Journal of Computational Social Science (2025) 8:22
https://doi.org/10.1007/s42001-024-00350-y

RESEARCH ARTICLE

Characteristics of two polarized groups in online social networks' controversial discourse

Amin Mahmoudi[1] · **Dariusz Jemielniak**[1] · **Leon Ciechanowski**[1,2]

Received: 25 May 2024 / Accepted: 9 December 2024

Abstract

In today's interconnected world, online social networks play a pivotal role in facilitating global communication. These platforms often host discussions on contentious topics such as climate change, vaccines, and war, leading to the formation of two distinct groups: deniers and believers. Understanding the characteristics of these groups is crucial for predicting information flow and managing the diffusion of information. Moreover, such understanding can enhance machine learning algorithms designed to automatically detect these groups, thereby contributing to the development of strategies to curb the spread of disinformation, including fake news and rumors. In this study, we employ social network analysis measures to extract the characteristics of these groups, conducting experiments on three large-scale datasets of over 22 million tweets. Our findings indicate that, based on network science measures, the denier (anti) group exhibits greater coherence than the believer (pro) group.

Keywords Social network analysis · Misinformation · Echo chambers · Network measures

✉ Amin Mahmoudi
amahmoudi@kozminski.edu.pl
https://scholar.google.com/citations?user=wQihe8oAAAAJ&hl=en

Dariusz Jemielniak
darekj@kozminski.edu.pl
https://scholar.google.com/citations?user=rqW2VUsAAAAJ&hl=en

Leon Ciechanowski
lciechanowski@kozminski.edu.pl
https://scholar.google.com/citations?user=94tY3HMAAAAJ&hl=en

[1] Management in Networked and Digital Societies (MINDS) Department, Kozminski University, Warsaw, Poland, 03-301, 57/59 Jagiellonska St

[2] MIT Center for Collective Intelligence, Massachusetts Institute of Technology, Cambridge, MA, USA

Published online: 30 December 2024

Introduction

Background

The anti-science movements in online social networks (OSNs) have been growing in popularity over the last decade. The COVID-19 pandemic has brought a growth in anti-vaxxer sentiments about this particular vaccine [1], partially inspired by bad-will, professional actors [2], but the overall vaccine hesitancy in online spaces has grown as well [3]. Many other topics are prone to misinformation on OSN, too, such as the war in Ukraine [4]. Even more exotic fringe theories abound: flat Earthers unite over the internet all around the globe [5]. Overall, the anti-intellectual and anti-scientific views are clearly on the rise [6–8].

Climate change, a phenomenon that has garnered significant attention in recent years, has been the subject of extensive research [9, 10]. This issue has led to the formation of echo chambers on social networks, characterized by two primary factions: those who perceive climate change as a serious threat to humanity's future, and those who dismiss it as a conspiracy theory.

Regarding anti-science movement and disinformation spreading, Russia has employed various tactics of information warfare to discredit Ukraine's sovereignty and legitimacy, sow discord among NATO allies and erode trust in Western institutions [11]. Similarly, during the coronavirus pandemic in Poland, Poland has faced numerous false or misleading claims about the virus origin, transmission, prevention and treatment that have fueled public confusion and anxiety [12, 13].

Recent scholarly efforts have sought to elucidate the characteristics of echo chambers, which are essentially formed by two polarized groups. Jiang et al. [14] explored the attributes and mechanisms of echo chambers, identifying four key mechanisms: homophily, recommender algorithms, confirmation bias, and cognitive dissonance. They further delineated five common attributes of echo chambers, namely, the creation of social trends, the propagation of conspiracy theories, the diffusion of misinformation, political polarization, and the emotional contagion of users.

Complementing this, Alatawi et al. [15] provided a comprehensive overview of echo chambers, focusing on their attributes, mechanisms, detection and modeling strategies, and mitigation and prevention approaches. Their work underscores the complexity of echo chambers and the multifaceted challenges they present in the context of social networks and information dissemination.

Motivation

The field of network science has yet to fully explore the characteristics of deniers and believers, a limitation that this study aims to address. While it may initially appear that analyzing social network content related to these groups would not contribute to resolving this issue, the opposite is true. Accurately identifying users who deny certain phenomena can enhance our understanding of their message dissemination strategies, their methods of reinforcing their beliefs, and their influence on others' thoughts. Furthermore, to mathematically model these users' behaviors, it is essential to measure their characteristics within the context of network science. This study's primary motivation is to examine these two main groups from three perspectives: inter-member connections, activity distribution, and group structures. It is important to note that some network users may not initially fall into a specific category. However, by generalizing common features within this user group, it is possible to identify other deniers using machine learning techniques, thereby creating a self-aware network.

Contribution

This paper explores the characteristics of deniers and believers in OSNs debates, utilizing fifteen measures derived from network science. These measures are categorized into three domains: connections, distributions, and segmentation. First of all, revealing the coherence within the deniers (anti) group justifies why, despite having fewer members, their voice in the network is strong, and they can be considered creators of echo chambers. By determining the characteristics of these two user groups based on the aforementioned categorization, we gain a more nuanced understanding of echo chamber formation in the context of OSNs' debates. Moreover, the network measures provided herein can be employed in structure-based approaches to detect echo chambers, a method followed by some studies.

The paper is structured as follows: Sect. "Related works" reviews the relevant literature. Sect. "Data" provides a statistical overview of the datasets used. Sect. "Theoretical framework" presents the theoretical framework of the study. Sect. "Measures" describes the measures used to highlight the differences between deniers (anti) and believers (pro). Sect. "Results and discussion" details the experiments conducted to test the proposed method, along with the results and discussions. In Sect. "Implication", we discuss the implications of the study. Finally, Sect. "Conclusion" concludes the paper.

Related works

Around the beginning of the second decade of the 21st century, the use of the Internet and online tools to investigate public opinions about controversial issues began. Before that, this issue was often discussed and analyzed in newspapers and television. [16] used the Google search engine to assess public responses to Japan's proposed climate change mitigation policies. Later, [17] emphasized the role of microblogs in studying public perceptions of climate change.

Gradually, with the spread of online social networks such as Twitter, this issue entered these media. in [18] analyzed hashtags related to Intergovernmental Panel on Climate Change's (IPCC), they found that, in general, the use of these hashtags on Twitter can make an unfamiliar topic like climate change more tangible to those interested in these topics.

In the context of social media network structures, several researchers have applied network science principles to investigate climate change discourse. Williams et al. [19] explored key aspects such as sentiment analysis, community analysis, and homophily within Twitter data. They discovered a direct correlation between polarization and personal bias, with active participants in online climate discussions exhibiting strong attitudes towards skepticism or belief, while neutral perspectives were largely absent. Their study revealed a high degree of homogeneity in climate change discussions on social networks, with users forming like-minded clusters and primarily interacting within these groups.

Holme and Rocha [20] attempted to bridge climate change and network science by defining network measures in three categories: centrality, vitality, and controllability. They sought to model the climate system as a complex network system. Baumann et al. [21] examined user activity levels and interaction rates, demonstrating that more active users tend to express extreme opinions. Crucially, they found that a user's opinions often mirror those of their network neighbors, suggesting a mechanism for the reinforcement and propagation of extreme and radical views.

As online social networks expanded, two primary groups emerged: those who accept climate change as anthropogenic, and those who deny it. Tyagi et al. [22] developed a framework to analyze the dialogue between these two competing Twitter user groups, using climate change-related tweets from the United Nations Climate Change Conference – COP24 (2018) as a case study. They found that both deniers and believers predominantly converse within their respective groups, a trend particularly pronounced among deniers. Deniers' messages primarily target believers in anthropogenic climate change. Similarly, Neff and Jemielniak [23] analyzed COP25 tweets to show the emergence of transnational public spheres. Van Eck et al. [10] demonstrated that believers primarily engage with mainstream climate blogs, while deniers favor climate-skeptical blogs. Cody et al.[24] analyzed Twitter responses to climate change news, finding that the majority of responses came from climate change activists rather than deniers.

The phenomenon of echo chambers in online social networks has also been a subject of extensive research. Cinelli et al. [25] quantified echo chambers on social media using two main elements: homophily in interaction networks and bias in information diffusion towards like-minded peers. Their study of Facebook and Twitter data revealed a strong tendency for users to form like-minded groups. Jasny and Fisher [26] reported that the climate denial movement has significantly hindered political progress on the climate crisis in the United States. They used a novel dataset collected from central policy actors involved in the US climate policy network in 2017 to analyze the formation of echo chambers. They found that conservative think tanks, public communications by fossil fuel companies, and private philanthropy have all contributed to the growth of these echo chambers.

Samantray and Pin [27] analyzed tweet texts and defined a measure for polarization and homophily. They found that an increase in homophily can reinforce individual beliefs, leading to the creation of echo chambers and increased polarization. They also suggested that increased polarization can lead to societal segregation based on differing beliefs, thereby naturally increasing the probability of like-minded communication. Walter et al. [9] showed that users typically align with dominant opinions in the media. They investigated factors that influence skepticism or support regarding the anthropogenic nature of climate change, including the risk of social isolation for individuals whose views diverge from the mainstream. This phenomenon, based on the spiral of silence theory, suggests that social media content can influence individual behavior, potentially leading individuals to conform to dominant behaviors on social media. The authors posited that echo chambers tend to reinforce existing opinions rather than promote dialogue and critical reasoning. They identified influencing factors at four levels: country, news outlet, individual journalist, and individual news story, and studied these factors across five countries. They concluded that echo chambers in climate change discourse can shape public opinion, and that climate scientists aiming to disseminate evidence-based knowledge must engage with a broader range of media outlets, including tabloid news and conservative media, to better inform the debate.

Data

We have acquired the data from George Washington University Libraries Dataverse, the *Climate Change Tweets Ids [Data set]* [28].[1] This dataset has been collected from the Twitter API using Social Feed Manager, and totalled to 39,622,026 tweets related to climate change. The tweets were collected between September 21, 2017 and May 17, 2019. However, there is a gap in data collection between January 7, 2019 and April 17, 2019. The tweets with the following hashtags and keywords were scraped: climatechange, #climatechangeisreal, #actonclimate, #globalwarming, #climatechangehoax, #climatedeniers, #climatechangeisfalse, #globalwarminghoax, #climatechangenotreal, climate change, global warming, climate hoax.

Due to Twitter's Developer Policy, only the tweet IDs were shared in the database, not the full tweets. Therefore, we had to *hydrate* the tweet ids with the use of Hydrator application [29]. Hydrating was carried out by us in June, 2020, and it allowed us to obtain 22,564,380 tweets (some tweets or user accounts are deleted or suspended by Twitter in its standard maintenance procedures). Challenges encountered during data hydration included dealing with deleted tweets or suspended user accounts, which is a common occurrence in Twitter's standard maintenance procedures. We addressed this by using the Hydrator application, which allowed us to recover as much data as possible within the constraints of Twitter's Developer Policy.

In order to comprehensively diagnose Polish social networks and to enable automated classification of Twitter users in terms of their attitude towards vaccinations, we collected a balanced, importance-wise database of Twitter users for manual annotation. The most important keywords used by groups that spread anti-vaccination propaganda were identified. Using our programming pipeline, databases of Polish social media on the topic of the pandemic and attitudes towards vaccinations were obtained. The raw data contained over 5 million tweets from almost 3600 users with the following hashtags related to the COVID-19 pandemic in Poland and the war in Ukraine: stopsegregacjisanitarnej, nieszczepimysie, szczepimysie, szczepienie, szczepienia, koronawirus, koronawiruswpolsce, koronawiruspolska,

1 https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi%3A10.7910%2FDVN%2F5QCCUU.

rozliczymysanitarystow, stopss, covid, covid19, sanitaryzm, epidemia, pandemia, plandemia, zelensky, zelenski, wojna, muremzabraunem, konfederacja, wojnanaukrainie, putin, ukraina, ukraine, rosja, russia, wolyn, bandera, upa. Twelve annotators rated the scraped Twitter users based on their posts on a nine-point Likert scale. Samples evaluated by annotators were partially overlapped in order to examine their consistency and reliability. Statistical tests performed on data before and after binning (in three- and two-category versions) confirmed significant annotator agreement. Fleiss' kappa, Randolpha, Kirchendorff alpha, and intracorrelation coefficients indicate non-random agreement among the competent judges (annotators).

Our initial data acquisition based on the abovementioned hashtags yielded 5,308,997 posts. To focus specifically on discussions related to COVID-19 and the war in Ukraine, we implemented a filtering process using Polish word stems relevant to these topics. This step reduced our dataset to 4,840,446 posts. The filtering was performed using regular expressions based on lemmatized versions of key terms. For war-related content, we used stems such as 'wojna' (war), 'inwazj' (invasion), 'ukrai' (Ukraine), and 'putin'. For COVID-related content, we used stems like 'mask' (mask), 'szczepi' (vaccine), and 'koronawirus' (coronavirus). This approach allowed us to capture various grammatical forms of these words.

Following this initial filtering, we removed three users who had no posts related to either COVID-19 or the war in Ukraine. This step left us with 3,597 users and 4,839,995 posts. Finally, to ensure consistency in our analysis, we selected only posts in the Polish language. This final step resulted in our dataset of 3,577,040 posts from 3,597 users. Before the tweets content analysis was performed, text lemmatization had been performed, special characters, links, and low-importance words based on a stop list (e.g. conjunctions) had been removed.

Data preprocessing has been carried out in Python programming language [30] with the use of specific libraries and our original code. The hydrated tweets were further cleaned by removing duplicates and all tweets that had no English language label. Some characters and technical expressions were then replaced with natural language terms (e.g., changing "&amp" into "and"). We have also created a couple of versions of the database, for various purposes—in some of them we have replaced emoji pictures with their descriptions (using the *demoji* library and our original code), for other database versions we have removed the emojis, hyperlinks, and special characters. This caused the dataset to comprise 24,083,452 tweets (7,741,602 tweets without retweets), which makes it the biggest database of social media data referring to climate change analyzed to date.

We created the social network directed graph with the use of RAPIDS cuGraph library in Python[2] for most of the network statistics calculations, and also with the use of the graph-tool [31]. The final graph visualization was created with the use of Gephi [32] after preparing and filtering the data in Python. The final graph had 4,398,368 nodes and 18,595,472 edges, after removing duplicates and self-loops.

The final label of "believer," "denier," or "neutral/unknown" was assigned to users present across annotators through the averaging of results from multiple annotators.

[2] https://github.com/rapidsai/cugraph.

In the Ukraine dataset, the term ‘anti-group’ refers to various tactics of information warfare aimed at discrediting Ukraine’s sovereignty and legitimacy, whereas the ‘pro-group’ consists of tweets that support Ukraine’s sovereignty and legitimacy. In the Vaccine dataset, ‘anti’ denotes a group of users who publish tweets against vaccination, while ‘pro’ users advocate for vaccination programs. In the Climate Change dataset, ‘denier’ users dismiss it as a conspiracy theory, while ‘believer’ users perceive climate change as a serious threat to the future of humanity.

Theoretical framework

Existing data in climate change and war datasets in this study show that the number of members in the deniers (anti) group is significantly lower than in the believers (pro) group. Despite this, the deniers group is often the main actor in creating and spreading controversial issues. Given that this group is capable of influencing a large number of nodes (users) in the network, our study aims to investigate the network structure of this group and explore why it plays a crucial role in spreading controversial topics despite having fewer members. Additionally, while existing studies on influence maximization have highlighted the importance of network structures, such as clustering coefficient, path, and node degree, several studies have focused exclusively on central or influential nodes. Berahmand et al. [33] introduced the DCL model to detect spreaders, emphasizing their importance in information diffusion within collective networks. Their model takes into account key location parameters, including node degree, the degree of its neighbours, common links between a node and its neighbours, and inverse cluster coefficient. Salavati et al. [34] proposed the GLR model to rank influential nodes, indirectly addressing information spreading. Their model enhances closeness centrality by leveraging the local structure of nodes. Rui et al. [35] introduced the Reversed Node Ranking (RNR) model, which assesses a node’s influence and the impact of its neighbours on that node. Additionally, Wen & Deng [36] devised the local information dimensionality (LID) model, which considers central nodes and their local structural properties in the context of information diffusion.

Many researchers are increasingly delving into the intricacies of network structures to harness their potential in solving information diffusion problems. In his work, Chen [37] shed light on the pivotal role of network structure and explored how the information diffusion process hinges on the quantity and positioning of seed nodes. To identify influential early adopters, Chen employed four centrality measures, including betweenness centrality, closeness centrality, K-shell, eigenvector, as well as two heuristics: the greedy algorithm and degree discount.

Liu et al. [38] emphasized the significance of the average path length in information diffusion. They proposed two heterogeneous nonlinear models—one for modeling the topologies of information cascade trees and another for simulating the stochastic process of information diffusion within social networks. Key features for modeling cascade tree topologies included the average path length and degree variance of cascade trees. Their study revealed that users with a large number of friends had a lower probability of disseminating information, particularly in terms of reading messages on WeChat.

In addition, we investigate the role of two network metrics in influence maximization —clustering coefficient (CC) and betweenness centrality (BC)—which significantly highlight network closure, path information, and coherency in the network. We developed an information diffusion model based on these two features, and the experimental results show that, on average (over 10 runs), our model activated 55% of nodes in the deniers group and 50% of nodes in the believers group. This finding indicates that in a network with higher values of CC and BC, the number of activated nodes is higher. Our information diffusion model takes as input a graph relevant to the network, consisting of nodes, edges, associated probabilities (randomly assigned to each edge in the network), and the k value (representing the seed node). The model first randomly selects a node and designates it as the seed node, initiating the information cascade, denoted as C. Within a while loop, the algorithm discovers the neighbors of the tail node of C. If the tail node has no neighbors, the algorithm terminates. Otherwise, the BC and CC are computed for each neighbor, and their identifiers, along with the computed BC and CC values, are appended to a data structure. In the second phase of the algorithm, we aim to find the best neighbor for the current node. There are two scenarios to consider: when there is more than one neighbor and when the current node has only one neighbor. In the first scenario, if there are multiple neighbors, the model determines the best neighbor's index based on the BC value. In the second scenario, when the current node has only one neighbor, it assesses the probability value. If the probability is greater than the threshold defined in this study (0.5), the information is diffused to the next node. Otherwise, the process is terminated.

Therefore, our hypothesis for this research is that the deniers group is more coherent compared to the believers group, such that despite the lower number of members in the deniers group, information propagates more smoothly, leading to the activation of more nodes. To do this, we investigate network measures in three categories such as connections, distributions and segmentation to highlight the coherency of deniers group.

Measures

In this section, we introduced some measures that deploy to address the characteristics of believers and deniers groups, these measures widely used in network science especially social network analysis. These measures fall into three main categories, namely, connections, distributions and segmentation, and are discussed under their respective category.

Connections

The connections category of social network measures encompasses the connection characteristics that exist between users in OSNs.

Assortativity. Assortativity coefficient (r) of a graph measure can be defined as the probability of the nodes in a graph linking to other nodes of similar degree [39–41]. To determine this measure, the Pearson correlation coefficient is computed for the degrees of the node pairs for all the edges in a graph, and this calculation returns values in a range between -1 and 1. An assortativity value of less than zero shows that nodes connect to others with dissimilar degrees. On the other hand, an assortativity value that is greater than zero indicates that nodes are likely to connect with those with similar degrees.

Network Closure. Allcott et al. in [42] stated that the concept of network closure is used to indicate the level of connectivity between friends of friends. So, if friends in a network have common friends (neighborhood), then this network is considered to exhibit a high level of network closure. In contrast, if the friends in a network have different friends, then low network closure is present. According to [43], *network_reciprocity* function from the *igraph* package can be used to compute network clouser, in fact a way to find the network clouser is computing reciprocity. This measure in network sciences refers to the likelihood of vertices to be mutually linked.

Motifs: Network motifs are sub-graphs that repeat themselves in a specific network or even among various networks. Each of these sub-graphs, defined by a particular pattern of interactions between vertices, may reflect a framework in which particular functions are achieved efficiently.

Density: The density measure is calculated as the number of existing connections in a network divided by the number of possible connections in a network [39, 44, 45].

Distributions

The distributions category contains measures that describe how users are distributed in a network. This is very important especially for identifying the influential nodes in a network.

Closeness centrality: The closeness for a vertex is the reciprocal of the sum of geodesic distances to all other vertices as in (1):

$$C_c(v) = \frac{1}{\sum_{t \in V \# v} d_G(v,t)} \tag{1}$$

where $d_G(v,t)$ is geodesic from v to t.

Path length: To establish the path length, the radius and diameter are calculated using the eccentricity of each node in a social graph. Eccentricity is defined as the maximum distance between a node and any other nodes in the graph. The radius is the minimum number of all eccentricities, while the diameter is the maximum. The average path length is simply the average of all-pairs-shortest-paths on the social graph [39, 45].

Network skewness: The skewness statistical function is utilized to show the asymmetry of the data distribution. Under positive skewness, most of the interactions occur at the start of the time period, whereas under negative skewness the majority of interactions take place at the end of the time period. The distribution of data is approximately symmetrical if the skewness value is between − 0.5 and 0.5. Therefore, this study considers distribution to be abnormal (asymmetrical) if the skew value is higher than 0.5.

Segmentation

The segmentation category of measures addresses the issue of clustering in OSNs, and the measures in this category, especially cliques, are widely used in community detection research.

Clustering Coefficient (transitivity): The cluster coefficient (CC) is a fraction of the possible interconnections in a network as in (2). The value of CC is between 0 and 1. The CC of a whole network is the average of CC of each user [39, 45]

$$CC(v) = \frac{2 * N_v}{k_v * (k_v - 1)} \tag{2}$$

where k_v is a degree of v and N_v is the number of connections between the neighbors of v.

Clique: A clique in a network is formed when all the nodes have connections with each other. This measure is usually used in community detection algorithms as it can show how dense the connections are in a community.

Community structure: Community detection (CD) studies attempt to develop methods to find a group of nodes which have a higher connection with each other than with other nodes in the rest of the network. The basic algorithm for detecting communities is the GN algorithm, which was proposed by [41]. The GN algorithm is based on the maximum betweenness between the nodes in each community and the lowest interconnection between the nodes in different communities. To evaluate the results of the GN algorithm, the authors also introduced the modularity measure (Q). Modularity is computed as the best division such that the greatest number of edges are within communities and the least are between communities.

Strongly connected component (SCC): The portion of a directed graph in which there is a path from each vertex to another vertex.

Weakly connected component (WCC): A directed graph is weakly connected If for its undirected equivalent graph, there is a path between its two vertices.

Table 1 Network measures in terms of connections

Measure		Dataset					
		Vaccine		Ukraine		Climate change	
		Anti	Pro	Anti	Pro	Deniers	Believers
1	Assortativity	– 0.09	– 0.29	– 0.15	– 0.30	– 0.034	– 0.021
2	Network reciprocity	0.0025	0.0012	0.005	0.001	0.0048	0.0000
3	Motifs	725	1690	2092	1058	1083	20
4	Density	0.004	0.004	0.006	0.003	0.005	0.001

Table 2 Network measures in terms of distribution

Measure		Dataset					
		Vaccine		Ukraine war		Climate change	
		Anti	Pro	Anti	Pro	Deniers	Believers
1	Closeness	0.044	0.015	0.07	0.06	0.06	0.008
2	Diameter	11	13	8	12	8	9
3	Skewness (duration)	0.63	0.52	0.30	0.61	– 2.36	– 2.39
4	Skewness (join date)	– 0.60	– 0.56	– 0.22	– 0.53	3.09	3.31

Results and discussion

Experimental setup

In this section, we present the results obtained from the experiments based on the aforementioned measures in Sect. "Theoretical framework". It is important to note that due to the disparity in the number of users between the two groups and consequently the variation in the number of connections within these groups, we employed the permutation method to test the results of the measures mentioned in Sect. "Theoretical framework". For this purpose, we randomly selected a subgraph with 200 connections 1000 times. In each iteration, the entire set was shuffled to ensure the distinctiveness of the subgraphs, resulting in different subgraphs each time. However, it was not possible to apply this approach to certain measures, such as SCC, diameter, and closeness, which rely on the path between nodes. The reason behind this limitation is that when selecting a subgraph for the permutation test, multiple components are formed within the believers' group. Consequently, when computing these measures using R language functions, the length of the path between two nodes within a component is considered as one, which is incorrect. Hence, the permutation test was not conducted for these three metrics.

To ensure the robustness and statistical significance of our findings, we computed the pseudo p-value for each metric. In the context of this study, the pseudo p-value is a measure used to determine whether the observed differences between the Deniers and Believers groups could be attributed to random chance. The pseudo p-value is calculated using the formula $\frac{k+1}{n+1}$ where k represents the number of iterations in which the metric value for the deniers (anti) group is consistent with our hypothesis and supports the observed results, as presented in Tables 1, 2, and 3. n is the total number of permutations. This approach involves repeatedly shuffling the data and recalculating the metrics to create a distribution of possible outcomes under the null hypothesis (i.e., no real difference between groups). The pseudo p-value quantifies the proportion of these permutations where the deniers (anti) group's metrics align with the hypothesis, relative to the total number of permutations. A low pseudo p-value suggests that the observed differences are unlikely to be due to random chance, thereby providing strong evidence that the structural characteristics of the deniers (anti) network, such as higher transitivity or density, are indeed distinct and significant.

Connection analysis

In this section, we analyze connections in the three aforementioned datasets. Table 1 presents the experimental results in terms of connection measures, including assortativity, motifs, density, and reciprocity. The assortativity values indicate that, in general, there is no tendency to form connections between nodes with the same degree in all datasets. However, a comparison of these measure values reveals that the Pro group in the vaccine and Ukraine datasets demonstrates a stronger inclination to interact with users who have different degrees. The assortativity values are − 0.09 and − 0.29 for the vaccine dataset and − 0.15 and − 0.30 for the Ukraine dataset, representing the anti and pro groups, respectively. Conversely, the Climate dataset exhibits the opposite trend. In this dataset, the deniers group shows a greater interest in establishing relationships with users who possess dissimilar degree values.

In terms of network closure, as mentioned earlier, it can be assessed through the reciprocity measure. It is important to note that a reciprocity value of 1 indicates a purely bidirectional network, while a value of 0 signifies a purely unidirectional network. The values of this metric in all three datasets indicate a more cohesive relationship within the deniers group. Specifically, for the deniers group, the reciprocity

Table 3 Network measures in terms of segmentation

		Dataset					
		Vaccine		Ukraine war		Climate change	
Measure		Anti	Pro	Anti	Pro	Deniers	Believers
1	Transitivity (local)	0.002	0.0008	0.006	0.0009	0.020	0.0002
2	Transitivity (global)	0.005	0.002	0.013	0.002	0.010	0.0005
3	Cliques	2.4	2.4	2.88	2.18	2.72	2
4	Modularity	0.01	0.12	0.02	0.06	0.10	0.15
5	SCC	0.01	0.04	0.02	0.04	0.007	0.001
6	WCC	0.0001	0.0007	0.0006	0.0005	0.0001	0.000008
7	Count components	29	28	7	36	18	152

values are 0.0025, 0.005, and 0.0048 in the vaccine, Ukraine, and climate datasets, respectively. On the other hand, for the believers group, the values are 0.0012, 0.001, and 0.000 for the corresponding datasets. Although the overall magnitude of this metric is relatively small, the difference between the two groups is statistically significant across all three datasets.

The number of motifs, which indicate repeating patterns in the network, is significantly higher for the deniers group compared to the believers group in at least two datasets. It is worth noting that we set the number of nodes in the 'count_motifs' function to 4. In fact, motifs serve as the building blocks of a network [46] and demonstrate that as the number of motifs increases for specific numbers of nodes, the number of subnets among those nodes also increases. Specifically, the deniers group exhibits 1083 motifs in the climate dataset and 2092 motifs in the Ukraine dataset, while the believers group shows 20 motifs in the climate dataset and 1058 motifs in the Ukraine dataset. Interestingly, in the vaccine dataset, the number of motifs is higher for the pro group (4690 motifs) than the anti group (725 motifs).

Another significant measure that indicates the cohesion of nodes based on the number of connections within a subgraph is density. This metric also reveals a significant difference between the two groups in the climate and Ukraine datasets, where the deniers group exhibits higher density values. In contrast, both groups demonstrate the same density value in the vaccine dataset. The measurement of this metric demonstrates that the deniers group possesses more connections compared to the believers group.

Distribution analysis

Closeness centrality is one of the significant measures that highlights the distribution within a network. As we have observed from the definition of closeness, a higher closeness value between two nodes signifies a shorter distance in terms of the shortest path. The test results of this metric demonstrate a similar pattern in both the deniers and believers groups across the three datasets, thus supporting our hypothesis. Specifically, for the deniers group, the closeness values in the climate change, Ukraine, and vaccine datasets are 0.06, 0.07, and 0.044 respectively. Conversely, for the believers group, the corresponding values are 0.008, 0.06, and 0.015 respectively. These results indicate that, on average, the distance between nodes is shorter within the deniers group, implying a higher level of cohesion among them.

Another metric that is closely tied to the length of the path is the diameter. In general, a lower value for this attribute indicates a shorter distance between two nodes. The test results reveal that the deniers group has a smaller diameter, indicating shorter distances between their nodes compared to the believers group. This further highlights the close relationship among the members of the deniers group.

We also examined and tested the distribution measure concerning the activity of the two groups. In terms of user activity, we initially analyzed the distribution from two perspectives: the duration of activity and the time of joining for users from different groups. We evaluated the skewness of activity duration and the time of joining the network in the two groups under study. It is important to note that positive

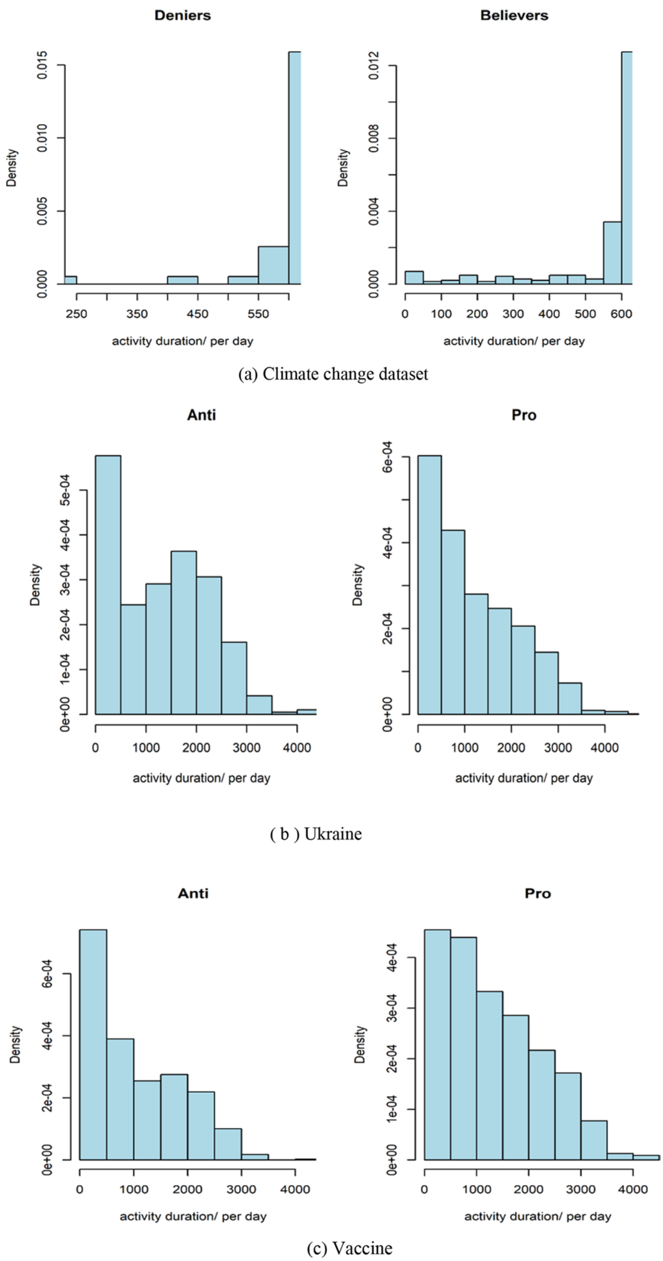


Fig. 1 Skewness of user activity. **a** In climate change, **b** Ukraine, **c** vaccine datasets

skewness indicates that the majority of interactions occur at the beginning of the time period, while negative skewness suggests that most interactions take place towards the end of the time period. Furthermore, as [47] states, if the skewness value falls between -0.5 and 0.5, the distribution of data is considered approximately symmetrical. Therefore, this study considers the distribution to be abnormal (asymmetrical) if the skew value exceeds |0.5|.

Figure 1 and Table 2 present the skewness of user activity duration, which refers to the length of time each user remains active in the network. We calculated this measure based on the first and last connection times. In both the Vaccine and Ukraine datasets, the skewness values for both the anti and pro groups are positive. Specifically, the skewness values are 0.63, 0.52, 0.30, and 0.61 for the anti and pro groups of the Vaccine and Ukraine datasets, respectively. These values indicate that users have not been active for the entire duration of the network, with the majority being active for less than 1000 days. Conversely, in the climate dataset, this metric exhibits a negative skewness, indicating that users have remained active until the last days of the available data. However, the difference in this metric between the two groups is minimal.

As shown in Fig. 2 and Table 2, the skewness values for the joining time of the two groups in Vaccine and Ukraine are negative. However, these values are negligible, as indicated in Table 2. Specifically, for the anti and pro groups of the Ukraine dataset, the skewness values are − 0.22 and − 0.53, while for the vaccine dataset, they are − 0.60 and − 0.56 respectively. Nonetheless, we can deduce that a significant portion of network joining in both datasets occurred in recent years. This trend is particularly prominent in the vaccine dataset, likely due to the COVID-19 pandemic's occurrence in recent years. On the other hand, it is evident that the majority of users joined the climate change network in its early days, with a minimal difference of 3.09 for deniers and 3.31 for believers.

Segmentation analysis

In this section, we test important metrics in the network that can greatly contribute to strengthening our hypothesis. The first metric is the clustering coefficient, which allows us to highlight the stronger relationships among nodes in the deniers ('Anti') group. According to the *igraph* documentation [48], two types of clustering coefficients can be examined in each network: local and global. The local clustering coefficient calculates the ratio of the number of triangles connected to a vertex to the number of triples centered on that vertex, while the global clustering coefficient calculates the ratio of the number of triangles to the number of connected triples in the entire graph. Table 3 presents the experimental results in terms of segmentation measures.

The results for both local and global clustering coefficients indicate that the deniers ('Anti') group has higher values compared to the believers ('Pro') group. In network science, this measure is used to demonstrate cycles in the network. This suggests that the network structure of denier users resembles a graph, while the structure of the believers group is more tree-like.

Specifically, for the believers group, the local clustering coefficient values for the Climate, Ukraine, and Vaccine networks are 0.0002, 0.0009, and 0.0008, respectively. Additionally, the global clustering coefficient values for the believers

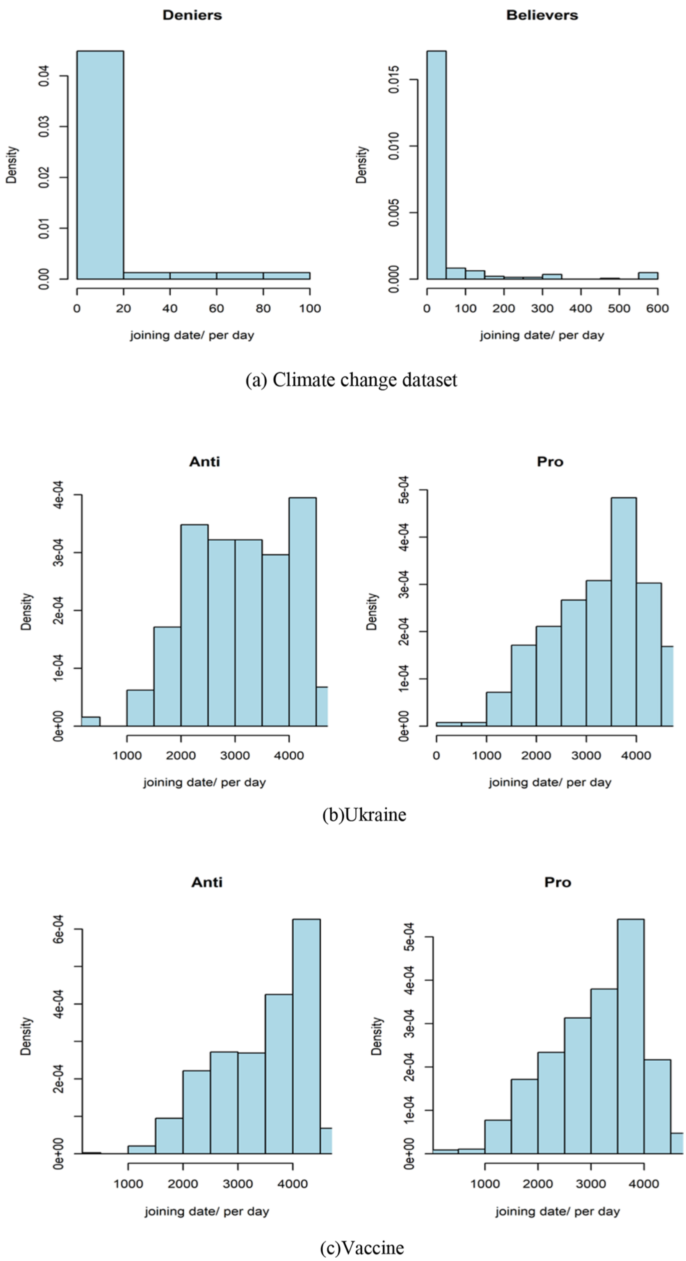


Fig. 2 Skewness plot of Joining the network. **a** Climate change, **b** Ukraine, **c** vaccine datasets

group in the Climate, Ukraine, and Vaccine networks are 0.0005, 0.002, and 0.002, respectively.

Another measure that highlights a strong and interconnected structure in the network is the clique. Testing this metric on the Climate and Ukraine networks further confirms the cohesive communication within the deniers group, as evidenced by the presence of multiple cliques. In the climate network, the average number of cliques for the deniers group is 2.72, while for the believers group it is 2. In the Ukraine network, these values are 2.88 for the deniers group and 2.18 for the believers group. In the vaccine network, both groups exhibit an average of 2.44 cliques.

Another intriguing observation we made while evaluating the measures pertains to the structure of communities within the two target groups. In community detection studies, achieving the highest modularity value is desirable, as it indicates well-structured communities. A well-structured community is characterized by a higher number of connections within the community compared to outside of it. In the datasets we examined, the modularity value for the deniers ('Anti') group is lower than that of the believers ('Pro') group. This suggests the presence of overlapping communities within the deniers group, as opposed to the believers group. This indicates a propensity among members of deniers communities to establish connections with members of other communities, thereby demonstrating stronger interconnections within the deniers group.

Specifically, for the Climate network, the modularity values for the deniers and believers groups are 0.10 and 0.15, respectively. In the Ukraine network, these values are 0.02 for anti and 0.06 for pro. Lastly, in the Vaccine network, the modularity values are 0.01 for anti and 0.12 for pro.

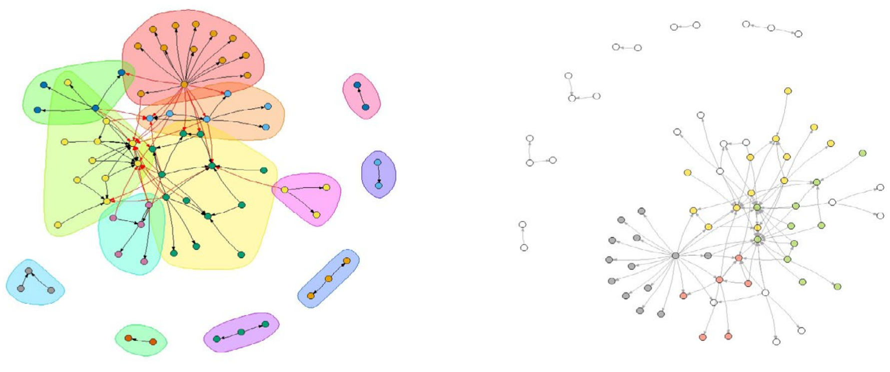
We have visualized the communities of the first 100 edges in each network, sorted based on the weight value between each pair of nodes. Figure 3 displays the communities in the vaccine network. It is worth noting that the modularity value for the believers ('Pro') group in this network is 0.52, while for the deniers ('Anti') group it is 0.36. The image further illustrates that the subnetworks of the believers group exhibit a higher degree of modularity, whereas there is overlapping within the communities of the deniers group.

Furthermore, Fig. 4 illustrates the community structure for the vaccine network. In this network, the presence of overlap and the lower modularity value within the 'Anti' group signify a higher degree of cohesion among its members.

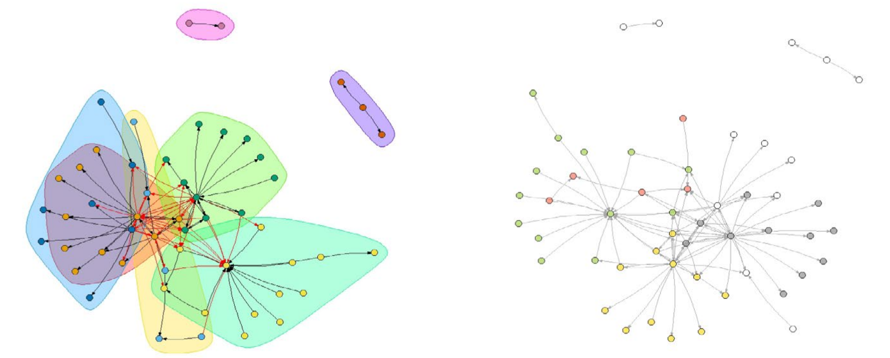
The culmination of this well-defined segregation is evident in the climate network. Figure 5 displays the majority of connections for the believers group, which is 0.83, compared to the deniers group, which is 0.51.

Furthermore, our findings indicate that the structure of the deniers (anti) network exhibits a more graph-like nature compared to the believers (pro) network. Although the overall network follows a sparse tree structure, the deniers network displays a greater level of graph-like connectivity. Figures 6, 7, and 8 depict the network structure of 100 edges with the highest weight in three datasets. It is evident that the network of believers (pro) is more segmented into distinct components, while the network of deniers (anti) users showcases a higher level of interconnectivity.

This observation is supported by the number of components in each group, as shown in Table 3. In the climate dataset, the deniers group has significantly fewer components compared to the believers group, with 18 and 152 components respectively. Similarly, for the anti and pro groups in the Ukraine dataset, there



(a) Pro group (Q=0.52)



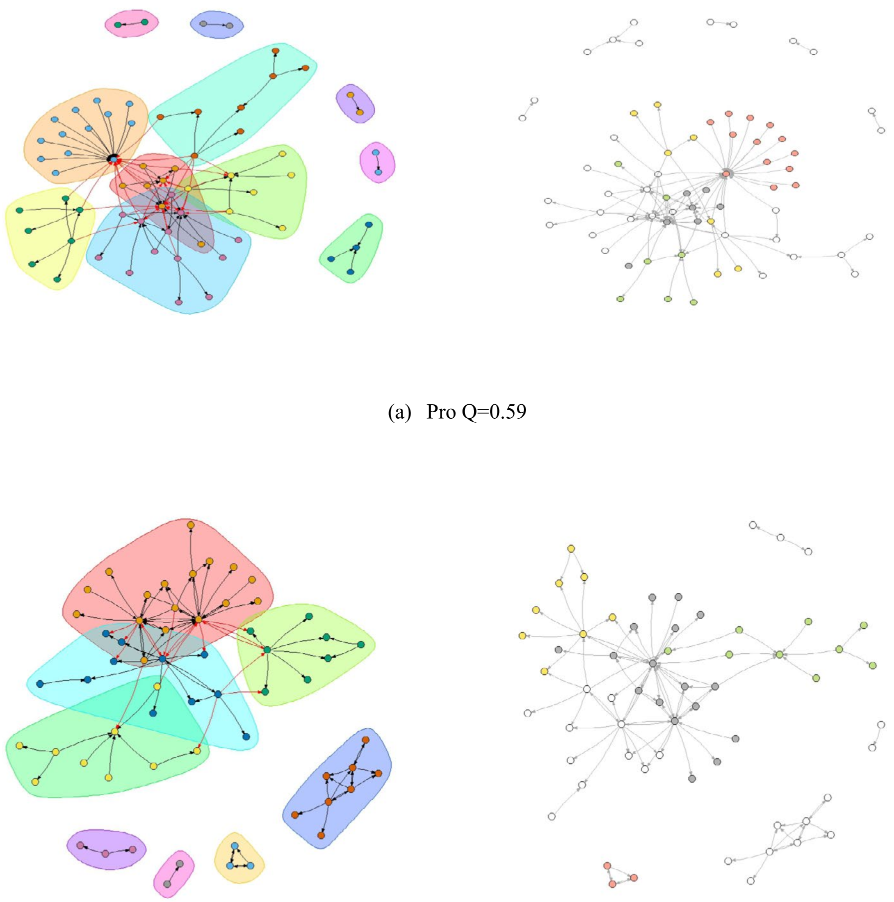
(b) Anti group (Q=0.36)

Fig. 3 The community structure and modularity value for 100 connections in the Ukraine dataset, based on the highest connection weight. Subfigure. **a** Represents the communities for the 'Pro' group, while subfigure, **b** depicts the communities for the 'Anti' group

are 7 and 36 components respectively. However, there is no significant difference in the number of components between the two groups in the Vaccine dataset.

Figure 6a and b depict the network structure of believers and deniers in the climate dataset, respectively. It is evident that the network of believers is highly segregated, with multiple disconnected components, while the deniers group forms a single component in Fig. 6. Specifically, the believers group exhibits 15 components, showcasing a higher degree of fragmentation in comparison.

Figure 7a and b illustrate the network structure of pro and anti in the Ukraine dataset, respectively. It is noticeable that the deniers group shows a smaller



(a) Pro Q=0.59

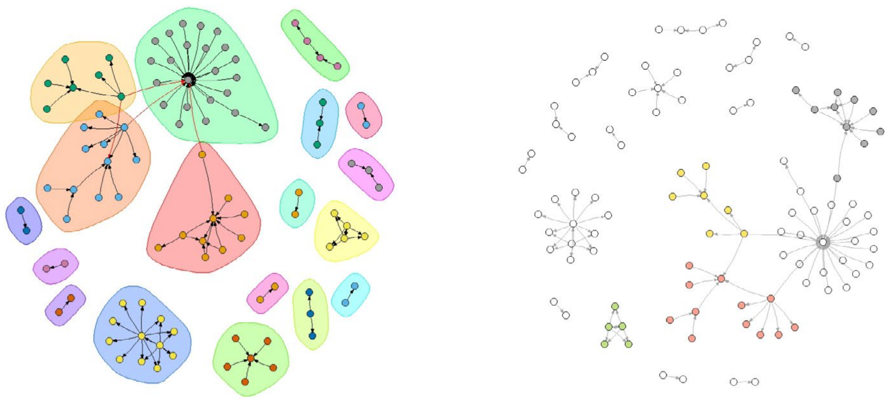
(b) Anti Q=0.48

Fig. 4 The community structure and modularity value for 100 connections in the Vaccine dataset, based on the highest connection weight. Subfigure. **a** Represents the communities for the 'Pro' group, while subfigure. **b** depicts the communities for the 'Anti' group

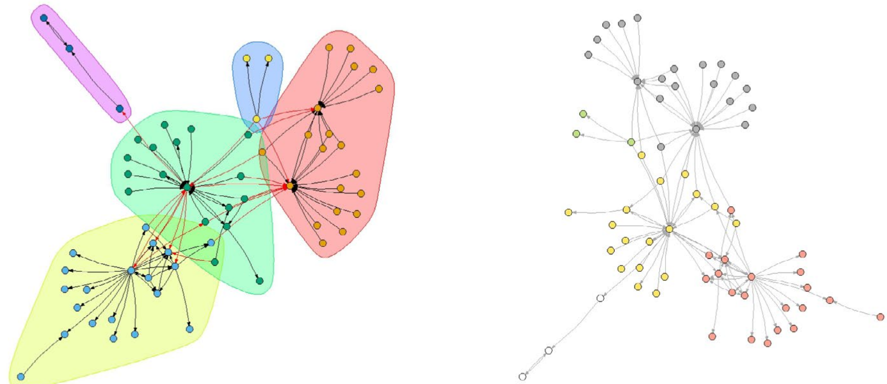
number of components compared to the believers group, indicating a higher level of cohesion within the deniers group.

Figure 8a and b display the network structure of pro and anti in the vaccine dataset, respectively. Similar to the previous datasets, there is a clear segregation observed within the Pro group, with multiple disconnected components. This further supports the notion of distinct subgroups within the Pro group in the vaccine dataset.

In addition, the pseudo p-value indicates that the difference in transitivity is statistically significant and unlikely to be due to random chance, which supports your hypothesis that the Deniers consistently have higher transitivity.



(a) believer Q=0.83

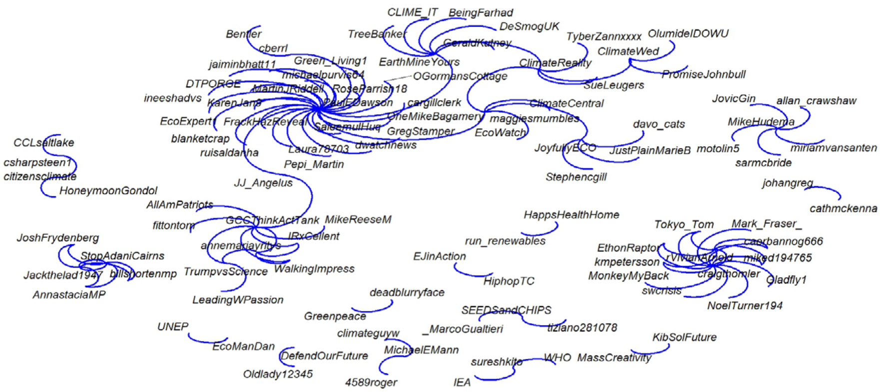


(b) Deniers Q=0.51

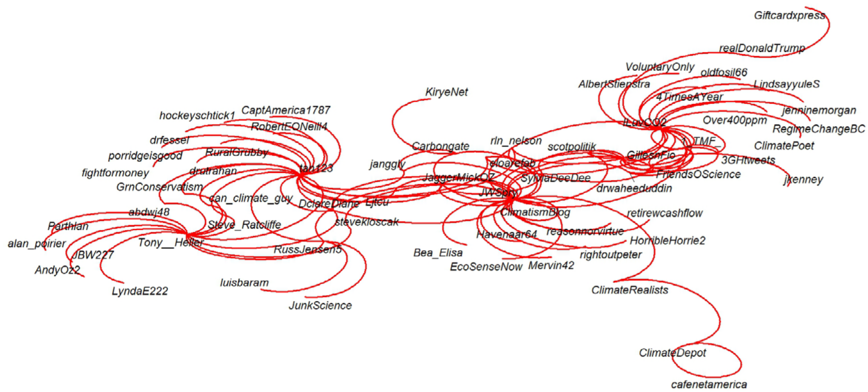
Fig. 5 The community structure and modularity value for 100 connections in the climate dataset, based on the highest connection weight. Subfigure (**a**) represents the communities for the believers group, while subfigure (**b**) depicts the communities for the deniers group

Pseudo *p*-value

To demonstrate the robustness of our results, we computed the pseudo *p*-values for various metrics, including local and global transitivity, the number of cliques, modularity, density, motifs, assortativity, and reciprocity. These pseudo p-values help



(a)Believer

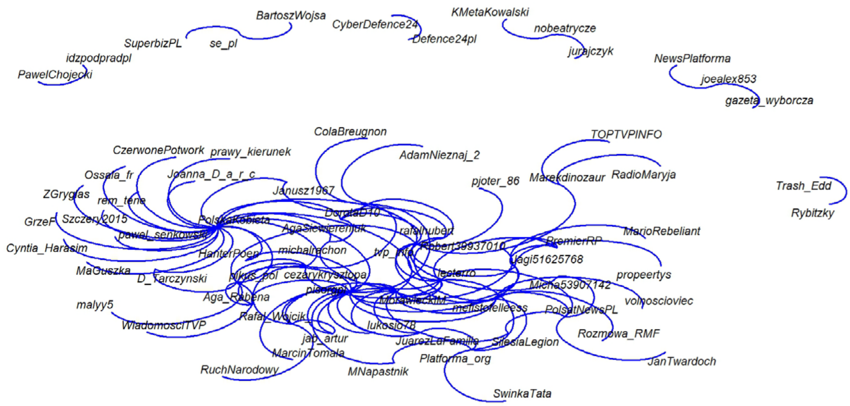


(b) deniers

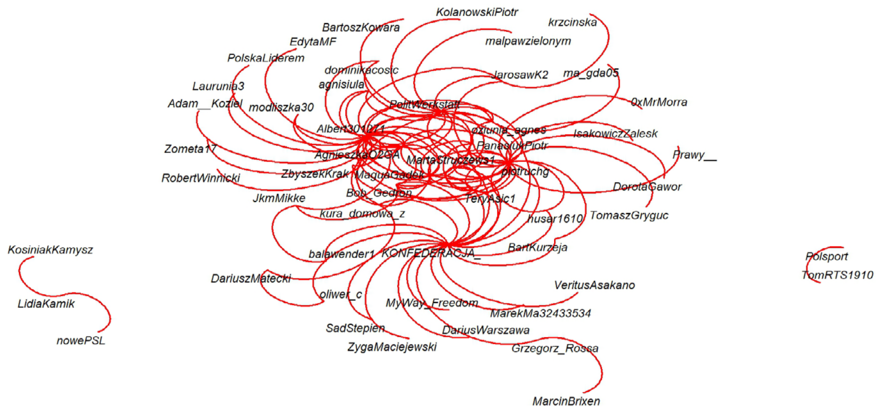
Fig. 6 Network structure of climate dataset. **a** Believers, **b** deniers

determine the statistical significance of observed differences between the two groups in each dataset. A low pseudo p-value indicates that the observed differences are unlikely to be due to random chance.

Table 4 presents the pseudo p-values for various network metrics across the three datasets. In summary, the results show extremely low pseudo p-values for both local and global transitivity, density, motifs, and reciprocity across all datasets, indicating highly statistically significant differences between the groups. The deniers' networks consistently exhibit higher transitivity and density than the believers' networks. Modularity shows statistically significant differences, particularly in the Climate Change dataset, though with less certainty compared to other metrics. Assortativity has relatively higher pseudo p-values, especially in the Ukraine War and Vaccine datasets, suggesting less pronounced differences between the groups for this metric.



(a) pro

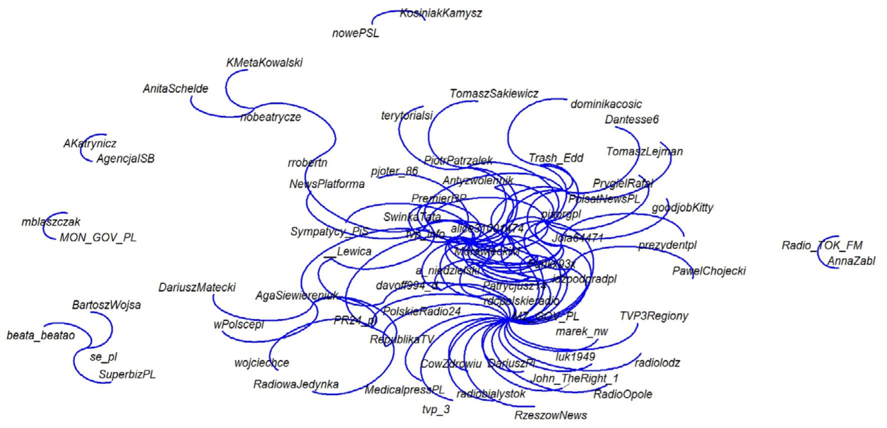


(b) anti

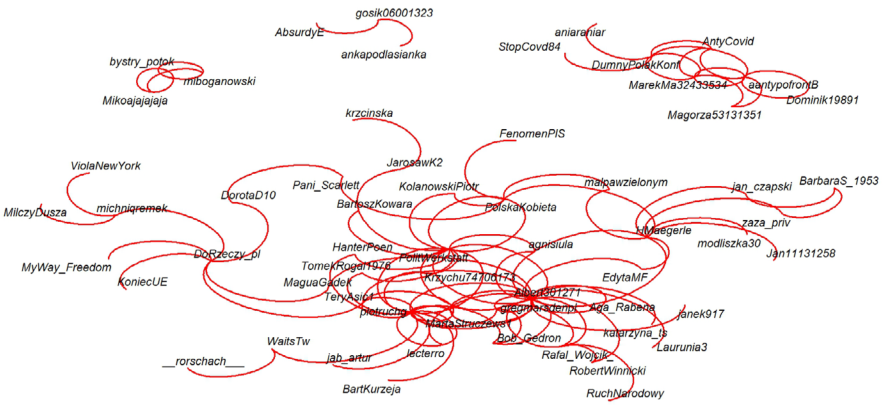
Fig. 7 Network structure of Ukraine dataset, **a** Pro, **b** anti

Implication

As this study identifies features based on network metrics, in this section, we aim to expand our finding by developing a machine learning model capable of predicting polarized groups using these network metrics. First, to gain insight into how these metrics differ between the two groups and their potential impact on model development, we employ pairplot visualization. The pairplot visualizes the relationships between various network metrics across two groups: believers or pro (in blue) and deniers or anti (in red). The metrics analyzed include density, closeness, transitivity (local and global), number of cliques, number of components, and assortativity. Figures 9, 10, and 11 present these plots for the climate change, war, and vaccine datasets, respectively. As evident from these figures, there is a



(a) pro



(b) anti

Fig. 8 Network structure of Vaccine dataset, **a** Pro, **b** anti

clear separation between the two groups. The consistent patterns observed in the pairplots support the possibility of the model achieving near-perfect metrics (precision, recall, F1-score), as reported later. The pairplot provides valuable insights into the relationships between network metrics for the two polarized groups. The observed patterns suggest that several features, particularly density, closeness, transitivity, and the number of cliques, are highly informative and should be central to the model development process. These features should be prioritized during model training, as they appear to carry significant information for distinguishing between the two groups. This visual analysis reinforces the expectation of high model performance, as indicated by the near-perfect metrics achieved in later evaluations.

Table 4 Pseudo P-values for network metrics across three datasets (climate change, Ukraine war, and vaccine)

	Transitivity (local)	Transitivity (global)	Cliques	Modularity	Density	Motif	Assortativity	Reciprocity
Climate change	0.00009	0.00009	0.00009	0.004	0.0009	0.0009	0.19	0.0009
Ukraine war	0.0009	0.0009	0.0009	0.1	0.0009	0.0009	0.61	0.0009
Vaccine	0.0009	0.0009	0.0009	0.09	0.0009	0.0009	0.55	0.0009

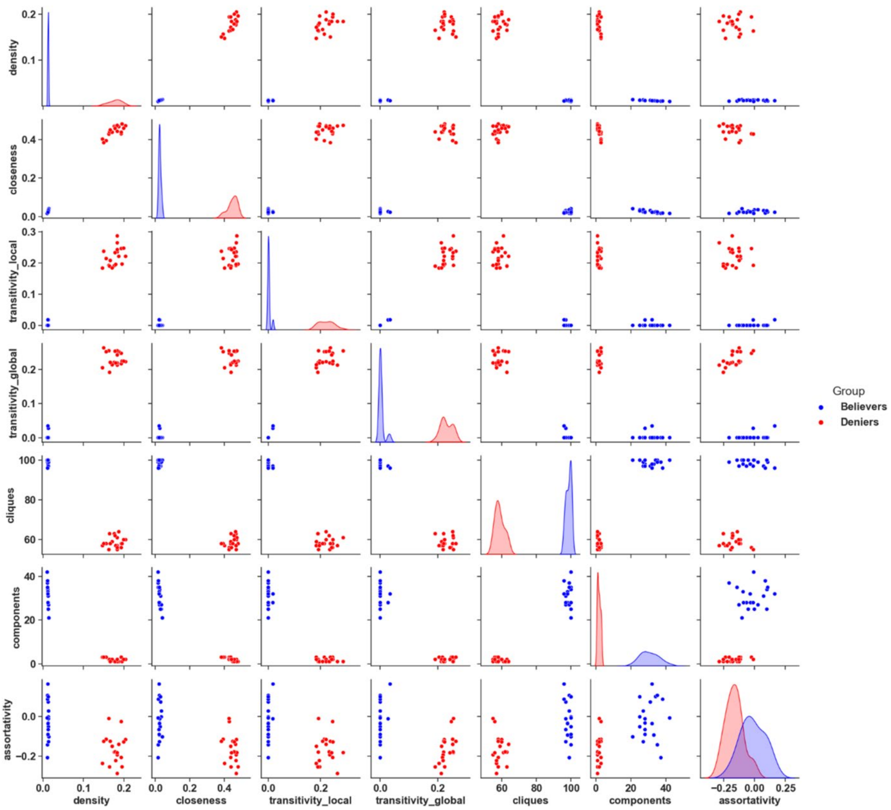


Fig. 9 Pairplot of network metrics for the climate change dataset, comparing believers and deniers groups. The plot shows distinct separations between the two groups across various network metrics such as density, closeness, transitivity, and cliques. The deniers group generally exhibits higher values in density, closeness, and transitivity, indicating a more tightly connected and clustered network structure compared to the believers group

After demonstrating how these features in the pairplot can effectively separate the two groups, we now propose our machine learning model, which uses these features to predict polarized groups. In our approach to predicting whether a subset of network data belongs to the “anti” or “pro” group, we utilize a combination of network metrics and machine learning techniques. The model leverages a RandomForest classifier to perform this classification task. This process involves several critical steps that contribute to the development of an effective predictive model. To do this, we start by dividing the dataset into two distinct groups based on their affiliations: “pro” and “anti”. This classification is derived from user status information where each user’s status is either “pro” or “anti”. We then further partition these groups into training, validation, and test subsets. This separation ensures that the model is trained on one portion of the data and evaluated on another, which helps in assessing its generalization capability. For each subset, we compute various aforementioned network metrics. These metrics are crucial as they capture the structural and

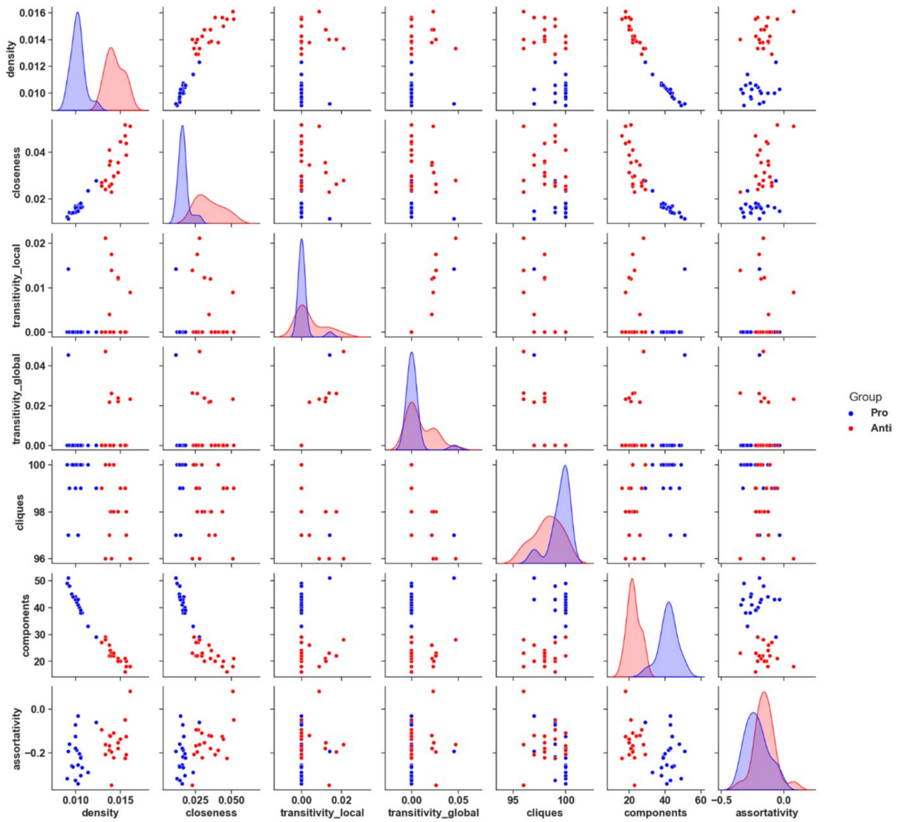


Fig. 10 Pairplot of network metrics for the Ukraine war dataset, highlighting differences between pro and anti groups. The anti group tends to have higher values in metrics like density, closeness, and transitivity, suggesting a more cohesive and clustered network. The pro group shows a more dispersed structure with lower density and fewer cliques, indicating a less centralized network

connectivity characteristics of the network. To find representative values for these metrics, we analyze subsets of the training data. By evaluating multiple random subsets, we derive approximate ranges and distributions of these metrics for both "pro" and "anti" groups.

Using the metrics computed from the training subsets, we train a machine learning model. This model learns to distinguish between the "pro" and "anti" groups based on the network metrics. To ensure robustness and avoid overfitting, we perform cross-validation during the training phase. Additionally, a separate hold-out validation set is used to evaluate the model's performance, ensuring that it generalizes well to unseen data.

After training, the model is tested on a separate test set that was not used during training or validation. This final test evaluates how well the model can predict the group classification for new, unseen subsets of data. We assess the model's performance using precision, recall, and F1-score. These metrics provide a comprehensive

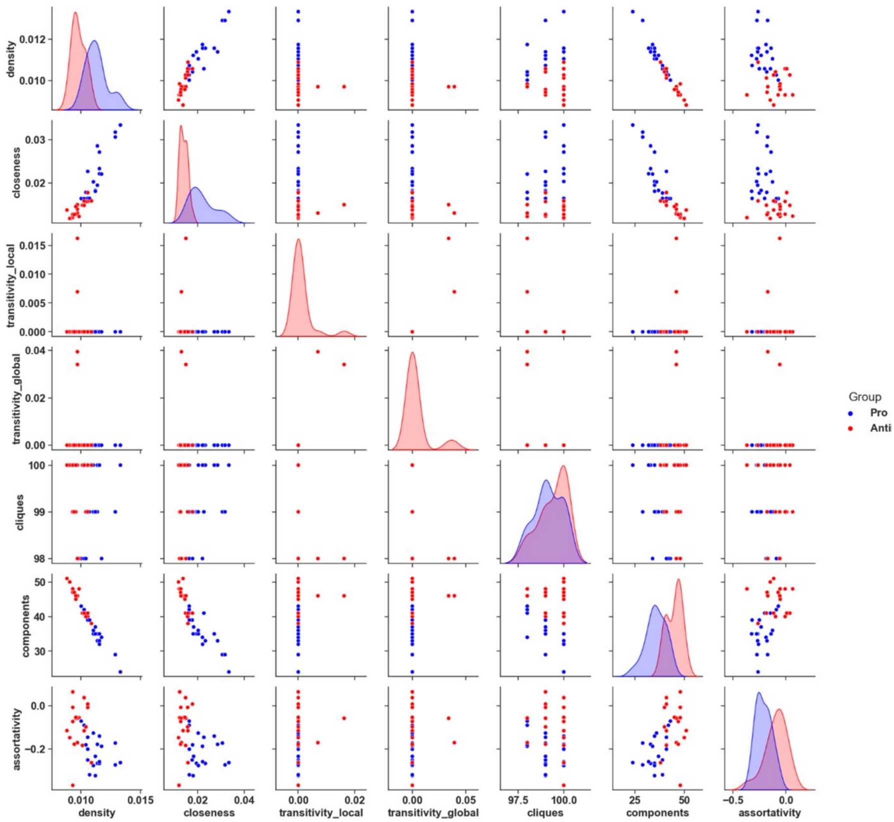


Fig. 11 Pairplot of network metrics for the vaccine dataset, showing the distribution of key metrics for pro and anti groups. The anti group displays higher density, closeness, and transitivity, similar to the other datasets, suggesting stronger internal connections and clustering. The pro group is characterized by lower metric values, reflecting a more loosely connected network structure

view of the model's predictive capability and help in understanding its effectiveness in distinguishing between the "pro" and "anti" subsets.

By using a range of network metrics, our model gains a nuanced understanding of the structural characteristics of different groups. This leads to more accurate predictions as the model can leverage multiple facets of network data to differentiate between "pro" and "anti" subsets.

The results presented in Table 5 demonstrate the model's strong performance across the climate change, war, and vaccine datasets. The model consistently achieves high precision, recall, and F1-scores during cross-validation, with values generally in the mid-90s range. For the climate change dataset, the model attains a precision of 0.96, recall of 0.95, and F1-score of 0.95. These metrics indicate that the model is highly effective at correctly identifying the believers and deniers groups within the climate change dataset, with minimal false positives and false negatives. In the war dataset, the model maintains its strong performance, achieving a

Table 5 Cross-validation performance metrics across different datasets

Dataset	Precision	Recall	F1-Score
Climate change	0.96	0.95	0.95
War	0.94	0.93	0.93
Vaccine	0.96	0.95	0.95

precision of 0.94, recall of 0.93, and F1-score of 0.93. This consistency across all metrics highlights the model's robustness in capturing the specific patterns within war-related data. The results for the vaccine dataset are equally impressive, with a precision of 0.96, recall of 0.95, and F1-score of 0.95. These results suggest that the model remains highly effective in distinguishing between pro and anti groups in vaccine-related discussions.

Given these metrics, we can conclude that the model has effectively learned the underlying patterns in the data, enabling it to accurately classify group affiliations across different contexts (climate change, war, and vaccine). The consistency between the cross-validation, validation, and test results further reinforces the model's robustness and generalization capability, suggesting that it is not overfitting and should perform well on new, unseen data.

Conclusion

This study aimed to examine the network structure of two active groups in social networks using 15 well-known metrics applied to three major networks. Based on the findings and experimental results, it can be concluded that the deniers group exhibits a more coherent network structure compared to the believers group. Various metrics, such as network reciprocity, density, diameter, closeness, transitivity (local and global), cliques, modularity, and number of components, consistently indicate a higher level of coherence and connectivity in the deniers (anti) group. Measures such as cliques, transitivity, number of components, and density reveal that the deniers (anti) group is more coherent and relatively well-connected. Additionally, the diameter, SCC (strongly connected components), WCC (weakly connected components), and closeness metrics indicate that users within the deniers group are closer to each other, suggesting the presence of cycles in their network. These cycles contribute to the amplification of tweets within echo chambers. On the other hand, the believers group exhibits a tree-like structure without cycles, and the longer path lengths between users in this group may lead to reduced impact and dissemination of their tweets within echo chambers, compared to the deniers group. The depiction of the number of components in Figs. 6, 7, and 8 further illustrates the degree of segregation within the believers group, emphasizing the stronger connectivity within the deniers group. Although the network structure for both groups is not purely tree-like, the higher density value and lower number of components in the deniers (anti) group suggest a more graph-like structure compared to the believers (pro) group.

These findings have implications for feature selection in machine learning models and can provide valuable insights for further research in this domain.

Funding This work was supported by Narodowe Centrum Nauki (National Science Centre, Poland) under Grant 2020/38/A/HS6/00066.

Data availability statement The data and codes used and/or analyzed during the current study available from the corresponding author on request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose. The authors have no conflicts of interest to declare that are relevant to the content of this article.

Ethics approval The authors declare that there is no any human, animals or live human cells involved in the study.

References

1. Germani, F., & Biller-Andorno, N. (2021). The anti-vaccination infodemic on social media: A behavioral analysis. *PLoS ONE, 16*, e0247642.
2. Jemielniak, D., & Krempovych, Y. (2021). An analysis of AstraZeneca COVID-19 vaccine misinformation and fear mongering on twitter. *Public Health, 200*, 4–6.
3. Neff, T., Kaiser, J., Pasquetto, I., Jemielniak, D., Dimitrakopoulou, D., Grayson, S., Gyenes, N., Ricaurte, P., Ruiz-Soler, J., & Zhang, A. (2021). Vaccine hesitancy in online spaces: A scoping review of the research literature, 2000–2020. *Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-82
4. Khaldarova, I., & Pantti, M. (2016). *Fake news* (pp. 891–901). Journalism Practice.
5. McIntyre, L. (2022). *How to talk to a science denier: conversations with flat Earthers, climate deniers, and others who defy reason*. MIT Press.
6. Marques, M. D., Kerr, J. R., Williams, M. N., Ling, M., & McLennan, J. (2021). Associations between conspiracism and the rejection of scientific innovations. *Public understanding of science, 30*, 854–867.
7. McDevitt, M., Parks, P., Stalker, J., Lerner, K., Benn, J., & Hwang, T. (2018). Anti-intellectualism among US students in journalism and mass communication: A cultural perspective. *Journalism, 19*, 782–799.
8. Okruszek, Ł, Piejka, A., Banasik-Jemielniak, N., & Jemielniak, D. (2022). Climate change, vaccines, GMO: The N400 effect as a marker of attitudes toward scientific issues. *PLoS ONE, 17*, e0273346.
9. Walter, S., Brüggemann, M., & Engesser, S. (2018). Echo chambers of denial: explaining user comments on climate change. *Environmental Communication, 12*, 204–217.
10. van Eck, C. W., Mulder, B. C., & van der Linden, S. (2021). Echo chamber effects in the climate change blogosphere. *Environmental Communication, 15*, 145–152.
11. Erlich, A., & Garner, C. (2023). Is pro-Kremlin disinformation effective? Evidence from Ukraine. *The International Journal of Press/Politics, 28*, 5–28.
12. Moy, Wesley R., Kacper Gradon, & Russian Covid- Effects. (2023). COVID -19 effects and Russian disinformation campaigns. https://www.hsaj.org/resources/uploads/2020/12/hsaj_Covid192020_COVID19EffectsRussianDisinformationCampaigns.pdf. Accessed June 26.
13. Nowak, B. M., Miedziarek, C., Pełczyński, S., & Rzymski, P. (2021). Misinformation, fears and adherence to preventive measures during the early phase of COVID-19 pandemic: A cross-sectional study in Poland. *International journal of environmental research and public health*. https://doi.org/10.3390/ijerph182212266

14. Jiang, B., Mansooreh K., Lu C., Black T., & Liu H. (2021). Mechanisms and attributes of echo chambers in social media.
15. Alatawi, F., Cheng L., Tahir A., Karami M., Jiang B., Black T., & Liu H. (2021). A Survey on echo chambers on social media: description, detection and mitigation.
16. Chay, S., & Sasaki, N. (2011). Using online tools to assess public responses to climate change mitigation policies in Japan. *Future Internet, 3*, 117–129.
17. Auer, M. R., Zhang, Y., & Lee, P. (2014). The potential of microblogs for the study of public perceptions of climate change. *Wiley interdisciplinary reviews: Climate change, 5*, 291–296.
18. Pearce, W., Holmberg, K., Hellsten, I., & Nerlich, B. (2014). Climate change on Twitter: Topics, communities and conversations about the 2013 IPCC Working Group 1 report. *PLoS ONE, 9*, e94785.
19. Williams, H. T. P., McMurray, J. R., Kurz, T., & Hugo Lambert, F. (2015). Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global environmental change: Human and policy dimensions, 32*, 126–138.
20. Holme, P., & Rocha J. C. (2021). Networks of climate change: Connecting causes and consequences. arXiv [physics.soc-ph]. arXiv.
21. Baumann, F., Lorenz-Spreen, P., Sokolov, I. M., & Starnini, M. (2020). Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters, 124*, 048301.
22. Tyagi, A., Babcock M., Carley K. M., & Sicker D. C. (2020). Polarizing tweets on climate change. In *Social, cultural, and behavioral modeling*, (pp. 107–117) Springer International Publishing.
23. Neff, T., & Jemielniak, D. (2022). How do transnational public spheres emerge? Comparing news and social media networks during the Madrid climate talks. *New Media & Society, 26*, 2066–2091.
24. Cody, E. M., Reagan, A. J., Mitchell, L., Dodds, P. S., & Danforth, C. M. (2015). Climate change sentiment on twitter: An unsolicited public opinion poll. *PLoS ONE, 10*, e0136092.
25. Cinelli, M., De Francisci Morales G., Galeazzi A., Quattrociocchi W. & Starnini M. (2021). The echo chamber effect on social media. In *Proceedings of the national academy of sciences of the United States of America*, 118. https://doi.org/10.1073/pnas.2023301118
26. Jasny, L., & Fisher, D. R. (2019). Echo chambers in climate science. *Environmental Research Communications, 1*, 101003.
27. Samantray, A., & Pin, P. (2019). Credibility of climate change denial in social media. *Palgrave Communications, 5*, 1–8.
28. Littman, J., & Wrubel, L. (2019). Climate change tweets Ids. Harvard Dataverse. https://doi.org/10.7910/DVN/5QCCUU
29. Documenting the Now. (2020). Hydrator (version 0.0.11).
30. Van Rossum, G., & Drake F. L. (2009). Python 3 reference manual createSpace.
31. Peixoto, Tiago P. (2014). The graph-tool python library. Figshare.
32. Bastian, M., Heymann S., Jacomy M. (2009). Gephi: an open source software for exploring and manipulating networks. In *Third international AAAI conference on weblogs and social media.*
33. Berahmand, K., Bouyer, A., & Samadi, N. (2019). A new local and multidimensional ranking measure to detect spreaders in social networks. *Computing, 101*, 1711–1733.
34. Salavati, C., Abdollahpouri, A., & Manbari, Z. (2019). Ranking nodes in complex networks based on local structure and improving closeness centrality. *Neurocomputing, 336*, 36–45.
35. Rui, X., Meng, F., Wang, Z., & Yuan, G. (2019). A reversed node ranking approach for influence maximization in social networks. *Applied Intelligence, 49*, 2684–2698.
36. Wen, T., & Deng, Y. (2020). Identification of influencers in complex networks by local information dimensionality. *Information sciences, 512*, 549–562.
37. Chen, Z. (2019). An agent-based model for information diffusion over online social networks. *Papers in Applied Geography, 5*, 77–97.
38. Liu, L., Bo, Qu., Chen, B., Hanjalic, A., & Wang, H. (2018). Modelling of information diffusion on social networks with applications to WeChat. *Physica A: Statistical Mechanics and its Applications, 496*, 318–329.
39. Wilson, C., Sala, A., Puttaswamy, K. P. N., & Zhao, B. Y. (2012). Beyond social graphs: User interactions in online Social networks and their implications. *ACM Transactions on the Web*. https://doi.org/10.1145/2382616.2382620
40. Mislove, Alan, Massimiliano Marcon, Krishna P. Gummadi, Peter Druschel, and Bobby Bhattacharjee. 2007. Measurement and analysis of online social networks. In *Proceedings of the ACM SIGCOMM Internet Measurement Conference*, (pp. 29–42) IMC.

41. Newman, M. E. J., & Girvan, M. (2003). Finding and evaluating community structure in networks. *Physics Review, 69*, 1–16.
42. Allcott, H., Karlan, D., Möbius, M. M., Rosenblat, T. S., & Szeidl, A. (2007). Community size and network closure. *The American economic review, 97*, 80–85.
43. R: Measures of network closure. (2023). https://search.r-project.org/CRAN/refmans/migraph/html/closure.html. Accessed June 20.
44. Himelboim, I. (2017). Social network analysis (Social Media). In J. Matthes, C. S. Davis, & R. F. Potter (Eds.), *The International Encyclopedia of Communication Research Methods* (pp. 1–15). John Wiley & Sons. https://doi.org/10.1002/9781118901731.iecrm0236
45. Santoro, N., Quattrociocchi W., Flocchini P., Casteigts A., & Amblard F. (2011). Time-varying graphs and social network analysis: Temporal indicators and metrics. In *AISB 2011: Social networks and multiagent systems*, (pp. 33–38).
46. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science, 298*, 824–827.
47. Bulmer, M. G. (1979). Principles of Statistics. Courier Corporation.
48. igraph R manual pages. (2023). https://igraph.org/r/html/1.3.4/transitivity.html. Accessed June 24.