UNIVERSITY OF MILANO-BICOCCA

MASTER'S THESIS

The Phenomenon of Echo Chambers in Social Media: Quantification and Reduction of Controversy

Author:
Federico COMOTTO

Supervisor:
Dr. Marco VIVIANI
Co-supervisor:
Prof. Gabriella PASI

*A thesis submitted in fulfillment of the requirements
for the degree of MSc in Data Science*

in the

Information and Knowledge Representation, Retrieval and Reasoning LAB
Department of Informatics, Systems and Communications

March 16, 2020

UNIVERSITA' DEGLI STUDI DI MILANO

BICOCCA

Declaration of Authorship

I, Federico COMOTTO, declare that this thesis titled, “The Phenomenon of Echo Chambers in Social Media: Quantification and Reduction of Controversy” and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a Master’s degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

"If your dreams do not scare you, they are not big enough."

Ellen Johnson Sirleaf

UNIVERSITY OF MILANO-BICOCCA

Abstract

Department of Informatics, Systems and Communications

MSc in Data Science

The Phenomenon of Echo Chambers in Social Media: Quantification and Reduction of Controversy

by Federico COMOTTO

The primary role of Social Media is to fulfill perceived social needs such as connecting with friends or creating new relationships. In the Information Age, these platforms have also become essential to society as a *news source*, given the amount of *User-Generated Content* that is disseminated. At the same time, Social Media have been criticized for their tendency to create polarized groups where each member "hear her/his own voice", also known as *echo chambers*. This is due to both technological and sociological properties of Social Media, i.e., *Selective Exposure*, *Homophily*, and *Personalized Recommendations*.

This thesis addresses the problem of *controversy reduction* on Social Media, by connecting opposite echo chambers. To this aim, we employ a pipeline which aims to *identify*, *quantify* and *reduce* the overall level of controversy between two echo chambers. The proposed approach is principally based on techniques framed in the context of *Social Network Analysis*. In particular: (i) we employ a *Graph Partitioning* algorithm to identify latent cohesive structures attributable to echo chambers; (ii) we assess the level of controversy through state-of-the-art graph topology-based *Controversy Measures*; and (iii) we employ *Link Prediction* algorithms to connect opposite communities with different points of view.

Beside the proof that connecting distinct echo chambers reduces the overall controversy level in a social network, another contribution of this thesis is the definition of two measures suitable for an edge-addition process in the context of controversy reduction. We refer to them as *Communicability Measures*, due to their intent to improve information diffusion and, as a consequence, to reduce controversy.

Acknowledgements

First, I would like to acknowledge my thesis supervisor, Dr. Marco Viviani. He always supported me throughout my thesis, giving me the right advice whenever I needed it. I also thank my co-supervisor, Prof. Gabriella Pasi.

I would like to thank all the people that I met during my Erasmus programme in Stockholm. In particular, Edith, Lorenzo, Jeremy and Matteo, I will never forget all the moments shared with you. This experience made me a better person.

Carlos and Eugenio, you deserve special acknowledgements. I was so lucky to meet you in Bicocca. Getting through this Master's with you guys has been one of the best thing that ever happened to me.

Finally, I must express my very profound gratitude to my family and friends for helping me in every situation. Life to me is a matter of moments, and with you, I have spent the best moments of my life. A special mention to Mattia, you deserve the best.

Contents

List of Figures

List of Tables

List of Abbreviations

ABM	**A**gent **B**ased **M**odeling
AI	**A**rtificial **I**ntelligence
CD	**C**ommunity **D**etection
KL	**K**ernighan-**L**in
LPA	**L**abel **P**ropagation **A**lgorithm
NLP	**N**atural **L**anguage **P**rocessing
PDF	**P**robability **D**ensity **F**unction
RS	**R**ecommender **S**ystems
SNA	**S**ocial **N**etwork **A**nalysis

Dedicated to all the doctors and nurses,
in particular to my Mom. You are the light in the darkest moments, you are the only hero.

ANDRÀ TUTTO BENE.

Chapter 1

Introduction

1.1 Context

In recent years, more known as *Information Age* (Floridi, 2014), people are exposed to a significant amount of information, available from both websites and social media. In particular, via social media platforms, such as Facebook[1] and Twitter[2] – which have become part of our daily lives – people have the possibility to be connected and directly share personal contents and opinions with friends and other members of a virtual community. The success of such platforms has contributed to a drastic shift in the size and velocity at which the information is communicated. Social media feeds are essential resources for accessing vast volumes of news, public opinion and information. Therefore, even if we might regret the time when people sat around a table sharing a cup of coffee and reading newspaper, we have to admit that today people have the opportunity to access a huge amount of information like never before, being updated about facts and relevant events.

However, this leads to the so-called *information overload* problem, which means that users often do not know how to discriminate relevant information from what is not. To face this problem, both *Web search engines*, such as Google[3] and Bing,[4] and *Recommender Systems* (RS)s have been developed. A Web search engine is a system that deals with the organization, storage, retrieval of *relevant* information from (textual) document repositories (in the form of Web pages) considering a user *query* and, in the case of *personalized search*, also her/his profile. A Recommender System provides automatic suggestions for items (also in the form of information) that are most likely of use to users (based on their past preferences). In both cases, even if in different ways, these systems *organize*, *filter*, and *prioritize* data according to the user's preferences and/or needs.

Personalized search engines and recommendation strategies are also implemented within social media to help users in retrieving useful information. Furthermore, social media platforms make it easy to connect with and access information from anyone, by emphasizing some of the congenital, social, and psychological traits

[1] https://www.facebook.com/.

[2] https://twitter.com/.

[3] https://www.google.com/.

[4] https://www.bing.com/.

of individuals. Among these traits, *homophily* and *selective exposure* are rather common. The former refers to the principle that a contact between similar people occurs at a higher rate than among dissimilar people (McPherson, Smith-Lovin, and Cook, 2001), the latter to the tendency of people to seek out information that reinforces their idea and to reject information that threatens it (Bessi, 2016).

Hence, as a consequence of filtered information, personalized recommendations, homophily, and selective exposure, users are likely (i) to receive information that mostly confirms their viewpoint and, in worse cases, (ii) to be "trapped" in a chamber with similar individuals. These two phenomena are better known as *filter bubbles* and *echo chambers*.

Filter bubbles are more related with the proliferation of technologies (e.g., personalized search and recommendation strategies) that enables the filtering and ranking of the information we see (Davies, 2018). These technologies hide part of the information to users, showing only what it is estimated "of interests or relevant" to them, based on their preferences and interests; users merely are in a "bubble" where they see only (a part) of all the possible information. Therefore, bubbles may threat democracy and information (Bozdag and Hoven, 2015).

Echo chambers refer to all that situations where people tend to create strong ties with a clique of similar individuals. In particular, to all that situations where people "hear their own voice" (Garimella et al., 2018). In fact, within a chamber, information variety is quite poor because similar people share similar beliefs and opinions. On social media, this tendency may even be emphasized by the above-mentioned technologies (i.e., personalized search and RSs). Echo chambers have long been criticized, especially for their ability to generate polarization (Del Vicario et al., 2016a) and, as a consequence, to increase the *controversy* among communities.

Many solutions have been proposed in the literature so far to counter both these phenomena. Among the techniques to mitigate bubbles, there is, for example, the one based on making users aware about the information consumed (Resnick et al., 2013). However, in this thesis, we focus on the study of the *echo chamber* problem. Although in the literature several studies have concerned the identification of chambers on social media and the measurement of the polarization level within them, few works have focused on the problem of how to reduce controversy among distinct polarized communities in order to mitigate the echo chamber effect. Therefore, this thesis mainly focuses on this problem. In particular, on the employment of a pipeline which aims to *identify*, *quantify* and *reduce* controversy in social media.

1.2 Research Goal

The idea behind this thesis is that, by connecting opposite polarized communities, it is possible to increase information diffusion among them and reduce, in this way, the overall level of controversy and the echo chamber phenomenon.
Therefore, the objectives that this thesis sets are the following:

- The *identification of echo chambers* on social media;
- The *quantification of the controversy level* among echo chambers;
- The *reduction of the controversy* among polarized communities, and, hence, the reduction of the echo chamber phenomenon.

These goals bring us to the principal research question:

What is the most effective way of reducing controversy?

This question leads to other two-sub questions:

1. Can we focus on *graph theory* and *Social Network Analysis* in order to evaluate and reduce controversy?
2. Can we *connect distinct polarized communities* by acting on the *properties* of the social network taken into account?

1.3 Contribution

As previously illustrated, the solution proposed in this thesis involves the implementation of a pipeline to identify, quantify and reduce controversy in social media. This pipeline is based on the combination of different techniques for *community detection*, *controversy evaluation* and *link prediction* that are framed in the context of Social Network Analysis.

Firstly, we preliminarily evaluated distinct social networks related to specific topics by investigating the presence of latent structures within social networks attributable to two cohesive echo chambers.

Secondly, we employed a novel community detection algorithm, called *FluidC*, in order to identify those two communities in the considered social networks. The algorithm produced satisfactory partitions, whose quality has been evaluated by using a proper metric (*coverage*).

Subsequently, we evaluated the level of controversy in the social networks by using three state-of-the-art controversy measures. In particular, the chosen metrics are *Random Walk Controversy*, *Embedding Controversy* and *Boundary Connectivity Controversy*. According to our results, we concluded that the latter measure was the best at capturing the real level of controversy within a graph.

Finally, we connected the opposite communities by employing and testing some link prediction algorithms. Given that the goal of our research is to improve communication among contradictory chambers, we considered only pairs of nodes belonging to different chambers. In other words, we only considered new connections *between* the chambers, rather that *within* the chambers. In order to rank pairs of nodes to be added, we used four well-known similarity measures (*Adamic-Adar Index*, *Jaccard Coefficient*, *Resource Allocation Index* and *Preferential Attachment*). We also implemented two novel measures, called *communicability measures*, based on the graph notions of *betweenness* among nodes and *effective size*. The foster aims to capture the importance of the node in terms of communicability, the latter to calculate the number of non-redundant connections of a node in the network. However, both may capture the influence of a node, within a community, in terms of information diffusion. Moreover, we combined similarity and communicability measures in order to create an *hybrid* version of the latter, which could benefit from the characteristics of both. We concluded the analysis by comparing all the mentioned measures and evaluating which of them generates the highest controversy reduction.

Our results show that, on average, adding edges using both similarity and communicability measures reduces the controversy level in the social networks. In particular, our novel communicability measure, based on effective size, outperforms the others, in case we use Boundary Connectivity as benchmark.

1.4 Thesis Outline

The rest of the thesis is organized as follows:

Chapter 2: We discuss the main concepts related to the problems addressed in this thesis. In particular, we give an overview of the main characteristics of both filter bubbles and echo chambers, and same related concepts, such as those of social media and social networks. We conclude the chapter by presenting some open issues related to filter bubbles and echo chambers.

Chapter 3: We concentrate on studies related to echo chambers, in particular on those that have proposed strategies to address the controversy reduction problem. Consequently, we provide some background theory that is necessary to understand the solution proposed in the thesis.

Chapter 4: We formally explain all the techniques implemented in this work, in order to reduce controversy in social media. In particular, we detail community detection algorithms, controversy measures, and link prediction algorithms adopted throughout our study.

Chapter 5: We report the results of the solution proposed in this thesis, and their discussion.

Chapter 6: We summarize the work that has been done in this thesis and we answer to the two questions stated at the beginning of this chapter.

Chapter 2

The Filter Bubble and the Echo Chamber Phenomena

The main problem that this thesis aims to address is that of the excessive polarization of opinions in social media, which leads to difficult communication and "contagion" of ideas between user groups with different interests, social backgrounds, and personal conditions within virtual communities. As briefly explained in Chapter 1, there are various psychological and technological motivations that lead to this situation. One reason is the formation of the so-called *echo chambers*, on which this thesis focuses. Despite this, in this section, we want to look at the whole picture. In fact, most of the times, the echo chamber phenomenon is related to a complementary one, the so-called *filter bubble* phenomenon. A lot of confusion has been done around these concepts, and it is rather difficult to describe echo chambers without having understood what filter bubbles are.

What is common between echo chambers and filter bubbles is their ability to get users polarized towards one specific idea or opinion about a topic. The terms filter bubble and echo chamber have been coined to refer to two different algorithmic pathways to opinion fragmentation, both related to the way algorithms filter and rank information. The first refers (mainly) to search engines and recommender systems, the second is more applied to social media feeds (Sasahara et al., 2019). The fact that users may get polarized frequently happens on social media platforms where, by attitude, people tend to keep in contact to those individuals who are similar to them. That is why we choose to study echo chambers on social media. However, as we are going to see in the next chapters, the propensity to get polarized also depends on topic nature: for example politics is one of the most controversial topics and it is quite simple to get polarized towards one side.

In this chapter, we are going to present, from a theoretical point of view, both filter bubbles and echo chambers. Furthermore, since some practical approaches have been proposed over the last years to identify, analyze and alleviate these phenomena, we will briefly introduce some *state-of-the-art solutions* and *methodologies* proposed so far to tackle both of them. In particular, in Section 2.1, we explain what a filter bubble is and how it originates, and, in particular, why filter bubbles could be potentially dangerous to individuals and society, and what has been done since today in order to mitigate their formation. Indeed, in Section 2.2, we give an overview on the echo chamber phenomenon. This section, in particular, has to be viewed as a preliminary and necessary introduction to the next chapters, where we are going to deeply analyze how to discover *echo chambers* and some possible solutions about how to evaluate and reduce *controversy* between different polarized groups of users within a virtual community.

2.1 Filter Bubbles: Main Concepts and State of the Art

In (Pariser, 2011), Eli Pariser treats the phenomenon of *filter bubbles* and explains what it really means. According to Pariser, users might get different search results for the same keyword and those with same friend lists can receive different updates. This is because information can be prioritized, filtered and hidden depending on a user's previous interaction with the system and other factors (Bozdag and Hoven, 2015). Although Pariser has coined the term filter bubble, which is defined by the author as: "the intellectual isolation that can occur when websites make use of algorithms to selectively assume the information a user would want to see, and then give information to the user according to this assumption", many other researchers have tried to study this phenomenon. Several approaches may be followed to understand and analyse such a topic.

2.1.1 Are Filter Bubbles just a matter of Technology?

Some of the works proposed in the literature, try to describe the filter bubble phenomenon from different perspectives. Despite Pariser thought, Davies (2018) shows that distortion in information can also be caused by the *social status* of individuals. By drawing a case study on young people coming from two sharply contrasting institutions (different cultural and social backgrounds), the author argues that people are actors who act in an environment based on culture, social class and technology. This means that people's experiences and understanding of society and technology are reflected in their digital practices (e.g., in their use of search engines), and that this has consequences for the content of their information ecospheres (Davies, 2018).

In particular, the Davies' paper shows how cultural and social backgrounds can affect filter bubbles: for example, the knowledge of other search engines available that are designed to prevent filter bubbling, the knowledge of Google's prioritising algorithms, informed scepticism about search results, and the knowledge and analytical skills to assess the authenticity of information are forms of technical and cultural capital that can be accumulated as a consequence of being a member of the professional class. This could mean that, although search engines have an effect on our life by filtering content, people might also avoid that behaviour by, for example, comparing results between different search engines (e.g., by employing meta-search engines), and/or investigating the veracity of the retrieved results.

Geschke, Lorenz, and Holtz (2019) build up a model in order to understand filter bubbles, but also their relation with echo chambers. According to the authors, filter bubbles are defined as "an individual outcome of different processes of information search, perception, selection, and remembering the sum of which causes individual users to receive from the universe of available information only a tailored selection that fits their pre-existing attitudes". Moreover, individuals tend to share a common social media bubble with like-minded friends; over time, such communities in which web content that confirms certain ideologies is echoed from all sides are particularly prone to processes of group radicalization and polarization (Vinokur and Burnstein, 1978); this phenomenon is also known as the echo chamber effect, which will be detailed in Section 2.2. Thus, according to the authors the filter bubble and echo chamber phenomena are strictly related, since echo chambers are "a social phenomenon where the filter bubbles of interacting individuals strongly overlap".

In their paper, the authors refer to filters in a very general way as a process that lead to a limitation of information that is available to individuals. In particular, the authors take into account filtering processes on three different levels: (i) *individual*, (ii) *social*, and (iii) *technological*.

The *individual* level is totally related to the psychological and cognitive process of human beings. In order to boost their social identity and confirm pre-existing attitudes, individuals are more likely to consume (*filter*) information that fit their beliefs and values and avoid conflicting point of views. Curiosity may, however, motivate individuals to have a preference for consuming information that is at least to some degree novel and surprising.

The *social* level is, instead, related to the attitude and tendency of human beings to build network structures based on common interests and characteristics. This trend seems to be particularly emphasized in recent years thanks to social media platforms. In the age of social media, information is often passed along such online networks (Bakshy et al., 2012), this means that information can be manipulated and filtered by this social layer especially in case of homogeneous networks. In fact, on social media, one common phenomenon is the so-called *homophily* (McPherson, Smith-Lovin, and Cook, 2001), which is the the principle that a contact between similar people occurs at a higher rate than among dissimilar people.

Finally, the *technological* level is exactly what Pariser treats in his book (Pariser, 2011). In fact, online media providers, such as Google or Facebook, compete for user attention. Therefore, they filter the provided information according to individual users' assumed wants and needs, leading to individually selected media offers. Basically, online filters, also known as *recommender systems*, show tailored result to users, hiding part of the content.

In order to carry out their goal, i.e., finding which combination of these three filters might facilitate or not the emergence of filter bubbles (and chambers), the authors use an *Agent-Based Modeling* (ABM). This kind of model is suited for studying interactions between individuals rather than interactions between variables and how such interpersonal influence processes play out. This approach allows the researchers to distinguish the individual effects from others, such as possible effect of the aforementioned filters. The authors, in particular, design a dynamic ABM where several individuals (together representing a society) position themselves in a two-dimensional attitude space based on attitudinal bits of information they hold in memory. The model rules define that an individual repeatedly receives new piece of information from different sources.[1] Individuals can also receive information through the network (social filter). The received information is finally *filtered* by cognitive process, i.e., individuals are more likely to hold information that fits their values.

Their simulations show that, even without any social (posting or refriending) or technological (recommender systems) processes involved, filter bubbles (and echo chambers) could emerge through the individual cognitive processes. If, however, additional social or technological processes occurred, these phenomena become more distinct; this would lead to even more fragmentation and polarization of society.

2.1.2 Surprising consequences

Although it is not trivial to establish whether or not filters bubbles are just a matter of technology, it is surely simpler to understand possible risks that we could eventually bear as online consumers. However, since most of the people do not know the existence of filter bubbles, they do not probably know their effects.

Bozdag and Hoven (2015) treat this problem from a more theoretical point of view. According to the authors, any kind of democracy might be jeopardized by filter bubbles. Democracy refers very roughly to a method of group decision making characterized by equality among the participants at an essential stage of the collective decision making.

Different *democracy theories* exist: they can be designed in different ways, each of them with their own rules and information requirements. Bozdag and Hoven (2015) identify four democracy theories: (a) *liberal*, (b) *deliberative*, (c) *republican and contestatory*, (d) *agonistic/inclusive political communication*. The authors accurately classify those theories and bring up possible downsides caused by filter bubbles. Table 2.1 shows the results of that analysis. It is easily understandable that all democracy forms have vulnerabilities that filter bubbles might exploit. Filter bubbles should be seen as worrying developments in the digital world from the point of view of democracy, different conceptions and models of democracy point to different undesired consequences of such bubbles, ranging from loss of autonomy to the diminishing epistemic quality of information (Bozdag and Hoven, 2015).

Another point of view on the consequences of filter bubbles is given by Nechushtai and Lewis (2019). In this case, the authors wonder if algorithms (i.e. *recommender*

[1] The sources represent the technological filters.

TABLE 2.1: Democracy models: norms and criticism of the filter bubble. From: Bozdag and Hoven (2015)

Model of democracy	**Norms**	**Criticism of the filter bubble**
Liberal	Awareness of available preferences Self-determination Autonomy Adaptive preferences Free media Respect human dignity	User is unaware of the availability of options User is restrained and individual liberty is curtailed The media is not free, it serves the interests of certain parties (e.g. advertisers) Powers are not separated (advertiser and the information provider are the same)
Deliberative	Discover facts, perspectives and disagreements Determine common interests Construct identity by self-discovery Refine arguments and provide better epistemic justifications Consensus Respect towards each other's opinions A collective spirit Free and equal participants Rationality	Epistemic quality of information suffers Civic discourse is undermined No need to have better epistemic justifications Respect for other opinions is decreased Legitimacy is more difficult to achieve. There is a loss of a sense of an informational commons Communication suffers as gaining mutual understanding and sense-making is undermined
Republican and contestatory	Freedom from domination by oppressors Contest matters effectively Be aware of the oppressor	Diminishes one's ability to contest Diminishes one's awareness of the oppressors and their potentially manipulative interventions
Agonistic/inclusive political communication	Conflict rather than consensus Passions rather than rationality Struggle rather than agreement Inclusion: Measures must be taken to explicitly include the representation of social groups, relatively small minorities, or socially or economically disadvantaged ones Measures must be taken so that antagonism is transformed into agonism	The adversary becomes the enemy The minorities are excluded from the democratic process, their voices are lost

systems) are good enough to be gatekeepers. A *gatekeeper* is a person who controls access to something. Generally speaking, gate-keeping, in the information domain, was conceptualized as a distinctly human activity, one focused on the process of selecting, writing, editing, positioning, scheduling, repeating, and otherwise massaging information to become news.

Usually, the role of gatekeepers is played by journalists.[2] Since with great power comes great responsibility, journalists had also the role of generating *cultural meaning* by conscientiously selecting piece of information to spread to individuals. The rise of the digital, mobile, social media and the emergence of online news during the past 25 years have not, on their own, eliminated traditional modes of news gatekeeping — editors still decide what to publish and where — but they have disrupted longstanding arrangements of prioritization. They have altered control over news hierarchy, resulting in new relationships between people and machines that shape how news is seen, circulated, and interpreted (Nechushtai and Lewis, 2019).

For example, a simple user on a social media platform can be a second-gatekeeper by sharing/re-posting information. At the same time, algorithms (machines) are defined gatekeepers due to their ability to prioritize and filter information according to users characteristics: they act like a human being who decides what is relevant or not for our cultural status. A primary concern about algorithms as news gatekeeper is their capacity to tailor information encounters, typically described as leading to filter bubbles or echo chambers; this means: can new gatekeepers create polarized group of individuals that live in a so-called bubble? By drawing an experimental on Google News, Nechushtai and Lewis (2019) try to answer that question.

Their findings indicate that, despite the ability of algorithms to provide much more personalized headlines than human editors, they might actually produce, at least in some cases, highly centralized and unifies news diets across diverse sets of users. This supports the filter bubbles theory.

2.1.3 When emotions make the difference

Abisheva, Garcia, and Schweitzer (2016) show what kind of role emotions might play in the generation of filter bubbles. A number of studies in social psychology, in particular those of Kühne (2014) and Gorn, Tuan Pham, and Yatming Sin (2001), show how emotions influence individual evaluations, judgements, and opinions, based on the theory of core affect (Russell and Barrett, 1999). According to this theory, emotions are characterized by two dimensions: *valence* which defines the feeling of pleasure or displeasure and *arousal* which encompasses a feeling of activation or deactivation, and quantifies mobilization and energy. Empirical evidence shows that the subjective experience of arousal motivates evaluation on the extremes.

[2] More in general we can say *editors* instead of journalists. The main point is that it was a pure human activity

Polarization might be higher when arousal is experienced in addiction to negative valence. This combination can lead to more polarized effect.

In their work, Abisheva, Garcia, and Schweitzer use real data[3] in order to asses some important hypotheses, in particular whether emotions may create polarization and whether this polarization can be increased by filtering mechanisms. In particular, the authors use several statistical methods, ranging from sentiment analysis to regression.

The results illustrated in the paper are consistent with the hypotheses that the expression of activating and negative feeling, such as anger or outrage, tend to create more polarized responses, in line with the theoretical argument that poses emotions as mechanisms to speed up evaluation processes at the expense of more extreme reactions.

2.1.4 Techniques to reduce filtering

This section illustrates a couple of solutions that have been put in place to alleviate the problem of over-filtering the information offered to users in a virtual community. Since the aim of this thesis is not to focus on filter bubbles, we describe only the works that have provided the main ideas to tackle the above-mentioned issue.

Promote diverse exposure

As seen before, filter bubbles might be seen as a combination of psychological, sociological and technological aspects. Media policy in the United States, however, has long focused on encouraging the access of citizens to diverse information. One rationale is that development of accurate beliefs requires some degree of exposure to information that challenges one's existing beliefs and opinions (Resnick et al., 2013).

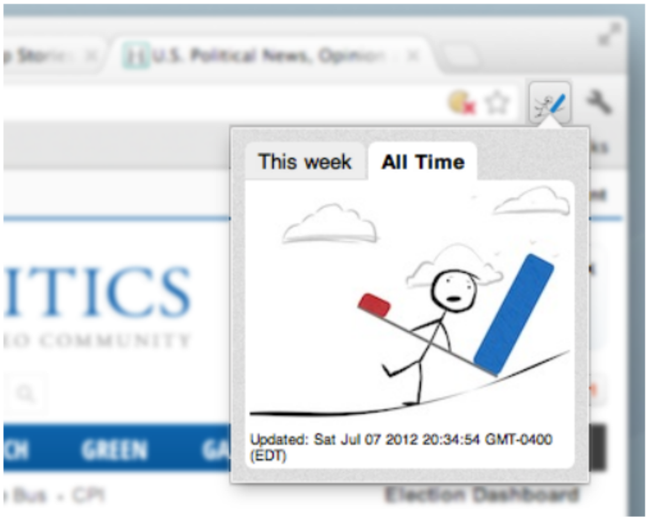
Generally speaking, individuals, particularly those in the minority, tend to think that their own ideas are more broadly shared than they actually are. Filters can enhance this phenomenon because they could potentially bring to bubbles where people are surrounded by like-minded individuals. Despite this, there are evidence that people prefer collections of news articles which include some counter-attitudinal articles over collection of purely agreeable items.

In general, also the most narrowed person sometimes could accept new conflicting items. This means that it may be possible to exploit this need and develop systems that nudge people toward more diverse exposure or encourage individuals to choose diverse exposure (Resnick et al., 2013).

Figure 2.1 shows a web extension used to make individuals aware about their choice in terms of right or left news sources. Initially analysis of experimental data suggest a small, but noticeable change in reading behavior, toward more balanced exposure, among users seeing the feedback, as compared to a control group.

[3] They crawl data from publicly accessible online communities such as *Youtube*, *Reddit* and *Imgur*.

In addiction to selecting information in a motivated way, people can also process information in a motivated ways. In this case, one possible solution is to make opposing individuals meet each other. In fact, interaction may lead people to be more open to different perspectives. That idea has encouraged the creation of pooling tools used in online newsrooms to inspire mindful engagement with alternative perspectives (Resnick et al., 2013).



Updated: Sat Jul 07 2012 20:34:54 GMT-0400 (EDT)

FIGURE 2.1: Balancer classifies pages based on their address: does this source tend to get linked to by mostly liberals or conservatives? Is it regularly visited by liberals, conservatives, or a mix?

Take into account Serendipity

Sometimes, in order to mitigate the filter's effect, it is necessary to act directly on the algorithm. Basically, this means to modify how the algorithm works.

In many cases, users' information is filtered by the so-called *Recommender Systems* - RS. A recommender system is a subclass of information filtering system that seeks to predict the "preference" a user would give to an item. RS aim to predict new items based upon user's information and user's previous interaction. Therefore, results are tailored and filtered according to users' characteristics.

The problem of training on the past without necessarily repeating it is an open problem in many collaborative filtering based recommendation contexts, particularly social networks, where, in the degenerate cases, users can get caught in filter bubbles, or model-based user stereotypes, leading to a narrowing of item recommendation variety (Pardos and Jiang, 2019).

In the above-mentioned paper, the authors has built a recommender system which aim to take into account *serendipity*. Serendipity is defined ad the ability to make fortunate discoveries by accident: you find something valuable without looking for it. In the mentioned paper, the authors introduce a "variant" into a production recommender system at a public university. This variant has been designed to surface serendipitous course suggestions. The authors refer to serendipity "as user perceived unexpectedness of result combined with successfulness".

The comparison of the original model with its serendipity variant leads to one important result: students of the public university have positively evaluated the variety of the recommendations. In fact, on average, the students found improvements in the recommended courses.

Beside the previous work, in (Maccatrozzo, 2012), the author has adopted the notion of serendipity as a performance measure for recommendation algorithms. Moreover, the author has proposed a user model that can facilitate and enables serendipitous recommendations by using semantic web techniques, in particular *Linked Open Data*.[4]

Therefore, despite it is important to recommend pertaining items to individuals, at the same time, recommender systems could take into account serendipity metrics in order to promote diverse contents. This could be another important way of getting out of the bubble.

2.2 The Echo Chamber Phenomenon on Social Media

As illustrated at the beginning of this chapter, the term *echo chamber* is employed, in particular, referred to the social media context. In this section, we are going to explain why. In fact, as stated in (Garimella et al., 2018), the echo chamber phenomenon refers to situations where people *hear their own voice*. This is a metaphor that emphasizes how the information circulates within a *closed system*. A typical environment where this kind of phenomenon has highly chance to grow up is the one of *social media* platforms. In fact, this kind of environment has the ability to connect similar people that tend to create closed networks in which the shared beliefs and opinions are facilitated to *echo*.

In the next sections, we describe the concept of *social media* and in particular of *online social network*, illustrating the main characteristics of online virtual communities, which are *complex networks*, which often leads to the generation of echo chambers. In addition to this, we are going to introduce those works discussing the *echo chamber* phenomenon from a general point of view, and illustrating the main characteristics of echo chambers. Specific solutions to identify echo chambers and to reduce controversy among the members of a social network will be illustrated in Section 3.1 of Chapter 3. Here, we are focusing more on general aspects related to the phenomenon, which have to be considered as an introduction and preliminary to the full understanding of the concepts described in Chapters 3, 4 and 5.

[4] In computing, linked data is structured data which is interlinked with other data so it becomes more useful through semantic queries.

2.2.1 Social Media, Social Networks, and their Characteristics

In this paragraph, we introduce the concepts of *social media* and (online) *social networks*. We have already pointed out the fact that the echo chambers formation deals with social media platforms: in these platforms, the combination of social dynamics and technologies could often lead to polarized communities.

Therefore, an introduction to the main concepts behind those systems, which may be the environment for the proliferation of echo chambers, has to be made. Social media can be defined as interactive computer-mediated technologies, such as websites and computer programs, that allow people to communicate and share information on the internet using a computer or mobile phone. Among the most popular social media, there are *Facebook*,[5] *Twitter*,[6] *YouTube*,[7] and *Instagram*.[8] Despite there exist different type of social media, they rely on the common idea that people are inclined to share contents and ideas with their friends and with other persons based on common interests. Like in real life, online, people make friends and create new relationships.

So, through the use of social media, it is possible, in the online scenario, to generate *social networks*. Therefore, an online social network is a virtual community in which its members are connected through different kinds of social relationships. Usually, a social network is represented in the form of a graph. In such a graph, the vertices can represent persons or other entities and the connections between them, in the form of edges, represent some form of social interaction, such as friendship. Hence, social media are the technologies through which online social networks can be generated. In this scenario, *Social Network Analysis* (SNA) is the process of investigating social structures through the use of networks and graph theory. It characterizes networked structures in terms of nodes (individual actors, people, or things within the network) and the ties, edges, or links (relationships or interactions) that connect them.

Social networks belong to the family of *complex networks*. Unlike regular networks, in which every node is connected to a fixed number of nodes, or random networks, in which connections among nodes are based on a certain probability p, complex networks are more sophisticated structures. Among the characteristics of complex networks, we find:

- the so-called *small world* property (also knon as the *six degrees of separation* property);
- the presence of both *strong* and *weak ties*;
- the *homophily* property;
- the *scale-free* property.

[5] https://www.facebook.com/.
[6] https://twitter.com/.
[7] https://www.youtube.com/.
[8] https://www.instagram.com/.

The *small world* property has been studied in multiple experiments performed by Stanley Milgram (Milgram, 1967) and other researchers investigating the average path length for social networks of people in the United States. These studies were revolutionary in the sense that they proposed that human society is a *small-world-type network* characterized by short path-lengths. The small-world experiments are often associated with the phrase "six degrees of separation". It is the idea that all people are six, or fewer, social connections away from each other. As a result, a chain of "a friend of a friend" statements can be made to connect any two people in a maximum of six steps.

Another important characteristic of complex network was studied by Granovetter (1977). In his studies, he observed that in social networks we have both *strong* and *weak ties*. In his paper, Granovetter refers to strong ties as the closest friends and weak ties as acquaintances. It is claimed that weak ties are responsible for the majority of the embeddedness and structure of social networks in society as well as the transmission of information through these networks. Weak ties are bridges to novel and fresh nodes which could lead to more opportunity than strong ties. Because our close friends tend to move in the same circles that we do, the information they receive overlaps considerably with what we already know. Conversely, acquaintances know people that we do not, therefore they are the vehicles towards new, unexplored connections.

Another fundamental characteristic of social networks is *homophily*. We briefly introduced this concept in Section 2.1. Homophily is the tendency of individuals to associate and bond with similar others. Individuals in homophilic relationships share common characteristics (beliefs, values, education, etc.) that promote contacts and the development of relationships. Homophily, by itself, may cause the formation of echo chambers because people, in this situation, tend to share and consume items that adhere to their system of beliefs.

Finally, the last important characteristic of complex network, and, hence, of social networks, is their propensity to be *scale-free network*. A scale-free network is a network whose degree distribution follows a power law, at least asymptotically. The most significant aspect in a scale-free network is the relative commonness of vertices with a degree that greatly exceeds the average. The highest-degree nodes are often called *hubs*. The main assumption is that, when a node has to establish a new connection, it prefers to do this with a node that already has many connections, i.e., a hub. Another important characteristics of hubs is their function of connecting parts of the graph that would otherwise be separate.

2.2.2 Echo Chambers and their main Characteristics

In this section, we contextualize the concept of echo chambers and their characteristics from a general point of view. Section 3.1, in Chapter 3, will introduce those works related to the identification of echo chambers and the quantification and reduction of controversy in order to tackle the echo chamber issue.

Sasahara et al. (2019) have studied the conditions in which echo chambers emerge. In order to analyse the emergence of echo chambers, they rely on a simple model based on two ingredients: *influence* and *unfriending*. The model dynamics indicate that the social network is rapidly evolving into isolated, homogeneous groups, even with small amount of influence and unfriending. Their findings indicate that the prevalence of online echo chambers can inevitably result from basic cognitive and social processes enabled by social media: the human propensity to be conditioned by the information and opinions to which one is exposed, and the avoidance of undesirable social relationships. Moreover, their approach shows how polarization and segregation arise without pretending that opinions are already polarized and that social media debates tend to polarize in exactly two opinion groups. They also find empirical evidence that in many cases (not always), the presence of users with many followers (*hub* node) affect the dissemination of the same messages. Furthermore, their study suggest that triadic closure connects individuals to friends of their friends, facilitating repeated exposure to the same opinion. Such *echoes* are a powerful reinforcing mechanism for the adoption of beliefs and behaviors.

Quattrociocchi, Scala, and Sunstein (2016) have studied the phenomenon of echo chambers on Facebook. Their goal was to demonstrate the presence of echo chambers in this particular social media. Actions such as "share," "comment," or "like" have distinctive meanings on Facebook. In most circumstances, a "share" communicates the desire to increase the exposure of a given piece of information; a "like" reflects a positive feedback to the post; while a "comment" suggests a contribution to an online debate that may include negative or positive feedback. Their findings show that users are highly polarized on Facebook and tend to focus their attention exclusively on information which adheres to their system of beliefs. They also find that users belonging to different communities prefer not interacting and tend to be connected only with like-minded people. This confirms the *homophily* theory previously described. Finally, the authors find evidence that users have a tendency to seek out information that strengthens their idea and to reject information that undermines it. This is also known as *confirmation bias* or *selective exposure*. Both are psychological factors that refer to the individuals' tendency to favor information which reinforces their pre-existing views while avoiding contradictory information.

Bessi (2016) has conducted an important study which aims to display common psychological characteristics among different echo chambers. The paper shows that different and contradictory communities are dominated by users showing similar psychological profiles, and that the dominant personality model is the same in different echo chambers. Psychologists describe personality using five dimensions known as the *Big Five*: *extraversion*, *emotional stability*, *agreeableness*, *conscientiousness*, and *openness*. Such five dimensions contain most known personality traits and represent the basic structure behind all personalities (Bessi, 2016). In order to assign a personality trait to each user, they use an unsupervised personality recognition approach which leverages a series of statistically significant correlations between linguistic features and personality traits. Their results show similarly distribution of personality traits within the polarized communities. Furthermore, within distinct echo chambers, they find very strong and relevant similarities between personality traits. Bessi shows that the prevalent personality model is the same in both the observed echo chambers. This prevalent personality model corresponds to a prototype-user who tends to enjoy interactions with close friends (low extraversion), is emotionally stable (high emotional stability), is suspicious and antagonistic towards others (low agreeableness), engages in antisocial behavior (low conscientiousness), and has unconventional interests (high openness).

Del Vicario et al. (2016a) have focused on the evolution of echo chambers within Facebook. They have proved that, despite different in contents, two different echo chambers evolve in a similar way and the behavior of their users is similar. The sizes of both communities have been evaluated in terms of their temporal evolution and fitted with models of classical population growth deriving from fields of biology and medicine. The behavior of users turns out to be similar for both categories, irrespective of the contents: both the communities reach a thresholding value in their sizes, after an almost exponential growth, in agreement with classical growth models.

They have also noticed that the emotional behavior of communities is affected by the involvement of users inside the echo chamber. To a higher involvement corresponds a more negative approach. In addition, they have observed that, on average, more active users show a faster shift towards the negativity than less active ones. Therefore, it seems that polarization within echo chambers depends on users' involvement and activity.

Baumann et al. (2019) have studied polarization and echo chambers by using a simple model of opinion dynamics. The model is based on three features: (i) *social influence*, (ii) *heterogeneity* activity of the users, and (iii) *homophily* in the interactions.

The model has been tested on real-data. In accordance to the results proposed by Del Vicario et al. (2016a), the model shows one frequently observed phenomena of polarized social networks: more active users, i.e., those more prone to engage in social interactions, tend to show more extreme opinions. Moreover, the model shows there is similarity between the opinion expressed by a user and those expressed by his/her neighbors in the social interaction network.

2.3 Issues Related to Filter Bubbles and Echo Chambers

Sometimes, the filter bubble and the echo chamber phenomena may lead to others correlated problems such as the diffusion of *misinformation*, for example *fake news* (Viviani and Pasi, 2017). Generally speaking, we can say that fake news are deliberately generated information proved to not be true. Social media platforms, such as Facebook or Twitter, have been hardly criticized for the proliferation of fake news through their systems. In fact, by re-posting information users may be directly or indirectly involved in the diffusion of fake news. In general, it should be best practice to check the source of the information and verify the content instead of *simply* reading the headline. However fake news, an more general misinformation, have been ranked by the World Economic Forum as one of the biggest problem in the recent years (Del Vicario et al., 2016b). Zimmer et al. (2019) investigates on the relations filter bubbles - fake news and echo chambers - fake news. In particular, their analysis focus on the following research question: *are echo chambers and filter bubbles of fake news man-made or produced by algorithms?* For the purpose of the study, the authors have chosen to carry out the analysis on Facebook.

Facebook has a pertinent recommender system [9] and it ranks post in descending order according to three factors: *affinity*, *weighting* and *timeliness*. Affinity is concerned with the user's previous interactions on the posting pages, whereas different interactions are weighted variously (Zimmer et al., 2019). Moreover, fresh posts are preferred to old ones. Thus, in a short time — with high activity on Facebook - an information diet may occur that presents users only those posts on top of their pages, whose creators they prefer. So it can be assumed that such personalized content representation leads to "partial information blindness", i.e., filter bubbles or echo chambers (Zimmer et al., 2019).

However, as Zimmer et al. state, it depends on the user to form "friendship" on Facebook and it is on the user to often select a subset of friend's posts for reading, liking, sharing and/or commenting. Facebook's pertinence ranking algorithm indeed may amplify existing behavioral patters of the users into filter bubbles and then into echo chambers, whereby the information behavior of the users plays the important primary role (Zimmer et al., 2019).

That means: there is a sort of interaction process between users and algorithms, especially on social media platforms. This interaction is often triggered by the user, but algorithms may cause wider effects. This is also true within the fake news context: it is not possible to argue that they are solely distributed by "bad algorithms," but by the active collaboration of the individual users (Zimmer et al., 2019). Therefore it is on the individuals themselves to accept or deny fake news, to verify or falsify them: a critical user may make the difference also in the era of filtered information.

2.4 Conclusions

In this chapter, we gave an overview on the concepts of *filter bubble* and *echo chamber*. We have introduced several studies that have been properly selected among those proposed in the literature.

[9] Pertinence ranking presupposes that the information system in question is able to identify the concrete user who works with the system; it is always subject dependent personalized ranking (Zimmer et al., 2019).

Filter bubble is a recent concept which has been coined, for the first time, by Pariser (2011). It refers to "the intellectual isolation that can occur when online algorithms filter and select information according to user's preferences and her/his previous interactions". Instead, echo chamber refers more in general to situation where people "hear their own voice". According to Pariser (2011), filter bubble is bounded to technological aspects, in particular to filtering processes. However, throughout Section 2.1 of this chapter, we have seen that also psychological and social processes may take part in this phenomenon.

On the other side, the phenomenon of echo chamber seems to be more related to psychological and social mechanisms, but it could be (highly) emphasized by algorithms. Therefore, the combination of *selective exposure* and *social homophily* could be the primary driver to the formation of segregated and polarized communities. The first refers to individual tendency to seek out information that strengthens one's idea and to reject information that undermines it, the second to the human tendency to be connected only with like-minded people.

We have also seen that, although there may exist different echo chambers, behaviours and personality traits of users within those chambers seem to be the same. Moreover, different echo chambers seem to evolve in the same way.

In the next three chapters we will only focus on the echo chamber phenomenon, and on the possible solutions to tackle this issue. In particular, we will illustrate a pipeline which aims to *quantify* and *reduce* controversy on social media.

Chapter 3

Quantifying and Reducing Controversy: State of the Art

In Chapter 2, we have introduced the general concepts of *filter bubble* and *echo chamber*, illustrating the way in which they are correlated and some general aspects connected to their formation. In particular, we have provided an in-depth analysis of the characteristics and issues of filter bubbles, while for the echo chamber phenomenon, we have introduced the general concept and provided a brief overview of the main characteristics connected to it in the social media context.

This choice has been undertaken because this work primarily concerns the study and analysis of echo chambers, and the reduction of the controversy among them. For this reason, the task of providing more detailed information about this phenomenon has been left to this chapter. In particular, here we are going to provide more technical details about echo chambers and the solutions that have been proposed to reduce controversy among them, by illustrating the related literature.

In general, the chapter is organized as follows. First, we are going to introduce, those works related to identification of echo chambers and quantification and reduction of controversy among them. Then, we are going to provide a more in-depth analysis about the pipeline and techniques that are currently used to tackle this problem. In fact, in the majority of the cases, the problem of how reduce controversy among echo chambers could be decomposed in the following phases:

1. *Finding communities*, i.e., identifying groups of users in the social network that can be associated to different echo chambers;

2. *Measuring controversy*, i.e., identifying suitable metrics able to assess the level of controversy among the identified communities;

3. *Reducing controversy*, i.e., finding suitable strategies in order to reduce controversy. This implies the re-evaluation of the controversy measures after the application of the technique adopted to reduce the controversy.

Due to the fact that one of the biggest problems of being stuck in a echo chamber is the high possibility to always get the same kind of information, our main purpose is to build a framework that aims at facilitating *information diffusion* and, in this way, reducing the controversy within the networks.

3.1 Related Work

The study and analysis of *controversy* among *echo chambers* has been explored by several authors and researchers in the literature. This topic has been receiving a lot of attention due to the fact that people may get polarized and subsequently influenced: knowing the structure and the characteristics of an echo chamber may represent an advantage in some cases. For example a political party could exploit a community[1] to spread (mis)information.

In Section 2.2 of Chapter 2, we illustrated that online social networks are the ideal environment where echo chambers are likely to appear, due to their main characteristics. In particular, even without the presence of filtering algorithms, people (i) tend to socialize with similar users, due to the *homophily* property, and (ii) consume items that confirm or strengthen one's prior personal beliefs or hypotheses, according to the *confirmation bias* phenomenon.

In this section, we are going to introduce and describe those works that are the most related to the problem addressed in this thesis. In the majority of the cases, the problem of controversy quantification and reduction begins with the identification of echo chambers on social media. In the literature, many different approaches have been proposed for identifying communities. In general, the *community detection* problem may be addressed in different ways. As it will be illustrated in this section, some of the works rely on the polarization of *social content* shared by users, others rely on the the analysis of the *topological structure* of social networks. Here, we describe those papers that we have carefully taken into account throughout our analysis.

3.1.1 Echo Chambers: Users Role, Polarization and Network Measures

Most of the time *echo chambers* are related to a specific topic and in many cases, the topic is the political one. This is not surprising because people, in the politics context, are more inclined to adhere to a specific political side (e.g, left or right). Hence, it is much more easy getting connected to people who share the same thought. Garimella et al. (2018) study the degree to which echo chambers exist in political discourse on Twitter, and how they are structured. They approach the study in terms of two components: the *opinion* that is shared by a user, and the the *chamber*, i.e., the social structure around the user, which allows the opinion to *echo* back to the user as it is also shared by others (Garimella et al., 2018). They say that an echo chamber exists if the political side of the content that users receive from the network agrees with that of the content they share. In order to study the opinion and the network, they define measures able to capture content produced by a user and the network position of a user, including their interactions with others (Garimella et al., 2018).

First, they measure the production and consumption polarity of users' content by analyzing their tweets. They look at tweets directly connected to external sources

[1] An echo chamber is nothing more than a community with specific features.

which their polarization have been previously calculated; this strategy simplifies the way to define polarization of users. Moreover, they associate the extremes of polarization score to two different political sides. This process allows to identify three different type of users:

1. *Partisan*, user who produce content of one political side;
2. *Bipartisan*, user who produce content of both political sides;
3. *Gatekeeper*, user who consume content of both sides, but produce content of just one political side.

The analysis of polarization shows the existence of political echo chambers because the distribution of production and consumption polarities of users were (i) bi-modal and (ii) highly correlated.

Secondly, they define measures that capture the position of users in a social network and their interaction with other users. Therefore, they consider measures such as *PageRank*, which gives an idea about importance of node in a network and the *clustering coefficient*, which defines how much connected is a node with its neighborhood. All the analysis have been done using a set of data coming from Twitter. In particular, the data sets contains tweets from both political and non-political topics. In this way, it has been possible proving that echo chambers are more likely to appear in political context than in others. The authors also highlight a worrying aspect of *echo chambers*, the so-called *price of bipartisanship*. Overall bipartisan users pay a price in terms of network centrality, community connection, and endorsements from other users (Garimella et al., 2018). It seems these users are less involved in the network. This suggests the existence of latent phenomena that effectively stifle mediation between the two sides. Finally they tried to examine the role of gatekeepers in the context of echo chambers.

Previous studies on Twitter showed that gatekeepers are typically ordinary citizens rather than officially active partisans (e.g., party members) (Garimella et al., 2018). Based on their findings, it is not clear if this kind of users are simple open-minded citizens or individuals who aim at getting informed and attacking the opponent opinions. However, gatekeepers would be good candidates users to nudge information from one side towards the opposing one. Therefore they used classification method in order to identify whether a user is a *partisan*, a *bipartisan* or a *gatekeeper*. Unfortunately, the model seems to work pretty well for the first two type of users, while "fails" to identify the last one.

Beside the work of Garimella et al. (2018), another important analysis on polarization measures and echo chambers has been done by Duseja and Jhamtani (2019). They carry out a comparative analysis of tweets from users in echo chambers (EC) versus tweets from users not in echo chambers (NE). Specifically, they compare some properties pertaining to tweet structure, lexical choices, and topics/attributes discussed in tweets; Table 3.1 lists all the tested hypotheses. The data set used to carry

TABLE 3.1: Hypotheses and results tested by Duseja and Jhamtani (2019).

Type	Hypothesis	Holds in data?
Tweet-structure	NE tweets are more likely than EC to cite external resource	Yes
Tweet-structure	EC tweets are more likely than NE to contain hashtags	Yes
Vocabulary	NE tweets are more likely than EC to express uncertainty	Yes
Vocabulary	NE tweets are more likely than EC to use swear words	Yes
Topical	Certain topics are talked about more in EC tweets and vice versa	Yes

out their analysis is part of the one used by Garimella et al. (2018). The original data sets consists of five main topics and they use the one pertaining to Obamacare politics.[2] In their work, the authors calculate user polarity scores; all the polarity scores are in the range $(-2.5, 2.5)$. A higher positive score represents more conservative viewpoint, while a more negative score represents a more liberal viewpoint (Duseja and Jhamtani, 2019). In order to define echo chambers, they use the following *homophily score* for a user u:

$$H(u) = \frac{|S(u)| - |D(u)|}{|S(u)| + |D(u)|}$$

where:

1. $S(u)$ is the set of followees of u having same polarity as the user u;

2. $D(u)$ is the set of followees of u having different polarity as the user u.

Thus $H(u) \in [-1;1]$. A score of $H(u) = 1$ means that all the followees of the user u have same polarity about the given topic as u herself/himself (Duseja and Jhamtani, 2019). In the paper, the authors define a threshold of such that users with $H(u)$ above this threshold are said to be in an echo chamber. In order to carry out their analysis, the authors compare tweets from moderate conservatives users in an echo chamber (EC) and moderate conservatives users not in an echo chamber (NE). As said before, they run experiments on different levels: *tweet structure*, *vocabulary* used and *topic* involved. Table 3.1 shows that all the hypothesis are hold in the data: they observe statistically significant difference in frequency of usage of uncertainty depicting words, hashtags, swear words, and external URL links,[3] as well as a difference in the aspects of Obamacare talked about frequently between the two types of tweets (Duseja and Jhamtani, 2019). Finally the results in Table 3.1 suggests that:

- Users not in an echo chamber are more likely to tweet or re-tweet with citing external news link or other sources. This follows the general idea that users in an echo chamber might feel less of a need to justify their claims or opinions as people (followees) around them echo with similar opinions (Duseja and Jhamtani, 2019);

[2] Obamacare is a United States federal statute in the healthcare context.
[3] They use chi-squared tests to check all the hypotheses.

- Users in an echo chamber are more likely to use hashtags within their tweets. It is worth noticing that hashtags are generally used to emphasize messages and to pile the range of the message;
- Tweets from users not in an echo chamber are more likely to contain uncertainty depicting words and swear words that express frustration on witnessing opposing viewpoints. This confirms the idea that people in echo chamber are generally surrounded by similar opinions that do not jeopardize their beliefs;
- Topics discussed by users in echo chambers and users not in echo chambers are different.

3.1.2 Quantification and Reduction of Controversy in Echo Chambers

The studies on controversy reduction are quite recent. In the following, we are going to illustrate the main solutions that have been proposed for both controversy quantification and reduction over the last years.

Controversy Quantification

To the best of our knowledge, one of the best works on controversy quantification in echo chambers is the one of Garimella et al. (2015). In this paper, the authors develop a framework to identify controversy regarding topics in any domain (e.g., political, economical, or cultural), and without prior domain-specific knowledge about the topics in question. Figure 3.1 shows the pipeline introduced in the mentioned paper. Each stage in the pipeline has a specific purpose:

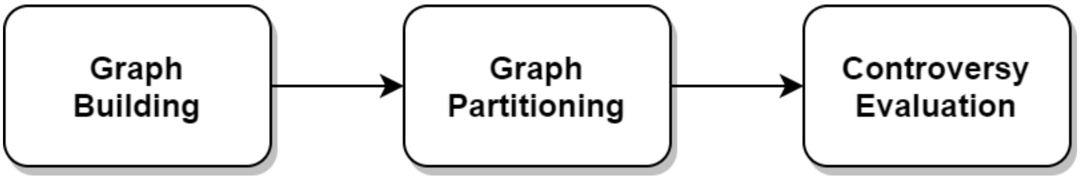


FIGURE 3.1: Pipeline introduced by Garimella et al. (2015) for quantifying controversy in a network.

1. *Stage 1 - Graph Building*: the purpose of this first stage is to build a *conversation graph* that represents activity related to a single topic of discussion. Each item related to a topic is associated with one user who generated it, and they build a graph where each user who contributed to the topic is assigned to one vertex. In this graph, an edge between two vertices represents endorsement, agreement, or shared point of view between the corresponding users.
2. *Stage 2 - Graph Partitioning*: this second stage aims at partitioning the graph in *two* distinct *communities*. In order to partition the *conversation graph*, they rely on state of art graph partitioning algorithms. Intuitively, the two partitions

defined by the algorithm correspond to two disjoint sets of users who possibly belong to different sides in the discussion.

3. *Stage 3 - Controversy Evaluation*: the third and last stage takes as input the graph built by the first stage and partitioned by the second stage, and evaluates controversy by means of a suitable controversy measure that capture how controversial the topic is. Intuitively, a controversy measure aims to capture how well separated the two partitions are (Garimella et al., 2015).

In summary, by using the pipeline illustrated in Figure 3.1, it is possible to generate a *conversation graph*, partition the graph to potentially find *echo chambers* and finally *evaluate controversy* in the network.

In order to properly create the *conversation graph*, Garimella et al. (2015) query Twitter[4] by using specif hashtags: each query refers to a topic and they choose a set of topics which is balanced between controversial and non-controversial ones, so as to test for both false positives and false negatives. As they suggest, often people on Twitter use a hashtag or a bunch of hashtags to discuss about one topic: that is why the pair topic-hashtag seems to be a suitable assumption. However for more details regarding the definition and creation of a topic from hashtags, we let the reader check the mentioned paper. For each topic, they build a graph G where they assign a vertex to each user who contributes to it, and generate edges according to four different types of endorsement. Based on these endorsements, they generate four types of graph: *retweet* graph, *follow* graph, *content* graph and *hybrid content & retweet* graph. The results suggest that both graph building methods, retweet and follow graph, are able to capture the difference between controversial and non-controversial topics while the others two type of graph have failed in this task.

Unlike the strategy based on polarization, Garimella et al. (2015) identify communities by using a different approach. In particular, they rely on the state-of-the-art *graph partitioning algorithms*. This category of algorithms has been developed in the research field of *Social Network Analysis* (SNA) and is based on the analysis of the *graph topology*. In particular, these graph partitioning algorithms exploit graph properties in order to partition a graph into several communities. The assumption here is that conversation graph of a controversial topic should have a clustered structure and this structure should be captured by a graph partitioning algorithms. Section 3.2 of this chapter will explain the mathematical fundamental underneath these approaches.

In relation to the identification of suitable controversy measures, we have dedicated Section 3.3 of this chapter to introduce those works related to this issue.

[4] Twitter is a natural choice for the problem at hand, as it represents one of the main fora for public debate in online social media, and is often used to report news about current events.

Controversy Reduction

An important work discussing how to reduce controversy and mitigate this way echo chambers is the one of Garimella et al. (2017). In particular, the authors employ the pipeline showed in Figure 3.1 in order to find two communities and then, using a novel link prediction algorithm, they reduce the overall controversy level within the network: they aim at finding the edges that produce the largest reduction in the controversy score (Garimella et al., 2017).

At the base of this process, there is the idea that connecting people of two disjoint communities may favour information diffusion. If there are no connections among the two communities, there is no possibility for any users in one community to get content of the other community and vice versa. In this situation, the phenomena previously described such as *confirmation bias*, *homophily* and *polarization* could be even more delineated.

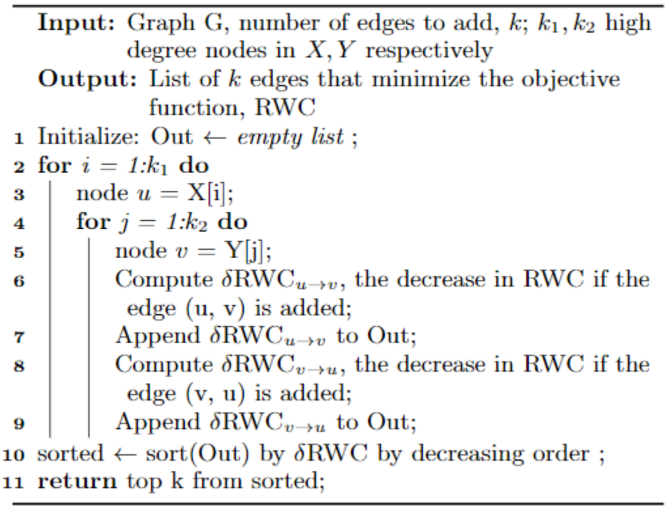
The findings described in Garimella et al. (2017) show how to reduce controversy through the addition of edges. One of their most innovative discovery is the possibility to solve controversy in a *language* and *domain agnostic* way for the first time. In fact the algorithm proposed by Garimella et al. (2017) simply rely on the graph structure of a network. They adopt controversy measures proposed by Garimella et al. (2015), as it is the most recent and valuable work discussing the considered problem.

Another finding of the paper starts from the observation that real networks often have a structure that resembles star-graphs in a certain way: a small number of highly popular nodes receive incoming edges from a large number of other nodes. This is based on the following model: in a controversial setting, there are thought *leaders* and *followers*. Most activity in the endorsement graph happens around retweeting and spreading the voice of the leaders across their side, on each side. This leads to a polarized structure which looks like a union of stars on each side of the controversy (Garimella et al., 2017).

This observation gives the fundamental to the authors to implement an algorithm which aims to consider edges between the high-degree nodes of each side instead of considering all possible pairs of nodes. They demonstrate that adding edges to connect high-degree nodes across the two sides of the network lead to the highest reduction in the controversy score. Probably, connecting those nodes with a lot of edges allows the information to spread out in the network.

Figure 3.2 shows the pseudo-code of the aforementioned algorithm. The algorithm, first identifies the high-degree nodes in both the communities,[5] then it iteratively calculates the controversy score and lists the pair of nodes in decreasing order in relation to the value of the controversy score: the pair of nodes which leads to the highest decrease in the score is listed first, this pair of nodes will be the first added in the graph. It is important to point out that just pairs of nodes made up by nodes in

[5] Communities, partitions or echo chambers refer to the same concept here.



```
Input: Graph G, number of edges to add, $k$; $k_1, k_2$ high
       degree nodes in $X, Y$ respectively
Output: List of $k$ edges that minimize the objective
        function, RWC
Initialize: Out ← *empty list* ;
for $i$ = *1:$k_1$* do
    node $u$ = X[i];
    for $j$ = *1:$k_2$* do
        node $v$ = Y[j];
        Compute $\delta\mathrm{RWC}_{u \to v}$, the decrease in RWC if the
          edge (u, v) is added;
        Append $\delta\mathrm{RWC}_{u \to v}$ to Out;
        Compute $\delta\mathrm{RWC}_{v \to u}$, the decrease in RWC if the
          edge (v, u) is added;
        Append $\delta\mathrm{RWC}_{v \to u}$ to Out;
sorted ← sort(Out) by $\delta$RWC by decreasing order ;
return top k from sorted;
```

FIGURE 3.2: *k*-EdgeAddition algorithm introduced by Garimella et al. (2017).

different communities are evaluated, i.e., only edges cross communities will be evaluated as possible candidates. Clearly, some bridges are more likely than others to materialize. For instance, people in the *middle* might be easier to convince than people on the two extreme, i.e., nodes in the boundary of the two communities should be easier to connect. They take this issue into account by modeling an acceptance probability for a bridge as a separate component of the model. They build such an acceptance model using user polarity. Intuitively, this polarity of a user takes values in the interval $[-1;1]$ and captures how much the user belongs to either side of the controversy. High absolute values (close to -1 or 1) indicate that the user clearly belongs to one side of the controversy, while middle values (close to 0) indicate that the user is in the middle of the two sides (Garimella et al., 2017).

The final algorithm proposed in the mentioned paper, takes into account also this probability value in order to rank the edges to add. Both the standard version and the one with acceptance probability model, seems to perform pretty well regarding most of the state-of-the-art link prediction algorithm.

3.1.3 Conclusions

The analysis of the most valuable works in the literature with respect to controversy reduction in social media, to tackle the echo chamber issue, lead to important conclusions that we have punctually summarized below:

1. On social media, there exist *controversial* and *non-controversial topics*. Usually the echo chamber phenomenon is likely to happen in a controversial environment because people often choose one side of the debate;

2. Under the assumption that *echo chambers* are more related to controversial topics, it is possible to identify these chambers (communities or partitions) analyzing either the content shared by users, or the graph topology of the social

network;

3. In the literature, there are several *metrics* that aim to measure the overall controversy level among communities that can have been identified as echo chambers. The choice of the best measures allows to have an idea about how controversial is the network;

4. Adding edges among different communities can reduce the overall controversy level. One way to get out of the echo chamber is bridging new fresh information of opposite sides;

5. Generally, nodes with a high centrality score may be good candidates for the application of an *edge addition* algorithm. It will be worth to add a sort of *acceptance probability* to the algorithm in order to get more realistic new connections. In fact, although it is hypothetically possible to connect nodes of different communities in order to maximize the reduction of the controversy, there are social reasons that make the addition of certain arcs more likely than others. These reasons will be explained in the next chapter.

In the rest of this chapter, we will provide some *background theory* that is necessary to understand the proposed approach. Since in this thesis we focus on the detection of different communities based on the study of topological aspects, we have applied the same pipeline illustrated in Figure 3.1. For this reason, in the next section, we will get through the theoretical part related to *graph partitioning*. In particular, the mentioned section will introduce the mathematical background of graph partitioning algorithms, some of which will be detailed in Section 4.1 of Chapter 4 that is dedicated to the description of the proposed approach. Furthermore, Section 3.3 will provide some theoretical background with respect to the definition and the identification of suitable *controversy measures*, while Section 3.4 will describe the theoretical issues related to *link prediction*. Also the specific measures for the controversy assessment and the specific algorithms for link prediction adopted in the proposed approach will be detailed in Sections 4.2 and 4.3 of Chapter 4.

3.2 Graph Partitioning

Computer scientists often use graphs as abstractions to model application problems. In fact, in many cases, it turns out that the natural relationships among objects has a graph structure: social networks are nothing more than relationships among people, scientific papers can be viewed as endorsements among researchers, routes and cities form networks by definition. One of the basic, but important, graph algorithmic operations is the division of a graph into smaller components. Even if the final application concerns a different problem, partitioning or clustering large graphs is frequently an important preliminary step that leads to complexity reduction or parallelization. Consequently, with the advent of so-called *big data* in many applications, suitable and effective graph partitioning and graph clustering techniques are becoming very important and strategic.

3.2.1 Community Detection Strategies

For the purpose of our analysis, we are going to use *graph partitioning* to reveal the presence of latent cohesive communities on social media platforms. As already mentioned, this hypothesis relies on the human attitude to make groups based on similar characteristics, attitude which could be emphasized by filtering algorithms. In this case, we may also use the word *community detection* (CD) algorithms. Community detection extracts structural information of a network in an unsupervised way. Communities are typically defined by sets of vertices densely interconnected which are sparsely connected with the rest of the vertices from the graph. Finding communities within a graph helps unveil the internal organization of a graph, and can also be used to characterize the entities that compose it (e.g., groups of people with shared interests, products with common properties, etc.)

Typically, graph partitioning is a challenging problem because it falls under the category of NP-hard problems.[6] Solutions for these applications are thus usually derived from heuristics and approximation algorithms (Schulz, 2016).

In this section, we are going to define what graph partitioning means, highlighting the mathematical theory related to this problem, and what is the difference between graph partitioning and graph clustering. We refer to Section 4.1 of Chapter 4, for a detailed explanation and motivation of the algorithm adopted for detecting communities in this thesis.

3.2.2 Preliminary definitions

Before introducing the mathematical concept of graph partitioning, we would like to start with a couple of definitions. In particular those of *graph* (and related definitions) and *cut*, due to their importance in this context.

Graph

A *graph* G consists of a set of nodes V and a set of edges $E \subset V \times V$ to represent relations between the nodes. In general, we write n for the number of nodes and m for the number of edges.

In a *weighted graph*, a weight is assigned to each edge. Such weights might represent for example costs, lengths or capacities, depending on the problem at hand.

[6] NP-hardness (non-deterministic polynomial-time hardness) is, in computational complexity theory, the defining property of a class of problems that are informally "at least as hard as the hardest problems in NP".

A graph $G(V, E)$ consisting of the set V of vertices and the set E of edges, which are ordered pairs of elements of V, is formally defined as a *directed*.

In an *undirected* graph, an edge $(u, v) \in E$ implies an edge $(v, u) \in E$ and that both edge weights are equal.

The set $\Gamma(u) := \{v : \{u, v\} \in E\}$ denotes the *neighbors* of a node u.

The *degree* $d(v)$ of a node v is the number of its neighbors. With Δ we denote the maximum degree of a graph. The weighted degree of a node is the sum of the weights of its incident edges.

A graph is *bipartite* if its node set can be divided into two disjoint sets U and V such that $u; v \in E$ implies $u \in U$ and $v \in V$ or vice versa.

A *subgraph* is a graph whose node and edge set are subsets of another graph.

Cut

A *cut* $C = (S, T)$ is a partition of V of a graph $G = (V, E)$ into two disjoint subsets S and T.

The *cut-set* of a *cut* $C = (S, T)$ is the set $\{(u, v) \in E \mid u \in S, v \in T\}$ of edges that have one endpoint in S and the other endpoint in T.

If s and t are specified vertices of the graph G, then a s - t cut is a cut in which s belongs to the set S and t belongs to the set T.

In an *unweighted undirected* graph, the size or weight of a cut is the number of edges crossing the cut.

In a *weighted undirected* graph, the size or weight is defined by the sum of the weights of the edges crossing the cut.

A cut is *minimum* if the size or weight of the cut is not larger than the size of any other cut while.

A cut is *maximum* if the size of the cut is not smaller than the size of any other cut.

3.2.3 Graph Partitioning Theory

We chose to use the mathematical definition of graph partitioning proposed by Schulz (2016). Given a number $k \in \mathbb{N}_{>1}$ and an undirected graph with non-negative edge weights, the *graph partitioning* problem asks for blocks of nodes $V_1, ..., V_k$ that partition the node set V, where:

1. $V_1 \cup ... \cup V_k = V$;
2. $V_i \cap V_j = \emptyset, \forall i \neq j$.

Sometimes, a *balance constraints* requires all the blocks to have equal size. A node $v \in V_i$ that has a neighbor $w \in V_j$, $i \neq j$, is a *boundary node*, An *edges* that runs between blocks is also called *cut edge*. The set $E_{ij} := \{\{u, v\} \in E : u \in V_i, v \in V_j\}$ is the set of cut edges between two blocks V_i and V_j. An abstract view of the partitioned graph is the so called *quotient graph*, where nodes represent blocks and edges are induced by connectivity between blocks, i.e. there is an edge in the quotient graph if there is an edge that runs between the blocks in the original, partitioned graph.

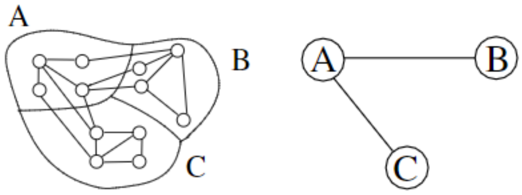


FIGURE 3.3: A partitioned graph with its quotient. Image taken from (Schulz, 2016).

Figure 3.3 shows an example of a graph partitioned into three blocks of equal size (four) and its quotient. The quotient may be seen as a sketch of the final result: it promptly tells us that the network has been divided into three different blocks. The structure also reveals a missing connection among partition C and B.

Objective Function

In practice, we often seek to find a partition that minimizes (or maximizes) an objective (Schulz, 2016). Minimizing the total *cut* is one of the most popular and used objective function. This means minimizing the following quantity:

$$\sum_{i<j}^{n} \omega(E_{ij})$$

In other words, the objective computes the sum of the weight of the cut edges. It is well-known that there exist more practical objective functions than minimizing the total cut, but for many applications this objective is still enough and guarantees good results.

It is worth mentioning that during the last decades minimizing the cut size has been adopted as a kind of standard (Schulz, 2016). However, the standard may be changing in the future since nowadays applications that need to partition social networks or web graphs are emerging. On this kind of networks, it is often required different objective functions.

3.3 Controversy Measures

According to Guerra et al. (2013), *polarization* in social sciences "is the social process whereby a social or political group is divided into two opposing sub-groups having conflicting and contrasting positions, goals and viewpoints, with few individuals remaining neutral or holding an intermediate position". Understanding and quantifying polarization is a long-term problem for researchers from various fields, and is also a key information for tasks like opinion analysis.

3.3.1 Controversy in Social Media

Referring to a specific topic on social media, a *controversy* or *polarization* measure aims to capture how controversial or polarized is the topic discussed. One option to catch the level of topic's polarization is through the use of *Natural Language Processing* (NLP) techniques. Technically speaking, NLP is a sub field of *Artificial Intelligence* (AI) and is all about teaching computers how to process and analyze large amounts of natural language data. In particular, there exists a particular set of NLP techniques, known as, *sentiment analysis* techniques, which aims exactly to synthesize the text's sentiment.

Another common option to identify polarization of communities comes from the field of Social Netowork Analysis, which we have already introduced in the previous sections. In general, researchers in the field of SNA assumes (either implicitly or explicitly) that a social network is polarized if nodes can be partitioned into two highly cohesive subgroups, reflecting, possibly, two contrasting viewpoints. In this case, a controversy measure aims at capturing how well separated the two partitions defined by a graph partitioning algorithm are.

Between the two different approaches, i.e., evaluating polarization by means of the use of Natural Language Processing techniques and Social Network Analysis, we did focus our attention on the second strategy. In this case, one of the most used measure to capture polarization is *modularity*. In fact, the modularity of a network quantifies the extent, relative to a random network, to which vertices cluster into community groups, and the higher its value, more modular the network is. However, Guerra et al. (2013) have shown that, although modularity is correlated to the social phenomenon of polarization, and highly modular networks are certainly linked with an increased likelihood of polarization of positions expressed by users who are part of the network, modularity is not a direct measure of polarization.

Therefore, there was the need to identify and possibly define new controversy metrics. In fact, to the best of our knowledge, there is a sort of lack in the literature. Despite a lot of works have been done in the identification of polarized groups, less studies have been focused on the quantification of controversy; with the term quantification, we mean synthesizing the controversy level through a proper metric. In the following, we illustrate three different studies that have identified and/or proposed one or more controversy measures.

3.3.2 Some State-of-the-art Controversy Measures

Guerra et al. (2013) have proposed a *polarization score* based on the notion of *internal* and *boundary* nodes. The main idea is to identify those nodes that effectively interact with the (potentially) opposing group. These nodes are part of a so-called *community boundary*, which they have defined, for a community G_i, as the subset of nodes in boundary $B_{i,j}$ that satisfies two conditions:

- A node $v \in G_i$ has at least one edge connecting to the opposing community G_j;

- A node $v \in G_i$ has at least one edge connecting to a member of G_i which is not connected to G_j.

Morales et al. (2015) have proposed a model to estimate opinions in which a minority of influential individuals propagate their opinions through a social network. They have further introduced an index to quantify the extent to which the social network is polarized. Their measure of polarization is inspired by the *electric dipole moment* - a measure of the charge system's overall polarity. For two opposed point charges, the electric dipole moment increases with the distance between the charges. Analogously, the polarization of two equally populated groups depends on how distant their views are.

To the extent of our knowledge, the most recent work on controversy quantification has been proposed by Garimella et al. (2015). In this study, Garimella et al. have tested both the polarization metrics introduced in the studies of Guerra et al. (2013) and Morales et al. (2015), but they have further proposed three novel measures based on the analysis of the graph topology. We refer to these measure as *random walk controversy*, *betweennees centrality controversy* and *embedding controversy*. The findings illustrated in Garimella et al. (2015) shows that, on average, the novel metrics are consistent with previous results on controversial graphs and, in most of the cases, outperform the state-of-the-art baselines.

In this section, we have given an overview on the state of the art in terms of controversy measure based on graph topology. We remind to Chapter 4, Section 4.2, for a more detailed explanation concerning the controversy measures that have been employed in the proposed approach. In particular, we will provide the motivations behind their choice, and their formal definitions.

3.4 Link Prediction

The addition of edges (i.e., links) between different communities is one of the solutions that can be applied to reduce the controversy between their members, thereby reducing the problem of echo chambers. The choice of which links to add is particularly important and must be made taking into account the characteristics of the social network itself.

3.4.1 The Link Prediction Problem

Link prediction is described as the action of inferring which new interactions among members of a social network are more likely to occur in the near future. Social networks are highly dynamic objects: they grow and change rapidly over time through the addition of new edges or the removal of old ones. Understanding the mechanisms by which they evolve is a fundamental thing and predicting the appearance of new interactions in the underlying social structure is even more important. Most of the time, it is possible to predict the evolution of a social network using features intrinsic to the network itself. One significant example is given by collaborations among scientists: two scientists who have never collaborated, but are "close" in the network will have colleagues in common, and will travel in similar circles: this could suggest that they are more likely to collaborate in the near future and perhaps a new connection between them is going to appear.

Successful *link prediction* methods could be used in a wide range of applications in order to better understand possible future interactions among entities. In addiction to its main functionality of predicting new future edges, *link prediction* algorithms could also be used to infer missing links from an observed network: in a number of domains, one constructs a network of interactions based on observable data and then tries to infer additional links that, while not directly visible, are likely to exist (Liben-Nowell and Kleinberg, 2007).

To the best of our knowledge, the use of link prediction algorithms to solve controversy is quite an innovative and recent application. Probably the first work that refer to link prediction in this context has been presented by Garimella et al. (2017).

However, methods and techniques implemented in the mentioned paper have already been described. Here, link prediction has been used to add pairs of nodes made up by nodes belonging to two different communities. We remind to Section 3.1 of Chapter 3, for a more detailed explanation.

3.4.2 Link Prediction Methods

All *link prediction methods* assign a connection weight $score(x,y)$, to pairs of nodes (x,y) based on the input graph $G(V,E)$, and then produce a ranked list in decreasing order of $score(x,y)$. Thus, they can be viewed as computing a measure of *proximity* or *similarity* between nodes x and y, relative to the network topology.

There exists a lot of methods in order to score pairs of nodes of a graph. Perhaps the most basic approach is to rank new pairs (x,y) by the length of their shortest path in G (in keeping with the notion that pairs of nodes have to be ranked in decreasing order of $score(x,y)$, in this case, $score(x,y)$ has to be defined as the negative of the shortest path length). Such a measure relies on the notion *small worlds*, in which individuals are related through short chains. Finally, we may add k number of new edges according to the ranked list produced using shortest path as similarity measure.

In general, it is possible to identify different *link prediction algorithms* according to their way to score possible new pairs of nodes. Liben-Nowell and Kleinberg (2007) has divided link prediction methods into two main categories:[7]

- *Methods based on node neighborhoods*. For a node x, let $\Gamma(x)$ denote the set of neighbors of x in G. These approaches are based on the idea that two nodes x and y are more likely to form a link in the future if their sets of neighbors

[7] The authors also describe the so-called *higher-level approaches* which are meta-approaches that can be used in conjunction with any of the methods mentioned in their paper.

$\Gamma(x)$ and $\Gamma(y)$ have large overlap. This follows the natural intuition that such nodes x and y represent authors with many colleagues in common, and hence are more likely to come into contact themselves;

- *Methods based on the ensemble of all paths*: these methods refine the notion of shortest-path distance by implicitly considering the ensemble of all paths between two nodes.

Techniques such as *PangeRank*, *Hitting time* or *SimRank* are the basis for the methods that use the ensemble of all paths. However, given that we did not use such as algorithms throughout our study, we will not describe them. For further details about the above-mentioned methods, please refer to the original paper (Liben-Nowell and Kleinberg, 2007).

This section aims to be a necessary, preliminary introduction to the concept of link prediction; concept which will be further described. In fact, in Section 4.3 of Chapter 4, we will mathematically describe those similarity measures used in our analysis; all the similarity measures throughout our study pertaining to those methods based on node neighborhoods. Moreover, in the mentioned section of Chapter 4, we will introduce two metrics that we have defined as "communicability measures".

Chapter 4

Implementing Controversy Reduction Solutions

In this chapter, we are going to explain how we have selected and combined some of the techniques proposed in Chapter 3 in order to quantify and reduce the controversy level in social media. This will be done through: (i) the detection of communities by using algorithms that focus on the structure of the social network; (ii) the measure of the level of controversy among communities through the choice and implementation of suitable measures; (iii) the choice and the implementation of link prediction algorithms. In the next sections, the choices made with respect to the three points illustrated above will be explained and justified.

It is worth mentioning that we partially follow the pipeline proposed by Garimella et al. (2015) and illustrated in Figure 3.1 in Chapter 3. In summary the pipeline is composed by three different phases: (i) *Graph Building*, (ii) *Graph Partitioning* and (iii) *Controversy Evaluation*. In Section 5.1 of Chapter 5 will be described the set of data used to carry out our analysis. These data are necessary to start with phase (i) and the next phases. Here we are going to concentrate on phases (ii) and (iii). However, given that the outline of this thesis is the reduction of the controversy level among polarized communities, we add two further phases to the already mentioned pipeline. This two phases are *Link Prediction* and *Controversy Re-evaluation*.

Figure 4.1 shows the final pipeline adopted to carry out the analysis.

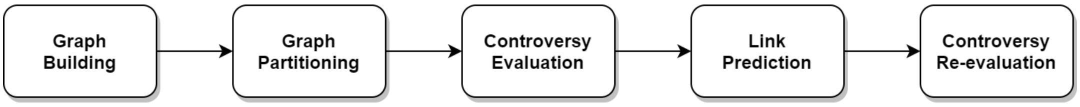


FIGURE 4.1: Framework used to carry out the analysis.

The last two phases, i.e., *Link Prediction* and *Controversy Re-evaluation*, aim to add new edges to the original graph and to re-evaluate the controversy level within the graph after that new connections have been added, respectively. Garimella et al. (2015) have showed that, through the pipeline illustrate in Figure 3.1, is possible to detect the echo chamber phenomenon and, moreover, to quantify the controversy. Here we show that, by following the pipeline depicted in Figure 3.1, it is possible to predict new edges in a efficient way, reducing this way the divergence among polarized communities.

The choices of which graph partitioning algorithms to employ and implement, in order to detect polarized communities, will be motivated in Section 4.1. The choices regarding the controversy measures to be used, in third and last phase, to evaluate how polarized is the social network, will be outlined in Section 4.2. Finally, the choices of which link prediction algorithms to employ in order to reduce the controversy level, will be illustrated in Section 4.3.

4.1 Community Detection

Grouping nodes, representing users in a social media platform, is the first essential steps towards the final goal of reducing controversy.[1] Therefore, the choices of the community detection algorithms and the number of potential communities to be found are two elements of extreme importance. In fact, these choices will affect the entire study, in particular the measurement of controversy in the social network.

4.1.1 How Many Communities?

Detecting communities is an unsupervised problem which requires prior knowledge about the structure, the interactions and connections among users. Defining the "right" number of final groups is always challenging because communities depend on the context and the topic of interest. Another important aspect is represented by the conditions and contexts in which echo chambers are more likely to appear: deciding one topic rather than another one could bring to unsatisfactory results.

However, in previous chapters (Chapter 2, Section 2.2 and Chapter 3, Section 3.1), we have introduced some important works that give useful clues about the issues above illustrated. Sasahara et al. (2019) depicted a world where people on social media debates tend to form exactly *two* communities. Therefore, it does not matter the kind of topic discussed online, what really matters is that, at the end, debates gravitate around two sides. Thinking about real cases, this seems to make sense and to be in accordance with the echo chamber hypothesis. Despite there could exist several different ways of thinking about one topic, in the long period, an individual has the attitude to get informed and gradually reach a specific opinion which is affected by the surrounded ones.

A scenario with multiple small echo chambers has reason to exist but it is less likely to appear because users, on social media platforms, tend to select claims that adhere to their system of beliefs and to ignore dissenting information. The consequence of this process could be a sort of homogeneity maximization which is better represented by two almost disjointed communities rather than various fragmented groups.

[1] Remember the real first step is creating a conversation graph.

Moreover, Garimella et al. (2018) support the idea that echo chambers in political context are more likely to emerge than in other contexts. Also this evidence seems to be reasonable because political debates, by nature, are divided into groups and most of the time these groups belong to two - left and right - different sides.

4.1.2 Topology-based or Content-based Community Detection?

Beside the number of possible communities to be found and the topic to be analyzed, others two important aspects that have to be taken into account are the computational cost of the community detection algorithm and its way of detecting community. Although the goal of this study is to demonstrate the existence of *echo chambers* and the possibility to reduce the divergence among them rather than build a real-time system able to detect community as fast as possible, we would like to identify communities in a reasonable amount of time. Another desirable thing would detect communities without putting any constraints on the algorithm: we would like to let the data talk and obtain communities which are the most similar to the one really present in the social network rather than force the groups to be of the same shape or have the same number of members.

Nonetheless, we want to explain a bit deeper why it has been choice to discard text mining approaches in order to find out echo chambers. Such techniques have been shown to be powerful tools and clustering algorithms based on text and words are often used in order to get data insights. In fact, beside the use of polarization and graph topology, another possible solution to detect communities in a social media platforms could be the use of text clustering algorithm. Here we go to the field of Natural Language Processing, as introduced in Section 3.3. Therefore, by applying a NLP technique, such as *Latent Semantic Analysis*, it would be possible to cluster text data. In this case, the assumption is that people belonging to different *echo chambers* use different words and jargon. Assumption confirmed by the work of Duseja and Jhamtani (2019). Using NLP instead of graph topology or polarization could be a valuable alternative. However, detecting communities in this way would require not only the social connections among participants of a social media platforms, but also the social interactions among them represented in the form of plain text; condition which in most of the cases described is also required for the polarization technique. Working with text requires a lot of preprocessing steps in order to clean and standardize words and phrases, even when the text is short like a tweet. Moreover, if we use text instead of the graph structure, we definitely depend on the language itself. Imagine for example the case of two users who belong to the same community but write in different language, e.g., Chinese and Arabic, it would be challenging building a NLP pipeline able to group this two individuals in the same chamber.

Using pure community detection algorithms coming from social networks theory allows to exploit the graph structure and therefore building a pipeline which is domain and language independent. This means that despite the type of users who are present in the social media platform, we will be always able to detect community by looking the social connections (edges) among participants (nodes).

As consequence of what said previously, we do rely on graph partitioning algorithms based on graph topology. In particular, Section 4.1 of Chapter 4 will fully explained the algorithms adopted in our study.

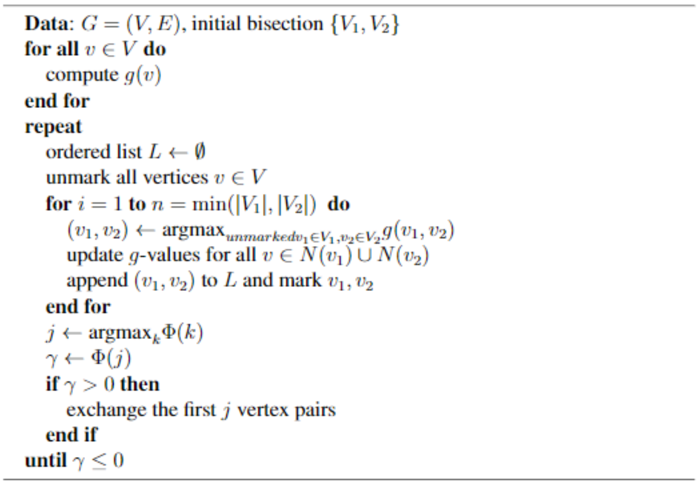
4.1.3 Graph Partitioning Algorithms

In this section, we are going to describe the *graph partitioning* algorithms, also from a mathematical point of view, which have been selected, implemented and used throughout our study. Therefore, this section does not contain an exhaustive list of all possible graph partitioning methods available in literature, but rather a focus on those algorithms that are more suitable with respect to the considered problem and that have been implemented in this thesis.

Kernighan-Lin Algorithm

The *Kernighan–Lin algorithm* is a heuristic algorithm for partitioning graphs into two parts. It was originally proposed by Kernighan and Lin (1970). Kernighan and Lin were probably the first that defined the graph partitioning problem and worked on local improvement methods for this problem (Schulz, 2016).

The input to the algorithm is an undirected graph $G = (V, E)$ with vertex set V, edge set E, and (optionally) numerical weights on the edges in E. The algorithm's goal is to divide V into two disjoint blocks A and B of equal size, in a way that minimizes the sum T of the weights of the subset of edges crossing from A to B. If the graph is unweighted, then instead the goal is to minimize the number of crossing edges. In each *step*, the algorithm preserves and enhances a partition using a *greedy strategy* for linking vertices of A with vertices of B, so that moving the paired vertices from one side to the other would improve the partition. After matching the vertices, it then performs a subset of the selected pairs to have the best overall effect on the quality of the solution T. Given a graph with n vertices, each step of the algorithm runs in $O(n^2\, log\, n)$ time. Hence, the Kernighan-Lin algorithm is an heuristic approach that consists of finding *good* sets (A, B) and exchanging the corresponding nodes until this exchange does not decrease the number of edges cut. A key concept for the Kernighan-Lin algorithm is the one of *node's gain* for v, i.e., the reduction in the cut when v is moved from one block to the other. Thus, when $g(v) > 0$, it is possible to decrease the cut by $g(v)$ by moving v to the opposite block. This concept is pivotal within algorithm in order to find those nodes that minimize the *cut*. The main idea here is minimizing the *cut* by swapping nodes between two different partition. Figure 4.2 shows the pseudo-code of the Kernighan-Lin algorithm taken from Schulz (2016).



```
Data: $G = (V, E)$, initial bisection $\{V_1, V_2\}$
for all $v \in V$ do
  compute $g(v)$
end for
repeat
  ordered list $L \leftarrow \emptyset$
  unmark all vertices $v \in V$
  for $i = 1$ to $n = \min(|V_1|, |V_2|)$ do
    $(v_1, v_2) \leftarrow \text{argmax}_{unmarked v_1 \in V_1, v_2 \in V_2} g(v_1, v_2)$
    update $g$-values for all $v \in N(v_1) \cup N(v_2)$
    append $(v_1, v_2)$ to $L$ and mark $v_1, v_2$
  end for
  $j \leftarrow \text{argmax}_k \Phi(k)$
  $\gamma \leftarrow \Phi(j)$
  if $\gamma > 0$ then
    exchange the first $j$ vertex pairs
  end if
until $\gamma \leq 0$
```

FIGURE 4.2: Pseudo-code of the Kernighan-Lin algorithm.

Label Propagation

Even though we did not use this algorithm throughout our study, we have employed an innovative method which is based on *label propagation*. Therefore, it is worth having a definition of it. Despite the *Kernighan-Lin* algorithm, that was developed in the 70's, the *Label Propagation Algorithm* (LPA) is much more recent. In fact, it was proposed for the first time by Raghavan, Albert, and Kumara (2007). As the original title of the paper suggest, it is a fast, near-linear time algorithm that locally optimizes the number of edges cut.

To the extent of our knowledge, a short, but satisfactory explanation of *label propagation* is given in Schulz (2016). Therefore, we are going to use the description given in the aforementioned paper. Initially, each node is in its own cluster/block, i.e., the initial block ID of a node is set to its node ID. The algorithm then works in rounds. In each round, the nodes of the graph are traversed in a random order. When a node v is visited, it is moved to the block that has the strongest connection to v, i.e., it is moved to the cluster V_i that maximizes $\omega(\{(v;u) \mid u \in N(v) \cap V_i\})$. Ties are broken randomly. The process is repeated until it has converged. One label propagation round can be implemented to run in $O(n+m)$ time.

Hence, the intuition behind the algorithm is that a single label can quickly become dominant in a densely connected group of nodes, but it will have trouble crossing a sparsely connected region. Labels will get trapped inside a densely connected group of nodes, and those nodes that end up with the same label when the algorithm finishes are considered part of the same community. Figure 4.3 gives a simple idea of the application of this algorithm.

FluidC

Previously, we have briefly introduced the propagation method, which represents the state-of-the-art in terms of computational cost and scalability (Yang, Algesheimer, and Tessone, 2016).

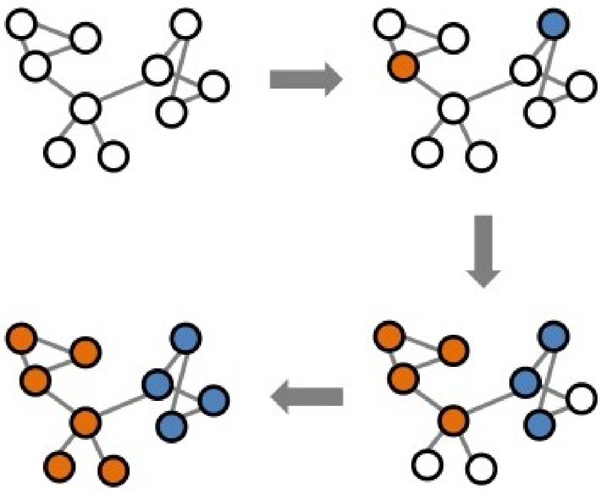


FIGURE 4.3: At each iteration LP updates node's label based on its neighbors.

For the purpose of our analysis, we wanted to implement a graph partitioning algorithm as computational efficient as LPA, but with the possibility to define k number of communities to find out. To the best of our knowledge, one of the first and most recent algorithm with these two properties is the one proposed by Parés et al. (2017). As the original paper declares "this algorithm mimics the behaviour of several fluids (i.e., communities) expanding and pushing one another in a shared, closed and non-homogeneous environment (i.e., a graph), until equilibrium is found". Hence, the so-called *FluidC* can identify any number of communities in a graph by initializing a different number of fluids in the environment. According to Parés et al. (2017), FluidC is the first propagation-based algorithm with this property, which allows the algorithm to provide insights into the graph structure at different levels of granularity.

Given a graph $G = (V, E)$ constituted by V and E, set of vertices and a set of edges respectively and k number of desired communities where $0 < k <= |V|$, then FluidC first initializes k fluid communities $C = \{c_1, .., c_k\}$. Just like label propagation, each community $c \in C$ is initialized in a different and random vertex $v \in V$. In addiction, each initialized community has an associated density d within the range $[0; 1]$. The density of a community is defined as the inverse of the number of vertices composing that community, formally:

$$d(c) = \frac{1}{v \in c} \tag{4.1}$$

Notice that a fluid community composed by a single vertex (e.g., every community at initialization) has the maximum possible density ($d = 1$). FluidC operates through the so-called *supersteps*. At each superstep, the algorithm iterates in random order over all vertices of V, updating the community each vertex belongs to using an update rule. The algorithm converges and ends when the allocation of vertices to communities does not change over two consecutive supersteps. The update rule for a specific vertex v returns the community or communities with maximum aggregated density within the ego network of v (Parés et al., 2017). An exhaustive explanation of the mathematical formula underneath the update rule is given in the mentioned paper.

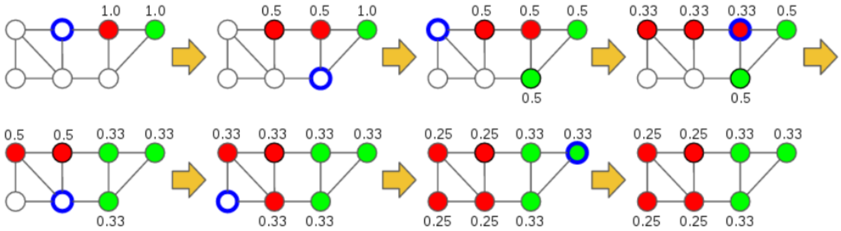


FIGURE 4.4: *Fluidc* initialized for finding $k = 2$ communities. Image taken from (Parés et al., 2017).

Figure 4.4 briefly shows how FluidC finds two communities (red and green) within a graph. Each vertex assigned to a community is labeled with the density of that community; the vertex highlighted in blue is the one evaluated at each step. After a complete superstep the algorithm converges.

Thus, FluidC is *asynchronous*, because the latest partial state of the graph is used to update each vertex (some vertices may have updated their label in the current superstep and some may not). A synchronous version of FluidC would not guarantee that densities are consistent with Equation 4.1 at all times (e.g., a community may lose a vertex but its density may not be increased immediately in accordance). Therefore, a community could lose all its vertices and be deleted from the graph. FluidC allows for the definition of the *number* of *communities* to be found, simply by initializing a different number of fluids in the graph. This is a valuable property for data analytics, as it allows the graph and its entities to be analyzed at various levels of granularity. This is not the unique community detection algorithm with this property, but, to the best of our knowledge, it is the first that use the efficient propagation method. Finally, FluidC avoids the creation of giant communities in a non-parametric manner. Due to the spread of density among the vertices that compose a community, a large community (when compared to the rest of communities in the graph) will only be able to keep its size and expand by having a favourable topology, i.e., having lots of intra-community edges which make up for its lower density (Parés et al., 2017). FluidC is designed for connected, undirected, unweighted graphs, but can be easily applied to a disconnected graph $G^{'}$ just by performing an independent execution of FluidC on each connected component of $G^{'}$ and appending the results.

Although FluidC is a novel and recent community detection algorithm, it is able to identify high quality communities, getting close to the current best alternatives in the state of the art (Parés et al., 2017). FluidC seems to be competent on large graphs, that is why it should be suitable for our analysis (social media platforms usually involve a lot of connections among people). The importance of this algorithm can be summarized both in terms of scalability and diversity. In terms of scalability because FluidC is a propagation-based algorithm and this family of community detection algorithms has been shown to be the most scalable. In terms of diversity because it is the first propagation-based algorithm which allows the definition of k number of desirable communities.

4.1.4 Discussion

Before going on with the description of the rest of the thesis work, we want to point out that the goal of our analysis is not the evaluation, in general, of several graph partitioning algorithms, but it rather wants to demonstrate the possibility to reduce the overall level of controversy within a social network. Therefore, even though *community detection* is an essential stage in our pipeline, the choice of which graph partitioning algorithms to adopt was guided by the popularity of the algorithms themselves and by the possibility of adapting well to the problem under consideration. For these reasons, we have chosen the well-known *Kernighan-Lin* graph partitioning algorithm, and a novel but particuarly effective approach, i.e., *FluidC*, which have been described in the previous sections. As consequences of what said in the introduction of this section, we have chosen to seek out two communities ($k = 2$). This choice is also motivated by other three factors:

(i) As stated in Section 3.3 of Chapter 3, polarization in social sciences refers to a social process whereby social debate is divided into *two* opposing groups;

(ii) All the *controversy metrics*, that will be formally introduced in the next section, are based on the above definition of polarization. Therefore, they rely on the presence of two partitions;

(iii) The *set of data* used to carry out the analysis has been collected by Garimella et al. (2015) whom have already validated the presence of two communities in most of the analysed topics.

However, in order to check that two is really a suitable number of final partitions, we did evaluate the quality of the partitions by using a quality metric called *coverage*. Coverage is the ratio of the number of intra-community edges by the total number of edges: by definition, an ideal cluster structure, where the clusters are disconnected from each other, yields a coverage of 1, as all edges of the graph fall within clusters (Fortunato, 2010). Moreover, we properly visualized all the graphs taken into considerations in our analysis. This qualitative evaluation has been taken in order to check the real presence of two communities.

Another important aspect taken into account during the analysis is the computational cost of the algorithms. It turned out that FluidC is extremely faster than Kernighan-Lin; Section 5.2 of Chapter 5 shows the computational costs of both the graph partitioning algorithms. Moreover, FluidC does not put any balancing constraints on the final result. This is a valuable property for a graph partitioning algorithm. As consequences, we do have reason to discard Kernighan-Lin in the evaluation process.

4.2 Controversy Measures

In Chapter 3, Section 3.3, we have mentioned at a high level some of the *controversy measures* that have been proposed in the literature. It is worth pointing out that all the above-mentioned controversy measures have been defined and tested for graphs in which two partitions have been identified, as in the case of this thesis. Formally, they rely on the presence of two partitions $X \cup Y = V \, and \, X \cap Y = \emptyset$, with V set of vertices. To the extent of our knowledge, metrics based on the assumption of more than two echo chambers have not been extensively studied yet in the literature, probably due to the fact, as illustrated in Section 3.1 of Chapter 3, virtual communities tend to divide into two groups highly polarized with respect to highly debated topics.

Another important fact analyzed in Section 3.2 of Chapter 3 is that, as in the case of graph partitioning algorithms, there are several families of measures that can consider different aspects, including the content diffused and the graph topology. Here, having focused on the use of community detection algorithms that are based on the structure of the social network, we have considered those controversy measures that consider the graph topology. As it was detailed in Section 3.3 of Chapter 3, this allows to the proposed solution to be domain- and language- independent.

4.2.1 Suitable Metrics

In this section, we will formally introduce five different controversy measures, all based on the study of the graph topology, which have been identified as the most suitable for the considered problem.

Random Walk Controversy

The first metric presented has been proposed by Garimella et al. (2015). Formally, given a graph $G = (V, E)$ and its two partition X, Y where $X \cup Y = V \, and \, X \cap Y = \emptyset$, we randomly select one partition (each with probability 0.5) and consider a random walk starting from a random vertex within the selected partition. The walk terminates when it visits any high-degree vertex. The measure captures the probability of being exposed to authoritative content from the opposite side by a random user on either side. The *authoritativeness* is captured by the degree of the node. Therefore, the *Random Walk Controversy* (RWC) measure is defined as follow: "consider two random walks, one ending in partition X and one ending in partition Y, RWC is the difference of the probabilities of two events: (i) both random walks started from the partition they ended in and (ii) both random walks started in a partition other than the one they ended in" (Garimella et al., 2015).

Thus, the measure is formally defined as:

$$\begin{aligned} &RWC = P_{xx}P_{yy} - P_{xy}P_{yx} \\ &where\ P_{AB}, A, B \in \{X, Y\}\ is\ the\ conditional\ probability \end{aligned} \tag{4.2}$$

$$P_{AB} = Pr[start\ in\ partition\ A | end\ in\ partition\ B]. \tag{4.3}$$

The *RWC* value tends to one when the probability of crossing sides is low, and tends to zero when the probability of crossing sides is equivalent to that of remaining on the same side. In other words we face a controversial situation when *RWC* is close to one.

Embedding Controversy

This measure is based on a low-dimensional embedding of graph G produced by the Gephi's *ForceAtlas2* algorithm (Jacomy et al., 2011) and it has been introduced by Garimella et al. (2015). The idea underneath this measure is the one of measuring the distance among pairs of vertices embedded in a new space produced by the *ForceAtlas2*. Graph embedding is an approach that is used to transform nodes, edges, and their features into vector space (a lower dimension) whilst maximally preserving properties like graph structure and information.

Given the two-dimensional embedding $\phi(v)$ of vertices $v \in V$ produced by *ForceAtlas2*, then we may determine the following quantities:

- d_x and d_y, the average embedded distance among pairs of vertices in the same partition, X and Y respectively;
- d_{xy}, the average embedded distance among pairs of vertices across the two partitions X and Y.

The *Embedding Controversy* measure EC is formally defined as:

$$EC = 1 - \frac{d_x + d_y}{2d_{xy}} \tag{4.4}$$

For controversial topics, ergo when we face better-separated graphs and higher degree of controversy, EC is close to one, while for non-controversial situation, EC is close to zero.

Betweenness Centrality Controversy

This metric has been proposed by Garimella et al. (2015) too.

Let us consider the set of edges $C \subseteq E$ in the *cut* defined by the two partitions X, Y. By using the notion of *edge betweenness*, this measure aim to catch how the betweenness of the cut differs from that of the other edges. The betweenness centrality $bc(e)$ of an edge e is defined as:

$$bc(e) = \sum_{s \neq t \in V} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}}, \tag{4.5}$$

where $\sigma_{s,t}$ the total number of shortest paths between vertices s, t in the graph and $\sigma_{s,t}(e)$ is the number of those shortest paths that include edge e. Edge betweenness is a centrality measure which quantifies the importance of an edge within the graph.

The hypothesis is that the cut should consist of edges that connect *structural holes* (Section 4.3 illustrates the concept of structural holes), if the two partitions are well-separated. In this situation, the shortest paths that connect vertices of the two partitions will pass through the edges in the cut, resulting in high values of betweenness for edges in C. On the other side, i.e., the two partitions are not well divided, the cut will consist of strong ties. In this case, the links that connect vertices between the two partitions will pass through one of the many edges in the cut, corresponding to betweenness values for C similar to the rest of the graph. Given the distributions of edge betweenness on the cut and the rest of the graph, the authors compute the Kullback-Leibler [2] divergence d_{kl} of the two distributions by using kernel density estimation to compute the PDF and sampling 10,000 points from each of these distributions with replacement (Garimella et al., 2015).

The *Betweenness Centrality Controversy* (BWC) is then formally defined as:

$$BWC = 1 - e^{-d_{kl}}, \tag{4.6}$$

which has values close to zero in case of controversy absence, and close to one for controversial topics.

Boundary Connectivity Controversy

This controversy measure relies on the notion of *boundary* and *internal* vertices. It was originally proposed by Guerra et al. (2013). Let $u \in X$ be a vertex in partition X; u is defined as boundary vertex of $X \, iff$ it is connected to at least one vertex of the other partition Y, and it is connected to at least one vertex in partition X that is not connected to any vertex of partition Y. Therefore, we are ready to define B_x, B_y, the set of boundary vertices for each partition, and $B = B_x \cup B_Y$ the set of all boundary vertices. While, vertices $I_x = X - B_X$ are said to be the internal vertices of partition X (similarly for I_Y). Let $I = I_x \cap I_y$ be all internal vertices in either partition. The reasoning for this measure is that, if the two partitions represent two sides of a controversy, then boundary vertices will be more strongly connected to

[2] Kullback-Leibler is a very useful way to measure the difference between two probability distributions.

internal vertices than to other boundary vertices of either partition (Garimella et al., 2015).

The *Boundary Connectivity Controversy* (*BCC*) measure is formally defined as:

$$BCC = \frac{1}{|B|} \sum_{u \in B} \frac{d_i(u)}{d_b(u) + d_i(u)} - 0.5, \tag{4.7}$$

where $d_i(u)$ is the number of edges between vertex u and internal vertices I, while $d_b(u)$ is the number of edges between vertex u and boundary vertices B. Lower values of the measure correspond to lower degrees of controversy.

Dipole Moment Controversy

This controversy measure was introduced by Morales et al. (2015), and is based on the physics notion of *dipole moment*, which, at a high level, can be defined as a measure of the charge system's overall polarity.

From a more formal point of view, let $R(u) \in [-1, 1]$ be a polarization values assigned to vertex $u \in V$. Intuitively, values of $R(u)$ close to its extreme (-1 and 1) correspond to users who have been polarized by one of the two sides; while values close to 0 correspond to neutral users. To set the values $R(u)$, it is possible to follow the process described in the original paper, i.e., setting $R(u) = \pm$ for the top-5% highest-degree vertices in each partition X and Y and setting the values for the rest of the vertices by using label-propagation. Furthermore, let $n^+ \, and \, n^-$ be the number of vertices V with positive and negative polarization values, respectively, and ΔA the absolute difference of their normalized size $\Delta A = |\frac{n^+ - n^-}{|V|}|$. Furthermore, let $gc^+(gc^-)$ be the average polarization values among vertices $n^+(n^-)$ and set d as half of their absolute difference, $d = \frac{|gc^+ - gc^-|}{2}$.

The *Dipole Moment Controversy* (*DMC*) measure can be formally defined as:

$$DMC = (1 - \Delta A)d \tag{4.8}$$

The idea behind this measure is simple but effective: if the two partitions X and Y are well separated, then label propagation will assign different extreme $(\pm)$ $R(u)$-values to the two partitions, resulting in higher values of the *MBLB* measure. Note also that larger differences in the size of the two partitions (reflected in the value of ΔA) lead to decreased values for the measure, which takes values between zero and one (Garimella et al., 2015).

4.2.2 Discussion

With respect to the choice of the controversy measures to adopt in the proposed approach among those detailed in the previous sections, we have decided to rely on three of them. In particular the choice has fallen into:

1. *Random Walk Controversy*;

2. *Embedding Controversy*;

3. *Boundary Connectivity Controversy*.

Random Walk Controversy and Embedding Controversy are measures that have been recently studied and applied to the problem considered in this thesis. It has been already shown that they perform very well compared to the state-of-the-art controversy measures. Therefore *novelty* and *performance* are the reasons of our choice. Instead, we did choose to use the so-called Boundary Connectivity Controversy as a sort of baseline. This is one of the first controversy measure adopted in literature. Our expectation is that its reliability and accuracy has already been proved and confirmed.

4.3 Link Prediction

The link prediction problem is the solution adopted in this thesis to reduce controversy among polarized communities. It can be summarized in three steps, as follows:

(i) Define a "measure", $score(x,y)$, generally a similarity measure, for each pair of nodes (x,y);

(ii) Sort, in descending order, the pairs of nodes according to the above measure;

(iii) Chose k, with k less than or equal to the maximum number of possible new edges to be added.

For the purpose of our analysis, we do evaluate just pairs of nodes (x,y) that belong to different partitions, i.e., $x \in X$ and $y \in Y$ or vice-versa, $X \cup Y = V \, and \, X \cap Y = \emptyset$, with V set of vertices. Since we are interested in reducing the overall controversy level within the graph by connecting two different communities, we do think that, connecting those links among pairs of nodes belonging to different partitions, could cut down controversy. In fact, as we have already said, this procedure should enhance information diffusion and help users to get out of the chamber.

Here, we want to concentrate on those methods which rely on *node neighborhoods* in order to predict the list of ranked pairs of nodes (x,y) that could be possibly added in the input graph G. We remind that the set of neighbors of node x in G is denoted with $\Gamma(x)$.

On social media, it is common to recommend friends of friends. Moreover using node neighborhoods, in order to predict new links, is a kind of baseline and we rely it could also be efficient within this context.

The assumption is the one of connecting nodes which belong to different chambers, but in somehow they have degrees of similarity: even though they are part of different sides, they have common neighbors. Therefore, it will be much more easier connecting these nodes in reality. Ideally, these node could be the ones in the boundary of one community. In general these nodes represent individuals who have not been (excessively) polarized towards one chamber: they are nothing more that people in the middle or perhaps gatekeeper, if we use the notation of Garimella et al. (2018). As a consequence, in the next section, you will find four different metrics used to score pairs of nodes; all of them are based on the notion of neighborhood.

4.3.1 Similarity Measures based on Node Neighborhoods

A list of the most popular similarity measures that are based on the concept of node neighborhoods are illustrated in the following.

Jaccard's Coefficient

The *Jaccard's coefficient* is a commonly used similarity metric in many disciplines and it measures the frequency that both x and y have a *feature* f, for a randomly selected feature f that either x or y has (Liben-Nowell and Kleinberg, 2007). When we take neighbors in G to be the *feature* in the Jaccard's Coefficient, we may define the following measure:

$$s^{J}_{x,y} = \frac{\Gamma(x) \cap \Gamma(y)}{\Gamma(x) \cup \Gamma(y)} \tag{4.9}$$

The *Jaccard's coefficient* is equal to 1 when the overlapping among neighbors of x and y in maximum, while it is equal to 0 when they do not have common neighbors.

Adamic Adar Index

Adamic and Adar (2001) have defined a measure to predict links in a social network. They smartly refined the simple counting of common neighbors among nodes by weighting nodes with few number of neighbors more heavily. The concept is that common elements with very large neighborhoods are less significant when predicting a connection between two nodes compared to elements shared between a small number of nodes. The *Adamic Adar Index* is defined as:

$$s^{AA}_{x,y} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{\log |\Gamma(z)|} \tag{4.10}$$

Resource Allocation Index

A variant of the *Adamic Adar Index* is represented by the *Resource Allocation Index*. In this case, nodes with very large neighborhoods are even more penalized due to the absence of the logarithm. The formula for this index is given by:

$$s^{RA}_{x,y} = \sum_{z \in \Gamma(x) \cap \Gamma(y)} \frac{1}{|\Gamma(z)|} \tag{4.11}$$

Preferential Attachment

Preferential Attachment relies on the aphorism "*rich get richer and poor get poorer*". A preferential attachment process is any of a class of processes in which a certain feature is distributed among a number of people or objects depending on how much they already have. Therefore, when the feature is defined by the neighbors of a node x in G, we can conclude that the probability that a new edge involves node x is proportional to $|\Gamma(x)|$, the current number of neighbors of x.
Moreover, on the basis of empirical evidence that the probability of co-authorship of x and y is correlated with the product of the number of collaborators of x and y, it is possible to define the following measure in order to predict possible new links:

$$s_{x,y}^{PA} = |\Gamma(x)| \times |\Gamma(y)| \tag{4.12}$$

4.3.2 Communicability Measures

All the previous measures have gotten excellent performance and they are considered to be a sort of baseline in the field of link prediction. Therefore, we do think that they may be also suitable for our specific goal of reducing the overall controversy level between two echo chambers.

However, we have decided also to adopt two further measures in order to rank new possible pairs of nodes to be added. We defined these measures to be *Communicability Measures* because they are based on metrics which aim to capture this characteristic. In particular, they are based on *Betweennes* and *Effective Size*, respectively. As previously done, also in this case we do consider only pairs of nodes that belong to different partitions.

Betweenness Index

For the first measure, we got inspired by Garimella et al. (2017), in particular by the algorithm 3.2 in Chapter 3 which use *degree* of nodes in order to identify edges that should reduce the controversy measure the most. Instead of considering those nodes with the highest degree, we take into account the *betweenness* of nodes.

We have already introduced the concept of edge betweenness, which is almost the same of node betweenness but, as the name suggests, it measures the bewteenness of edges and not of vertex.

Equation 4.13 shows the mathematical formula for *node betweenness*. Here, $\sigma_{s,t}$ is the total number of shortest paths from node s to node t, and $\sigma_{s,t}(v)$ is the number of those paths that pass through node v.

$$g(v) = \sum_{v \neq s \neq t} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \tag{4.13}$$

Betweenness is a centrality measure which aims to capture the importance of a vertex in terms of communicability: it represents the degree to which nodes stand between each other. In other words, nodes with high values of betweenness should be the ones responsible of information diffusion since they act like bridges among nodes in a social network. Consequently, we claim that connecting nodes with high values of betweenness within one community to those nodes with high values of betweenness in the other community should positively affect the final information diffusion and reduce the controversy level.

To the best of our knowledge, this is the first time that such as method has been applied in the field of controversy reduction. We have individually developed the same index which was applied for the first time by Zhang et al. (2015).

Motivated by the idea that communicability should favour information diffusion and reduce controversy, we hypothesize that the probability of two nodes to reduce the overall controversy is related to their ability to spread information within their own community.

We may conclude that if the sum of two nodes' betweenness, belonging to different communities, is bigger, the two nodes are more likely to spread opposite content. We are now ready to define our *Betweenness Index* as:

$$s^B_{x,y} = g(x) + g(y) \tag{4.14}$$

Despite Zhang et al. (2015) whom have calculated betwenness considering the entire graph G, we did use another approach. In fact, given a node, v, its betweenness value, $g(v)$, has been calculated considering not the whole graph, but only its community. This means that, given a community X of G, we have created a subgraph of G considering only edges and nodes in X. This allows to measure the betweenness of v using only the subgraph of G.

The main motivation is that we want to identify those nodes, within each community, which are the most important in terms of communicability. Therefore, considering only the community of a node allows to get its betweenness within the community, i.e., the importance in terms of communicability within the chambers.

Effective Size Index

The second measure is based on the concept of *structural holes*. Ronald Burt (Burt, 2009) coined and popularized the term "structural hole" to refer to some very important aspects of positional advantage/disadvantage of individuals that result from how they are embedded in neighborhoods. Figure 4.5 shows what being in a better position means.

In the figure, If we compare node A with node B, although they have the same number of links, A has more chance to get novel information than B. The location of node A lets it act as a *bridge* or *broker* between three separate clusters; so node A will probably receive non-redundant information from its contacts. Instead, B and

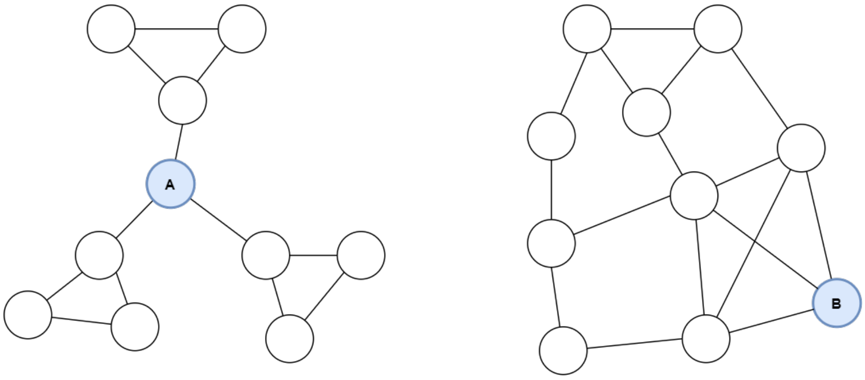


FIGURE 4.5: A is in a better position than B.

its neighbors are densely connected among each other. Therefore, any information that any of them could get from B, it could easily get from other nodes as well. In this situation, the information received by B, and any nodes in its neighborhood, is likely to be overlapping: the connections are redundant. The term "structural holes" is used for the separation between non-redundant contacts. In this case without node A, there would be a hole among the clusters and information diffusion would be undermined. As consequence, we do believe that connecting nodes with non-redundant connections should improve information diffusion and reduce controversy.

Burt (2009) proposed a measure of network's redundancy; this measure is called *Effective Size*. A node's network has redundancy to the extent that its contacts are connected to each other as well. The non-redundant part of a node's relationships it is the effective size of its network.

The effective size $e(v)$ of a node u is defined by:

$$e(u) = \sum_{v \in \Gamma(u) \setminus \{u\}} (1 - \sum_{w \in \Gamma(v)} p_{uw} m_{vw}); \tag{4.15}$$

where $p_{uw} m_{vw}$ is the defined as Redundancy (Burt, 2009).

Stimulated by the same idea that communicability should favour information diffusion and reduce controversy. If the sum of two nodes' effective size, belonging to different communities, is bigger, the two nodes are more likely to spread opposite content. Therefore, we define the *Effective Size Index* as:

$$s^E_{x,y} = e(x) + e(y) \tag{4.16}$$

The effective size of a given node u has been calculated using the same approach followed for the betweeness, i.e., using the subgraph of G defined by the community of u. The motivation is always the same: we want to get the importance of a node, in terms of communicability, within a community and not for the entire graph. The goal is to improve communication between two communities, therefore, we need to identify those important nodes for each community. Important nodes for the entire graph G are not necessary important also for the individual communities and vice-versa.

4.3.3 Discussion

Adding crossing edges from one polarized community to another one in a social network should improve communication among these chambers and reduce the controversy. By using specific measure, our claim is to add edges, pertaining to opposite groups, in a realistic (*similarity measures*) or efficient (*communicability measures*) way.

In particular, adding edges through the evaluation of similarity measures should mimic the most likely evolution of the graph, with the only constraint of adding nodes belonging to opposite groups. Instead, adding edges through the evaluation of communicability measures should reduce the controversy the most.

However, a further step has been done in our analysis. Basically, the idea is to combine the features of similarity and communicability. We aim to create a measure which is both realistic and efficient at the same time. Realistic because it takes into account similarity between two nodes. Efficient because it "weights" pairs of nodes according to the their importance in terms of communicability.

A Hybrid Measure to Link Prediction

As consequence, we have combined all the similarity metrics introduced in this section with the betweenness index. Given the betweenness index, $B_{x,y} = s^B_{x,y}$; if we generally refer to one of the four similarity measures with: $S_{x,y} = s_{x,y}$, then we could determine our *Hybrid index* as:

$$h_{x,y} = B_{x,y} \times S_{x,y} \tag{4.17}$$

Equation 4.17 aims to capture both communicability and similarity. In particular, to higher values of both $B_{x,y}$ and $S_{x,y}$ correspond higher values of $h_{x,y}$. In this case, (x, y) should be a pair of nodes which share common neighbors (they are similar) and with high value of betweeness (they are central in terms of communicability for their communities).

To the extent of our knowledge, this is the first time, at least in the context of echo chambers, that this index has been used. In Chapter 5, we will evaluate the performance of all these measures on real data.

Chapter 5

Evaluations

In this chapter, we are going to present and discuss the results of the evaluation of different aspects connected to the proposed solution to controversy reduction. In particular, the chapter is organized as follow.

In Section 5.1, we will introduce and describe the datasets that we have used in our study.

Section 5.2 will be devoted to the presentation of the results connected to the efficiency of the *community detection* algorithms; in particular, on a specific example, we illustrate how the the *Kernighan-Lin* algorithm is "computationally inefficient" compared to *FluidC* (as had already been pointed out in Section 4.1 of Chapter 4) and, for this reason, the remaining evaluations with respect to the usage of *controversy measures* and *link prediction* algorithms will be performed on the communities obtained by the use of FluidC.

In Section 5.3, we will evaluate, in general, the effectiveness of the controversy measures employed in this work. In particular, as detailed in Section 4.2 of Chapter 4, the choice has fallen on the following controversy measures: *Random Walk Controversy* (RWC), *Embedding Controversy* (EC) and *Boundary Connectivity Controversy* (BCC).

Finally, in Section 5.4, we will evaluate the effectiveness of the *link prediction* algorithms considered in this work, which are based on the *similarity*, *communicability*, and *hybrid* measures that have been presented in Section 4.3 of Chapter 4. Their effectiveness will be measured by re-evaluating the above-mentioned controversy measures after that the link prediction algorithms are applied to the communities that have been identified by the FluidC algoritm.

The code and data that have been employed to develop and evaluate the solutions proposed in this thesis are public available.[1]

5.1 Data

For the purpose of our study, we have used a subset of the data collected and described by Garimella et al. (2015).[2] In Section 3.1 of Chapter 3, we explained how

[1] `https://github.com/Comollo/echo-chambers/`.

[2] The original dataset can be downloaded at the following link: `https://github.com/gvrkiran/controversy-detection/tree/master/networks/`.

these sets of data have been collected. In particular, the process is based on the collection of Twitter data for a sets of specific hashtags. Given a hashtag, it is possible to create a graph by using the set of *social interactions* among people who used it. This graph is called "conversation graph".

As previously stated, it is possible to use four different types of social interactions: *retweet*, *follow*, *content*, *hybrid content & retweet*. According to the authors, only "retweet" and "follow" are able to reveal controversial and non-controversial patterns. As a consequence, we use that part of the data related to the "follow graphs" for nine different hashtags.

Hashtag	|V|	|E|	Topic	Collection Period (2015)
#baltimoreriots	1441	28291	Riots in Baltimore after police kills a black man	Apr 28–30
#gunsense	1821	103840	Gun violence in U.S.	Jun 1–30
#netanyahuspeech	4292	297136	Netanyahu's speech at U.S. Congress	Mar 3–5
#ukraine	3383	84035	Ukraine conflict	Feb 27–Mar 2
#russia_march	1189	16471	Protests after death of Boris Nemtsov ("march")	Mar 1–2
#sxsw	4558	91356	SXSW conference	Mar 13–22
#ultralive	2113	16070	Ultra Music Festival	Mar 18–20
#germanwings	2111	7329	Germanwings flight crash	Mar 24–26
#1dfamheretostay	3151	20275	Last OneDirection concert	Mar 27–29

TABLE 5.1: Datasets used throughout the analysis. According to Garimella et al. (2015), the top five hashtags refer to controversial topics, while the bottom four ones refer to non-controversial topics.

Table 5.1 shows the selected set of *hashtags*. For each hashtag, we have reported the number of vertices ($|V|$) and the numbers of edges ($|E|$) present in the follow graph.[3] According to the authors, the top five hashtags refer to controversial topics, while the bottom four refer to non-controversial ones. In particular, as stated by the authors, in all the controversial topics, the debate leads to two distinct groups which represent two echo chambers. The data have been collected in 2015 and the last column in the table highlights the exact collection period.

The choice of using both controversial and non-controversial topics is motivated by the need to understand which of the adopted controversy measures is the best in terms of reveling the controversial phenomenon.

Therefore, for each of the nine hashtags present in the mentioned table, we did create a conversation graph by using the vertices and edges present in the follow graph.

5.2 Community Detection Evaluation

To evaluate the solutions for controversy reduction proposed in this thesis, we rely on the innovative *FluidC* community detection algorithm. As stated in the preface of this chapter, we do not employ the *Kernighan-Lin* algorithm due to its "computational inefficiency" compared to FluidC. The expensive computational cost of the

[3] These statistics refer to the data available in the original repository (https://github.com/gvrkiran/controversy-detection/tree/master/networks), at the time we carried out the analysis.

Kernighan-Lin algorithm is known in the literature (Schulz, 2016), however, in this section, we still wanted to make a simple experiment to evaluate the difference, in terms of efficiency, between the two community detection algorithms.

5.2.1 Computational Efficiency

During the analysis, we have reported that, as expected, FluidC is extremely fast w.r.t. the KL algorithm. Table 5.2 shows the computational efficiency of both the algorithms.

topic\algorithm	**Fluidc**	**Kernighan-Lin**
#russia_march	0.21 sec	345.63 sec
#sxsw	2.18 sec	113,476.54 sec

TABLE 5.2: In both the tests, *FluidC* has outperformed *Kernighan-Lin* in terms of computational efficiency.

In order to test the computational efficiency, we have applied both the community detection algorithms to the smallest and the biggest graph in terms of nodes. In fact, *#russia_march* and *#sxsw* contain the minimum and the maximum number of nodes, respectively. The results show how *FluidC* outperforms *Kernighan-Lin* in terms of efficiency. These results are in line with our expectations because FluidC is a community detection algorithm based on label propagation, which is known as one of the most efficient methods. Moreover, as previously stated, FluidC is characterized by some important properties for a community detection algorithms. In particular, the possibility to define the number of communities to seek out, together with the ability to avoid the creation of giant communities, and, hence, of imbalanced communities.

5.2.2 Qualitative Pre-evaluation and Quantitative Evaluation

Before partitioning the graphs using FluidC, we have performed a qualitative pre-evaluation of the graphs with the purpose of displaying their patterns. Our expectation is to reveal the presence of two communities for the controversial topics. After having partitioned the graphs using FluidC, we perform a quantitative evaluation of the partitions. In this case, with the purpose of assessing the "quality" of the partitions created by FluidC.

Qualitative Pre-evaluation

The qualitative evaluation has been performed by employing the ForceAtlas2 algorithm (Jacomy et al., 2011) implemented in Gephi. As the paper states, "ForceAtlas2 is a force directed layout: it simulates a physical system in order to spatialize a network. Nodes repulse each other like charged particles, while edges attract their nodes, like springs. These forces create a movement that converges to a balanced state. This final configuration is expected to help the interpretation of the data".

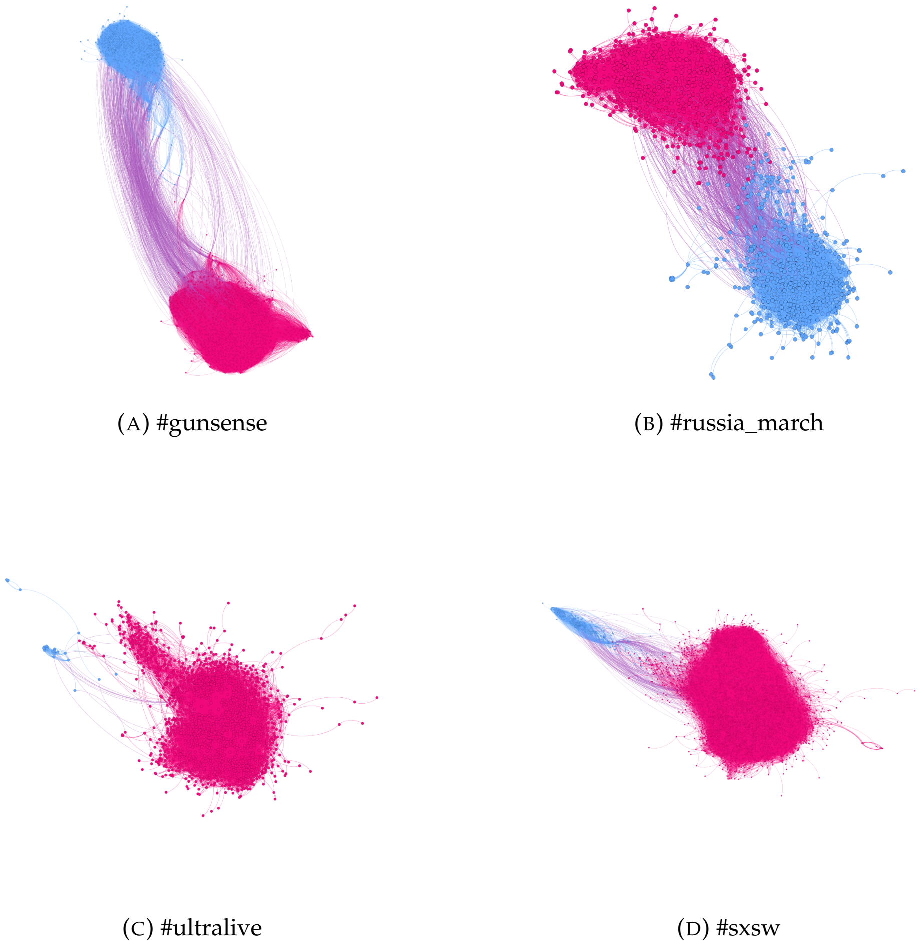
Moreover, Noack (2009) has shown that force-directed layouts optimize *modularity* helping to reveal communities. Therefore, the use of Gephi helps the interpretation of the graphs by revealing their real patterns. As a consequence, we rely on this feature in order to understand the extent and the nature of the communities in our dataset.

In particular, the qualitative pre-evaluation consists in the visualization of all the nine graphs using Gephi. The purpose of this process is double:

1. We want to reveal the presence of communities in the controversial graphs;

2. We want to show the absence of communities in the non-controversial graphs.

In fact, it will not make sense to partition the graphs into two communities even when these communities are not present. Therefore, for both controversial and non-controversial topics, we have visualized their graph using the mentioned layout algorithm (ForceAtlas2).



(A) #gunsense

(B) #russia_march

(C) #ultralive

(D) #sxsw

FIGURE 5.1: Illustration of four different graphs. The colors, red and blue, correspond to visualization patterns of Gephi. This type of visualization has been created for all the graphs.

Figure 5.1 displays some of the graphs created with Gephi. The top two images shows the graphs of two controversial topics (#*gunsense* and #*russia_march*), while the bottom two ones refer to two non-controverial topics (#*ultralive* and #*sxsw*).

In the figure, both the controversial graphs (Figure 5.1a and Figure 5.1b) reveal two well-separated communities. Instead, the non-controversial graphs do not display any patterns in terms of groups. As previously stated, we have performed this evaluation for all the nine graphs. As result, not all the declared controversial topics have shown the presence of two communities. In fact, in two cases, the graphs lead to patterns similar to non-controversial situations. Therefore, at the end of this evaluation process, we reject the hypothesis of two communities for two out of the five declared controversial graphs. These graphs refer to the #*baltimoreidiots* and #*ukraine* hashtags. As consequence, these two graphs have not been used for further evaluations.

Nonetheless, Figure 5.1 shows another important aspect. From the figure, we may conclude that the size of the communities depends on the graph itself. In fact, despite both the controversial graphs show the presence of two distinct groups, Figure 5.1a and Figure 5.1b have different type of communities. In the first case, we denote a remarkable distinction between the two communities. Moreover, one community seems to be bigger than the other. Instead, in the second case, the two communities are closer to each others and seem to have a comparable number of nodes. Therefore, avoiding a graph partitioning algorithm based on a balancing constraint, such as Kernighan-Lin, in favour of FluidC, which instead is more flexible, seems to have been a proper choice. We remind that a balancing constraint forces the partitions to equal size, situation that is not commonly present in real cases.

Quantitative Evaluation

The quantitative evaluation consists in measuring the "quality" of the partitions obtained by FluidC (over the three controversial graphs related to the #*gunsense*, #*russia_march* and #*netanyahuspeech* hashtags) by using a specific metric called *coverage*.[4] Coverage is the ratio of the number of intra-community edges by the total number of edges. Ideally, good partitions leads to value of the coverage close to 1 because all edges of the graph fall within clusters. Table 5.3 displays the coverage obtained for all the controversial graphs. The results suggest that the two partitions created by FluidC are satisfactory, having values close to 1. Particularly accurate is the partitioning of graph related to #*gunsense*; in this case, it is likely that almost all the edges of the graph fall within the two partitions. This results supports the fact that FluidC has correctly identified the two communities present in the conversation graphs.

[4] Given the absence of communities in the non-controversial graphs and in the two discarded hashtags (#*baltimoreidiots* and #*ukraine*), we did omit to partition these graphs.

Graph	Coverage
#gunsense	0.9927
#netanyahuspeech	0.9766
#russia_march	0.9621

TABLE 5.3: Coverage metrics for all graphs that have been proven to be controversial. The measure has been calculated using the two partitions of each graph.

5.3 Controversy Measures Evaluation

In this section, we evaluate the controversy measures adopted in our study. As previously stated, for the purpose of this work, we choose three metrics in order to understand the level of controversy within the graphs. These three measures are: *Random Walk Controversy* (RWC), *Embedding Controversy* (EC) and *Boundary Connectivity Controversy* (BCC). The first two are the more recent metrics presented in the literature, while the last one has a long history and can bed considered as a sort of baseline.

In this section, we will not evaluate the controversy measures before and after the application of the link prediction technique. This will be done in the following section. Here, we will asses which of the mentioned metrics, according to our results, seems to be better in terms of ability at differentiating controversial and not controversial topics, from a general point of view.

Assessing which of the controversy measures better differentiate controversial and non-controversial situations is pivotal for the next stage, i.e., reducing the controversy through the link prediction phase. In fact, if our goal is to reduce the controversy level within a social network, then we need to understand which of the adopted controversy metrics we should rely on the most.

Therefore, after having partitioned the three controversial graphs and the four non-controversial graphs (with FluidC), we measure the level of controversy between the two communities through the mentioned metrics. Then, we evaluate RWC, EC and BCC using their average value, differentiating when the topic is controversial or not.

Figure 5.2 displays the results of the evaluation. For each controversy measure, the green dots correspond to the their average value measured considering only the three controversial graphs (*#gunsense*, *#russia_march* and *#netanyahuspeech*). Conversely, the blue dots represents their average value measured considering only the four declared non-controversial topics (*#germawings*, *#1dfamheretostay*, *#ultralive* and *#sxsw*).

In the figure, all the measures seems to be good at differentiating controversial and non-controversial social networks. As the figure suggests, the values of RWK, EC and BCC are higher when we face polarized graphs (green dots), while they are

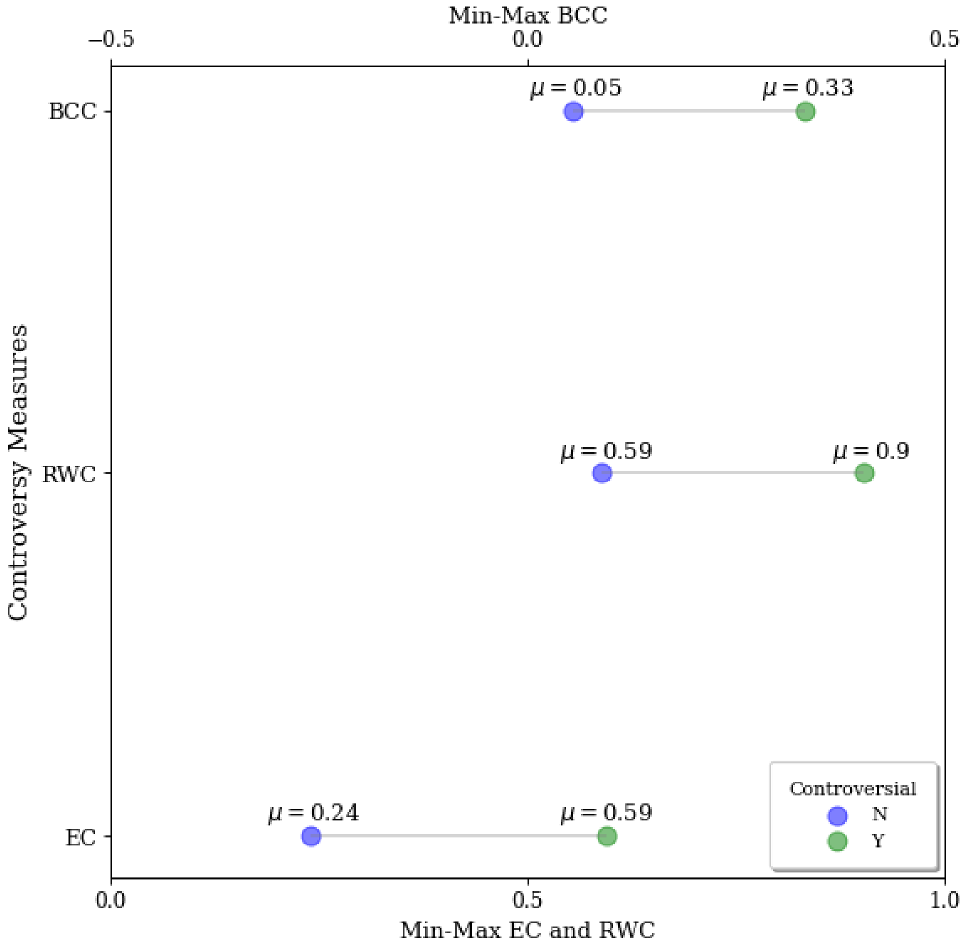


FIGURE 5.2: Evaluation of the metrics: *BCC*, which refers to *boundary connectivity*, seems better at differentiating controversial and non-controversial topics.

lower in case of not polarized social networks (blue dots). According to our results,[5] on average, we may conclude that RWK, EC and BCC differentiate controversial and non-controversial social networks.

However, on average, RWK seems to overestimate the level of controversy, especially in case of non-controversial situations: the blue dot indicates an average value close to 0.6 even though the topics were not controversial. This value is particular significant considering that RWK takes values in $[0, 1]$ and 0 refers to absence of controversy.

On the other side, EC, which takes values in the same range of RWK, seems to underestimate the level of controversy in case of polarized social networks: the green dot which refers to controversy topics, indicates an average value close to 0.5.

BCC, instead, lies in the range $(+0.5, -0.5)$. According to the original paper (Guerra et al., 2013), a BCC value below 0 indicates not only lack of polarization, but also that nodes in the boundary are more likely to connect to the other side;

[5] Due to the low number of samples available, we could not test these results using statistical tests.

a BCC value close to 0 means a neutral situation, because all boundary nodes established the same number of connections to internal nodes and to nodes from the alternate community; a BCC value greater than 0 indicates that, on average, nodes on the boundary tend to connect to internal nodes rather than to nodes from the other group, indicating that antagonism is likely to be present.

Therefore, on average, BCC shows a neutral situation for the non-controversial topics, since the average value (blue dot) is almost 0; while, for controversial topics, BCC catches the polarization phenomenon because the average value (green dot) is equal to 0.33 which is pretty close to its maximum (0.5).

As a consequence, we conclude that, according to our results, the controversy measure that we trust the most is BCC. Hence, in the next section, when we are going to evaluate the measures to reduce controversy, we highly rely on this metric for the comparison of *similarity* and *communicability* measures implemented in the link prediction phase.

5.4 Link Prediction Evaluation

In this section, we outline the most significant outcomes of our analysis. In particular, this section has to be considered the core of our experimental evaluation. The research goal of this thesis was the development of a pipeline aimed at reducing the controversy in a social network. Here, we will demonstrate that, even with the addiction of a minimum number of new edges, it is possible to reduce the controversy level among communities. As previously stated, we have applied different link prediction algorithms that are based on different measures for ranking pairs of nodes among which edges can be added. We refer to these measures as:

- *Similarity Measures*, whether they aim to catch the degree of similarity between two nodes. Here the concept of similarity is measured by using the concept of *neighbors*;
- *Communicability Measures*, whether their purpose is to improve *communication* among nodes.

The *Jaccard Coefficient*, *Adamic-Adar Index*, *Preferential Attachment* and *Resource Allocation Index* are the considered similarity measures, while *Betweenness Index* and *Effective Size Index* are employed as communicability measures. The first group concerns well-known measures which are a sort of baseline in the literature, while the second one focuses on novel approaches that have been proposed for the first time in this work.

However, in Section 4.3 of Chapter 4, we have also described another novel strategy (*hybrid*) which is aimed at combining communicability measures in particular, the Betweenness Index with the similarity measures.

We remind that, for the link prediction problem, it has only been considered pairs of nodes which nodes belong to opposite communities. This has been fully motivated in Chapter 4, Section 4.3.

5.4.1 Similarity and Communicability Measures: Results

In Section 5.3 of this chapter, we have demonstrated that BCC is the controversy measure which, according to our results, better discriminates controversial and non-controversial networks. As a consequence, here, we will particularly focus on this metric to compare our outcomes. The results referred to the use of the other controversy measures will be illustrated in Appendix A.

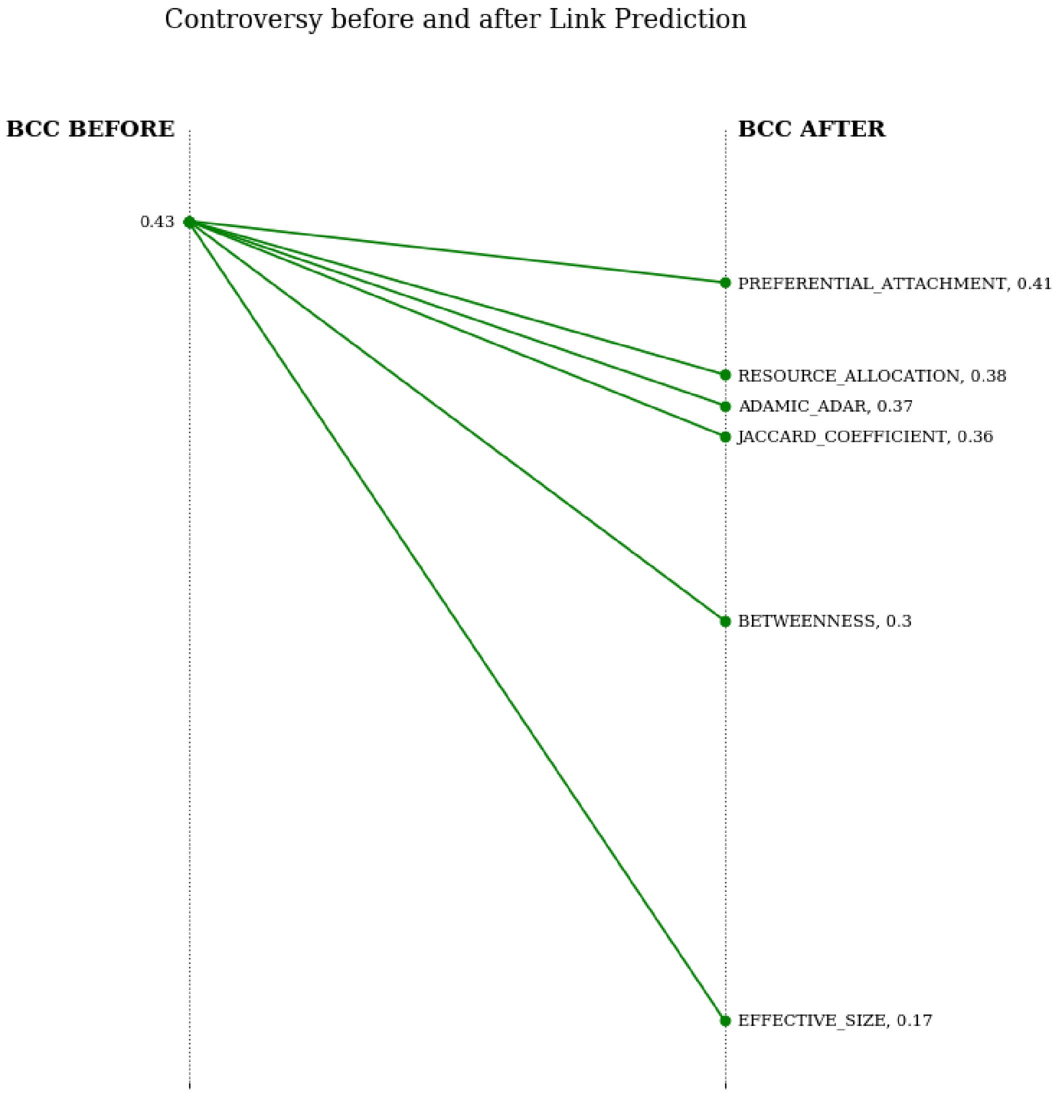


FIGURE 5.3: BCC values before and after link prediction. These values have been calculated using *#gunsense* as graph.

Figure 5.3 shows the values of BCC before and after link prediction. These values have been calculated using the graph of the hashtag *#gunsense*. In fact, as Figure 5.1a suggests, this network is highly polarized into two groups. Therefore, this graph seems to be a good candidate for our analysis. The measurese for all the graphs are reported in Appendix A. On the left of the figure, there is the original value of boundary connectivity (BCC), i.e., the value measured using the original graph *before* added edges. BCC fully catches the controversial situation of the graph: its value before link prediction (0.43) is very close to its maximum (0.5).

On the right, it is possible to see BCC's values *after* the addition of 1,000 edges, which corresponds to only the 0.0006% of new possible edges between the two communities. The graph shows also the name of the measure which has been used in the link prediction process. If we compared all the measures, for this specific graph, we may conclude that our novel approaches, based on communicability metrics, lead to lower values of controversy. In particular, the use of *betweenness* and *effective size* leads to the lowest (0.17) and the second lowest (0.3) value of BCC. The result is even more remarkable because the link prediction process has involved simply the 0.0006% of new possible edges between the two communities.

Due to the low number of controversial samples (3), we can not statistically test our results. However, in Figure 5.4, we have reproduced the same visualization, but using the average values of BCC along three controversial networks, before and after link prediction. This, in fact, should give a better representation of the real controversy reduction produced by the link prediction of new edges using different measures.

In particular, Figure 5.4 shows the average BCC reduction produced by both the similarity and communicability measures. As the figure suggests, adding edges through the *effective size* still generates the highest BCC reduction, i.e., 0.20 on average (from 0.33 to 0.13). On average, the second highest BCC reduction is produced by *betweenness*, i.e., $0.33 - 0.22 = 0.11$. Therefore, the consideration of the proposed communicability measures, on average, seem to perform better. Indeed, all the similarity measures have almost the same, less effective, behaviour. In particular, on average, *Adamic-Adar* performs better than *Preferential Attachment*, *Jaccard Coefficient* and *Resource Allocation*. However, they generate similar reductions, *Adamic-Adar* minus 0.6, *Resource Allocation* and *Jaccard Coefficient* minus 0.5, while *Preferential Attachment* minus 0.4, which is the lowest reduction.

We may conclude that, when BCC is used as controversy measure, adding new edges through the proposed link prediction algorithms based on communication measures reduce the controversy the most. This result is even more symbolic due to the average number of edges added over the three controversial graphs (*#gunsense*, *#russia_march* and *#netanyahuspeech*). In this case, on average, merely the 0.0007% of edges has been added.

5.4.2 Hybrid Measures: Results

The same procedure has been used to check the results of the proposed *hybrid* measures. In particular, we have calculated four different hybrid measures as result of

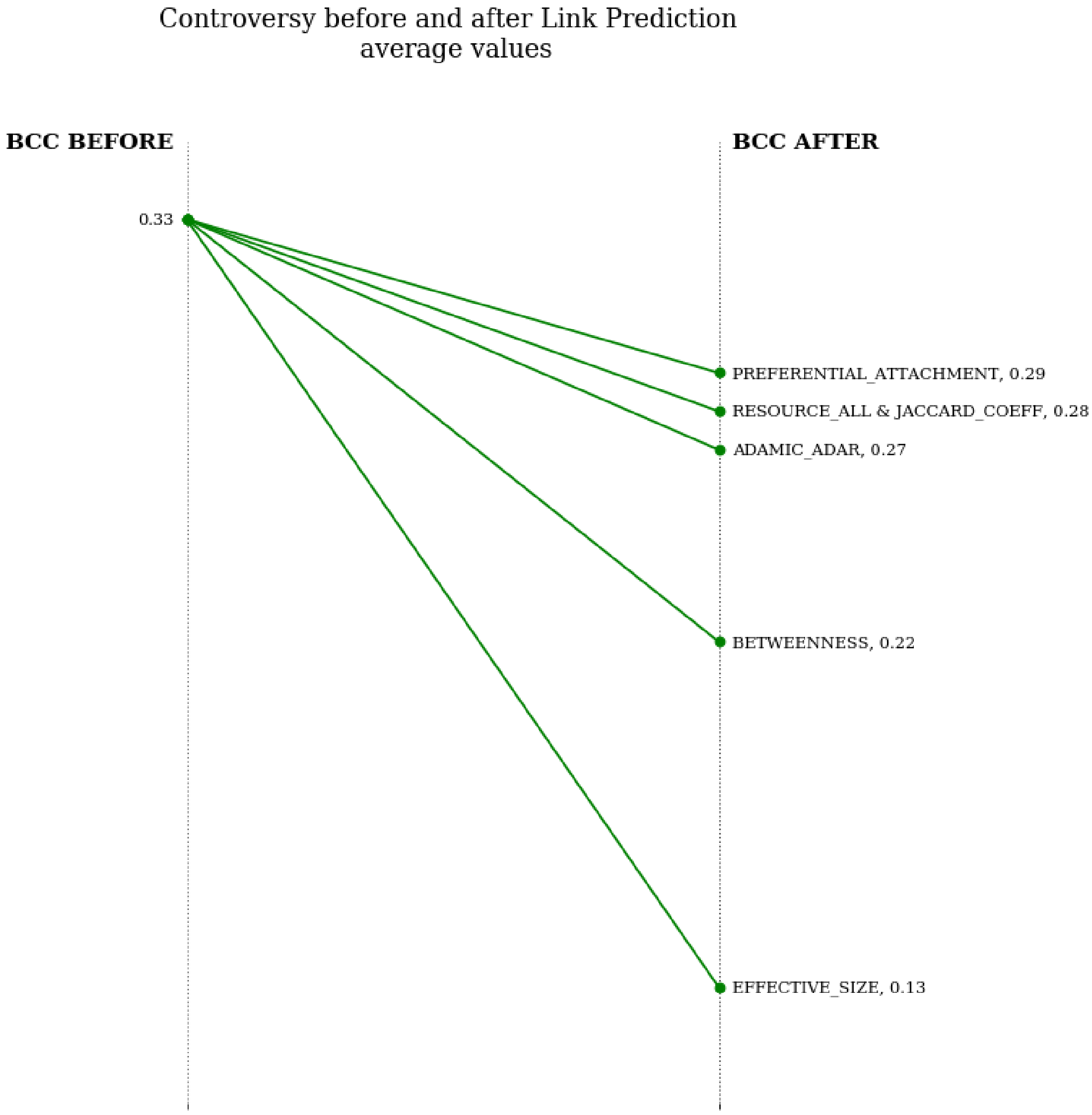


FIGURE 5.4: Average BCC values before and after link prediction. These values correspond to the average values of BCC, before and after link prediction, calculated using all the controversial networks (*#gunsense*, *#russia_march* and *#netanyahuspeech*).

the combination of the *Betweenness Index* with *Jaccard Coefficient*, *Adamic-Adar Index*, *Preferential Attachment* and *Resource Allocation Index*, respectively. The idea was to combine the benefits in terms of controversy reduction of the communicability measure with the accuracy of the similarity measures.

Figure 5.5 shows the values of BCC before and after link prediction. Again, these values have been calculated, in the illustrated example, using the graph of the *#gunsense* hashtag. Here, we do not focus on the lowest values, but more on possible improvements of some of the link prediction techniques based on similarity measures. In fact, as the figure suggests, pure Betweenness Index and Effective Size Index still guarantee better reductions. Comparing this figure with Figure 5.3, we do not see any particular improvement. The only marginal improvement is performed by the combination of *Preferential Attachment* with *Betweenness Index*. In fact, in Figure 5.3, pure link prediction based on preferential attachment lead to a BCC value of 0.41, while the combination of this similarity measure with the communicability measure brings the BCC value to 0.39. A small extra reduction of 0.02. However, all the other similarity measures combined with the Betweenness Index get a worse performance. Again, this results have been produced by adding simply 1000 edges, which corresponds to the 0.0006% of new edges among all the possible ones between the two communities.

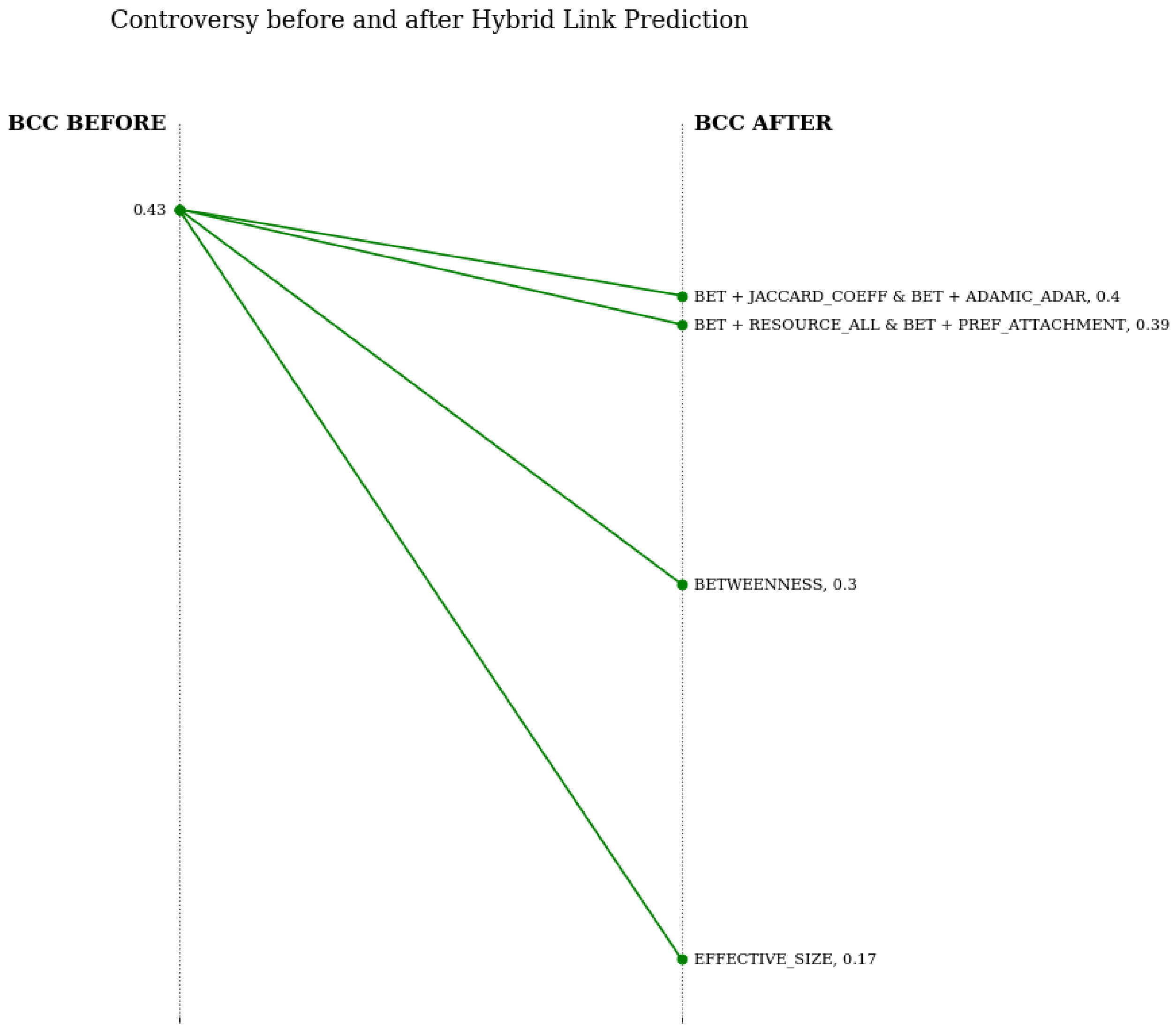


FIGURE 5.5: BCC values before and after hybrid link prediction. These values have been calculated using #gunsense as graph.

In order to better understand if there are any improvement in the hybrid similarity measures respect to their pure version, we compare the average reduction in the BCC along all the controversial networks (#*gunsense*, #*russia_march* and #*netanyahuspeech*) for both their standard and hybrid version. Table 5.4 shows the results of this comparison. For each similarity measure, we calculated the average BCC values using both hybrid and standard version. The first column in the table refers to the average value of BCC (0.33) among the three mentioned controversial graphs. The second and the third columns refer to the average BCC values calculated after the link prediction using respectively the hybrid and the standard version of a given similarity measure.

As the table suggests, on average, we denote slightly improvement for *Jaccard Coefficient* and *Preferential Attachment*, while for *Resource Allocation Index* and *Adamic-Adar Index*, we got the same results. Although, we can not statistically test these results, Table 5.4 supports the idea that combining the *Betweenness Index* with the similarity measures does not improve their performance in terms of controversy reduction.

	BCC Original Graph	BCC Hybrid Version	BCC Standard Version
Resource Allocation Index	$\mu = 0.33$	$\mu = 0.28$	$\mu = 0.28$
Jaccard Coefficient	$\mu = 0.33$	$\mu = 0.27$	$\mu = 0.28$
Adamic-Adar Index	$\mu = 0.33$	$\mu = 0.27$	$\mu = 0.27$
Preferential Attachment	$\mu = 0.33$	$\mu = 0.27$	$\mu = 0.29$

TABLE 5.4: Comparison between the Hybrid Version and the Standard Version of the similarity measures using BCC as benchmark. In each cell, the number refers to the average BCC value calculated using all the controversial networks.

5.4.3 Evaluation of the Algorithms: Looking at the Whole Picture

In the previous sections, we have evaluated the measures implemented in the link prediction phase with respect to one controversy measure (*Boundary Connectivity Controversy*). For completeness of the work, in this section we provide the results over all the controversy measures. Despite the "lesser reliability" of *Random Walk Controversy* and *Embedding Controversy*, it is worth displaying the performance of both *Communicability* and *Similarity* measures for all the controversy measures. In particular, we want to understand whether or not one of the measures implemented in the link prediction is *always* better than the others.

Previously, we have demonstrated that the *hybrid* version of the similarity measures does not improve their *standard* form. Therefore, the mentioned comparison of both communicability and similarity measures have been done using the standard version of the latter. Moreover, we compared the cited measures by using their average performance over the three controversial networks, which refer to the *#gunsense*, *#russia_march* and *#netanyahuspeech* hashtags.

Figure 5.6 exhibits the conclusive evaluation of the link prediction measures over the three controversy metrics. On the y-axis of the figure, we find the *average controversy reduction* produced, over the three cited graphs, by both communicability and similarity measures. The x-axis shows the *controversy measures* taken into account in our thesis; the three groups of bars represent the different link prediction algorithms and connected measures. By *looking at the whole picture*, we may establish which measure, on average, has performed better, given a specific controversy measure.

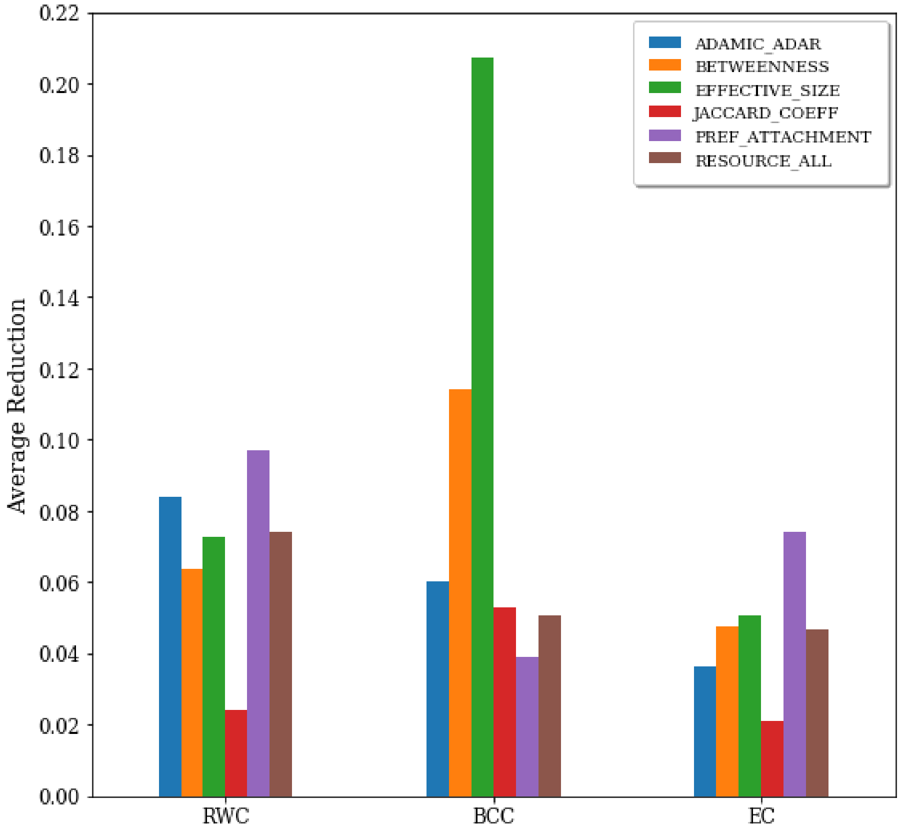


FIGURE 5.6: Comparison of the Link Prediction Measures over the three controversy measures.

In particular, if the same colored bar is always the highest over RWC, BCC and EC, then one of the measure has always outperformed the others. However, this is not the case in our situation. Rather, we denote that, on average:

- *Preferential Attachment* (purple bar) generate the highest controversy reduction if we take into account *Random Walk Controversy* and *Embedding Controversy* as metrics. However, it is the worst, if we consider *Boundary Connectivity Controversy* as controversy benchmark.
- *Effective Size Index* (green bar) outperform the other measures, when we take into account BCC. However, in all the other cases, this measure still generates good results, In fact, it produces the second highest reduction in case of EC and the fourth in case of RWC.

Therefore, on average, we cannot conclude that one of link prediction measures outperforms the others. Rather, we denote different behaviours over the three controversy measures. However, the accuracy of BCC, tested in Section 5.2, still allows to strongly rely on those link prediction metrics which generate the highest reduction for it.

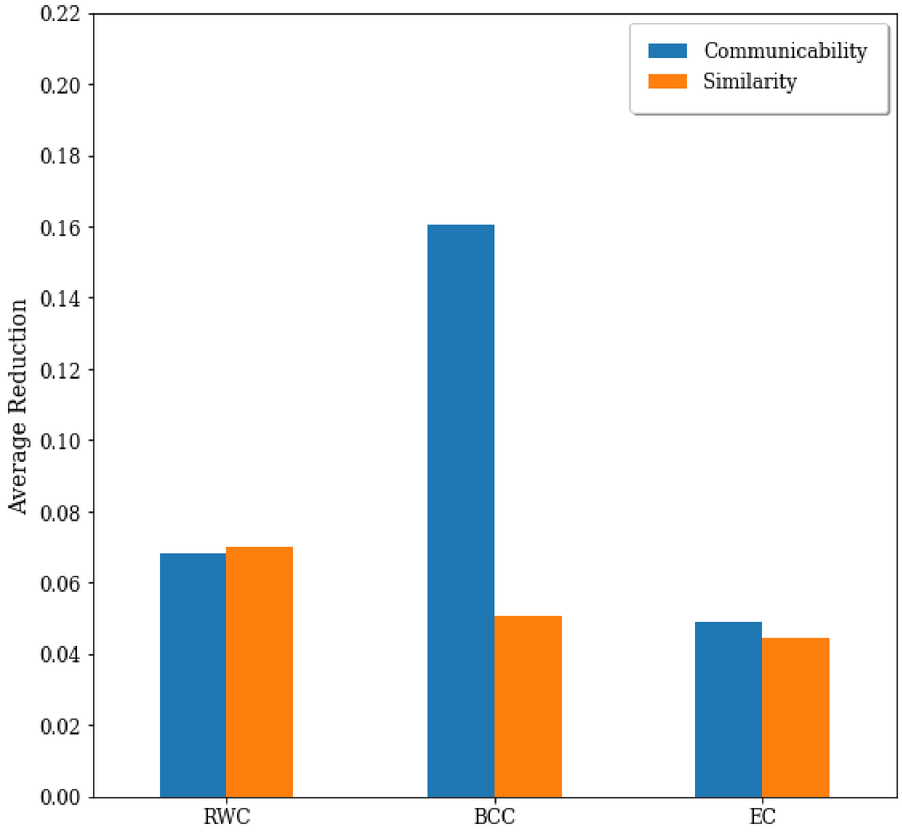


FIGURE 5.7: Comparison of the Link Prediction Categories over the three controversy metrics.

However, we also performed a further analysis. In particular, we grouped the link prediction measures according to their nature: *Communicability* or *Similarity*. The idea is to inspect the average performance of these two categories over the three controversy measures. Figure 5.7 shows the results of this comparison. We follow the same strategy of before in order to display the average controversy reduction of both communicability and similarity measures. The figure shows that, on average, it is quite difficult to establish which category generates the highest reduction for RWC and EC. For EC, the communicability measures are slightly better, while for RWC, it is the opposite, even if the results are quite comparable. However, on average, the two *Communicability* measures outperforms the four *Similarity* measures, if we take into account BCC.

For completeness of the results, all these assumptions should be statistically tested in future works. Here, we are limited by the low number of samples available.

Chapter 6

Conclusions

In this thesis, we developed a pipeline in order to identify, quantify and reduce controversy among echo chambers. This pipeline is based on different techniques that mainly concern the field of Social Network Analysis. In particular, the approach relies on three methodologies: *Community Detection* algorithms, *Controversy Measures* based on graph topology and *Link Prediction* methods. The pipeline has been tested using Twitter data, and we obtained positive results in terms of its effectiveness in (i) identifying echo chambers in a social network, (ii) quantifying the level of controversy and (iii) reducing the controversy level among echo chambers. We therefore come back to the questions stated in Chapter 1 and tackle them individually in the light of our results.

6.1 RQ1: Can we focus on graph theory and Social Network Analysis in order to evaluate and reduce controversy?

In Chapter 5, we demonstrated the possibility of reduce the controversy level in a social network by using some of the state-of-the-art graph techniques among those present in the literature.

Firstly, we confirmed that graph partitioning is a suitable and effective methodology to discover latent cohesive structures attributable to echo chambers. In particular, in our thesis, we used *FluidC*, which is a quite innovative algorithm based on label propagation. The partitions generated by FluidC have been tested using a proper quality measure called "coverage". The partitions were satisfactory, having values of coverage close to its maximum (1).

Secondly, we confirmed that all the employed controversy measures (*Random Walk Controversy*, *Embedding Controversy* and *Boundary Connectivity Controversy*) are good at differentiating controversial and non-controversial social networks. However, Boundary Connectivity, which is based on the concepts of internal and boundary nodes, seems better than the others.

Finally, our results confirmed the theory that connecting opposite communities reduces controversy in a social network. In particular, we considered only edges *between* communities, rather than edges *within* communities, in the link prediction phase. We ranked pairs of nodes according to two type of measures: (i) *Similarity* and (ii) *Communicability*. Both families of metrics cut down the overall level of controversy between two echo chambers.

Therefore, with respect to the first research question, we positively answer that *Social Network Analysis* may play a pivotal role in the reduction of controversy on social media. Moreover, the use of SNA allowed to exploit only the graph topology (and, in this case, not the content diffused) and build a pipeline which is domain- and language-independent.

6.2 RQ2: Can we connect distinct polarized communities by acting on the properties of the social network taken into account?

To reply this question, in Chapter 4, we entirely devoted Section 4.3 to the presentation of four well-known *Similarity Measures* and two novel *Communicability Measures*. The foster are based on the concept of node's *neighbors*, while the latter on the properties of *Betweenness* and *Structural Holes*.

In Chapter 5, we proved that, despite the presence of two distinct communities, there exist nodes, pertaining to contradictory communities, which share common features. In this case the feature is represented by the number of friends or, in general, one-hop distance members of the network (neighbors). These nodes, likely nodes in the boundary of the communities, are good candidates for a link prediction phase. Moreover, we showed that, the addition of edges through these similarity measures does reduce the overall level of controversy among echo chambers. In particular, among those chosen in this thesis, *Preferential Attachment* exhibited better results (Chapter 5, Figure 5.6).

In the same chapter, we demonstrated that the connection of nodes, important for a community in terms of "communicability", promote information diffusion and reduce controversy. Particular significant has been the reduction, generated by the *Communicability Measure* based on the *Effective Size*, when we took *Boundary Connectivity* as controversy benchmark. In fact, this measure, which captures phenomena related to *Structural Holes*, outperforms all the other measures in the reduction of the *Boundary Connectivity Controversy* (Chapter 5, Figure 5.6). The result is even more symbolic, considering that, on average, only a marginal number of new edges have been added.

We concluded that using graph properties, such as node's neighbors or node's importance in terms of communicability, promote satisfactory results in the field of controversy reduction.

6.3 Limitations and Future Work

In this thesis, we encountered two main conditions which limit the generalization of our results. Firstly, the number of controversial social networks tested (i.e., the three networks based on the following hashtags: *#gunsense*, *#russia_march* and *#netanyahuspeech*) was actually not enough for any statistical hypothesis tests.

Secondly, the approach proposed by this thesis merely considered social networks with two contradictory communities. Even if we stated that, in most of the cases, real echo chambers appear when only two different polarized communities are present, future works could investigate if there are situations where multiple polarized communities can exist in social networks, and extend the problem to situations with more than two echo chambers.

Moreover, we want to discuss an important consideration. The *Communicability Measures*, introduced for the first time in this thesis, have been defined and tested with respect to the aim of reducing controversy. One of their limitation is the fact that, in real cases, it should be necessary to find strategies that effectively can lead to the connection of nodes with the highest communicability values. We tried to overcome this limitation by introducing *hybrid* measures, with the intent of combining the properties of *similarity* and *communicability*. Future research, however, may consider other solutions for "a more practical" implementation of our *Communicability Measures*.

Finally, a further development could be to propose a hybrid approach that uses both the structural information of the graph that represents the social networks considered, and the contents that are disseminated through the social networks themselves. This way you could actually consider the concepts of content polarity and more details about the topics discussed.

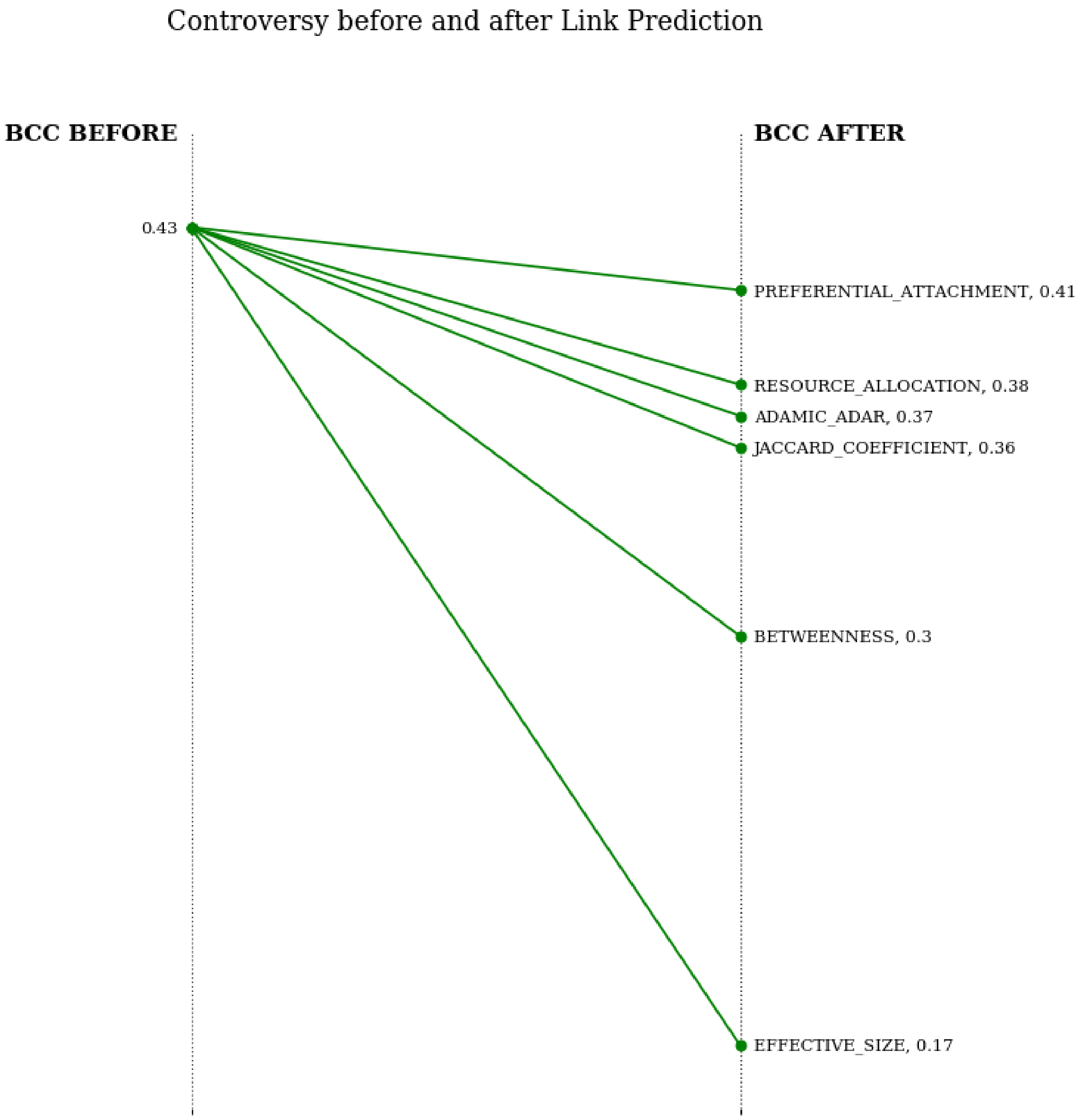
Appendix A

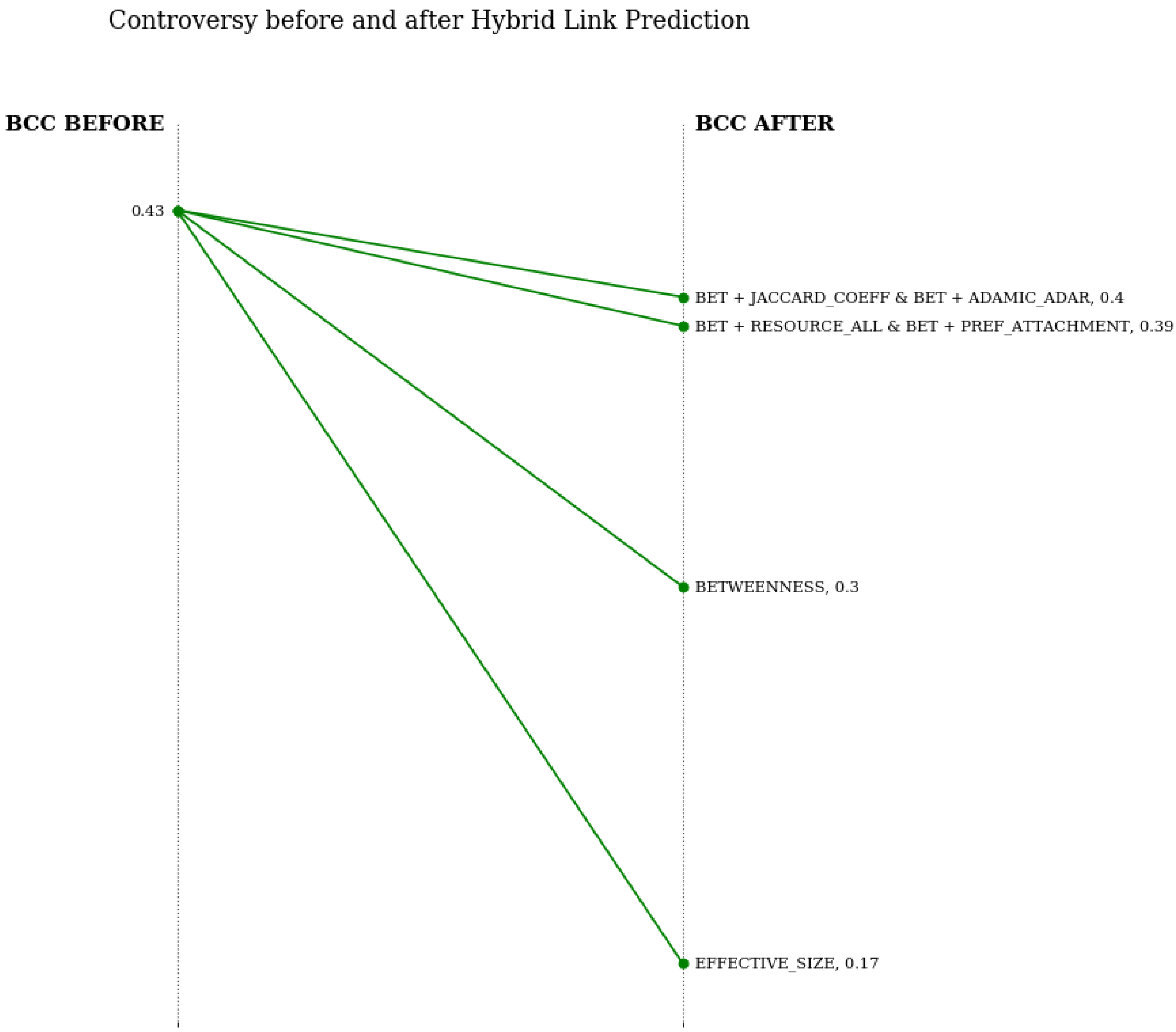
Controversy Reduction: Evaluation of Controversy Measures and Link Prediction over Controversial Graphs

For completeness of our analysis, in this appendix, we provide the results of all the measures employed in the link prediction phase. The results are divided according to the three controversial networks (#*gunsense*, #*russia_march* and #*netanyahuspeech*) and the three controversy measures (*Random Walk Controversy*, *Embedding Controversy* and *Boundary Connectivity Controversy*).

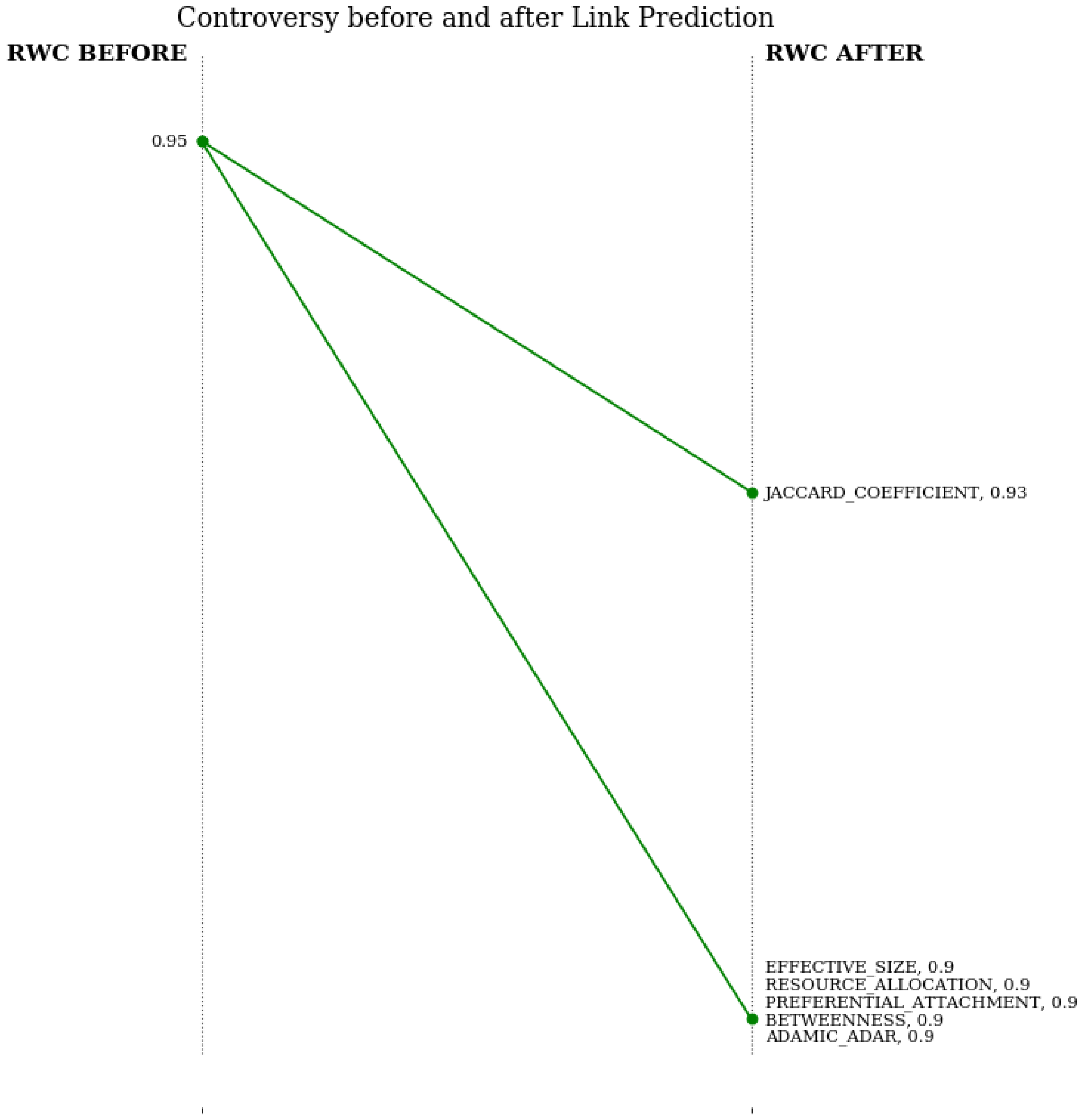
A.1 Hashtag #*gunsense*: Results

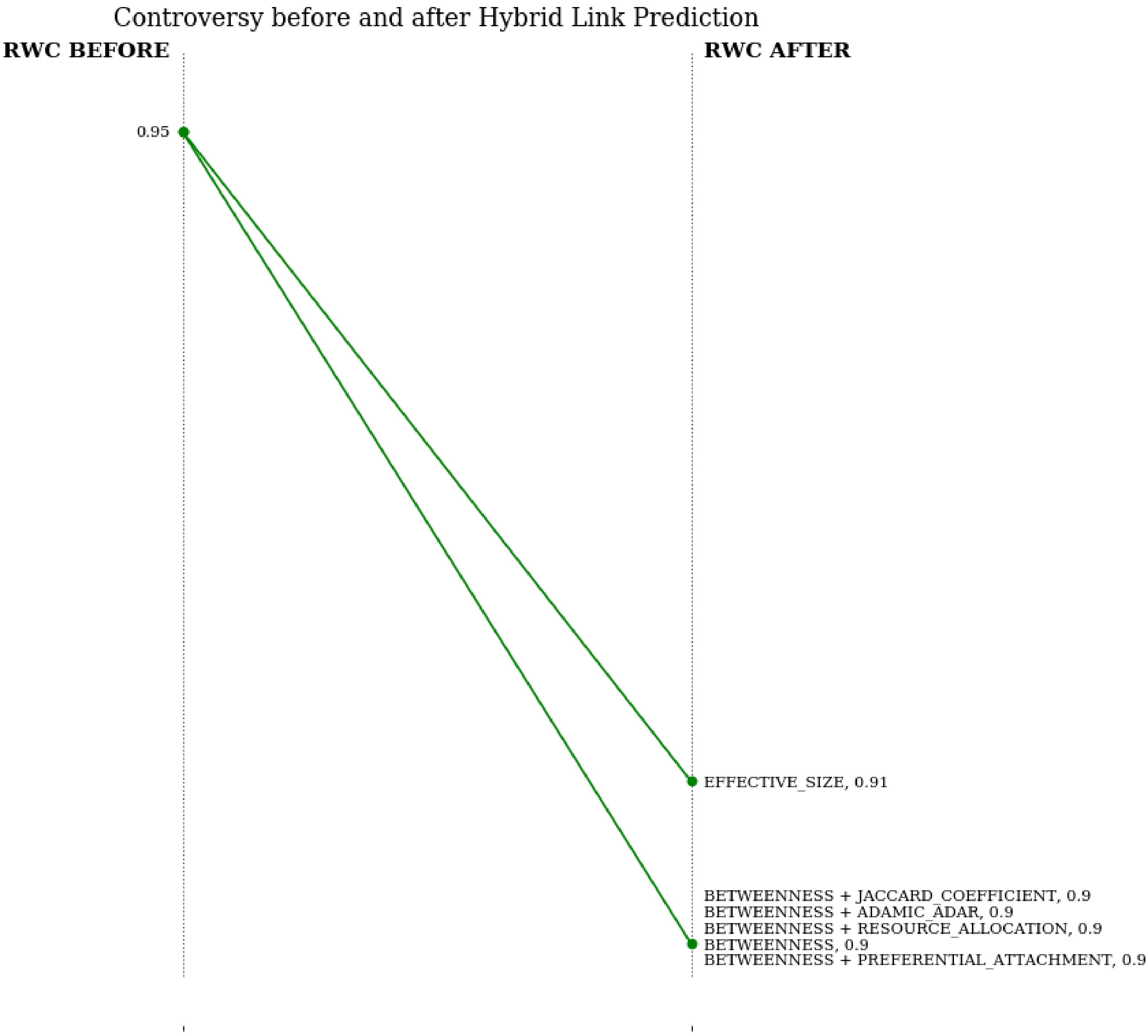
A.1.1 Reduction of Boundary Connectivity Controversy



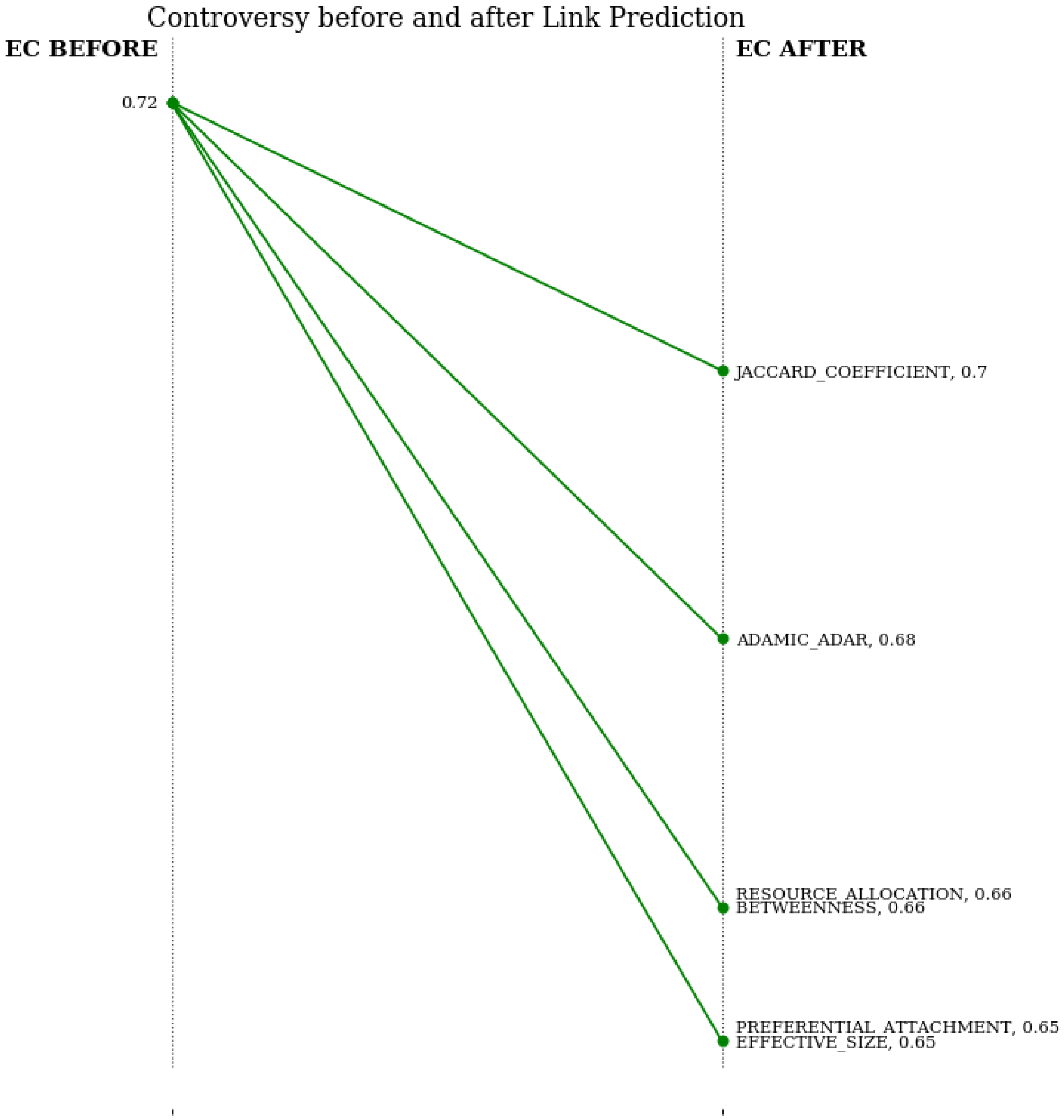

Controversy before and after Hybrid Link Prediction
BCC BEFORE
BCC AFTER
0.43
BET + JACCARD_COEFF & BET + ADAMIC_ADAR, 0.4
BET + RESOURCE_ALL & BET + PREF_ATTACHMENT, 0.39
BETWEENNESS, 0.3
EFFECTIVE_SIZE, 0.17

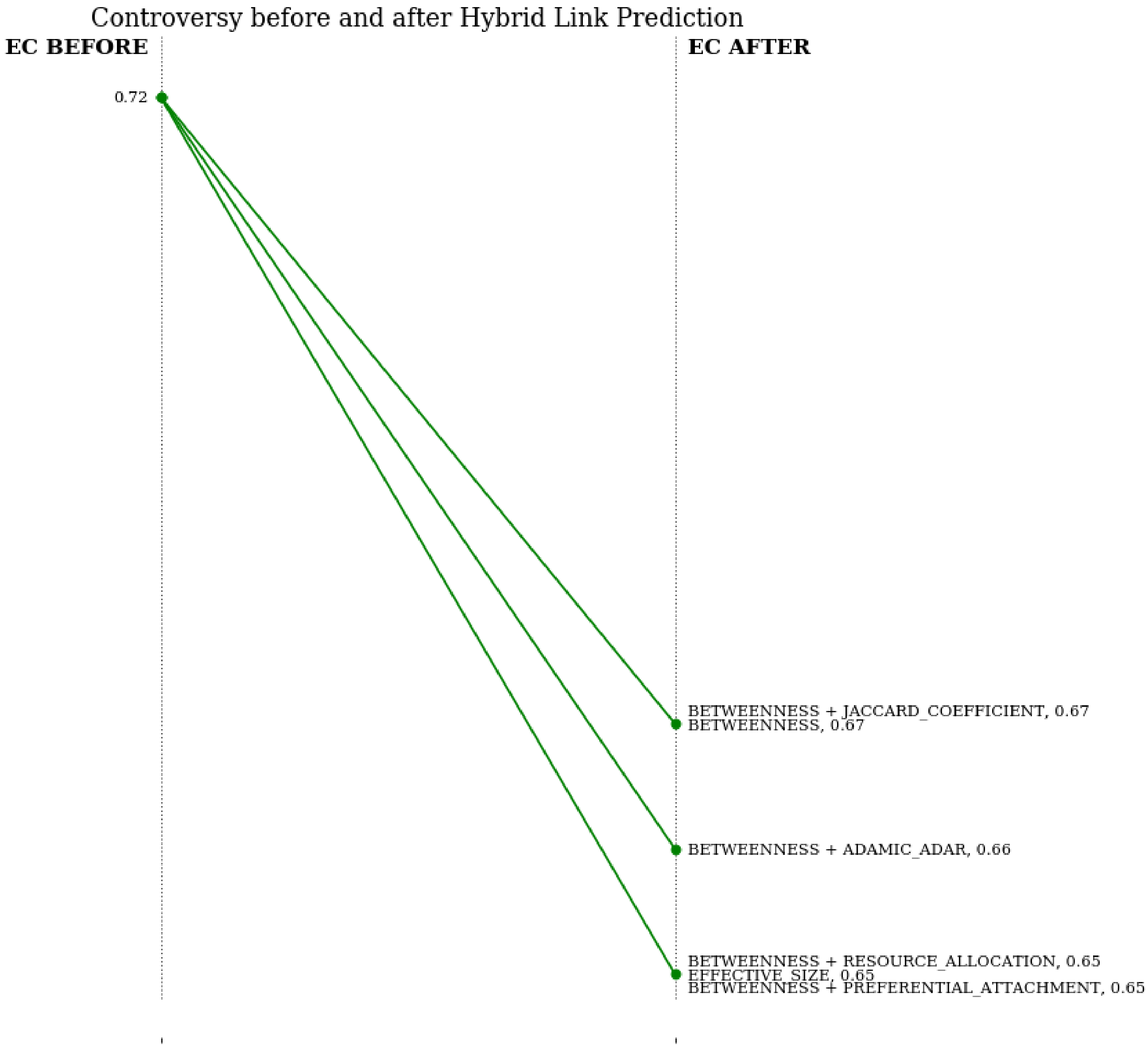
A.1.2 Reduction of Random Walk Controversy




Controversy before and after Hybrid Link Prediction
RWC BEFORE
RWC AFTER
0.95
EFFECTIVE_SIZE, 0.91
BETWEENNESS + JACCARD_COEFFICIENT, 0.9
BETWEENNESS + ADAMIC_ADAR, 0.9
BETWEENNESS + RESOURCE_ALLOCATION, 0.9
BETWEENNESS, 0.9
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.9

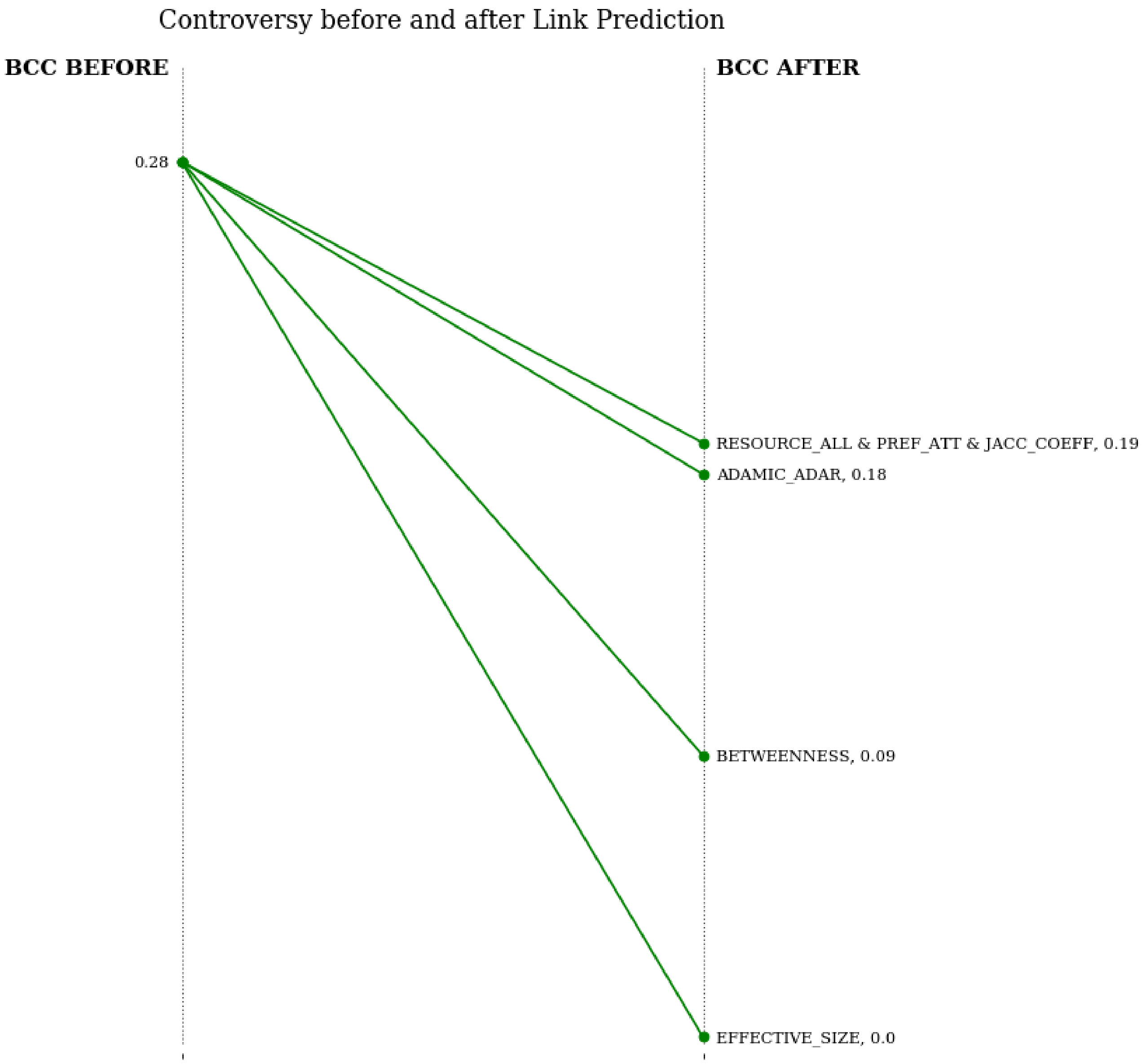
A.1.3 Reduction of Embedding Controversy

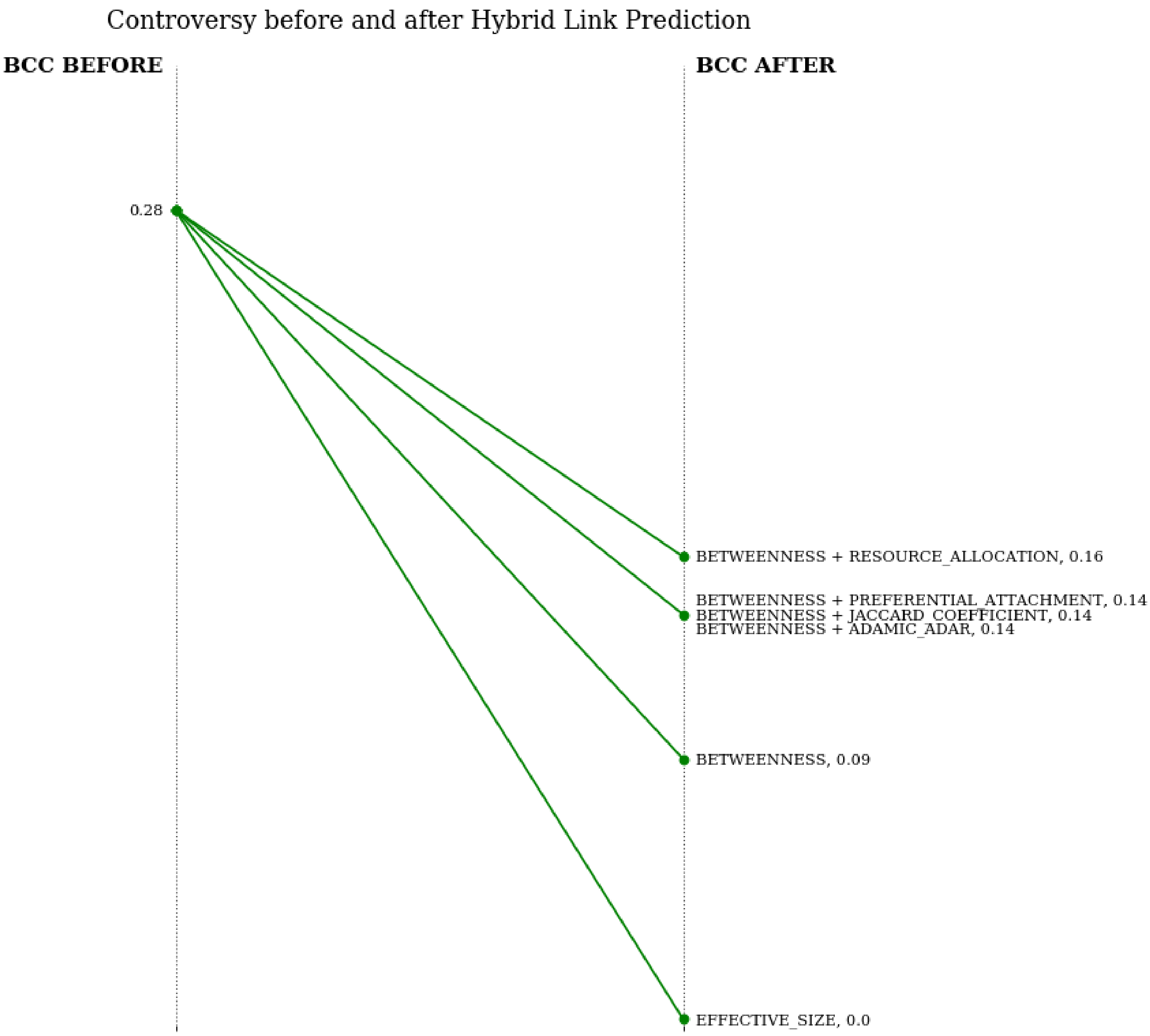



Controversy before and after Hybrid Link Prediction
EC BEFORE
EC AFTER
0.72
BETWEENNESS + JACCARD_COEFFICIENT, 0.67
BETWEENNESS, 0.67
BETWEENNESS + ADAMIC_ADAR, 0.66
BETWEENNESS + RESOURCE_ALLOCATION, 0.65
EFFECTIVE_SIZE, 0.65
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.65

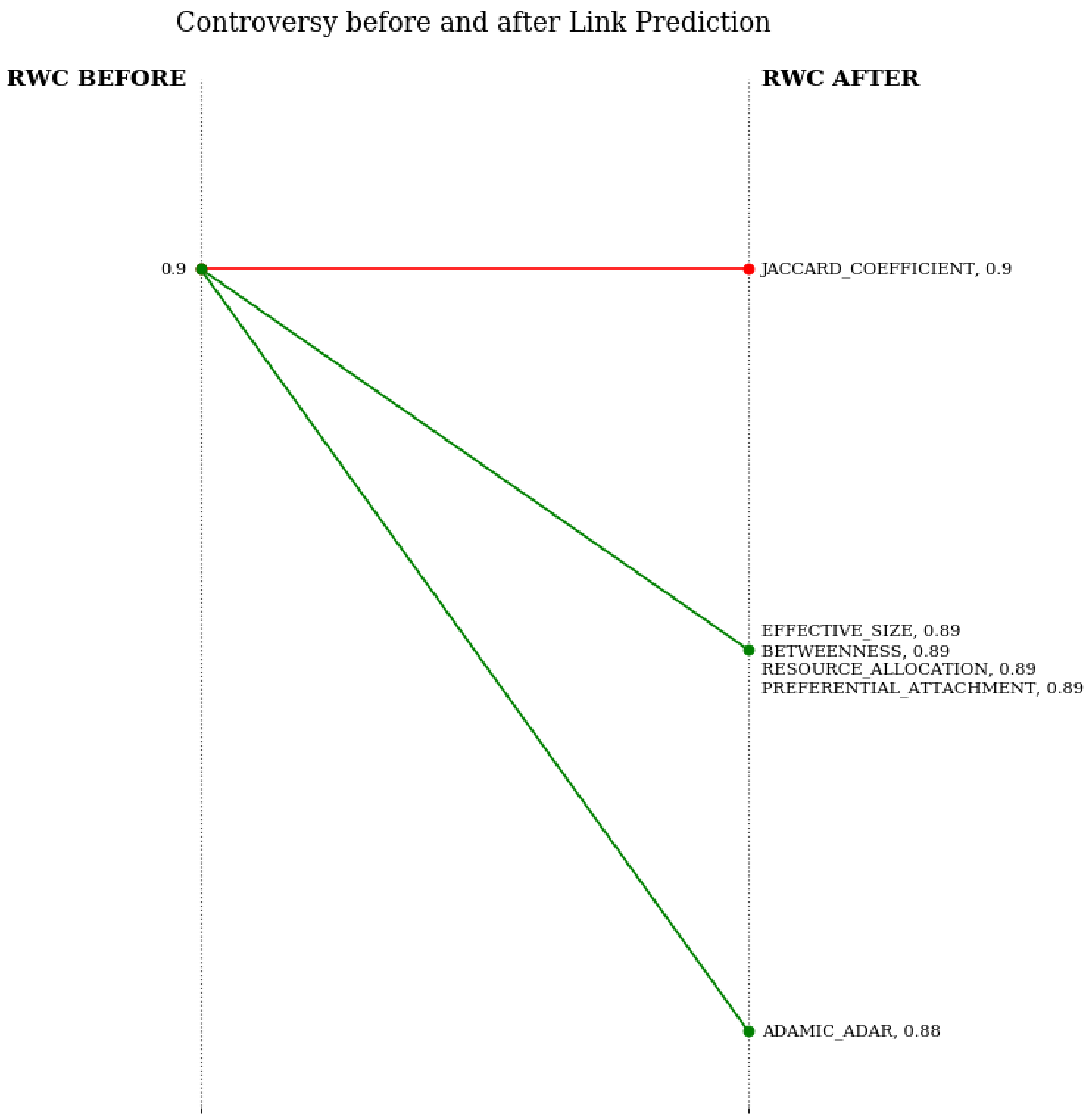
A.2 Hashtag #*russia_march*: Results

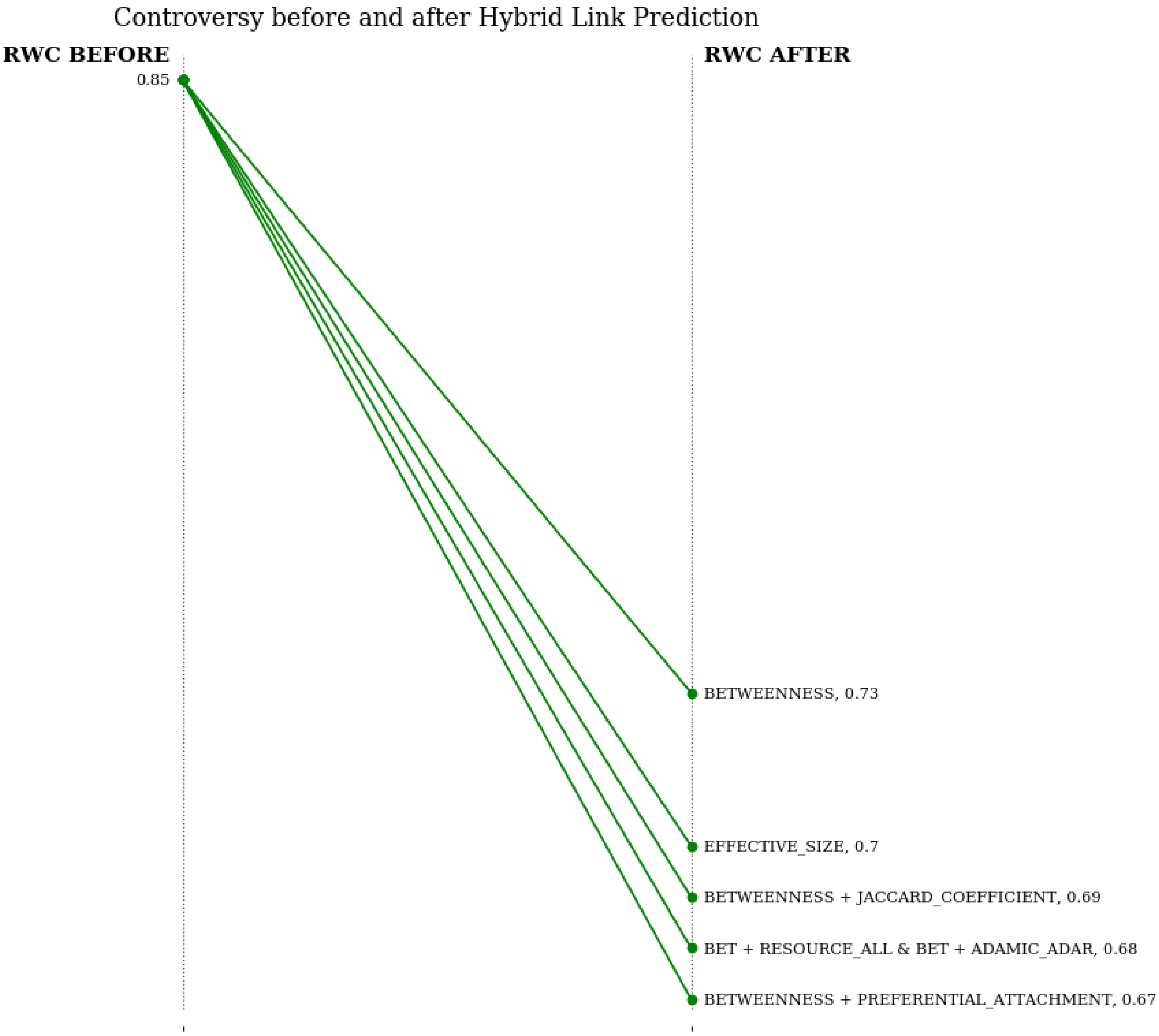
A.2.1 Reduction of Boundary Connectivity Controversy



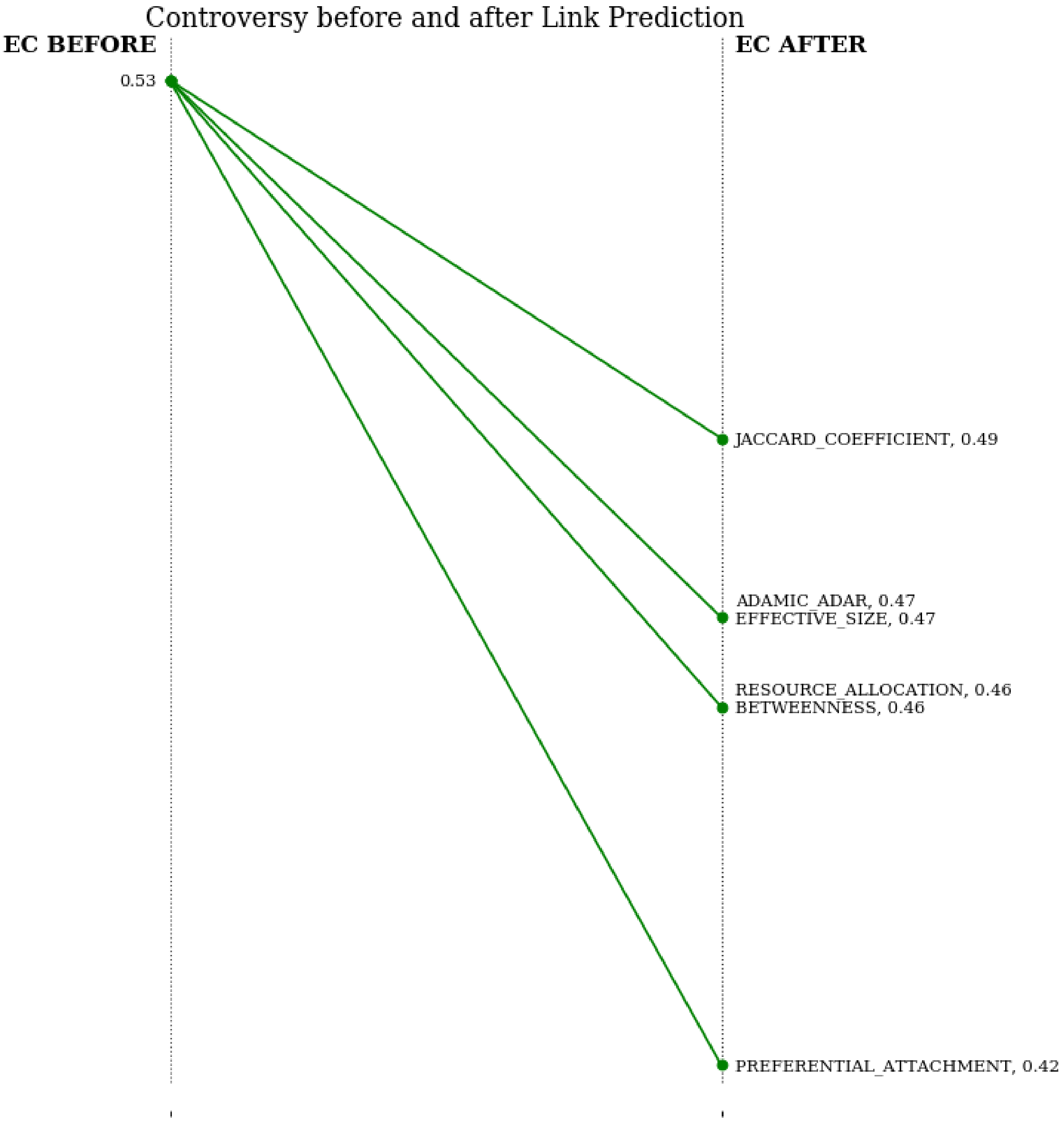

Controversy before and after Hybrid Link Prediction
BCC BEFORE
BCC AFTER
0.28
BETWEENNESS + RESOURCE_ALLOCATION, 0.16
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.14
BETWEENNESS + JACCARD_COEFFICIENT, 0.14
BETWEENNESS + ADAMIC_ADAR, 0.14
BETWEENNESS, 0.09
EFFECTIVE_SIZE, 0.0

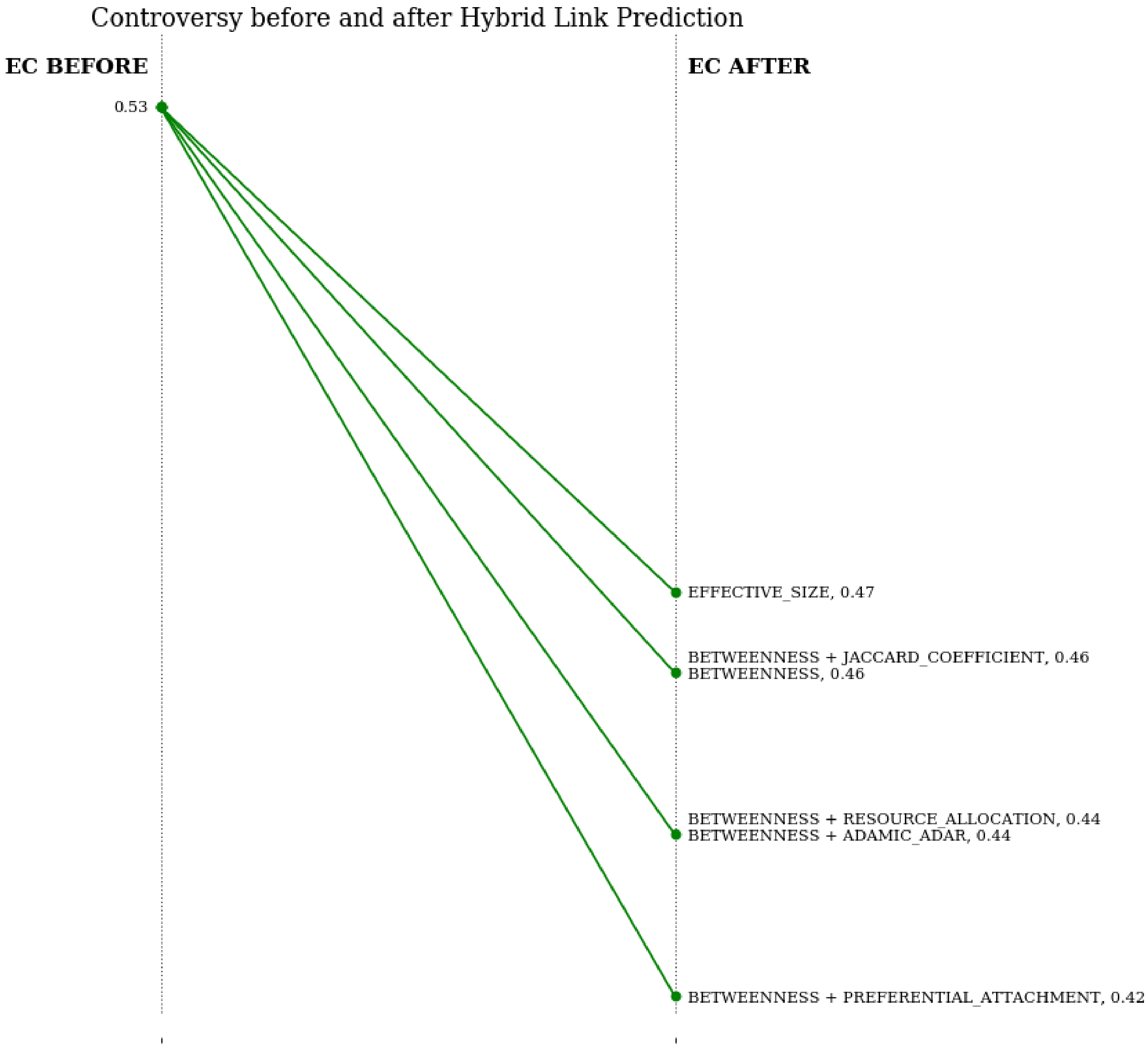
A.2.2 Reduction of Random Walk Controversy




Controversy before and after Hybrid Link Prediction
RWC BEFORE
RWC AFTER
0.85
BETWEENNESS, 0.73
EFFECTIVE_SIZE, 0.7
BETWEENNESS + JACCARD_COEFFICIENT, 0.69
BET + RESOURCE_ALL & BET + ADAMIC_ADAR, 0.68
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.67

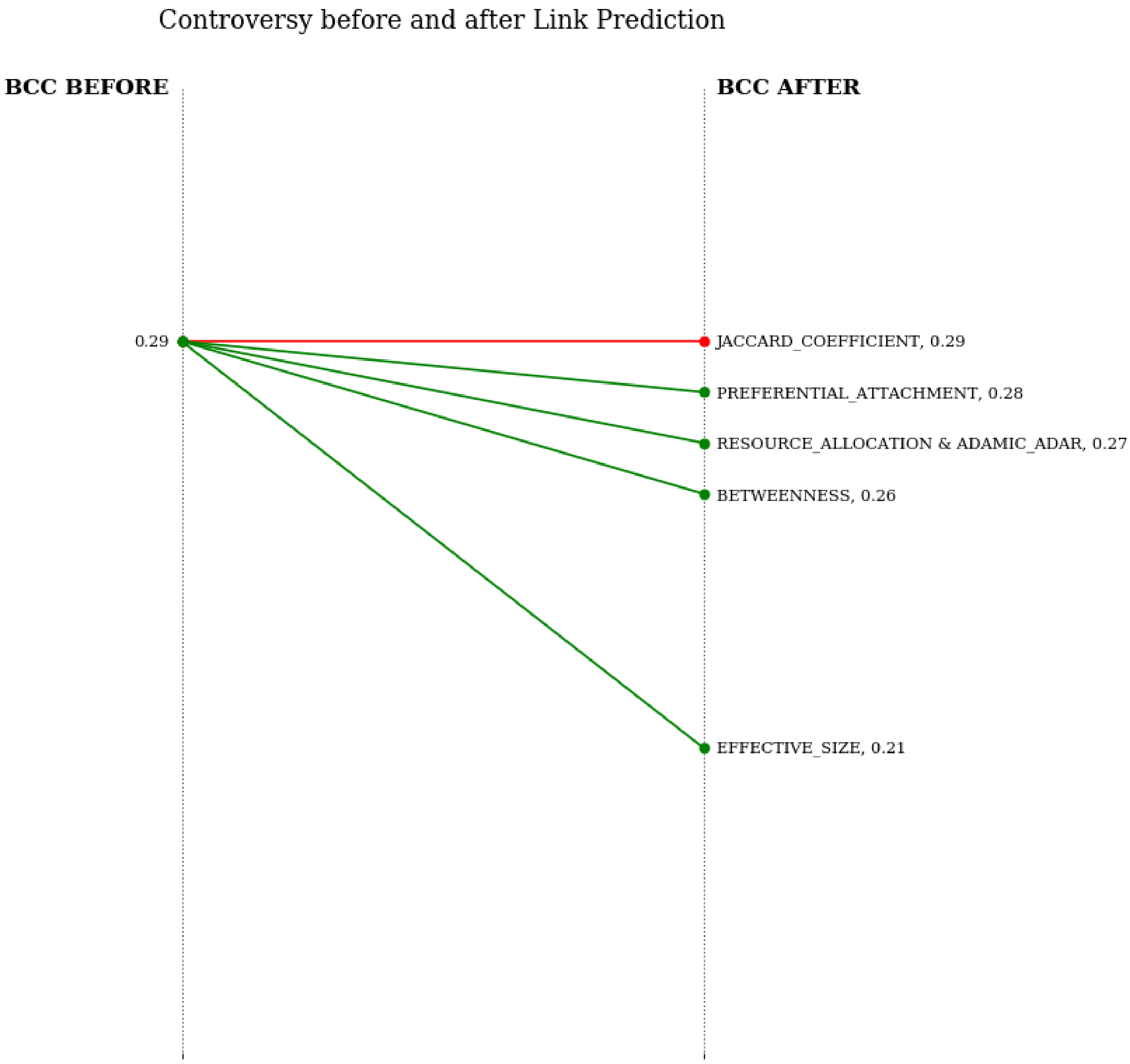
A.2.3 Reduction of Embedding Controversy

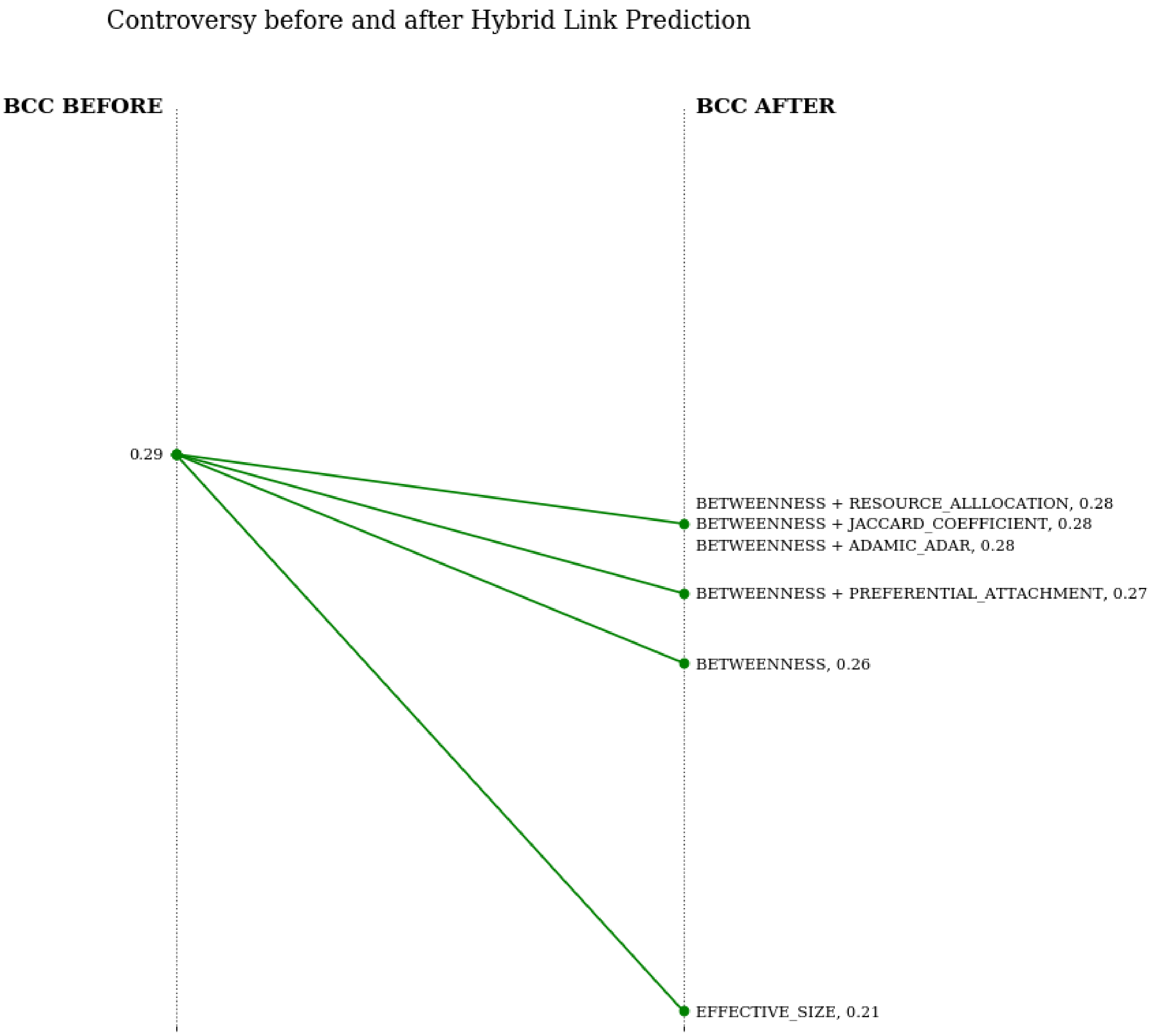



Controversy before and after Hybrid Link Prediction
EC BEFORE
EC AFTER
0.53
EFFECTIVE_SIZE, 0.47
BETWEENNESS + JACCARD_COEFFICIENT, 0.46
BETWEENNESS, 0.46
BETWEENNESS + RESOURCE_ALLOCATION, 0.44
BETWEENNESS + ADAMIC_ADAR, 0.44
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.42

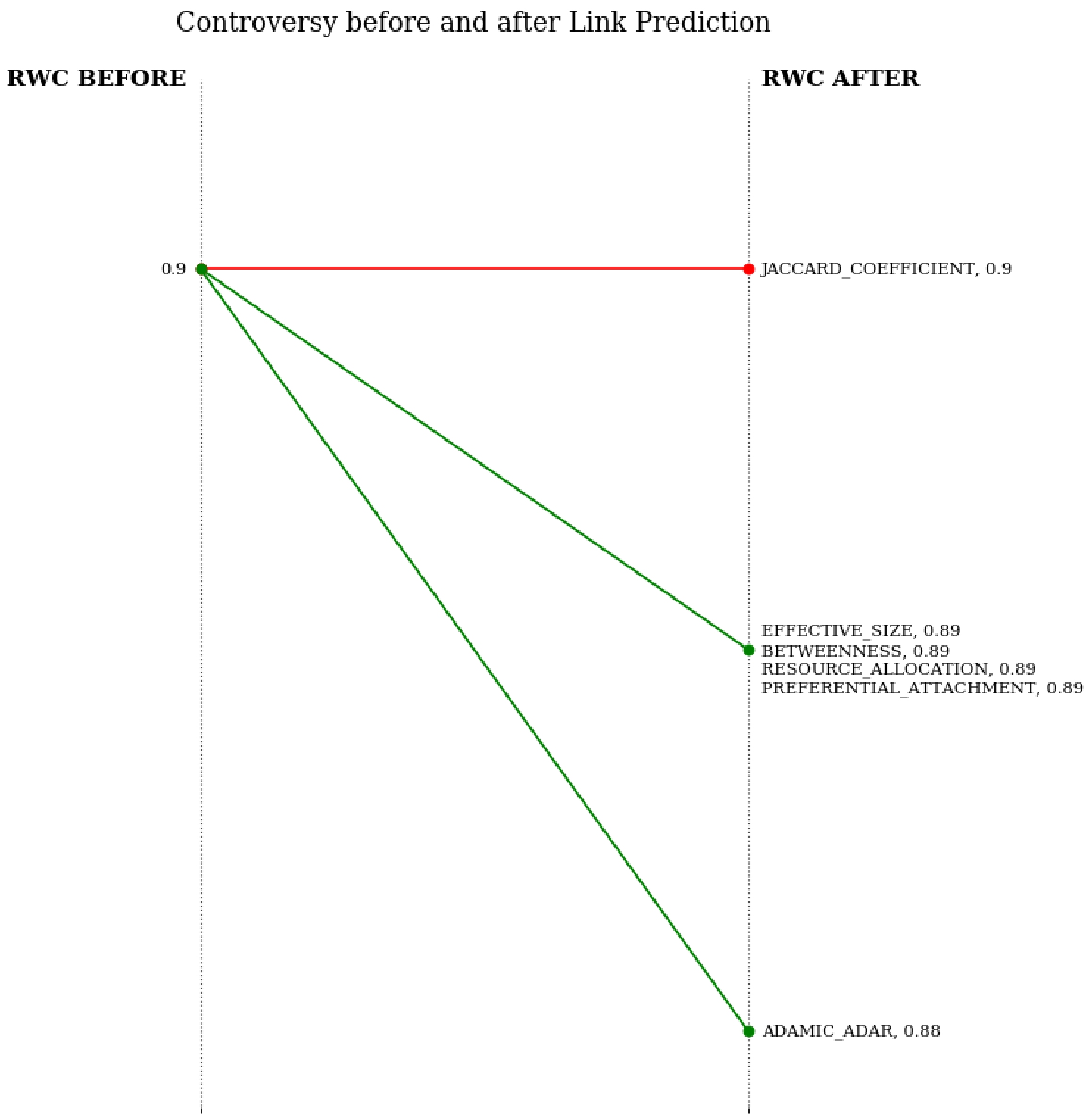
A.3 Hashtag *#netanyahuspeech*: Results

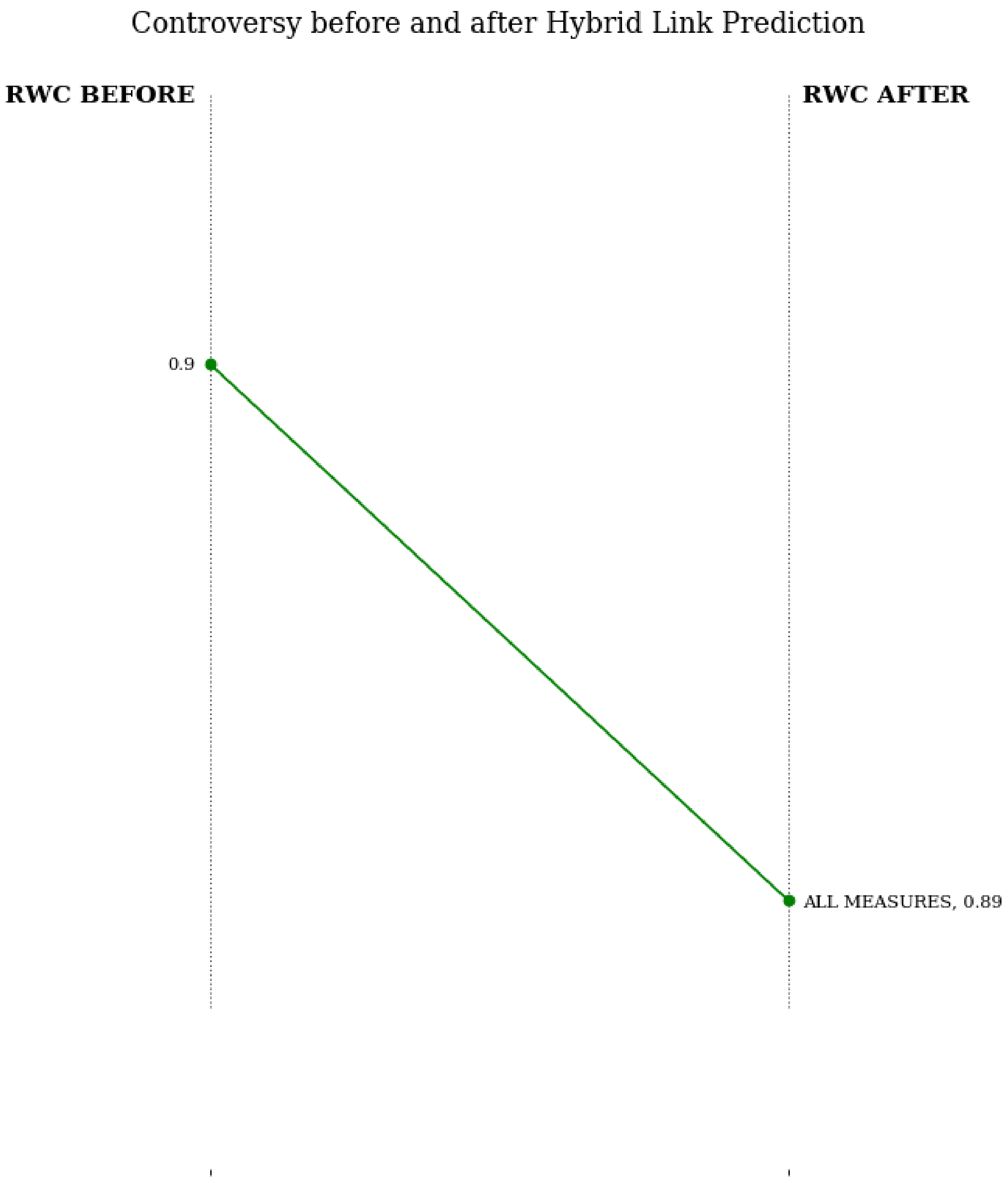
A.3.1 Reduction of Boundary Connectivity Controversy



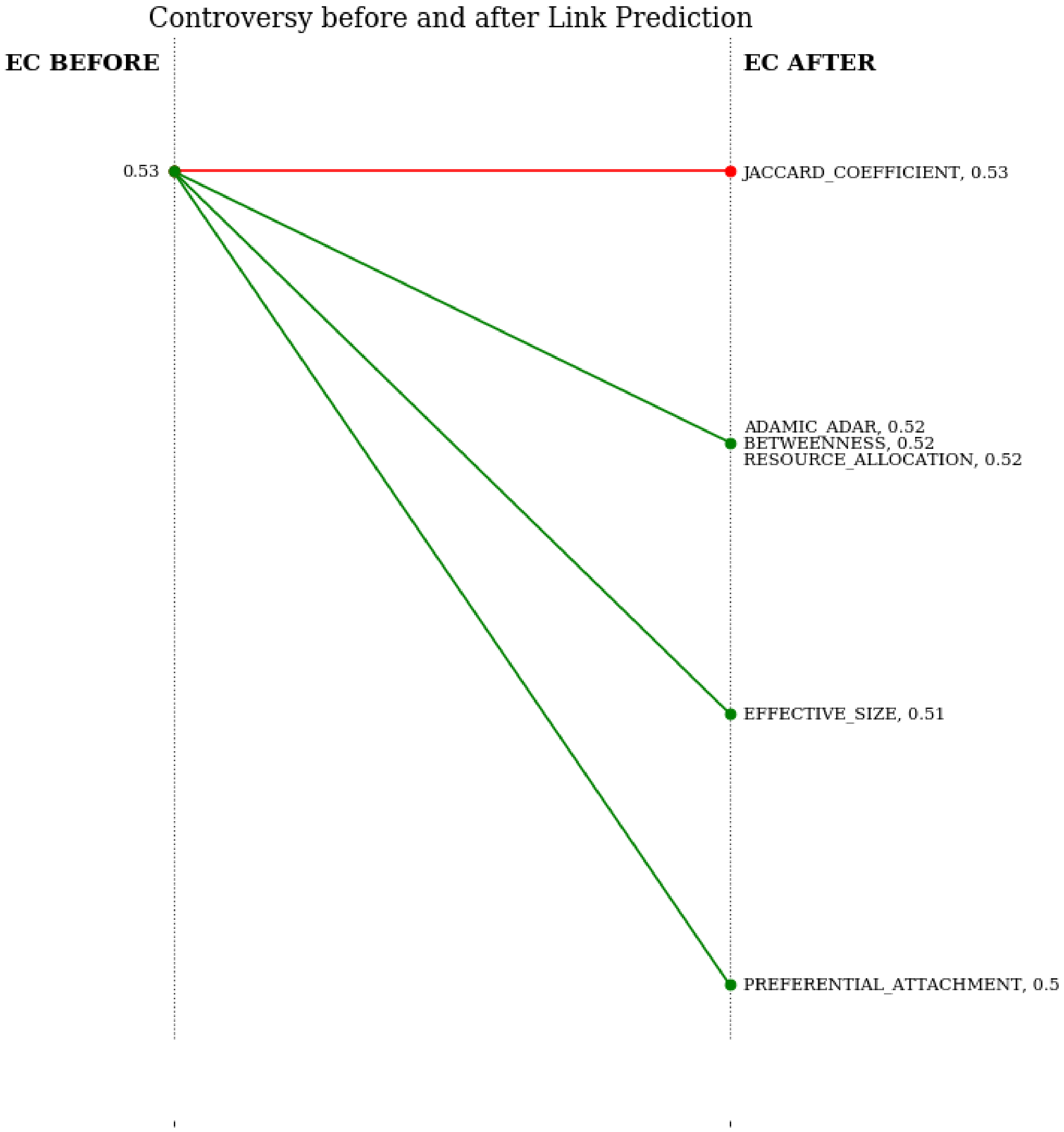

Controversy before and after Hybrid Link Prediction
BCC BEFORE
BCC AFTER
0.29
BETWEENNESS + RESOURCE_ALLLOCATION, 0.28
BETWEENNESS + JACCARD_COEFFICIENT, 0.28
BETWEENNESS + ADAMIC_ADAR, 0.28
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.27
BETWEENNESS, 0.26
EFFECTIVE_SIZE, 0.21

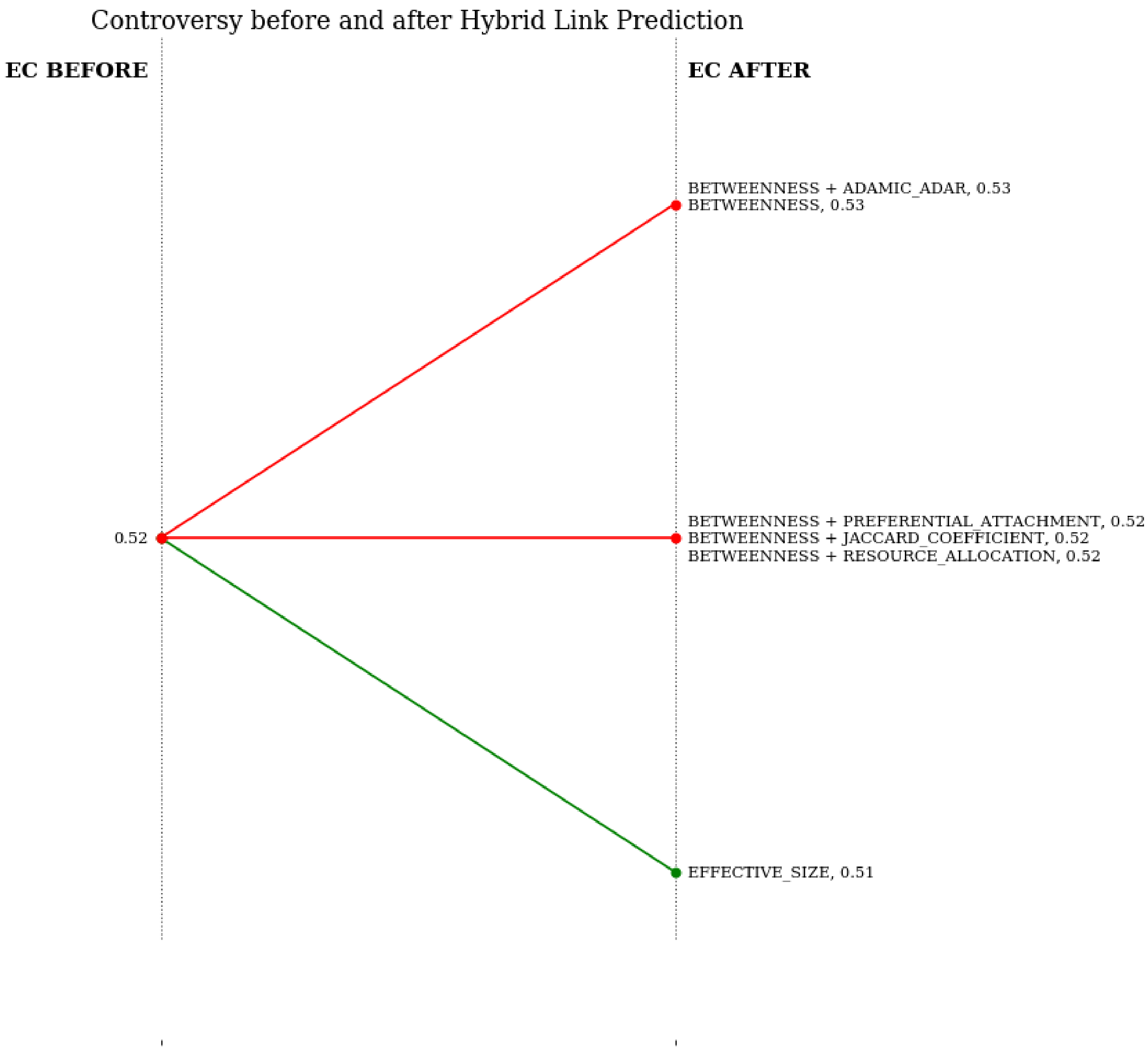
A.3.2 Reduction of Random Walk Controversy




Controversy before and after Hybrid Link Prediction
RWC BEFORE
RWC AFTER
0.9
ALL MEASURES, 0.89

A.3.3 Reduction of Embedding Controversy




Controversy before and after Hybrid Link Prediction
EC BEFORE
EC AFTER
BETWEENNESS + ADAMIC_ADAR, 0.53
BETWEENNESS, 0.53
0.52
BETWEENNESS + PREFERENTIAL_ATTACHMENT, 0.52
BETWEENNESS + JACCARD_COEFFICIENT, 0.52
BETWEENNESS + RESOURCE_ALLOCATION, 0.52
EFFECTIVE_SIZE, 0.51

Bibliography

Abisheva, Adiya, David Garcia, and Frank Schweitzer (2016). “When the filter bubble bursts: collective evaluation dynamics in online communities”. In: *Proceedings of the 8th ACM Conference on Web Science*. ACM, pp. 307–308.

Adamic, Lada A. and Eytan Adar (2001). “Friends and neighbors on the Web”. In: *Social Networks* 25, pp. 211–230.

Bakshy, Eytan et al. (2012). “The role of social networks in information diffusion”. In: *Proceedings of the 21st international conference on World Wide Web*, pp. 519–528.

Baumann, Fabian et al. (2019). “Modeling echo chambers and polarization dynamics in social networks”. In: *arXiv preprint arXiv:1906.12325*.

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

Bozdag, Engin and Jeroen van den Hoven (2015). “Breaking the filter bubble: democracy and design”. In: *Ethics and Information Technology* 17.4, pp. 249–265.

Burt, Ronald S (2009). *Structural holes: The social structure of competition*. Harvard university press.

Davies, Huw C (2018). “Redefining Filter Bubbles as (Escapable) Socio-Technical Recursion”. In: *Sociological Research Online* 23.3, pp. 637–654.

Del Vicario, Michela et al. (2016a). “Echo chambers: Emotional contagion and group polarization on facebook”. In: *Scientific reports* 6, p. 37825.

Del Vicario, Michela et al. (2016b). “The spreading of misinformation online”. In: *Proceedings of the National Academy of Sciences* 113.3, pp. 554–559.

Duseja, Nikita and Harsh Jhamtani (2019). “A Sociolinguistic Study of Online Echo Chambers on Twitter”. In: *Proceedings of the Third Workshop on Natural Language Processing and Computational Social Science*, pp. 78–83.

Floridi, Luciano (2014). *The fourth revolution: How the infosphere is reshaping human reality*. OUP Oxford.

Fortunato, Santo (2010). “Community detection in graphs”. In: *Physics reports* 486.3-5, pp. 75–174.

Garimella, Kiran et al. (2015). *Quantifying Controversy in Social Media*. arXiv: 1507.05224 [cs.SI].

Garimella, Kiran et al. (2017). “Reducing controversy by connecting opposing views”. In: *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pp. 81–90.

— (2018). "Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship". In: *Proceedings of the 2018 World Wide Web Conference*. International World Wide Web Conferences Steering Committee, pp. 913–922.

Geschke, Daniel, Jan Lorenz, and Peter Holtz (2019). "The triple-filter bubble: Using agent-based modelling to test a meta-theoretical framework for the emergence of filter bubbles and echo chambers". In: *British Journal of Social Psychology* 58.1, pp. 129–149.

Gorn, Gerald, Michel Tuan Pham, and Leo Yatming Sin (2001). "When arousal influences ad evaluation and valence does not (and vice versa)". In: *Journal of consumer Psychology* 11.1, pp. 43–55.

Granovetter, Mark S (1977). "The strength of weak ties". In: *Social networks*. Elsevier, pp. 347–367.

Guerra, Pedro Calais et al. (2013). "A measure of polarization on social media networks based on community boundaries". In: *Seventh International AAAI Conference on Weblogs and Social Media*.

Jacomy, Mathieu et al. (2011). "Forceatlas2, a continuous graph layout algorithm for handy network visualization". In: *Medialab center of research* 560, p. 4.

Kernighan, Brian W and Shen Lin (1970). "An efficient heuristic procedure for partitioning graphs". In: *The Bell system technical journal* 49.2, pp. 291–307.

Kühne, Rinaldo et al. (2014). "Political news, emotions, and opinion formation: Toward a model of emotional framing effects". In: *Annual conference of the international communication association (ICA), Phoenix, AZ*.

Liben-Nowell, David and Jon Kleinberg (2007). "The link-prediction problem for social networks". In: *Journal of the American society for information science and technology* 58.7, pp. 1019–1031.

Maccatrozzo, Valentina (2012). "Burst the filter bubble: using semantic web to enable serendipity". In: *International Semantic Web Conference*. Springer, pp. 391–398.

McPherson, Miller, Lynn Smith-Lovin, and James M Cook (2001). "Birds of a feather: Homophily in social networks". In: *Annual review of sociology* 27.1, pp. 415–444.

Milgram, Stanley (1967). "The small world phenomenon". In: *Psychol. Today* 2, pp. 60–67.

Morales, AJ et al. (2015). "Measuring political polarization: Twitter shows the two sides of Venezuela". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25.3, p. 033114.

Nechushtai, Efrat and Seth C Lewis (2019). "What kind of news gatekeepers do we want machines to be? Filter bubbles, fragmentation, and the normative dimensions of algorithmic recommendations". In: *Computers in Human Behavior* 90, pp. 298–307.

Noack, Andreas (2009). "Modularity clustering is force-directed layout". In: *Physical Review E* 79.2, p. 026102.

Pardos, Zachary A and Weijie Jiang (2019). “Combating the Filter Bubble: Designing for Serendipity in a University Course Recommendation System”. In: *arXiv preprint arXiv:1907.01591*.

Parés, Ferran et al. (2017). “Fluid Communities: A Community Detection Algorithm”. In: *CoRR* abs/1703.09307.

Pariser, Eli (2011). *The filter bubble: What the Internet is hiding from you*. Penguin UK.

Quattrociocchi, Walter, Antonio Scala, and Cass R Sunstein (2016). “Echo chambers on Facebook”. In: *Available at SSRN 2795110*.

Raghavan, Usha Nandini, Réka Albert, and Soundar Kumara (2007). “Near linear time algorithm to detect community structures in large-scale networks”. In: *Physical review E* 76.3, p. 036106.

Resnick, Paul et al. (2013). “Bursting your (filter) bubble: strategies for promoting diverse exposure”. In: *Proceedings of the 2013 conference on Computer supported cooperative work companion*. ACM, pp. 95–100.

Russell, James A and Lisa Feldman Barrett (1999). “Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant.” In: *Journal of personality and social psychology* 76.5, p. 805.

Sasahara, Kazutoshi et al. (2019). “On the inevitability of online echo chambers”. In: *arXiv preprint arXiv:1905.03919*.

Schulz, Christian (2016). “Graph Partitioning and Graph Clusteringin Theory and Practice”. In: *Institute for Theoretical Informatics Karlsruhe Institute of Technology (KIT).–May* 20, pp. 24–187.

Vinokur, Amiram and Eugene Burnstein (1978). “Novel argumentation and attitude change: The case of polarization following group discussion”. In: *European Journal of Social Psychology* 8.3, pp. 335–348.

Viviani, Marco and Gabriella Pasi (2017). “Credibility in social media: opinions, news, and health information—a survey”. In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 7.5, e1209.

Yang, Zhao, René Algesheimer, and Claudio J Tessone (2016). “A comparative analysis of community detection algorithms on artificial networks”. In: *Scientific reports* 6, p. 30750.

Zhang, Pengyuan et al. (2015). “A method of link prediction based on betweenness”. In: *International Conference on Computational Social Networks*. Springer, pp. 228–235.

Zimmer, Franziska et al. (2019). “Fake News in Social Media: Bad Algorithms or Biased Users?” In: *Journal of Information Science Theory and Practice* 7.2, pp. 40–53.