considered as spring attractive forces and negative edges are anti-spring repulsive forces, as follows.

$$\boldsymbol{\pi}^T L_r \boldsymbol{\pi} = \sum_{i,j} A_{ij}^+ |\pi_i - \pi_j|^2 + \sum_{i,j} A_{ij}^- |\pi_i - \pi_j|^2. \quad (2)$$

In higher dimensions, the algorithm associates each node in the network with a position (i.e., an embedding  $\boldsymbol{\pi}_i \in \mathbb{R}^N$ ), generated using the first  $k$  eigenvectors of the repelling Laplacian, such

that nodes connected by positive edges are placed closer together, and nodes connected by negative edges are placed further apart. The algorithm also provides a method to identify the optimal dimension for the embedding, by minimizing a generalized version of the Hamiltonian in Eq. 2 (for more details see Babul and Lambiotte (2024)). This method, when applied to a signed network of bill co-sponsorship frequency in the US House of Representatives, is successful at recovering the political ideology of the House members on a continuous spectrum (Babul and Lambiotte (2024)).

We apply the SHEEP embedding method for each topic to both the unipartite signed network of *user-to-user* votes and the bipartite signed networks of *user-to-news outlet* votes. We use a normalized version of the adjacency matrix, where the entries are divided by the total degree of the network. Following this procedure, we obtain an embedding for each user and news outlet, which we use for further analysis. Since nodes with a high number of negative votes are pushed away in one of the dimensions from all other nodes, we project each embedding into one dimension using a principal component analysis (PCA) projection.

## 2.5. Correspondence analysis

CA, first theorized by Hirschfeld (Hirschfeld (1935)) and later applied by Benzécri et al., (Benzécri and Bellier (1973)), is a widely used method to obtain lower-dimensional representations of data, especially networks. As described in Greenacre's works (Greenacre (1984, 2017)), CA is a statistical technique that produces embeddings of categorical data in a lower-dimensional space. Similar to the PCA, it is based on the singular value decomposition (SVD) of the interaction matrix. For example, in our case, we have a matrix where the rows are the users, the columns are the news outlet domains, and the entries are the number of times each user interacted with the news outlet. In detail, given the interaction matrix  $\mathbf{I} \in \mathbb{N}^{a \times b}$ , where  $a$  is the number of users, in our case, and  $b$  is the number of outlets, we first compute the correspondence matrix  $\mathbf{P} = \frac{1}{n}\mathbf{I}$ , where  $n = \sum_i \sum_j I_{ij}$  is the sum of all the entries. Then, we compute the matrix of standardized residuals.

$$\mathbf{S} = \mathbf{D}_r^{-\frac{1}{2}} (\mathbf{P} - \mathbf{r}\mathbf{c}^T) \mathbf{D}_c^{-\frac{1}{2}}, \quad (3)$$

where  $\mathbf{r} = \mathbf{P} \mathbf{1}$  and  $\mathbf{c} = \mathbf{P}^T \mathbf{1}$ , i.e., the so-called row and column masses, and  $\mathbf{D}_x = \text{diag}(\mathbf{x})$ . Then we calculate the SVD of the matrix  $\mathbf{S} = \mathbf{U}\Lambda\mathbf{V}^T$  such that  $\mathbf{U}\mathbf{U}^T = \mathbf{V}\mathbf{V}^T = \mathbf{1}$ . The diagonal matrix  $\Lambda$  contains the singular values and is used to determine the embeddings of rows (users) and columns (either users in the *user-to-user* network or news outlets in the *user-to-news outlet* network) following

$$\begin{aligned} \mathbf{CA}_r &= \mathbf{D}_r^{-\frac{1}{2}} \mathbf{U} \Lambda \\ \mathbf{CA}_c &= \mathbf{D}_c^{-\frac{1}{2}} \mathbf{V} \Lambda \end{aligned} \quad (4)$$

We use the first dimension of the embeddings recovered in Eqs. 4 for further analysis. In this work, we employ the *prince* Python package (Halford (nd)), which performs CA among other statistical techniques.

## 2.6. Ideology of news outlets

To validate the results in the *user-to-news outlet* network of the two methods described above, we created two independent “ground truth” measures of the left-right ideological position of news outlets. First, we calculated ideological positions using Twitter (now X) data. We used the Tweepy Python package (Harmon et al., (2023)) and the Twitter API v2 with Academic Research access. We collected all tweets from the main Spanish political parties that are influential in terms of popular votes, excluding regionalist parties that primarily tweet in languages other than Spanish. The

analyzed parties were *PP*, *PSOE*, *CS*, *PODEMOS*, *IU*, *VOX*, *MasPais*, and *PACMA*. Additionally, we identified the 20 most mentioned accounts by each party's account and manually filtered out accounts not associated with politicians or institutions linked to the same party. The complete list of Twitter handles can be found in the Appendix A.5.

Next, we counted the number of tweets per political party mentioning one of the 40 Spanish news outlets (by website domain) that are more popular on Meneame.<sup>5</sup> We then applied CA to the interaction matrix between political parties and news outlets. The resulting CA embedding for political parties aligns with the left-right division in Spain (see Figure A3 in the Appendix). We use the first dimension of the CA embedding as our “ground truth” measure of the left-right ideological position of news outlets. This procedure for obtaining ideological positioning is closely linked to latent space models (Barberá et al., (2015)).

As a robustness test, we used the media positioning provided by Political Watch.<sup>6</sup> The qualitative analysis of PoliticalWatch evaluates 30 media outlets' characteristics, such as the wording and fact-checking standards on a sample of articles to assess their ideological leaning. We found a very high correlation (96%) between the ideological position of news outlets determined by PoliticalWatch and our method based on Twitter (Figure A4 in the Appendix).

The media positioning provides us with two key opportunities. First, it allows us to test the performance of SHEEP and CA on the user-to-news outlet network. Second, it allows us to validate our analysis on the *user-to-user* network. For example, users classified as left-wing according only to their interaction with other users can be validated by their vote on left-wing stories in Menéame.

### 3. Results and discussion

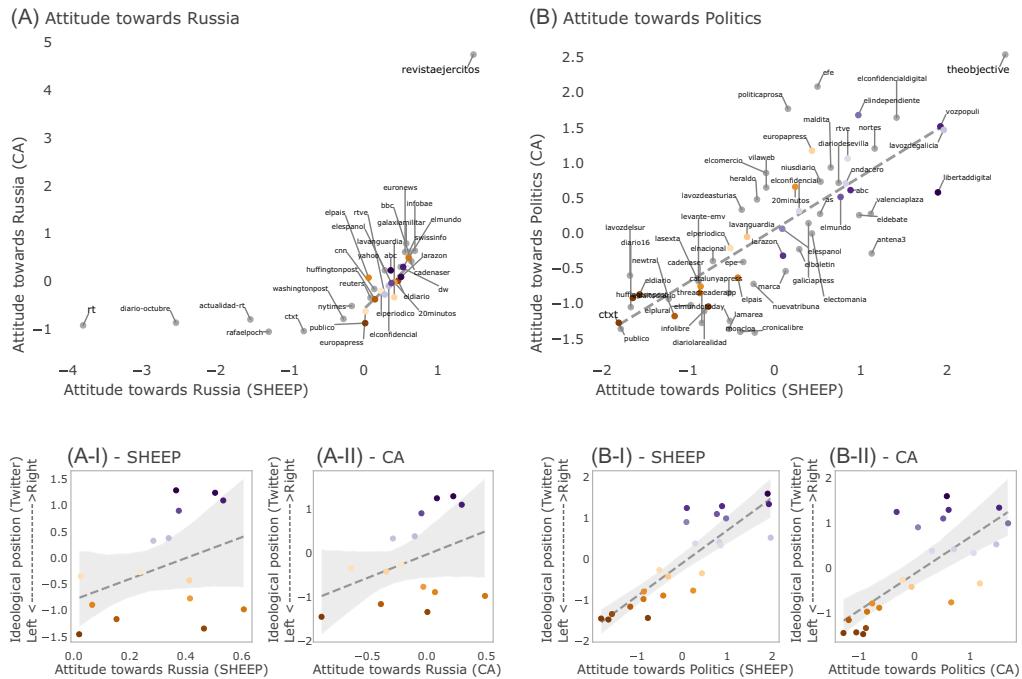
#### 3.1. Mapping the ideological landscapes of news outlets

Understanding the ideological positioning of news outlets is crucial in today's polarized media environment, especially to discern how contentious topics shape public discourse. This ideological positioning is reflected in the *user-to-news outlet* network, where users are connected to a news outlet domain by an edge with an associated weight obtained by aggregating their votes on stories linked to that outlet. Users with an ideological positioning close to the outlet will be more likely to vote positively, while users with an ideological positioning far away from the outlet will be more likely not to vote or vote negatively. Since the ideological positioning might be issue-dependent (e.g., individuals from both the left and right spectrum may support Ukraine in the Ukraine-Russia war), we focus on the two biggest macro-topics on the platform during the specified time frame: *Russia-Ukraine war* and *Broad Politics*.

We use SHEEP and CA to quantify the structural position of each news outlet and compare these results to external benchmarks of left-right ideology (See Methods Section 2.6). We find that incorporating negative ties allows us to uncover patterns of structural polarization that would otherwise remain hidden, especially when observing divisive issues, such as the Russia-Ukraine war.

##### 3.1.1. Ideological mapping in the Russia-Ukraine war

We compare the embeddings of the news outlets nodes obtained by performing SHEEP and CA on the bipartite *user-to-news outlet* network generated by stories classified as belonging to the *Russia-Ukraine war*. We find a moderate Pearson correlation of 58% between CA and SHEEP (Figure 2A). The moderate correlation is due to the unique ability of SHEEP to correctly identify *Russia Today* (*rt*), *Diario Octubre*, and *Actualidad RT* as extreme in the pro-Russia faction. In contrast, CA maps independent left-wing outlets such as *CTXT* or *Publico* at a similar level of attitude as Russia-funded outlets. This divergence between methods highlights the role of negative ties in distinguishing left-wing positions from pro-Russia. Both methods highlight the news outlet *Revista Ejércitos*—an outlet that, according to their website “aims to influence political agendas,



**Figure 2. Comparing news outlets' attitudes toward the Russia-Ukraine war and politics.** The main panels (A–B) display the embeddings for each news outlet, obtained by both positive and negative ties (SHEEP) and only positive ties (CA). The smaller panels (x-I) and (x-II) compare the two embedding techniques with the ideology retrieved from Twitter for a subset of news outlets. Colors represent the Twitter ideology in all panels, ranging from left-wing (brown) to right-wing (dark purple), while news outlets not classified are colored in gray. In the case of Russia (panel A), both CA and SHEEP identify the army-related news outlet *Revista Ejércitos* as an outlier, whereas only SHEEP distinguishes the pro-Russia news outlets *Russia Today*, *Diario Octubre*, and *Actualidad RT* from the left-leaning media such as *ctxt* or *publico*. Panel B shows that the two methods are highly correlated, both identifying ideology in accordance with Twitter (see panels B-I and B-II) when considering the politics topic. In contrast, panels A-I and A-II show that both SHEEP and CA are less correlated with Twitter ideology in the case of the Russia-Ukraine topic.

highlight defense gaps, and promote public investment in Spain's defense industry"—as the most extreme in the other direction.

We then compared the similarity between the embeddings of SHEEP (Figure 2A-I) and CA (Figure 2A-II) with the validated left-right ideology identified from Twitter data (see the Methods section 2.6 for more details). The ideological positioning obtained from SHEEP and CA shows only a moderate Pearson correlation (38% and 39%, respectively) with left-right ideology. This finding underscores that, while Twitter-based ideologies derived from the outlets cited by political parties reflect general political leanings, they do not fully capture issue-specific stances in polarized topics. For example, the left-leaning outlets *eldiario*, *20minutos*, and *cadenaser* are found to be strongly opposed to Russia in this context.

### 3.1.2. Ideological mapping in broad politics

After finding that negative ties allow us to uniquely find pro-Russia outlets, we looked at the embeddings of news outlets on the *user-to-news outlet* network generated from stories in the "Broad Politics" topic, which is a macro-topic incorporating stories about Spanish politics. We find a high Pearson correlation (76%) between the embeddings generated by CA and SHEEP (Figure 2B). The results are also highly correlated (86.6% for SHEEP and 73.1% for CA) with the ideology retrieved from Twitter (Figures 2B-I and B-II).

These results suggest that when applied to the broad politics topic, both methods produce results that align with the left-right political spectrum, but on the Russia–Ukraine War, the embeddings diverge, reflecting issue-specific dimensions that are captured by negative ties.

### 3.2. Examining structural polarization at the user level

This section brings us closer to the central question of the paper: *what do negative interactions—downvotes—reveal about polarization that remains hidden when we consider only positive engagements?* We estimate the SHEEP and CA embeddings for the users using the unipartite *user-to-user* networks, where users vote on the comments of other users. We consider the same two topics as in the analysis at the outlet level: *Russia-Ukraine War* and *Broad Politics*.

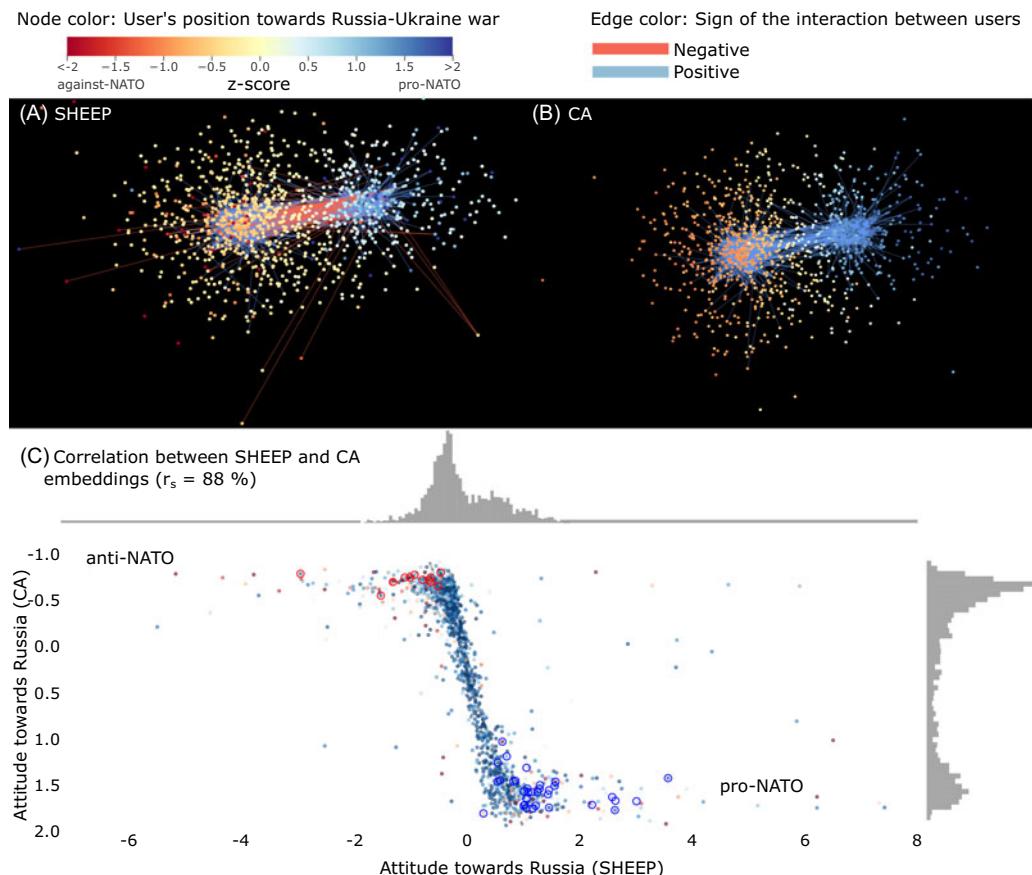
#### 3.2.1. Polarized factions and antagonism in the Russia-Ukraine topic

We begin by visualizing the network using the Fruchterman-Reingold force-directed algorithm (Fruchterman and Reingold (1991)), considering only positive interactions (Figure 3). This algorithm (often called Spring Layout) works by using repulsion (between non-connected nodes) and attraction (between connected nodes with positive interactions) to position nodes in two dimensions. Both the visualization produced by Spring Layout and the embeddings generated by SHEEP and CA (indicated by node colors in Figure 3A-B) reveal two distinct factions. By manually reading the comments, we identify these factions as against NATO and pro-arming Ukraine.<sup>7</sup> While SHEEP and CA demonstrate significant agreement in the overall classification of users within these factions, they differ significantly in identifying the most extreme individuals. SHEEP pulls disengaged users (those with few votes) toward the center of the embedding, while this pattern is less pronounced for CA—the Spearman correlations between degree centrality and SHEEP and CA embeddings are 36% and 22% respectively. As a result, SHEEP identifies extreme users as those with a more active, hostile behavior, i.e., with more negative links (dark red nodes in Figure 3A).

Examining the similarities and differences between SHEEP and CA, we find that the embeddings generated by both methods are highly correlated (Figure 3C), with a Spearman correlation of 88%. As observed in the network visualization, SHEEP and CA only disagree in identifying the most extreme users. In the context of the *Russia-Ukraine war*, both SHEEP and CA exhibit bimodal distributions (histograms in Figure 3C), which correspond to two primary ideological factions on Menéame: pro-NATO users and anti-NATO users. Both methods identify a larger prevalence of users with a negative stance towards NATO. While both methods exhibit a bimodal distribution, indicative of polarization, the SHEEP embedding displays a long tail at both extremes, while CA does not. The differences in the tails are driven by negative votes. This is evident in the color coding in Figure 3C, where red indicates users who predominantly cast negative votes, and blue indicates those who predominantly vote positively. Users located at the tail-ends of the SHEEP embedding are typically associated with negative voting behavior and frequently interact with the two news outlets identified as polarizing: *RT* and *RevistaEjércitos*. Users who vote positively for these outlets are respectively marked with red and blue circles in Figure 3C. As a result, CA often fails to distinguish extremist pro-Russia users from left-wing users who criticize the role of NATO in the years preceding the Russian invasion.

To understand to what extent SHEEP and CA can recover ideological factions with homogeneous voting behavior, we analyze the normalized voting probability (Figure A5C–D) and average vote sign (Figure A5E–F) as a function of the embedding created by each method. These matrices are used to identify ideological factions using the k-means clustering method (See Appendix A.6 for more details) and to facilitate the interpretation of the results. We then examined the voting propensity of the three factions (anti-NATO, moderate, pro-NATO) identified by both SHEEP (boxes in Figure A5E and histograms) and CA (boxes in Figure A5F and histograms).

The majority of the votes of each faction are made to others in the same faction (within the boxes in Figures A5C–D). Interestingly, SHEEP identifies a larger “moderate” faction. This is due



**Figure 3. Visualization of user-to-user network for the Russia-Ukraine war topic.** Both panels share the same layout, generated using the Fruchterman-Reingold force-directed algorithm, but not all the nodes appear on both networks. A random sample of 3,000 nodes is shown. Edges represent interactions: positive in blue, negative in red (only in SHEEP), and those with absolute weight smaller than 3 are filtered out. Node colors show standardized SHEEP (A) and CA (B) embedding values, capped at 2, with red indicating anti-NATO and blue pro-NATO attitudes. The layout reveals two ideological factions with notable cross-faction interaction. Note that for SHEEP (A), the most extreme users (darker shades of blue and red) are located within the network, while for CA (B), they tend to be the ones with a few votes and thus no visible edges. Panel C compares SHEEP and CA in determining user attitudes: each point is a user, with the color indicating their tendency to vote positively (dark blue) or negatively (red). Blue and red circles highlight users voting positively on *revista ejércitos* and *Russia today*, respectively. Histograms show user distribution across the embedding space for each method. Note that SHEEP identifies extreme negative voters in both factions, and the methods are strongly correlated, with a Spearman correlation of 88%.

to the difference in the mapping of users with few votes. While CA considers users with a few votes to extreme users to be themselves extreme, SHEEP considers these users as moderate, as there is not enough information about their latent ideology to pull them far away from the center (see also Figure 3).

The voting propensities match well with the sign of votes cast by users. The pro and anti-NATO factions vote positively for users with similar attitudes and negatively for users with different attitudes (Figures A5E–F). Moreover, while the moderates do not vote negatively for either the pro or anti-NATO factions, they only vote positively for the anti-NATO faction, indicating a closer ideological affinity with this group. Interestingly, this pattern is also found for the CA method, in which negative ties are excluded. This indicates that the information encoded by negative ties

is partially available in positive ties—i.e., the absence of interaction is related to the propensity to vote negatively.

Both CA and SHEEP can identify the ideological polarization on the Russia-Ukraine war on the platform, in terms of the bimodal distribution in the embeddings, while only SHEEP describes the inter-faction hostility that we observe as negative voting patterns between extreme users in opposing factions.

### 3.2.2. Polarized factions and antagonism in the broad politics topic

We perform the same analysis as in the previous section, using the *user-to-user* network representing the politics topic. Unlike in the case of the Russia-Ukraine topic network, the visualization of the politics network does not show two clear ideological factions (Figure 4). Instead, we see a more continuous transition from left-wing to right-wing users.<sup>8</sup> As in the case of the Russia-Ukraine network, SHEEP tends to pull active hostile users towards the extremes (i.e., with more negative links (dark red nodes in Figure 4A), resulting in a higher correlation between degree centrality and ideological position (9% for SHEEP versus 4% for CA).

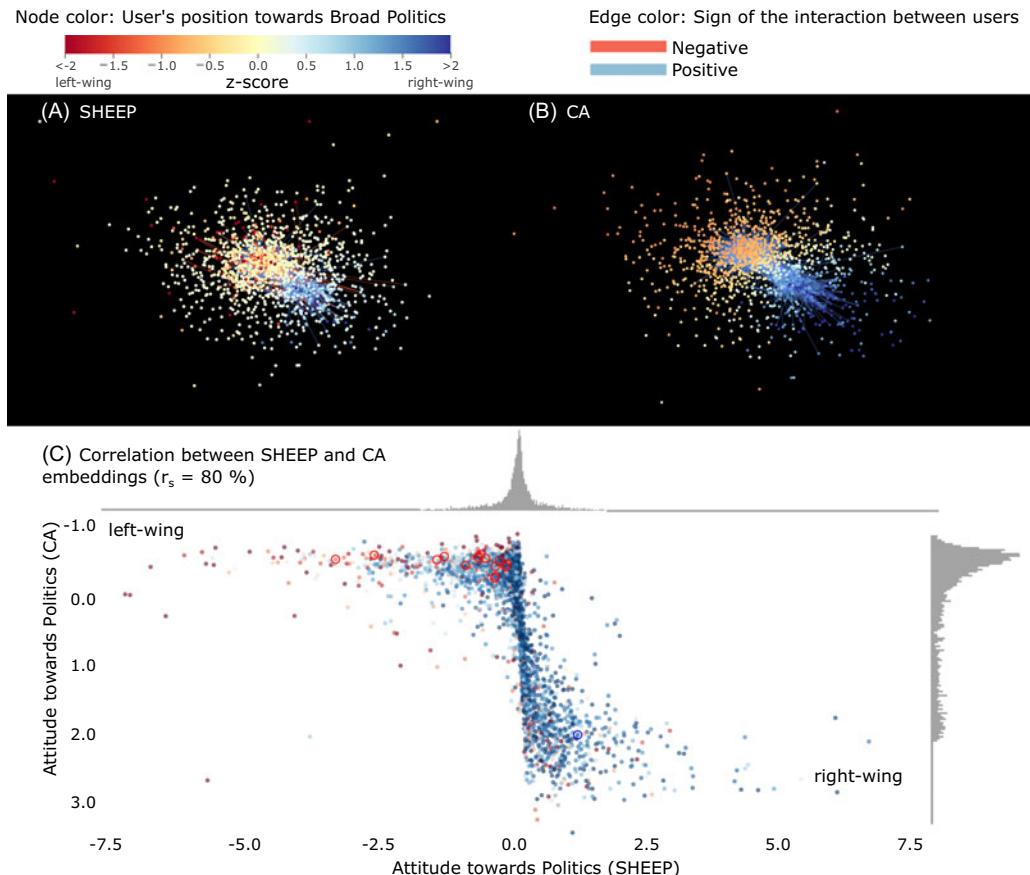
In the case of the broad politics topic, we find more significant differences between the embeddings created by SHEEP and CA (Figure 4C). In the Russia-Ukraine topic, both embeddings display a bimodal distribution. However, in the broad politics topic, only the CA embedding is bimodal, dominated by a large majority of left-wing users. In contrast, the SHEEP embedding exhibits a unimodal distribution with long tails, particularly extending toward far-left users (Figure A6A). While the user distributions in the SHEEP and CA embeddings have different shapes, the position of users within the embedding is remarkably similar (Spearman correlation of 80%). The main differences occur at the tails, where the methods identify different users as extreme. As in the Russia-Ukraine case, the difference is created by negative votes. Only SHEEP can separate far-left users from other left users (the red circles in Figure 4C indicate users who vote positively towards far-left media).

Performing a k-means clustering on the matrices (see Appendix A.6 for details) of votes and signs identifies four factions of users: far-left, left-leaning, right-leaning, and far-right. Users on the left and the right are more likely to vote on comments of users with a similar structural position in the network (Figure A6C–D), but only far-left and far-right users vote negatively towards the opposite faction (Figure A6E–F). Interestingly, the most extreme far-left users also exhibit a high propensity to vote negatively to far-right users (bottom-right corner of Figure A6E–F), but the opposite is not true: far-right users do not engage as significantly with far-left users. This is likely a platform-specific effect, resulting from the asymmetric distribution of left and right-wing users.

In summary, our analysis reveals that the platform’s audience skews toward left-leaning users. In general, users engage positively with others who share similar views, while far-left extremists use negative votes strategically to target the opposing extreme faction. We find that negative interactions are necessary to detect these extreme users.

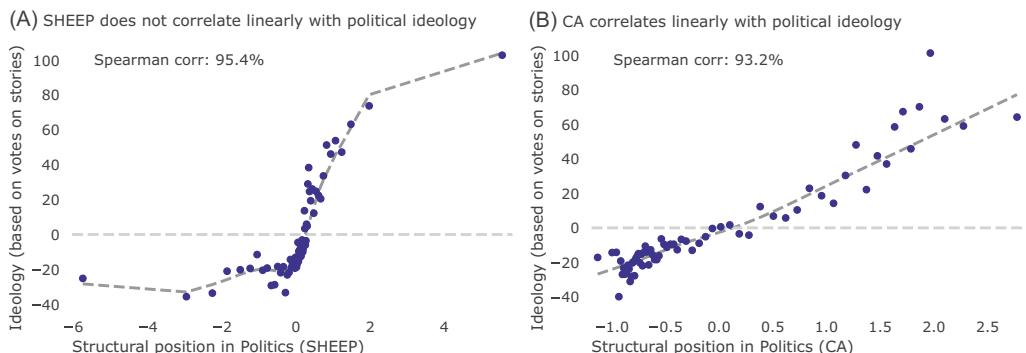
We finally examine whether the structural positions generated by SHEEP and CA align with left-right ideological positions. By comparing the uni-partite *user-to-user* network embeddings to estimates of ideological positioning (see Figure 5) obtained using the *user-to-news outlet* bipartite network, we found that CA exhibits a linear correlation with left-right political ideology, effectively placing users along the ideological spectrum. Conversely, the SHEEP embedding demonstrates a non-linear relationship with ideology, suggesting that SHEEP captures additional dimensions beyond ideological alignment.

Given that extreme users identified by SHEEP use negative votes more frequently than other users, and that negative interactions may signal emotional hostility (Beigi et al., (2020)), SHEEP likely captures underlying elements of this hostility. This phenomenon is particularly pronounced in highly polarizing topics, such as the Russia-Ukraine war, while appearing less pronounced in broader political discussions—except among certain extreme users.



**Figure 4. Visualization of user-to-user network for the broad politics.** Both panels share the same layout, generated using the Fruchterman-Reingold force-directed algorithm. A random sample of 3,000 nodes is shown. Edges represent interactions: positive in blue, negative in red (only in SHEEP), and those with absolute weight smaller than 5 are filtered out. Node colors show standardized SHEEP (A) and CA (B) embedding values, capped at 2. The layout reveals two ideological factions with notable cross-faction interaction. Note that for SHEEP (A), the most extreme users (darker shades of blue and red) are located within the network, while for CA (B), the most extreme users are those with few votes (and thus have no edges visible). Panel C compares SHEEP and CA in determining user attitudes towards politics. Each point represents a user, with color indicating their tendency to vote positively (dark blue) or negatively (red). Red circles are users who vote positively for *far-left media*. The histograms show the distribution of users across the embedding space for each method. We note that SHEEP places extremely negative voters in the left faction, while in general, the two methods are consistent in their identification of most users, with a Spearman correlation of 80%.

Methodologically, we note that SHEEP and CA take as input two different networks; SHEEP takes a signed network, while CA takes an unsigned network. To confirm that the differences we find using the two methods in the following analysis is indeed due to the addition of the negative ties, rather than the choice of algorithm, we reproduce the above analysis comparing SHEEP and CA, where SHEEP takes as input a null model in which we preserve the existing positive ties and add artificial negative ties to all missing edges, simulating for SHEEP what CA already assumes: that not having a connection implies some level of dissimilarity. We find for both topics that the embeddings generated by the SHEEP null model are highly correlated with those from CA, far more than with the embeddings generated by the original version of SHEEP (see Appendix A.9). These results confirm that the differences we observe between the signed and unsigned networks are not artifacts of the algorithm choice but are indeed coming from the addition of negative ties, which offer a more nuanced understanding of the platform polarization.



**Figure 5.** Comparison of the structural positioning derived from two embedding methods: (A) SHEEP and (B) Correspondence Analysis (CA), with ideological positioning. The x-axis represents the binned structural positioning obtained from the embeddings, while the y-axis indicates the average vote of users in each bin toward stories (using the bipartite network described in the previous section). Votes are weighted by the average number of votes per domain and the sign of those votes (see Methods). Points represent ideological bins, with a dashed gray line showing a smoothed regression (lowess) to highlight trends. Positive y-values indicate a higher propensity for and positivity in voting for right-wing news outlets, or a lower propensity for and positivity in voting for left-wing news outlets. Note that CA captures political ideology more linearly compared to SHEEP, which exhibits a non-linear pattern.

Overall, the two methods reveal different aspects of both topics. Although negative votes make up only 3% of all votes on the platform, they are highly relevant for detecting antagonistic hostility in extremely polarized subpopulations—e.g., pro-Russia communities that are conflated with the anti-NATO general faction. Extremist users, both right- or left-leaning, often use negative votes as a tool against the other faction, while more moderate users show support for their ideological camp.

#### 4. Conclusion

As social media becomes one of the primary mechanisms for ideological exchange and emotional expression, it is an increasingly pressing challenge to understand and quantify polarization in digital environments. This study addresses a key gap by exploring what negative ties reveal about polarization that positive interactions alone cannot. Using data from Menéame, a Spanish social media platform, we investigated the dynamics of polarization through a dual-method approach and created a publicly available dataset for future research (available at <https://doi.org/10.5281/zenodo.15682068>).

By comparing the results of SHEEP on the signed network and CA on the unsigned network, we identified both ideological divisions and critical interaction patterns, such as antagonism between extreme factions. Both CA and SHEEP successfully map users and outlets along a political spectrum (verified against independent ideology measures from Twitter and PoliticalWatch), with CA somewhat more effectively mapping ideological polarization among users. However, by including negative links through SHEEP, we reveal behaviors that remain hidden in unsigned networks. For example, in the Russia-Ukraine topic, we detect extreme users who are not only involved in antagonistic interactions but consistently upvote content from Russia-affiliated outlets like RT. This highlights the value of negative ties for identifying users who may play a disproportionate role in spreading polarizing or foreign-influenced narratives.

Our findings show that negative ties are crucial for identifying extreme users who engage in antagonistic behavior. Moreover, we find that far-left users on Menéame are more likely to interact across ideological lines through negative votes, as compared to far-right users, who tend to remain isolated in their interaction patterns. In general, the dual-method approach we take here indicates that the signed and unsigned networks provide complementary insights into the

structure of the polarization, and that both can be used together to better understand ideological dynamics—revealing not only users’ ideological stances, but how they engage with opposing views and contribute to antagonistic activity.

While the methods we employ here can be applied to any signed social network, some aspects of the analysis may not generalize to other online platforms. One limitation is that our findings are based on a platform with a predominantly left-leaning user base, which may affect the representativeness of the results. Oversampling right-leaning users is not straightforward, as ideological scores come from the embeddings themselves, and oversampling users who frequently upvote right-leaning outlets would also require additional assumptions. We see this as an interesting direction for future work. More broadly, future research could replicate this framework on other online platforms with negative interactions to allow for comparisons of user behavior and political leanings. The methods we present could also be modified to combine both CA and SHEEP into one embedding, to allow for an “interpolation” between the information we obtain from each, which could be a rich area for future investigation.

Our study advances the theoretical understanding of online polarization, offering methods to identify extreme users and their behaviors. Our conclusions could inform strategies for mitigating the negative effects of online social media and fostering healthier, more constructive online conversations. Ultimately, this work underscores the complex interplay between ideology, emotion, and interaction in digital spaces, contributing to the broader literature on polarization in online networks and methodologies for studying signed interaction data.

**Acknowledgements.** We are grateful to R. Lambiotte, J. Pougué-Biyong, and Dorian Quelle for useful conversations and feedback on this work.

**Competing interests.** None.

**Funding statement.** S.A. Babul was supported by EPSRC grant EP/W523781/1, and The Alan Turing Institute’s Enrichment Scheme 2023-2024. J. Garcia-Bernardo acknowledges support from the Dutch Research Council (NWO, grant VI.Veni.231S.148).

**Data availability statement.** The code associated with this research is publicly available at the following URL: <https://doi.org/10.5281/zenodo.15496623>. This repository includes the code implementations of the methods discussed in the paper. The dataset is also available on Zenodo at the URL: <https://doi.org/10.5281/zenodo.15682068>.

**Ethical statement.** The data was collected using a custom scraper of [meneame.net](https://www.meneame.net). Ethical approval for this study was obtained from the Faculty Ethics Review Board of Utrecht University (Approval number 23-0111). The data published in this article does not include usernames and has been aggregated to avoid possible identification.

**Author contributions.** Conceptualization: JGB, SAB, AB, EC; Data curation: JGB, EC; Formal analysis: SAB, EC, JGB; Methodology: all; Supervision: JGB, AB; Visualization: JGB, EC; Writing—original draft: SAB, EC; Writing—review & editing: all.

**Declaration of generative AI and AI-assisted technologies in the writing process.** During the preparation of this work, the author(s) used <https://www.deepl.com/write> and ChatGPT 4.0 for copy-editing and to improve readability. After using this tool/service, the authors reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

## Notes

1 <https://www.meneame.net/>

2 The Menéame dataset is available at the following link <https://doi.org/10.5281/zenodo.15682068>

3 <https://www.meneame.net/faq-es>

4 We assume that the absence of interaction is comparable to a null sum of interactions (number of positive votes equal to the number of negative votes), motivated by a low number of such “neutral” interactions, as they entail only 0.74 % of the dataset.

5 *rtve, abc, elmundo, atresplayer, cope, okdiario, larazon, ondacero, telecinco, vozpopuli, youtube, elespanol, europapress, elconfidencial, telemadrid, cuatro, canalsur, eltorotv, elindependiente, cadenaser, eleconomista, elpais, lavanguardia, esdiario, libertaddigital, 20minutos, elperiodico, lasexta, lavozdegalicia, eldiario, huffingtonpost, facebook, twitch, infolibre, publico, laultimahora, elsaltodiario, gaceta*

6 <https://politicalwatch.es/blog/political-watch-publica-primer-media-bias-chart-espana-2021/>

7 An example comment of each side is “You speak as if Russia is preventing Ukraine from joining UNICEF. NATO is not just any “supranational body,” it is a military alliance with its missiles pointed at Russia.” and “Letting Russia do whatever the hell it wants is much more dangerous than arming Ukraine. Among other things because otherwise in a year you would have 3 or 4 other major countries taking example and going over the top of international laws knowing that they have more legitimacy than before.”

8 Two representative comments of extreme users are “of course, you talk like “left-wing voters,” but in the end, you end up saying that you’ll let “the right” win. Very logical, all of it. Then it’s four years of eating shit.” and “Isn’t that exactly what Sánchez [Spanish PM] wants to do to govern? Despite having fewer votes than the PP [main left-wing party], it seems he’ll be the one governing. I haven’t seen any criticism about that here. . . Could it be that everything the left does seems fine to you, and everything the right does seems wrong?”

9 Given two vectors  $\vec{a}$  and  $\vec{b}$ , the cosine similarity is defined as  $\frac{\vec{a} \cdot \vec{b}}{||\vec{a}|| ||\vec{b}||}$ .

## References

- Abramowitz, A., & Saunders, K. (2005). Why can't we all just get along? The reality of a polarized America. *The Forum*, 3(2), 1–22. doi: [10.2202/1540-8884.1076](https://doi.org/10.2202/1540-8884.1076)
- Abramowitz, A. I., & Saunders, K. L. (2008). Is polarization a myth? *The Journal of Politics*, 70(2), 542–555.
- Adamic, L. A., & Glance, N. (2005). The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pp. 36–43.
- Aragón, P., Gómez, V., & Kaltenbrunner, A. (2017). To thread or not to thread: The impact of conversation threading on online discussion. In *Proceedings of the International AAAI Conference on Web and social media*, Vol. 11, 1, 12–21.
- Aref, S., & Neal, Z. (2020). Detecting coalitions by optimally partitioning signed networks of political collaboration. *Scientific Reports*, 10(1), 1506.
- Aref, S., & Neal, Z. P. (2021). Identifying hidden coalitions in the US house of representatives by optimally partitioning signed networks based on generalized balance. *Scientific Reports*, 11(1), 19939.
- Aref, S., Wilson, M. C., & Estrada, E. (2019). Balance and frustration in signed networks. *Journal of Complex Networks*, 7(2), 163–189.
- Babul, S., & Lambiotte, R. (2024). SHEEP, a signed Hamiltonian eigenvector embedding for proximity. *Communications Physics*, 7(1), 8.
- Barberá, P. (2015). Birds of the same feather tweet together: bayesian ideal point estimation using twitter data. *Political Analysis*, 23(1), 76–91.
- Barberá, P. (2020). Social Media, Echo Chambers, and Political Polarization. In Tucker, J. A., & Persily, N. (eds), *Social Media and Democracy. SSRC Anxieties of Democracy*. Cambridge: Cambridge University Press, 34–55.
- Barberá, P., Jost, J. T., Nagler, J., Tucker, J. A., & Bonneau, R. (2015). Tweeting from left to right: is online political communication more than an echo chamber? *Psychological Science*, 26(10), 1531–1542.
- Beigi, G., Tang, J., & Liu, H. (2020). Social science–Guided feature engineering: a novel approach to signed link analysis. *ACM Transactions on Intelligent Systems and Technology*, 11(1), 1–27.
- Benzécri, J., & Bellier, L. (1973). L'analyse des données: Benzécri, J.-P. et al., L'analyse des correspondances. L'analyse des données: leçons sur l'analyse factorielle et la reconnaissance des formes, et travaux du Laboratoire de statistique de l'Université de Paris VI. France: Dunod. <https://books.google.nl/books?id=sDTwAAAAMAAJ>.
- Bird, S., Loper, E., & Klein, E. (2009). *Natural Language Processing with Python*. O'Reilly Media Inc.
- Blei, D. M., & Lafferty, J. D. (2009). Topic models. In *Text Mining*. Chapman and Hall/CRC.
- Bramson, A., Grim, P., Singer, D. J., Berger, W. J., Sack, G., Fisher, S., Flocken, C., & Holman, B. (2017). Understanding polarization: meanings, measures, and model evaluation. *Philosophy of Science*, 84(1), 115–159.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1–27.
- Cartwright, D., & Harary, F. (1956). Structural balance: a generalization of Heider's theory. *Psychological Review*, 63(5), 277.
- Cinelli, M., De Francisci Morales, G., Galeazzi, A., Quattrociocchi, W., & Starnini, M. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, 118(9), e2023301118.
- Clinton, J., Jackman, S., & Rivers, D. (2004). The statistical analysis of roll call data. *American Political Science Review*, 98(2), 355–370.
- Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., & Flammini, A. (2011). Political polarization on twitter. In *Proceedings of the international aaai conference on web and social media*, Vol. 5, 1, 89–96.
- Davis, J. A. (1967). Clustering and structural balance in graphs. *Human Relations*, 20(2), 181–187.

- Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, 113(3), 554–559.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, <https://arxiv.org/abs/1810.04805>.
- DiMaggio, P., Evans, J., & Bryson, B. (1996). Have Americans' social attitudes become more polarized? *American Journal of Sociology*, 102(3), 690–755.
- Doreian, P., & Mrvar, A. (2009). Partitioning signed social networks. *Social Networks*, 31(1), 1–11.
- Enelow, J. M., & Hinich, M. J. (1984). The Spatial Theory of Voting: An Introduction. Cambridge: CUP. <https://arxiv.org/abs/2311.18535>.
- Esau, K., Choucair, T., Vilkins, S., Svegaard, S. F. K., Bruns, A., O'Connor-Farfán, K. S., & Lubicz-Zaorski, C. (2024). Destructive polarization in digital communication contexts: a critical review and conceptual framework. *Information, Communication & Society*, 28(8), 1–22.
- Esmailian, P., & Jalili, M. (2015). Community detection in signed networks: the role of negative ties in different scales. *Scientific Reports*, 5(1), 14339.
- Falkenberg, M., Galeazzi, A., Torricelli, M., Di Marco, Nò, Larosa, F., Sas, M., Mekacher, A., Pearce, W., Zollo, F., Quattrociocchi, W., & Baronchelli, A. (2022). Growing polarization around climate change on social media. en. *Nature Climate Change*, 12(12), 1114–1121.
- Falkenberg, M., Zollo, F., Quattrociocchi, W., Pfeffer, J., & Baronchelli, A. (2023). Affective and Interactional Polarization Align across Countries. doi: [10.48550/arXiv.2311.18535](https://arxiv.org/abs/2311.18535). arXiv: 2311.18535 [physics] (accessed 5 December 2024).
- Ferraz de Arruda, H., Maciel Cardoso, F., Ferraz de Arruda, G., R. Hernández, A., da Fontoura Costa, L., & Moreno, Y. (2022). Modelling how social network algorithms can influence opinion polarization. *Information Sciences*, 588, 265–278.
- Flamino, J., Galeazzi, A., Feldman, S., Macy, M. W., Cross, B., Zhou, Z., Serafino, M., Bovet, A., Makse, H. A., & Szymanski, B. K. (2023). Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. *Nature Human Behaviour*, 7(6), 904–916.
- Fruchterman, T. M. J., & Reingold, E. M. (1991). Graph drawing by force-directed placement. *Software: Practice and Experience*, 21(11), 1129–1164.
- Garamella, K., De Francisci Morales, G., Gionis, A., & Mathioudakis, M. (2017). Reducing controversy by connecting opposing views. In *Proceedings of the tenth ACM international conference on web search and data mining*, 81–90.
- Gerlach, M., Peixoto, T. P., & Altmann, E. G. (2018). A network approach to topic models. *Science Advances*, 4(7), eaao1360.
- Gómez, V., Kappen, H. J., Litvak, N., & Kaltenbrunner, A. (2013). A likelihood-based framework for the analysis of discussion threads. *World Wide Web*, 16(5), 645–675.
- Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis*. Academic Press. <https://books.google.nl/books?id=LsPaAAAMAAJ>.
- Greenacre, M. (2017). *Correspondence analysis in practice*. Chapman and Hall/crc.
- Grootendorst, M. (2022). BERTopic: neural topic modeling with a class-based TF-IDF procedure, <https://arxiv.org/abs/2203.05794>.
- Halford, M. *Prince* [Computer software]. <https://github.com/MaxHalford/prince>.
- Harary, F. (1953). On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2), 143–146.
- Harmon, R. J., & Other Contributors (2023). Tweepy. Zenodo. doi: [10.5281/ZENODO.7860636](https://doi.org/10.5281/ZENODO.7860636) (accessed 24 September 2024).
- Heider, F. (1946). Attitudes and cognitive organization. *The Journal of Psychology*, 21(1), 107–112.
- Hirschfeld, H. O. (1935). A connection between correlation and contingency. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(4), 520–524.
- Hohmann, M., Devriendt, K., & Coscia, M. (2023). Quantifying ideological polarization on a network using generalized Euclidean distance. *Science Advances*, 9(9), eabq2044.
- Honnibal, M., Montani, I., Van Landeghem, S., & Boyd, A. (2020). spaCy: Industrial-strength Natural Language Processing in Python. doi: [10.5281/zenodo.1212303](https://doi.org/10.5281/zenodo.1212303).
- Huang, Z., Silva, A., & Singh, A. (2021). POLE: Polarized Embedding for Signed Networks. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining (WSDM '22)*. Association for Computing Machinery, New York, NY, USA, 390–400. doi: [10.1145/3488560.3498454](https://doi.org/10.1145/3488560.3498454).
- Kaltenbrunner, A., Gonzalez, G., Ruiz De Querol, R., & Volkovich, Y. (2011). Comparative analysis of articulated and behaviourual social networks in a social news sharing website. *New Review of Hypermedia and Multimedia*, 17(3), 243–266.
- Keuchenius, A., Törnberg, P., Uitermark, J., & Guidi, B. (2021). Why it is important to consider negative ties when studying polarized debates: a signed network analysis of a Dutch cultural controversy on Twitter. *PloS one*, 16(8), e0256696.
- Kirkley, A., Cantwell, G. T., & Newman, M. E. J. (2019). Balance in signed networks. *Physical Review E*, 99(1), 012320. doi: [10.1103/physreve.99.012320](https://doi.org/10.1103/physreve.99.012320).
- Leik, Y. (2016). Mass polarization: manifestations and measurements. *Public Opinion Quarterly*, 80(S1), 392–410.
- Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010). Predicting Positive and Negative Links in Online Social Networks. In *Proceedings of the 19th International Conference on World Wide Web*, Raleigh North Carolina USA: ACM, 641–650.
- Leskovec, J., Lang, K. J., Dasgupta, A., & Mahoney, M. W. (2009). Community structure in large networks: natural cluster sizes and the absence of large well-defined clusters. *Internet Mathematics*, 6(1), 29–123.

- Marchi, R. (2012). With Facebook, blogs, and fake news, teens reject journalistic objectivity. *Journal of communication inquiry*, 36(3), 246–262.
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *Journal of Open Source Software*, 3(29), 861. doi: [10.21105/joss.00861](https://doi.org/10.21105/joss.00861).
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), doi: [10.21105/joss.00205](https://doi.org/10.21105/joss.00205).
- Moody, J., & Mucha, P. J. (2013). Portrait of political party polarization. *Network Science*, 1(1), 119–121.
- Neal, Z. P. (2020). A sign of the times? Weak and strong polarization in the US Congress, 1973–2016. *Social networks*, 60, 103–112.
- Offer, S. (2021). Negative social ties: prevalence and consequences. *Annual Review of Sociology*, 47(1), 177–196.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O. . . . Duchesnay, E. (2011). Scikit-learn: machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Peixoto, T. P. (2014). Hierarchical block structures and high-resolution model selection in large networks. *Physical Review X*, 4(1), 011047.
- Peralta, A. F., Ramaciotti, P., Kertész, J., & Iñiguez, G. (2024). Multidimensional political polarization in online social networks. *Physical Review Research*, 6(1), 013170.
- Poole, K. T. (2005). Spatial Models of Parliamentary Voting. In *Analytical Methods for Social Research*. Cambridge; New York: Cambridge University Press.
- Poole, K. T., & Rosenthal, H. (1985). A spatial model for legislative roll call analysis. *American Journal of Political Science*, 29(2), 357–0384.
- Pougué-Biyong, J., Gupta, A., Haghghi, A., & El-Kishky, A. (2023). Learning stance embeddings from signed social graphs. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, 177–185.
- Pougué-Biyong, J., Semenova, V., Matton, A., Han, R., Kim, A., Lambiotte, R. & Farmer, D. (2021). DEBAGREEMENT: A comment-reply dataset for (dis) agreement detection in online debates. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*. [https://openreview.net/forum?id=udVUN\\_gFO](https://openreview.net/forum?id=udVUN_gFO).
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings Using Siamese BERT-Networks. In K. Inui, J. Jiang, V. Ng & X. Wan (eds.), *EMNLP/IJCNLP (1)* (pp. 3980–3990). Association for Computational Linguistics.
- Ribeiro, M. H., Calais, P. H., Almeida, V. A., & Meira, W., Jr (2017). Everything I disagree with is# FakeNews: Correlating political polarization and spread of misinformation, <https://arxiv.org/abs/1706.05924>.
- Richardson, M., Agrawal, R., & Domingos, P. (2003). Trust management for the semantic web. In The SemanticWeb-ISWC 2003: Second International SemanticWeb Conference. Sanibel Island, FL, USA, October 20–23, 2003. Proceedings 2. Springer, 351–368.
- Salloum, A., Chen, THY., & Kivelä, M. (2021). Separating polarization from noise: comparison and Normalization of structural polarization measures. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW1, Article 115, 33. doi: [10.1145/3512962](https://doi.org/10.1145/3512962).
- Shi, G., Altafini, C., & Baras, J. S. (2019). Dynamics over signed networks. *SIAM Review*, 61(2), 229–257.
- Silva, J. A. B. e (2008). An Exploratory Study About Issue and Attribute SalienceWithin the System of News Promotion. Menéame.
- Pérez, A., Ariza, E. T., Gesuelli Pinto, L., & Mazuecos, M. (n.d.). *Paraphrase-Spanish-Distilroberta* [Model card]. Hugging Face. Available at <https://huggingface.co/somosnlp-hackathon-2022/paraphrase-spanish-distilroberta> (accessed 4 August 2025).
- Talaga, S., Stella, M., Swanson, T. J., & Teixeira, A. S. (2023). Polarization and multiscale structural balance in signed networks. *Communications Physics*, 6(1), 1–15.
- Törnberg, P. (2022). How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences*, 119(42), e2207159119.
- Traag, V. A., & Bruggeman, J. (2009). Community detection in networks with positive and negative links. *Physical Review E*, 80(3), 036115.
- Waugh, A. S., Pei, L., Fowler, J. H., Mucha, P. J., & Porter, M. A. (2009). Party Polarization in Congress: A Network Science Approach, 8 December. doi: [10.48550/ARXIV.0907.3509](https://doi.org/10.48550/ARXIV.0907.3509) (accessed 8 December 2023).
- Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2020). Political polarization on the digital sphere: a cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, 38, 1–42.
- Yu, X., & Rodríguez, A. (2021). Spatial voting models in circular spaces: a case study of the U.S. House of representatives. *The Annals of Applied Statistics*, 15(4), 1897–1922.

## Appendix A.

### Appendix A.1. Reciprocal voting analysis

We analyzed mutual voting behavior between users for the topics Russia and Politics. For each user pair, we calculated the number and mean of votes in each direction, retaining only pairs with at least three votes both ways. Tables A1–A2 show a strong degree of agreement and symmetry between users' voting patterns.

**Table A1.** Cross-tabulation of binned average vote scores (Topic: *Russia-Ukraine war*)

	(-1.01, -0.899]	(-0.899, -0.3]	(-0.3, 0.3]	(0.3, 0.899]	(0.899, 1.01]
(-1.01, -0.899]	116	2	2	1	0
(-0.899, -0.3]	2	2	1	0	1
(-0.3, 0.3]	2	1	0	1	1
(0.3, 0.899]	1	0	1	0	17
(0.899, 1.01]	0	1	1	17	2892

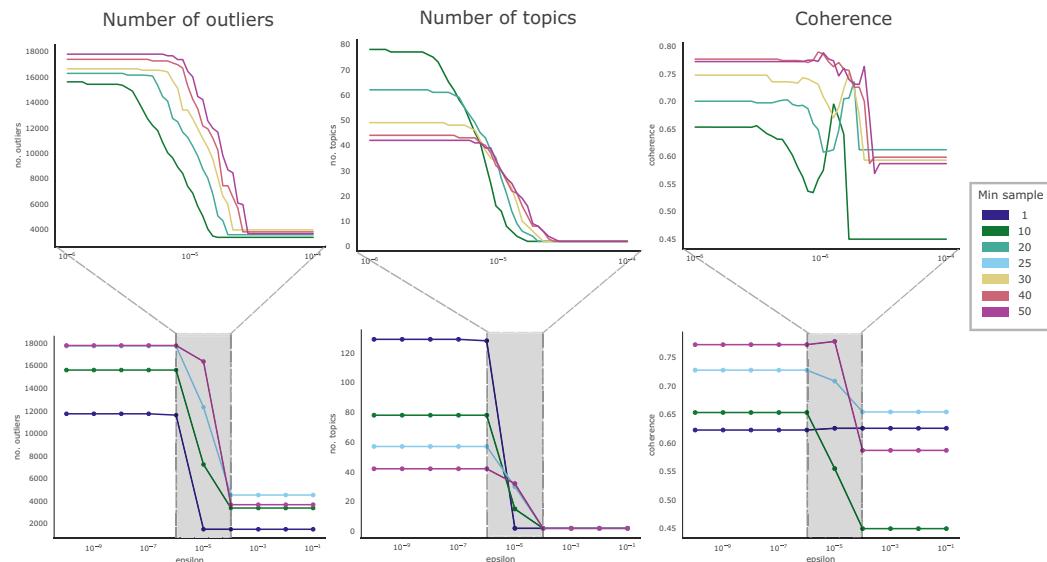
**Table A2.** Cross-tabulation of binned average vote scores (Topic: *Broad Politics*)

	(-1.01, -0.899]	(-0.899, -0.3]	(-0.3, 0.3]	(0.3, 0.899]	(0.899, 1.01]
(-1.01, -0.899]	96	15	6	4	3
(-0.899, -0.3]	15	8	4	5	6
(-0.3, 0.3]	6	4	0	5	5
(0.3, 0.899]	4	5	5	24	96
(0.899, 1.01]	3	6	5	96	5060

### Appendix A.2. BERTTopic algorithm

In this section, we describe the steps followed by the BERTTopic algorithm and the rationale behind each parameter choice:

1. **Sentence embeddings:** This step transforms a text corpus into a collection of vectors, where each vector identifies a text in a multidimensional space. We use SentenceTransformers (Reimers and Gurevych (2019)), a variation of the BERT (Bidirectional Encoder Representations from Transformers) architecture specialized for creating sentence embeddings. In particular, given that our text corpus is in Spanish, we choose a model *paraphrase-spanish-distilroberta* (Pérez *et al.* n.d.) pre-trained on Spanish text data. The outputs are 768-dimensional vectors for each document.
2. **Dimensionality reduction:** Given the high dimensionality of the vectors produced by the previous step, we need to represent the corpus in a lower-dimensional space before proceeding with the clustering task. We use the default technique of BERTTopic, UMAP (McInnes *et al.*, (2018)) (Uniform Manifold Approximation and Projection), choosing the parameters such as *n\_neighbors* = 50, which is a parameter that aims to balance the importance of local (low values) versus global (high values) patterns in the data; *n\_components* = 75, which is a parameter that controls the dimension of the output vectors, to reduce to approximately 10% of the initial dimensions; and last, we chose as similarity metrics the cosine similarity.<sup>9</sup>
3. **Clustering:** after reducing the text embeddings to a lower dimensionality with UMAP, we cluster them using HDBSCAN (McInnes, *et al.* (2017)) (Hierarchical Density-Based Spatial Clustering of Applications with Noise) algorithm. We chose the cosine similarity as metrics again, then we set *min\_cluster\_size* = 50, to avoid the presence of too granular topics (i.e., few stories assigned per topic). Two parameters critical in determining cluster sizes and



**Figure A1.** Variation of the number of outliers, the number of topics, and the coherence, respectively varying the parameters  $\epsilon$  and min sample. As we would like a situation with not many outliers and a "reasonable" number of topics, we choose the following values for the parameters: min\_sample = 1 and cluster\_selection\_epsilon =  $3 \times 10^{-6}$ .

number of outliers are *min\_sample* and *cluster\_selection\_epsilon*. *min\_sample* determines how conservative a clustering procedure should be, i.e., if high, more texts will be found as outliers; *cluster\_selection\_epsilon* defines the radius within which two clusters will be merged. A.2.1 shows the performed tests to understand how the results are affected by this parameter choice. We finally chose *min\_sample* = 1 and *cluster\_selection\_epsilon* =  $3 \times 10^{-6}$ , which ensures a trade-off between the number of topics and outliers.

4. **Vectorizer:** In the previous steps, we subdivided the text corpus into groups. From now on, we want to represent each topic with relevant keywords. Here, we use CountVectorizer from the *sci-kit learn* package (Pedregosa et al., (2011)) to convert the text corpus of each topic into a matrix of token counts (i.e., we want to find the most popular words for each topic). We remove the Spanish stopwords, we limit the number of features to 1000, and we consider words appearing more than 100 times.
5. **c-TFIDF:** from the previous step, we obtained matrices of frequencies of each word in each document of a certain topic. Here, we want to find the most relevant words per topic; therefore, we use the c-TFIDF, a modified version of TFIDF that accounts for a topic-level measurement instead of a document-level measurement.
6. **Representation Tuning:** given the keywords for each topic, we want to perform an additional step to bind the results steps (1)–(3) to the ones of steps (4)–(5). Until now, the same keywords could be the most relevant for all the topics, as we determined them for each topic independently. We employ the method *MaximalMarginalRelevance* to maximize the diversity of those keywords.

#### Appendix A.2.1. Sensitivity analysis of HDBSCAN parameters

Given the high number of model parameters that need fine-tuning, we performed a sensitivity analysis for the clustering step of the BERTopic algorithm. The clustering is performed using the HDBSCAN method, which has, among others, two parameters: *cluster\_selection\_epsilon* and *min\_sample*. Figure A1 shows the results for the number of outliers, topics, and the coherence value, ranging for several values of *cluster\_selection\_epsilon* and *min\_sample*.

### Appendix A.3. hSBM topic model steps

In this section, we describe the steps followed to perform the hSBM topic modeling:

1. **Text preprocessing:** We remove the stopwords using the NLTK (Bird et al., (2009)) Spanish stopwords list and create tokens using the Spacy package (Honnibal et al., (2020)).
2. **Graph creation:** We create the word-document bipartite graph, where the edge weights are given by the frequency of each word in each document.
3. **hSBM fit:** We fit the community detection model on the word-document network, resulting in hierarchical subdivisions of words and documents into non-overlapping clusters.
4. **Topic representation:** Similarly to BERTopic, we apply the CountVectorizer and compute the c-TF-IDF. Then, we identify each topic's top 10 highest-scoring terms and assign them as topic representations.

### Appendix A.4. Topic modeling results

In this section, we show the results of the intermediate topic modeling, before fine-tuning. Tables A3 and A4 show the topic representations obtained with BERTopic and TM-hSBM, respectively.

### Appendix A.5. Ideology from Twitter

The study analyzed tweets from the following political parties and their associated Twitter accounts:

- PACMA (@partidopacma) (Animalist Party Against Mistreatment of Animals): @lau\_duart, @sanchezcastejon, @yolanda\_morpe, @crisgarsalazar
- Más País (@maspais\_es): @ierrejon, @Monica\_Garcia\_G, @EduardoFRub, @isabanes, @Rita\_Maestre, @Equo, @P\_GomezPerpinya, @compromis, @htejero\_, @MasPais\_Es, @MasMadrid\_\_
- VOX (@vox\_es): @Santi\_ABASCAL, @Jorgebuxade, @Ortega\_Smith, @ivanedlm, @monasterioR, @Igarrigavaz, @juan\_ggallardo, @MeerRocio, @VOX\_Congreso, @\_patricia\_rueda
- IU (@izquierdaunida): @agarzon, @sirarego, @EnriqueSantiago, @iuandalucia, @InmaNietoC, @joanmena, @iucyl, @Toni\_Valero, @Congreso\_Es, @ma\_bustamante84, @Roser\_Maestro, @iurioja, @elpce
- PODEMOS (@PODEMOS): @PabloIglesias, @IreneMontero, @ionebelarra, @isaserras, @Yolanda\_Diaz\_, @PabloEchenique, @PabloFdez, @MayoralRafa, @AleJacintoUrang, @Pam\_Angela\_, @MazelLilith, @SofCastanon, @VickyRosell, @nachoalvarez\_, @juralde, @jessicaalbiach, @m\_tere\_perez, @JA\_DelgadoRamos
- Ciudadanos (@ciudadanoscs): @InesArrimadas, @BalEdmundo, @carrizosacarlos, @GuillermoDiazCs, @begonavillacis, @FranciscoIgea, @CiutadansCs, @PatriciaGuaspB, @jordi\_canyas, @MelisaRguezH, @Beatriz\_Pino\_, @Nmartinblanco
- PSOE (Spanish Socialist Workers' Party, @PSOE): @sanchezcastejon, @Adrialastra, @saldavordilla, @mjmonteroc, @felipe\_sicilia, @NadiaCalvino, @carmencalvo\_, @abalosmeco, @isabelrguez, @Pilar\_Alegria, @\_JuanEspadas, @luistudanca, @santicl
- PP (Partido Popular, @ppopular): @pablocasado\_, @TeoGarciaEgea, @cucagamarra, @NunezFeijoo, @IdiazAyuso, @Aglezterol, @AlmeidaPP\_, @alferma1, @abeltran\_ana, @alejandroTGN, @eliasbendodo, @anapastorjulian, @erodriguez\_2019, @JuanMa\_Moreno, @jaimededeolano

**Table A3.** Topics obtained with BERTopic before any outlier reduction

Topic	No. of stories	Topic Representation
-1	16633	juan, empresa, ayuntamiento, china
0	2734	ucrania, rusia, ruso, ucraniano
1	2129	elecciones, voto, votos, electoral
2	1754	perros, animales, animal, arqueologos
3	1573	novela, videojuegos, videojuego, libros
4	1284	policia, agentes, arrestado, agresion
5	1085	prision, condena, condenado, delito
6	1003	musica, cancion, banda, musical
7	979	bancos, inflacion, interes, deuda
8	882	huelga, medicos, hospitales, urgencias
9	769	temperaturas, temperatura, climatico, calentamiento
10	754	luna, mision, universo, planeta
11	752	futbol, jugadores, deporte, liga
12	700	electricidad, renovables, hidrogeno, solar
13	619	laboral, trabajadores, trabajador, empleados
14	572	chatgpt, inteligencia, gpt, lenguaje
15	533	alquiler, viviendas, pisos, propietarios
16	520	miguel, antonio, juan, alberto
17	503	coches, vehiculos, automoviles, coche
18	488	violencia, prostitucion, feminist, sexuales
19	418	carne, comida, alimentos, productos
20	403	abortion, embarazo, castilla, protocolo
21	362	donana, aguas, ecologica, cuenca
22	361	alumnos, educacion, profesores, escuelas
23	356	musk, tuit, mensajes, publicidad
24	348	fotografia, artista, fotografias, fotos
25	341	israel, israeli, civiles, soldados
26	327	constitucional, cgpj, reforma, tribunal
27	291	coronavirus, muertes, enfermedad, pandemia
28	288	cocaina, medicamentos, droga, drogas
29	285	whatsapp, proteccion, microsoft, informacion
30	282	periodismo, desinformacion, profesion, news
31	281	manifestantes, protesta, activistas, protestas

**Table A3.** Continued

Topic	No. of stories	Topic Representation
32	279	puerto, barcos, aguas, costa
33	252	impuesto, impuestos, fiscal, tributaria
34	241	vehiculo, accidente, coche, trafico
35	239	transportes, transporte, pasajeros, movilidad
36	230	cancer, enfermedad, enfermedades, científicos
37	229	incendios, incendio, bomberos, fuego
38	228	jubilacion, reforma, ipc, ingresos
39	225	lengua, castellano, palabra, catalan
40	199	fallecio, murió, actriz, cancer
41	197	amazon, móviles, telefonica, microsoft
42	197	franquista, víctimas, democrática, histórica
43	194	iglesia, abuso, víctimas, sexualmente
44	186	imperio, batalla, reina, soldados
45	169	comisión, donana, censura, calle
46	168	sexual, sexuales, relaciones, mujeres
47	168	vuelos, aviones, piloto, aeropuerto
48	165	sueldo, concejales, alcalde, salario
49	165	económicos, supermercados, inflación, económica
50	154	fiscalía, corrupción, investiga, fraude
51	154	instagram, perfil, contenido, internet
52	147	migrantes, inmigrantes, frontera, méjico
53	143	hospital, urgencias, médico, paciente
54	140	supermercados, alimentos, precios, compra
55	140	terremoto, turquía, siria, costa
56	140	criptomonedas, inversores, digitales, dólares
57	136	bancaria, clientes, empleo, instancia
58	134	palacio, presupuesto, ultraderechista, rodríguez
59	133	enfermedades, medicamentos, sangre, enfermedad
60	133	gonzález, polonia, espionaje, vicepresidente
61	131	empleados, recortes, trabajadores, microsoft
62	130	instagram, comida, lugares, contenidos
63	130	bebé, hermano, foto, propias
64	129	nazis, alemanes, detenidos, soldados

Table A3. Continued

Topic	No. of stories	Topic Representation
65	129	espectadores, television, tve, tv
66	128	suicidio, adolescentes, muertes, causa
67	127	iglesia, articulos, publicar, científicos
68	127	china, oeste, central, provincia
69	125	pobreza, poblacion, porcentaje, estadistica
70	121	marruecos, occidental, autonomia, relaciones
71	121	iglesia, catolica, basura, ferrovial
72	118	venezuela, reuniones, colombia, xunta
73	116	franquismo, infantil, trabajar, ultraderecha
74	112	chino, aerea, espionaje, misiles
75	111	digital, moviles, instagram, tecnologias
76	110	arboles, especies, verdes, reserva
77	108	licencia, edificios, ayuntamiento, viviendas
78	104	censura, candidato, sanchez, debate
79	103	fiestas, festival, navidad, sanidad
80	103	arabia, saudi, siria, relaciones
81	102	franco, flores, franquista, negociacion
82	102	trump, expresidente, jurado, republicano
83	102	bolsonaro, lula, brasil, expresidente
84	101	miedo, relaciones, bienestar, individuo
85	99	periodismo, droga, sovietica, incendios
86	99	japon, literatura, historias, dirigida
87	96	residuos, contaminacion, toneladas, ambiental
88	96	memoria, financiera, riqueza, dispositivos
89	95	acoso, profesor, escuela, alumnos
90	94	netflix, tiktok, anuncios, contenido
91	93	cientifico, científica, representacion, dispositivos
92	93	hogares, familias, pobreza, ingresos
93	93	hollywood, humanos, derechos, agente
94	91	peru, protestas, virgen, abogados
95	90	riqueza, empresarios, economica, empresarial
96	89	tráfico, vehiculo, coche, sancion
97	89	contratos, millon, euros, contrato

**Table A3.** Continued

Topic	No. of stories	Topic Representation
98	87	entrevista, escritor, sanchez, habla
99	86	revista, publicacion, felipe, asociacion
100	86	suiza, franquismo, prostitucion, investigado
101	86	contaminacion, sustancias, salud, muertes
102	86	concierto, festival, responsabilidad, amenazas
103	85	motor, motores, combustible, pobreza
104	85	mapa, urbano, corea, empresarios
105	84	comisaria, martinez, acusa, sancion
106	84	turquia, comicios, elecciones, independencia
107	83	lugares, hijos, familiares, comentarios
108	83	europea, bruselas, conservadores, politicos
109	81	dolares, nacion, envio, suiza
110	81	viral, javier, radio, cadiz
111	81	ruta, sierra, montana, naturales
112	81	muerte, condenado, protestas, ejecucion
113	79	parlamento, europeo, marruecos, escandalo
114	79	microsoft, software, lenguaje, sistemas
115	79	enfermedades, especies, enfermedad, humanos
116	79	padre, murio, felipe, franquismo
117	78	restaurante, carbon, peru, barrio
118	78	bruselas, educacion, odio, restricciones
119	77	imagen, ciudadano, alfonso, diversos
120	76	republicanos, republicano, news, expresidente
121	75	conflictos, invasion, civiles, george
122	75	ilegal, detencion, pablo, inmigrantes
123	74	agricultura, represion, contratos, socialistas
124	74	supermercados, produccion, agricultura, vender
125	74	colores, lluvias, simbolo, palacio
126	74	italia, republica, dictadura, franco
127	74	ejercicio, actividad, adolescentes, bienestar
128	74	radio, contenidos, rtve, documental
129	72	espionaje, republicano, normativa, conservadores
130	71	ortega, nino, lengua, bolsonaro

Table A3. Continued

Topic	No. of stories	Topic Representation
131	71	reina, palacio, felipe, cruz
132	70	racismo, racistas, discriminacion, nazis
133	70	mexico, lopez, venezuela, chile
134	70	valladolid, marruecos, cancer, suicidio
135	68	negocios, trabajar, sector, condiciones
136	68	news, francesa, fraude, argentina
137	67	peru, bolsonaro, brasil, dictadura
138	65	tareas, dictadura, imperio, brasil
139	65	economicas, clases, revolucion, cultural
140	65	drones, costa, policias, asociaciones
141	64	reaccion, comentarios, discriminacion, ultraderechista
142	64	artistas, eta, viva, academia
143	64	reto, daniel, positivo, cristina
144	64	odio, tecnica, acoso, ultraderechista
145	64	tratado, donana, representan, multinacional
146	63	sueldo, edificios, cronica, juventud
147	63	tecnico, gastos, barcelona, financiero
148	63	historicos, restaurante, franco, novela
149	63	alemanes, america, extremadura, hipotecas
150	62	arquitectura, carbono, economicas, cientifico
151	60	moral, policial, protestas, religion
152	60	arboles, parque, ampliacion, obras
153	59	pensamiento, conocimiento, clasico, duda
154	59	bildu, terrorismo, daba, palabras
155	59	movilidad, trafico, ciudades, automoviles
156	59	nazi, qatar, oferta, britanica
157	59	arquitectura, edificios, torres, construccion
158	58	norma, sexual, tribunales, vigor
159	58	corto, demasiado, vineta, minimo
160	58	competencia, telefonica, consumidores, comision
161	57	electoral, martinez, junta, fernandez
162	57	trafico, limites, coches, accidente
163	57	libros, digital, dificultades, pensar

**Table A3.** Continued

Topic	No. of stories	Topic Representation
164	57	inversores, inversion, prevencion, efectivo
165	57	consumidores, ciento, hielo, comercial
166	56	comision, bruselas, europea, deuda
167	56	nazis, extrema, alemanes, odio
168	55	canciones, sindicato, laborales, municipal
169	55	odio, virgen, agresiones, expresion
170	55	armada, saudi, construccion, cristina
171	54	parlamento, terroristas, pobres, relaciones
172	54	coalicion, republica, presidencia, franquismo
173	54	indemnizacion, videojuegos, galicia, condena
174	54	salamanca, colombia, regional, represion
175	53	conservador, financiacion, polonia, territorios
176	53	salarios, salario, inflacion, sueldo
177	53	prostitucion, republicanos, electoral, alcaldia
178	52	victimas, terrorismo, bildu, tratamiento
179	52	michael, reyes, jugador, teoria
180	51	cerebro, memoria, laboratorio, conciencia

**Table A4.** Topics obtained with hSBM

Level	Topic	Representation
0	0	energetico, eolico, gw, electricidad, fotovoltaico
0	1	vladimir, ucraniano, paz, invasion, ruso
0	2	recep, arabia, magnitud, erdogar, siria
0	3	cinematografico, guionista, serie, ficcion, hollywood
0	4	streaming, telefono, meta, web, internet
0	5	alec, yeremi, biyin, fast, payaso
0	6	faso, irak, velo, arma, mahsa
0	7	complutense, curso, facultad, estudiante, educacion
0	8	fascismo, adolf, republica, holocausto, fascista
0	9	lago, metro, kilometro, oceano, marino
0	10	comisario, mediador, juez, audiencia, presunto
0	11	escritor, calle, aquel, san, nombre

Table A4. Continued

Level	Topic	Representation
0	12	operacion, detener, hachis, civil, kilo
0	13	salud, centro, atender, clinico, hospitalario
0	14	pandemia, enfermedad, depresion, anar, poblacion
0	15	precio, alquilar, inmueble, inquilino, propietario
0	16	generativo, herramienta, chatbot, lenguaje, gpt
0	17	mohamed, saharauis, polisario, eurodiputado, rabat
0	18	partido, candidato, pp, elección, ayuntamiento
0	19	comunidad, facultativo, pediatra, consejeria, huelga
0	20	sabor, menu, comer, alimento, patata
0	21	senor, gustar, alguien, nadie, querer
0	22	deportivo, jugar, copa, futbolista, torneo
0	23	luis, baltar, jo, ourense, consistorio
0	24	television, tve, radio, tv, mediaset
0	25	inversor, bitcoin, bancario, entidad, suisse
0	26	demasiado, tira, corto, caracter, manel
0	27	habiar, familia, escuela, bebe, nino
0	28	hablar, siempre, alguien, aprender, gente
0	29	vox, discurso, democracia, partido, politica
0	30	album, cancion, concierto, cantante, musical
0	31	ley, embarazo, aborto, derecho, tran
0	32	bajmut, bakhmut, rusia, soldado, donetsk
0	33	natalidad, elevado, rico, riqueza, economia
0	34	hallazgo, ac, bronce, antiguedad, piedra
0	35	tierra, foto, cielo, elemento, paisaje
0	36	it, you, for, canción, on
0	37	autor, literario, historia, artista, escritor
0	38	sequia, hielo, cambio, calor, oceano
0	39	satelit, spacex, nave, utc, luna
0	40	condenar, victima, abuso, agresion, delito
0	41	rusia, petrolero, licuado, exportación, gnl
0	42	trabajador, paro, cgt, sindical, patronal
0	43	abc, variación, error, exagerar, economista
0	44	oiea, planta, residuo, fukushima, energia

**Table A4.** Continued

Level	Topic	Representation
0	45	gaza, judío, benjamín, palestina, cisjordania
0	46	pozo, andalucía, hidráulico, parque, río
0	47	patrimonio, renta, tributo, fortuna, sucesión
0	48	crucero, titanic, velero, embarcación, orca
0	49	mosquito, microorganismo, genoma, especie, humano
0	50	kmh, circular, coche, radar, carretera
0	51	calle, detener, local, incidente, disparo
0	52	marca, bmw, combustion, tesla, fabricante
0	53	salario, ingreso, jornada, empleado, despido
0	54	aplicación, ley, penal, condenado, rebajar
0	55	formigal, pirineo, ecologizar, sostenible, ribera
0	56	barrio, edificio, calle, habitante, vecino
0	57	consello, monbus, contrato, xunto, tsxg
0	58	derribar, espía, pentágono, tripulado, dron
0	59	filtración, wikileaks, sabotaje, assange, cia
0	60	internet, periodismo, digital, idea, social
0	61	acción, anterior, recaudar, cifra, neto
0	62	contagio, granja, hn, oms, coronavirus
0	63	episcopal, santo, vaticano, diócesis, papa
0	64	mexicano, venezolano, guaidó, mexico, obrador
0	65	texas, aborto, disney, demócrata, legislador
0	66	euskadi, oskar, otegi, abertzale, navarra
0	67	joe, capitolio, carlson, news, republicano
0	68	macarén, iglesia, abascal, pablo, ramón
0	69	partido, popular, líder, investidura, pedro
0	70	plataforma, mastodon, red, blue, tuit
0	71	atari, rol, spectrum, ordenador, nintendo
0	72	vii, muralla, antiguo, ac, dc
0	73	trastorno, cerebro, medicamento, vacuna, farmaco
0	74	insecto, lobo, iberico, fauna, conservación
0	75	reforma, judicial, órgano, vocal, tribunal
0	76	desquadrado, detenido, agredir, hombre, agresión
0	77	recuperación, der, von, comunitario, bruselas

**Table A4.** Continued

Level	Topic	Representation
0	78	eeuu, corea, bric, xi, jinping
0	79	sargento, acuartelamiento, tejera, jarava, cuartel
0	80	sector, empresa, ferrovial, mercado, inversion
0	81	tormenta, meteorologia, calido, aemet, lluvia
0	82	galardon, gala, sabbath, nobel, nominado
0	83	consumidor, compra, producto, mercadona, carrefour
0	84	pasajero, ferrocarril, trayecto, transporte, via
0	85	ministro, anticrisi, consejo, decreto, aprobar
0	86	borbon, monarca, abu, infanta, reina
0	87	estadistico, economia, tasa, decima, crecimiento
0	88	vox, tuberculosis, bovino, latido, junta
0	89	mojacar, comicio, correos, melilla, jec
0	90	pacma, cinegetico, bienestar, jabali, animalista
0	91	software, linux, aplicacion, desarrollador, microsoft
0	92	mercado, electricidad, regulado, mayorista, gas
0	93	tierra, espacial, particula, webb, agujero
0	94	ruanda, boris, johnson, irlanda, rishi
0	95	derogar, aprobo, unidas, diputado, aprobar
0	96	contenciosoadministrativo, indemnizacion, juzgado, indemnizar, recurso
0	97	explosion, transportar, quimico, descarrilar, southern
0	98	adquisitivo, discontinuo, bruto, subida, minimo
0	99	secretaria, violencia, ley, irenir, feminista
0	100	votante, sociologica, sondeo, eleccion, barometro
0	101	ponsati, junquera, independentista, borra, proz
0	102	planta, reciclar, co, tonelada, envase
0	103	aguirre, pp, yolanda, madrilén, comunidad
0	104	arder, roda, evacuar, monte, llama
0	105	morada, coalicion, belarra, vicepresidenta, montero
0	106	paulo, inacio, luiz, brasileno, brasilia
0	107	provincial, acusado, condenar, fiscalia, condena
0	108	infanteria, entrenamiento, abrams, ucrania, blindado
0	109	deposito, central, inflacion, tipo, euribor
0	110	scholz, bukelir, naciones, violacion, onu

**Table A4.** Continued

Level	Topic	Representation
0	111	vladimir, putin, paramilitar, ruso, yevgeny
0	112	pienso, leche, cosecha, cereal, cultivo
0	113	arbitral, bartomeu, negreirar, rref, fc
1	0	grado, mas, solar, mar, espacial
1	1	invasion, otan, militar, putin, trump
1	2	civil, peru, presidente, brasil, derechos
1	3	tener, vida, hacer, mas, ser
1	4	compania, plataforma, millón, electrico, usuario
1	5	aviar, emerito, gripe, juan, felipe
1	6	alquiler, sanitario, hospital, sindicato, laboral
1	7	si, él, derecho, jo, violencia
1	8	dos, padre, él, jugador, habiar
1	9	menor, victima, guardia, hombre, agente
1	10	tumba, yacimiento, imperio, arqueologico, ac
1	11	sanchez, ayuso, alberto, nunez, electoral
1	12	gas, mwh, ohio, fontdevila, grafico
1	13	ave, universo, humano, estrella, telescopio
1	14	software, apple, pc, ordenador, windows
1	15	economia, trimestre, euros, subida, petroleo
1	16	estadounidense, china, wagner, rusia, incendio
1	17	regadio, agricultor, parque, alimento, sequia
1	18	gobierno, galicia, xunta, fiscal, reforma
1	19	bildu, partido, igualdad, congreso, tribunal
2	0	primero, si, poder, hacer, tener
2	1	hacer, tener, euros, romano, precio
2	2	ley, madrid, vox, él, poder
2	3	rusia, ucrania, ruso, primero, europeo
3	0	primero, si, poder, hacer, tener

#### Appendix A.6. Clustering user voting behaviors

To analyze user interactions based on their ideological positioning, we visualized two interaction matrices: one representing the total number of votes from users in ideological bin X to those in ideological bin Y, and another showing the average sign of those votes. The number of bins was determined by taking the square root of the total number of users (as there are  $B \times B$  interactions, where  $B$  is the number of bins) and dividing by a normalization constant, which we set to 1.5.

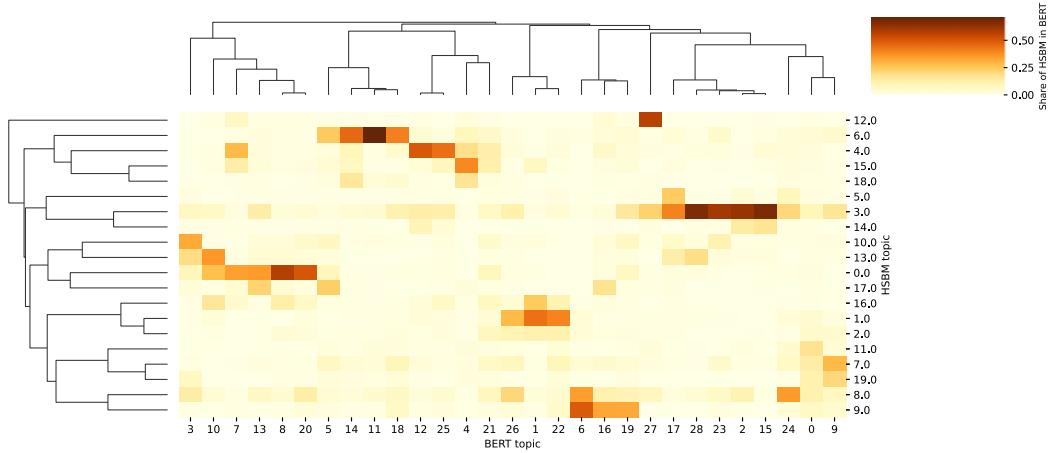


Figure A2. Confusion matrix with colors indicating the overlap between topics identified by BERTTopic and hSBM.

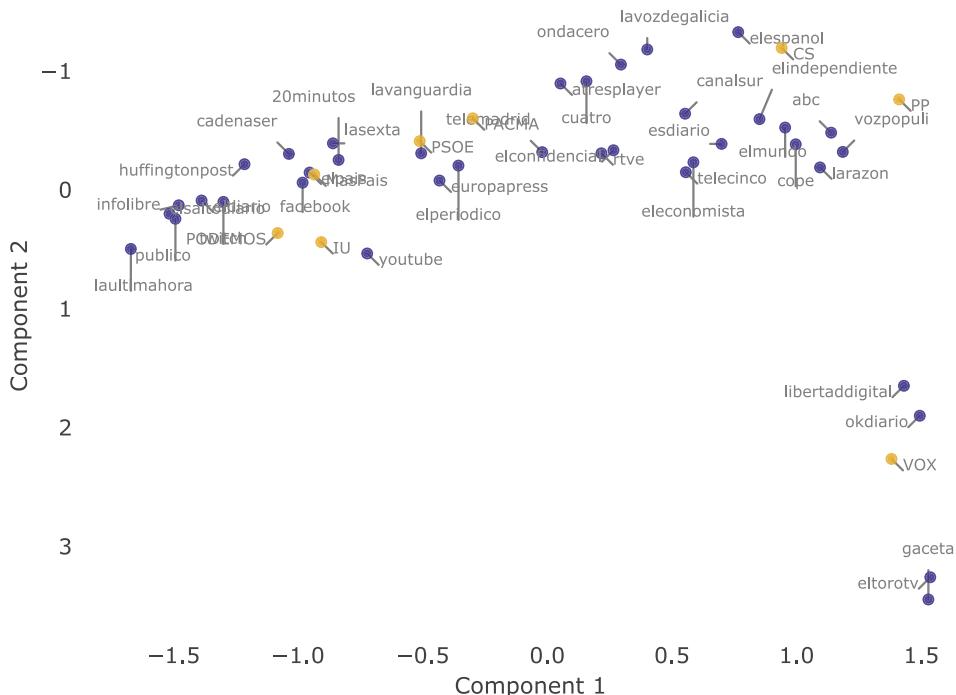
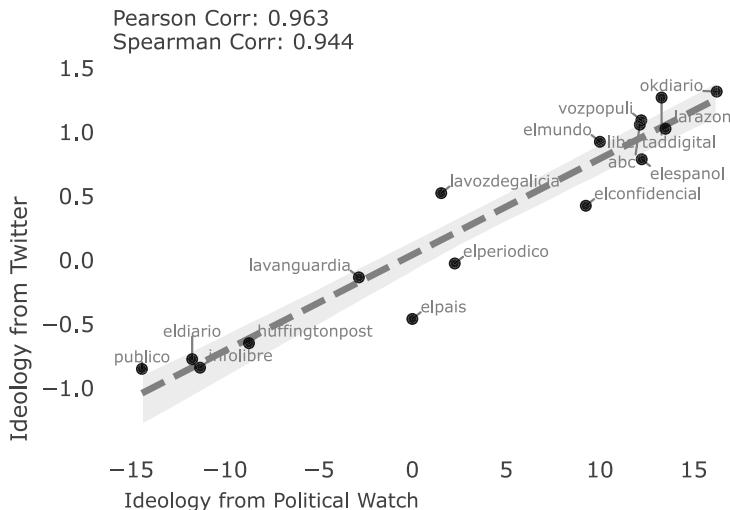


Figure A3. Visualization of the first two components of the Correspondence Analysis for both news outlets (dark purple) and political parties (yellow). We interpret the first dimension as left-right ideology, and the second dimension (not used in this paper) as mainstream-radical.

To ensure comparability and reduce biases caused by variations in interaction activity, we applied an iterative normalization process to the total vote matrices. This process balances the influence of row and column totals, preventing highly active users from skewing the patterns. Specifically, each row and column of the matrix was normalized iteratively by dividing by its



**Figure A4.** Comparison of the ideology of media outlets and Twitter accounts. The x-axis represents the ideology of media outlets as reported by Political Watch, while the y-axis represents the ideology of Twitter accounts. The dashed line represents the regression line.

respective sum, ensuring each sum to one. This normalization reflects the relative strength of interactions independent of individual user activity levels, allowing for a fair comparison of voting behaviors and enabling meaningful clustering and visualization of user patterns across ideological groups.

We used the two matrices—normalized vote count and average vote sign—as inputs for a K-Means clustering algorithm. Each ideological bin was characterized by its normalized voting probabilities with all other bins. The optimal number of clusters was determined using the Variance Ratio Criterion (Calinski-Harabasz score) (Calinski and Harabasz (1974)), which evaluates the ratio of between-cluster dispersion to within-cluster dispersion.

The resulting clusters were directly used in visualizations. In cases where clusters overlapped (which occurred only at boundary bins), the overlap was removed to maintain clarity.

#### ***Appendix A.7. Validation of political ideology***

To determine if the structural positioning of users in the *user-to-user* network corresponds to left-right political ideology, we employed the following steps:

### Appendix A.7.1. Step 1: Binning user positions

Users were grouped into bins of 50 individuals each based on their structural positions as derived from the Signed Hamiltonian Eigenvector Embedding for Proximity (SHEEP) and Correspondence Analysis (CA) methods. This binning approach ensured that each bin represented a manageable and consistent sample size for subsequent analysis.

### Appendix A.7.2. Step 2: Calculating voting behavior

For each bin, we analyzed the voting behavior of users by calculating how often users in that bin voted positively for each media outlet, compared to the average positive voting behavior across the platform. Let  $V_{u,m}$  represent the average voting sign of user  $u$  towards media outlet  $m$ , and  $\bar{V}_m$

the average voting sign for  $m$  across all users. The deviation  $D_{u,m}$  was computed as:

$$D_{u,m} = V_{u,m} - \bar{V}_m$$

Positive deviations indicate a higher share of positive votes towards  $m$ , while negative deviations indicate a lower share of positive votes.

#### Appendix A.7.3. Step 3: Weighting by media ideology

Each deviation  $D_{u,m}$  was weighted by the ideological positioning of the media outlet  $I_m$ , which was extracted from the *user-to-news outlet* network (using Twitter ideology yields highly similar results). The weighted deviation for user  $u$  and outlet  $m$  is given by:

$$W_{u,m} = D_{u,m} \cdot I_m$$

This step provides an estimation of the ideological preference of user  $u$ . Higher  $W_{u,m}$  indicates a higher share of positive votes towards right-wing media or a lower share of positive votes toward left-wing media.

#### Appendix A.7.4. Step 4: Adjusting for voting frequency

To account for variations in the number of votes cast toward different outlets, the weight was further scaled by the ratio of the user's voting frequency  $F_{u,m}$  for outlet  $m$  to the average frequency  $\bar{F}_m$  across all users:

$$R_{u,m} = W_{u,m} \cdot \frac{F_{u,m}}{\bar{F}_m}$$

This adjustment ensures that outlets receiving disproportionately high or low attention are appropriately weighted in the analysis.

We calculate  $R_{bin,m}$  as the average  $R_{u,m}$  for all users in the bin and  $F_{bin,m}$  as the voting frequency of users in the bin toward outlet  $m$ .

#### Appendix A.7.5. Step 5: Aggregating ideological estimates at the bin level

We aggregated the average weighted deviation for each bin, which was calculated as:

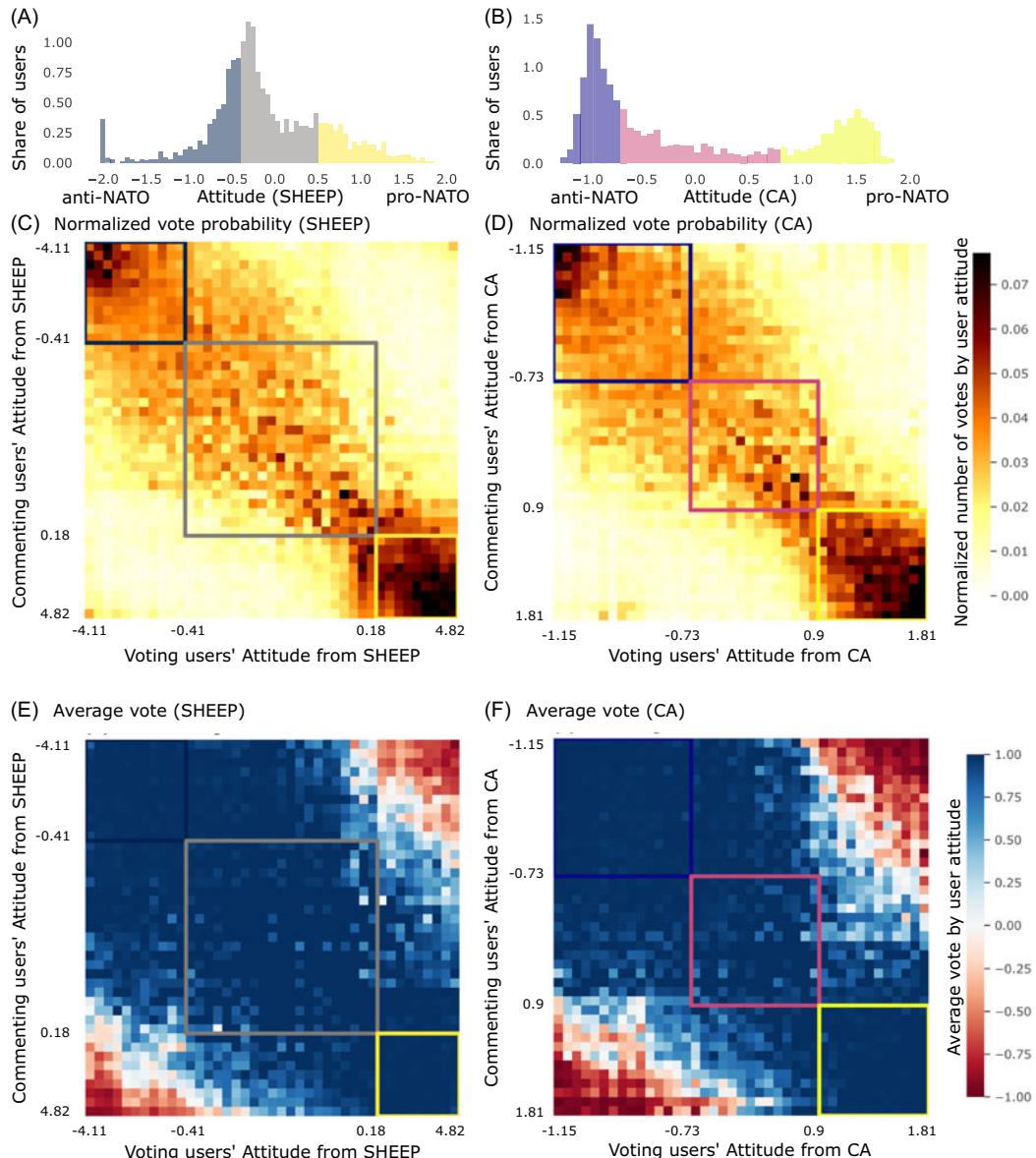
$$I_{bin} = \frac{\sum_m R_{bin,m} \cdot F_{bin,m}}{\sum_m F_{bin,m}}$$

This value represents the aggregated ideological positioning of users within the bin, based on their voting behavior towards media outlets.

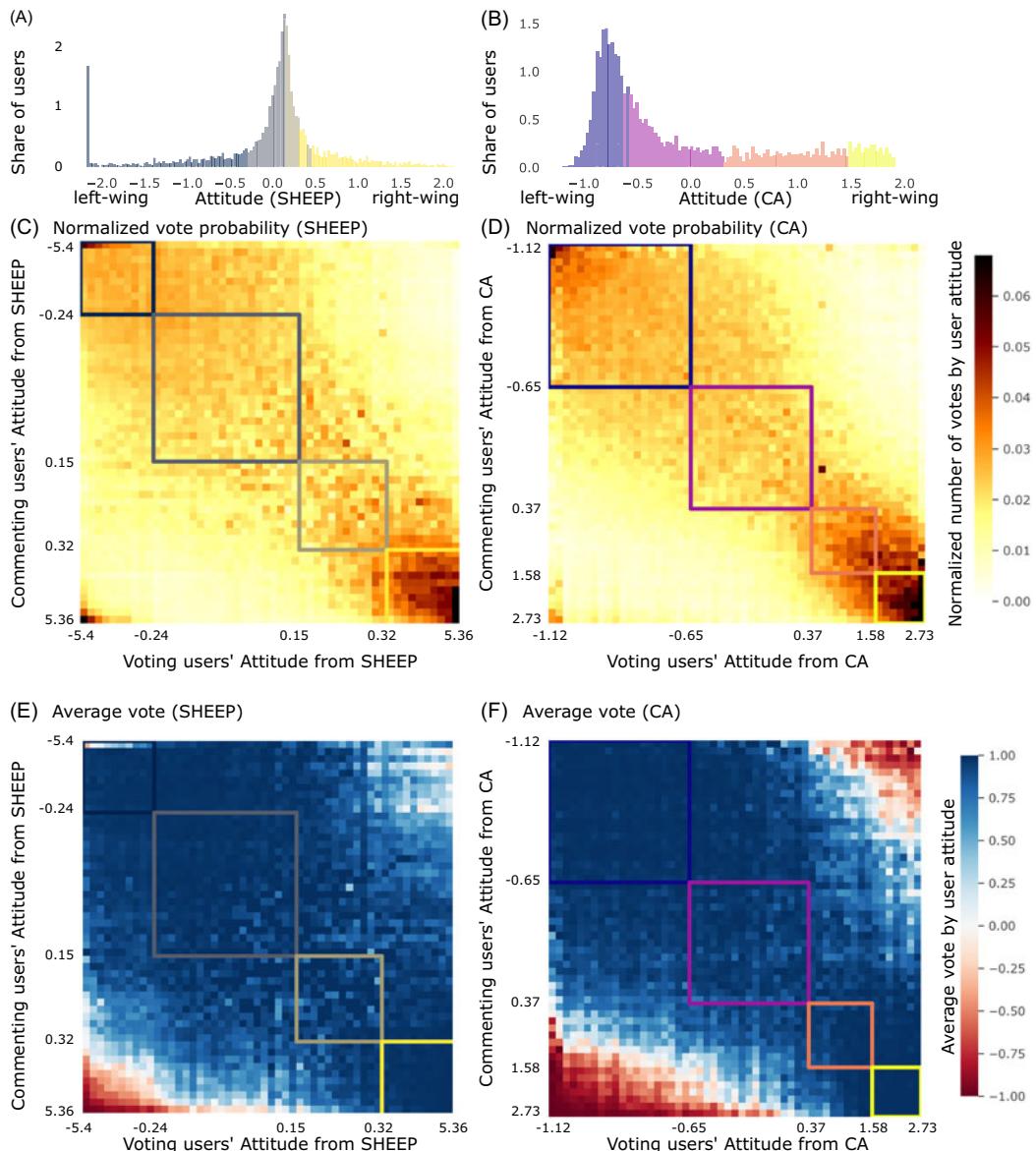
### Appendix A.8. Voting behavior and political ideology

#### Appendix A.9. SHEEP null model comparison

SHEEP is a method designed to take signed networks as input, requiring both positive and negative edges to generate meaningful embeddings, while CA requires unsigned networks as input. In this section, we investigate whether the different embedding results we obtain from the two methods for the *user-to-user* network are indeed due to the additional information gained from negative ties or just a result of the algorithmic choice. To that end, we create a signed network null model to input into the SHEEP algorithm, which obscures the information about the real negative ties by replacing all non-links in the unsigned network with an artificial negative link of weight -50, a similar magnitude to the largest positive links in the unsigned graph. This null model makes the same assumption as CA, that disconnected nodes are inherently dissimilar.

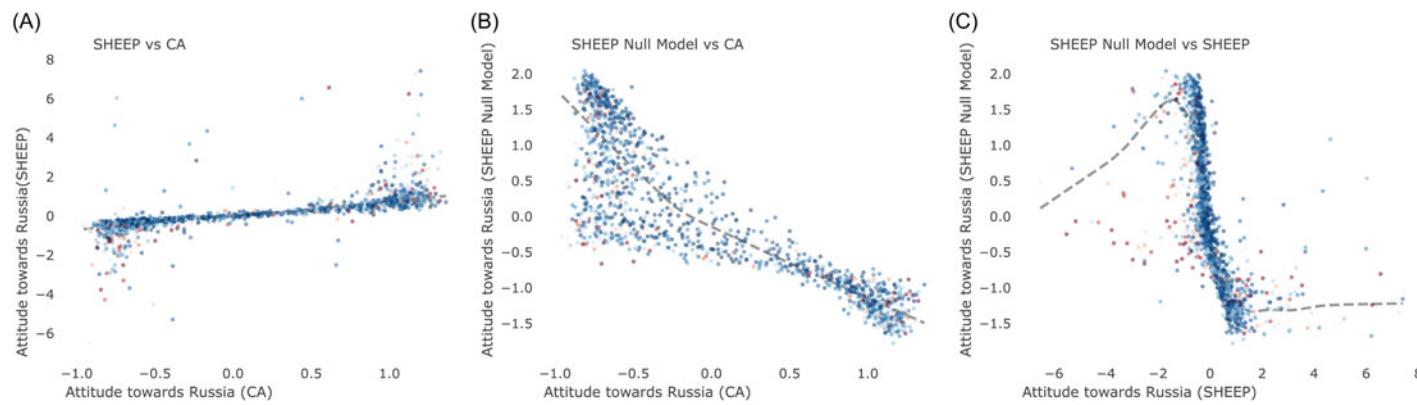


**Figure A5. Analyzing users' views on Russia-Ukraine war.** Panels A and B show the distribution of users across the embedding space for SHEEP and CA, respectively. Bar colors reflect k-means clustering (see Appendix A.6). Note that histogram scales differ from Figure 3 as the bins have uniform size. Panels C and D display heatmaps of normalized vote probabilities, with rows representing voters' attitudes and columns representing the attitudes of users they vote on. Note that voting tends to occur between users with similar attitudes (votes often lie close to the diagonal). Panels E and F show average votes cast on stories and comments, ranging from  $-1$  (all negative) to  $+1$  (all positive). Users generally vote positively, except for extreme users, who downvote the opposite extreme. Clusters from k-means (matching panel A/B colors) are shown as boxes: anti-NATO (blue, dark purple), moderate (gray, pink), and pro-NATO (yellow). See Appendix A.7 for details on binning, vote normalization, and cluster interpretation.

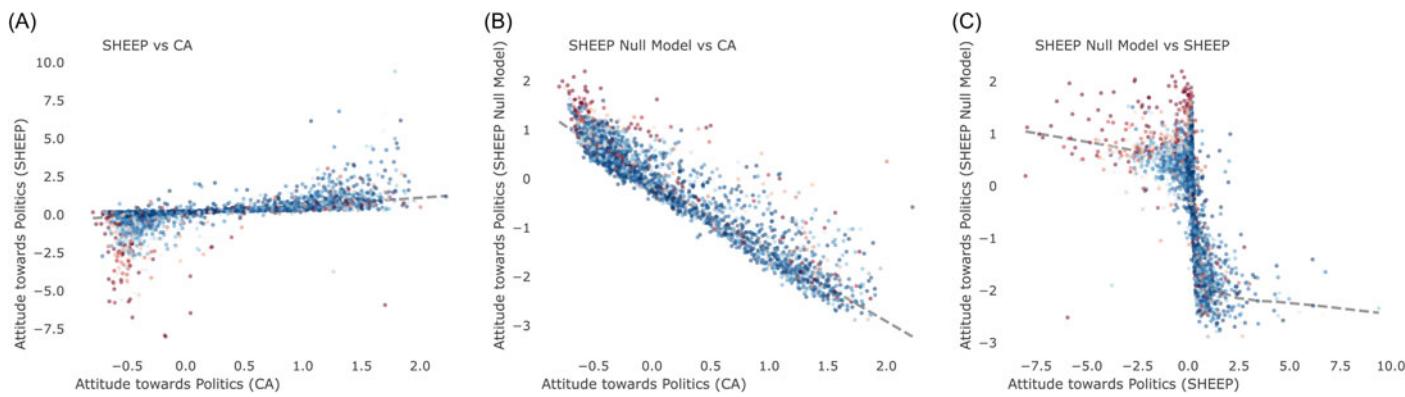


**Figure A6. Analyzing users' views on broad politics.** Panel A and B show again the distribution of users across the embedding space, for SHEEP and CA, respectively. Bar colors reflect k-means clustering (see Appendix A.6). Note that histogram scales differ from Figure 4 as the bins have uniform size. Panels C and D display heatmaps of normalized vote probabilities, with rows representing voters' attitudes and columns representing the attitudes of users they vote on. Panels E and F show average votes cast on stories and comments, ranging from  $-1$  (all negative) to  $+1$  (all positive). Users generally vote positively, except for extreme left-wing users, who vote negatively towards the opposite extreme. Clusters from k-means (matching panel A/B colors) are shown as boxes: far-left (dark blue, dark purple), left-wing (light blue, pink), right-wing (gray, orange), and far-right (yellow). See Appendix A.7 for details on binning, vote normalization, and cluster interpretation.

Figures A7 and A8 show the position of the user in the embedding space generated for the *user-to-user* network across the three models. The color indicates the propensity of the user to vote positively (dark blue) or negatively (red) in the real signed network. For both the Russia and politics topic, we find that the embeddings generated by the SHEEP null model are extremely aligned



**Figure A7.** User-to-user network embeddings on Russia topic. Comparing user embeddings generated by CA, SHEEP, and SHEEP null model. As in Figure 3, the color indicates the tendency of the user to vote positively (dark blue) or negatively (red). The Pearson correlation for SHEEP vs CA is 0.68, SHEEP null model vs CA is 0.83, and SHEEP null model vs SHEEP is 0.61.



**Figure A8.** User-to-user network embeddings on politics topic. Comparing user embeddings generated by CA, SHEEP, and SHEEP null model. As in Figure 4, the color indicates the tendency of the user to vote positively (dark blue) or negatively (red). The Pearson correlation for SHEEP vs CA is 0.56, SHEEP null model vs CA is 0.94, and SHEEP null model vs SHEEP is 0.54.

with those of CA, much more than with the embeddings generated by SHEEP acting on the real signed network. In both cases, the Pearson correlation is highest between the SHEEP null model and CA (0.83 for Russia and 0.94 for politics), indicating that the SHEEP null model produces an embedding that is highly linearly correlated with that produced by CA. These results once again support the conclusion that the differences we observe when incorporating negative ties are not simply artifacts of the algorithmic choice, but rather that the addition of the real negative tie information offers a more nuanced perspective on polarization.