Article

Toward a Standard Approach for Echo Chamber Detection: Reddit Case Study

Virginia Morini [1], Laura Pollacci [2] and Giulio Rossetti [1,*]

1   KDD Lab, CNR-ISTI, 56126 Pisa, Italy; virginia.morini@isti.cnr.it
2   KDD Lab, University of Pisa, 56126 Pisa, Italy; laura.pollacci@di.unipi.it
*   Correspondence: giulio.rossetti@isti.cnr.it

Featured Application: The framework proposed in this paper could be used to detect echo chambers in a standard way across multiple online social networks (i.e., leveraging features they commonly share). Such application then allows for comparative analysis between different platforms, thus discovering if some are more polarized than others. Further, a standard echo chamber characterization could be a starting point for designing a Recommendation System able to recognize and mitigate such a phenomenon.

Abstract: In a digital environment, the term echo chamber refers to an alarming phenomenon in which beliefs are amplified or reinforced by communication repetition inside a closed system and insulated from rebuttal. Up to date, a formal definition, as well as a platform-independent approach for its detection, is still lacking. This paper proposes a general framework to identify echo chambers on online social networks built on top of features they commonly share. Our approach is based on a four-step pipeline that involves (i) the *identification of a controversial issue*; (ii) the *inference of users' ideology* on the controversy; (iii) the *construction of users' debate network*; and (iv) the *detection of homogeneous meso-scale communities*. We further apply our framework in a detailed case study on Reddit, covering the first two and a half years of Donald Trump's presidency. Our main purpose is to assess the existence of Pro-Trump and Anti-Trump echo chambers among three sociopolitical issues, as well as to analyze their stability and consistency over time. Even if users appear strongly polarized with respect to their ideology, most tend not to insulate themselves in echo chambers. However, the found polarized communities were proven to be definitely stable over time.

Keywords: social network analysis; echo chambers; polarization; community discovery

1. Introduction

During the last decade, the ease of access and ubiquity of online social media and online social networks (OSNs) has rapidly changed how we are used to searching, gathering, and discussing any kind of information. Such a revolution, as the promise of equality carried by the World Wide Web since its first appearance, has enriched us all. It has made geographical distances vanish and empowered all internet users, letting their voices be heard [1,2]. However, at the same time, it has also increased our chances of encountering misleading behaviors. Indeed, the unlimited freedom of generating contents and the unprecedented information flooding we are used today are seeds that, if improperly managed, can make online platforms into fertile grounds for polluted realities [3,4]. Among them, several studies [5–7] claim that overpersonalization enhanced by OSNs, leveraging the human tendency to interact with like-minded individuals, might lead to a self-reinforcing loop confining users in echo chambers. Although a formal definition of the phenomenon is still missing, an echo chamber (EC) is commonly defined as a polarized situation in which beliefs are amplified or reinforced by communication repetition inside a closed system and insulated from rebuttal.

Such phenomenon might prevent the dialectic process of "thesis–antithesis–synthesis" that stands at the basis of a democratic flow of opinions, fostering several related alarming episodes [8] (e.g., hate speech, misinformation, and ethnic stigmatization). Furthermore, since discussions, campaigns, and movements taking place in online platforms also resonate in the physical world, it is no more possible to relegate their effect only to the virtual realm. For such reasons, a large body of scientific works [9–14] has addressed the issue of echo chamber detection over the last decade, moving from content-only characterization toward a structural/topological analysis of the phenomenon. However, the lack of an actionable definition of echo chambers and a standard strategy to support their identification has led to conflicting experimental observations [15]. Further, most of the abovementioned works leverage platform-specific features or resources to identify ECs, thus strongly limiting their generalizability to the entire range of social media platforms.

Moving from such literature, the purpose of this paper is twofold. Firstly, we propose a general framework to identify echo chamber on OSNs built on top of features they commonly share. Secondly, we aim to enrich the body of knowledge on EC detection, presenting a detailed case study on Reddit (i.e., the least explored social platform from an EC point of view).

Starting from online social platform data, the approach to detect echo chambers presented in this paper can be summarized into a four-stage pipeline. *(i)* Since opinion polarization generally arises in the presence of topics that trigger a significant difference of opinions, we propose starting from *controversial issue identification*. *(ii)* Then, because an EC key feature is the homogeneous group thinking, the second step consists of *inferring users' ideology on the controversy* from posts shared on the platform. *(iii)* People inside an EC tend to interact with like-minded individuals, thus insulating themselves from rebuttal. To assess this requirement, we propose to *define the users debate network* retrieving all posts' comments and labeling users with their leaning on the controversy. *(iv)* Lastly, the fourth step consists of *homogeneous meso-scale users' clusters identification*. In other words, we look for areas of the network that are homogeneous from an ideological and topological point of view.

Subsequently, we provide a case study of the proposed framework on Reddit. We focus on the debate between Pro-Trump supporters and Anti-Trump citizens during the first two and a half years of Donald Trump's presidency and look for ECs among three sociopolitical topics (i.e., gun control, minorities discrimination, and political discussions). We find that Gun Control users, even if they show a strong tendency toward Pro-Trump beliefs, do not mostly insulate themselves in polarized communities. On the other hand, for the topics of Minorities Discrimination and Political Sphere, we are able to detect homogeneous ECs among different semesters. Moreover, we assess their stability and consistency over contiguous semesters, finding that ECs members have a high probability of interacting with like-minded individuals over time.

The rest of the paper is organized as follows. In Section 2, we discuss the literature involving echo chamber detection first in more general terms, then focusing on the echo and chamber dimension of the phenomenon. Section 3 provides an overview of our four-step framework for EC detection, describing its rationale and providing examples of its applicability to multiple platforms. Then, in Section 4, we test such a framework in a specific case study involving the identification and analysis of echo chambers on Reddit. We conclude in Section 5 with a discussion on results and directions for future work.

2. Related Work

2.1. Echo Chamber Detection

Detecting and characterizing the echo chamber phenomenon is of utmost importance since it is the first step toward the deployment of actionable strategies to mitigate its effects. Although attempts to identify such a phenomenon have existed for several years now, there has been an exceptional growth of research efforts in the last decade that span over

a wide range of different digital services (e.g., blogosphere [9], movie recommendation services [12], E-commerce platforms [11], and music streaming applications [10]).

For the sake of this paper, we will focus on approaches proposed to address this issue on online social networks. Since a formal definition of echo chamber, as well as a standard methodology to assess its existence, is still missing, previous works follow different strategies to deal with it. We attempt to categorize them first by considering their focus on the *echo* or *chamber* dimension of the phenomenon, i.e., respectively the debated content/opinion and the network which allows them to echo among users. The *content-based* approach relies on the assumption that polarized environments are detectable by looking at the leaning of content shared or consumed by a user as well as analyzing its sentiment on the controversy, regardless of its interactions with others. For instance, An et al. [16] explore the US debate between Liberals and Conservatives on Facebook and Twitter, looking for partisan users, i.e., users who share news articles conforming to their political beliefs. Bakshy et al. [14] follow a similar strategy on Facebook, but they also consider users' exposure to crosscutting contents from the news feed or friends. On the other hand, the *network-based* strategy mainly focuses on finding clustered topologies in users' interactions rather than on their content homophily. In dealing with it, the authors of [17], first define the conversational network of Facebook users discussing the 2014 Thai election and then partition it into close-knit communities from a topological point of view. Nevertheless, most researchers have deployed a hybrid methodology to detect ECs, taking into account both users' ideology as well as their interactions with each other. One of the first works in this direction is attributable to Barberá et al. [18] who explored whether online communication resembles an EC, collecting several million tweets concerning twelve political and nonpolitical issues. Authors infer users' ideology relying on their using *follow* for popular controversial accounts then define their interaction network via *retweet*. Similarly, Garimella et al. [19] first estimate users' leaning on political controversy based on the media slant that they share and consume and thus define the debate network through the *follow* relationship. To the best of our knowledge, only Morales et al. [20] have tackled such a task on Reddit, focusing on the debate between Republicans and Democrats during the 2016 elections.

Different approaches have also led to a difference in scale in the detected echo chambers. We define *micro-scale ECs* the outputs of those approaches in which EC detection is accomplished, relying on the online behavior of single users, thus losing their aggregate dimension [14,16]. On the contrary, *macro-scale ECs* are identified looking at the users' interaction network on an aggregated level, not taking into account differences within certain areas of the network. For instance, [20,21] check if the overall network is strongly characterized by two insulated groups of users, i.e., the two sides of the controversy. In addition, Conover et al. [22] search for a similar output, using a community detection approach but forcing the algorithm to find exactly two communities. Differently, a *meso-scale EC* can be considered as a subset of nodes in the overall network that resembles an echo chamber. This implies that in the overall debate network, it is possible to identify multiple ECs having the same ideological leaning. For example, Grömping [17] leverages the Modularity function (see Definition 2) to discover several cohesive clusters across Facebook pages. In Table 1, we give a summary of related works based on the two categorizations proposed above.

Table 1. Summary of related works for echo chamber detection on online social networks.

Paper	Network	Content	Micro-Scale EC	Meso-Scale EC	Macro-Scale EC
[14,16]	✗	✓	✓	✗	✗
[13]	✗	✓	✗	✗	✓
[17]	✓	✗	✗	✓	✗
[22]	✓	✓	✗	✗	✓
[18]	✓	✓	✗	✗	✓
[19–21]	✓	✓	✗	✗	✓

Summing up, according to Dubois and Blank [15], there are two key methodological issues with how echo chamber work has been conducted. First, the adoption of so many different approaches for EC detection has led to conflicting results over the years. Second, most of works leveraged platform-specific features or resources to identify ECs, limiting their generalizability, e.g., *mention* and *retweet* [18,22], *follow* [19,21], *friend* [13], users' clicks [14], pre-annotated datasets [14,22].

Moving from the literature, in this paper we propose a framework to identify *meso-scale ECs* built on top of features commonly shared by most OSNs. In doing so, we leverage both the content and network dimension of the phenomenon. In the following, we provide a brief overview of techniques both for ideology and community detection used in this work.

2.2. Ideology Detection

Estimating user ideology on a controversy is challenging since online social platform users are rarely explicitly associated with a specific ideological label. Several works focusing on political EC [19,21,23] rely on the media slant consumed or shared by a user to infer its ideology. However, such a strategy is not feasible in other domains and, further, there is no guarantee that users only consume news outlets they agree with. On the other hand, the authors in [24,25] have shown that different sides of a controversy are also detectable via the lexicon used. For instance, Mejova et al. [26] state that the language used in these contexts is often characterized by strongly biased terms as well as negative statements. Following this last line of reasoning, we estimate users' ideology on a controversy based on how they speak/write about it. Thus, we model such an issue as a text classification task.

When dealing with textual data, it is of utmost importance to take into account both the suitable type of word representation and the proper type of classifier. Since traditional word representations (i.e., bag-of-words model) encode words as discrete symbols not directly comparable to others [27], they are not fully able to model semantic relations between words. Instead, word embeddings (e.g., Word2vec [28] and Glove [29]), mapping words to a continuously valued low dimensional space, can capture their semantic and syntactic features. Moreover, their structure makes them suitable to be deployed with Deep Learning models, fruitfully used to address Natural Language Processing-related (NLP) classification tasks. Specifically, Recurrent Neural Networks (RNNs) have been proven to be extremely successful for sequence learning [30]. Among them, Long Short-Term Memory (LSTM) [31] network can maintain long-term dependencies through an elaborate gates mechanism, overcoming the vanishing gradient problem of standard RNNs. Another Deep Learning network dealing with NLP-related tasks is Convolutional Neural Network (CNN) [32]. Even though CNNs are primarily used in computer vision problems, e.g., image recognition [33,34], they have also been proven to perform well with textual data, including generic NLP tasks, e.g., Part-Of-Speech Tagging, Chunking, Named Entity Recognition and Semantic Labeling [35], sentence modeling [36], search queries and document retrieval [37], and sentence classification [38]. Compared to the sequential architecture of RNNs, CNN has a hierarchical architecture able to extract position-invariant features instead of modeling context dependencies. Accordingly, they are most suitable in such tasks where keyword extraction is involved.

More recently, in the literature have been introduced the so-called Transformer models that, differently from the previous ones, can process each word in a sentence simultaneously via the attention mechanism. In particular, autoencoding transformer models such as Bidirectional Encoder Representations from Transformers (BERT) [39] and the many BERT-based models spawning from it (e.g., RoBERTa [40], DistilBERT [41]), have proven that leveraging a bidirectional multi-head self-attention scheme yields state-of-the-art performances when dealing with sentence-level classification. However, also autoregressive monodirectional Transformer models like OpenGPT [42,43] can be fruitfully leveraged for such kind of task [44]. Unlike BERT-based models, they leverage masked self-attention, which models the overall meaning of a sentence by only looking at the left context.

2.3. Community Detection

When dealing with meso-scale topologies, it turns out to be necessary to define a strategy to partition the overall interaction network in homogeneous clusters. By doing so, we can better identify different cohorts of individuals sharing a common set of features.

In complex networks, a widely popular way to address node clustering relies on community detection (CD) approaches. However, there is no single, standard definition of what a community should look like. Indeed, several algorithms have been proposed so far to efficiently partition graphs into connected clusters, often maximizing specifically tailored quality functions [45]. As regards echo chamber-related tasks, Conover et al. [22], to discover the political structure of defined *retweet* and *mention* networks, rely on the *Label Propagation* algorithm, i.e., assigning an arbitrary cluster membership to each node and then updating it according to the label shared by most of its neighbors. In [17], authors exploit the *Internal-Density* CD approach, partitioning the Facebook pages graph using the Modularity function. Further, Cossard et al. [46] look for highly segregated communities of users in the Italian vaccination debate via the Infomap algorithm. The aforementioned works model the issue mainly identifying an accurate partition of nodes from a topological point of view, regardless of nodes homophily.

From our perspective, echo chamber's members carry valuable semantic information (i.e., their opinion on the controversy) that has to be taken into account when computing cohesive communities. For such a reason, we will leverage a specific instance of the CD problem in the present work, namely *Labeled Community Detection* (LCD). Formally, given a labeled graph $\mathcal{G}$, LCD algorithms aim to find a node partition $C = \{c_1, \ldots, c_n\}$ of $\mathcal{G}$ that maximizes both topological clustering criteria and label homophily within each community.

3. Detecting Echo Chambers in Online Social Platforms

The lack of a formal and actionable echo chamber definition and of a standard strategy to support their identification has, over the years, led to conflicting experimental observations, namely scientific studies whose comparison is rather unfair due to their hard-wiring to platforms' specific characteristics. In this regard, here we propose a general framework to detect echo chambers built on top of features commonly shared by most of online social platforms.

Before discussing our framework, it is important to fix—in an actionable way—the object of our investigation along with its expected properties.

Definition 1 (**Echo Chamber**). *Given a network describing users' interactions centered on a controversial topic, an echo chamber is a subset of the network nodes (users) who share the same ideology and tend to have dense connections primarily within the same group.*

Following 1, to assess an EC's existence we cannot rely on a single user's digital traces (i.e., following a *micro-scale approach*), nor suppose that all users in the network belong to a polarized community (i.e., assuming a *macro-scale approach*). The proposed EC definition relies on *meso-scale* topologies. We focus on this particular definition of EC since we are interested in the role that group dynamics play in the increase in polarization. Moreover, we believe that it is quite unrealistic that all the users involved in a controversial debate insulate themselves in an echo chamber. Our framework composes of four steps to identify EC starting from online social platform data, namely (i) controversial issue identification; (ii) users' ideology inference; (iii) debate network construction; (iv) homogeneous meso-scale users' clusters identification.

In this section, we provide an overview of each of such steps, describing their rationale and supporting our claim of framework generality by providing examples of its applicability to multiple platforms. Subsequently, in Section 4 we propose a specific case study involving the identification and analysis of echo chambers in a well-known online social platform, Reddit. There, we describe a specific instance of the proposed framework and discuss the choices made while implementing its pipeline.

3.1. Step 1. Controversial Issue Identification

The first step in our pipeline consists in identifying a controversial issue. With this terminology, we refer to questions, subjects, or problems that can create a great difference of opinion among people discussing it. Of course, they can include topics that may have political, social, environmental, or personal impacts on a community. We rely on controversial issues due to the fact that the polarization of opinions generally emerges in such situations, turning divergent attitudes into ideological extremes [3].

Controversial issues are debated both in offline and online realms. However, online social platforms are probably the most frequently used open space for discussions. Further, their structure and functionalities make quite easy to identify online debates about a wide range of different issues. For instance, on Twitter, it is possible to search for a specific hashtag to discover users debating a topic. Further, both Reddit and Gab, thanks to their structure divided into subreddits or groups, make it even simpler. Additionally, they allow searching for topic-related communities (e.g., gaming, politic, entertainment) via several lists available on the platforms. Accordingly, it is also possible to retrieve such kinds of data, leveraging platforms API (e.g., Twitter (Twitter API: https://developer.twitter.com/en/docs/twitter-api (accessed on 1 June 2021)), Reddit and Gab (Reddit and Gab API: https://github.com/pushshift/api) (accessed on 1 June 2021)), or external released datasets (DocNow Catalog of Twitter Datasets: https://catalog.docnow.io/ (accessed on 1 June 2021)).

3.2. Step 2: Users' Ideology Inference

Throughout the slightly different definitions of echo chambers given over the years, the concept of homogeneous group thinking always emerges. Indeed, once we have identified a controversial issue and users discussing it, the second step of our approach consists of estimating users' leaning on the controversy. Following the rationale explained in Section 2.2, we rely on users' posts and comments to infer users' ideology. Indeed, all online social platforms provide some sort of text publishing as a basic functionality to their users (e.g., Twitter, Facebook, Reddit, Gab). Thus, we model the task of predicting the ideology of a user on a controversy as a text classification problem. In other words, given a controversial topic (e.g., the debate around Gun Control), we encode the textual content of users' posts into a low-dimensional vector representation and use it to train a classifier aimed at predicting users' ideology (e.g., *Anti-Gun*, *Pro-Gun*). Notice that the selection of one text classification model over the other strictly depends on data source peculiarities. Accordingly, it is of utmost importance to perform ad hoc model selection and fine-tuning to consider context-specific features. For instance, the size of available data is a crucial part of selecting a model. Usually, Deep Learning models need a huge amount of data to learn model parameters as well as impart generalization. However, pre-trained models such as Transformers are able to target the classification task even with low data resources thanks to their extensive pre-training. Another aspect to consider is whether we are interested in capturing sentence-level semantics or, rather, in extracting specific information from the sentence. In the first case, methods able to model long-range dependencies are the most suitable (e.g., LSTM, Transformer models), while in the second, architecture able to extract position-invariant features (e.g., CNN) are preferable [47].

Moreover, a controversy may not necessarily induce binary opinions about it, e.g., wrong/right, pro/against. Suppose, for instance, that we are interested in exploring the political debate in the US. Beyond the simple Republican–Democrat analysis, we can identify other popular ideologies such as Liberalism, Conservatism, Libertarianism, and Populism. Accordingly, in such a scenario, our framework can be instantiated with a multi-class text classifier.

Since we tackle a supervised approach, it requires a ground truth of sample posts labeled with respect to their opinion on the controversy. Even if it seems a tricky step in our pipeline, it is not so difficult to find among different online social platforms, publicly known polarized user collectors in which subscribed users support a specific leaning. For

instance, we may rely on Twitter lists, Facebook pages, Reddit subreddits, or Gab groups. Further, on those online platforms in which the majority of users do not use a nickname (e.g., Twitter and Facebook), we may also rely on public figures supporting a specific idea to define a ground truth. For example, following the previous example on the US political debate, we may retrieve all posts shared by the most prominent exponents of those parties.

3.3. Step 3: Debate Network Construction

Once we have identified a set of users labeled with respect to their ideology on the controversy, the next step consists of defining their interaction network. The concepts of exposure, affiliation, and interaction with like-minded individuals is a crucial aspect of echo chambers. Accordingly, it becomes necessary to define when an interaction between two users takes place. Most of the previous approaches [19,20,48] rely on the follow or friend relationship to define the connections among users. However, not all online social platforms allow discovering who follows whom and, thus, retrieving such information, e.g., Reddit. Further, we believe that it is quite unrealistic that a user has a direct relationship with all the users they follow or that consume all their contents.

For such reason, we build the interaction graph among the previously labeled users through the who–comment–whom relation, since such a feature is available across all the main online platforms, as in the case of textual contents. Formally, we define the interaction network as a graph $\mathcal{G}$ where each node represents a user, and two nodes u and v are connected if and only if u directly replies to a post or a comment of v or vice versa. Each node u is also associated with a discrete label $a_u \in \{l_1, l_2, \ldots, l_n\}$ where n is the number of sides/ideologies in the considered controversy. Further, each edge (u, v) is described by a weight $w_{u,v} \in \mathbb{N}$ that represents the number of interactions between two users. In our case study, we model the interaction network as an undirected graph. Depending on the specific task it would make sense to use a directed network instead; however, such a choice mainly depends on how strict is the given definition of echo chamber. In our rationale, we believe that the echo chamber components should somehow be insulated from opposite views. Thus, neither incoming information from the rest of the network nor outgoing ones are allowed to a certain extent.

3.4. Step 4: Homogeneous, Meso-Scale, Users' Clusters Identification

Up to this step, we have described how to label a set of users with respect to their leaning on controversy and how to define their interaction network. In other words, we have identified the basic bricks to discuss both the *echo* and the *chamber* dimensions of a given phenomenon. However, conforming to the proposed Definition 1, we handle echo chambers as meso-scale topologies. Thus, we aim to detect subsets of nodes in the network that are homogeneous from an ideological and topological point of view.

As stated in Section 2.3, we model such a problem relying on a specific instance of the CD problem, namely Labeled Community Detection. Among the LCD algorithms, we leverage Eva [49], a bottom-up low complexity algorithm designed to identify network meso-scale topologies by optimizing structural and attribute-homophilic clustering criteria. From a structural point of view, Eva leverages the modularity score (Definition 2) to incrementally update community membership. Such an update is then weighted in terms of cluster Purity (Definition 3), another function tailored to capture the ideological cohesion of a community.

Definition 2 (Modularity). *Modularity is a quality score that measures the strength of the division of a network into modules. Formally:*

$$Q = \frac{1}{(2m)} \sum_{uv} \left[M_{uv} - \frac{k_u k_v}{(2m)} \right] \delta(c_u, c_v) \tag{1}$$

where *m* is the number of graph edges, $M_{u,v}$ is the entry of the adjacency matrix for $u, v \in V$, k_u, k_v the degree of u, v and $\delta(c_u, c_v)$ identifies an indicator function taking value 1 iff u, v belong to the same community *c*, 0 otherwise.

Definition 3 (Purity). *Given a community $c \in C$, its Purity is the product of the frequencies of the most frequent labels carried by its nodes. Formally:*

$$P_c = \prod_{a \in A} \frac{max(\sum_{u \in c} a(u))}{|c|} \tag{2}$$

where A is the label set, $a \in A$ is a label, $a(u)$ is an indicator function that takes value 1 iff $a \in A(u)$.

Once Eva has identified the candidate communities, we now discuss how to evaluate them to determine to what extent a community qualifies as an echo chamber. Since the analyzed phenomenon consists both of the *echo* and the *chamber* dimensions, we propose to evaluate them relying respectively on Purity and on Conductance (Definition 4).

Definition 4 (Conductance). *Given a community $c \in C$, its Conductance is the fraction of total edge volume that points outside the community.*

$$C_c = \frac{c_s}{2m_s + c_s} \tag{3}$$

where c_s is the number of community nodes and m_s is the number of community edges.

Thus, we propose to set ideological and topological constraints on P_c and C_c by means of two thresholds i.e., $P_c > p_0$ and $C_c < c_0$ where p_0 and c_0 can be tuned according to the strictness of the definition of echo chamber. For instance, in the Reddit case study, we set the Purity score ≤ 0.7 to make sure that most of the users in an echo chamber share the same ideological label. Meanwhile, for Conductance, we set a threshold equal to 0.5 to ensure that more than half of the total edges remain within the community boundaries.

Further, we can also define the risk for a community to be an echo chamber through a function that takes into account both Purity and Conductance. The most straightforward function is the difference between the two terms. In Figure 1 we show the results obtained by subtracting Conductance from Purity and then normalizing values in a range from 0 to 1. Intuitively, we notice that such a risk is maximized when Purity is equal to 1 and Conductance equal to 0.

Given the theoretical foundation set in this section, in the following we test our proposed framework on Reddit and discuss each of the pipeline steps.
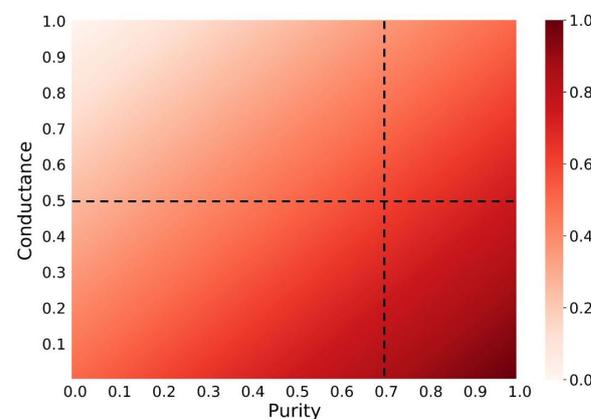

Figure 1. Risk for a community to be an echo chamber defined as the difference between Purity and Conductance. Red color implies a higher risk while white a lower one. The two dotted lines represent the constraints imposed in the Reddit case study.

4. Reddit: Politics, Gun Control, and Minorities Discrimination

This section discusses how to apply our proposed approach to Reddit data and provide further insights on the results. As stated by its slogan *'The front page of the internet'*, Reddit is a social platform that allows its users to post content to individual forums called *subreddits*, each dedicated to a specific topic. Currently, it is the nineteenth most visited website on the internet and the seventh in the USA (https://www.alexa.com/topsites (accessed on 1 June 2021)). We decide to build our approach upon Reddit mainly because it is the least explored from an echo chamber point of view. Further, since users can write anonymously and posts are not limited in length, this platform is particularly active in controversial discussions [50]. The data and the code used for this case study are available on a dedicated Github repository (https://github.com/virgiiim/EC_Reddit_CaseStudy (accessed on 1 June 2021)).

4.1. Controversial Issue Identification

To proceed with our approach, we first have to select a controversial topic, i.e., an issue in which single attitudes tend to diverge into ideological extremes. For such a purpose, we decide to focus on the debate between Trump supporters and Anti-Trump citizens during the first two and half years of Donald Trump's presidency (i.e., January 2017–July 2019). Indeed, the political rise of Donald Trump has further exacerbated the divide between Republicans and Democrats, making the debate even more polarized and uncivil [51]. To deal with this, we built a ground truth of polarized posts with respect to Pro-Trump and Anti-Trump beliefs, then we identified discussions themes likely to favor the formation of chambers related to such dichotomic attitude. In particular, we look for subreddits related to sociopolitical issues—GUN CONTROL, MINORITIES DISCRIMINATION—and general discussions on the POLITICAL SPHERE.

We retrieve Reddit data through the Pushshift API [52] that offers aggregation endpoints to explore Reddit activity from June 2005 to nowadays. More details about the ground truth dataset as well as the sociopolitical ones are given below. Further, in Table 2 we provide a description of each dataset in terms of number of selected subreddits, number of posts, and number of users.

Table 2. Description of the datasets.

Dataset	# Subreddit	# Post	# User
Ground Truth	3	302,762	68,113
GUN CONTROL	6	180,170	65,111
MINORITIES DISCRIMINATION	6	223,096	52,337
POLITICAL SPHERE	6	431,930	72,399

Polarized Ground Truth. To define a text classifier able to infer post ideology with respect to Trump stances, we need a ground truth of polarized posts. For such purpose, we rely on a set of subreddits, known to be either Pro-Trump or Anti-Trump. Thus, based on subreddits descriptions and on Reddit List (http://redditlist.com/\https://www.reddit.com/r/ListOfSubreddits/wiki/listofsubreddits (accessed on 1 June 2021)), we select posts belonging to r/The_Donald for the first group and r/Fuckthealtright and r/EnoughTrumpSpam for the second. To have a balanced dataset, for the Anti-Trump data we merge the last two subreddits, strictly related both on the users and the keywords (https://subredditstats.com/r/EnoughTrumpSpam; https://subredditstats.com/r/Fuckthealtright (accessed on 1 June 2021)).

For each selected submission, we collect the *id*, *author*, *self text*, and *title* fields, respectively, the identifier, the author username, the content, and the title of the submission (*id* and *author* where were pseudonymised through the adoption of an irreversible hash function during data collection). We merge contents and titles in a unique set to use as input for text classifier, because the *self text* of a post may be empty or just a reference to the *title* itself. By doing so, we make sure to have text capturing what the user is actually trying to convey. Then, we assign to each post a label of 1 if it belongs to the Pro-Trump subreddit,

0 otherwise. Further, we notice that several posts are composed of only a few words. This is probably because, during data extraction, we only select textual data, removing all multimedia content related to each submission. To avoid affecting classifier performance, we remove all posts shorter than six words from our original dataset.

Sociopolitical Topics. For each of the three topics in which we want to find evidence of an EC, we selected several subreddits via Reddit List. In such a way, we attempt to cover different points of view. For instance, for GUN CONTROL we select both subreddits that expressly support gun legalization and those against it. Meanwhile for MINORITIES DISCRIMINATION, we identify groups that promote gender/racial/sexual equality and those showing more conservative attitudes. Last, concerning POLITICAL SPHERE we try to cover different US political ideologies such as Republicans, Democrats, Liberals, and Populists. Then, for each of them, we define two datasets composed of all the posts and comments shared from January 2017 to July 2019. Regarding posts, we apply the same pre-processing of the ground truth, while for comments, we retrieve all the fields necessary to define the interaction network. These include *id*, *author*, *link_id*, *parent_id*, respectively, the comment's identifier, its author, the identifier of the post that this comment is in, the identifier of the parent of this comment.

4.2. Users' Ideology Inference

Once we have gathered data, the second step consists of inferring users' ideology on the controversy. As stated in Section 3.2, we model the task of predicting the political alignment of users' posts as a text classification problem. Since we consider two sides of the controversy in this case study (i.e., Pro-Trump and Anti-Trump), we model the text classification task as a binary problem.

Among the suitable NLP approaches discussed in Section 2.2, we test two different Deep Learning models (i.e., LSTM and BERT) which have been widely used to predict the political leaning both of OSN contents [53–55] and news articles [56,57]. Concerning LSTM, we have already carried out some experiments in a preliminary work [58]. In the following, for both models, we discuss the updated experimental setup, the model evaluation phase, and the prediction results on sociopolitical topics.

Experimental Setup. To train and test both models, we rely on the Polarized Ground Truth dataset defined in Section 4.1. To create the training and validation sets, we randomly select 80% of the whole dataset (242,762 posts) in such a way to guarantee the balancing between the Pro-Trump and Anti-Trump classes. The remaining 20% (60,000 posts) is used as the test set. During model selection, we perform a 3-fold Cross-Validation trying different hyper-parameters configurations of both models. For LSTM, we varied the number of LSTM units $[32, 64, 128]$ as well as the type of word embeddings with a fixed dimension of 100, i.e., GloVe pre-trained word embeddings and embeddings directly learned from the texts. In both scenarios, we vectorize each input submission creating a lexicon index based on word frequency, where 0 represents padding: indexes are assigned in descending order of frequency. Moreover, since the Out-of-Vocabulary (OOV) rate is rather low (i.e., 1.36%), we mark OOV tokens with a reserved index. In all settings, we use a dropout regularization of 0.3, adam optimizer, the sigmoid activation function, and as loss function, the binary cross-entropy. We obtain the best performances on the validation set using GloVe word embeddings and 128 LSTM units, obtaining an average accuracy of 82.9%. For BERT, we leverage the pre-trained model $BERT_{BASE}$ presented by Devlin et al. [39]. Specifically, we rely on the PyTorch implementation publicly released by Hugging Face [59]. Then, we varied the length of the input $[64, 128, 256, 512]$, the learning rate $[2 \times 10^{-5}, 3 \times 10^{-5}, 5 \times 10^{-5}]$, and the type of text pre-processing, i.e., with or without punctuation. We obtain the best results on the validation set, leaving punctuation, 512-token input (note that 512 tokens is also the maximum input length supported by BERT; nevertheless, in this case study, we deal with OSN posts that tend to be relatively short, i.e., only 0.4% of total posts are longer than the 512 limit), and a learning rate of 2×10^{-5}, reaching an average accuracy of 86.3%. Comparing the two models' validation results, we

notice that BERT reaches a definitely higher accuracy than LSTM. Thus, we select such a model for this case study.

Model Evaluation. We assess BERT performances on the test set, obtaining an accuracy of 85.6%. Then, to verify to what extent the model is able to generalize on less polarized posts, we further evaluate it on the sociopolitical datasets. However, in this scenario, we do not have any labeled data with respect to Trump's beliefs. Thus, we search for users belonging to our ground truth datasets, label them accordingly, and then apply the model to their posts. Table 3 shows model evaluation results for the test set and the three topics. Even if the model suffers from the domain change, it can generalize quite well, reaching an accuracy greater than 70% among all sociopolitical datasets. Despite the task of predicting the political affiliation of user/post having been successfully performed on popular OSNs (e.g., Facebook [60], Twitter [53–55]), we have not found any similar work on Reddit with which to compare our results.

Table 3. BERT performances on evaluation samples.

Evaluation Sample	# Post	Accuracy
Test Set	60,000	0.856
GUN CONTROL	6093	0.713
MINORITIES DISCRIMINATION	26,187	0.707
POLITICAL SPHERE	99,920	0.723

Predictions on Sociopolitical Topics. Once the expected accuracy of our model was evaluated, we applied it to each of the sociopolitical datasets in order to infer posts' leaning on the controversy for the entire population. For each post, we obtain model predictions ranging from 0 to 1 (i.e., the model confidence), where 1 means that the post aligns itself with Pro-Trump ideologies while 0 with Anti-Trump ones. Lastly, for each user u belonging to a specific topic we compute their *leaning score*, L_u, as the average value of their post's leaning as follows:

$$L_u = \frac{\sum_{i=1}^{n} \text{Prediction Score}(p_i)}{n} \tag{4}$$

where $p_i \in P_i$ is a post shared by a user u and $n = |P_i|$ is the cardinality of the set of u's posts. In Figure 2 we show the authors' leaning score distribution for each topic. Both MINORITIES DISCRIMINATION and POLITICAL SPHERE follow the typical U-shaped distribution of polarized issue, i.e., underlying a neat prevalence of extreme values. Thus, we can assert that in these two topics, we can find both sides of the controversy. On the other hand, GUN CONTROL users are strongly polarized with respect to Pro-Trump ideas and less sided with Anti-Trump citizens.
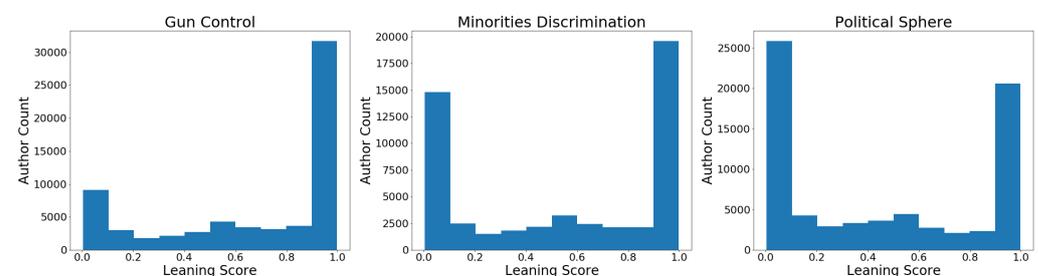

Figure 2. Authors' leaning score distribution for each sociopolitical topic.

4.3. Debate Network Construction

During the previous step, we observed that, across different sociopolitical issues, most users tend to assume a polarized position on the controversy rather than a moderate one. Starting from such insight provided by the analyzed data, we now have to answer

a more specific question: *Do the observed polarized users also tend to interact prevalently with like-minded individuals, or are they open to discussion with peers sharing opposing views?*

To answer such a question we define, for each topic, a proper users' debate network. To take into account the evolution of ideologies in time, we look for echo chambers on semester basis rather than in the whole period. Indeed, users may change their opinion on the controversy during two and a half years.

Accordingly, we define users' interaction network for each topic and semester following the approach proposed in Section 3.3. Each node of the network represents a user, and an edge between two users exists if one directly replies to a post or a comment of the other. We set each edge weight to represent the total number of comments exchanged between two users. We also label users (i.e., nodes) with their *leaning score* L_u. To do so, we discretize such leanings into three intervals: *Anti-Trump* if $L_u \leq 0.3$; *Pro-Trump* if $L_u \geq 0.7$; while *Neutral* if $0.3 < L_u < 0.7$. We add the third label, mainly because it is quite possible that some posts are not politically charged and thus not openly sided in the controversy. In Table 4, for each topic we provide average statistics of the networks across the five considered semesters.

Table 4. For each topic, network statistics averaged across semesters: size of the network in terms of nodes and edges, network density, number of users with a Pro-Trump, Anti-Trump, or Neutral leaning score.

Topic	# Nodes	# Edges	Density	# Pro-Trump	# Anti-Trump	# Neutral
Gun Control	12,609.4	151,813.0	0.0019	9165.6	3291.4	153.4
Minorities Discrimination	8133.6	134,362.6	0.0040	4544.2	2615.4	974.0
Political Sphere	8494.6	66,261.8	0.0018	3368.0	3935.0	1191.6

4.4. Homogeneous, Meso-Scale, Users' Clusters Identification

Once we have defined labeled interaction networks, we focus on discovering if they present cohesive meso-scale topologies both from a structural and ideological perspective. Following the rationale in Section 3.4, we rely on Eva, a CD algorithm belonging to Labeled Community Detection approaches. Eva is tailored to detect communities both maximizing their internal density through Modularity (see Definition 2) and their labels homogeneity relying on Purity (see Definition 3).

Accordingly, we apply Eva to each topic and semester, thus identifying our candidate communities. Then, we evaluate them by means of Conductance (see Definition 4) and Purity. In this case study, we set the Conductance score ≤ 0.5 to ensure that more than half of the total edges remain within the community boundaries. Meanwhile, for Purity, we set a threshold equal to 0.7 to make sure that most of the users in an echo chamber share the same ideological label. In Figure 3 we show the communities evaluation process for Gun Control, minorities discrimination, and Political Sphere topics. In each scatter plot, we can classify as echo chambers those communities that lie above the horizontal red line.

The difference in results between the three topics is stark. Gun Control (Figure 3a) does not present strongly polarized communities among different semesters. Indeed, on average, only 6.7% of total users fall in an echo chamber. Among them, 97.1% of members show a Pro-Trump tendency (i.e., in the first semester), while only the 2.9% have an Anti-Trump leaning (i.e., in the fourth semester). Moreover, we can observe that Pro-Trump users tend to form communities composed of like-minded individuals, even if not sufficiently ideologically homogeneous. Instead, Anti-Trump users mainly interact with the opposite side of the controversy, probably because they are a minority with respect to the totality of users (see Table 4). As regards minorities discrimination (Figure 3b), the overall scenario is definitely different. Indeed, on average, more than half of total users (i.e., 53.8%) are trapped in echo chambers. Further, we can observe in all semesters both Pro-Trump and Anti-Trump ECs, even if the first group outnumbered the second (i.e., 85% of Pro-Trump ECs members and 15% of Anti-Trump ones). Moreover, different from the other two topics, this one presents more than one EC of the same ideological leaning. On the other hand,

POLITICAL SPHERE (Figure 3c) shows a strong tendency toward Anti-Trump polarization (i.e., 23.3% of total users belong to Anti-Trump ECs). Indeed, Pro-Trump individuals are not ideologically homogeneous enough to be classified as echo chambers.


(**a**) Gun Control

(**b**) Minorities Discrimination
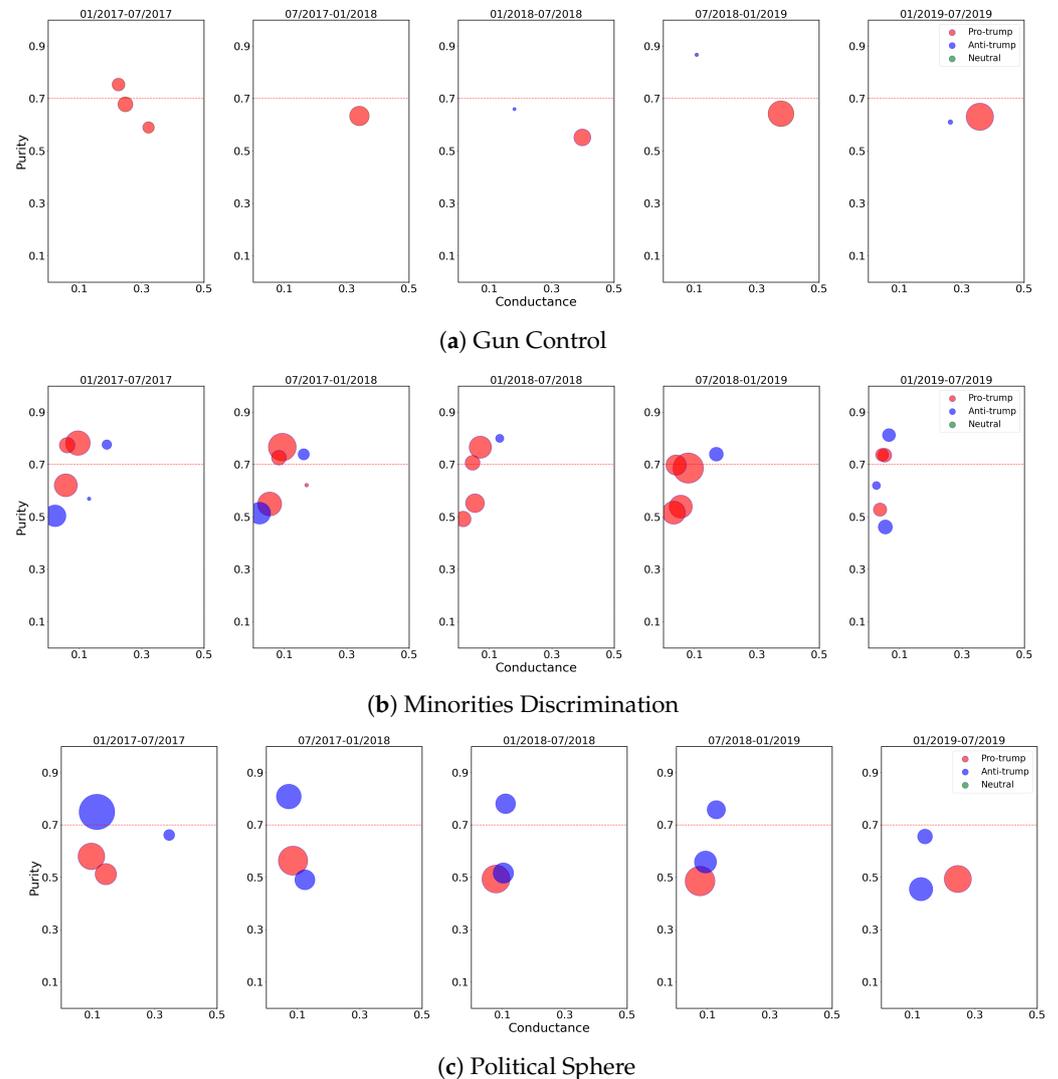
(**c**) Political Sphere

Figure 3. Communities evaluation for GUN CONTROL (**a**), MINORITIES DISCRIMINATION (**b**), and POLITICAL SPHERE (**c**) topics. The scatter plots display, for each semester, the Conductance (*x*-axis) and Purity (*y*-axis) scores for each detected community. Circles represent EVA communities, where red denotes communities populated by most Pro-Trump users, blue by Anti-Trump ones, and green Neutral. The horizontal red line marks the Purity threshold (0.7). Thus, the communities lying above it can be classified as *strong* echo chambers. Note that here, we are just plotting those communities that satisfy the Conductance constraint (0.5).

Furthermore, we have also noticed an interesting trend among all topics and semesters: Neutral users do not fall either in an EC or form a community with a Neutral majority. We further investigate such an aspect to verify if they tend to communicate prevalently with Pro-Trump or Anti-Trump users. We found that both in MINORITIES DISCRIMINATION and POLITICAL SPHERE topics, ~80% of Neutral users join Pro-Trump communities while, concerning GUN CONTROL, they are equally distributed among opposite-side communities. Following such results, we can assume that in this case study, having a Neutral leaning score means being somewhat undecided.

4.5. EC Analysis: Stability and Persistence over Time

Up to this point, we have explored users' leanings on the controversy (see Figure 2) and echo chambers (see Figure 3) in static temporal snapshots. Now, we are interested in analyzing their stability and consistency through time. In other words, we aim to answer the following questions:

i. *How do users' ideology evolve over time? Are users stable and consistent with the same ideology, or do they tend to change opinion?*

ii. *How do echo chambers evolve over time? Do members tend to fall again in a polarized community, or are they open to debate with opposite-leaning users?*

To answer both questions, we model such issues in terms of transition probabilities. In other words, for each user, we compute their probability p_{ij} to move from *state i* to *j* over contiguous semesters. In the first question, *state* stands for user ideology (i.e., Pro-Trump, Anti-Trump, and Neutral). In the second one, with *state* we refer to the leaning of community the user belongs to (e.g., Pro-Trump EC, Anti-Trump EC, Pro-Trump community, ...).

i. Ideology Stability over Time. In the heat maps in Figure 4, we show the stability of users' ideology over contiguous semesters. As regards GUN CONTROL, transition probabilities could explain why we have not found ECs or communities composed by a majority of Anti-Trump users. Indeed, users have an equal probability of remaining in their position and changing their leaning toward a Pro-Trump one, thus proving to be not strongly polarized. On the contrary, most Pro-Trump users seem to be rooted in their position, with a probability to remain in their state greater than 0.64 in all semesters. For MINORITIES DISCRIMINATION, the scenario is quite similar. Indeed, even if Anti-Trump users are more polarized across semesters (i.e., $p_{AA} \geq 0.5$), some of them change their opinion in favor of the opposite side of the controversy. On the contrary, POLITICAL SPHERE users aligned with Trump have a strong tendency to move to a Neutral position. However, except for the last couple of semesters, it is pretty unlikely that they change their leaning to Anti-Trump. Instead, with the exception of the last couple of semesters, Anti-Trump users are the ones with both the highest probability to remain in their state and lower to change in favor of Pro-Trump ideas (i.e., respectively $p_{AA} \geq 0.86$ and $p_{AP} \leq 0.04$). Lastly, for what concerns Neutral users, both GUN CONTROL and MINORITIES DISCRIMINATION insights confirm our hypothesis that they are somewhat undecided. Indeed, in both cases, they have a higher probability of changing state instead of remaining Neutral. On the other hand, POLITICAL SPHERE Neutral users are definitely more rooted in their position with a tendency toward Anti-Trump beliefs.

	Gun Control					Minorities Discrimination					Political Sphere			
A -> A	0.4	0.41	0.42	0.41		0.52	0.53	0.52	0.49		0.87	0.86	0.89	0.71
A -> N	0.15	0.17	0.15	0.16		0.16	0.17	0.15	0.17		0.089	0.1	0.092	0.26
A -> P	0.45	0.42	0.43	0.42		0.32	0.3	0.33	0.34		0.04	0.037	0.017	0.024
N -> A	0.28	0.36	0.31	0.33		0.3	0.31	0.35	0.29		0.44	0.31	0.33	0.3
N -> N	0.24	0.22	0.2	0.23		0.28	0.27	0.23	0.26		0.5	0.54	0.65	0.61
N -> P	0.48	0.42	0.48	0.44		0.42	0.42	0.42	0.45		0.058	0.15	0.02	0.091
P -> A	0.18	0.21	0.22	0.2		0.23	0.23	0.23	0.22		0.13	0.11	0.18	0.31
P -> N	0.12	0.15	0.13	0.12		0.15	0.14	0.16	0.17		0.57	0.5	0.53	0.46
P -> P	0.7	0.64	0.66	0.67		0.62	0.62	0.61	0.61		0.3	0.39	0.29	0.23

(columns: 1/17-7/17 - 7/17-1/18 · 7/17-1/18 - 1/18-7/18 · 1/18-7/18 - 7/18-1/19 · 7/18-1/19 - 1/19-7/19)

Figure 4. For each topic, transition probabilities of users' ideology over contiguous semesters. *A* stands for Anti-Trump users, *P* for Pro-Trump ones, and *N* for Neutral.

ii. Echo Chamber Stability over Time. To analyze ECs' consistency over time, we only take into account those two topics (i.e., MINORITIES DISCRIMINATION, POLITICAL SPHERE) in which we have detected an EC across different semesters. Indeed, GUN

CONTROL users fall in a Pro-Trump EC only in the first semester, thus proving to be not stable over time. Additionally, in this analysis, we do not consider the Neutral group of users since they never form an EC or aggregate in a community with a Neutral majority.

We decide to compute transition probabilities both for echo chambers (*EC*) and communities that do not satisfy ideological or topological constraints (*C*). In the heat maps, we show obtained results for MINORITIES DISCRIMINATION (Figure 5a) and POLITICAL SPHERE (Figure 5b). Differently from ideologies, in both topics, echo chambers prove to be strongly consistent over semesters.

(a) Minorities Discrimination
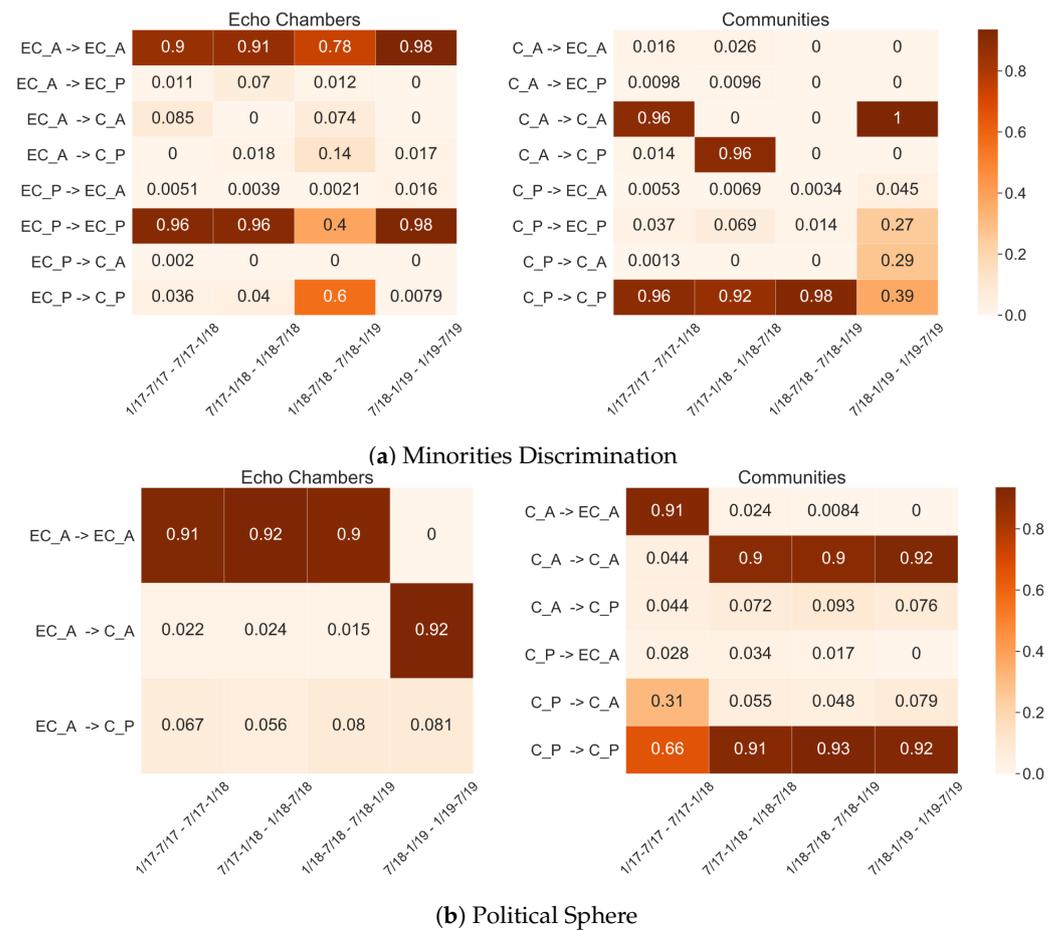
(b) Political Sphere

Figure 5. MINORITIES DISCRIMINATION (**a**) and POLITICAL SPHERE (**b**) transition probabilities for echo chamber (*EC*) and communities (*C*) over contiguous semesters.

Indeed, in most samples, ECs have a high probability to not change state, thus remaining in the same-leaning polarized systems. Further, in the few cases in which ECs change state, they never move to a community with most opposite-leaning users. On the contrary, communities are less stable over contiguous semesters. Moreover, with respect to echo chambers, they show a higher probability to change state in favor of the opposite side of the controversy.

5. Discussion, Conclusions and Future Works

In this work, we proposed a formal definition of echo chamber and a general framework to assess their existence in online social networks. In doing so, we performed two choices that make our study quite different from the previous. Firstly, we handle echo chambers as meso-scale topologies. In other words, to detect ECs we do not rely on a single user's digital traces as in [14,16] (i.e., following a *micro-scale approach*), nor suppose that all users in the network belong to a polarized community as in [18–22] (i.e., assuming a *macro-scale approach*). We focus on such peculiar EC since we are interested in the role

that group dynamics play in increasing polarization. Further, we believe that it is quite unrealistic that all the users involved in a controversial debate insulate themselves in an echo chamber.

Secondly, our framework is built upon features and resources commonly shared by most social networks, thus allowing its applicability across different OSNs and domains. In detail, such an approach consists of four main steps, namely: (i) controversial issue identification; (ii) users' ideology inference; (iii) debate network construction; (iv) homogeneous meso-scale users' clusters identification. Consequently, the basic bricks that we should find on OSNs to apply such an approach are the presence of discussions about controversial topics; the *post* feature in order to infer users ideology on the controversy based on how they write about it; the *comment* feature to define users debate network. To the best of our knowledge, all of these requirements are satisfied by the most popular OSNs (e.g., Twitter, Facebook, Reddit, Gab).

Moreover, we applied our framework in a detailed case study on Reddit covering the first two and a half years of Donald Trump's presidency (January 2017–July 2019). In such settings, our main aim was to assess the existence of Pro-Trump and Anti-Trump ECs among three sociopolitical issues. As concerns users' ideology, we have found that in the overall period, users tend to assume strongly polarized positions on the controversy rather than moderate ones across all topics. However, analyzing the stability of users' ideology over contiguous semesters, we noticed that users are not so rooted in their positions as expected. Indeed, an exception is made for Anti-Trump users discussing political issues, for which the majority of users have an ~30/40% probability of changing their leaning.

Regarding echo chambers, we have found that for GUN CONTROL, MINORITIES DISCRIMINATION, and POLITICAL SPHERE only 6.7%, 53.8%, 23.3% of total users fall in polarized communities. Among them, we observed that both the first and the second topics have a stronger tendency toward Pro-Trump beliefs, while the third to Anti-Trump ones. Additionally, we also assess ECs' stability and consistency over contiguous semesters, finding that EC members have a higher probability of interacting with like-minded individuals with respect to the other communities in the overall network.

Comparing our results to the ones obtained by the only EC detection work on Reddit [20], we find both commonalities and differences. Even if the authors focus on a different period (i.e., 2016 presidential elections) and on a slightly different controversy (i.e., Republicans vs. Democrats), we also noticed that Reddit users, compared to those of other OSNs, show a lower tendency to insulate themselves from opposite viewpoints. This attitude could be attributable to the Reddit structure, which is more a social forum than a traditional social network (e.g., Twitter, Facebook). However, differently from us, they conclude that Reddit political interactions do not resemble an echo chamber at all. Such a difference could be imputable to the difference in scale between approaches. Indeed, authors have identified ECs looking at the users' interaction network on an aggregated level, thus not considering differences within specific meso-scale network regions.

Approach weaknesses and limitations. As with all frameworks, our proposal suffers of a few known limitations and weaknesses that need to be carefully taken into account while instantiating it. A first limitation lies in the loss of contextual details derivable from platform-specific features. Indeed, we define and identify echo chambers by means of common features and resources shared by multiple platforms, thus providing, in some sense, a high-level representation of the EC phenomenon. However, generalizability does not come for free and, further, we believe that *posts* and *comments* are good proxies for respectively inferring users' ideology and defining the debate network. Moreover, we acknowledge that different data sources, although possessing the set of features required by our framework, can require context-specific tuning for each proposed step. In particular, to infer the political leaning from textual data, we assume that ad hoc model selection and fine-tuning have to be performed to account for data source peculiarities. Political leaning classifiers are hardly transferable among different contexts, and we can assume that a "no free lunch" solution exists to this challenging task. Finally, another limitation of the

proposed framework lies in the absence of a general rule to select the proper community discovery algorithm to identify ECs. Community Discovery is an ill-posed problem, and alternative algorithmic solutions are known to optimize different quality function differences that highly affect the resulting node clustering. In this paper, we opted for a model, Eva, designed to balance both topological and semantic information while partitioning the social graph. Indeed, such a choice, although reasonable, is not the only valid alternative to address node cluster identification.

On the other hand, the major weakness of the proposed framework lies in the structural absence of a strong result validation strategy. Such an issue—affecting the majority of EC detection and political polarization studies—lies in the absence of reliable ground truth for individuals' political leaning labels. As in other works [13,16,18–21], leaning labels are assigned making strong assumptions on the political orientation of users that take part to the discussions of polarized clusters—not separating real supporters from their opponents and trolls, nor relying on users' provided information. Indeed, the absence of a ground truth annotation does not allow making a final observation and, rather, to underline that—considering the chosen methodological proxy—ECs emerge.

Research outlooks. As future research directions, we have both short- and long-term plans. Firstly, to further support our claim of framework generalizability, we are currently designing other case studies on popular OSNs (e.g., Twitter, Facebook, Gab). Consequently, we would like to perform a comparative analysis of obtained results in such a way as to assess if some environments are more polarized than others. Secondly, given the information retrieved in the previous step, we would like to characterize ECs by describing their DNA (e.g., the characteristics of the users composing them in terms of online activities) and use such footprints to design an echo chamber-aware Recommendation System able to foster pluralistic viewpoints in suggestions.

Abbreviations

The following abbreviations are used in this manuscript:

EC	echo chamber
OSN	online social network
RNN	Recurrent Neural Network
LSTM	Long Short-Term Memory
CNN	Convolutional Neural Network
BERT	Bidirectional Encoder Representations from Transformers
CD	Community Detection
LCD	Labeled Community Detection
NLP	Natural Language Processing
OOV	Out-of-Vocabulary

References

1. Kim, Y.; Hsu, S.H.; de Zúñiga, H.G. Influence of social media use on discussion network heterogeneity and civic engagement: The moderating role of personality traits. *J. Commun.* **2013**, *63*, 498–516. [CrossRef]
2. Brundidge, J. Encountering "difference" in the contemporary public sphere: The contribution of the Internet to the heterogeneity of political discussion networks. *J. Commun.* **2010**, *60*, 680–700. [CrossRef]
3. Sunstein, C.; Sunstein, C.R. *# Republic*; Princeton University Press: Princeton, NJ, USA, 2018.
4. Sunstein, C.R. *Republic.Com 2.0*; Princeton University Press: Princeton, NJ, USA, 2007.
5. Pariser, E. *The Filter Bubble: What the Internet Is Hiding from You*; Penguin UK: London, UK, 2011.
6. Smith, H.J. Ethics and information systems: Resolving the quandaries. *ACM Sigmis Database Database Adv. Inf. Syst.* **2002**, *33*, 8–22. [CrossRef]
7. Tam, K.Y.; Ho, S.Y. Understanding the impact of web personalization on user information processing and decision outcomes. *MIS Q.* **2006**, 865–890. [CrossRef]
8. O'Connor, C.; Weatherall, J.O. *The Misinformation Age: How False Beliefs Spread*; Yale University Press: New Haven, CT, USA, 2019.
9. Gilbert, E.; Bergstrom, T.; Karahalios, K. Blogs are echo chambers: Blogs are echo chambers. In Proceedings of the 2009 42nd Hawaii International Conference on System Sciences, Big Island, HI, USA, 5–8 January 2009; pp. 1–10.
10. Hilbert, M.; Ahmed, S.; Cho, J.; Liu, B.; Luu, J. Communicating with algorithms: A transfer entropy analysis of emotions-based escapes from online echo chambers. *Commun. Methods Meas.* **2018**, *12*, 260–275. [CrossRef]
11. Ge, Y.; Zhao, S.; Zhou, H.; Pei, C.; Sun, F.; Ou, W.; Zhang, Y. Understanding echo chambers in e-commerce recommender systems. In Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, Xi'an, China, 25–30 July 2020; pp. 2261–2270.
12. Nguyen, T.T.; Hui, P.M.; Harper, F.M.; Terveen, L.; Konstan, J.A. Exploring the filter bubble: The effect of using recommender systems on content diversity. In Proceedings of the 23rd International Conference on World Wide Web, Seoul Korea, 7–11 April 2014; pp. 677–686.
13. Quattrociocchi, W.; Scala, A.; Sunstein, C.R. Echo Chambers on Facebook. Available online: https://ssrn.com/abstract=2795110 (accessed on 1 June 2021).
14. Bakshy, E.; Messing, S.; Adamic, L.A. Exposure to ideologically diverse news and opinion on Facebook. *Science* **2015**, *348*, 1130–1132. [CrossRef]
15. Dubois, E.; Blank, G. The echo chamber is overstated: The moderating effect of political interest and diverse media. *Inform. Commun. Soc.* **2018**, *21*, 729–745. [CrossRef]
16. An, J.; Quercia, D.; Crowcroft, J. Partisan sharing: Facebook evidence and societal consequences. In Proceedings of the Second ACM Conference on Online Social Networks, Dublin, Ireland, 1–2 October 2014; pp. 13–24.
17. Grömping, M. 'Echo Chambers' Partisan Facebook Groups during the 2014 Thai Election. *Asia Pac. Media Educ.* **2014**, *24*, 39–59. [CrossRef]
18. Barberá, P.; Jost, J.T.; Nagler, J.; Tucker, J.A.; Bonneau, R. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychol. Sci.* **2015**, *26*, 1531–1542. [CrossRef]
19. Garimella, K.; De Francisci Morales, G.; Gionis, A.; Mathioudakis, M. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In Proceedings of the 2018 World Wide Web Conference, Lyon, France, 23–27 April 2018; pp. 913–922.
20. Morales, G.D.F.; Monti, C.; Starnini, M. No echo in the chambers of political interactions on Reddit. *Sci. Rep.* **2021**, *11*, 1–12.
21. Cinelli, M.; Morales, G.D.F.; Galeazzi, A.; Quattrociocchi, W.; Starnini, M. The echo chamber effect on social media. *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2023301118. [CrossRef]
22. Conover, M.; Ratkiewicz, J.; Francisco, M.; Gonçalves, B.; Menczer, F.; Flammini, A. Political polarization on twitter. In Proceedings of the International AAAI Conference on Web and Social Media, Barcelona, Spain, 17–21 July 2011; Volume 5.
23. Iyengar, S.; Hahn, K.S. Red media, blue media: Evidence of ideological selectivity in media use. *J. Commun.* **2009**, *59*, 19–39. [CrossRef]

24. Dori-Hacohen, S.; Allan, J. Detecting controversy on the web. In Proceedings of the 22nd ACM International Conference on Information & Knowledge Management, San Francisco, CA, USA, 27 October–1 November 2013; pp. 1845–1848.

25. Jang, M.; Foley, J.; Dori-Hacohen, S.; Allan, J. Probabilistic approaches to controversy detection. In Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, Indianapolis, IN, USA, 24–28 October 2016; pp. 2069–2072.

26. Mejova, Y.; Zhang, A.X.; Diakopoulos, N.; Castillo, C. Controversy and sentiment in online news. In Proceedings of the Computation+ Journalism Symposium, New York, NY, USA, 24–25 October 2014.

27. Kowsari, K.; Jafari Meimandi, K.; Heidarysafa, M.; Mendu, S.; Barnes, L.; Brown, D. Text classification algorithms: A survey. *Information* **2019**, *10*, 150. [CrossRef]

28. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient Estimation of Word Representations in Vector Space. In Proceedings of the 1st International Conference on Learning Representations, ICLR 2013, Scottsdale, AZ, USA, 2–4 May 2013.

29. Pennington, J.; Socher, R.; Manning, C.D. Glove: Global vectors for word representation. In Proceedings of the Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October2014; pp. 1532–1543.

30. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef] [PubMed]

31. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

32. LeCun, Y.; Haffner, P.; Bottou, L.; Bengio, Y. Object recognition with gradient-based learning. In *Shape, Contour and Grouping in Computer Vision*; Springer: Berlin/Heisenberg, Germany, 1999; pp. 319–345.

33. Schmidhuber, J. Multi-column deep neural networks for image classification. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 3642–3649.

34. Ciresan, D.C.; Meier, U.; Masci, J.; Gambardella, L.M.; Schmidhuber, J. Flexible, high performance convolutional neural networks for image classification. In Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence, Barcelona, Spain, 16–22 July 2011.

35. Collobert, R.; Weston, J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland, 5–9 July 2008; pp. 160–167.

36. Grefenstette, E.; Blunsom, P. A convolutional neural network for modelling sentences. In Proceedings of the ACL, Baltimore, MA, USA, 22–27 June 2014.

37. Shen, Y.; He, X.; Gao, J.; Deng, L.; Mesnil, G. Learning semantic representations using convolutional neural networks for web search. In Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea, 7–11 April 2014; pp. 373–374.

38. Kim, Y. Convolutional Neural Networks for Sentence Classification. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; Association for Computational Linguistics: Stroudsburg, PA, USA, 2014; pp. 1746–1751.

39. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Minneapolis, MIN, USA, 2–7 June 2019; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; Volume 1, pp. 4171–4186.

40. Liu, Y.; Ott, M.; Goyal, N.; Du, J.; Joshi, M.; Chen, D.; Levy, O.; Lewis, M.; Zettlemoyer, L.; Stoyanov, V. Roberta: A robustly optimized bert pretraining approach. *arXiv* **2019**, arXiv:1907.11692.

41. Sanh, V.; Debut, L.; Chaumond, J.; Wolf, T. DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter. *arXiv* **2019**, arXiv:1910.01108.

42. Radford, A.; Narasimhan, K.; Salimans, T.; Sutskever, I. Improving Language Understanding by Generative Pre-Training. 2018. Available online: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf (accessed on 1 June 2021).

43. Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I. Language models are unsupervised multitask learners. *OpenAI Blog* **2019**, *1*, 9.

44. Minaee, S.; Kalchbrenner, N.; Cambria, E.; Nikzad, N.; Chenaghlu, M.; Gao, J. Deep Learning–based Text Classification: A Comprehensive Review. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–40. [CrossRef]

45. Plantié, M.; Crampes, M. Survey on social community detection. In *Social Media Retrieval*; Springer: Berlin/Heisenberg, Germany, 2013; pp. 65–85.

46. Cossard, A.; Morales, G.D.F.; Kalimeri, K.; Mejova, Y.; Paolotti, D.; Starnini, M. Falling into the echo chamber: The Italian vaccination debate on Twitter. In Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, GA, USA, 11–15 January 2020; Volume 14, pp. 130–140.

47. Vu, N.T.; Adel, H.; Gupta, P.; Schütze, H. Combining recurrent and convolutional neural networks for relation classification. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, San Diego, CA, USA, 12–17 June 2016.

48. Garimella, K.; Morales, G.D.F.; Gionis, A.; Mathioudakis, M. Quantifying controversy on social media. *ACM Trans. Soc. Comput.* **2018**, *1*, 1–27. [CrossRef]

49. Citraro, S.; Rossetti, G. Eva: Attribute-aware network segmentation. In Proceedings of the International Conference on Complex Networks and Their Applications, Lisbon, Portugal, 10–12 December 2019; pp. 141–151.

50. Medvedev, A.N.; Lambiotte, R.; Delvenne, J.C. The anatomy of Reddit: An overview of academic research. *Dynamics on and of Complex Networks*; Springer: Berlin/Heisenberg, Germany, 2017; pp. 183–204.

51. Bartels, L.M. Partisanship in the Trump era. *J. Politics* **2018**, *80*, 1483–1494. [CrossRef]

52. Baumgartner, J.; Zannettou, S.; Keegan, B.; Squire, M.; Blackburn, J. The pushshift reddit dataset. In Proceedings of the International AAAI Conference on Web and Social Media, Atlanta, GA, USA, 8–11 June 2020; Volume 14, pp. 830–839.

53. Ansari, M.Z.; Aziz, M.; Siddiqui, M.; Mehra, H.; Singh, K. Analysis of political sentiment orientations on twitter. *Procedia Comput. Sci.* **2020**, *167*, 1821–1828. [CrossRef]

54. Ayata, D.; Saraçlar, M.; Özgür, A. Political opinion/sentiment prediction via long short term memory recurrent neural networks on Twitter. In Proceedings of the 2017 25th Signal Processing and Communications Applications Conference (SIU), Antalya, Turkey, 15–18 May 2017; pp. 1–4.

55. Mapes, N.; White, A.; Medury, R.; Dua, S. Divisive language and propaganda detection using multi-head attention transformers with deep learning BERT-based language models for binary classification. In Proceedings of the Second Workshop on Natural Language Processing for Internet Freedom: Censorship, Disinformation, and Propaganda, Hong Kong, China, 3–4 November 2019; pp. 103–106.

56. Baly, R.; Da San Martino, G.; Glass, J.; Nakov, P. We Can Detect Your Bias: Predicting the Political Ideology of News Articles. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Online, 16–20 November 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 4982–4991.

57. Stefanov, P.; Darwish, K.; Atanasov, A.; Nakov, P. Predicting the topical stance and political leaning of media using tweets. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Online, 5–10 July 2020; pp. 527–537.

58. Morini, V.; Rossetti, G.; Pollacci, L. Capturing Political Polarization of Reddit Submissions in the Trump Era. In Proceedings of the Italian Symposium on Advanced Database Systems, Villasimius, Italy, 21–24 June 2020.

59. Wolf, T.; Debut, L.; Sanh, V.; Chaumond, J.; Delangue, C.; Moi, A.; Cistac, P.; Rault, T.; Louf, R.; Funtowicz, M.; et al. Transformers: State-of-the-Art Natural Language Processing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations, Online, 3 June 2020; Association for Computational Linguistics: Stroudsburg, PA, USA, 2020; pp. 38–45.

60. Chang, C.C.; Chiu, S.I.; Hsu, K.W. Predicting political affiliation of posts on Facebook. In Proceedings of the 11th International Conference on Ubiquitous Information Management and Communication, Beppu, Japan, 5–7 January 2017; pp. 1–8.