PLOS COMPLEX SYSTEMS

RESEARCH ARTICLE

Trends and topics: Characterizing echo chambers' topological stability and in-group attitudes

Erica Cau[1,2☯], **Virginia Morini**[1,2☯], **Giulio Rossetti**[2*]

1 Computer Science Department, University of Pisa, Pisa, Italy, **2** Institute of Information Science and Technologies "A. Faedo" (ISTI), National Research Council (CNR), Pisa, Italy

☯ These authors contributed equally to this work.
* giulio.rossetti@isti.cnr.it

Citation: Cau E, Morini V, Rossetti G (2024) Trends and topics: Characterizing echo chambers' topological stability and in-group attitudes. PLOS Complex Syst 1(2): e0000008. https://doi.org/10.1371/journal.pcsy.0000008

Editor: Fabiana Zollo, Universita Ca' Foscari, ITALY

Received: January 20, 2024

Accepted: July 16, 2024

Published: October 3, 2024

Data Availability Statement: Data and code are available at the following GitHub link https://github.com/ericacau/Trends-Topics_case_study.

Funding: This work is supported by: the EU NextGenerationEU programme under the funding schemes PNRR-PE-AI FAIR (Future Artificial Intelligence Research) (to CE and RG); the EU – Horizon 2020 Program under the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities" (G.A. n.871042) "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu) (to RG and MV); PNRR-

Abstract

Nowadays, online debates focusing on a wide spectrum of topics are often characterized by clashes of polarized communities, each fiercely supporting a specific stance. Such debates are sometimes fueled by the presence of echo chambers, insulated systems whose users' opinions are exacerbated due to the effect of repetition and by the active exclusion of opposite views. This paper offers a framework to explore how echo chambers evolve through time, considering their users' interaction patterns and the content/attitude they convey while addressing specific controversial issues. The framework is then tested on three Reddit case studies focused on sociopolitical issues (gun control, American politics, and minority discrimination) during the first two years and a half of Donald Trump's presidency and on an X/Twitter dataset involving BLM discussion tied to the EURO 2020 football championship. Analytical results unveil that polarized users will likely keep their affiliation to echo chambers in time. Moreover, we observed that the attitudes conveyed by Reddit users who joined risky epistemic enclaves are characterized by a slight inclination toward a more negative or neutral attitude when discussing particularly sensitive issues (e.g., fascism, school shootings, or police violence) while X/Twitter ones often tend to express more positive feelings w.r.t. those involved into less polarized communities.

Author summary

Since their introduction, Social Networks have revolutionized human interactions, allowing for instantaneous interactions with others. Despite their advantages, social networks have introduced many drawbacks. In this paper, we focus on echo chambers, polarized environments where like-minded people interact with each other, actively excluding people with dissenting opinions, thus insulating them from rebuttal. Understanding this phenomenon is highly relevant, as echo chambers characterize many real-world events, such as election interference. Currently, few works focus on detecting the temporal evolution of echo chambers, and even less focus on their evolution over a long temporal window. In addition, many existing works are often structured as case studies centered on one event

"SoBigData.it - Strengthening the Italian RI for Social Mining and Big Data Analytics" - Prot. IR0000013 (to RG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

in a specific social network, so their proposed methodology is hardly reusable. We introduce a framework and apply it to two different case studies to track the temporal evolution of echo chambers through time and examine the related discussions of users. The framework relies on the network structure of social relations and Natural Language Processing algorithms to analyze the discussed topics and the underlying tendencies of users in their utterances inside and outside echo chambers. The framework was applied to data from Reddit, collected over the first two and a half years of Donald Trump's presidency, and from X/Twitter during a debate related to Black Lives Matter (BLM) coinciding with a sports event. These case studies revealed interesting insights about the persistence of echo chambers over time and users' attitude while discussing sensitive topics within these closed systems.

1 Introduction

The emergence of Online Social Network sites (OSNs) in the *information age* [1] reshaped several aspects of our everyday lives. Those platforms have made the exchange of information and opinions almost instantaneous—increasingly accelerating those spreading processes that happen at a slower pace in the offline world. Such a change of paradigm that introduces unprecedented social opportunities also led to novel issues. An example lies in the *information overload* [2] to which users are exposed when accessing online spaces. The massive amount of conflicting information found online may lead users to experience a mental discomfort—often called *cognitive dissonance* [3]. Consequently, to avoid this discomfort, online users are more prone to filter and consume only pieces of information confirming their beliefs and ideas—a pattern often supported and reinforced by recommendation systems [4, 5].

However, despite the importance of opinion heterogeneity for creating meaningful debates, OSNs represent a perfect breeding ground for human and algorithmic biases that may interfere with safe dialectic processes and knowledge formation [6–8]. The rise and, most of all, the exacerbation of these biases contribute to creating pollution in online spaces [9, 10]. Furthermore, this issue has grown in importance because of the loss of an evident boundary between the online and offline world, thus resulting in potentially harmful consequences of online behaviors that might resonate in the offline world [11, 12]. Among the most studied pollutants, opinion polarization has raised several concerns due to the features offered by OSNs, as they tend to exacerbate the ideological positions [13] of users by allowing for easier connections with people with the same interests and exposing them to content aligned with their thoughts.

This paper focuses on identifying and analyzing those meso-scale topologies that—with a different degree of likelihood, or *risk*—can be considered *echo chamber* (henceforth, EC): polarized environments where like-minded people interact, actively excluding people with dissenting opinions, thus insulating them from rebuttal. Although there is no agreement on how to identify and quantify echo chambers in online social contexts [14, 15], their presence has been observed across multiple domains and topics. Recent studies that analyze political debates often argue that ECs are major players in spreading misinformation [16], promoting pseudoscience narrations, and exacerbating political discourse. However, although often referred to as pollutant realities, it is important to underline that ECs are not *per se* signs of users' epistemic vices [17]. They are, conversely *communities of practice* [18–21], epistemic structures that, once established a tacit knowledge [22–24], support their members by enforcing an in/out-group dialectic which is functional to "protect" a community from external interferences—often reflected in out-group abusive communication patterns [25]. Indeed, such insulation

might lead to feed confirmation bias—and, in turn, increasing polarization. However, it is not given that their presence is a sufficient condition to define a given social context polluted [17]. Existing works to date have analyzed the (allegedly) effects of ECs and assessed their presence in online spaces (underlying that the users participating in them are often a small fraction of the online population [26]), but often from a modelistic perspective or through case studies that generally do not consider their temporal unfolding.

The contributions of this paper are threefold:

- It formalizes a platform-independent framework for investigating those social dynamics of ECs that have often been ignored in the literature;
- It defines a methodology to investigate the topics and the attitude of users inside and outside context at risk of acting as ECs;
- It enriches the body of work on ECs dynamics by offering case studies on Reddit and X/Twitter sociopolitical discussions—respectively, during the first two years and a half of Donald Trump's presidency and a sportive event.

In this paper, potential ECs are detected, tracked through time, and analyzed by considering both the social interactions and the content produced by online users—focusing on capturing the emotional component of EC users' discussions. This study extends the framework described in [27]—targeting the identification of EC—by integrating (i) a focus on ECs dynamics and (ii) a pipeline to characterize the contents discussed (along with their emotional valence) by users inside/outside ECs.

The paper is organized as follows. In the Related works, we introduce and discuss the literature on ECs, focusing on their detection. Subsequently, the Framework section describes the proposed framework for tracking and analyzing ECs dynamics over time, emphasizing both the relations and topics of discussion. The framework is tested on OSN data extracted from Reddit and X/Twitter and the results obtained are discussed. The Conclusion section summarizes the main findings of this project and discusses weaknesses and future developments.

2 Related works

This section discusses the recent literature impacting our research, focusing on the three main topics: echo chamber detection, dynamic community discovery, and natural language processing.

Echo chamber detection

As the concept of ECs is widely discussed, much debate exists on how these polarized systems create and develop. This information is necessary to allow their identification and, subsequently, their mitigation. In recent decades, an ever-growing body of research has focused on quantifying the extent to which discussions are polarized [28–31] and, consequently, are deemed a fertile ground for polluted phenomena. Traces of ECs have been found in forums [32], blogs [33], and, generally, in those online spaces employing recommendation systems, such as e-commerce platforms [34].

Traditional EC detection methods can be coarsely grouped into two different approach families. The former focuses on the textual content shared by users (i.e., the *echo* dimension), which is used as a proxy to capture the debates echoing among users sharing the same ideology. For example, in [35], the authors attempted to understand whether users were exposed to crosscutting content by investigating the news articles they shared on Facebook. Likewise, in [36], over 10 million U.S. Facebook users publicly sharing their political leaning in their profile

information were classified into two categories depending on whether they discussed/shared more or less polarized news. In [37], the authors defined approaches to categorize user-generated content shared within ECs—estimating the content stance on a given topic and determining the emotion conveyed. Certainly, one of the main issues of content-based approaches lies in the need for high-quality data annotations—a requisite often addressed through unsupervised Natural Language Processing pipelines, which does not ensure the correctness of the labels.

The latter family of methods leverages the network of interactions designed by users while debating in online forums. In this scenario, the *chamber* dimension is investigated, which opens up the reverberation of the opinion. Conversely, from the contents-based approaches, the analysis focuses on modeling the users' social graph to inspect their relations at various semantic levels (i.e., *retweet network*, *comment network*), eventually with the possibility of enhancing the analysis through the mining of additional information from the text—such as in [38], where the authors estimated the users' political leaning from their shared contents and then proceeded with the construction of the interaction network.

Network-based approaches can be further grouped based on the *topological* scales at which the ECs are detected, i.e., *macro-scale*, *meso-scale*, and *micro-scale*. Macro-scale studies focus on the interaction networks on an aggregate level to identify two well-distinguished clusters of users with opposite leaning in the network. For example, in [39], two ideologically contrapposed communities were identified solely using the HITS algorithm [40]. Similarly, in [41], the authors reconstructed and analyzed the interaction network between Donald Trump and Hillary Clinton supporters. Meso-scale studies, instead, focus on studying the organization of network nodes into clusters, usually by leveraging a community detection algorithm, to detect echo-chamber-like structures composed of nodes sharing a common opinion/ideology. An example of a hybrid meso-scale and content-based approach was described in [42], where the authors explored the presence of ECs in tweets about COVID-19 by constructing the interaction network and by applying a community detection algorithm (METIS [43]), which allowed to partition the network into two distinct communities. The communities were then evaluated according to traditional community evaluation fitness functions and controversy measures. Ultimately, the micro-scale analysis focuses on investigating the leaning of individual users and their relations to the one adopted by the members of their 1-hop neighborhood—such as in [44], where the authors leverage homophily to assess the presence of ECs, moved by the idea that users surrounded by people with a similar leaning are consequently exposed to similar content(s), thus increasing the likelihood of ECs formation.

A common limitation of network-based studies is their focus on detecting ECs by seldom characterizing them by their topology without considering the content they vehiculate and their emotional valence. Moreover, such studies usually rely on platform-specific features, thus making them difficult to replicate in slightly different scenarios. Finally, both families of approaches often neglect another invisible yet fundamental component of every complex system: time. Temporal un-awareness is often mitigated by designing case studies with a relatively short timespan or, following a completely different perspective, by leveraging what-if simulating (e.g., relying on *opinion dynamics* modeling [10]). Regrettably, such flattened representations, keeping together interactions potentially distant in time and disregarding their temporal ordering, often describe a complex phenomenon simplistically—introducing the risk of overestimating users' sociality and failing to capture the dynamics behind online debates.

(Dynamic) Community Discovery

Studies focusing on meso-scale ECs usually rely on community detection (henceforth, CD) algorithms to identify homogeneous clusters of users sharing common features. Since there is

no common agreement on what a *community* should represent, several algorithms have been proposed to identify communities [45] and, most importantly, to track their temporal unfolding [46] in dynamic settings. Focusing on dynamic communities identification and characterization adds a layer of complexity to an already *ill-posed* problem: in this scenario, topological perturbations—e.g., nodes and edges appearing/vanishing—can reverberate at the mesoscale level, thus introducing instability and clusters' events [47–49]. While a few case studies employing CD algorithms in the context of polarized information systems detection [28] exist, there is a scarcity of works employing dynamic community detection (DCD) to define EC identification and tracking frameworks. Among the exceptions is the work by Kopacheva *et al.* [50], DCD is leveraged to analyze the evolution of users' communities on Twitter revolving around the refugee crisis in Sweden in 2015.

Since users trapped within an EC are also characterized by a shared opinion/tacit knowledge, it is useful to account for such additional semantics while approaching the CD task. This can be accomplished by modeling the interaction network as a *node-attributed graph*—where each node is associated with an attribute expressing its leaning on a given topic—and using such semantic augmented representation to extract annotated communities [51]. This study focuses on designing a pipeline to work on interaction networks modeled as snapshot series graphs. To such an extent, we leverage a *Labeled Community Detection* method [52] able to extract—from each temporal snapshot—a set of communities balancing topological quality and label homogeneity. Indeed, the selected method is only one of the possible candidates that can be selected for the specific task; therefore, we can consider it as a parameter of the proposed approach.

Topic modeling and Valence analysis

Moved by the intention to analyze ECs in more depth once they have been identified, in this section, we briefly present the state of the art of the two approaches usually employed to gain insights on the subjects and tone of the discussions unfolding in online platforms: topic modeling and valence analysis.

Topic modeling. The former approach, namely *topic modeling*, is related to the extraction of the most relevant topics covered in a textual *corpus*, i.e., from a collection of documents. This task has been widely employed in the literature, even in research fields not immediately related to linguistics, e.g., bioinformatics [53] or in computer vision [54]. Many approaches and algorithms have been formalized and implemented to address this task. Among these approaches, one of the first and most well-known is Latent Dirichlet Allocation (LDA) [55], a probabilistic model that assumes that a statistical process generates each document. Under such an assumption, each document is characterized by its distribution of topics, and consequently, each topic can be characterized by the probability of specific words appearing in it.

Several evolutions of LDA have been proposed, such as LDA2vec [56], which incorporates Word2Vec [57] into the LDA model. Moreover, topic modeling algorithms have also been devised to capture the dynamics of topics in documents over time, i.e., DTM [58] and TTM [59]. As an effect of the trend of research on deep learning models, *Transformers* [60] have revolutionized how documents are represented as vectors. In 2018, Bidirectional Encoder Representations from Transformers (BERT) [61] was released, followed shortly by more specialized/better-performing ones [62–64]. These models are well known for being trained over a massive amount of data and for implementing the so-called *attention* mechanism [65], which builds for each word a particular and contextual word embedding while accounting for both the words to its right and its left (thus, being defined as a *bidirectional* model). This revolution in

language representation models has led to their massive application in several NLP tasks, e.g., text generation, classification tasks, Named-Entities Recognition, and Topic modeling.

Valence analysis. Valence has often been investigated in Natural Language Processing, psychology, and cognitive sciences as one of the factors defining the meaning of words. In such contexts, *Valence* is often paired with two other meaning-related dimensions: *Arousal* and *Dominance*. According to [66], the values of such measures might be employed to characterize a text through the primary emotions it conveys. Typically, measures quantifying Valence, Arousal, and Dominance are extracted through manually annotated datasets, such as in the case of ANEW [67] and its extension by Warriner *et al.* [68]. Another dataset is the VAD lexicon [69], which consists of around 20.000 English words manually annotated. For each word in an input text, VAD values summarize in the range [0, 1]—namely, the negativeness and positiveness—the emotional value (expressed in valence, arousal, dominance) it possesses.

3 Echo chambers diachronic analysis

In this section, we formalize a framework to track and analyze the dynamics of those mesoscale topologies at risk of acting as EC while addressing the content-related aspect of the debate described by the data. The first step of the pipeline is based on the platform-agnostic framework from Morini *et al.* [27], to which we add two additional steps that handle the temporal analysis and the content characterization. The original framework that we enhance lies within approaches that investigate meso-scale topologies. Before defining the four-step framework, it is worth properly defining the network model we adopt to represent online social interactions (e.g., debates) and, as a consequence, what we consider to be at risk of being an *echo chamber* in the context of this study.

Definition 1 (Interaction Graph) *Online debates are modeled as feature-rich networks* $G = (V, E, A)$ *describing a set* V *of users interacting on a topic* η *(thus establishing edges* $(u, v) \in E$ *with* $u, v \in V$*) each having an opinion* $A(v)$ *on* η.

Edges in G can describe directed or undirected social interactions depending on the assumptions made on the online social traces used to infer the network structure. For simplicity and without loss of generality, we assume edges in G to be undirected in this work.

It is worth noticing that the framework we introduce in this work can be considered to be "platform-independent" only under the assumption that the social structure it is applied on semantically describes an interaction graph, namely a topology where edges connecting user pairs are proxies for dialogues occurring among them (i.e., post-comment, mention, direct message relations). Whenever such edge semantic is not preserved (i.e., while considering follower-followee relations or friendship ties), the framework—although technically applicable—loses its anchor to the founding assumption that explicit communications involving a specific topic are taking place. However, online social platforms are often designed to support the exchange of information among their users (although sometimes following slightly different protocols or media content types); therefore, the proposed framework is general enough to be applied to most analytical scenarios.

Leveraging such modeling and assumptions, in the rest of the paper, we will consider the following definition of Echo Chamber:

Definition 2 (Echo Chamber) *Given an interaction graph* $G = (V, E, A)$*, an echo chamber is a densely connected subset* $E_i \subseteq V$*, composed by users' sharing the same opinion* $a_i \in A$ *on* η*, that is loosely connected w.r.t. the rest of the network.*

The proposed definition incorporates proxies for both characteristics usually associated with ECs: (i) being polarized environments composed of interacting like-minded individuals and (ii) being structures that actively exclude outsiders by insulating them from rebuttal/

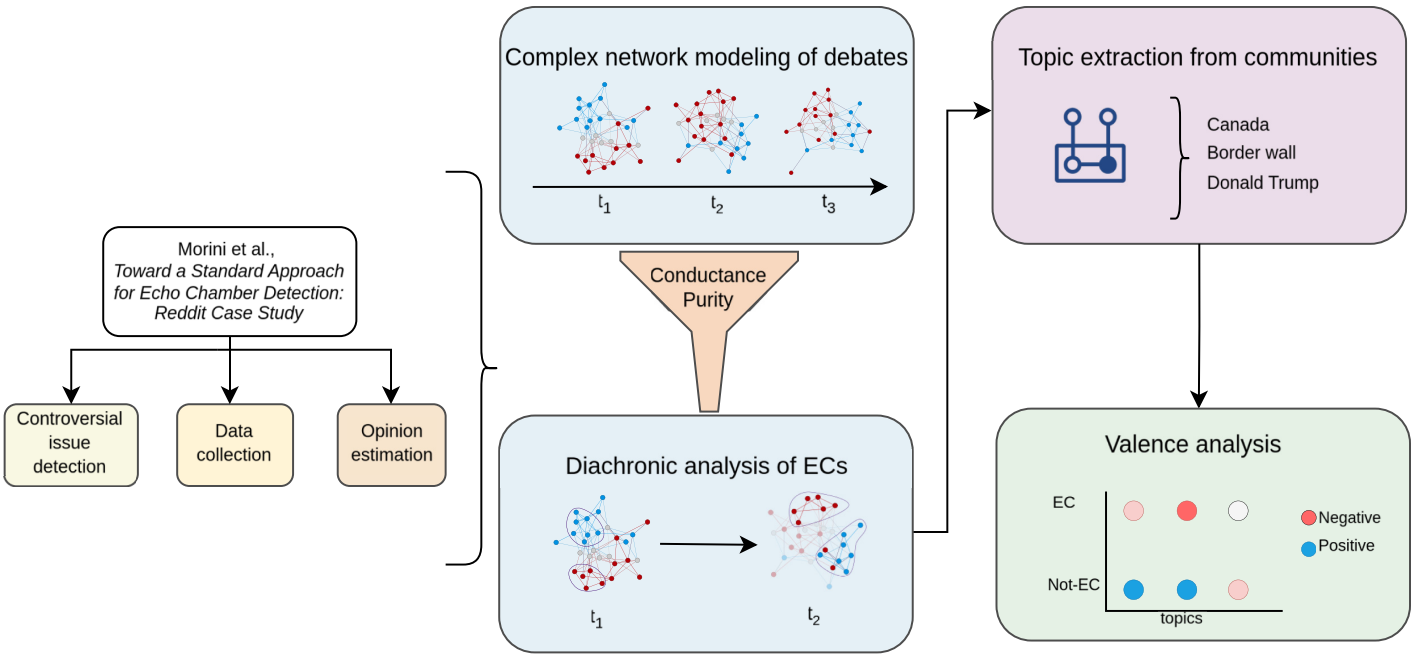

Fig 1. Pipeline visual representation. From left to right: (i-ii) data collection, annotation and network modeling; (iii) dynamic community detection and analysis; (iv) valence analysis.

https://doi.org/10.1371/journal.pcsy.0000008.g001

opposite viewpoints. The former characteristics are captured by the request that an EC is composed of a connected subset of nodes sharing the same label (e.g., identifying a common political leaning); the latter—insulation—is proxied through physical distancing, namely a reduced/absent social interaction with the *alters* (e.g., the rest of the networks).

By leveraging such definitions, our pipeline reorganizes the four analytical steps described in [27] and formalizes two additional steps. EC risk detection and topological analysis is structured as follows: (i) data collection and opinion estimation in the context of online polarized discussions (former steps i and ii of [27]); (ii) network modeling of online debates (former step iii and iv of [27]); (iii) users' group's identification (leveraging an identify-and-match DCD approach), and lifecycle analysis; (iv) topic extraction and valence analysis. Fig 1 reports a graphical representation of the proposed pipeline.

3.1 Step 1: Data collection and annotation

The original pipeline's starting point is identifying a target online debate—e.g., an online discussion spanning from politics to social and environmental issues. Once the data related to the online debate have been obtained, it is necessary to focus on the ideological characterization of users. Typically, obtaining data with a clear user's leaning toward a controversy is difficult. Hence, there is a need to define a user's classification methodology to estimate their learning about the debate according to the issue under analysis.

In [27], the task is modeled as a *text classification* problem. A carefully selected subset of raw textual data conveyed by user-generated interactions (i.e., posts and comments) are embedded into vectors, which—after a preliminary annotation phase—are then used to train a classification model specific to the context. This choice grants a higher level of generalizability while keeping the framework independent of any platform-specific feature—e.g., the number of *likes* or *retweets*, in Twitter/X or the *reactions* in Facebook. Among the various approaches

to classification, the authors propose choosing between Deep Learning models or pre-trained Transformers while considering the amount of data and the type of information to be extracted.

3.2 Step 2: Modeling online debates

One of the least considered dimensions by the echo chamber identification and analysis literature is time. Debates in online and offline arenas unfold through time, and the stances of those participating in them change as time goes by. Therefore, it is important to longitudinally analyze the interactions that lead to creating polarized communities—e.g., echo chambers—to understand the observed system dynamics better. Since ECs are epistemic systems in constant evolution, it is necessary to define an actionable way to model their topological counterpart.

We model time-evolving interaction networks as a series of attributed snapshot graphs—extending the classical model defined in [70]. Formally,

Definition 3 (Attributed Snapshot Graph)

$$\mathcal{G} = \langle G_1, G_2 ... G_t \rangle \tag{1}$$

where each snapshot $G_i = (V_i, E_i, A_i)$ *is a feature-rich graph univocally identified by the set of nodes* V_i*, edges* E_i*, and node labels* A_i*—as described in Definition 1.*

Indeed, identifying a proper snapshot timespan is a key factor to generating reasonable modeling of social interactions: if too large, it might induce information loss in terms of nodes and social dynamicity; if too short, the interaction graph might end up being unstable and noisy—therefore, making impossible to observe temporal correlation among topological and opinion processes [71]. Snapshot identification methodologies depend on both interaction semantics and the quality of the available data temporal annotations. One of the most adopted criteria while approaching such a task is to aim at supporting interpretability of the modeled phenomena either by producing fixed-frequency or fixed-width thresholding [72–74]. Therefore, the proposed framework is parametric on the snapshotting strategy to be applied, which needs to be fitted to the data available and the phenomenon to be modeled.

3.3 Step 3: Identify Groups and their dynamics

So far, we have defined our reference model for dynamic interactions as annotated graphs capturing social agents enriched by some semantic information (e.g., their stance/opinion in a debate). The next step is to describe how to handle the extraction of meso-scale structures from such a complex system leveraging Dynamic Community Detection. Coherently with [27], our choice for the CD algorithm fell on EVA [52], a Labeled Community Detection algorithm that—in our settings—we apply on each snapshot graph. Eva is designed to extract communities that balance structural cohesion and intra-community label homogeneity; therefore, it simultaneously aims at both constraints (topological and semantic) we explicated in our EC definition. The algorithm extends the Louvain algorithm [75] to node-attributed graphs. On the one hand, it maximizes Newman's modularity and, on the other, a measure defined in its paper, known as *Purity* (in our experiments we set EVA parameter α to 0.5, aiming to weight equally the contributions of both measures in the greedy optimization schema). Moreover, as empirically shown in [52], the multi-criteria optimization performed by EVA alleviate the well-known resolution limit [76] that affects modularity optimization approaches like Louvain. It should be noted, however, that one of the limits of EVA—inherited by Louvain—is being single-resolution.

The two quality functions optimized by EVA are defined as follows:

Definition 4 (Modularity) *Modularity quantifies the observed number of edges inside the given partition minus the expected number of edges if they were distributed following a null model of a random graph. The modularity has values ranging in* $\left[-\frac{1}{2}, 1\right]$. *It is formalized as follows*:

$$Q = \frac{1}{2m}\sum_{vw}\left[A_{uw} - \frac{k_v k_w}{(2m)}\right]\delta(c_v, c_w) \tag{2}$$

where: $m = |E|$ *is the number of edges in* G; $A_{uw} = 1$ *if the edge* $(u, w) \in E$, 0 *otherwise*; k_* *is the degree of node* $*$; $\delta(c_v, c_w)$ *is the Kronecker delta function that equals* 1 *if* c *and* w *belong to a same community* c, 0 *otherwise*.

Definition 5 (Purity) *Purity was defined in* [52], *and it is calculated as the product of the frequencies of the most frequent labels carried by its node. This function lies within the range [0, 1].*

$$P_c = \prod_{a \in A}\frac{\max\left(\sum_{v \in c} I(a, v)\right)}{|c|} \tag{3}$$

where $I(a, v)$ *is an indicator function that equals* 1 *if* $A(v) = a$, 0 *otherwise, and* c *is the community to which* v *belongs*.

EVA is applied to every snapshot graph independently (implementing a *two-step* DCD pattern [46]): after node clusterings are identified, "potential" ECs are detected—snapshot wise—by following the rationale in [27].

As suggested in [27], for each snapshot, the *risk* of acting as an EC is measured for each identified community in terms of its *conductance* and *purity*. The former measure estimates the volume of the edges exiting the community w.r.t. the total ones established by the nodes belonging to it; the latter assesses the quality of the cluster in terms of attribute homogeneity.

Definition 6 (Conductance) *The conductance of a community,* C, *is the ratio of the edges pointing out of it w.r.t. the total one incident to community nodes.* C *lies in* [0, 1] *where* 0 *identifies a cluster not connected to the rest of the graph*, 1 *a set of disconnected nodes*.

$$Conductance_c = \frac{|E_{OC}|}{2|E_C| + |E_{OC}|} \tag{4}$$

where $|E_{OC}|$ *is the number of edges exiting the community and* $|E_C|$ *is the number of edges remaining inside the community* [77].

According to these two measures, the risk of a community being an echo chamber is maximized when *conductance* is minimized—it tends to 0—and *purity* is maximized—it tends to 1. The original paper [27] considers EC communities the ones having, at the same time, *conductance* less than 0.5 and *purity* equal to or greater than 0.7.

It is worth noting that neither threshold should be considered as clear-cut but rather coarse-grained filters—for such a reason, in the forthcoming analysis we impose more stringent values for the former, setting its lower bound to 0.3 instead of 0.5.

Given a specific node cluster C_i and its conductance and purity values, a more sound interpretation we adopt relates such measurements to C_i *"risk"* of acting as an EC. In particular, the loosen is C_i connection to the rest of the network (low conductance), and the higher its purity, the more likely it is to behave as an EC. In practice, the conductance and purity thresholds are used to filter out those communities having a relatively low risk of acting as EC, while their actual values identify the individual EC-likelihood degree of the remaining mesoscale topologies. Furthermore, to reduce potential noise, we maintain only communities of relevant sizes (e.g., composed by at least 20 users in Reddit, 10 in X/Twitter).

Once the potential ECs are identified, we perform longitudinal analysis—aimed at assessing evolutive patterns—by computing the pairwise Jaccard index among communities of adjacent snapshots: an approach which has already been used in [70] to identify the most likely evolution of partitions based on similarity. Before proceeding with this step, to reduce the noise further, we preprocess the community sets by removing those users who joined online discussions by posting/commenting once. Moreover, we retain only the users each community shares with adjacent ones for each snapshot, thus focusing on "stable" sub-populations. We analyze the temporal development of ECs and of those communities that are below the EC-risk threshold using a trend line plot—each line representing the evolution of the similarity between adjacent partitions through timestamps. In addition, each line is enriched with a marker representing the *status* of the community in a specific timestamp: triangles representing communities labeled as ECs, dots communities that are not. This type of plot allows, on the one hand, to assess the stability and evolution of *individual* communities and, on the other, to observe the difference between all the ECs extracted using the approach previously described.

3.4 Step 4: Topic extraction and analysis

After assessing the stability of potential ECs (and not-ECs) over time, the proposed pipeline focuses on (i) identifying the topics discussed within them and (ii) computing the cluster-wide attitude towards such topics.

Topic modeling. To carry out topic modeling, we decided to leverage an approach based on embeddings, i.e., BERTopic [78]: a fine-tuned implementation of the BERT model designed to support topic modeling tasks. The motivations behind the choice are twofold: the guarantees offered by a transformer architecture—which allow for a better representation of words in context—and the competitive results of BERTopic w.r.t. alternative topic modeling algorithms [79]. In particular, BERTopic is robust and independent of the language model employed despite performing or not performing the fine-tuning phase. First, it generates the embedding of the input text using a language model and—to improve cluster quality—it reduces the data dimensions via UMAP [80] to avoid the curse of dimensionality [81]. Secondly, it clusters the embeddings through HDBSCAN [80], which has the feature to consider noisy topics as outliers. Finally, leveraging a class-based version of *tf-idf* extracts the most meaningful words from each identified cluster.

To evaluate the quality of the obtained topics, we rely on two measures as proxies for an indicative—and subjective—human evaluation, as highlighted in [78]: namely, topic *coherence* [78] and *diversity* [82]. The former estimates the coherence of the extracted topics by using Normalized Pointwise Mutual Information (NPMI) [83]. It spans the range [-1, 1], being 1 a perfect association with scores given by human annotators. The latter describes, for each topic, the percentage of unique words and lies within [0, 1].

Valence analysis. Our aim is also to investigate the emotional component of interactions among users within potential ECs and outside such closed systems. To address this issue, we rely on the VAD Lexicon and KeyBERT [84], a method for keyphrase extraction that exploits BERT embeddings and cosine similarity to identify the most likely keywords describing a raw text. The main idea is to extract a set of keywords describing each post/comment and then proceed by calculating the valence score of the topic they are associated with. As a first step, keywords from the cleaned texts included in each topic are extracted. Subsequently, the valence score is extracted for each keyword having a match in the VAD lexicon (e.g., 0.931 for "*travel*"—a word associated to a positive valence -, or 0.115 to "*chaotic*"—a word associated to

negative valence). The final score returned as output for each post/comment is the average valence of its matched keywords.

4 Case study: Reddit socio-political datasets

In this section, we apply the proposed framework to a specific case study, discuss the obtained results, and evaluate its effectiveness and limitations.

The dataset we focus on is introduced in [27, 85], where the authors assessed the presence of ECs, which we decided to track further from a temporal perspective. The dataset is composed of Reddit discussions about three socio-political topics. It focuses on the pro-/anti-Trump debate between January 2017 and June 2019, as it sharply exacerbated the clash between the two factions of Democrats and Republicans.

Reddit is currently the seventh most used social network in the world [86]. It is particularly suitable as a source of data since it consists of *subreddits*, topic-specific forums devoted to a single topic where users may freely discuss both general matters and more specific topics within various niches. Since the platform encourages anonymity, its users may be motivated to talk more openly and, therefore, even express extreme positions while confronting controversial discussions. The pseudononymized datasets we analyzed are available on a dedicated Github repository (Datasets: https://github.com/ericacau/Trends-Topics_case_study). Given such characteristics, Reddit has increasingly received attention from researchers focusing on debate polarization and EC analysis (e.g., [41, 85, 87, 88]), leading to a body of literature often having conflicting results due to the specific thematic focus/communities they were focusing on.

Data collection and annotation

The three analyzed datasets focus on specific socio-political issues: *gun control*, *minorities discrimination*, and *politics*. The specific subreddit used to create the thematic three datasets and all the preprocessing made are described in [27].

In [27], an additional dataset describing a *polarized population* was created—collecting posts and comments from subreddits that openly support or antagonize the Trump presidency (and that explicitly specify in their moderation sections banning strategies for users/contents that do not adhere the subreddit specific stance). That dataset, composed of balanced samples of post/comments from *r/The_Donald* and, *r/Fuckthealtright*, *r/EnoughTrumpSpam*, was employed to train and test a classification model, namely $BERT_{BASE}$ (reaching an accuracy greater than 70%), aimed at infer users' political leanings.

$BERT_{BASE}$ was then applied to the three Reddit sociopolitical datasets. Each post/content was classified as either Pro- or Anti-Trump, and the prediction confidence (ranging in [0, 1]) was used to assign a continuous value to the identified class. Posts/comments with prediction confidence equal to 1 were considered perfectly aligned with a pro-Trump ideology, while the ones on the other extreme aligned with anti-Trump ones. Individual scores were subsequently averaged at the user level to compute the *leaning score*

$$L_u = \frac{\sum_{i=1}^{n} PredictionScore(p_i)}{n}$$

where $p_i \in P_i$ corresponds to a post shared by the user u and $n = |P_i|$ indicates the cardinality of the posts the users publish. Finally, the resulting users' leaning scores values were discretized into intervals, as follows: if $L_u \leq 0.3$, then the posts are classified as *antitrump*, as *protrump* if $L_u \geq 0.7$ and *neutral* otherwise (for additional details on scores/leanings distributions refer to

Table 1. Datasets statistics. For each considered dataset, the number of subreddits, the number of posts, and the number of users included.

Dataset	# Subreddit	# Post	# User
Gun control	6	180,170	65,111
Minorities discrimination	6	223,096	52,337
Politics	6	431,930	72,399

https://doi.org/10.1371/journal.pcsy.0000008.t001

[27]). These thresholds, arbitrarily in nature, have been maintained to align our results with the ones of the original study but may be increased or decreased according to the dataset.

Table 1 reports basilar descriptive statistics of the three thematic datasets.

Network modeling and EC identification

Five temporal snapshots were extracted from the three datasets, each covering a semester. Starting from such a temporal discretization, a dynamic network was reconstructed as a snapshot sequence where a labeled user u had an edge pointing towards user v at time t, if and only if u interacted with a post/comment by user v or vice versa during semester t. Each undirected edge (u, v, t) is then enriched with the weight of that tie, equal to the number of times the interaction between u, v occurs during t. Table 2 provides an overview of the network.

After network construction, communities were extracted from each snapshot through the Labeled Community Detection. As previously discussed, the chosen algorithm was EVA since, by design, it optimizes both modularity and label homogeneity. The scatterplots in Figs 2, 3 and 4 underlines—respectively for *Gun Control*, *Minorities discrimination* and *Politics*—the presence of polarized communities (having different nuances of associated risk of behaving as ECs) in each of the temporal snapshot. In each scatter plot, the radius of the points corresponds to the size of the community. Moreover, the horizontal/vertical line identifies the cut-off threshold we set to separate potential ECs from non ECs ($Purity >= 0.7$), ($1 - Conductance >= 0.7$). Section 6.1 further discusses the impact such threshold have on the percentage of risky communities/users in each dataset.

Ideally, the most polarized echo chambers are placed in the upper right corner in such graphical representations. To align with the analysis of [27] we leverage such thresholds to separate "potential" ECs from not-ECs.

Given the temporal nature of observed phenomena, we computed the Jaccard index to identify the most similar clusters in temporally adjacent partitions—thus reconstructing the most likely evolution of a node cluster c_i from t to $t + 1$.

Table 2. Reddit network statistics. Averaged number (per semester) of nodes, edges, degree, and density of the networks, and distribution of neutral and pro-/anti-Trump nodes' leaning attributes.

	Gun Control	Minorities Discrimination	Politics
$\|V\|$	11,388	6,617	7,912
$\|E\|$	114,262	80,497	57,463
avg. degree	17.10	19.36	17.36
avg. density	0.003	0.003	0.001
Pro-Trump nodes	2,803	2,150	3,837
Anti-Trump nodes	7,385	3,676	2,923
Neutral nodes	1,199	790,6	1,151

https://doi.org/10.1371/journal.pcsy.0000008.t002

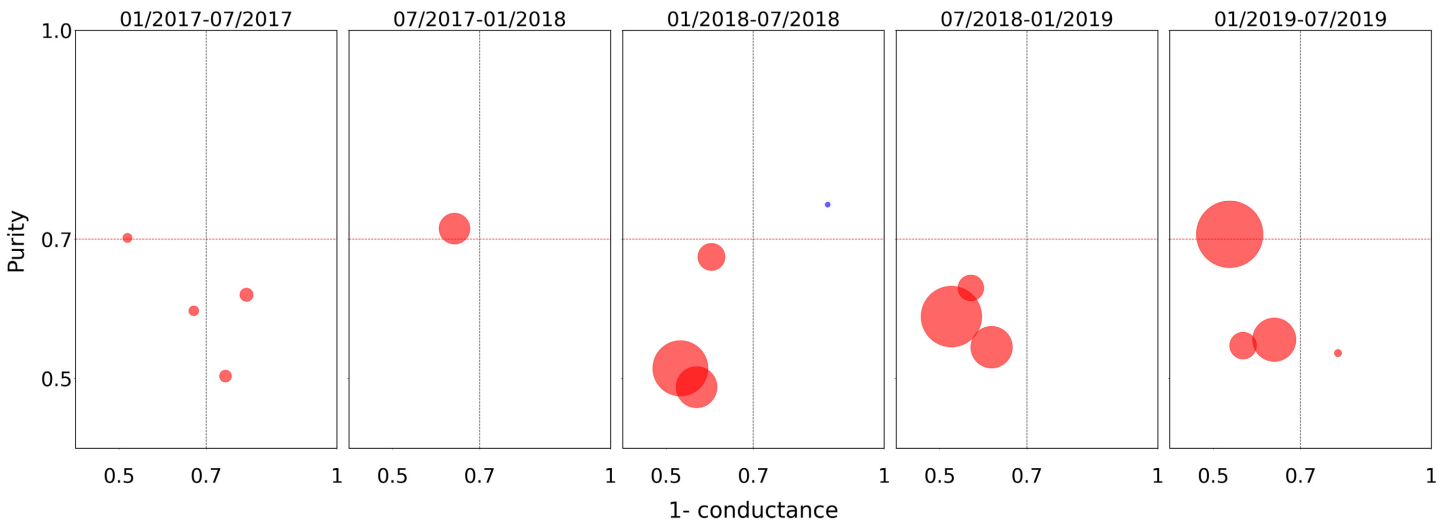

Fig 2. EC risk. Gun control. Insulation and opinion coherence of Reddit communities extracted in the five temporal snapshots. Colors identify the community's prevalent opinion (red for *Pro-Trump*, blue for *Anti-Trump* respectively). Circle sizes are proportional to the number of users. Horizontal and vertical lines identify the coarse-grained EC threshold. The higher values for the *purity* and 1 − *conductance* scores, the higher the risk for the community to act as an EC.

https://doi.org/10.1371/journal.pcsy.0000008.g002

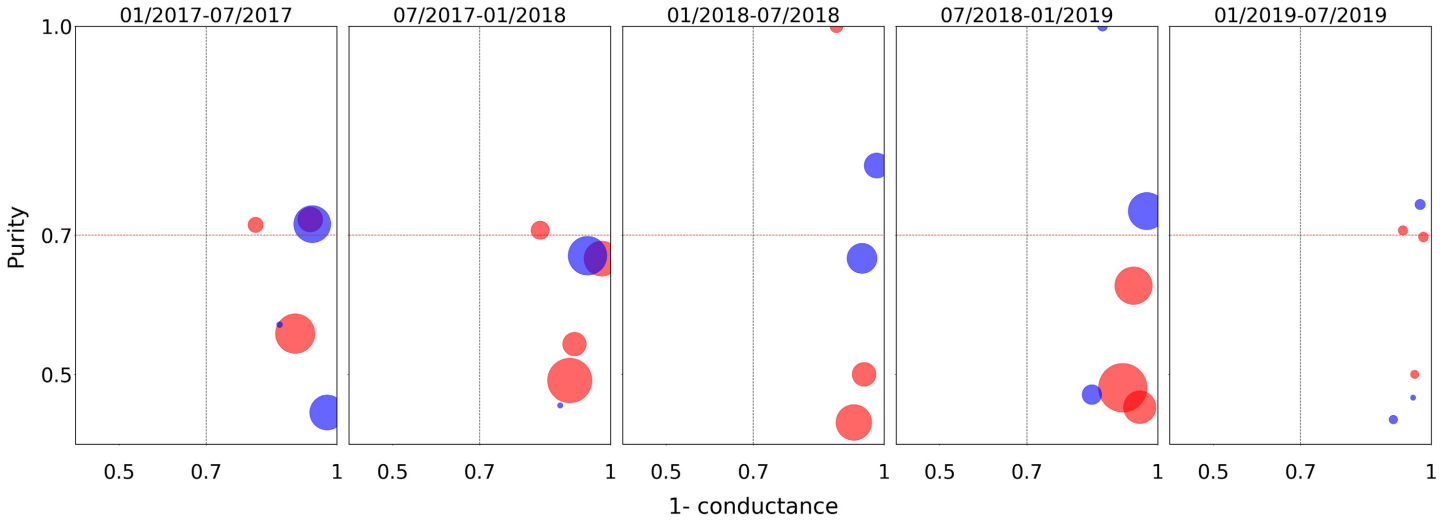

Fig 3. EC risk. Minority Discrimination. Insulation and opinion coherence of Reddit communities extracted in the five temporal snapshots. Colors identify the community's prevalent opinion (red for *Pro-Trump*, blue for *Anti-Trump* respectively). Circle sizes are proportional to the number of users. Horizontal and vertical lines identify the coarse-grained EC threshold. The higher values for the *purity* and 1 − *conductance* scores, the higher the risk for the community to act as an EC.

https://doi.org/10.1371/journal.pcsy.0000008.g003

Echo chambers' stability analysis

The analysis now moves toward understanding ECs' internal dynamics to answer two research questions.

- RQ1: *Are echo chambers stable over time w.r.t. the users that compose them?*
- RQ2: *Do echo chambers keep or lose their polarization as time passes?*

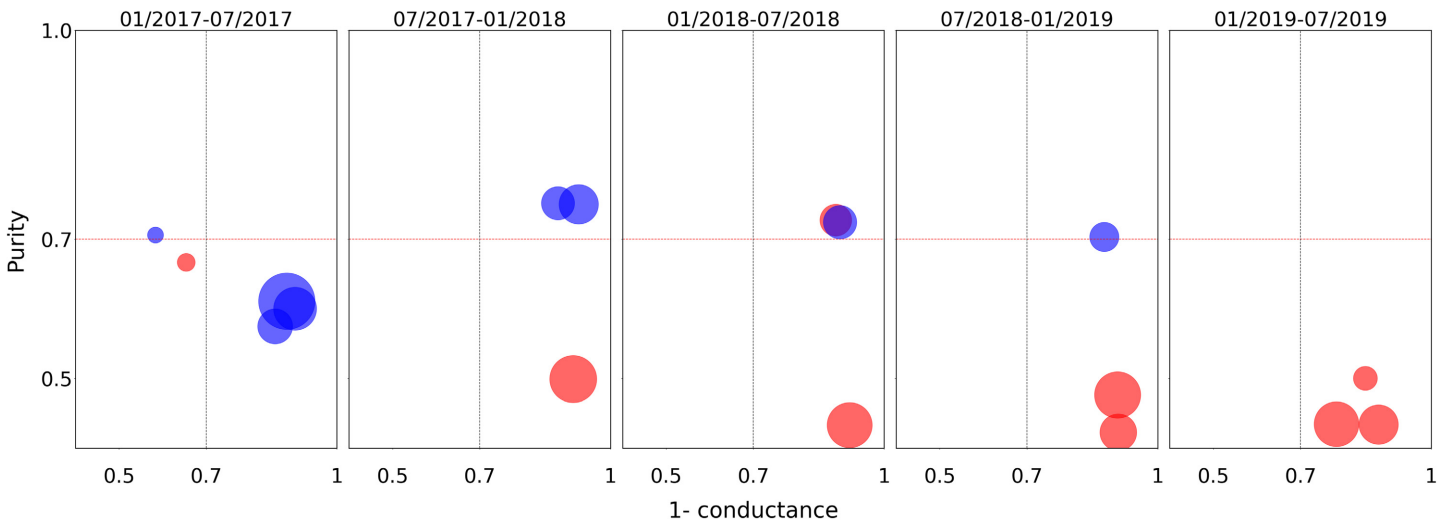

Fig 4. EC risk. Politics. Insulation and opinion coherence of Reddit communities extracted in the five temporal snapshots. Colors identify the community's prevalent opinion (red for *Pro-Trump*, blue for *Anti-Trump* respectively). Circle sizes are proportional to the number of users. Horizontal and vertical lines identify the coarse-grained EC threshold. The higher values for the *purity* and 1 − *conductance* scores, the higher the risk for the community to act as an EC.

https://doi.org/10.1371/journal.pcsy.0000008.g004

While RQ1 focuses on the stability of EC—i.e., their ability to persist even in contexts where high dynamicity of the users involved in a specific debate in time is expected—RQ2 approaches a more complex, often neglected behavior. The philosophical model proposed by Nguyen [17] postulates that ECs are impossible to break/depolarize unless their members are willing to pass through a "cognitive reboot". Following the RQ2 perspective—and assuming [17] as a reference model—allows us to shed light on the real nature of ECs in specific online debates: Are they unbreakable/unrecoverable realities or more ephemerals—event-driven—epistemic bubbles whose behavior might change as time goes by?

The results, shown in Figs 5, 6 and 7, differed slightly for the three main categories of discussions analyzed. It should be noted that the potential ECs belonging to *Politics* and *Minorities discrimination* appear to be stable even over a long period, with variations that may be ascribed to the different topics discussed and to the temporal segmentation chosen for the snapshots. Instead, *Gun control* potential ECs—as shown in Fig 5—behave differently than those in the other two topics. In this case study, the Jaccard similarity between adjacent timestamps is low—except for a single EC that turns into a lesser polarized community between the end of 2017 and the beginning of 2018 (reaching a Jaccard of 0.53, against the 0.24 of the previous semester pairs).

Such a behavior is less pronounced in *Minorities discrimination* and *Politics* (Figs 6 and 7). In both cases, the internal stability is very high throughout the monitoring, reaching its highest peak, 93%, during the first year of discussions about *Minorities discrimination*. Then, as time passes, the internal similarity decreases slightly, and in certain cases, mesoscale topologies at risk of acting as ECs transit into communities that do not have strong ideological cohesion. In *Minorities discrimination* (Fig 6), as an example, it is interesting to note that the potential EC with the lowest percentage of common users between the first two semesters (75%) is also the one with the longest lifecycle (as potential EC), as it maintains its status until the very last pair of analyzed snapshots. Such behavior might also be identified in *Politics* (Fig 7), which has a similar case of an EC emerging during the second semester of 2017 that maintains such behavior until the end of our monitoring. Furthermore, it seems that the less polarized communities

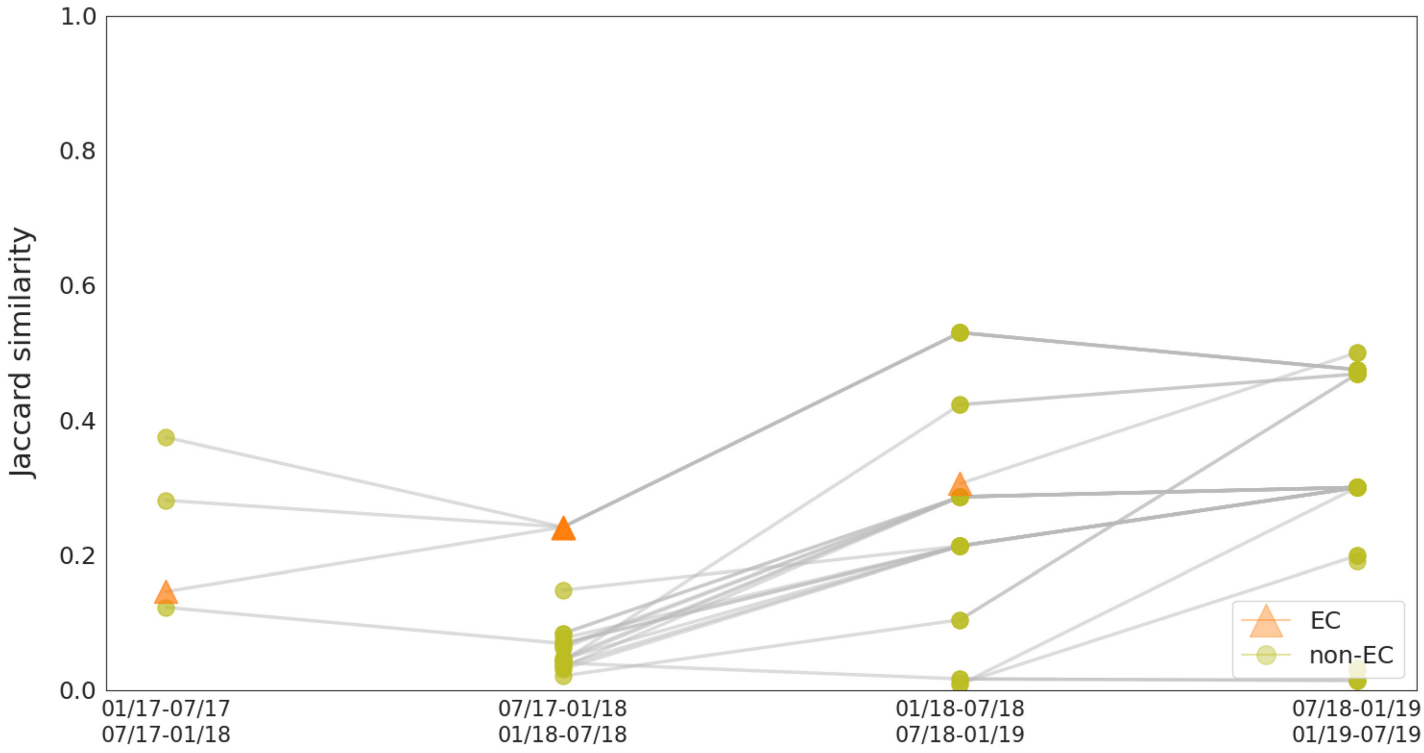

Fig 5. Gun Control Communities evolution through pairs of adjacent semesters. For each topic, the plot reports individual communities' Jaccard index (*y-axis*) trends through a pair of adjacent semesters (*x-axis*). Triangles mark the community as a "potential" echo chamber, while circles are non-ECs.

https://doi.org/10.1371/journal.pcsy.0000008.g005

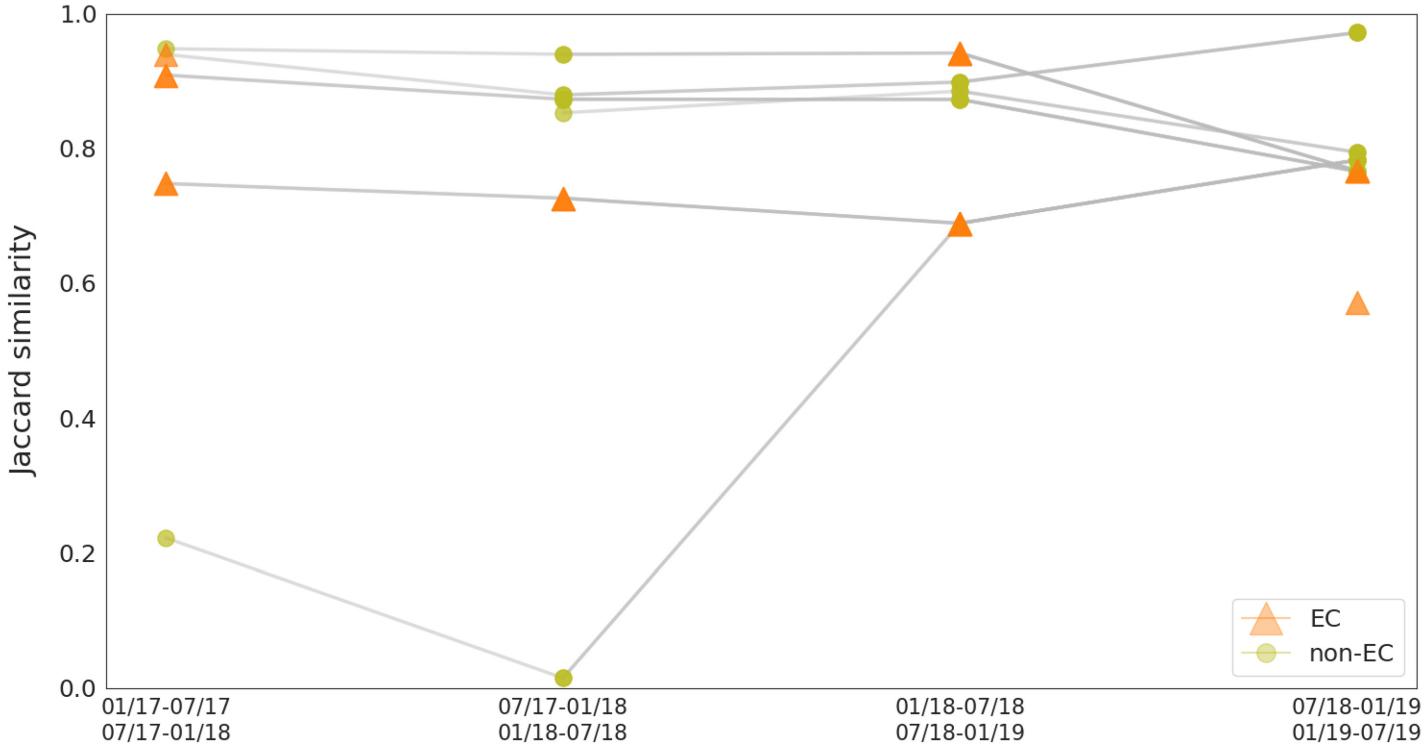

Fig 6. Minorities discrimination Communities evolution through pairs of adjacent semesters. For each topic, the plot reports individual communities' Jaccard index (*y-axis*) trends through a pair of adjacent semesters (*x-axis*). Triangles mark the community as a "potential" echo chamber, while circles are non-ECs.

https://doi.org/10.1371/journal.pcsy.0000008.g006

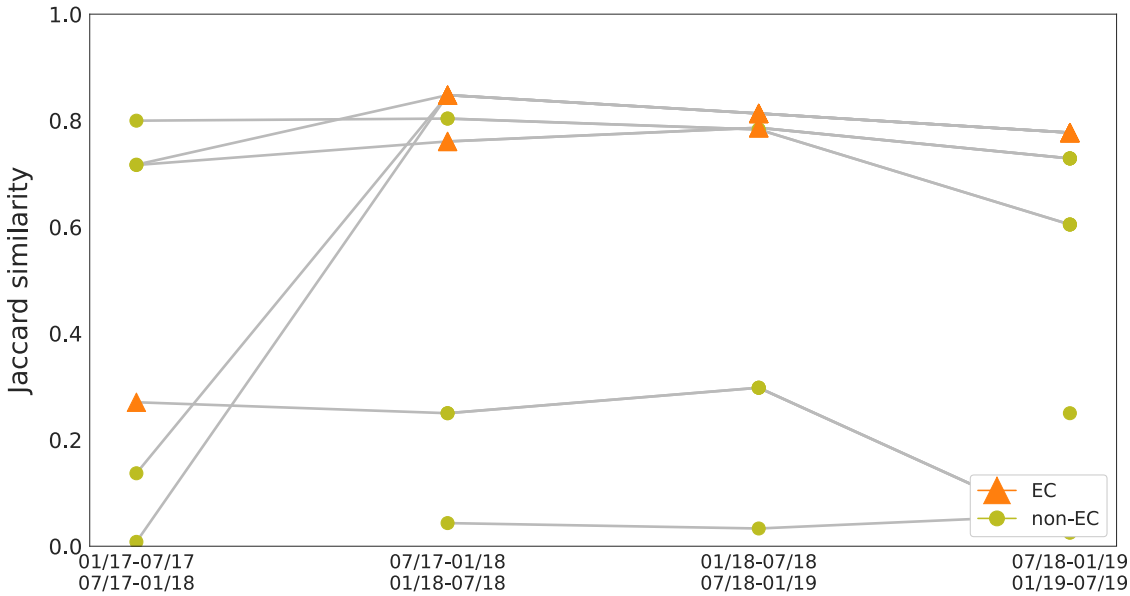

Fig 7. Politics Communities evolution through pairs of adjacent semesters. For each topic, the plot reports individual communities' Jaccard index (*y-axis*) trends through a pair of adjacent semesters (*x-axis*). Triangles mark the community as a "potential" echo chamber, while circles are non-ECs.

https://doi.org/10.1371/journal.pcsy.0000008.g007

(e.g., the ones characterized by a lower "risk") derived from ECs do not become ECs again. Moreover, similarity trends also underline the presence of "consistent" potential ECs that hold their status over shorter, still concerning timespans—e.g., six months.

Focusing on Not-ECs, from Fig 5, it can be noted that *Gun control* communities experience an increase in their internal stability between the second and the third semester, without reaching the stability peaks observed for potential ECs. Diversely, in both *Minorities discrimination* and *Politics*, communities' temporal stability tends to be very high—with a few exceptions, i.e., one community of the former dataset and two of the latter.

Topic modeling on polarized systems

After observing the overall persistence over time and the intrinsic stability of most potential ECs in terms of user compositions, we inspect the textual productions made by the users. To such an extent, we created a dataset containing the texts produced by active users for each topic and semester—augmented with information on whether the users are likely to be members of clusters acting as ECs. Therefore, we applied BERTopic (*Vectorizer_model* = *CountVectorizer*(*ngram_range*(1, 3)), *max_df* = 0.5), *min_topic_size* = 120, *nr_topics* = 13, *hdbscan_model* = *KMeans*, *representation_model* = *MaximalMarginalRelevance*(*diversity* = 0.4)) on all the documents (i.e., covering texts produced inside ECs and non-ECs users), thus identifying 13 topics. The corpus of user-generated textual data was preprocessed by normalizing, cleaning (expanding abbreviations and shortened forms, removing markdown characters), and lemmatizing (using Wordnet [89]) the raw text.

Finally, a set of representative keywords was extracted for each dataset post using KeyBERT [84], thus creating a fine-tuned vocabulary to label the identified topics better.

To reduce the number of outliers originally identified by BERTopic on the available data, we decided to substitute the default clustering algorithm it employs (HDBSCAN [80]) with K-Means [90]. Moreover, the minimum cluster size was set to 120 to avoid smaller—and noisier—topics.

Table 3. BERTopic. Topic coherence and diversity scores for the three case studies.

	Topic coherence	Topic diversity
Guncontrol	0.1661	0.8376
Minorities discrimination	0.2533	0.9145
Politics	0.1997	0.9316

https://doi.org/10.1371/journal.pcsy.0000008.t003

Finally, the clusters were further diversified by employing the *Maximum Marginal Relevance* [91] ranking algorithm, which allowed us to identify the most meaningful and diverse words describing each topic. Table 3 reports the *Coherence* and *Diversity* values for the identified topics.

Gun control. In both potential ECs and less polarized communities, it was possible to assess the presence of discussions about gun and ammunition brands and requests for advice from expert users. The over-representation of such topics leaves small room for others related to real-world events. Interestingly, at the beginning of 2018, a single high-risk EC (purity$\geq$0.8, conductance$\leq$0.2) focused mainly on a particularly controversial topic: the War in Syria. In the following semester, users of the same EC discussed the 2018 Firearms Amendment Act instead; then, the focus shifted back to the War in Syria.

Minorities discrimination. One interesting topic discussed by potential EC members from this dataset is Gamergate, an online social movement, for which one of the subreddits included in the dataset, namely *r/KotakuInAction*, represents the main discussion hub on Reddit, as stated on the subreddit homepage. The campaign started in 2014 to harass female journalists and developers involved in the video game industry who experienced doxing, rape, and death threats [92]. Despite lacking leaders or internal organization, it rapidly evolved into a broader movement targeting *Social Justice Warrior* ("Social Justice Warrior", Cambridge Dictionary. Last accessed July 18, 2023. Available: https://shorturl.at/zoH7y) activists and the perceived excess of political correctness in video games. According to Massanari [88]—who described the movement as an "echo chamber of anger"—it comprises people sharing the same core values of toxic masculine gaming culture, who may see the presence of women in the game industry as a threat. In addition, such a movement has also been addressed as ideologically near to the alt-right wing of the political spectrum [88]. Other controversial issues that emerged in this context are related to *anti-fascist* movements, often discussed along with the protests in Berkeley during 2017. Events started in 2017 when the right-wing supporter Milo Yiannapoulos was invited as a speaker at UC Berkeley. He encountered opposition from a group of armed anti-fascists who turned regular student opposition into a violent riot, damaging the university infrastructure and assaulting police forces. This first event turned into a chain of violent events that culminated at the end of August in the *Rally Against Hate*, in which far-left protesters clashed with far-right supporters. Outside ECs, recurring discussions were focused on more general but still polarizing topics, e.g., the *white privilege*, Canadian politics, and gender equality.

Politics. Users belonging to potential ECs mainly discussed abortion and the *Mueller special counsel investigation*, conducted to assess the interference of Russia in the 2016 U.S. elections. The latter topic was the main focus of the discussion in the echo chamber, with the longest life-cycle shown in Fig 7. Outside ECs, the communities seem to discuss other issues besides those discussed in ECs, e.g., news about Trump and politics, Obamacare, and Libertarianism. Further details on topic modeling can be found in the Supplementary Materials.

Valence analysis

After assessing polarizing topics in the selected case studies, our analysis deepened to gain insights into their perceived pleasantness/unpleasantness as vesiculated by the users' produced

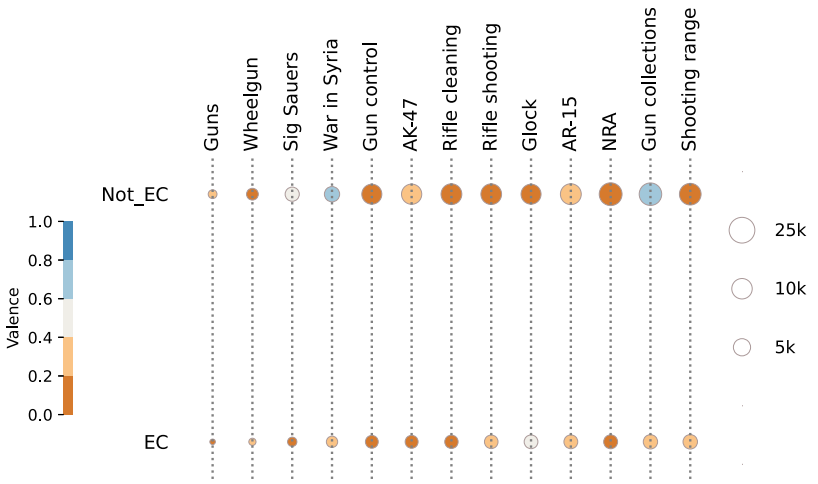

Fig 8. Gun Control—Topic valence (*x-axis*) for EC and non-ECs (*y-axis*) Reddit users' clusters. Colors describe the attitudes conveyed in texts. Polarized topics are characterized by a blue or dark orange hue, the former for the positive, the latter for negatively connotated ones. Circle sizes capture topic text volume.

https://doi.org/10.1371/journal.pcsy.0000008.g008

texts. To such an extent, the emotive attitude of users was investigated leveraging the NRC-Valence Lexicon [69], which allowed for computing the average Valence score of texts belonging to each topic. Specifically, for each topic, the average valence was calculated as the ratio of the sum of the valences of its keywords annotated in the VAD *w.r.t.* the total number of keywords in the lexicon. The results were then visually investigated with a scatterplot—Figs 8, 9 and 10. Each topic is colored according to its average valence score—orange for negativeness, pearl white for neutral, and blue for positiveness. The observed patterns reflect the inherently polarized nature of the topics under analysis, with only slight differences between ECs and communities.

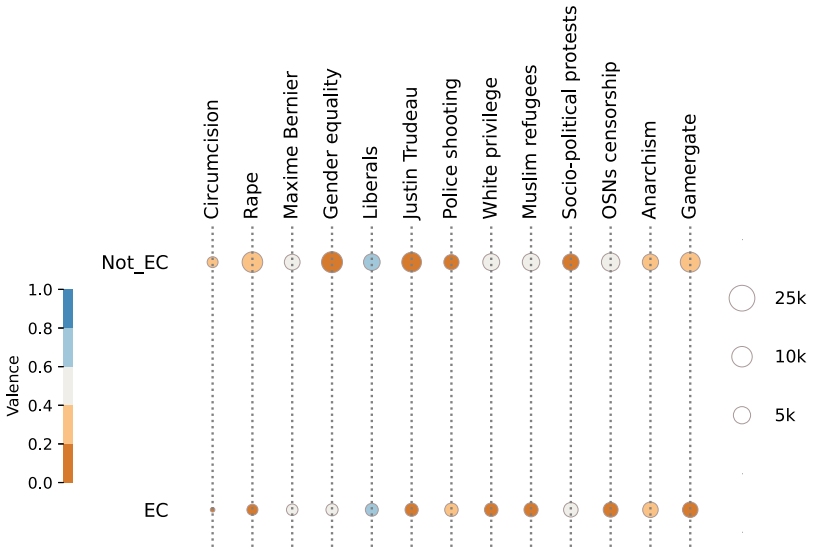

Fig 9. Minority discrimination—Topic valence (*x-axis*) for EC and non-ECs (*y-axis*) Reddit users' clusters. Colors describe the attitudes conveyed in texts. Polarized topics are characterized by a blue or dark orange hue, the former for the positive, the latter for negatively connotated ones. Circle sizes capture topic text volume.

https://doi.org/10.1371/journal.pcsy.0000008.g009

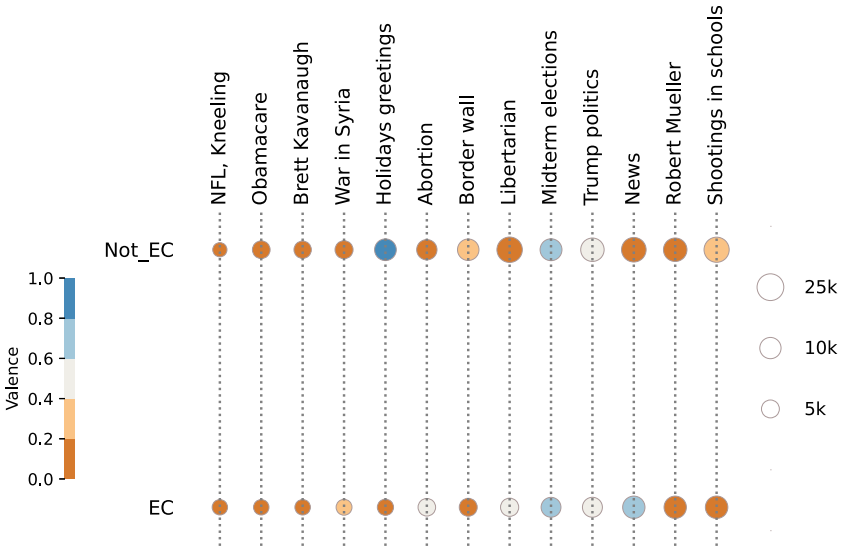

Fig 10. Politics—Topic valence (*x-axis*) for potential EC and non-ECs (*y-axis*) Reddit users' clusters. Colors describe the attitudes conveyed in texts. Polarized topics are characterized by a blue or dark orange hue, the former for the positive, the latter for negatively connotated ones. Circle sizes capture topic texts volume.

https://doi.org/10.1371/journal.pcsy.0000008.g010

Gun control. Fig 8—the difference between the two systems (potential ECs and not-ECs) is not as stark as we expected, as users outside ECs appear to discuss using more negative words than users inside ECs, except when talking specifically about *Gun collections* or *War in Syria*, where users outside ECs often use terms associated with more positive meaning.

Minorities discrimination. Fig 9—it is worth noticing that potential ECs and not-ECs users express positive attitudes only on a single topic—namely, discussions involving the center-left wing of the political spectrum. Moreover, we can also observe how topics that tend to be strongly negatively polarized outside ECs—e.g., police shootings against minorities and socio-political-protests (where the most frequently used terms are "nazi", "fascist" and "white privilege")—are more neutral inside ECs. Such a result, which might appear contradictory at first glance, might be related to the fact that in ECs, users tend to have less negative or condemning opinions on fascism and sensitive issues—a pattern we were able to assess from the available data qualitatively and that we will further inspect quantitatively in future works. Furthermore, the Gamergate controversy is characterized by an increase in wording negativeness, which the misogynistic nature of the movement may justify. Thus, we can infer that users are prone to condemn and attack women and minorities using negatively connoted language. This result is reflected and magnified by an ever more polarized negative attitude toward the topic "OSNs censorship", representing another core argument discussed by GamerGate supporters against SJWs. In particular, the topic refers to the Twitter ban on journalists involved in the movement, including Milo Yiannapoulos.

Politics. Fig 10—A negative connotation emerges in school shootings and debates on the Mexican border wall. Moreover, similarly to what was observed in the *Minority* dataset, some topics are treated as less negative in potential ECs w.r.t. not-ECs—e.g., War in Syria and abortion. At the same time, a more negative attitude emerges toward the border wall between Mexico and the United States.

Despite these results, the average valence score alone seems insufficient to highlight a clear distinction in sentiment between potential ECs and non-ECs. To address such a limitation, we plan, in a future study, to characterize the valence of cross-community topic-specific

Table 4. X/Twitter network statistics. Averaged number (per snapshot) of nodes, edges, degree, and density of the networks, and distribution of neutral and pro-/cons- taking the knee nodes' leaning attributes.

$\|V\|$	$\|E\|$	avg. degree	avg. density	Pro	Cons
3444	5279	2.54	0.001	2074	1290

https://doi.org/10.1371/journal.pcsy.0000008.t004

interactions to confirm what has been observed in other domains [25]—namely, the prevalence of out-group abusive comunication patterns.

5 Case study: X/Twitter BLM @ EUR0 2020

To support the platform-independence claims, we replicated the experiment detailed in the previous section on an X/Twitter dataset focused on the EURO2020 sportive event, where Italian users expressed their stances on the controversy around taking a knee in favor of the Black Lives Matter protests. The dataset used has been introduced in [31]—where the authors proposed a first study on ECs tracking and characterization—and subsequently leveraged in [10] to empirically validate a model focusing on the media roles in opinion formation processes.

The dynamic network analyzed comprises six snapshots—each covering the debate happening among Italian supporters after the games played by their national team. Interactions are built on top of the *mention* relation—therefore capturing a direct exchange of content among users. To each tweet in the dataset, we assigned a label (either "Pro", "Cons") to reflect the alignment of its content w.r.t. the BLM discussion: labels—and related numerical scores—were classified following the same strategy employed for the Reddit case study (where the positive/negative classes where manually annotated using known polarized hashtags used by the two factions). Users' leaning scores are then assigned—independently for each snapshot in which each user was active—following the same procedure defined for the Reddit datasets. Summary statistics are reported in Table 4.

EC risk and temporal stability

Fig 11 reports the communities identified in each snapshot characterized by their risk of acting as ECs. Assuming a coarse threshold of 0.7 for both the measures, we can observe that—starting from the first observation—"problematic" communities are always present for both the pro- and against-kneeling positions, although their relative sizes peek during the third and fourth observations, periods. These findings are particularly relevant as they correspond with the matches that sparked more intense debates. Indeed, during the third match (Italy vs. Wales), five Italian players chose to kneel, while others stood. Subsequently, in the fourth match (Italy vs. Austria), the Italian team made the final decision that they would kneel only if the opposing team did so, further inflaming the debate between supporters and opponents in Italy. Fig 12 underlines an interesting pattern, which focuses on the temporal stability of EC-like clusters. Conversely, from what was observed in the Reddit case studies, X/Twitter risky communities tend to remain as such during the considered period. However, there is a lower stability in terms of the users composing them, likely due to the viral and volatile nature of the discussions, which are tied to events that are short-lived and bounded by time.

Topic modeling and Valence analysis

Users discussions mostly revolved around the *take the knee* debate without straying from the central theme of the discussion (language="italian", vectorizer_model = paraphrase-multilingual-MiniLM-L12-v2, min_topic_size = 120, nr_topics = 6, hdbscan_model = cluster_model.

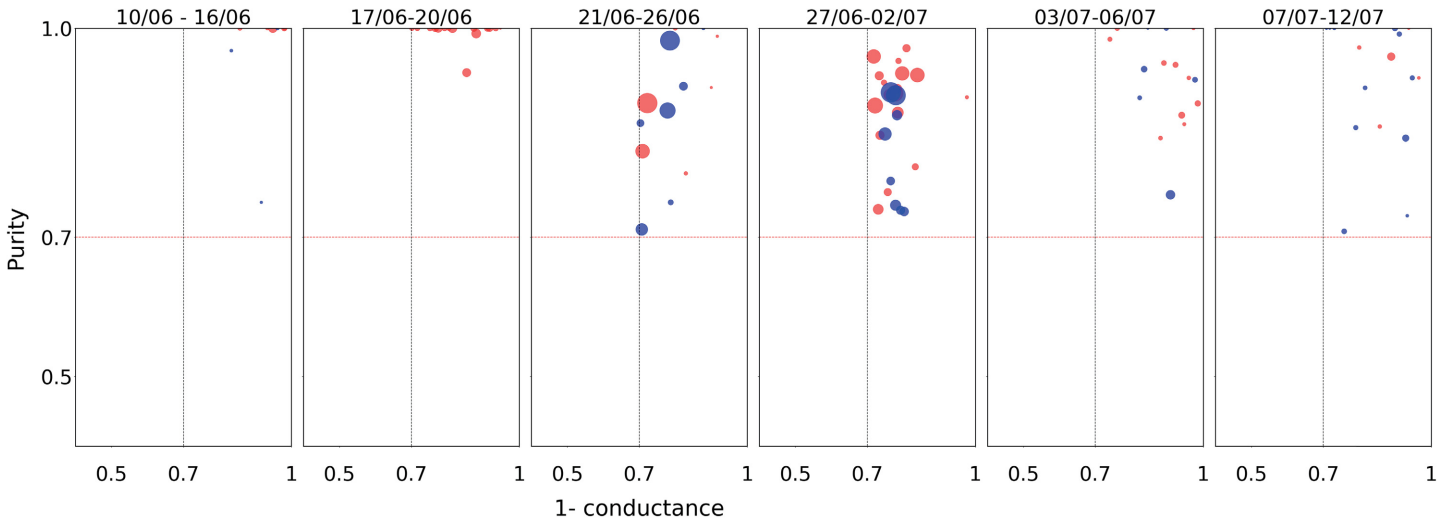

Fig 11. X/Twitter EURO2020. EC risk—Insulation and opinion coherence of Reddit communities extracted in the five temporal snapshots. Colors identify the community's prevalent opinion (red for *pro*, blue for *against*-kneeling respectively). Circle sizes are proportional to the number of users. Horizontal and vertical lines identify the coarse-grained EC threshold.

https://doi.org/10.1371/journal.pcsy.0000008.g011

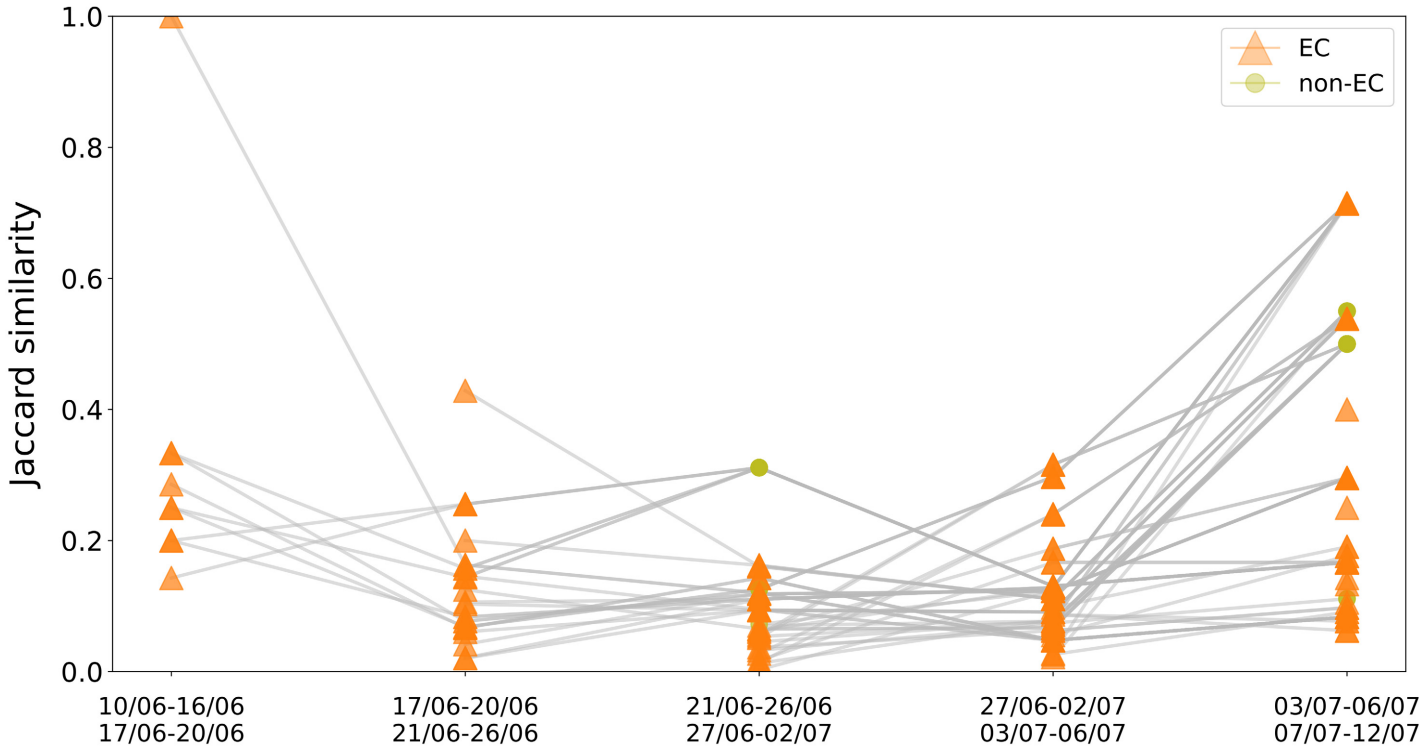

Fig 12. X/Twitter EURO2020. Communities evolution through pairs of adjacent observations—Triangles mark the community as a "potential" echo chamber.

https://doi.org/10.1371/journal.pcsy.0000008.g012

Table 5. BERTopic. Topic coherence and diversity scores for the Euro2020 case study.

	Topic coherence	Topic diversity
Euro2020	0.1355	0.944

https://doi.org/10.1371/journal.pcsy.0000008.t005

We focused on 6 topics only due to the smaller size of the dataset w.r.t. the Reddit one). The quality of topics extracted was examined through coherence and diversity (Table 5), which showed a high level of variety in the topics.

From the identified topics (see Fig 13) emerged the presence of two well-distinguished stances, pro and against kneeling in support of BLM, particularly referring to the press statements made by the Italian football players Giorgio Chiellini (*kneel only because other football teams do it*) and Federico Chiesa, who did not take the knee along with other four teammates during Italy—Wales. Another group of Twitter users condemned all the attention received by the BLM and the act of kneeling during Euro2020, comparing it to the feminicide of Pamela Mastropiero in early 2018. The event, although occurring in 2018, has often been used by Italian right-wing parties to express disapproval of illegal immigration, given the origins of the murderer identified in the trial. Regarding the valence attributed in texts by Twitter users, most topics were described positively in ECs, especially clusters supporting players with a stance against kneeling. Outside ECs, all the topics were instead discussed with a neutral valence.

6 Conclusions

In this work, we proposed a platform-independent pipeline to capture the topological dynamics of ECs and inspect the content shared by users associated with them—while characterizing users' topics of discussion w.r.t. their expressed valence. Our framework comprises four steps and leverages only users' interactions as extrapolated from textual exchanges—features common to most OSNs that can be leveraged to capture the topology of relations and the analyzed content shared.

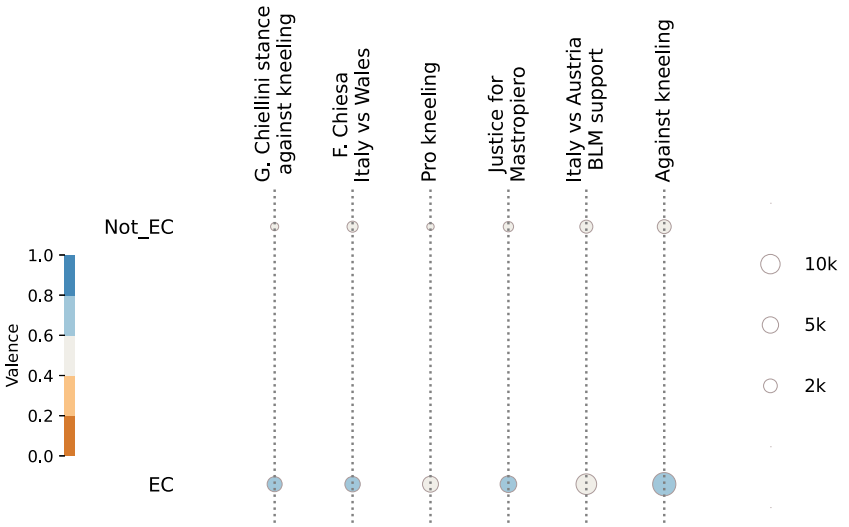

Fig 13. Topic valence (*x-axis*) for potential EC and non-ECs (*y-axis*) X/Twitter users' clusters. Colors describe the attitudes conveyed in texts. Polarized topics are characterized by a blue or dark orange hue, the former for the positive, the latter for negatively connotated ones. Circle sizes capture topic texts volume.

https://doi.org/10.1371/journal.pcsy.0000008.g013

The proposed pipeline is designed to work on a commonly shared EC definition—e.g., it considers ECs as closed systems of like-minded users mainly interacting with one another and actively avoiding alternative views. Such a qualitative definition is concretized by identifying those meso-scale topologies at risk of acting as ECs—controlling for internal ideological homogeneity and topological separation from the rest of the system—starting from node-attributed snapshot graphs.

We tested our methodology on three thematic case studies built on Reddit and one on X/Twitter data. Generally speaking, the results obtained applying our pipeline to social media data are aligned to the ones of recent literature [44, 87]—supporting the existence of ECs, although underlying the existence of different nuances of risks and diversifying their existence (and behavior/stability) on the basis platform/theme of discussion. Moreover, our analyses underline that Reddit potential ECs are relatively stable over time since they are able to retain a large portion of their users. Such a concerning trend is particularly evident for the *Minorities discrimination* discussions whose epistemic enclaves are shown to maintain high stability for up to two years. Conversely, X/Twitter ones experience a higher variability in their participants, although guaranteeing a more stable risky connotation.

Once we identified users' clusters at risk of behaving as ECs, we leveraged topic modeling to identify discussion themes. Such a characterization was then enriched with emotion analysis, thus quantifying the valence of the users' generated texts. Topic and valence analyses underlined that in Reddit potential ECs, most topics convey negative feelings, while in X/Twitter, usually more positive ones w.r.t. to what is observed in non-ECs. Moreover, we also observed that often, when sensible topics are taken into account—i.e., racism—users participating in risky social environments tend to use neutral wording rather than the negative one used by the rest of the population.

6.1 Notes on the approach applicability

The proposed analytical pipeline composes of multiple steps, each of them requiring some degree of data-driven fine tuning to be performed. Although the methodology can be applied in all those context where the available data met the minimum requirements discussed in Section 3 it is worth noticing that the specific configurations adopted in the presented case study cannot be considered as one-fits-all solutions.

As an example, the leaning annotation performed with BERT can be replaced with alternative transformer models finetuned on platform specific data, or even by other approaches (e.g., rule based ones). In that direction, we strongly encourage the analysts to carefully design and validate their classifiers by explicitating the assumptions made on the data used as "ground truth" during the training (being aware that they only represent proxies for a target, often unknown, variable).

Concerning community detection, we suggest using a feature-rich CD algorithm (as EVA) to leverage directly node semantic information while searching for choese mesoscale topologies. However, being the CD problem ill-posed, the selection of the best performing algorithm for a given network might also be subject to alternative choices focusing on different rationale than balancing structural and semantic partition quality. Similarly, when selecting the temporal scale for the analysis, we strongly advise to perform a preliminar study on the users' interaction frequency (to generate temporal snapshots covering comparable users' activity rates) and/or to identify semantically meaningful timespan based on the phenomenon analyzed (e.g., as done for the semester of the Trump presidency and the intra-game time window for the EURO 2020 scenario).

Fig 14. Echo Chambers thresholds. Cumulative percentage of communities/users considered at risk of acting/belongin as/to ECs. Each cell (i,j) reports the percentage of identified communities (red scale) or users (blue scale) having at least a Purity $\geq i$ and 1-Conductance $\geq j$. Each column identify one of the four analyzed datasets.

https://doi.org/10.1371/journal.pcsy.0000008.g014

Moreover, while addressing threshold selection, we advise focusing on mesoscale structures of relevant size, characterized by a $1 - Condactance \geq 0.7$ (to guarantee a reasonable separation from the rest of the network) and a $Purity \geq 0.7$ (to account for potential noise in the data). To better select both threshold values, we encourage to study the cumulative percentage of communities (and users' within them) considered at risk of acting as ECs. In Fig 14 we visually report, for all the four datasets analyzed the variation of such indicators values. As expected, only a small fraction of communities are associated to a high EC risk (e.g., taking as a cut-off value of 0.7 for both measures only 7%, 32%, 6% and 44% of, respectively, Gun Control, Minority, Politics and EURO2020); however, the users within such structures often represent a non-negligible part of the observed population. For the sake of simplicity, in our case studies, we imposed the same thresholds lower bounds—i.e. 0.7—to all the analyzed datasets. It is worth noticing, however, that the different risk distributions observed in the four datasets might have allowed to impose restrictive values, while still capturing risky behaviors of a reasonable portion of the users (e.g., setting 0.9 threshold values in Gun Control and 0.8 in Minority still guarantee a 16–20% coverage).

Due to the lack of a ground truth annotation it is also worth noticing that the final labeling (EC, not-EC) along with the conductance/purity level have to be considered as an indication of *risk*, not as an evidence. Finally, while characterizing the topics (reguarding the usage of BERTtopic the same rationale for the classification step applies) and valence of the users' produced contents we encourage the analysts to semantically validate the output by crossing them with known event that took place during the observed temporal span.

6.2 Approach weaknesses and limitations

The proposed analytical pipeline has weaknesses and limitations that must be considered. Firstly, although qualitative definitions of EC are usually shared and agreed on by the scientific community, a quantitative framework to assess their existence/measure their strength is something on which there is currently no consensus. Similarly, the proposed pipeline leverages Community Detection—a problem well-known in network science literature to be *ill-posed*—to identify topological-attribute homogeneous clusters of users to be used as proxies for ECs. Secondly, the framework generally lacks rigorous validation related to the absence of ground truth for labels describing the users' leanings. The labels adopted in the case studies are inferred through a classifier trained on polarized ground truth and act as a mere *proxy* to understand people's real—and multi-faceted—political leaning. Moreover, an intrinsic limitation is associated with the topic modeling stage: given the stochastic nature of UMAP—a dimensionality reduction methodology employed by BERTopic—the identified thematic clusters might be subject to slight instabilities across different extractions over the same input data [93].

6.3 Future developments

To better understand the complex nature of ECs, we plan to perform a more in-depth analysis of their users' network topology and textual data. In particular, for what concerns the underlying topological representation of social interactions, we plan to move from pairwise to high-order interactions to account for group dynamics explicitly. In this way, we can capture a wider range of interactions that might provide insight into homophilic behaviors related to the phenomenon, e.g., peer pressure. Furthermore, we aim to enhance content analysis by integrating and studying the *stance* of users towards the controversy in which ECs are detected by leveraging *Stance detection*, an NLP approach that fits well with the concept of echo chambers as it is related to the prediction of users' viewpoints toward a target [94]. Finally, we plan to validate the introduced pipeline on alternative case studies from other OSNs (e.g., focusing on understudied platforms like Bluesky Social [95]) to properly observe whether similar patterns characterizing ECs can be found in polarized/less polarized discussions.

Author Contributions

Conceptualization: Erica Cau, Virginia Morini, Giulio Rossetti.

Data curation: Erica Cau, Virginia Morini, Giulio Rossetti.

Formal analysis: Erica Cau.

Investigation: Erica Cau, Virginia Morini, Giulio Rossetti.

Methodology: Erica Cau, Virginia Morini.

Supervision: Giulio Rossetti.

Visualization: Erica Cau, Virginia Morini.

Writing – original draft: Erica Cau.

Writing – review & editing: Erica Cau, Virginia Morini, Giulio Rossetti.

References

1. Floridi L. The Fourth Revolution: How the Infosphere is Reshaping Human Reality. Oxford University Press UK; 2014.

2. Fu S., Li H., Liu Y., Pirkkalainen H., Salo M. Social media overload, exhaustion, and use discontinuance: Examining the effects of information overload, system feature overload, and social overload. Information Processing & Management, 2020. 57(6), 102307. https://doi.org/10.1016/j.ipm.2020.102307
3. Festinger L. Cognitive Dissonance. Scientific American. 1962; 207(4):93–106. https://doi.org/10.1038/scientificamerican1062-93 PMID: 13892642
4. Schwind C., Buder J., Cress U., Hesse F. W. Preference-inconsistent recommendations: An effective approach for reducing confirmation bias and stimulating divergent thinking? Computers & Education, 2012, 58(2), 787–796. https://doi.org/10.1016/j.compedu.2011.10.003
5. Bhadani, S. Biases in recommendation system. In Proceedings of the 15th ACM Conference on Recommender Systems, 2021, pp. 855–859.
6. Eslami, M., Aleyasen, A., Moghaddam, R. Z., Karahalios, K. Friend grouping algorithms for online social networks: Preference, bias, and implications. In Social Informatics: 6th International Conference, SocInfo 2014, Barcelona, Spain, November 11-13, 2014. Proceedings 6 (pp. 34–49).
7. Peralta A. F., Neri M., Kertész J., Iniguez G. Effect of algorithmic bias and network structure on coexistence, consensus, and polarization of opinions. Physical Review E, 2021, 104(4), 044312. https://doi.org/10.1103/PhysRevE.104.044312 PMID: 34781537
8. Stoica, A. A., Riederer, C., Chaintreau, A. Algorithmic glass ceiling in social networks: The effects of social recommendations on network diversity. In Proceedings of the 2018 World Wide Web Conference, 2018, (pp. 923–932).
9. Sîrbu A, Pedreschi D, Giannotti F, Kertész J. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. PloS one. 2019; 14(3):e0213246. https://doi.org/10.1371/journal.pone.0213246 PMID: 30835742
10. Pansanella V, Sirbu A, Kertész J, Rossetti G. Mass Media Impact on Opinion Evolution in Biased Digital Environments: a Bounded Confidence Model. Scientific Reports. 2023. https://doi.org/10.1038/s41598-023-39725-y PMID: 37670041
11. Branley D. B., Covey J. Is exposure to online content depicting risky behavior related to viewers' own risky behavior offline? Computers in Human Behavior, 2017, 75, 283–287. https://doi.org/10.1016/j.chb.2017.05.023
12. Kreski N. T., Chen Q., Olfson M., Cerdá M., Martins S. S., et al. Experiences of Online Bullying and Offline Violence-Related Behaviors Among a Nationally Representative Sample of US Adolescents, 2011 to 2019. Journal of school health, 2022, 92(4), 376–386. https://doi.org/10.1111/josh.13144 PMID: 35080013
13. Lim S. L., Bentley P. J. Opinion amplification causes extreme polarization in social networks. Scientific Reports, 2022, 12(1), 18131. https://doi.org/10.1038/s41598-022-22856-z PMID: 36307510
14. Terren L., Borge R. Echo chambers on social media: A systematic review of the literature. Review of Communication Research, Vol. 9 2021 https://doi.org/10.12840/ISSN.2255-4165.028
15. Bruns A. Echo chambers? Filter bubbles? The misleading metaphors that obscure the real problem. In Hate speech and polarization in participatory society, 2021, (pp. 33–48). Routledge.
16. Santos BR. Echo chambers, ignorance and domination. Social epistemology. 2021; 35(2):109–119. https://doi.org/10.1080/02691728.2020.1839590
17. Nguyen CT. Echo Chambers and Epistemic Bubbles; 2020.
18. Ardichvili A, Page V, Wentling T. Motivation and barriers to participation in virtual knowledge-sharing communities of practice. Journal of knowledge management. 2003; 7(1):64–77. https://doi.org/10.1108/13673270310463626
19. Cox A. What are communities of practice? A comparative review of four seminal works. Journal of information science. 2005; 31(6):527–540. https://doi.org/10.1177/0165551505057016
20. Dubé L, Bourhis A, Jacob R. The impact of structuring characteristics on the launching of virtual communities of practice. Journal of Organizational Change Management. 2005; 18(2):145–166. https://doi.org/10.1108/09534810510589570
21. Henri F, Pudelko B. Understanding and analysing activity and learning in virtual communities. Journal of Computer Assisted Learning. 2003; 19(4):474–487. https://doi.org/10.1046/j.0266-4909.2003.00051.x
22. Collins H. Bicycling on the moon: Collective tacit knowledge and somatic-limit tacit knowledge. Organization Studies. 2007; 28(2):257–262. https://doi.org/10.1177/0170840606073759
23. Lynch M. At the margins of tacit knowledge. Philosophia Scientiæ Travaux d'histoire et de philosophie des sciences. 2013;(17-3):55–73.
24. Turner S. Taking the collective out of tacit knowledge. Philosophia Scientiæ Travaux d'histoire et de philosophie des sciences. 2013;(17-3):75–92.

25. Falkenberg M, Zollo F, Quattrociocchi W, Pfeffer J, Baronchelli A. Affective and interactional polarization align across countries. arXiv preprint arXiv:231118535. 2023;.
26. Wojcieszak M, Casas A, Yu X, Nagler J, Tucker JA. Most users do not follow political elites on Twitter; those who do show overwhelming preferences for ideological congruity. Science advances. 2022; 8 (39):eabn9418. https://doi.org/10.1126/sciadv.abn9418 PMID: 36179029
27. Morini V, Pollacci L, Rossetti G. Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study. Applied Sciences. 2021; 11(12):5390. https://doi.org/10.3390/app11125390
28. Conover M, Ratkiewicz J, Francisco M, Goncalves B, Menczer F, Flammini A. Political Polarization on Twitter. Proceedings of the International AAAI Conference on Web and Social Media. 2021; 5(1):89–96. https://doi.org/10.1609/icwsm.v5i1.14126
29. Adamic LA, Glance N. The political blogosphere and the 2004 U.S. election: divided they blog. In: Proceedings of the 3rd international workshop on Link discovery. KDD05. ACM; 2005. p. 36–43. Available from: http://dx.doi.org/10.1145/1134271.1134277.
30. Guerra P, M W Jr, Cardie C, Kleinberg R. A Measure of Polarization on Social Media Networks Based on Community Boundaries. Proceedings of the International AAAI Conference on Web and Social Media. 2021; 7(1):215–224. https://doi.org/10.1609/icwsm.v7i1.14421
31. Buongiovanni C, Candusso R, Cerretini G, Febbe D, Morini V, Rossetti G. Will You Take the Knee? Italian Twitter Echo Chambers' Genesis During EURO 2020. In: International Conference on Complex Networks and Their Applications. Springer; 2022. p. 29–40.
32. Edwards A. (How) do participants in online discussion forums create 'echo chambers'? Argumentation in political deliberation. 2013; 2(1):127–150. https://doi.org/10.1075/jaic.2.1.06edw
33. Gilbert E, Bergstrom T, Karahalios K. Blogs are Echo Chambers: Blogs are Echo Chambers. In: 2009 42nd Hawaii International Conference on System Sciences; 2009. p. 1–10.
34. Ge Y, Zhao S, Zhou H, Pei C, Sun F, Ou W, et al. Understanding Echo Chambers in E-commerce Recommender Systems. In: Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval. SIGIR'20. ACM; 2020. p. 2261–2270. Available from: http://dx.doi.org/10.1145/3397271.3401431.
35. An J, Quercia D, Crowcroft J. Partisan sharing. In: Proceedings of the second ACM conference on Online social networks. ACM; 2014. p. 13–24. Available from: https://doi.org/10.1145/2660460.2660469.
36. Bakshy E, Messing S, Adamic LA. Exposure to ideologically diverse news and opinion on Facebook. Science. 2015; 348(6239):1130–1132. https://doi.org/10.1126/science.aaa1160 PMID: 25953820
37. Calderón FH, Cheng LK, Lin MJ, Huang YH, Chen YS. Content-based echo chamber detection on social media platforms. In: Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM; 2019. p. 597–600. Available from: https://doi.org/10.1145/3341161.3343689.
38. Garimella K, Morales GDF, Gionis A, Mathioudakis M. Political Discourse on Social Media. In: Proceedings of the 2018 World Wide Web Conference on World Wide Web—WWW '18. ACM Press; 2018. p. 913–922. Available from: https://doi.org/10.1145/3178876.3186139.
39. Kratzke N. How to Find Orchestrated Trolls? A Case Study on Identifying Polarized Twitter Echo Chambers. Computers. 2023; 12(3):57. https://doi.org/10.3390/computers12030057
40. Kleinberg JM, Kumar R, Raghavan P, Rajagopalan S, Tomkins AS. The web as a graph: Measurements, models, and methods. In: Computing and Combinatorics: 5th Annual International Conference, COCOON'99 Tokyo, Japan, July 26–28, 1999 Proceedings 5. Springer; 1999. p. 1–17.
41. Morales GDF, Monti C, Starnini M. No echo in the chambers of political interactions on Reddit. Scientific Reports. 2021; 11(1).
42. Villa G, Pasi G, Viviani M. Echo chamber detection and analysis. Social Network Analysis and Mining. 2021; 11(1). https://doi.org/10.1007/s13278-021-00779-3 PMID: 34457082
43. Abou-Rjeili A, Karypis G. Multilevel algorithms for partitioning power-law graphs. In: Proceedings 20th IEEE International Parallel & Distributed Processing Symposium. IEEE; 2006. p. 10–pp.
44. Cinelli M, DF Morales G, Galeazzi A, Quattrociocchi W, Starnini M. The echo chamber effect on social media. Proceedings of the National Academy of Sciences. 2021; 118(9). https://doi.org/10.1073/pnas.2023301118 PMID: 33622786
45. Fortunato S. Community detection in graphs. Physics reports. 2010; 486(3-5):75–174. https://doi.org/10.1016/j.physrep.2009.11.002
46. Rossetti G, Cazabet R. Community Discovery in Dynamic Networks. ACM Computing Surveys. 2018; 51(2):1–37. https://doi.org/10.1145/3172867

47. Palla G, Barabási AL, Vicsek T. Quantifying social group evolution. Nature. 2007; 446(7136):664–667. https://doi.org/10.1038/nature05670 PMID: 17410175

48. Cazabet R, Rossetti G. Challenges in Community Discovery on Temporal Networks. In: Computational Social Sciences. Springer International Publishing; 2019. p. 181–197. Available from: https://doi.org/10.1007/978-3-030-23495-9_10.

49. Rossetti G, Pappalardo L, Pedreschi D, Giannotti F. Tiles: an online algorithm for community discovery in dynamic social networks. Machine Learning. 2017; 106:1213–1241. https://doi.org/10.1007/s10994-016-5582-8

50. Kopacheva E, Yantseva V. Users' polarisation in dynamic discussion networks: The case of refugee crisis in Sweden. PLOS ONE. 2022; 17(2):e0262992. https://doi.org/10.1371/journal.pone.0262992 PMID: 35139109

51. Chunaev P. Community detection in node-attributed social networks: a survey. Computer Science Review. 2020; 37:100286. https://doi.org/10.1016/j.cosrev.2020.100286

52. Citraro S, Rossetti G. Identifying and exploiting homogeneous communities in labeled networks. Applied Network Science. 2020; 5(1). https://doi.org/10.1007/s41109-020-00302-1

53. Liu L, Tang L, Dong W, Yao S, Zhou W. An overview of topic modeling and its current applications in bioinformatics. SpringerPlus. 2016; 5(1). https://doi.org/10.1186/s40064-016-3252-8 PMID: 27652181

54. Hospedales T, Gong S, Xiang T. Video Behaviour Mining Using a Dynamic Topic Model. International Journal of Computer Vision. 2011; 98(3):303–323. https://doi.org/10.1007/s11263-011-0510-7

55. Blei DM, Ng AY, Jordan MI. Latent Dirichlet Allocation. J Mach Learn Res. 2003; 3(null):993–1022.

56. Moody CE. Mixing Dirichlet Topic Models and Word Embeddings to Make lda2vec; 2016. Available from: https://arxiv.org/abs/1605.02019.

57. Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed Representations of Words and Phrases and Their Compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems—Volume 2. NIPS'13. Red Hook, NY, USA: Curran Associates Inc.; 2013. p. 3111–3119.

58. Blei DM, Lafferty JD. Dynamic topic models. In: Proceedings of the 23rd international conference on Machine learning—ICML '06. ACM Press; 2006. p. 113–120. Available from: https://doi.org/10.1145/1143844.1143859.

59. Iwata T, Watanabe S, Yamada T, Ueda N. Topic Tracking Model for Analyzing Consumer Purchase Behavior. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence. IJCAI'09. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.; 2009. p. 1427–1432.

60. Acheampong FA, Nunoo-Mensah H, Chen W. Transformer models for text-based emotion detection: a review of BERT-based approaches. Artificial Intelligence Review. 2021; p. 1–41.

61. Devlin J, Chang MW, Lee K, Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding; 2018. Available from: https://arxiv.org/abs/1810.04805.

62. Liu Y, Ott M, Goyal N, Du J, Joshi M, Chen D, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach. ArXiv. 2019;abs/1907.11692.

63. Lewis M, Liu Y, Goyal N, Ghazvininejad M, Mohamed A, Levy O, et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In: Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. Online: Association for Computational Linguistics; 2020. p. 7871–7880. Available from: https://aclanthology.org/2020.acl-main.703.

64. Sanh V, Debut L, Chaumond J, Wolf T. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. ArXiv. 2019;abs/1910.01108.

65. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is All You Need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. NIPS'17. Red Hook, NY, USA: Curran Associates Inc.; 2017. p. 6000–6010.

66. Russell JA. Core affect and the psychological construction of emotion. Psychological Review. 2003; 110(1):145–172. https://doi.org/10.1037/0033-295X.110.1.145 PMID: 12529060

67. Bradley MM, Lang PJ. Affective Norms for English Words (ANEW): Instruction Manual and Affective Ratings; 1999.

68. Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13, 915 English lemmas. Behavior Research Methods. 2013; 45(4):1191–1207. https://doi.org/10.3758/s13428-012-0314-x PMID: 23404613

69. Mohammad S. Obtaining Reliable Human Ratings of Valence, Arousal, and Dominance for 20,000 English Words. In: Proceedings of the 56th Annual Meeting of the Association for Computational

Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics; 2018. p. 174–184. Available from: https://aclanthology.org/P18-1017.

70. Greene D, Doyle D, Cunningham P. Tracking the Evolution of Communities in Dynamic Social Networks. In: 2010 International Conference on Advances in Social Networks Analysis and Mining. IEEE; 2010. p. 176–183. Available from: https://doi.org/10.1109/asonam.2010.17.

71. Caceres RS, Berger-Wolf T, Grossman R. Temporal Scale of Processes in Dynamic Networks. In: 2011 IEEE 11th International Conference on Data Mining Workshops. IEEE; 2011. p. 925–932. Available from: https://doi.org/10.1109/icdmw.2011.165.

72. Salama M, Ezzeldin M, El-Dakhakhni W, Tait M. Temporal networks: a review and opportunities for infrastructure simulation. Sustainable and Resilient Infrastructure. 2020; 7(1):40–55. https://doi.org/10.1080/23789689.2019.1708175

73. Cazabet R, Rossetti G. Challenges in Community Discovery on Temporal Networks. In: Temporal Network Theory. Springer; 2023. p. 185–202.

74. Chiappori A, Cazabet R. Quantitative evaluation of snapshot graphs for the analysis of temporal networks. In: Complex Networks & Their Applications X: Volume 1, Proceedings of the Tenth International Conference on Complex Networks and Their Applications COMPLEX NETWORKS 2021 10. Springer; 2022. p. 566–577.

75. Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. Journal of Statistical Mechanics: Theory and Experiment. 2008; 2008(10):P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

76. Fortunato S, Barthelemy M. Resolution limit in community detection. Proceedings of the national academy of sciences. 2007; 104(1):36–41. https://doi.org/10.1073/pnas.0605965104 PMID: 17190818

77. Yang J, Leskovec J. Defining and evaluating network communities based on ground-truth. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics; 2012. p. 1–8.

78. Grootendorst M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure; 2022. Available from: https://arxiv.org/abs/2203.05794.

79. Egger R, Yu J. A topic modeling comparison between lda, nmf, top2vec, and bertopic to demystify twitter posts. Frontiers in sociology. 2022; 7:886498. https://doi.org/10.3389/fsoc.2022.886498 PMID: 35602001

80. McInnes L, Healy J, Saul N, Großberger L. UMAP: Uniform Manifold Approximation and Projection. Journal of Open Source Software. 2018; 3(29):861. https://doi.org/10.21105/joss.00861

81. Bellman R. Dynamic programming. Science. 1966; 153(3731):34–37. https://doi.org/10.1126/science.153.3731.34 PMID: 17730601

82. Dieng AB, Ruiz FJR, Blei DM. Topic Modeling in Embedding Spaces. Transactions of the Association for Computational Linguistics. 2020; 8:439–453. https://doi.org/10.1162/tacl_a_00325

83. Bouma G. Normalized (pointwise) mutual information in collocation extraction. Proceedings of GSCL. 2009; 30:31–40.

84. Grootendorst M. KeyBERT: Minimal keyword extraction with BERT.; 2020. Available from: https://doi.org/10.5281/zenodo.4461265.

85. Morini V, Pollacci L, Rossetti G. Capturing Political Polarization of Reddit Submissions in the Trump Era. In: SEBD; 2020. p. 80–87.

86. Top Websites Ranking—Most Visited Websites in June 2023;. Available from: https://www.similarweb.com/top-websites/.

87. Waller I, Anderson A. Quantifying social organization and political polarization in online platforms. Nature. 2021; 600(7888):264–268. https://doi.org/10.1038/s41586-021-04167-x PMID: 34853472

88. Massanari A. #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. New Media and Society. 2016; 19(3):329–346. https://doi.org/10.1177/1461444815608807

89. Fellbaum C, editor. WordNet: An Electronic Lexical Database. Language, Speech, and Communication. Cambridge, MA: MIT Press; 1998.

90. Lloyd S. Least squares quantization in PCM. IEEE Transactions on Information Theory. 1982; 28(2):129–137. https://doi.org/10.1109/TIT.1982.1056489

91. Carbonell J, Goldstein J. The use of MMR, diversity-based reranking for reordering documents and producing summaries. In: Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval. ACM; 1998. p. 335–336. Available from: https://doi.org/10.1145/290941.291025.

92. Jhaver S, Chan L, Bruckman A. The view from the other side: The border between controversial speech and harassment on Kotaku in Action. First Monday. 2018. https://doi.org/10.5210/fm.v23i2.8232

93. McInnes L, Healy J, Melville J. Umap: Uniform manifold approximation and projection for dimension reduction. arXiv preprint arXiv:180203426. 2018;.

94. Alturayeif N, Luqman H, Ahmed M. A systematic review of machine learning techniques for stance detection and its applications. Neural Computing and Applications. 2023; 35(7):5113–5144. https://doi.org/10.1007/s00521-023-08285-7 PMID: 36743664

95. Failla, A., Rossetti, G. "I'm in the Bluesky Tonight": Insights from a Year Worth of Social Data. arXiv preprint arXiv:2404.18984 (2024).