University of Milano-Bicocca

Department of Informatics, Systems and Communication

Master's degree program in Data Science

Analysis of dynamic metrics for the study of echo chambers in social media

Supervisor: Marco Viviani

Co-supervisor: Davide Mancino

Master's degree thesis by:

Marta Bernardelli

851178

Academic Year 2023-2024

Contents

List of Figures

Abstract

In recent years, the vast amount of data disseminated online has lead to the phenomenon of information overload affecting individuals. To manage this problem, customized filters and algorithms are being adopted to help users navigate by relevant content according to their preferences. However, the use of such algorithms in social media, now major sources of information, can lead to the formation of so called “echo chambers” closed virtual environments where people interact primarily with likeminded individuals who share the same opinions and beliefs.

This thesis proposes a dynamic analysis of polarization phenomena by leveraging state-of-the-art echo chamber detection metrics. Focusing on major controversial issues from the past decade, the study examines data from multiple platforms, including Twitter and Reddit, across both short- and long-term topics. The primary goal is to investigate the formation and evolution of echo chambers over time, particularly in response to significant triggering events. By doing so, the research seeks to uncover common patterns and varying rates of polarization associated with different controversial topics.

1. Introduction

The possibility of accessing more information than ever before, coupled with the increased popularity of social media as both an entertainment and information tool, has raised numerous concerns about how this digital revolution may have impacted the way society informs itself and creates its own opinions. The vast amount of data available and the risk of information overload force individuals to use filters and customisation mechanisms to select the information they need. However, the use of such algorithms in social media can lead to the formation of so-called "echo chambers" closed virtual environments where people interact primarily with like-minded individuals who share the same opinions and beliefs. This phenomenon has various implications and consequences that must be understood, especially due to its significant influence on the process of opinion formation. One of the main risks is the potential for a distorted view of reality, leading to a decline in the quality of the news we are exposed to. In this scenario, echo chambers prove to be particularly fertile grounds for the spread of false information. It is therefore vital to analyse this phenomenon and fully understand its dynamics so that mitigation techniques can be put in place.

1.1 The Context

The origins and effects of the echo chamber phenomenon have been extensively analysed in the literature. Over the past decade, much research has been invested in understanding the origins and factors influencing this phenomenon. Indeed, an attempt was made to understand how much the digital revolution may have impacted on the spread of this phenomenon since it is closely related to concepts well before the advent of digitisation such as homophily. Different types of platforms (blogs, forums, social media) and different topics of conversation, particularly the medical field and politics, were also investigated.

Further work has focused on individual users to understand the relationship between personality and echo chambers. In particular, they sought to understand which personality

traits were most prevalent within echo chambers and how certain particular individuals interact with this type of dynamic, such as bipartisans and gatekeepers.

However, the area of greatest interest in the literature is certainly that of echo chamber detection techniques. Indeed, algorithms have been presented that allow the identification of the communities most at risk by taking into account the structural conformation of the conversation graph, the content expressed by the posts of the selected users, or both. Analyses have also been carried out to understand the generalisation capacity of these methodologies, performing analyses on different social platforms as well as on different types of domains and contexts. What, on the other hand, has still been little contemplated in the literature is the involvement of the time variable and in particular how the phenomenon of echo chambers forms and evolves over time according to the type of event considered.

1.2 The Contribution

This thesis work proposes from the study of state-of-the-art echo chamber detection metrics to conduct a dynamic type of analysis of polarization phenomena on some of the major controversial issues of the last decade. In particular both short- and long-term topics will be considered in order to gather quantitative evidence on both types of events. The former refer to particular events that often make people talk about themselves for only a few weeks, while the latter refers to major debates that have been going on for decades. In order to obtain a comparison also from a platform perspective, the same event will be analysed with data from both Twitter and Reddit. Twitter is a microblogging platform where users can write short messages and use hashtags to contribute to the conversation on specific topics, on the other hand, Reddit is a forum-like platform organised in subreddits (sub-forums), each dedicated to a multitude of topics and communities. With Twitter's recent change of ownership resulting in a name change to X, it was decided to keep the old platform name for a matter of recognisability and continuity in relation to previous studies. Once the most significant time intervals have been identified for each event, different types of metrics will be monitored according to the three most important types of

approaches for echo chamber detection: Micro-scale, Meso-scale, and Macro-scale.

1.3 Organization of the Work

The rest of this thesis work will be structured as follows:

- **Chapter 1: Background Concepts**. The aim of this chapter is to present the basic concepts, fundamental to understanding the experimental pipeline adopted in this thesis. In the first part of the chapter, the context in which the concepts of filter bubbles and echo chambers emerged is introduced. In addition, the origins and formal definitions of these phenomena will be recounted, and some implications and strategies to mitigate their effects will also be presented. The second part of the chapter presents the main tools of Social Media Analysis needed to identify echo chambers within social media. In particular, the techniques of Social Network Analysis and Social Content Analysis will be presented.

- **Chapter 2: Related Works on Echo Chambers.** This chapter illustrates the state-of-the-art primary methods for detecting echo chambers. It begins by summarizing some of the key research conducted in the field of echo chambers and then concentrates later on the echo chamber detection task. In this part of the chapter, the main criteria for classifying echo chamber detection methodologies are introduced with the goal of explaining the theoretical approaches that will be most relevant for the experimental section.

- **Chapter 3: An Analysis of Echo Chamber Detection over Time.** The purpose of this chapter is to introduce the experimental part of this thesis. The objectives of the research will be presented, as well as the work pipeline that was conceived at each of its stages: from the definition of the analysis events, passing through the modelling of the datasets, ending with the actual analysis methodology. Finally, the datasets taken into consideration will be presented.

- **Chapter 4: Experimental Evaluation.** This chapter will present the results obtained from the application of the previously illustrated methodology and highlight

the main insights gained from the analysis of echo chambers from a temporal perspective for both short-term and long-term events.

- **Conclusions and Further Research.** In this chapter, the work is summarised and limitations and possible future developments are discussed.

2. Background Concepts

This chapter aims to present the basic concepts that will be useful in fully understanding the context and the fundamental tools for the experimental part of this thesis. In the first section (Section 2.1) the concepts of Echo Chamber and Filter Bubble will be presented in order to understand their meaning, but also to highlight their similarities and differences. The second section (Section 2.2) will present the tools needed to carry out a comprehensive and relevant analysis in the field of Social Media. In particular, as far as Social Network Analysis is concerned, there will be an outline of Graph Theory and Community Detection Algorithms. The objective will be to provide the elements for understanding the structure of relations within Social Media. As regards Social Content Analysis, Sentiment Analysis, and Topic Modeling techniques will be presented in order to make the most of the content expressed by users.

2.1 The Concepts of Echo Chambers and Filter Bubbles in the Information Age

This section examines the technological and information context (Subsection 2.1.1) in which the concepts of filter bubble and echo chamber emerge (Subsection 2.1.2). Subsequently, the implications of these phenomena and some strategies designed to handle them are discussed (Subsection 2.1.3).

2.1.1 Context Analysis

In today's Information Age [1], individuals are inundated with a vast array of online content. Suffice it to say that enough data are generated each day to fill all the American libraries more than eight times. Thanks to ICT, we have entered the age of the zettabyte (10^{21}) [1], and this number is set to grow exponentially and uninterruptedly for the foreseeable future. According to the Word Economic Forum in 2020, we have touched 44

zettabytes, which means that there are 40 times more bytes than stars in the observable part of our universe [2].

What marked the beginning of the Information Age was definitely the birth of the Internet and in particular the birth of one of the most important resources provided by Internet: the Web. The arrival of the Web has profoundly revolutionized the ways we communicate and consume information, eliminating limits related to distance and time. In particular, one of its evolutionary stages called Web 2.0 [3] can be traced back to the origin of social media, platforms that have increased the possibilities for communication and content sharing among individuals. It differs from Web 1.0, which instead consisted solely of static Web pages, where the user could not interact with the content in any way.

Although the advent of Web 2.0 has given rise to various social media, Andreas Kaplan and Michael Haenlein [4] have identified the main four characteristics that a social media must have to be defined as such:

- Social media are interactive Web 2.0 applications.
- Content is generated by users, such as text posts or comments, digital photos or videos, and other data generated through all online interactions
- Users create service-specific profiles for the website or app designed and managed by the social media organization
- Social media facilitates the development of online social networks by linking a user's profile with those of other individuals or groups

Having a medium that allows for rapid and instantaneous communication without geographical limits has changed not only our personal relationships and entertainment, but also everything in the world of news and information.

Considering the analysis conducted in 2023 by the European Parliament on media habits [5] it can be seen that social media is on par with radio as the primary source of information for 37% of respondents, preceded only by news platforms at 42% and television at 71%. What is interesting to note, however, is that compared to 2022, the primary information figure for social media has increased by as much as 11 percentage

points. Also interesting to note is how citizens see online news platforms and social media channels, including influencers, are increasingly trusted sources.

2.1.2 Filter Bubbles and Echo Chambers

The vast amount of data that characterized this decade has created a phenomenon of *information overload* for individuals. The term was coined by Bertram Gross [6] but it became popular with Toffler [7] who said that *Information overload occurs when the amount of input to a system exceeds its processing capacity. Decision-makers have fairly limited cognitive processing capacity. Consequently, when information overload occurs, it is likely that a reduction in decision quality will occur*. More specifically, Klapp [8] defined information overload as *an excessive amount of information that the receiver can no longer process efficiently without distraction, stress, increased errors, or other costs that reduce the efficient use of the information.*

To manage this problem, customized filters and algorithms are being adopted to help users navigate relevant content according to their preferences. The goal of platforms is to develop systems that help the user to identify information relevant to their needs. In this scenario, two are the main systems used to select and access relevant information:

- **Information Retrieval Systems (IRS, Search Engines)**. It consists of a software program that facilitates a user in finding the information the user needs. The gauge of the success of an information system is how well it can minimize the cost for a user to find the needed information [9]. They require the formulation of a query to translate a user's information needs into an expression of a formal query language. Usually, however, in order to make the user's search result more accurate, some context information is added to that given by the individual query.

- **Information Filtering Systems (IFS, Recommender Systems)**. Unlike search engines, they require user profiles, i.e., descriptions of specific needs dynamically updated, also based on the user behavior and not by the formulation of a query.

However, these algorithms that greatly simplify our digital experience have consequences. This type of technology offers us many advantages and plays a huge role in our

daily online lives. Having said that, it is also important to be aware of how information is filtered to us, but also what selection criteria are followed by these algorithms, and how these choices affect our lives.

A hint is given by Mark Zuckerberg when speaking to a journalist about the logic of Facebook's news feed, he said: *"A squirrel dying in front of your house may be more relevant to your interests right now than people dying in Africa"*. So Facebook is showing us what it thinks we want to see, but not necessarily what we need to see [10].

This biased selection done by the algorithms makes everyone live in a so-called "*filter bubble*" as defined by Eli Pariser [11]. Those algorithms create " a unique universe of information for each of us which fundamentally alters the way we encounter ideas and information". According to Pariser, the choices made by the filters instead of a "balanced information diet", can end up surrounding us with "information junk food". The main problem with filters is our human tendency to think that what we see is all there is, without realizing that what we see is being filtered, so a wrong "information diet" can have bad consequences on people's thoughts and actions.

The concept of filter bubble is often associated and confused with that of *echo chamber*, a closed virtual environment where people interact primarily with like-minded individuals who share the same opinions and beliefs. A possible definition of echo chamber is the one provided in the literature by Bruns [12]: *An echo chamber comes into being where a group of participants chooses to preferentially connect with each other, to the exclusion of outsiders. The more fully formed this network is [...] the more isolated from the introduction of outside views is the group, while the views of its members are able to circulate widely within it*.

Echo chambers primarily stem from the structural characteristics of networks, while filter bubbles predominantly arise from the behavioral tendencies exhibited by individuals within those networks. This does not mean that there is no connection between the two concepts: the existence of an echo chamber means that it becomes much easier for a filter bubble to emerge, but it does not automatically imply the opposite [12].

For both phenomena, there are several studies confirming that user experience is limited nowadays. For instance, according to a 2016 Pew Center study [13], conducted during

the presidential election, showed that the 23% of American Facebook users and 17% of American Twitter users say that most of their contacts' views are similar to their own. In addition, 20% have changed their minds about a political or social issue because of interactions on social media. At the same time, 39% of social media users say they have changed their settings to filter out political posts or block certain users in their network; this could definitely be perceived as an effort to create a personal echo chamber, but it also indicates according to Burns that the filtering mechanisms are still quite ineffective [12].

Moreover, both phenomena hark back to human tendencies that predate the appearance of the Web. Cognitive and social psychology has long identified the phenomena that may be behind filter bubbles and echo chambers. For example, considering the individual's point of view, the cognitive tendency that can be related to filtering bubbles is *confirmation bias*: this is the tendency to focus only on information that confirms our prior beliefs, instead of information that provides us with an alternative point of view [14]. It is a phenomenon that although on different intensities characterizes each human being.

Instead, in the context of echo chambers the confirmation bias is added to other group dynamics such as *homophily* and *polarization*. The first is the tendency to surround oneself with like-minded people, and this can lead to the formation of homogeneous groups in terms of opinions, the second one happens when in a group whose members have similar starting points and the discussion eventually causes people to arrive at beliefs more extreme than those they had at the beginning, so that mutual reinforcement drives homogeneous groups to radicalize [15].

In light of this line of reasoning, the question arises as to what extent the web has increased the spread of this phenomenon, in this regard Corazza [16] says that it is not possible to establish with certainty because while we have a lot of data on user interactions on social media, the same cannot be said for traditional forms of communication. Despite that Gentzkow and Shapiro [17] try to analyze online news consumption by constructing an Isolation index, the bigger the more there is ideological isolation.

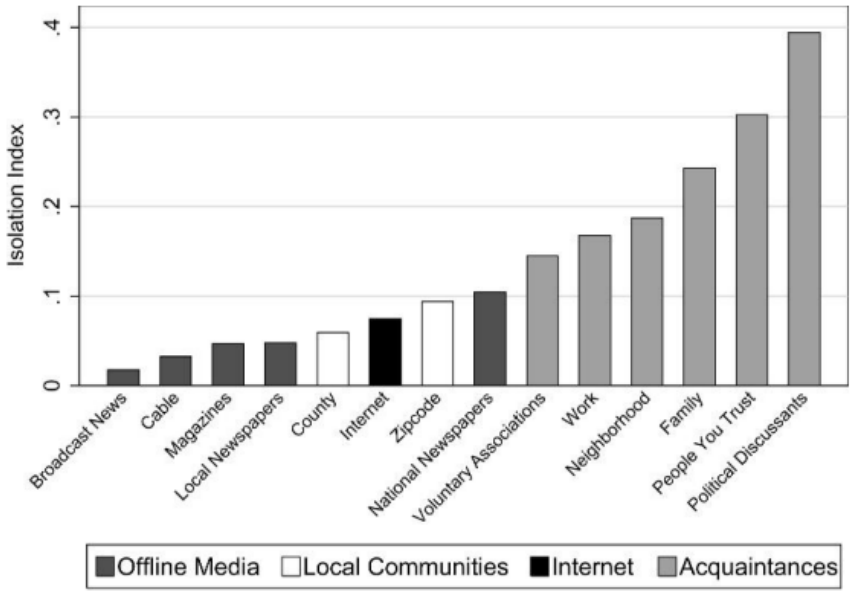

Figure 2.1: Ideological Segregation by Medium and Type of Interaction

As it is possible to see from 2.1, while it is true that news consumption online shows more segregation as compared to broadcast television and local news, it is still lower than national newspaper consumption. Moreover, Internet news consumption was found to be significantly lower than the network formed by acquaintances. The same study also found that overall news diversity is increasing. About a quarter of all U.S. adults receive news from two or more sites, up from 18% in 2016 and 15% in 2013.

2.1.3 Implications and Consequences

In addition to the previously discussed problem of information overload, the lack of traditional trusted intermediation in content sharing has led to another major problem: *informational disorder* [18]. By this term is meant the three types of non-genuine information that pollute the current online ecosystem:

- **Dis-information**: Information that is false and deliberately created to harm a person, social group, organization, or country.
- **Mis-information**: Information that is false, but not created with the intention of causing harm.

- **Mal-information**: Information that is based on reality, used to inflict harm on a person, organization or country

Thus, fake news is just one of many types of false information which can be found online, but the term has become famous due to the fact that one could see in recent events how they can manipulate public opinion. Suffice it to say that according to a recent study [19] limited to Twitter, fake news travels faster than real news.

In this scenario, echo chambers prove to be particularly fertile grounds for the spread of false information. This is because within an echo chamber, a feedback loop develops, in which the sharing of content increases the credibility of that information among group members. As a result, misinformation tends to be reinforced and considered increasingly true and reliable [20]. One of the main examples of the above is the measles outbreak in Disneyland, California, in 2014. The outbreak stemmed in part from escalating parental apprehension regarding vaccination efficacy and potential side effects. Various anti-vaccine groups and influential figures have effectively spread misinformation and instilled fear, resulting in diminished vaccination rates. Consequently, more individuals have contracted the entirely preventable measles disease [21].

Another example of how the filter bubble system is affecting society is given by the political field. Pariser argues that the negative effects of filter bubbles are damaging democracy as a whole. Democracy relies on the participation of the entire population in decision-making, empowering the people by prioritizing the will of the majority. In order for democracy to be an effective as a system, each person needs to be informed on the issues they're voting on, and be able to have critical discussions. In this regard, the article Your Filter Bubble is Destroying Democracy [22] by Mostafa M. El-Bermawy tries to explain how the fact that all major political polls failed to predict Donald Trump's election in 2016 is due to the filter bubble phenomenon.

Finally, it is important to point out that there are several solutions for living with these phenomena. First of all, education is crucial to address issues related to filter bubbles and echo chambers. It's important to teach people to seek out knowledge, how to research issues, and that not everything online is true. If people get a well-rounded education and learn to have meaningful, learning-oriented conversations the negative effects will

be greatly decreased [23]. With the aim of increasing awareness of these phenomena, Allsides[1] is a platform created to provide a more balanced view of news and events from different political perspectives. It collects news from a wide variety of sources and then classifies them according to their political leanings, which can be left, center, or right. The platform allows users to compare news from different political orientations, providing a more comprehensive and balanced view of the contemporary media landscape.

New algorithms and technologies were also created for the same purpose, an example are some search engines as: Qwant and DuckDuckGo, that are starting to consider and deal with the issues related to the filter bubble, for example, Qwant[2] " is the search engine that doesn't know anything about you" and DuckDuckGo[3] filter the information with the goal of improving the relevance.

2.2 Social Media Analytics

This section presents the main tools of Social Media Analysis needed to identify echo chambers within social media. A definition of Social Media Analytics is given by Gartner [24]:

It is monitoring, analyzing, measuring, and interpreting digital interactions and relationships of people, topics, ideas, and content. Interactions occur in the workplace and external-facing communities. Social analytics include sentiment analysis, natural language processing, and social networking analysis (influencer identification, profiling, and scoring), advanced techniques such as text analysis, predictive modeling and recommendations, and automated identification and classification of subject/topic, people, or content.

As will be illustrated in the following subsections, social media analyses can be divided into two overarching categories: (i) *Social Network Analysis* (Subsection 2.2.1), and *Social Content Analysis* (Subsection 2.2.2).

[1] `https://www.allsides.com/unbiased-balanced-news`

[2] `https://www.qwant.com/`

[3] `https://duckduckgo.com/`

2.2.1 Social Network Analysis

Social Network Analysis refers to the process of investigating social structures through the use of network theory and graphs. Network theory is the study of structures that represent symmetrical or asymmetrical relationships between discrete objects. In computer science, network theory is a part of graph theory: a network can be defined as a graph of nodes (vertices) interconnected through links (edges).

In the first subsection the most important concepts of Graph Theory will be examined, then in the second the task of Community Detection will be approached, one of the main applications for echo chamber detection.

Graph Theory

Graph theory [25] was introduced by the Swiss mathematician Euler. In this section, the most important notions of graph theory will be introduced. A graph is a pair $G = (V, E)$ of sets such that $E \subseteq [V]^2$ hence, the elements of E are 2-element subsets of V. The elements of V are the vertices of the graph G, the elements of E are its edges (or arcs). There is also the definition of a weighted graph $G = (V, E, w)$ in which the function $w : E \to R$ associates a value or cost to each arc.

A graph is said to be directed if the edges have an orientation, which means that all the edges are directed from one vertex to another; otherwise, it is said to be undirected, i.e., all the edges are bidirectional.

A loop is an edge whose extreme vertices are the same vertex. Multiple edges are different edges with the same pair of extreme vertices. If a graph G has no loops and no multiple edges, G is called a simple graph, if not, it is called a multigraph.

The degree $d(v)$ of the generic node v is defined as the number of arcs incident on that node, if the graph is oriented, the outgoing degree is calculated, given by the number of arcs of which v is the head vertex of the node, and vice versa the incoming degree given by the number of arcs of which v is the tail vertex. The degree is a particularly important measure that is used to evaluate the centrality of the elements within the graph.

When analyzing social media networks, the complexity of the graphs, which consist

of numerous edges (connections) and vertices (nodes), requires the use of various metrics to gain insights. These metrics can be broadly categorized into three main families [26]:

- **Connection**: They have to do with the ways in which social network entities connect with each other. In this category one can find all the metrics connected with the concept of homophily, previously mentioned. Homophily is related to the concept of Assortativity (or assortative mixing) which refers to the tendency of nodes in a network to connect with other nodes that have similar characteristics or properties. It quantifies the correlation between the attributes or degrees of connected nodes in a network [27].

- **Distribution**: They have to do with the way in which information can flow within a social network. The most important metrics in this group are the Centrality ones, a group of metrics that aim to quantify the importance or influence of a particular node or group of nodes within a network. Also crucial in this scenario are the density metrics, which can be defined as the percentage of effective links in a network out of the total possible number.

- **Segmentation**: They have to do with the ways of ”clustering” the components of the social network.

Centrality Metrics

In graph theory and network analysis, centrality indicators are used to identify the most important vertices within a graph. These indicators are used to detect the most influential individuals in social networks but also for example key infrastructure nodes in urban systems.

The historically first and simplest centrality indicator is *degree centrality*, defined as the number of edges incident to a node. Degree centrality can be interpreted in terms of the “immediate risk” of a node catching or spreading whatever is flowing through the network. For directed graphs, two separate measures of degree centrality are defined: in-degree and out-degree. When ties represent positive interactions, such as friendship or collaboration, the in-degree is often interpreted as a form of popularity, while the out-degree is seen as a

propensity to follow the behavior of others. Beyond degree centrality, there are two other centrality measures based on shortest (geodesic) paths:

- **Betweenness Centrality**: This measure quantifies the extent to which a vertex lies on the shortest paths between other vertices. It indicates how much a node controls the flow of information across the network. The first formal definition of this indicator was given by Freeman (1977) and is given by the following expression [28]:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

  where $\sigma_{st}(v)$ represents the number of shortest paths from s to t that pass through v and σ_{st} represents the total number of shortest paths from node s to node t

- **Closeness Centrality**: This measure reflects how close a vertex is to all other vertices in the network, based on the shortest paths. It captures the ability of a node to quickly interact with all other nodes in the network. The first definition of this metric was defined by Bavelas (1950) [29] and is expressed through the following formula

$$c_C(v) = \frac{1}{\sum_u d(v,u)}$$

  where $d(v,u)$ represents the shortest distance between the node v and the node u.

Community Detection

The identification of communities within a network is one of the primary tasks of Social Network Analysis (SNA), with numerous applications. Communities are cohesive sets of elements that interact more frequently with each other than with the rest of the nodes in the graph. The primary objective of community detection is to identify densely connected communities within a network or graph. Communities can be disjoint, meaning they do not share any elements, or overlapping, where at least one node is shared between different groups.

Broadly speaking, community detection methods can be divided into four non-exclusive categories [30]:

- **Node-Centric Community**: strategies that require nodes to satisfy certain properties such as complete mutuality, reachability of members, and nodal degrees.
- **Group-Centric Community**: are based on the properties that can be defined at the level of individual clusters. The group has to satisfy certain properties without zooming into node level.
- **Network-Centric Community**: methods in which partitions are based on the similarity of nodes where all the communities are disjoint. The interest is referred to partition the whole network into several disjoint sets.
- **Hierarchy-Centric Community**: strategies that build a hierarchical community structure based on the network topology. In particular, there are two different approaches: divisive hierarchical clustering, based on iterating the partitioning process to find groups gradually of smaller size, and agglomerative hierarchical clustering based on joining groups that satisfy certain characteristics. A hierarchical structure of communities is built and the community examination is made at different granularity.

The modularity measure is used to assess the quality of a community partition by comparing the actual number of connections within communities to what would be expected by chance. The formula is as follows:

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i k_j}{2m} \right] z_i z_j$$

where m is the number of edges, $A_{i,j}$ is the adjacency matrix, k_i and k_j are the degrees of nodes i and j, and z_i and z_j indicate the community membership of nodes i and j respectively.

When the observed connections within communities are significantly higher than expected, the modularity score is positive, indicating the presence of meaningful communities. If it is close to zero, the network's community structure is not well-defined. The modularity score typically falls within a range of -1 to 1. In particular:

- $Q = 1$: A high positive modularity score indicates that the network's community structure is pronounced and meaningful.

- $Q = 0$: A modularity score close to zero means that the network's community structure is not significantly different from what would be expected by random chance.

- $Q < 0$: A negative modularity score indicates that the network's community structure is less pronounced than what would be expected by random chance.

From several studies [31] emerges that modularity does not capture so clearly echo chamber aspects, probably because not targeted at the problem at hand. In fact, the identification and impact evaluation of echo chambers in online social networks is a task that cannot be easily addressed relying only on standard CD methodologies.

An approach that could be useful in this scenario, could be EVA (LOUVAIN Extended to Vertex Attributes), a bottom-up low-complexity algorithm designed to identify network hidden meso-scale topologies by optimizing structural and attribute-homophilic clustering criteria. The main family of algorithms to which EVA [32] belongs is Labeled Community Discovery (LCD).

Let $G = (V, E, A)$ be a labeled graph where V is the set of vertices, E the set of edges, and A a set of categorical attributes such that $A(v)$, with $v \in V$, identifies the set of labels associated to v. The labeled community discovery problem aims to find a node partition $C = c_1, ..., c_n$ of G that maximizes both topological clustering criteria and labels homophily within each community. LCD focuses on obtaining topologically well-defined partitions (as in CD) that also results in homogeneous labeled communities. EVA is designed as a multi-objective optimization approach. It adopts a greedy modularity optimization strategy, inherited by the LOUVAIN algorithm, pairing it with the evaluation of intra-community label homophily. EVA's main goal is maximizing the intra-community label homophily while assuring high partition modularity. EVA is also designed to handle networks with more than one label.

The most important measure used in conjunction with modularity is *Purity*: Given $c \in C$ its purity is the product of the frequencies of the most frequent labels carried by its nodes [32]. To take into account both modularity and purity while incrementally identifying a network partition, it combines them linearly, thus implicitly optimizing the following score: $Z = \alpha P + (1 - \alpha)Q$ where α is a trade-off parameter that allows tuning the impor-

tance of each component. Moreover, the partitions obtained with $\alpha = 0.8$ and $\alpha = 0.9$ are the ones that usually better harmonize the two quality functions, so more importance is given to Purity.

2.2.2 Social Content Analysis

The activity of Social Content Analysis is understood to be the analysis of content generated by users on social media. The goal is to understand the opinions, emotions, and behaviors of individuals who interact on social platforms. In this subsection, the Sentiment Analysis and Topic Modeling tasks will be explored.

Sentiment Analysis

From the large amount of data we have today, it is possible to extract from texts emotions and sentiments expressed by people. In this scenario, sentiment analysis deals with analyzing and determining the sentiment or emotional tone expressed in a piece of text, such as a review, social media post, or customer feedback. Existing approaches to sentiment analysis can be grouped into three main categories [33]:

- **Knowledge-based techniques**: Text can be categorized into different emotional states by identifying specific words such as “happy” or “sad”. This classification often relies on lexicons for emotion and sentiment analysis. However, a significant drawback of knowledge-based methods is their limited ability to recognize emotions when linguistic rules come into play. For instance, while a knowledge-based system can accurately label a sentence like “today was a happy day” as expressing happiness, it may struggle with sentences like “today wasn’t a happy day at all”.

- **Supervised methods**: It consists of feeding a machine learning algorithm a large training corpus of labelled texts.

- **Hybrid approaches**: They exploit both knowledge-based techniques and supervised methods to perform tasks such as emotion recognition and polarity detection. Given lexicons can be used as additional features in supervised approaches to improve the effectiveness of the considered model.

Sentiment analysis, while a powerful technique, encounters several difficulties related to the correct interpretation of natural language. It is indeed very difficult for an algorithm to catch polysemy, irony, or even some idiomatic phrases.

One of the main tools used to perform Sentiment Analysis on social media is VADER (Valence Aware Dictionary and sEntiment Reasoner) [34], a lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media. The model outputs a score that is calculated by summing the valence scores of each word in the lexicon, adjusted according to the rules, and normalized to be between -1 (maximum negative) and +1 (maximum positive). VADER not only scores words individually, but also takes the context of the sentence into account. The score of the individual word can in fact be increased or decreased depending on certain contextual information such as the proximity to an intensifier ('very nice' vs. 'nice') or a negation, which will consequently invert the sentiment. This tool is very specialized in analyzing texts from social media because is very accurate as it can recognise emoji, slang but also initialisms and acronyms.

Topic Modeling

Topic modeling is an unsupervised machine learning technique in natural language processing that automatically identifies and groups recurring themes (topics) within a set of documents [35]. By analyzing documents to detect patterns of words and phrases that frequently appear together, it discovers the main concepts in the text without requiring predefined labeling or classification. The two most important topic modelling methods are

- **Latent Semantic Analysis (LSA)**: The core idea is to take the Document-Term matrix and decompose it into a separate Document-Topic matrix and a Topic-Term matrix. This technique, based mainly on Single Value Decomposition was patented by Deerwester et al. (1990) [36].

- **Latent Dirichlet Allocation (LDA)**: Each document is considered a mixture of topics and each word in a document is considered randomly drawn from the document's topics. The topics are considered hidden (latent) and must be uncovered via

analysing joint distribution to compute the conditional distribution of hidden variables (topics) given the observed variables, and words in documents. The name is given by the fact that the distribution of topics in a document and the distribution of words in topics are Dirichlet distributions. This Bayesian model was first presented by Blei et al. (2003) [37].

- **Neural network-based topic modeling** is an advanced approach to identifying and grouping themes within a set of documents using neural networks. Unlike traditional topic modeling techniques such as LDA, they leverage the power of deep learning to capture more complex patterns and relationships in the data.

One of the main topic modelling algorithms is BERTopic, which uses embedding techniques to mathematically represent texts in a semantic space with the aim of identifying their main topics. More specifically unlike LDA or LSA models, BERTopic generates document embedding with pre-trained transformer-based language models and then clusters these embeddings to generates topic representations [38]. Transformer models are able to obtain richer representations, resulting in more accurate and interpretable results.

Disregarding the approach, the output of a topic modelling algorithm is a list of topics with associated clusters of words (and their probabilities). Three are the main metrics used to evaluate topic modeling:

- **Perplexity** is a measure of uncertainty, meaning the lower the perplexity better the model. It aims to capture how "surprised" a model is with new data it has not seen before. This metric was introduced in the field of speech recognition by Jelinek et al. (1977) [39].

- **Coherence** is the measure of semantic similarity between top words in our topic. The higher the coherence better the model performance. These measurements help distinguish between topics that are semantically interpretable topics and topics that are artifacts of statistical inference. An early version of this metric was proposed by Mimno et al. (2011) [40]

- **Diversity** evaluates whether topics are diverse and not redundant. An example of

this metric is the one proposed by Dieng et al. (2020) [41], which consists of the percentage of unique words in the top 25 words of all topics.

3. Related Works on Echo Chambers

After presenting the main techniques behind Social Media Analytics, this chapter aims to conclude the state-of-the-art overview by presenting the main methods of Echo Chamber Detection. In the first part of this chapter (Section 3.1), some of the work that has been done in echo chamber research will be presented. Secondly, the main classification criteria of echo chamber detection methodologies will be introduced at a high level (Section 3.2) and then go into detail in explaining the theoretical approaches that will be most useful in the experimental part (Section 3.3)

3.1 Main Studies Concerning the Origin and Formation of Echo Chambers

Echo chambers are a topical and highly relevant phenomenon to assess the interactions and opinions formed on social media. Over the past decade in addition to echo chamber identification methods, much research has been invested in understanding the origins and factors influencing this phenomenon. In this regard, Sasahara et al. (2021) [42] studied the conditions under which echo chambers emerge and suggested some mitigation ideas. Their model outlines how an online community can easily become an echo chamber even with small amounts of influence and unfriending.

Other works have focused on individual users to understand the relationship between personality and echo chambers. In one of these studies, conducted by Song et al. (2024) [43], the impact of personality was investigated using Digman's Big Five model [44], which comprises five categories: extraversion (E), neuroticism (N), agreeableness (A), conscientiousness (C) and openness (O). The study utilizes an unsupervised personality recognition method to assign personality models to users, aiming to explore the distribution differences between echo chambers and personality traits across different platforms and topics. This model leverages textual data from social media, which captures a wide array of real-life expressions and thus serves as a rich resource for analyzing personal-

ity traits [45]. The variables analyzed in the textual data include the usage of pronouns, prepositions, commas, and emoticons. The results revealed that although the patterns of users' personality traits show similar distributions across topics, there are differences between the two platforms considered: Weibo and Twitter. Specific personality traits attract like-minded individuals who engage in discussions on certain topics, eventually forming homogeneous communities [43]. Still, on the relationship between echo chambers and personality, Bessi (2016) identified in his study on Facebook the prevailing personality in an echo chamber, i.e. an individual who has low extroversion and is suspicious and antagonistic towards others (low agreeableness) [46]. Also in the context of work that investigates the role of the individual within echo chambers, it is interesting to explore the work of Garimella et al. (2018) [47] that examines the role and behaviour of two types of users in echo chambers concerning political discourses: gatekeepers and bipartisan. The gatekeeper is a user who consumes content with diverse leanings but produces partisan content (with a single-sided leaning) [47]. In a certain way, the gatekeeper is a user who filters one side of the information. The bipartisan, on the other hand, is a user that produces content of different political orientations (if one considers a bipolar situation, it can be said to produce content of both orientations). The work showed how producing content that expresses opinions aligned with both sides has a cost in terms of centrality in the network and engagement rate. Gatekeepers, on the other hand, have high centrality within the network, as they receive input from both leanings but are not completely tied to a single community. This result highlights a worrying aspect of echo chambers, as it suggests the existence of phenomena that prevent mediation between the two parties [47]. Political topics along with health and religion are in fact those most likely to form echo chambers, as they are usually polarised into two points of view [48].

3.2 Echo Chamber Detection Methods Classification

Before moving on to the presentation of the main approaches for the identification of echo chambers, it is important to provide a guideline to critically define each of the methods presented.

In the literature, echo chamber detection techniques are mainly divided into three macro-groups [49].

- **Network-based methods** These methods extract information on the level of polarisation solely by considering the relationships between users, analysing the network of relationships at a structural level. They prefer to disregard the content as it is very dependent on both context and language [49]. This type of approach measures disputes arising from network partitioning performed with Community Detection techniques.

- **Content-based methods** These methods ignore the structural characteristics of the network and use User-Generated Content (UGC) to find polarised environments. The main techniques used are NLP and Sentiment Analysis.

- **Combined methods** These methods use both user-generated content and the structure of the relationship network to obtain better results for echo chamber searches.

A classification for echo chamber detection methods based on scale differences has also been proposed in the literature, which consists of the following [50]:

- **Micro-scale ECs**: Approaches that focus on the analysis of the individual user, neglecting the aggregate dimension of the network.

- **Macro-scale ECs**: Approaches based on considering the network of interactions as a whole. It usually consists of partitioning the network into two well-polarised groups assuming that they are the two sides of the controversy. In this way, however, differences that may arise in subgroups of the network are not taken into account.

- **Meso-scale ECs**: These approaches aim to identify clusters of nodes within the network that have the characteristics of an echo chamber. Unlike the previous approach, it does not merely represent users divided into two opposing network factions.

3.3 Network-Based Methods

According to Garimella et al. (2018) [51], many of the network-based methods follow a pipeline structured in three basic steps. These steps can be described as follows:

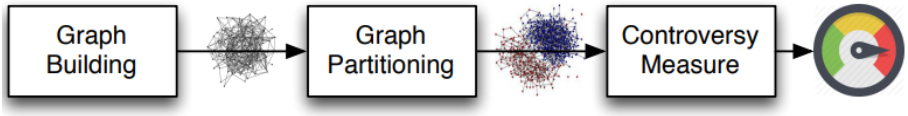

Figure 3.1: Pipeline of network-based methods defined by Garimella et al. (2018)

In particular, the pipeline shown in Figure 3.1, is introduced as part of the investigation of controversial topics. This terminology is defined by Garimella et al.(2018) as questions, topics or issues that can create a large difference of opinion between the people discussing them [51]. In the following sections, each stage will be examined in detail.

3.3.1 Graph Building

The first step in the pipeline consists of modelling the network representing user interactions around a specific topic of conversation. Usually, the latter is identified through hashtags and keywords, identifying all posts and comments that contain that word. The hashtag is then used as a query to isolate all posts that do not talk about the topic of interest [51]. The objective of this phase is to obtain a conversation graph where the nodes represent the users while the arcs represent the type of interaction between them. However, depending on the type of platform considered and the objective of the analysis, it is possible to model user interaction in different ways. In particular, in the work of Garimella et al. (2018) [51], it emerges from Twitter that there are two types of graph that are best suited to the identification of controversy:

- **Follow-graph**: According to this approach, given users who have used a certain hashtag, an edge exists between users u and v if u follows v or vice-versa. This technique is based on the fact that following each other implies homophily and therefore people following each other agree on the same topic. However, this type of information is more difficult to obtain and requires a lot of effort during the data extraction phase.

- **Retweet-graph** This approach identifies a connection between two users if there have been at least two retweet actions between the users, regardless of direction. It is based on the assumption that retweeting another user's post is an endorsement action and results in the estimation.

Both approaches have been considered by Garimella in a non-oriented manner, but there is nothing to prevent the graph from being constructed in an oriented manner and perhaps putting different weights that express the strength of the relationship between two users: for example, if the follow is reciprocated between two users, the weight of the relationship will be greater.

In this regard, the work of Cossard et al. 2020 [52], in addition to using oriented graphs, proposes another type of criterion that can be used to model the conversation network, namely the action of including a user's username in a tweet (together with replies). In particular, there is a connection between two users if one mentions the other in a tweet and the weight of this connection is equal to the number of mentions. Using mention, however, can lead to different results from the previous methods as it often leads to linking two users with very different ideas about a particular topic as mention is often used precisely as a form of 'attack'.

All the network modelling methods presented above are created and based on the Twitter platform, not all of which can be extended to other platforms such as Reddit. Often the most generic form of modelling that can be extended to almost any platform present online is that which considers comments: in Morini et al. (2021) [50] approach for identifying echo chambers on Reddit defines a graph 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 [50].

Regardless, however, of what type of modelling is chosen for the conversation graph, it is important to emphasise as stated by Natarajan et al. (2013), that a connection between two users increases the probability that they have common interests, although it does not guarantee it. Moreover, users who are not connected might still share similar interests [53].

3.3.2 Graph Partitioning

After modelling the conversation graph, various partitioning techniques come into play, which in the work of Garimella et al. (2018) aim to divide the network into two factions, thus placing themselves in the context of macro-scale ECs. In the aforementioned work, it emerges how the METIS algorithm and the techniques of community detection via label propagation turn out to be the best in the partitioning of non-oriented graphs [51].

If, on the other hand, one wants to refer to network-based techniques that are not limited to the bipartition of the conversation network, the approach of Coletto et al. (2017b) allows a more detailed analysis of specific areas of the graph, thus enabling the identification of meso-scale ECs [54]. The approach is based on the assumption that conversation graphs are characterised by motifs, i.e. recurring patterns of user interactions, which can be used to identify controversial content. The frequency of these motifs is used as a predictor of controversy.

3.3.3 Controversy Measure

The last step in Garimella et al. (2018)'s pipeline involves measuring the level of controversy of the partitions created through certain metrics that will be discussed in this section. The main metrics that will be implemented in the experimental part of this work are presented as follows:

- **Random Walk Controversy**: The metrics, introduced by Garimella et al. (2018), considers two partitions X and Y of the graph $G = (V,E)$ (such that $X \cup Y = V$, and $X \cap Y = 0$) and two random walks, one ending in partition X and the other ending

in partition Y [51]. The formula is the following:

$$RWC = P_{XX}P_{YY} - P_{XY}P_{YX}$$

where P_{AB} with $A,B \in X,Y$ is a conditional probability defined as follow: $P_{AB} = P[Start\ in\ A | ends\ in\ B]$. The value of this metric ranges between 0 and 1. The closer it is to 0, the more likely it is to switch to the other partition (no controversy); the closer it is to 1, the more likely it is to stay in the original partition (presence of controversy) [51].

- **Authoritative Random Walk Controversy**: The measure proposed by Villa et al. (2021) is a variation of the RWC measure. In RWC, the starting nodes are randomly selected from the users of the two different partitions. In contrast, the ARWC measure begins only from nodes defined as authoritative. A node is considered authoritative if its degree ranks within the top 15% of its community. The underlying hypothesis of ARWC is that when an authoritative node is influenced by an authoritative node from the other community, it can more effectively sway the non-authoritative nodes within its own community, thereby reducing controversy [31].

- **Displacement Random Walk Controversy**: This metric, also proposed by Villa et al. (2021), focuses on the ratio between the number of steps during a fixed-length random walk that results in a change of community, and the total length of the walk, which is defined as twice the average shortest path in the graph. If a node never changes community during its walk, it indicates a strong connection to its own community, signifying high controversy between the two communities. Conversely, if the node frequently crosses between the two communities, it indicates low controversy. Thus, higher values of this measure correspond to greater controversy between communities, and lower values correspond to lesser controversy [31].

- **Boundary Connectivity**: The metric proposed by Guerra et al. (2013) [55] is based on the concepts of internal and boundary vertices. Given a graph G, it considers $u \in X$ as a vertex in partition X; u belongs to the boundary of X if and only if it is connected to at least one vertex in partition Y and to at least one vertex in partition X

that is not connected to any vertex in partition Y [55]. The set of boundary vertices is therefore defined as $B = B_X \cup B_Y$; conversely, the set $I_X = X - B_X$ is the set of internal vertices of partition X. The set of internal vertices is therefore defined as the set $I = I_X \cup I_Y$. If the two partitions would constitute echo chambers, the whole B should be made up of vertices that are more strongly connected with the elements of I rather than with elements of B. The formula is the following:

$$P = \frac{1}{|B|} \sum_{v \in B} \left(\frac{d_i(v)}{d_i(v) + d_b(v)} - 0.5 \right)$$

where $d_i(u)$ is the number of edges between the vertex u and the elements of the set I, and $d_i(u)$ is the number of edges between the vertex u and the elements of the set B. The metric lies in the range $[-0.5, 0.5]$ which means that a value greater than zero points out that nodes on the boundary tend to connect more to internal nodes rather than to nodes from the other group, indicating that controversy is likely to be present [55].

3.4 Content-Based Methods

In this section, techniques will be analysed that allow the identification of echo chambers from the content posted by individual users. The majority of content-based methods focus on the identification of EC micro-scales, identifying polarised environments through the analysis of content shared or consumed by individual users. These approaches are highly domain-dependent and require prior knowledge of the domain in order to interpret the results.

One of the works to identify the opinion of each user and construct a bias index is the one carried out by Matakos et al. (2017) [56]. They follow Friedkin and Johnsen's (1990) [57] model of opinion formation, that each person i in the network has an internal opinion s_i and an expressed opinion z_i, which is the most important one. expressed opinion z_i which depends on both his internal opinion s_i and the opinions expressed by his neighbours. After estimating internal opinion by means of sentiment analysis or topic modelling

techniques, it is possible for each user it is possible to define the expressed opinion z_u as:

$$z_u = \frac{w_{uu}s_u + \sum_{j \in N(u)} w_{uj} z_j}{w_{uu} + \sum_{j \in N(u)} w_{uj}}$$

where w_{uj} is a weight representing the strength of the link between node u and node j. The z_u will be high for users whose social circle expresses different or moderate opinions, and low for users whose neighbours express extreme and similar opinions. The polarization index is defined as follow: Given a network $G = (V, E)$ and the vector of expressed opinions z, Polarization index is defined as:

$$\pi(z) = \frac{z^2}{n}$$

where n is the number of element in the opinions vector.

Another content-based approach is to implement the sentiment measures proposed by Al-Ayyoub et al.(2017) [58]. The three metrics that yielded the best results among controversial topics on some Twitter data are reported below.

- **Ratio of positive to negative sentiment scores(PN)**: It is calculated as the ratio between the number of tweets with positive sentiment to those with negative sentiment. It can of course be extended to a measurement between 0 and 1.

- **Ratio between numbers of summation of positive and negative tweets to total number of tweets (PNT)**: Considering that a tweet can take on positive, negative or neutral connotations, this metric considers the ratio of the number of positive and negative tweets to the number of total tweets. The measure takes values between 0 and 1 and the higher the value, the higher the level of controversy.

- **Ratio between sentiment scores count of the ratio of positive and negative tweet count to total number of tweets (PNPNT)**: metric that combines the two measures previously defined above.

3.5 Combined Methods

Hybrid approaches combine aspects concerning the structure of the conversation network with aspects related to the user's content. In this section, two hybrid methods that include

content-related information in the modelling phase will be considered.

In this scenario, the approach of Morini et al. (2021) [50] places itself in the context of meso-scale ECs by trying to combine the ideological with the structural side of the network on Reddit datasets. This work is also developed on a four-step pipeline:

- **Controversial Issue Identification** : It consists in identifying a controversial topic. They can include for example political, social, or environmental issues.

- **Users' Ideology Inference**: This phase aims to determine each individual's ideology. Usually, this consists of classifying by means of sentiment analysis or topic modelling techniques each of the user's textual contributions. Usually, the individual's leaning is an average of the leaning of each of the user's contributions.

- **Debate Network Construction**: the interaction network is defined as a graph 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 edge (u,v) is described by a weight $w_{u,v}$ that represents the number of interactions between two users [50].

- **Homogeneous Users' Clusters Identification**: At this point, EVA, a Labeled Community Detection algorithm, previously described in Chapter 2 of this thesis, is applied.

Having obtained the partition of the network through EVA, each of the communities obtained is evaluated through two metrics: the previously defined Purity and Conductance. The latter is defined as follows: Given a community $c \in C$, its Conductance is the fraction of the total edge volume that points outside the community.

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

where c_s is the number of community nodes and m_s is the number of community edges [50].

Having obtained the values of the two metrics for each community, thresholds are set based on which there is a risk that it is an echo chamber. Morini et al. (2021) in their

work set the thresholds for purity at 0.7 and conductance at 0.5: in the case of purity, the risk is higher if the threshold is exceeded; in the case of conductance, however, exactly the opposite happens.

It is possible to define the difference between Purity and Conductance as the risk for a community to be an echo chamber. Figure 3.2 helps to explain which cases make the echo chamber risk index high.

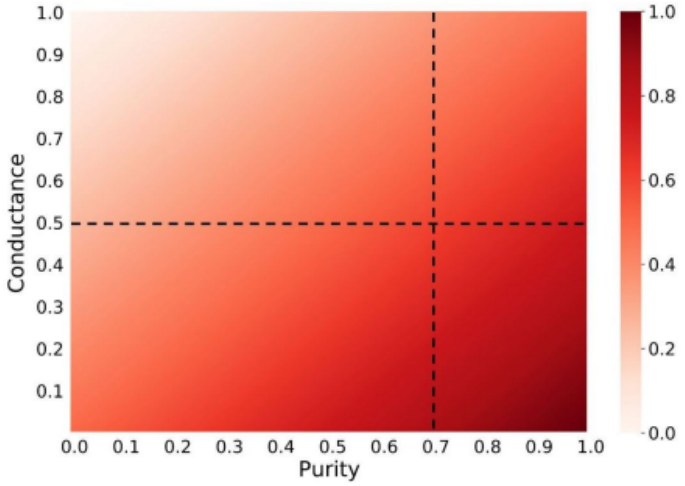

Figure 3.2: Risk for a community to be an echo chamber red color implies a higher risk while white a lower one (Morini et al. (2021))

Another hybrid method that wants to integrate user content information within the conversation network is the one proposed by Villa et al. (2021) [31]. It proposes to incorporate semantic information into the graph via weights on the arcs that include sentiment, topic, or both:

- **Sentiment-based modeling**: First of all a sentiment score $s(v)$ to be associated with each user it is defined. It may vary between the range values $[-\alpha, +\alpha]$. Then it is possible to compute a sentiment similarity score between two users v_i and v_j that already have a connection in the network:

$$ss(v_i, v_j) = 2\alpha - |s(v_i) - s(v_j)|$$

This way, a value from 0 to 2 is attributed to such couples of users, depending on their sentiment similarity. Then the sentiment similarity score is used to compute

the weight to be used inside the network:

$$w_s s = 1 + ss(v_i, v_j)$$

In this formulation, when two users have completely opposing sentiments, the edge weight is set to 1. Consequently, the focus shifts exclusively to the network's topology, disregarding sentiment nuances.

- **Topic-based modeling**: In this graph modeling, to each user v in the graph, is associated a set of topics $T(v)$ that is the subset of topics discussed by v within the global set of topics T. The topic similarity $ts(v_i, v_j)$ between two users v_i and v_j already connected in the graph, is computed by considering the overlap of their topics as follows:

$$ts(v_i, v_j) = |T| - |T(v_i) \Delta T(v_j)|$$

where Δ denotes the symmetric difference between the two sets. In this way, a topic similarity value equal to $|T|$ is assigned, for users with a total overlap of topics discussed. Then the topic similarity score is computed in the same way as before:

$$w_s s = 1 + ts(v_i, v_j)$$

- **Hybrid based modeling**: The hybrid weight takes into account both the sentiment similarity score and the topic similarity score:

$$w_s s = 1 + ts(v_i, v_j) + ss(v_i, v_j)$$

3.6 Method Comparison

At this point, it may be natural to ask which metrics are most effective and which are best suited for different platforms or use cases.

In this regard, the work of Impicciché and Viviani (2024) [59] aims to examine and compare different methodologies for identifying echo chambers in social networks, using data from both Twitter and Reddit. The work showed that structural metrics are more effective in identifying echo chambers on Twitter, as on this type of platform the structure of

relationships carries more weight than on Reddit, where the same metrics however were found to be inadequate. Indeed, interaction between two users who are not ideologically close is much more common on Reddit. The involvement of semantic aspects improves these metrics, suggesting that content plays a more predominant role on Reddit in identifying polarised communities. Hybrid metrics are actually the ones with more consistent results between the two platforms.

4. An Analysis of Echo Chamber Detection over Time

After presenting the concept of echo chambers and illustrating the complete overview of echo chamber detection techniques, the aim of this chapter is to introduce the experimental part of this thesis. Section 4.1 will present the research objectives, while in Section 4.2 the chosen analysis methodology will be illustrated. Finally, Section 4.3 will introduce the datasets taken into consideration.

4.1 Research Goals

The primary objective of this work is to observe the behaviour of some echo chamber detection metrics on several controversial datasets, in order to further test their generalisability from the perspective of the topics addressed and the platforms used. The added value of this work is to address the issue of the temporal dimension and thus highlight new ways of measuring the evolution of echo chambers over time. In particular, it aims to answer the following research questions:

RQ1. How does user ideology evolve over time? Does it tend to change opinion?

RQ2. How do echo chambers evolve over time? How quickly does polarisation occur around a given controversial topic?

The purpose of the analysis is to study the formation of echo chambers over time and to monitor their behaviour in relation to particular triggering events with the aim of understanding whether common patterns and different rates of polarization of controversial topics can be revealed.

4.2 Methodology and Work Pipeline

In order to accomplish the objectives described above, it was necessary to structure the work pipeline represented in Figure 4.1, which will be described below in each of its points

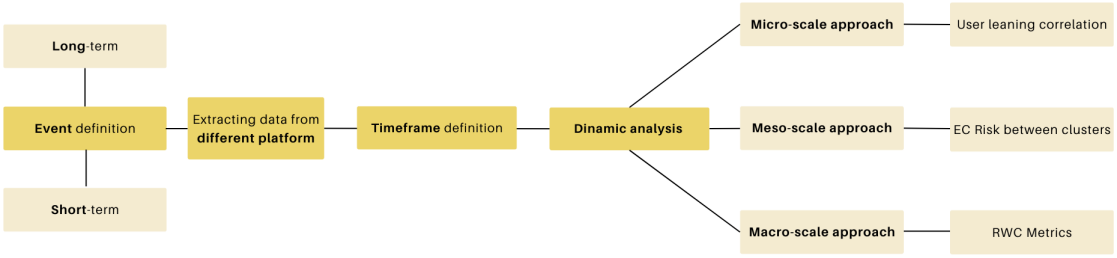

Figure 4.1: Concept map summarising the methodology applied in the experimental part of this thesis

4.2.1 Event Definition

In order to be able to perform the temporal analysis in the best possible way and to be able to differentiate the developments that a controversial event may have over time, it was decided to divide the work into two case studies:

- **Long-term events**: Typically big debates that have been going on for decades such as climate change, abortion, or vaccines.
- **Short-term events**: They refer to a particular event that often makes people talk about it for only a few weeks, as in the case of the post-Brexit referendum.

It is important to keep the two event types separate as they require two different analysis timeframes, especially when assessing the speed of echo chamber formation. In fact, it was assumed that different effects may come into play and that echo chamber detection metrics may also behave differently, especially when looking at the time series.

4.2.2 Extracting Data from Different Platforms

Since the aim of the work is also to analyse the differences between platforms, data from both Twitter and Reddit will be considered, but with different methodologies. In fact, given the new regulations concerning Twitter's API (currently X), it was decided to use Twitter datasets already present in the literature and instead use the API extraction of the same topics but from Reddit's platform. Indeed, the latter offers an official API to which data can be freely accessed by querying the database. The wish was to obtain for each Twitter dataset the comparable Reddit dataset in terms of time, but this was not always possible due to the extraction limits of the Reddit API. To summarise, for each topic chosen, regardless of whether it is long or short-lived, conversation data on the topic will be extracted from both Twitter and Reddit, trying to make the two-time intervals as comparable as possible.

It is also important to emphasise that only users for whom it was possible to have information on both the content (i.e. at least one post on which leaning could be estimated) and the relationships they had within the conversation network were taken into account.

Having obtained the raw data, it is important to define how it was modelled in order to obtain the conversation graph. Given the different nature of the two platforms (Twitter and Reddit), two different modulations of the network were chosen. When creating conversation graphs, it is important to define how the link between two users will be mapped. Based on the platform and the available data, the following criteria were adopted:

- **Twitter**: In the datasets taken from Twitter, mention activity was taken into account: a relationship exists between two generic users of the platform if there has been at least one mention activity between the two. The graph is not oriented and each link has a weight relative to the number of mention activities between the two users.

- **Reddit**: In the datasets taken from Reddit, comments were instead taken into account: there is a relationship between two generic platform users if user1 responded to a comment by user2, or vice versa. The weight on this arc corresponds to the total number of responses recorded between user1 and user2.

Based on the topic covered by the chosen dataset, one of the following methods will be used to analyze the content of the posts: VADER as regards sentiment and BERTopic with regard to the topic.

It was decided to determine according to the topic of conversation the technique for extracting content-related information because there are topics where sentiment analysis is more suitable than topic modelling and vice versa. For example, when dealing with topics such as the COVID-19 pandemic, almost all comments refer to concern and turn out to have negative sentiment, thus making it necessary to go deeper with topic modelling. On the contrary, in big debates where one can take sides either for or against, it was decided that sentiment is a good approximation to capture the user's leaning concerning the conversation.

4.2.3 Timeframe Definition

Since the focus of this thesis is time and how echo chamber detection metrics evolve over time, it is necessary from each dataset obtained to identify significant time segments. This step is crucial for constructing the reference time series and can greatly affect the outcome of the experiment itself. Customised time intervals were identified for each dataset in order to highlight its individual characteristics. The following criteria were used to choose the optimal time segment:

- **Volumes**. An attempt was made to take into account a division that would make the time clusters comparable and homogeneous from the point of view of volumes. A time interval comprising few users would not have been able to produce meaningful results and could not be compared to another interval with higher volumes.

- **Type of events: long and short-term**. The subdivision of the time segments also tried to consider the type of event taken into consideration. For example, for long-term debates, it was decided to consider the year while for short-term events the month was considered sufficient. However, since each debate is linked to events that may have had an impact on users' interest in the conversation, an attempt was made to construct time segments capable of capturing these aspects. For example in

the case of Brexit, if new reforms were implemented in January, it is necessary that that month be isolated in order to capture the impacts.

- **Platform comparability**. Since the same event will be monitored on several platforms, it is important that the time intervals are as comparable as possible so that the same phenomenon can be observed over the same time span. This, unfortunately, was not always possible due to the limitations of the API, which did not allow the extraction of point-in-time endpoints. However, every effort was made to have a consistent comparison.

4.2.4 Dynamic Analysis

For each dataset considered and for each time interval considered, the following analysis will be carried out:

- **Micro-scale EC approach**: Correlation between neighborhood leaning and individual leaning.
- **Meso-scale EC approach**: risk of being an echo chamber (a measure based on Purity and Conductance). The idea is to replicate the approach of Morini et al. (2021) [50].
- **Macro-scale EC approach**: Partition of the conversation network based on both structure and content. The idea is to replicate the approach of Villa et al. (2021) [31].

The main idea is to consider three different echo chamber detection methodologies in order to see if the results are consistent with each other. The objective is then for each dataset to construct a time series with each of the methods selected above.

In the following sections, each of the approaches will be analysed in detail, in order to best define the experiment.

Micro-Scale EC Approach Definition

This type of approach focuses on the individual user and the idea of measuring the correlation index between individual leaning and neighbourhood leaning is aimed at seeing how much the individual user's thinking on a given topic correlates with that of the users he or she has come into contact with. The process that will be applied is as follows:

1. For each set time interval, the leaning is calculated: in the case of sentiment is estimated for each post using VADER. If the Twitter dataset already had a sentiment or topic label, that was taken into account for this step. In this case, the same dataset is used as a training set to build a classification model with the aim of classifying the corresponding dataset from Reddit.

2. For each user, the median of the posts written by them is taken to obtain an estimate of the leaning of the individual. In the case of a dataset where a categorical variable is considered the most frequent label for the individual user is taken into account.

3. The correlation between the individual leaning of the user and the leaning of the users with whom he or she has come into contact within that time interval is calculated. In the case of the categorical variable where it is not possible to calculate the correlation, the probability of being in contact with someone with the same leaning is taken into account.

4. Once the correlation (or probability) value is obtained for each time interval, it is then constructed the time series and compared with the other datasets.

Meso-Scale EC Approach Definition

The meso-scale approach does not look at the individual network but work on individual communities.

Following Morini et al. (2021) work [50] the individual leaning estimated above is considered and then EVA (Louvrain Extended to Vertex Attributes) is applied as a Labeled Community Discovery algorithm.

The EC Risk index is then estimated by means of the previously defined Purity and Conductance. In particular is defined as:

$$ECRisk = Purity - Conductance$$

This approach is applied to the two largest communities obtained and is repeated for all available time intervals in order to obtain the time series.

Macro-Scale EC Approach Definition

This last part of the analysis aims to replicate the approach of Villa et al. (2021) [31]. It takes the entire network into account and is characterised by the fact that weights are considered in the community detection algorithm that maps the similarity between users at the level of the content expressed in the posts. In this case, the type of weight taken into account will depend on the type of analysis used to estimate the content leaning for each dataset: for example, in the case of the Brexit dataset, the topic weight will be used, while for the vaccines dataset, the sentiment one will be used.

After modelling with METIS, the evolution over time of the following 4 metrics, already defined in previous chapters, is monitored:

- **Random Walk Controversy**
- **Authoritative Random Walk Controversy**
- **Displacement Random Walk Controversy**
- **Boundary Connectivity**

4.3 Dataset Presentation

In this section, the tested datasets will be presented. In particular, it was decided to consider the following controversial topics:

- **Long-term events**
  - The use of vaccines
  - The COVID-19 pandemic
- **Short-term events**
  - The Brexit phenomenon
  - Opinions on ChatGPT

There are therefore 4 controversial topics and, at the end of the extraction process 8 datasets (4 from Twitter and 4 extracted from Reddit) were obtained. As below, the first two of the list will be analysed as long-term events, while the last two as short-term events.

The choice of these topics was mainly conditioned by the availability of data. In particular, It has been tried to consider datasets that were already used in the literature and had a large enough timeframe to allow for a dynamic analysis.

In the following subsections, the context and the events taken into consideration will be explained for each dataset. In addition, the extraction criteria used for using the Reddit API and the way in which the content information was extracted will be explained.

4.3.1 Vaccinations

As the first long-term dataset, it was decided to opt for the topic of vaccinations, a controversial topic that has been going on for many years.

The starting dataset was the Twitter Vaccination Dataset freely downloadable from the Kaggle platform[1]. The data refer to tweets containing the string 'vaccination' in the text,

[1] `https://www.kaggle.com/datasets/keplaxo/twitter-vaccination-dataset?select=master.csv`

collected from December 2006 to October 2019. As the dataset is international, it also contains tweets in several foreign languages. In order not to have problems with handling tweets from different languages, only those in English were selected.

At the end of the process of extraction and selection of users according to the previously explained criteria, it was possible to obtain data from approximately 7000 users. The timeframes for the analysis were then decided in order to have a homogeneous number of users per period.

As for the Reddit platform, the same kind of extraction was performed by trying to extract as many posts containing the keyword "vaccination" as possible with the official API. In this case, data from 3190 users were obtained after the extraction and selection process.

For an issue such as vaccinations, it was decided to extract content-related information through sentiment, as it is an issue where it is possible to have positive sentiment and thus take sides in favor or conversely to have negative sentiment about it and take sides against it.

4.3.2 COVID-19

The second long-term dataset to be considered refers to COVID-19. Here again, the starting point was the *COVID-19 Twitter Dataset* downloadable from Kaggle. In this work, data were extracted for three different time periods referring to the 3 main phases of the pandemic: The first phase following the WHO declarations (April-June 2020), the second phase (August-October 2020) and the arrival of the vaccines (April-June 2021). In all three time periods, tweets containing at least one of the following keywords were considered: #covid-19, #coronavirus, #covid, #covaccine, #lockdown, #homequarantine, #quarantine-center, #socialdistancing, #stayhome, #staysafe, resulting in approximately 1mln tweets. As mentioned earlier, however, not all were considered, but only those for which was available leaning and relationship-building information within the network. Specifically, only 22k users were considered, subsequently trying to find the time split that would allow as comparable a number of users as possible.

In addition to tweet-related information such as mentions and user code, this dataset

also has the particularity of containing a Sentiment column: i.e. an indicator between 0 and 1 where 1 corresponds to the highest level of positive sentiment and 0 to the highest level of negative sentiment.

This data source is also part of the work of Chakraborty et al. (2021) where Sentiment Analysis of Tweets concerning COVID-19 was carried out using a Classification Model based on LSTM [60].

Although the valuable sentiment information was already present, it was noticed how it had little relevance in describing and characterizing the pandemic since most of the comments were united by negative sentiment and concern about the phenomenon. In this regard, it was decided to extract topics, coming to the conclusion that all the comments found could be divided into the following three topics:

- **Effects**: Comments mainly talking about the effects of COVID, particularly the symptoms on their body
- **Political**: All comments referring to anti-COVID measures that have been taken by various governments
- **Conspiracy**: All comments that refer to conspiracy theories, such as those related to BigPharma

In order to make a comparison between platforms, the official Reddit API was used with the aim of extracting the greatest number of posts concerning the topic within different subreddits and extracting first and second-level comments. In particular, the maximum possible number of comments containing at least one of the keywords analysed for Twitter was taken. An attempt was made to get as close as possible to Twitter's timeframes, but this was not always possible given the limitations of Reddit's API. Again, after the extraction and selection process, it was possible to identify 3k users, who will be the subject of the analysis. In order to have the same topics highlighted above in the Reddit dataset, the Twitter dataset with the respective topics found through topic modelling techniques was used as a training set in the construction of a classification model which was then applied to the Reddit dataset.

4.3.3 Brexit

As a first short-term event, it was decided to analyse the Brexit phenomenon. Although it can be considered a long-term event, given that debates on the UK's choice to leave the European Union have been going on for a decade now, it was decided to address this issue by focusing on the early months of 2022, when new regulations were triggered that further detached the UK from the European Union. To do this, the dataset *Brexit Polarity Tweets [Anti + Pro]*[2] was taken into account, which contains about 20k tweets covering the first 3 months of 2022. The interesting thing about this dataset is that each tweet is already labelled as 'ProBrexit' or 'AntiBrexit' taking into account the profile bio of each user. After the data selection process, only data from 5k users were considered suitable for the purpose.

In order to be consistent with what it was seen in this dataset, were extracted as many posts containing the word 'Brexit' as possible with the Reddit API, taking into account all existing subreddits, and then derived first and second-level comments as in the previous extractions. Here too, an attempt was made to come as close as possible to the timeframe of the Twitter dataset in order to have comparable datasets. Furthermore, in order to have in the Reddit dataset the same labels seen in the Twitter dataset (ProBrexit vs AntiBrexit) it was decided to use the second dataset to train an AdaBoost classification model and then make predictions on the Reddit dataset. In this way, it was possible to obtain two datasets from different platforms, but consistent in terms of the labels characterising their content. In particular, after the extraction and selection process for the Reddit dataset, 5k users were identified.

4.3.4 ChatGPT

The last topic considered as a short-term event concerns the introduction of GenAI and in particular ChatGPT. Here again, The starting point was a Twitter dataset called *500k*

[2] `https://www.kaggle.com/datasets/visalakshiiyer/twitter-data-brexit?select=TweetDataset_AntiBrexit_Jan-Mar2022.csv`

ChatGPT-related Tweets Jan-Mar 2023[3]. It contains approximately 500k tweets with mentions and hashtags referring to ChatGPT in the three months following the launch of the platform, i.e. from January 2023 to March 2023. It was considered useful to monitor the ChatGPT launch event also in the light of the updates and new releases produced in the following months. The dataset is also part of an analysis carried out by Agarwal Kartik where he used NLP techniques to enhance some useful insights regarding the introduction of ChatGPT [61]. Of the 500k tweets, however, only those from users of which there was at least one piece of information concerning leaning were selected, thus only considering 9k users.

Again, the Reddit API was used to extract as many posts containing the ChatGPT-related keyword as possible, and then analysed the first and second-level comments. An attempt was made to align the time period as closely as possible to the Twitter dataset, within the limits imposed by the API. Again, after the extraction and selection phase, it was possible to consider for the Reddit dataset 3k users.

As with the vaccination issue, it was decided to use sentiment to capture content-related information in order to capture both enthusiasts of this new technology and skeptics.

[3] `https://www.kaggle.com/datasets/khalidryder777/500k-chatgpt-tweets-jan-mar-2023/data`

5. Experimental Evaluation

In this chapter, the results obtained from the application of the previously explained methodology will be presented. The goal is also to highlight the main insights that emerge from the analysis of echo chambers from a temporal perspective. In Section 5.1, the application of the approaches to long-term events will be analysed, before moving on to short-term events in Section 5.2. Finally, some general considerations will be made in Section 5.3, summarising the results of all the approaches used.

5.1 Long-Term Events

In this section, the results obtained for each of the three approaches for the two long-term themes: COVID-19 and vaccination will be analysed. An effort will be made to highlight the differences between platforms, despite the limitations mentioned in the preceding chapter.

5.1.1 Micro-Scale Approach Results on Long-Term Events

Let's start the analysis by considering the leaning of each individual to see if it correlates with that of the users with whom they come into contact.

As mentioned earlier in the case of vaccinations, a continuous index related to sentiment was considered, while for the COVID dataset, topic-related labels were considered. In the former case it was possible to calculate the correlation while in the latter a probability measure was created. The results for the Twitter datasets can be seen here:

Twitter

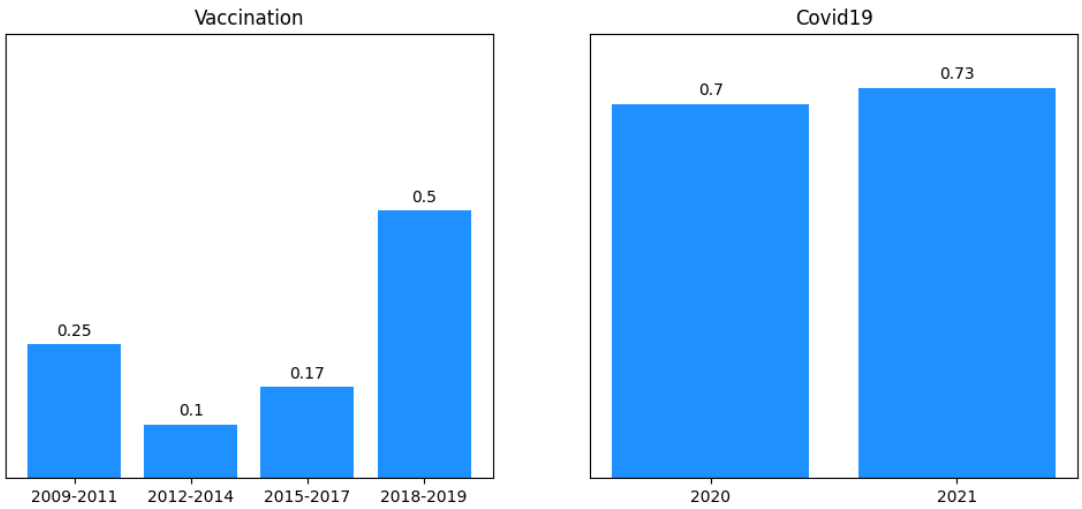

Figure 5.1: On the left the evolution over time of the correlation index between individual and neighborhood leaning for the Vaccination Twitter dataset, on the right for the COVID-19 Twitter dataset the probability of having a connection with a user of the same leaning

In Figure 5.1 it can be seen that for the vaccination graph, largely positive correlation values can be seen, especially for the time segment 2018-2019. The choice of this type of segmentation wanted to highlight some of the main events that have played a key role in the vaccination debate: the 2015-2017 segment for example is the one on which the debate related to the invention of the vaccine for Malaria is focused, the 2018-2019 time segment on the other hand refers to the discussions that took place on Ebola and the discovery of its vaccine. From this first analysis of Twitter, it can be seen that there is certainly a significantly higher level of correlation in the most recent year available. It is not hard to think that the greater exposure to social media that characterizes that time segment compared to earlier ones may amplify the effect of echo chambers in this type of debate.

As for the COVID-19 graph, on the other hand, a high likelihood of interacting with someone with the same leaning can be seen, which again, although less importantly, increases with time.

Reddit

If for the case of the COVID-19 dataset it was possible with the Reddit API to extract content with the same temporal endpoint with respect to the Twitter dataset analyzed earlier so as to allow a cross-platform comparison, this was not possible for the Vaccines dataset, which instead referred to a very large time period and far back in time for a social platform like Reddit that had just been founded in 2006. However, an attempt was made to perform the same kind of analysis with the available data as seen previously.

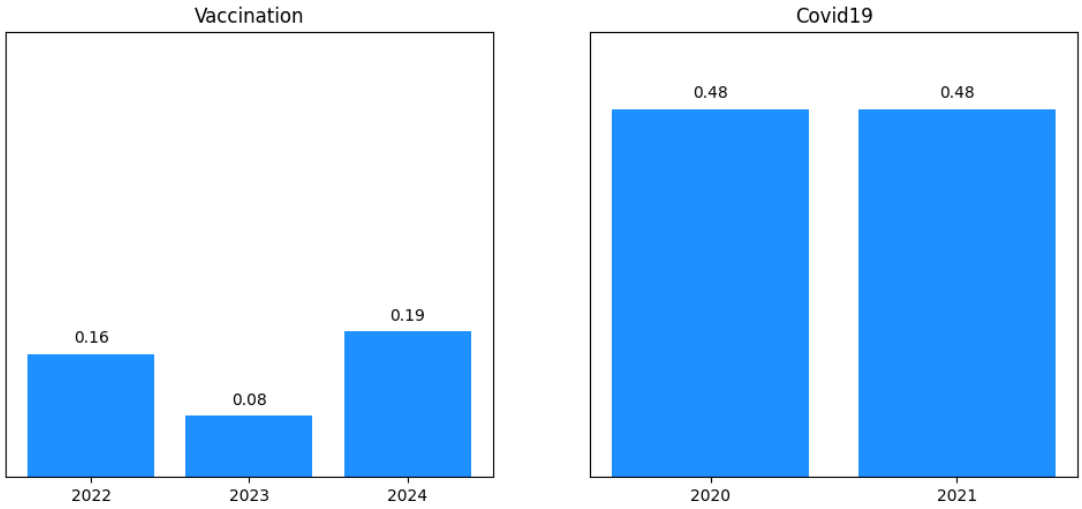

Figure 5.2: On the left the evolution over time of the correlation index between individual and neighborhood leaning for the Vaccination Reddit dataset, on the right for the COVID-19 Reddit dataset the probability of having a connection with a user of the same leaning

Looking at the graph related to COVID-19 shown in Figure 5.2, it is immediately noticeable how at the same time segments the values obtained are definitely lower than what it was analyzed before. This can certainly be an aspect due to the nature of the platform, Twitter in fact is much more related to an algorithm that suggests content, while Reddit is based on subreddits on which it is not so unusual to find situations of confrontation between users with different opinions.

Although these are different time segments, even in the case of the Vaccines dataset, slightly lower values can be seen than in the case of Twitter, and, consistent with the above, the correlation value referring to the most recent year is the highest.

5.1.2 Meso-Scale Approach Results on Long-Term Events

In this second approach, the presence of echo chambers will be assessed by considering some aggregations obtained by EVA within the conversation network, measuring for each cluster the risk of being an echo chamber, obtained through Purity and Conductance measurements. For each topic, the two largest aggregations for each sentiment expressed will be considered. The results for the Twitter dataset can be seen here:

Twitter

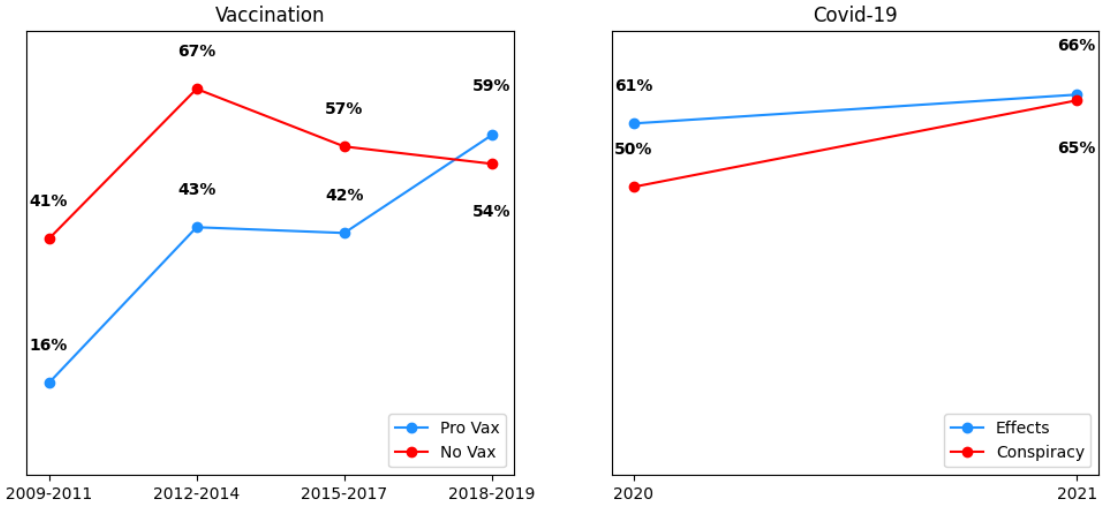

Figure 5.3: For each dataset (Vaccination and COVID-19 from Twitter), the risk of being an echo chamber over time for the two largest echo chambers identified within the analyzed data is represented

From the vaccination graph shown in Figure 5.3 it is evident how the risk of echo chambers is on average higher for the community related to the No Vax, likewise however the Pro Vax community has an increasing Echo chamber index over time, while the No Vax have a peak in 2012-2014.

On the other hand, with regard to the values obtained for the COVID-19 dataset, it can be seen that both communities have increasing values, particularly the conspiracy community that rises in one year from a risk of 50% to one of 65%. This result partly confirms what was observed in the micro-scale analysis, namely that biases and the risk of echo chambers increase with time.

To best interpret this data, it is necessary to investigate the metrics that led to these echo chamber risk values.

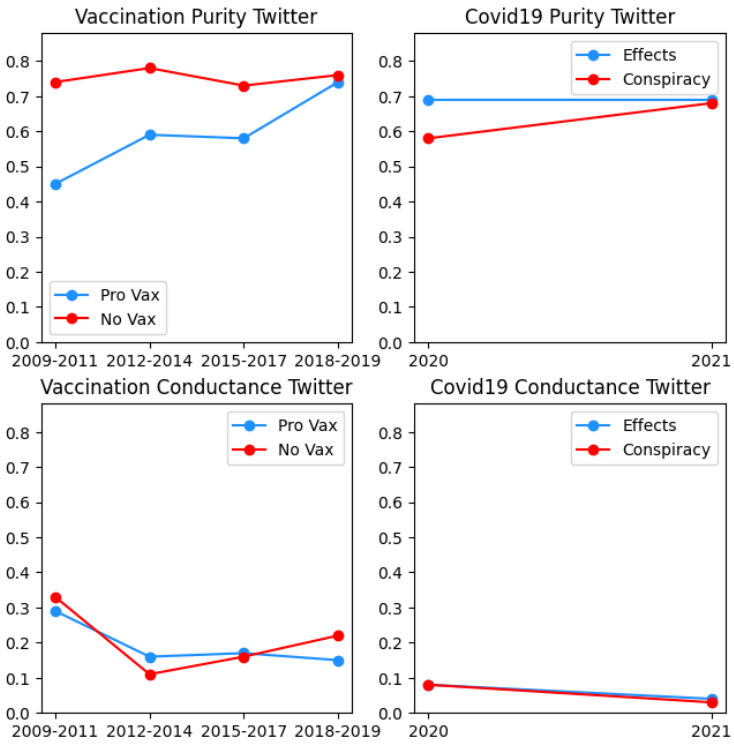

Figure 5.4: This graph is helpful in understanding how the two measures that make up the echo chamber risk index, Purity and Conductance, have changed over time for each of the two datasets (Vaccination and COVID-19 from Twitter)

For both datasets shown in Figure 5.4, it can be seen that the true discriminant that causes the risk index to vary is purity, while conductance remains at constant values. In the case of vaccinations, for the echo chamber related to Pro Vax the Purity index increases steadily over time, while the same metric but measured on No Vax remains more stable. As far as conductance is concerned, a similar trend can be observed for both modes, and in both cases the conductance values of the most recent year are well below those measured in 2009-2011. It is intuitive to think that with the advancement of new profiling algorithms this makes it even more difficult to talk to individuals outside of the community they belong to. In the case of COVID, on the other hand, while the purity index remains constant for the cluster that speaks of the effects of diseases, an increase is observed for the cluster relating to conspiracy theories in 2021. In contrast, the conductance for both

communities falls equally.

Reddit

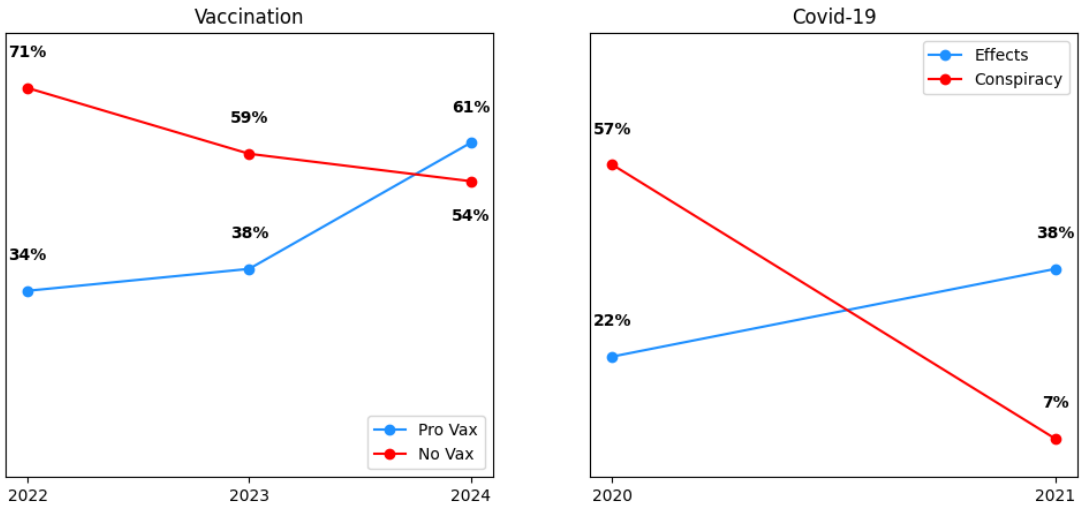

Figure 5.5: For each dataset (Vaccination and COVID-19 from Reddit), the risk of being an echo chambers over time for the two largest echo chambers identified within the analyzed data is represented

In the case of the Reddit dataset, shown in Figure 5.5 one finds once again, albeit over different time intervals, an important increase in the risk of echo chambers for the Pro Vax cluster, while on the other hand unlike Twitter one observes an important decline for the index relating to No Vax. It is possible to think that moving temporally away from the latest vaccination-related debates and with the increasing amount of scientific evidence in favor of vaccination, has led to a less heated debate, causing confidence in vaccination to grow instead. This may be related to what one expects to see in short-term events: as one moves away from major events, echo chamber detection rates tend to decrease.

Similarly, despite the fact that the Reddit COVID dataset shares the same time frame compared to the Twitter COVID dataset resulted in inconsistent results for the community related to conspiracy theories. For the community related to COVID effects, a consistent and increasing behaviour of the risk index can be observed, similarly confirming the theory that on the social Reddit one obtains values of the echo chamber detection metric that tend to be lower. In the case of the conspiracy theories cluster, on the other hand, a drop in

the risk index can be observed, which can be explained mainly by a much more medicine-focused sample and the presence of not very robust volumes. Also in this case, to better interpret this data, it is necessary to investigate the metrics that led to these echo chamber risk values.

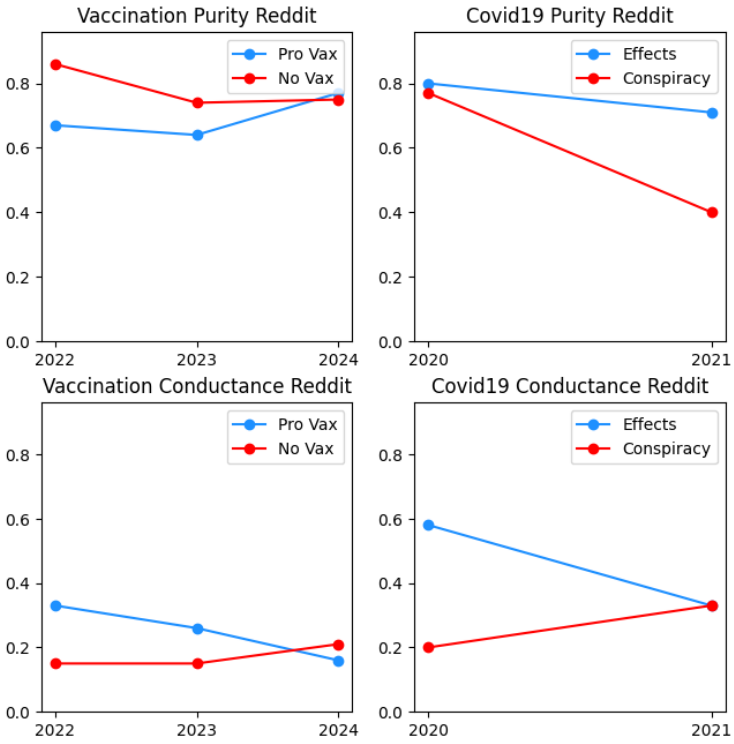

Figure 5.6: This graph is helpful in understanding how the two measures that make up the echo chamber risk index, Purity and Conductance, have changed over time for each of the two datasets (Vaccination and COVID-19 from Reddit)

As it can be seen in Figure 5.6, unlike in the case of Twitter, here there does not seem to be a measure that contributes most to the changes in the risk index seen above. There is certainly a decrease in conductance for the Pro Vax community, which over the years seems to come into less and less contact with Anti Vax users. As far as COVID is concerned, however, it can be seen that for both clusters purity is decreasing and thus the growth of the index for the COVID effects cluster is mainly driven by a decrease in conductance.

5.1.3 Macro-Scale Approach Results on Long-Term Events

The final approach analyzes the entire conversation graph partition to monitor its evolution over time. Metrics will be used to understand the flow between users in the two partitions obtained via METIS. However, the graphs may not be fully comparable due to differences in time segments and construction methods. In fact, Twitter graphs are based on mentions, while Reddit graphs are built from comments.

Twitter

Let's first examine the vaccination theme's conversation graphs by time segment:

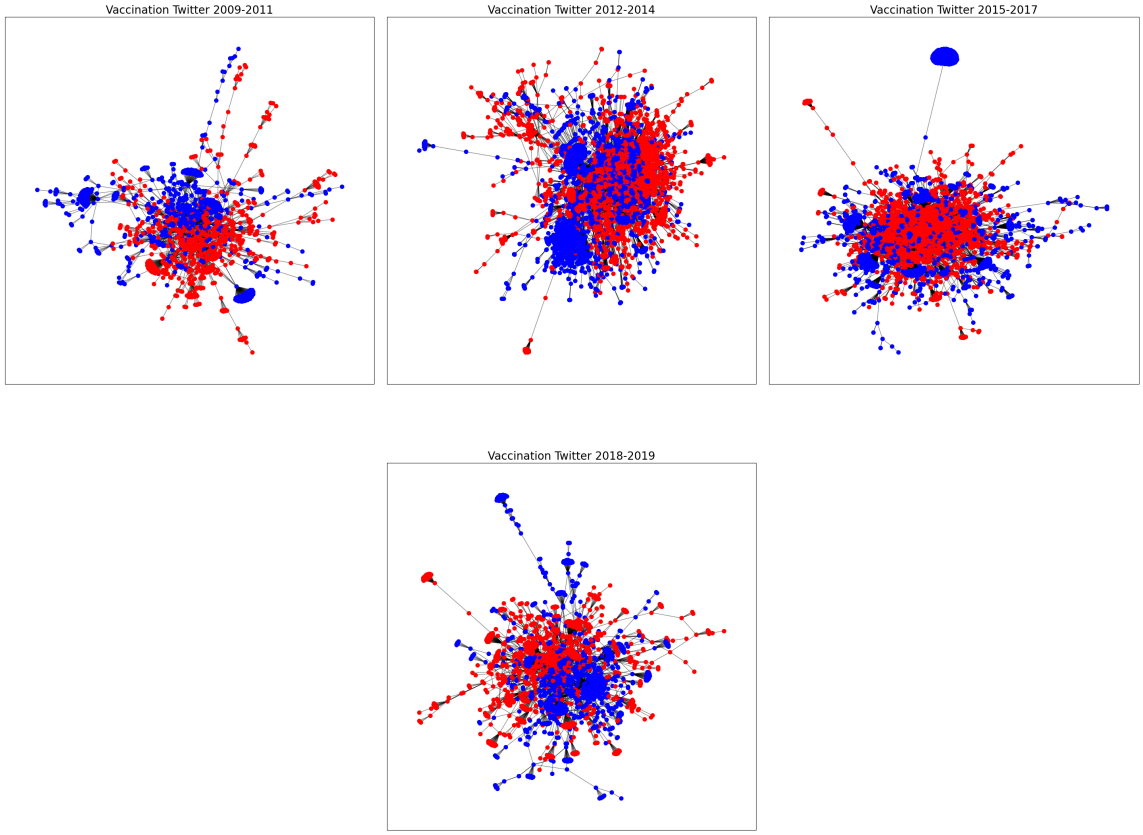

Figure 5.7: Each time segment's conversation graph is partitioned using METIS with content-based weighting for the Vaccination Twitter dataset

All 4 graphs shown in Figure 5.7 are actually quite similar and represent a single

agglomeration where the two communities are not sharply divided. The only graph that seems to present a difference is the one for the years 2015-2017, where the red community seems to be completely encompassed within the blue one and instead do not seem to be divided into two parts like the other three cases analyzed.

However, it is important to accompany the graphical evaluation with one based on some echo chamber detection metrics.

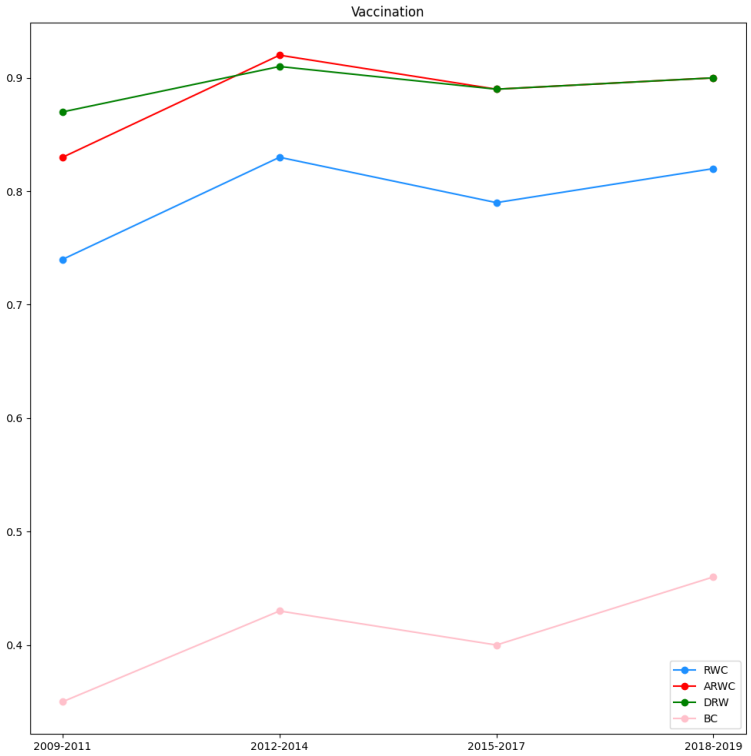

Figure 5.8: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization for the Vaccination dataset from Twitter

The graph shown in Figure 5.8 is of fairly linear interpretation and shows very high values of controversy and bias. Each of the metrics analyzed is increasing and peaks in the most recent time segment going to confirm what has been seen with the other methodologies particularly the micro-scale one. Next, the same analysis will be considered but with Twitter's COVID dataset.

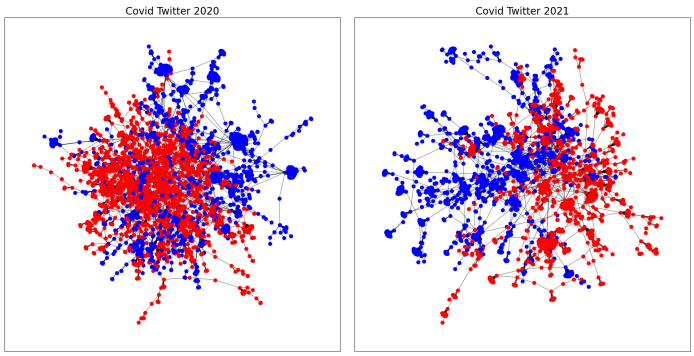

Figure 5.9: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for the Twitter COVID dataset

Comparing the graphs shown in Figure 5.9, both show a single agglomeration of communities without any significant visual separation. However, while the two communities almost completely overlap in the 2020 graph, a smaller area of overlap can be seen in 2021, suggesting that there are fewer connections between users from different communities.

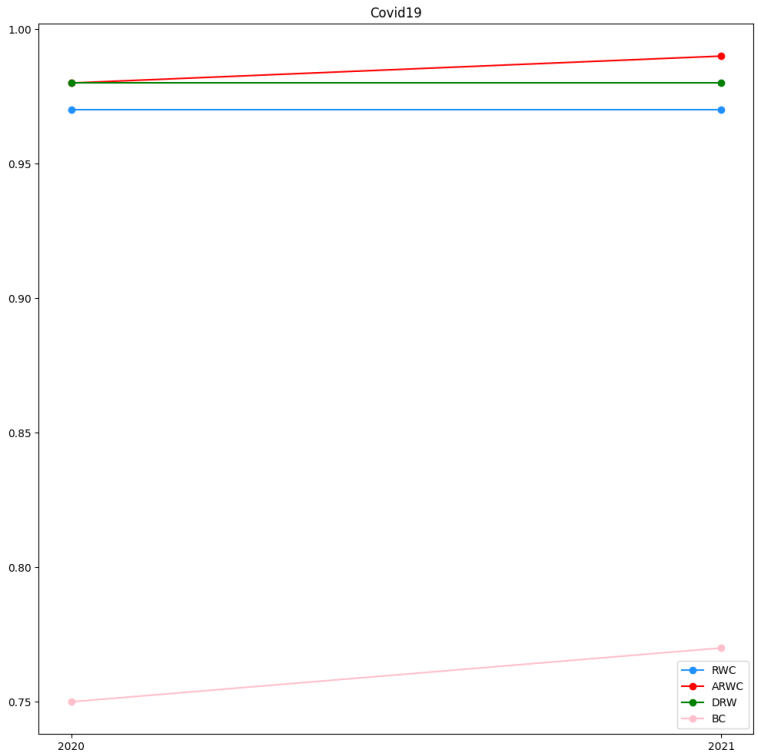

Figure 5.10: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization for the Twitter COVID dataset

Looking at Figure 5.10 it can be seen a similar situation to the meso-scale approach, in fact there is an increase in metrics between 2020 and 2021, particularly for the BC and ARWC measures, confirming once again how the risk of echo chamber formation tends to increase over time.

Reddit

First, let's look at the conversation graphs for each time segment for the vaccination theme

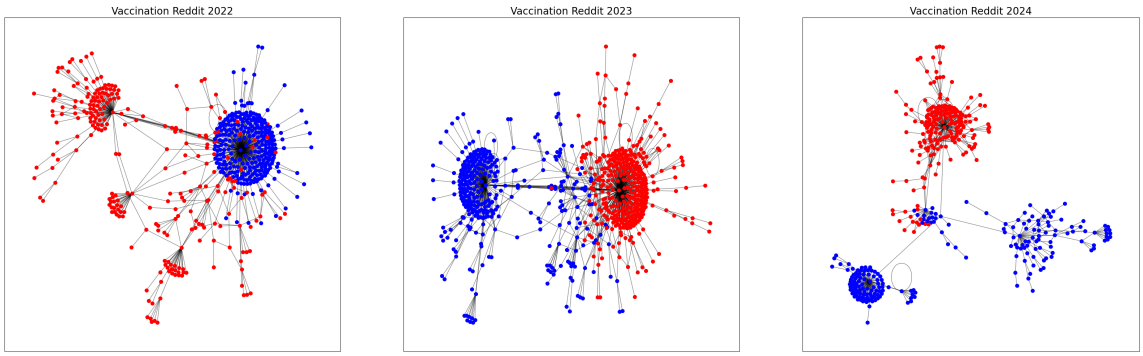

Figure 5.11: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for the Vaccination dataset from Reddit

In these graphs shown in Figure 5.11, however, unlike those seen previously, a clearer distinction between the two communities can be seen. In fact, there are many more intra-community connections than between elements of different communities, leading to the graphic creation of two distinct areas.

This effect seems to be greater as time changes: if in the first graph one notices some red users in the blue area, already in the second time segment one can see that this effect is no longer there. Finally, in the 2024 graph, not only is the division even sharper, but the blue community is even more divided.

However, it is important to accompany the graphical evaluation with one based on some echo chamber detection metrics.

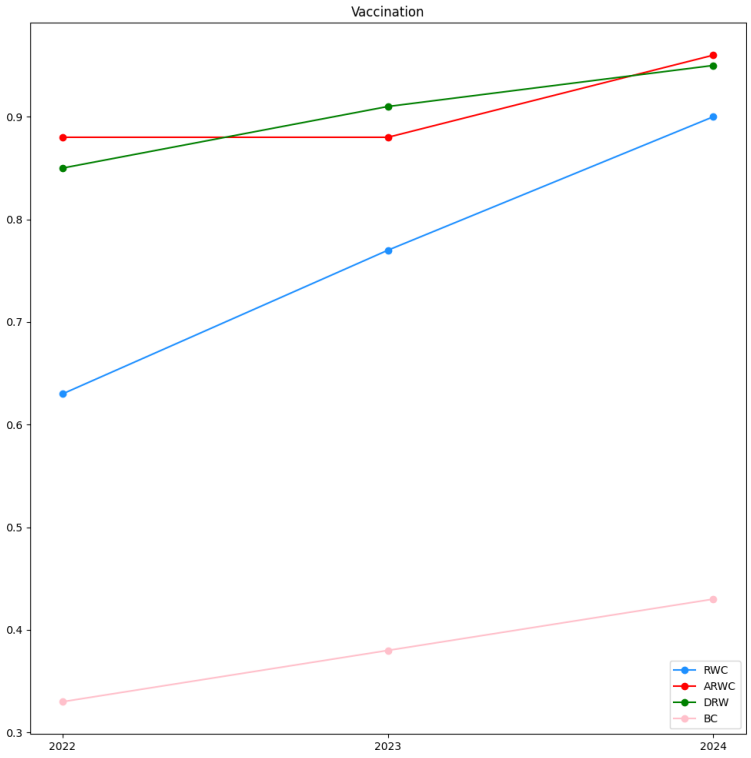

Figure 5.12: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization for the Vaccination dataset from Reddit

Consistent with previous approaches, in Figure 5.12 can be noted that all metrics are growing during the time segments considered, in particular there seems to be an exponential growth of RWC, which in 2022 stood at just above 0.6, while in 2024 it touches a value close to 0.9. Next the same analysis with the Reddit COVID dataset will be considered.

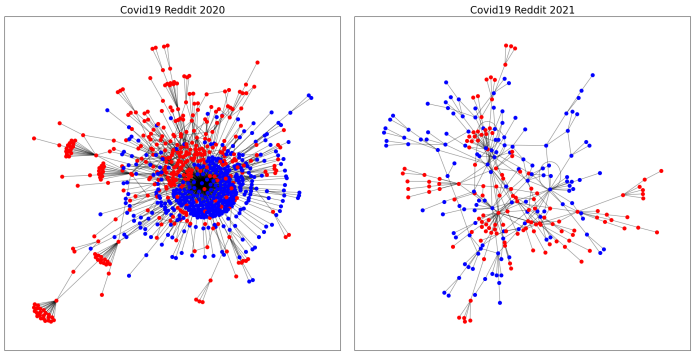

Figure 5.13: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for the Reddit COVID dataset

In Figure 5.13 it can be seen that if in the meso-scale approach, although considering the same time span, differences could be observed with respect to what was analysed on Twitter, in this case the graphical evaluation of the macro-scale approach reveals several similarities with what was observed on Twitter. Also in this case a single core divided into two sections can be noted although 2021 has fewer users.

However, it is important to accompany the graphical evaluation with one based on some echo chamber detection metrics.

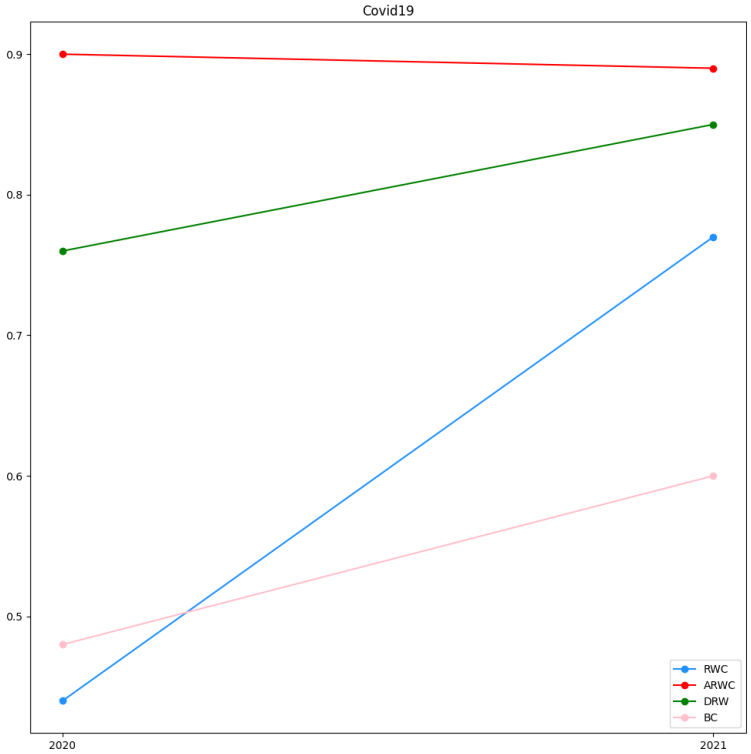

Figure 5.14: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization for the Reddit COVID dataset

The metrics shown in Figure 5.14 confirm the similarities intuited in the graphical evaluation. In fact, from the graph a coerence with the Twitter's result can be seen. The metrics considered tend to grow over time, in particular that of Random Walk Controversy, which from values of less than 0.3 in 2021 has reached a value of 0.6. Interestingly, all 4 metrics have higher values on Twitter than those observed on Reddit, further confirming the intuition that there is a platform effect.

5.2 Short-Term Events

This section will analyse the results of the three approaches for the two short-term topics: Brexit and ChatGPT. Again, the focus will be on highlighting the differences between platforms and on offering some contextual information related to the events under analysis.

5.2.1 Micro-Scale Approach Results on Short-Term Events

Again, the first analysis refers to the individual user and to measuring the relationship between the individual's thinking with that of the users who come into contact with them. It is useful to mention that to identify user leaning in the case of Brexit a topic-related label (Pro/Aganist) was used while in the case of ChatGPT the sentiment indicator obtained through VADER. For this reason in the case of ChatGPT it was possible to use the correlation measure, while in the case of Brexit it was necessary to use a probability measure, understood as the probability that the neighborhood has the same leaning as the user under consideration, simply calculated as favorable cases over total cases.

Twitter

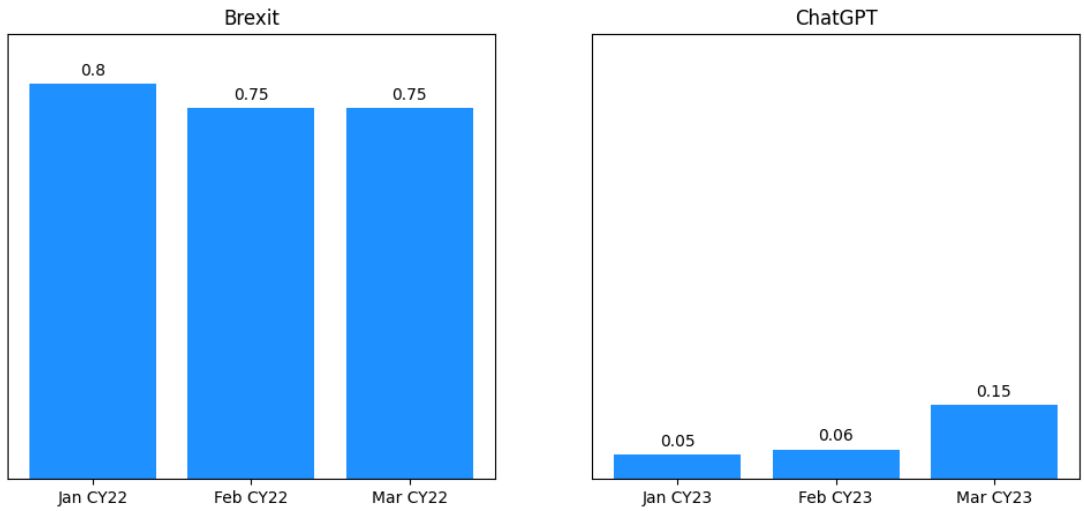

Figure 5.15: On the left the evolution over time of the probability of having a connection with a user of the same leaning for the Brexit Twitter dataset, on the right the correlation index between individual and neighborhood leaning for the ChatGPT Twitter dataset

As mentioned earlier, the Brexit dataset refers to the discussion around new rules that came into effect on January 2022. Thus, In Figure 5.15 it can be seen that the probability goes down with time, almost in the opposite way as seen in long-term events. However, it is intuitive to think that with the passage of time in relation to the event under consideration passes, the Twitter conversation drops in interest leading consequently to a lowering of indices. One can observe something similar on ChatGPT as well: in this case, it was decided to monitor a time frame that goes by the period after the launch of the software and the reaction of users to some updates available from the end of February 2023.

Reddit

Again, the Reddit API did not allow us to extract data for the same temporal endpoint, an objective made even more difficult by the fact that it was required to extract a much narrower time span than before. While in the case of ChatGPT it was possible to get close enough to the required temporal endpoint and managed to extract data on a monthly basis, this was not possible for the Brexit case, where one is forced for volume reasons to consider three years. Therefore, the opportunity was taken to also try to analyze the Brexit phenomenon as a long-term event, to see if interesting patterns of analysis could emerge.

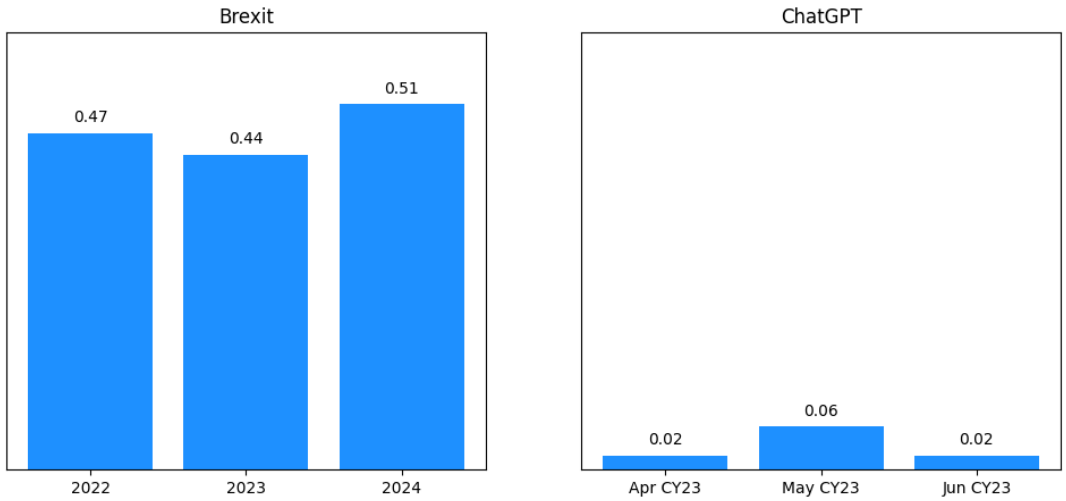

Figure 5.16: On the left the evolution over time of the probability of having a connection with a user of the same leaning for the Brexit Reddit dataset, on the right the correlation index between individual and neighborhood leaning for the ChatGPT Reddit dataset

If, on the other hand, one considers Brexit as a long-term event, one can see that the pattern shown in Figure 5.16 repeats itself, namely that the most recent year coincides with the year with the highest probability. This is probably also due to the fact that the Brexit debate has certainly intensified with the 2024 general election in the United Kingdom.

On the topic of ChatGPT, it is further noticeable how moving away from the reference event continues to have lower and lower correlation values despite the fact that it is clear that the theme related to the AI platform is much less controversial than the others considered. In general, again, the values obtained for Reddit are lower than those obtained for Twitter.

5.2.2 Meso-Scale Approach Results on Short-Term Events

In this second approach, the presence of echo chambers will be assessed by considering some aggregations obtained from EVAs within the conversation network and measuring for each cluster the risk of being an echo chamber. For each topic the two largest aggregations for each sentiment expressed will be considered.

Twitter

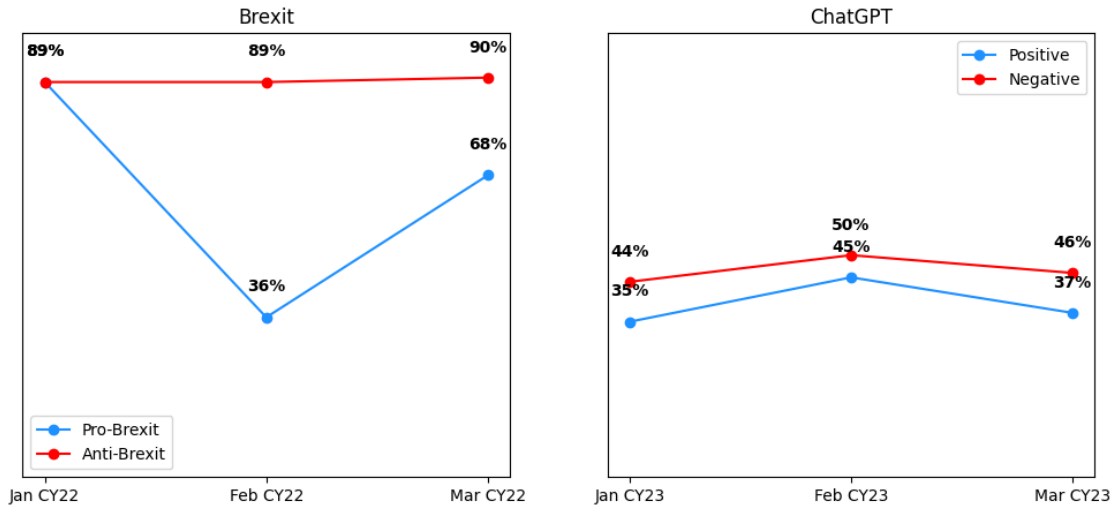

Figure 5.17: For each dataset, the risk of being an echo chambers over time for the two largest echo chambers identified within the analyzed data (Brexit Twitter dataset and ChatGPT Twitter dataset) is represented

Regarding the Brexit issue, In Figure 5.17 it can be seen that the risk index for AntiBrexit is high and constant throughout the period. On the other hand, as far as ProBrexit is concerned, the index seems to be more variable over the period considered and peaks precisely in January 2022, i.e., the month on which the main event in the dataset falls.

On the other hand, as far as ChatGPT is concerned, lower values of the index can be seen immediately, a further indication that the Gen AI theme is much less controversial than the others considered. Moreover, the two clusters seem to have similar behavior, with a slight peak in February 2023, the month in which the updates referred to earlier were launched To best interpret this data, it is necessary to investigate the metrics that led to these echo chamber risk values.

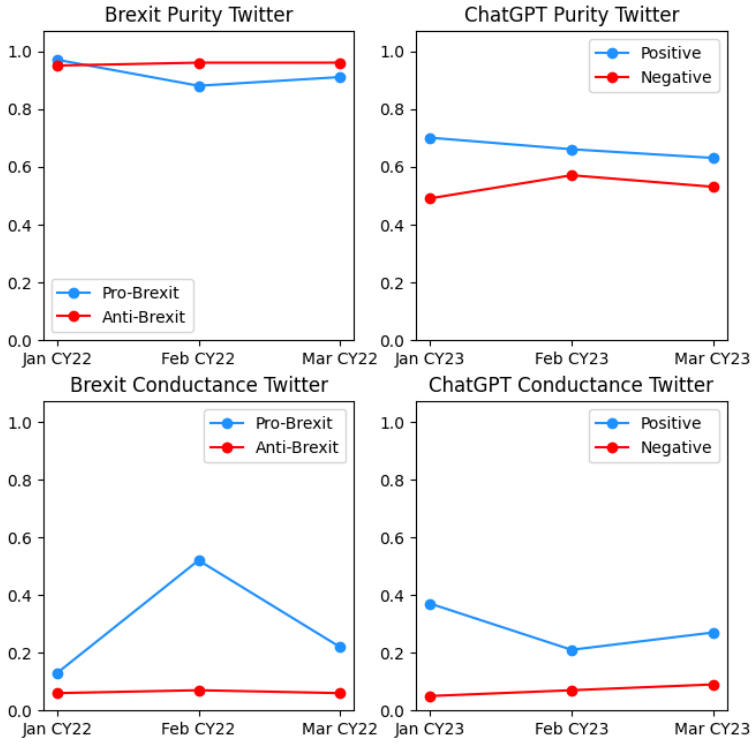

Figure 5.18: This graph is helpful in understanding how the two measures that make up the echo chamber risk index, Purity and Conductance, have changed over time for each of the two datasets (Brexit Twitter dataset and ChatGPT Twitter dataset)

In the case of Brexit, In Figure 5.18 it can be seen that the drop in the metric for ProBrexit is entirely explained by the change in conductance, which peaked at 0.50 in February. In contrast, purity seems to remain stable, so the conclusion can be drawn that

in February the ProBrexit cluster had more connections with the AntiBrexit cluster.

On the other hand, with regard to the ChatGPT theme, the explanation of the two February peaks for the two communities is explained in a completely different way. For the community with negative sentiment, the values of the indices seen above are explained by a peak in Purity in February, while conductance is albeit on low values always increasing. For the community with a positive feeling, however, the situation is diametrically opposed: with decreasing purity throughout the period while a particularly low value of conductance in February 2023 that consequently raises the risk index in that month.

Reddit

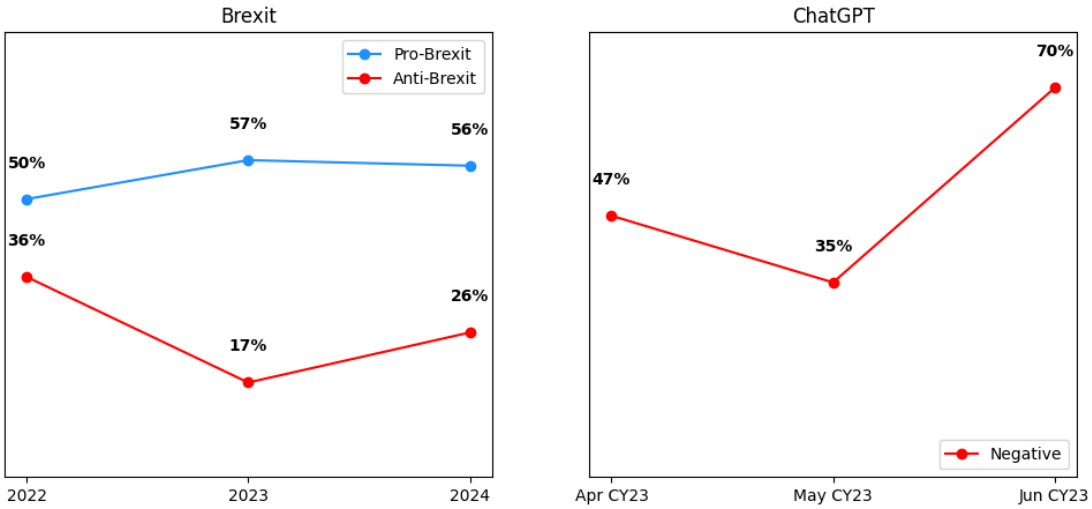

Figure 5.19: For each dataset, the risk of being an echo chambers over time for the two largest echo chambers identified within the analyzed data is represented. For the case of ChatGPT only one line is shown because no additional significantly numerous communities are identified

In part of Figure 5.19 regarding ChatGPT, only one curve was presented since an equally present community was not found for those who speak positively about the topic. These users are more scattered within the network and do not form unique aggregations.

Analysing Brexit as a long-term event, one notices a very different situation compared to the analysis with a monthly timeframe. First, the ProBrexit community seems to have a higher echo chamber risk index throughout the period than the AntiBrexit community, which seems to have its peak precisely in 2022, the year previously analyzed in detail.

On the other hand, with regard to the community with negative sentiment, the only one intercepted in the ChatGPT dataset, there is an important growth in the risk index in June 2023. Investigating the history of the platform, this may be justified by a sudden drop in users that the platform suffered in June after the controversy there had been in the previous months over privacy issues.

Also in this case, to better interpret this data, it is necessary to investigate the metrics that led to these echo chamber risk values.

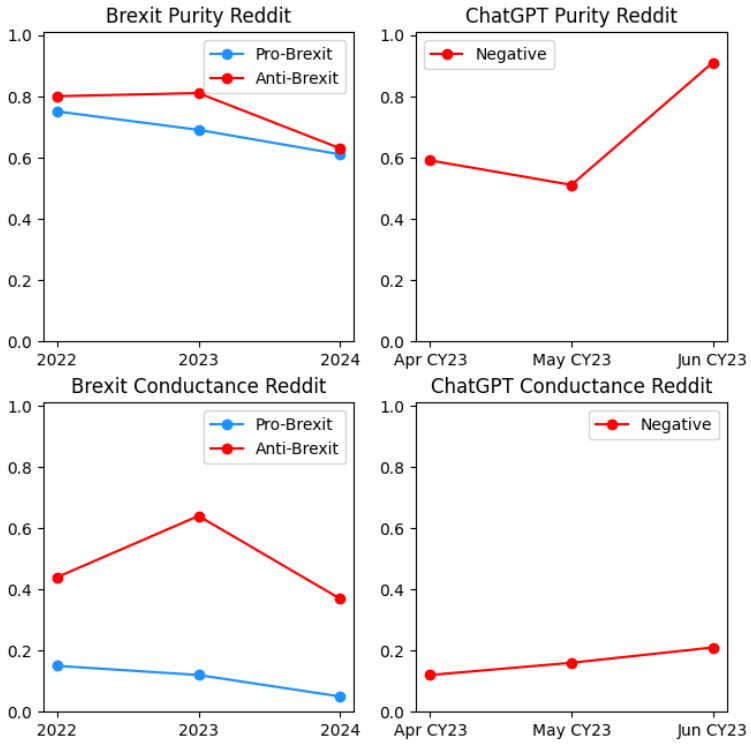

Figure 5.20: This graph is helpful in understanding how the two measures that make up the echo chamber risk index, Purity and Conductance, have changed over time for each of the two datasets (Brexit Reddit dataset and ChatGPT Reddit dataset

Looking at the Brexit-related graphs of Figure 5.20 it can be seen a decline in purity for both communities, again the main variations come from conductance which is steadily decreasing for the ProBrexit community while it is less linear and peaking in 2023 for the Antibrexit community. In contrast, the June peak for the ChatGPT dataset is fully explained by a major growth in Purity, in contrast to a continuously increasing conductance.

5.2.3 Macro-Scale Approach Results on Short-Term Events

Finally, the last approach considers the entire partition of the conversation graph, with the goal of monitoring its evolutions over time. Some metrics will be considered to understand the behavior of the conversation flow between users of the two different partitions obtained through METIS. Again, it is likely that the graphs are not entirely comparable both because of the different time segments used but also because they are constructed differently.

Twitter

First, let's look at the conversation graphs for each time segment for the Brexit theme

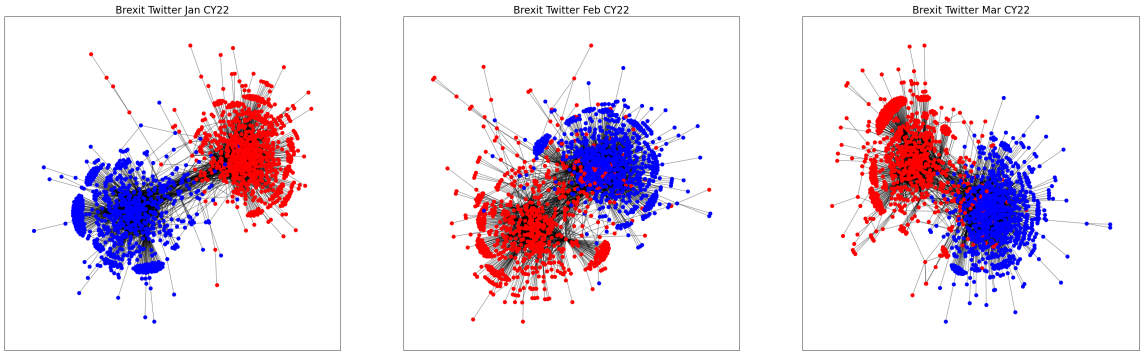

Figure 5.21: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for the Brexit Twitter dataset

Already from the graphical evaluation shown in Figure 5.21, one can see a strongly polarised situation, in which the two communities are quite distinct and connected by a few arcs. Particularly in the graph of January 2022 the greatest polarisation is observed, where the two communities are further apart and with fewer links between them. Already in the following months, there seems to be more interaction between the communities, confirming the theory that in the case of short-term events, the peak of polarisation and thus of being inside an echo chambers risk is greater in the periods corresponding to the event of interest, and then fades with the passage of time.

However, it is important to accompany the graphical evaluation with one based on some echo chamber detection metrics.

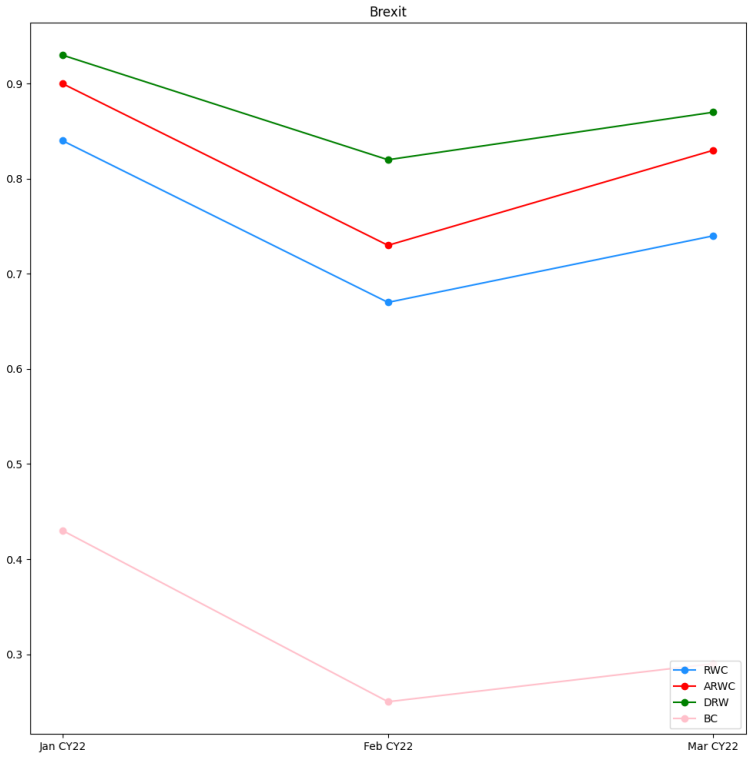

Figure 5.22: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization for the Brexit Twitter dataset

Again, the metrics shown in Figure 5.22 seem to underline a consistent situation with the previous approaches and the graphical evaluation. The peak of all 4 metrics occurs at January 2023, underlining how after the peak period the risk of echo chambers has declined. Next, the same analysis will be considered but with the ChatGPT Twitter dataset

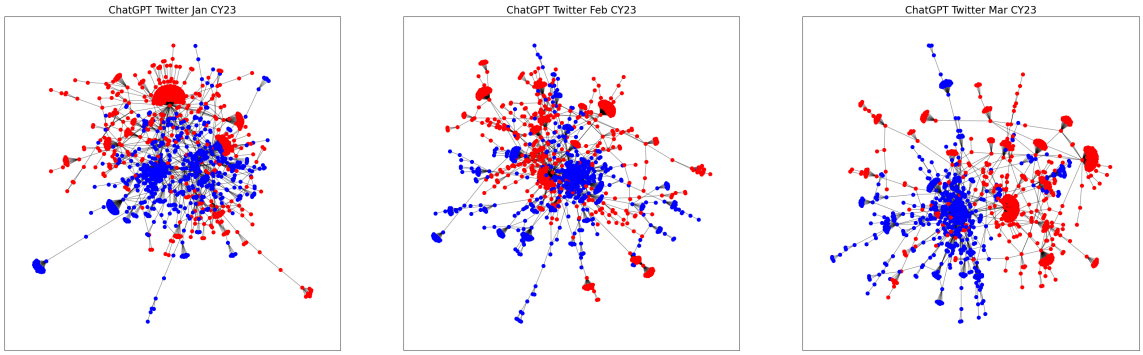

Figure 5.23: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for ChatGPT Twitter dataset

The approach in Figure 5.23 confirms that this topic is less controversial than the others. The March 2023 conversation graphs show a clear division between the two communities, though a central core where both coexist is present in all three graphs. However, this graphical analysis should be supported by echo chamber detection metrics.

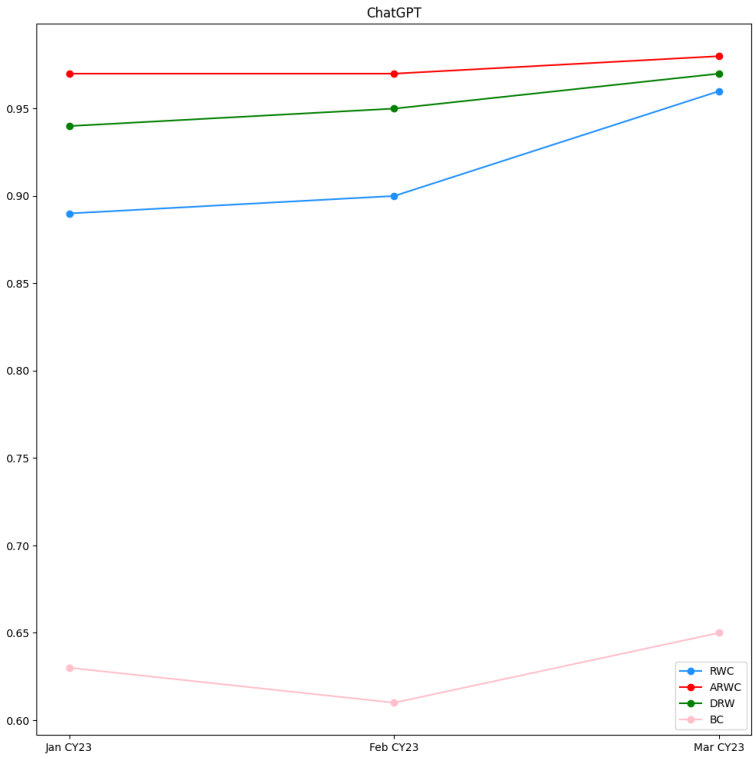

Figure 5.24: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization for the ChatGPT Twitter dataset

Also at this juncture, in Figure 5.24 it can be see that the main metrics of the Random Walk Controversy family have an increasing trend over time. Despite having identified from previous analyses that the topic of ChatGPTs and GenAI is something less controversial than topics such as Brexit or COVID, the Random Walk Controversy measures still have very high values close to 0.90. The only metric that deviates is that of Boundary Connectivity, which drops in February 2023 before rising and reaching its highest value in March.

Reddit

First, let's look at the conversation graphs for each time segment for the Brexit theme

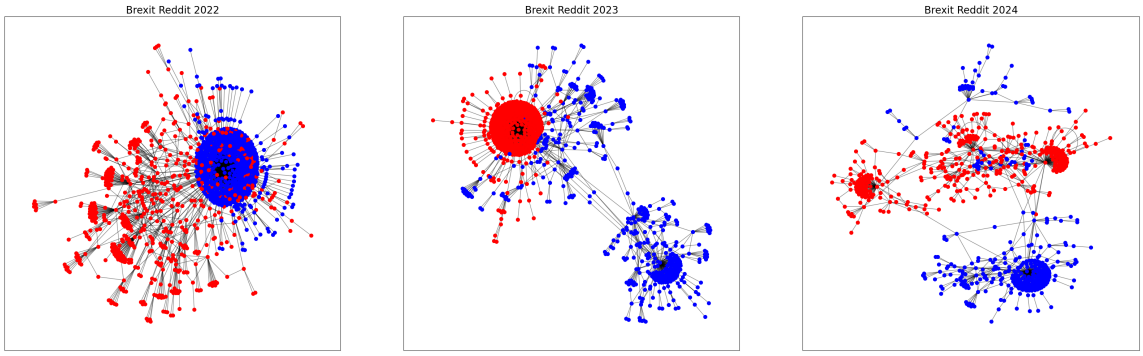

Figure 5.25: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for the Brexit Reddit dataset

Although at different times from Twitter, here too in Figure 5.25 one can see well-defined communities without overlapping. The main difference from Twitter is that the two communities have different conformations. In the first graph of 2022, it can be seen that the blue community is all about one influential user, while the red community is much more dispersed and not focused on one user. Again, there seems to be a reduction in the arcs connecting the two communities over time.

However, it is important to accompany the graphical evaluation with one based on some echo chamber detection metrics.

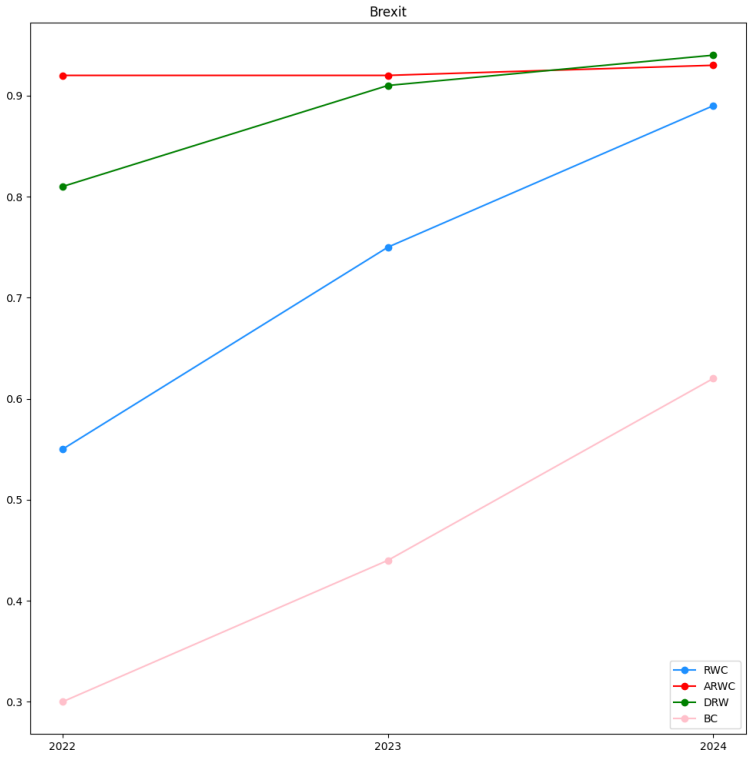

Figure 5.26: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization of the the Brexit Reddit dataset

The metrics shown in Figure 5.26 confirm what the graphical analysis said, which grow considerably during the period considered, confirming an increase polarisation with the passage of time. The only metric that is an exception is the ARWC, which is instead stable. Next, the same analysis will be considered but with the ChatGPT Reddit dataset

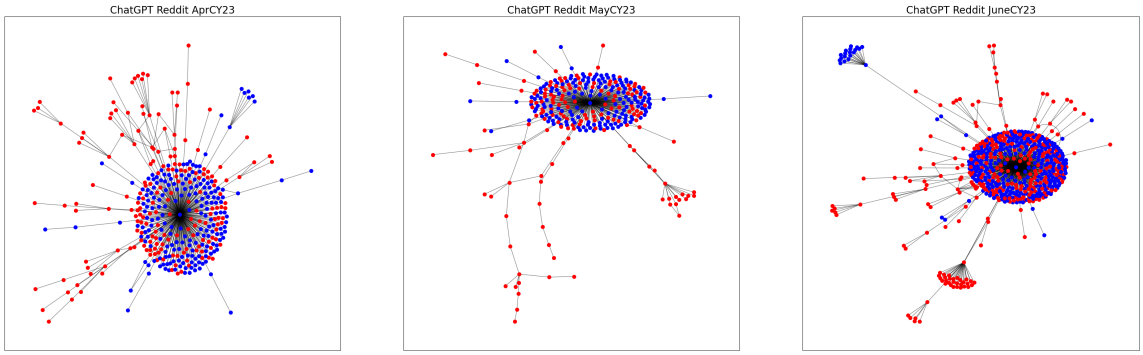

Figure 5.27: For each time segment, one can see the partition of the conversation graph using METIS with the addition of the weight regarding the content for the ChatGPT Reddit dataset

In spite of the fact that for reasons related to the nature of the graph, a different conformation from the one seen with Twitter is noticeable, it is evident in Figure 5.27 that there are not two well-defined and graphically divided communities. For this reason, lower metric values are to be expected than with the other analysis topics.

However, it is important to accompany the graphical evaluation with one based on some echo chamber detection metrics.

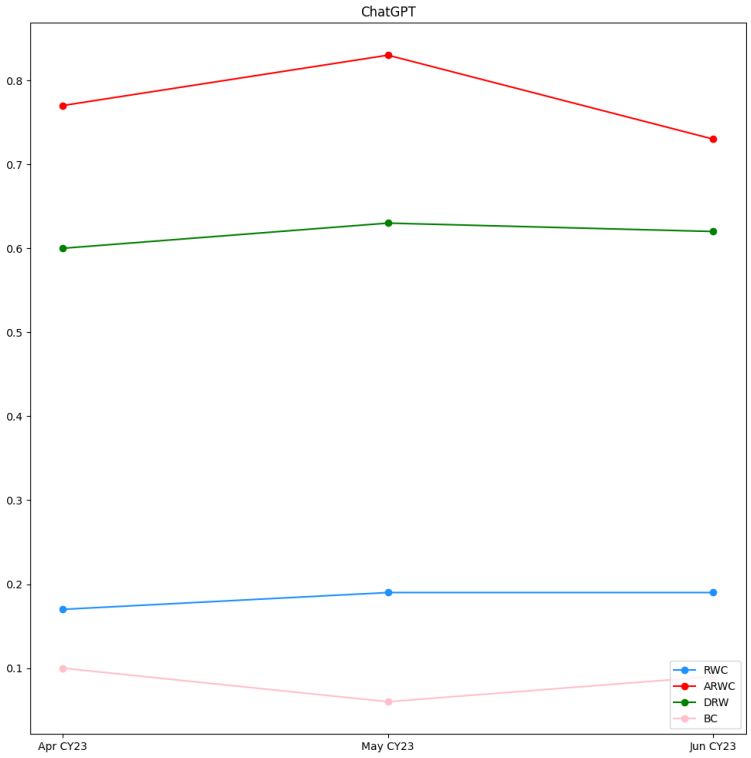

Figure 5.28: The graph represents for each time interval the metrics used to measure the degree of controversy and polarization of the ChatGPT Reddit dataset

As was to be expected, the values of the metrics shown in Figure 5.28, although they may still seem high, are much lower than in previous experiments. The metrics generally present constant values or at least non-significant year-on-year variations, the only exception being the Authoritative Random Walk Controversy, which seems to have a peak around May and then falls in June, thus not confirming what was seen in the meso-scale approach where there was a peak in June.

5.3 Main General Evidences

After dwelling for each approach on the outcome of the individual dataset, this section aims to take an overview and extract some more general considerations that may apply more broadly to the analysis and monitoring of echo chambers over time.

In the light of the analyses carried out, three main pieces of evidence stand out:

- For the two long-term events analysed, i.e. COVID-19 and the issue of vaccinations, it can be seen that regardless of methodology and platform, the last year taken into account is the one with the highest values of the various polarisation metrics considered. In fact, an increasing curve can be seen over time, particularly for vaccinations, suggesting that the passage of time favours polarisation and remaining immersed in echo chambers creates a vicious circle that leads to further isolation and contact with fewer different leaning users.

- In the two short-term events, i.e. Brexit and ChatGPT, the aim was to analyse individual events that took place over a shorter period of time and monitor their behaviour in the following months. For example, in the case of Brexit, the aim was to monitor the evolution of the conversation after the introduction of certain regulations that came into force in January 2022. This type of analysis confirmed that in the hot months in which the reference event takes place, echo chamber detection metrics tend to assume higher values. The more current and debated a topic is, the greater the bias and risk of being in an echo chambers. Conversely, as time passes since the reference event, the indicators decline as does the interest around the event.

- The last piece of evidence concerns what emerged from the comparison of data from two different platforms on the same topic. Although there are differences in terms of both the time segments chosen and the methodology used to create the conversation graph, it was evident in several scenarios that the values of the echo chambers metrics obtained were significantly higher on Twitter than the observation of the counterpart on Reddit, although they still presented consistent trends. The result is even more evident in the measurements that were made regarding the two

COVID-19 datasets where the time segments were the same between the two platforms. Having used hybrid echo chamber detection methods and thus having taken into account both structural and content aspects, it was decided to look for the reason for this delta between the characteristics of the two platforms. It is known that Twitter is a platform that is much more conducive to the formation of echo chambers than Reddit precisely because of the way it is constructed. Reddit is based on the discussion of topics via subreddits as if it were a real forum, favouring in-depth discussions and making interaction between people with different thoughts much easier. In contrast, Twitter, focuses on quick and trending conversations. Discussions on Reddit turn out to be more thoughtful while Twitter's algorithm favours quick and instant response conversations. These are some of the elements that may explain the different results between the two platforms, although more experiments would need to be carried out to fully investigate this aspect.

6. Conclusions and Further Research

In this thesis work, an attempt was made to introduce the time variable to the different methodologies in the literature for the identification of echo chambers in social networks. After an initial phase introducing the problem and then a phase presenting the state-of-the-art in the field of echo chamber detection, a comparative analysis was carried out to monitor certain controversial topics such as Brexit, opinions on ChatGPT, vaccinations and the COVID-19 pandemic over time. In order to perform the temporal analysis in the best possible way and to be able to differentiate the developments that a controversial event may have over time, the topics were divided into two case studies: long-term events, i.e. debates that have been going on for decades, and short-term events, i.e. events for which there is great debate only for a short period of time. In addition, in order to test possible differences between platforms, it was decided to extract data on the topic in question from both Twitter and Reddit.

Among the echo chamber detection techniques available in the literature, hybrid techniques were chosen, as they were found to be the most capable of generalisation between platforms and controversial topics. Three types of approaches were tested: micro-scale, where the level of correlation between individual leaning and neighbourhood leaning is investigated, meso-scale, where the risk of being an echo chambers of each identified community is calculated according to the approach of Morini et al. (2021) [50] and finally a macro-scale approach where the entire conversation graph is taken into account and partitioned according to the approach of Villa et al. (2021) [31].

The results obtained showed that for long-term events, regardless of the techniques used and the platform taken into consideration, the echo chamber detection metrics tend to grow over time, peaking at the most recent year available. On the contrary, for short-term events, a peak of the same metrics was noted, but unlike in the previous case in correspondence with the period encompassing the core event of analysis, and then gradually fading out as one moves away in time. Furthermore, the cross-platform comparison resulted in significantly lower metrics on Reddit than on Twitter, for reasons that are mainly

attributable to the structure of the platform, which favours the exchange of opinions and consequently leads to lower polarisation indices.

However, it is important to emphasise that this thesis work has limitations, mainly due to the limited availability of data. In addition to the fact that the chosen topics do not exhaustively represent the contexts in which one can deal with echo chambers, the choice was forced due to the limitations of the data extraction methods of both Twitter and Reddit. With regard to the former, it is no longer possible to freely extract data from the platform, so it was necessary to select datasets that already existed and were used in the literature. With regard to Reddit's platform, although it is possible to freely extract a large amount of data from different subreddits, but it was not possible to consider precise temporal endpoints. This resulted in not having a comparison between platforms with the same timeframe. Furthermore, all the main choices made in conducting this experiment were made on the basis of suggestions from the literature. All the tools used were selected because they had already been tested with positive results, but this does not mean that they are the best in this context and for these datasets.

The realisation of these limitations gives rise to a reflection on the future development of this work. Surely one of the goals of future research must be to focus on a greater diversification of the data and tools used, perhaps expanding experiments to platforms beyond Twitter and Reddit, especially at this time with the advent of short video platforms such as TikTok. Furthermore, it would be interesting to be able to go beyond the limitations of data mining tools and be able to perform the same analysis by precisely comparing time segments between platforms.

Bibliography

[1] L. Floridi. *The Fourth Revolution: How the Infosphere is Reshaping Human Reality*. OUP Oxford, 2014. ISBN: 9780199606726. URL: https://books.google.it/books?id=hOF_AwAAQBAJ.

[2] Jeff Desjardins. *How much data is generated each day?* 2019. URL: https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/.

[3] Tim O'Reilly. "What Is Web 2.0: Design Patterns And Business Models For The Next Generation Of Software". In: *University Library of Munich, Germany, MPRA Paper* 65 (Jan. 2007).

[4] Andreas Kaplan and Michael Haenlein. "Users of the World, Unite! The Challenges and Opportunities of Social Media". In: *Business Horizons* 53 (Feb. 2010), pp. 59–68. DOI: 10.1016/j.bushor.2009.09.003.

[5] European Parliament. *Media & News Survey 2023*. 2023. URL: https://europa.eu/eurobarometer/surveys/detail/3153.

[6] B.M. Gross and B.M. Gross. *The Managing of Organizations: The Administrative Struggle*. The Managing of Organizations v. 1. Free Press of Glencoe, 1964. URL: https://books.google.it/books?id=boSFAAAAMAAJ.

[7] A. Toffler. *Future Shock*. CBC learning systems v. 644. Random House, 1970. ISBN: 9780394425863. URL: https://books.google.it/books?id=atGCm-nvJVEC.

[8] Orrin Edgar Klapp. *Overload and boredom: Essays on the quality of life in the information society*. Greenwood Publishing Group Inc., 1986.

[9] Gerald J Kowalski. *Information retrieval systems: theory and implementation*. Vol. 1. springer, 2007.

[10] D. Kirkpatrick. *The Facebook Effect: The Inside Story of the Company That Is Connecting the World*. Business book summary. Simon & Schuster, 2011. ISBN: 9781439102121. URL: https://books.google.it/books?id=PxTvbM-VCPEC.

[11] E. Pariser. *The Filter Bubble: What the Internet is Hiding from You*. Penguin Books, 2012. ISBN: 9780241954522. URL: https://books.google.it/books?id=Qn2ZnjzCE3gC.

[12] Axel Bruns. "Echo chamber? What echo chamber? Reviewing the evidence". In: *6th Biennial Future of Journalism Conference (FOJ17)*. 2017.

[13] Lee Rainie and Aaron Smith. *The Political Environment on Social Media*. Pew Research Center. 2016. URL: https://www.pewresearch.org/internet/2016/10/25/the-political-environment-on-social-media/.

[14] Raymond S Nickerson. "Confirmation bias: A ubiquitous phenomenon in many guises". In: *Review of general psychology* 2.2 (1998), pp. 175–220.

[15] Cass R Sunstein. "The law of group polarization". In: *University of Chicago Law School, John M. Olin Law & Economics Working Paper* 91 (1999).

[16] Pietro Corazza. "Filter Bubbles and Echo Chambers: Pre-digital Origins and Elements of Novelty-Reflections from a Media Education Perspective". In: *Formazione & insegnamento* 20.1 Tome II (2022), pp. 856–867.

[17] Matthew Gentzkow et al. "Social media and the spread of misinformation: Evidence from the 2016 US presidential election". In: *Journal of Economic Perspectives* 31.2 (2017), pp. 211–36.

[18] Claire Wardle and Hossein Derakhshan. *Information disorder: Toward an interdisciplinary framework for research and policymaking*. Vol. 27. Council of Europe Strasbourg, 2017.

[19] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". In: *science* 359.6380 (2018), pp. 1146–1151.

[20] Cass Sunstein. *# Republic: Divided democracy in the age of social media*. Princeton university press, 2018.

[21] Harald Holone. “The filter bubble and its effect on online personal health information”. In: *Croatian medical journal* 57.3 (2016), p. 298.

[22] Mostafa M. El-Bermawy. “Your Filter Bubble is Destroying Democracy”. In: *Wired* (2016).

[23] Eli Pariser. *Beware online “filter bubbles”*. TED: Ideas worth spreading. Mar. 2011. URL: https://www.ted.com/talks/eli_pariser_beware_online_filter_bubbles.

[24] *Social Analytics*. Gartner. URL: https://www.gartner.com/en/information-technology/glossary/social-analytics (visited on 05/07/2024).

[25] Reinhard Diestel. *Graph theory*. 4th. Berlin, Heidelberg: Springer, 2010.

[26] Marco Viviani. *Metrics for Social Network Analysis*. Lecture slides. Course on Social Media Analytics, University of Milano Bicocca. 2023.

[27] Mark EJ Newman. “Assortative mixing in networks”. In: *Physical review letters* 89.20 (2002), p. 208701.

[28] LC Freeman. “A set of measures of centrality based on betweenness”. In: *Sociometry* (1977).

[29] Alex Bavelas. “Communication patterns in task-oriented groups”. In: *The journal of the acoustical society of America* 22.6 (1950), pp. 725–730.

[30] Sumit Kumar Gupta, Dhirendra Pratap Singh, and Jaytrilok Choudhary. “A review of clique-based overlapping community detection algorithms”. In: *Knowledge and Information Systems* 64.8 (Aug. 2022), pp. 2023–2058. DOI: 10.1007/s10115-022-01704-6. URL: https://doi.org/10.1007/s10115-022-01704-6.

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

[32] Salvatore Citraro and Giulio Rossetti. “Identifying and exploiting homogeneous communities in labeled networks”. In: *Applied Network Science* 5.1 (2020), p. 55.

[33] Marco Viviani. *Sentiment Analysis*. Lecture slides. Course on Social Media Analytics, University of Milano Bicocca. 2023.

[34] Clayton Hutto and Eric Gilbert. “Vader: A parsimonious rule-based model for sentiment analysis of social media text”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 8. 1. 2014, pp. 216–225.

[35] Marco Viviani. *Topic Modeling*. Lecture slides. Course on Text Mining and Search, University of Milano Bicocca. 2023.

[36] Scott Deerwester et al. “Indexing by latent semantic analysis”. In: *Journal of the American society for information science* 41.6 (1990), pp. 391–407.

[37] David M Blei, Andrew Y Ng, and Michael I Jordan. “Latent dirichlet allocation”. In: *Journal of machine Learning research* 3.Jan (2003), pp. 993–1022.

[38] Maarten Grootendorst. “BERTopic: Neural topic modeling with a class-based TF-IDF procedure”. In: *arXiv preprint arXiv:2203.05794* (2022).

[39] Fred Jelinek et al. “Perplexity—a measure of the difficulty of speech recognition tasks”. In: *The Journal of the Acoustical Society of America* 62.S1 (1977), S63–S63.

[40] David Mimno et al. “Optimizing semantic coherence in topic models”. In: *Proceedings of the 2011 conference on empirical methods in natural language processing*. 2011, pp. 262–272.

[41] Adji B Dieng, Francisco JR Ruiz, and David M Blei. “Topic modeling in embedding spaces”. In: *Transactions of the Association for Computational Linguistics* 8 (2020), pp. 439–453.

[42] Kazutoshi Sasahara et al. “Social influence and unfollowing accelerate the emergence of echo chambers”. In: *Journal of Computational Social Science* 4.1 (2021), pp. 381–402.

[43] Xiaolei Song, Siliang Guo, and Yichang Gao. “Personality traits and their influence on Echo chamber formation in social media: a comparative study of Twitter and Weibo”. In: *Frontiers in Psychology* 15 (2024), p. 1323117.

[44] John M Digman. “Higher-order factors of the Big Five.” In: *Journal of personality and social psychology* 73.6 (1997), p. 1246.

[45] Mitja D Back et al. “Facebook profiles reflect actual personality, not self-idealization”. In: *Psychological science* 21.3 (2010), pp. 372–374.

[46] Alessandro Bessi. “Personality traits and echo chambers on facebook”. In: *Computers in Human Behavior* 65 (2016), pp. 319–324.

[47] Kiran Garimella et al. “Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship”. In: *Proceedings of the 2018 world wide web conference*. 2018, pp. 913–922.

[48] Fabian Baumann et al. “Modeling echo chambers and polarization dynamics in social networks”. In: *Physical Review Letters* 124.4 (2020), p. 048301.

[49] Hanif Emamgholizadeh et al. “A Framework for Quantifying Controversy of Social Network Debates Using Attributed Networks: Biased Random Walk (BRW)”. In: *Social Network Analysis and Mining* 10 (Dec. 2020). DOI: 10.1007/s13278-020-00703-1.

[50] Virginia Morini, Laura Pollacci, and Giulio Rossetti. “Toward a standard approach for echo chamber detection: Reddit case study”. In: *Applied Sciences* 11.12 (2021), p. 5390.

[51] Kiran Garimella et al. “Quantifying controversy on social media”. In: *ACM Transactions on Social Computing* 1.1 (2018), pp. 1–27.

[52] Alessandro Cossard et al. “Falling into the echo chamber: the Italian vaccination debate on Twitter”. In: *Proceedings of the International AAAI conference on web and social media*. Vol. 14. 2020, pp. 130–140.

[53] Nagarajan Natarajan, Prithviraj Sen, and Vineet Chaoji. “Community detection in content-sharing social networks”. In: *Proceedings of the 2013 IEEE/ACM international conference on advances in social networks analysis and mining*. 2013, pp. 82–89.

[54] Mauro Coletto et al. “Automatic controversy detection in social media: A content-independent motif-based approach”. In: *Online Social Networks and Media* 3 (2017), pp. 22–31.

[55] Pedro Guerra et al. “A measure of polarization on social media networks based on community boundaries”. In: *Proceedings of the international AAAI conference on web and social media*. Vol. 7. 1. 2013, pp. 215–224.

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

[57] Noah E Friedkin and Eugene C Johnsen. “Social positions in influence networks”. In: *Social networks* 19.3 (1997), pp. 209–222.

[58] Mahmoud Al-Ayyoub et al. “Studying the controversy in online crowds’ interactions”. In: *Applied Soft Computing* 66 (2018), pp. 557–563.

[59] Paola Impicciché and Marco Viviani. “Comparing Echo Chamber Detection Metrics”. In: (2024).

[60] Arunava Chakraborty, Sourav Das, and Anup Kolya. “Sentiment Analysis of Covid-19 Tweets Using Evolutionary Classification-Based LSTM Model”. In: Jan. 2021, pp. 75–86. ISBN: 978-981-16-1542-9. DOI: 10.1007/978-981-16-1543-6_7.

[61] Kartik Agarwal. *The ChatGPT Phenomenon: Unraveling Insights from 500,000 Tweets Using NLP*. 2024. URL: https://medium.com/@ka2612/the-chatgpt-phenomenon-unraveling-insights-from-500-000-tweets-using-nlp-8ec0ad8ffd37.