Longitudinal Study of Polarization in Social Media Platforms

A Thesis

submitted to the designated
by the General Assembly
of the Department of Computer Science and Engineering
Examination Committee

by

Chrysoula Terizi

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN DATA AND COMPUTER
SYSTEMS ENGINEERING

WITH SPECIALIZATION
IN DATA SCIENCE AND ENGINEERING

University of Ioannina
February 2021

Examining Committee:

- **Evaggelia Pitoura**, Professor, Department of Computer Science & Engineering, University of Ioannina (Supervisor)
- **Panayiotis Tsaparas**, Associate Professor, Department of Computer Science & Engineering, University of Ioannina
- **Nikos Mamoulis**, Professor, Department of Computer Science & Engineering, University of Ioannina

Acknowledgements

First and foremost, I would like to express my deep and sincere gratitude to my research supervisor, Evaggelia Pitoura, Professor at the department of Computer Science & Engineering, University of Ioannina and, for giving me the opportunity to do research and providing invaluable guidance through out this research. Her daily encouragement, vision, sincerity and motivation have deeply inspired me. She has taught me the methodology to carry out the research and to present the research works as clearly as possible. It was a great privilege and honor to work and study under her guidance. I am extremely grateful for what she has offered me.

My completion of this thesis could not have been accomplished without the support of my parents through their love, caring and sacrifices for educating and preparing me for my future. Also, I express my thanks to my sister for her continuing support to complete this thesis successfully.

Table of Contents

List of Figures

List of Tables

List of Algorithms

Abstract

Chrysoula Terizi, M.Sc. in Data and Computer Systems Engineering, Department of Computer Science and Engineering, School of Engineering, University of Ioannina, Greece, February 2021.
Longitudinal Study of Polarization in Social Media Platforms.
Advisor: Evaggelia Pitoura, Professor.

Over the last decade, there has been an increase in the number of users who participate in various discussions on social media; discussions on political issues, sports, religion, etc. Therefore, their participation in controversial topics of discussion is quite common. There is controversy, if users disagree on a topic of discussion and there ar several extreme points of views. The occurrence of polarization in the network is also common. There is polarization if there are two diametrically opposed points of views and users are divided into two communities where either there is no communication between them or when communication exists, it expresses disagreement.

In this thesis, we study controversy and polarization in Reddit. Reddit classifies the topics of discussion into categories namely, subreddits or communities. We investigate the following research questions: Is there controversy or polarization inside a community (intra−polarization), or between two or more subreddits that discuss the same topic (inter−polarization)? Does controversy increase polarization, i.e., are controversial posts more prone to polarization than non−controversial ones? What are the common motifs of user interaction in the comments of a discussion in the case of controversial and non−controversial posts?

To address these questions, we exploit the structure of the user interactions. Specifically, we construct aggregated user conversation graphs (AUG), whose nodes are users and there is an edge between two users if the corresponding users responded to each other. We study both unsigned AUGs and signed AUGs where the sign of an

edge is induced by the controversy score of the comments (a plus sign on an edge express agreement and a minus signed disagreement between the two users connected by the edge). We apply various algorithms that quantify controversy and polarity in a network, including algorithms based on random walks, embeddings, betweenness centrality of the edges and eigenvectors of the adjacency matrix of the AUG graphs. Finally, we analyze the most frequent motifs in AUG and in discussion graphs.

We apply these algorithms to discussions in Reddit about four current events. We find that there is some degree of (unsigned) inter−polarization across subreddits expressing different sides on the event, which is more evident in the case of controversial posts. We also find some degree of intra−polarization in the signed graphs. Finally, the discussion motifs are similar in both controversial and non−controversial posts.

Εκτεταμενη Περιληψη

Χρυσούλα Τεριζή, Δ.Μ.Σ. στη Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πολυτεχνική Σχολή, Πανεπιστήμιο Ιωαννίνων, Φεβρουάριος 2021.
Μελέτη πόλωσης σε πλατφόρμες κοινωνικής δικτύωσης.
Επιβλέπων: Ευαγγελία Πιτουρά, Καθηγήτρια.

Την τελευταία δεκαετία, έχει παρατηρηθεί μια αύξηση του πλήθους των χρηστών που συμμετέχουν ενεργά σε διάφορες συζητήσεις στα μέσα κοινωνικής δικτύωσης. Συζητήσεις οι οποίες είτε σχετίζονται με πολιτικά ζητήματα ή με τον αθλητισμό είτε ακόμη και με την θρησκεία ή τα δικαιώματα των γυναικών και των ανδρών στην σημερινή εποχή κλπ. Επομένως, η συμμετοχή τους σε θέματα συζήτησης τα οποία χαρακτηρίζονται ως αμφιλεγόμενα από την πλειψηφία είναι πολύ πιθανή. Ένα θέμα συζήτησης θεωρείται αμφιλεγόμενο όταν οι χρήστες οι οποίοι συμμετέχουν στην συζήτηση διαφωνούν και υπάρχουν πολλές ακραίες απόψεις. Επίσης, η εμφάνιση πόλωσης σε ένα δίκτυο είναι επίσης αρκετά συχνή σε περιπτώσεις αμφιλεγόμενων θεμάτων συζήτησης. Λέμε ότι εμφανίζεται πόλωση σε ένα δίκτυο χρηστών εάν υπάρχουν δύο διαμετρικά αντίθετες απόψεις και οι χρήστες χωρίζονται σε δύο κοινότητες όπου είτε δεν υπάρχει επικοινωνία μεταξύ τους είτε όταν υπάρχει επικοινωνία δηλώνει διαφωνία.

Στην τρέχουσα διατριβή, μελετάμε την ανίχνευση διαμάχης και πόλωσης στο Reddit. Το Reddit είναι ένα είδος κοινωνικής δικτύωσης με μορφή forum, δηλαδή η κύρια ενέργεια των χρηστών είναι να αλληλεπιδρούν μεταξύ τους μέσω των απαντήσεων τους. Ακόμη, το Reddit ταξινομεί τα θέματα συζήτησης σε κατηγορίες, τα οποία ονομάζονται subreddits. Άρα, μια σειρά από ερευνητικά ερωτήματα αναπτύσσεται προς μελέτη και αυτά είναι τα εξής: Υπάρχει διαμάχη ή πόλωση σε μια κοινότητα (εσωτερική πόλωση) ή μεταξύ δύο ή περισσότερων κοινοτήτων που συζητούν το ίδιο θέμα (εξωτερική πόλωση); Η ύπαρξη διαμάχης αυξάνει και την ύπαρξη πόλωσης μέσα στο δίκτυο; δηλαδή, οι αμφιλεγόμενες αναρτήσεις είναι πιο επιρρεπείς σε πόλωση από τις μη αμφισβητούμενες; Μπορούν να ανιχθευθούν κοινά μοτίβα στον τρόπο αλληλεπίδρασης των χρηστών στα σχόλια μιας συζήτησης σε περίπτωση αμφιλεγόμενων και μη αμφιλεγόμενων δημοσιεύσεων;

Για την αντιμετώπιση αυτών των ερωτήσεων, εκμεταλλευόμαστε τη δομή αλληλεπίδρασης των χρηστών μέσα στις συζητήσεις. Συγκεκριμένα, κατασκευάζουμε συγκεντρωτικά γραφήματα συνομιλίας χρηστών (*AUG*), των οποίων οι κόμβοι είναι χρήστες και υπάρχει μια σύνδεση (ακμή) μεταξύ δύο χρηστών εάν οι αντίστοιχοι χρήστες ανταποκρίθηκαν ο ένας στον άλλο. Μελετάμε δυο είδη *AUG* γραφημάτων, μη προσημασμένα (unsigned) *AUG*s και προσημασμένα (signed) *AUG*s, όπου το πρόσημο στην σύνδεση προκύπτει από τη βαθμολογία διαφωνίας των σχολίων (ένα σύμβολο "+" δηλώνει συμφωνία και "-" δηλώνει διαφωνία μεταξύ των δύο χρηστών που συνδέονται από την σύνδεση). Έπειτα, εφαρμόζουμε διάφορους αλγόριθμους που ποσοτικοποιούν τη διαμάχη και την πολικότητα σε ένα δίκτυο, συμπεριλαμβανομένων αλγορίθμων που βασίζονται σε τυχαίες διαδρομές, αναπαράσταση γράφων σε χαμηλότερη διάσταση, κεντρικότητα ακμών και ιδιοδιανύσματα της μήτρας γειτονίας των γραφημάτων *AUG*. Τέλος, αναλύουμε τα πιο συχνά μοτίβα σε *AUG* γραφήματα και σε γραφήματα συζήτησης.

Εφαρμόζουμε αυτούς τους αλγόριθμους σε συζητήσεις στο Reddit σχετικά με τέσσερα τρέχοντα γεγονότα. Συγκεκριμένα, επιλέγουμε να μελετήσουμε τα εξής θέματα: Μετατροπή της Αγιάς Σοφιάς που βρίσκεται στην Κωνσταντινούπολη σε ισλαμικό τέμενος. Η μελέτη επικεντρώθηκε σε τρεις κοινότητες (subreddits), στην Greece, Turkey και Islam. Το δεύτερο θέμα συζήτησης προέρχεται από την σύγκρουση για την περιοχή Ναγκόρνο−Καραμπάχ μεταξύ των χωρών Αρμενίας και Αζερμπαιτζάν. Επιλέγουμε να κάνουμε την ανάλυση μεταξύ των κοινοτήτων Armenia, Azerbaijan και Turkey. Το τρίτο αμφιλεγόμενο θέμα συζήτησης σχετίζεται με την ανεξέλεγκτη άσκηση αντυνομικής βίας στην Αμερική και συγκεκριμένα, με την ανθρωποκτονία ενός Αφροαμερικάνου τραγουδιστή ονόματι Τζορτζ Φλόιντστις από έναν λευκό αστυνομικό ονόματι Ντέρεκ Σόβιν. Οι κοινότητες τις οποίες επιλέγουμε να αναλύσουμε είναι οι εξής Unpopularopinion, Bad_Cop_No_Conut και BlackLivesMatter. Τέλος, ένα αρκετά πρόσφατο θέμα συζήτησης το οποίο μέχρι και σήμερα βρίσκεται στο επίκεντρο των συζητήσεων είναι για την πανδημία COVID−19. Οι κοινότητες χρηστών που μελετήθηκαν είναι οι China_Flu και Coronavirus.

Με την ολοκλήρωση των πειραμάτων που δρομολογήθηκαν και υλοποιήθηκαν καταλήγουμε σε κάποια συμπεράσματα. Συγκεκριμένα, διαπιστώνουμε ότι υπάρχει κάποιος βαθμός (unsigned) εξωτερικής πόλωσης μεταξύ των subreddits που εκφράζουν διαφορετικές πλευρές στο συμβάν, κάτι που είναι πιο εμφανές στην περίπτωση αμφιλεγόμενων δημοσιεύσεων. Βρίσκουμε επίσης, κάποιο βαθμό εσωτερικής πόλωσης σε προσημασμένα γραφήματα. Τέλος, μοτίβα που προκύπτουν από την ανάλυση της αλληλεπίδρασης των χρηστών μέσα σε μια συζήτηση είναι παρόμοια τόσο σε αμφιλεγόμενες όσο και σε μη αμφιλεγόμενες δημοσιεύσεις.

Chapter 1

Introduction

1.1 Motivation

During the last decades, millions of people daily interact with their fellow human beings and express their views and feelings on a variety of topics, such as politics, economy, religion, education, travel, entertainment, sports, music, climate change and men's and women's rights, via social media platforms e.g. Facebook, Twitter, Reddit, Instagram, etc. As expected, during a discussion between individuals different aspects of views have been observed. These different aspects can be either for, against or neutral to the topic of discussion. For instance, in the case of *United States presidential election 2020* topic of discussion there are two opinion sides, one in favor of Donald Trump and the other in favor of Joe Biden.

This tendency of people leads to various negative social phenomena within a network, such as polarization and controversy. These phenomena, *polarization* and *controversy* are identified and used in the same sense by users. However, by definition they differ and are expressed in dissimilar ways by the users. The definition of the term *polarization* based on the Cambridge dictionary and [1, 2] is the fact of people or opinions being divided into two opposing groups. In social platforms, this could be detected within a community that is, users within a community disagree and two opinion bubbles are created, or cross two communities that is, the users of each community communicate positively with other members of the same community (intra−polarization) and negatively or not at all with members of the opposite community (inter−polarization). Also, the definition of the term *controversy* based on the Cambridge dictionary and [3, 4] declares a lot of disagreement or argument about something, usually because it affects or is important to many people. In other words, there are many extreme views on a topic of discussion. In social platforms, this disagreement can be expressed through the existence of many comments of different perspective on the same topic of discussion.

Why is it important to study this kind of social phenomena? Identifying controversial topics is useful for exploring the space of public discourse and understanding the issues of current interest. Furthermore, knowing how much controversy is a topic of discussion, more targeted recommendations systems can be constructed that foster a healthier "news diet" on social media. Moreover, given the ever increasing impact of polarizing figures in our daily politics and the rise in polarization in the society, it is important to not restrict ourselves to our own "bubbles" or "echo chambers". Hence, focusing on the problem of detection and analysis of controversy and polarization on social media can be used as building blocks for designing such systems to reduce polarity on social media by connecting social media users with content outside their own bubbles.

Thus, a number of recent studies of these phenomena have focused on the problem of quantifying and identifying polarity and controversy in social media mostly based on textual analysis [5, 6, 7, 8, 9, 10, 11, 3]. Some studies approach the problem in a different way and independent of the text. These studies rely on the global network structure [12, 4, 2, 1, 13, 14]. That is, the authors try to determine how polarized a network is just looking how users are connected to the network and when they interact. Another approaches exploit the presence of network motifs, i.e., local patterns of user interaction [15, 16]. Also, some recent researches combine the network structure and the content of the text [17, 18, 19, 20].

As it is an active field of research to date, we identify some limitations in these studies. More specifically, the majority of studies work on specific social networks i.e., Twitter is the most commonly used social media. So, a reasonable question would be whether the algorithms and techniques proposed for Twitter can be applied to other social networks that do not have the same structure as Twitter, such as Reddit. Moreover, the quantification of controversy and the detection of polarized groups is based on algorithms that use static graphs and not temporal graphs. Therefore, it would be interesting to study the impact of these algorithms on temporal graphs as well. Last but not least, due to the identification of the two terms *polarization* and *controversy*, we would like to study if there really could be some identification, i.e. if a conversation is controversial then polarization is expected in the network.

1.2 Objectives of the thesis

In this current thesis, we make a longitudinal study of polarization and controversy in Reddit platform. Why did we choose Reddit platform and how it differs from Twitter?

Reddit vs Twitter. Twitter and Reddit are websites that collect news in the form of posts. But Twitter is only a microblogging and social networking service website. They tweets tweeted need not always be true. On the other hand, Reddit is a social news aggregation website where there are content rating and discussion services. The segregation and maintenance of the posts are also different for the two. In the case of Reddit, the posts are voted on by the users depending on the importance of the news or message being delivered. They are then arranged on the website on votes. Whereas on Twitter, the tweet needs to be trending for one to be able to get the post. The trending ones need not always be important. The pattern for the following is also different for the two websites. One only needs to follow a "subreddit" or "community" related to the topic, such as news, politics, science, sports, etc., that they are searching for news for in Reddit and it would only generate the posts related to those topics in the feed. The news would reach much faster. In the case of Twitter, you need to follow an individual or a group. If you need tweets related to music, you will have to follow the artists itself. Twitter users can tweet, retweet and like the posted tweet. On the other hand, Reddit users can create, comment and vote positive or negative on others' posts and comments.

At this point, we have a first sense of the differences between these two social networks, Reddit and Twitter. Therefore, we choose to study some algorithms for detection and quantification of polarization in Reddit. Initially, we need to clarify how we define polarization in Reddit forum. As already mentioned that the submissions in Reddit are organized in “subreddits”, we define a general definition of polarity and afterward two sub−definitions of this namely, unsigned and signed polarity either within a specific “subreddit” (intra−polarization) or cross more “subreddits” (inter−polarization).

- **Polarity.** People or opinions are divided into two opposing groups of individuals or sights. Either there is no communication between them or if it exists, it declares dispute.

- **Unsigned Polarity.** There is unsigned polarization within a community (intra − polarization) if members of the community are divided into two groups with opposing views on a specific topic of discussion and there is no communication between them. Also, there is polarization cross two or more communities (inter − polarization) if there is no connection between the members of communities in which users discuss the same issue.

- **Signed Polarity.** There is signed polarization within a community (intra − polarization) if members of the community are divided into two subgroups with diametrically opposed viewpoints. Users within each subgroup agree with each other and a positive thinking prevails while there is no interaction between users from the other subgroup. In the case of interaction cross subgroups, we assume that the connection declares disagreement. Disagreement can occur either using downvotes or controversial text. Furthermore, there is signed polarization cross two or more communities (inter − polarization), if the members within each community agree between them and disagree with members of the opposite community. Note, we assume that only one kind of opinion prevails in each community.

In this current study, our main aim is to verify if controversial topics of discussion in real world retain their polarization in the world of social media. As already mentioned above, we consider the problem from two points of views, intra − polarization and inter − polarization in Reddit. Therefore, the three main queries that will be answered follows,

- **RQ1.** Is there polarization within a subreddit or cross two or more subreddits?

- **RQ2.** Does controversy increase polarization i.e., are controversial posts more prone to polarization than non−controversial ones?

- **RQ3.** What are the common motifs of user interaction in the comments of a discussion in the case of controversial and non−controversial posts?

At first, our approach focuses on the structure of the network and it consists of several stages. The initial and most important stage is the generation of a user graph from one or more communities who talk about a common theme. So, we build an aggregated user conversation graph AUG where vertices indicate users and edges indicate connections between users. Afterward, we apply five proposed algorithms that quantify the unsigned structure polarity in a network [12, 1, 21] without considering the type of interaction between users, i.e., if they agree or disagree. Moreover, we measure polarity in signed networks applying five proposed algorithms by [2]. As there are references that the network motifs can determine whether a topic of discussion is controversial or non−controversial, we also study the existence of specific temporal motifs in the network [16]. Our approach also looks for motifs in the structure of the discussions. At the end, a case study of content analysis has been accomplished.

1.3 Outline of the thesis

The dissertation contains 8 chapters. In Chapter 2, a brief description of the existing literature studying the problem of polarization occurs. After that, the description of the main problem is presented in Chapter 3. Chapter 4 follows in which the explanation of the graph construction steps are given. Next, Chapter 5 makes an introduction to all the algorithms with which unsigned (Section 5.1) and signed (Section 5.2) polarization is calculated. In addition, the pseudo−code for each one of them added. Chapter 6 follows in which the general methodology of our approached mentioned. Afterwards, the process of collecting and initializing hyperparameters of the algorithms is marked out in Chapter 7. Subsequently, Chapter 8 discusses the main findings of the research queries. Finally, Chapter 9 indicates the conclusions and some future improvements.

Chapter 2

Related Work

The study of polarization and controversy in online social media has been expanded by many aspects. There are many references focuses on the connection of users within the network. Equally numerous are the studies which look at the type of interactions in social networks. These studies catch up on signed networks where the sign of an interaction is induced either by user opinion and can be friendly or antagonistic. Other approaches reach the problem either from mining motifs from the network or analyzing the meaning and significance of the text.

2.1 Controversy and polarization in unsigned networks

The dynamics and consequences of polarization in human groups have been studied for decades. The political leaning explains behavior patterns in the adoption of positions. The flow of information between different groups leads to conflict in [22]. The study by [23] showed that users who shared the same political orientation interacted more and cited more the contents of those users. On the other hand, the interaction between groups with different points of view was low. The need to quantify the impact of the conflict in social networks led to a sequel of works whose objective was to improve the understanding of this phenomenon.

Similar conclusions have been drawn by [1] where the authors claim that the controversy is related to the level of interaction between influencers with opposite views. Using hashtags to retrieve the network of retweets of a particular topic, they have proposed the use of graph partition algorithms such as METIS ([24]) to detect communities with opposing views. They proposed a metric based on random walks that quantified the level of coupling between two partitions and detecting controversial and non−controversial topics of discussions in Twitter. A case study of Venezuela ([21]) confirms that just a small group of influential individuals propagating their opinions through a social network are needed to produce polarization.

In addition, [12] showed that the use of modularity was not useful to detect polarized communities since non−polarized networks could also be partitioned using modularity. They observed that polarized communities had a more significant presence of high degree users in their boundaries, and therefore they introduced a metric that quantified the presence of highly connected nodes at the boundaries of the candidate communities. Also, [1] noticed that betweenness centrality of edges can detect polarized groups in a network. They claim that if users are well−separated in network then betweenness centrality of cuts edges differ from the rest of the network.

Moreover, a metric based on cluster cohesion is proposed using GENE graph representation which is able to segment the user network according to their polarity [19]. GENE framework is based on three main steps. Initially, a user−entity model that allows to quantify the bias of each user with respect to the most common entities appearing in the corpus is generating. Second step takes advantage of the user−entity model to build a multi−relational graph. Using embeddings computed from the graph, GENE generates a polarized network of intervening users in a discussion that involves specific entities. Finally, a third step analyses the interactions in the generated graph, leading to the identification of controversy. Consequently, the detection of controversy is more predictable at the beginning of an event, and as more users intervene, the task becomes more complicated.

2.2 Controversy and polarization in signed networks

Signed graphs have been used to model interactions in social networks, which can be either positive (friendly, accordance) or negative (antagonistic, dissent). [25] modeled the political leanings of U.S. Congress members from their opinions on specific topics using signed bipartite networks. They conducted a node labeling task on U.S. congressional records, showing that the voting intent of the congressmen was predictable from this model. Using techniques in duality theory and linear algebra, a local spectral approach by [13] finds polarized communities that are related to a small set of seed nodes provided as input. Seed nodes may consist of two sets, which constitute the two sides of a polarized structure. Additionally, [2] a subset of the network vertices is discovered in a signed network where two communities are polarized when within communities there are mostly positive edges while across communities there are mostly negative edges. The approach uses properties of signed adjacency matrices.

2.3 Analyzing text

In some studies, the detection of controversial topics of discussions focus on sentiment and stance analysis [26]. The significant difference is that in sentiment analysis, systems determine whether a piece of text is positive, negative, or neutral. However, in stance detection, systems determine author's favorability towards a given target and the target even may not be explicitly mentioned in the text. A case study of stance detection [27] uses unigrams, initial g− grams (first n words of the sentence), comment length and ECSD dictionaries created mainly for the purpose of entity−related polarity detection. A polarity analysis framework for Twitter messages has been studied in [28]. The knowledge base contains emoticons and sentiment−based words. They label part of the messages to train a classifier such as Naive Bayes, SVM, KNN and J48. Using Twitter as a starting point, the task of identifying controversial events is investigated [3]. The authors claims that a two−step blended system performs best with precision 95% in which a text analysis is applied in a set of tweets per different time snapshots. The analysis includes linguistic, structural, sentiment and controversy features from the tweets.

2.4 Mining network motifs

Identifying controversial topics is useful for exploring the space of public discourse and understanding the issues of current interest. Thus, a number of recent studies focus on the problem of controversy identification in social media mostly based on the analysis of textual content or rely on global network structure. Such approaches have strong limitations due to the difficulty of understanding natural language, especially in short texts, and of investigating the global network structure. [15] shows that it is possible to monitor the evolution of controversy in a conversation over time thus discovering changes in user opinion. They observe that long trees with multiple branches indicating the different threads of the discussions in controversial discussions. Also, more engagement among users is detected in controversial content. Furthermore, [20] claims that even when only a handful of comments are available discussion features often add predictive capacity to strong content−and−rate only baselines. Leveraging features drawn from both the textual content and the tree structure of the early comments such as number of comments, time of first reply, max depth/total comment ratio, average node depth, average branching factor etc., that initiate the discussion can be applied for an early prediction of controversy conversations.

2.5 Conflicts in Reddit

Users of one community in Reddit are mobilized by negative sentiment to comment in another community [18]. A model which combines graph embeddings, user, community and text features achieves AUC of 76% and can be used to create early−warning systems for community moderators to prevent conflicts. In the same direction, [4] extract inter−community (community − to − community) conflicts in Reddit by aggregating users who behave differently depending on the community they interact with. In two subreddits where users discuss the same topic through different ideologies (as determined through text analysis) have very low author overlap. Also,subreddits that target multiple others, will shift their main conflict focus over time. A rather different and original approach to detect inter−community conflict in r/place subreddit has been done by [14]. In r/place subreddit is a canvas where Reddit users can recolor pixels on the canvas but is rate−limited to one such recoloring every five minutes. They assume that conflicts between communities are expressed in the pixel placements of hundreds of Reddit users. The findings show that multiple communities are involved in many conflicts and not one–on–one conflicts only exist.

A recent study of a large–scale characterization of Reddit users' responses to the COVID–19 pandemic has been investigated in [17] cross China_Flu and Coronavirus subreddits. By comparing users' activity and language usage, the authors observe an increasing difference between them in many aspects. Notably, they look at overlapping and common news articles, the founders of these two subreddits and check which communities they came from before joining these two cimmunities, the keywords, the most downvoted comments, and replies to moderators in each community. Finally, they conclude that users in China_flu care more about China–related topics, generate more racist comments, and are more likely to be active in other extreme communities.

A case study of Brexit discussions is investigating in [9]. Especially, they investigate whether the stance of users with respect to contentious subjects is influenced by the online discussions that they are exposed to, and by the interactions with users supporting different stances that is, given the user description at the present time, the purpose is to predict his stance in the future. The perspective looks at this problem as a supervised machine learning problem using different predictive feature sets describing user interactions with information diffusions such as, user activity, user activity per stance, structure of diffusion etc. Moreover, they trained a model on Twitter data that applied on Reddit training data set. Finally, the experiments show notably that the opinions of the users involved in the same diffusions as him allow better to predict his opinion than the content he exchanges.

Chapter 3

Problem definition

In the dissertation the general subject of study is the detection and quantification of polarization in Reddit. Polarization within a network can be caused by controversial topics of discussion. A topic of discussion annotated as controversial if there is disagreement between users and more than one aggressive sights emerge during it [3, 4].

Further, polarization can be defined based on the main idea of the problem being studied each time. The general idea is that in a polarized network there are two opinions about a conversation and the members participating in conversation are divided into two disjoint groups. Particularly, [1, 12, 21] defines that there is polarization in a network if the members of polarized groups are strongly connected with members within their group and weakly connected with members outside their group. Based on this idea, the strong connection signifies agreement and the weakly connection signifies disagreement. Also, the quantification of polarization using this definition needs only unsigned networks.

Another definition of polarization has been investigated by [2, 25, 13] . From this perspective, two polarized groups reappear but not necessarily throughout the network that is, there are two subsets of the network in which the members are more polarized. Contrary to the previous definition, there is not only connection between members who agree but there is also interaction between members who disagree. That is, individuals from each polarized community communicate positively with other individuals within community (intra−polarization) and negatively or not cross individuals of the opposite community (inter−polarization). The polarization detection in this sense of polarization uses signed networks. In the most cases, a plus sign between two users express agreement and a minus signed disagreement between the two connected users.

In this current thesis, we need to define our own definitions of polarization. Following the lines of the above definitions, we delimitate two type of polarization namely, unsigned and signed polarization. These types of polarization can be either intra (intra−polarization) that is, within a community or inter (inter−polarization) that is, cross two or more communities. Our definitions of polarization are listed below,

- **Unsigned Polarity.** There is unsigned polarization within a community (intra − polarization) if individuals of the community are splitted in two subgroups with opposing views on a specific topic of discussion and there is no communication between them. Also, there is polarization cross two or more communities (inter − polarization) if there is no interaction between the members of communities in which users discuss the same issue.

- **Signed Polarity.** There is signed polarization within a community (intra − polarization) if members of the community are divided into two subgroups with diametrically opposed viewpoints. Users within each one agree with each other and a positive thinking prevails while there is no interaction between users from the other subgroup. In the case of interaction cross subgroups, we assume that the connection declares disagreement. Furthermore, there is signed polarization cross two or more communities (inter − polarization), if members within each community agree between them and disagree with members of the opposite community.

A summary of all the definitions is given in Table 3.1. Having presented a detailed description of the basic definitions that the current dissertation uses, the next step is an introduction to the questions that are scrutinized.

Overview of research questions. The general proposal of this dissertation is an longitudinal study of polarization in online social meida (for short OSN). Particularly, the main aim is to verify if controversial conversations nowadays, equally lead social users in two polarized bubbles in Reddit social platform. People of Reddit can participate in discussions which are classified into categories (subreddits) based on the

Table 3.1: Definitions of polarization and controversy.

Definitions	**Description of definitions**
Polarity	People in a network are divided into two non−overlapping groups. There are two perspectives, one is adopted by one group and the other one by the members of the opposite group.
Intra−polarity	Polarization exists only within a community.
Inter−polarity	Polarization exists cross two or more communities.
Unsigned Polarity	People in a network are divided into two non−overlapping groups. Members of one group do not communicate with members of the opposite group.
Signed Polarity	People in a network are divided into two non−overlapping groups. There is agreement between the members within groups and disagreement cross the members of the two polarized groups.
Controversy	There is disagreement between users participating in a discussion. More than one extreme point of views emerge during the discussion.
Controversial posts	A post is controversial in Reddit if the upvote rate is higher than 0.62.
Non−Controversial posts	A post is controversial in Reddit if the upvote rate is less than 0.62.

respective topic. Therefore, our interest was inspired by this classification. That is, a longitudinal study of polarization within and cross communities in Reddit. Having this general idea as a guideline and combining it with what we mentioned in Chapter 2 i.e., Is intra or inter polarization identified by controversial topics of discussion in Reddit?, we can detect polarization if it is based on the number of positive and negative votes, upvotes and downvotes respectively? on controversial and non−controversial dialogues do users tend to act in a specific way towards the other members of the discussion? such as, if a person's answer is negative from the public, will that person continue to interact with the others in the conversation? etc. Therefore, at this point, we set out the main queries of this thesis,

- **RQ1.** Is intra or inter polarization detected in Reddit?
- **RQ2.** Does controversy increase polarization?
- **RQ3.** What are the common motifs of user interaction in the comments of a discussion in the case of controversial and non−controversial posts?

As already mentioned in Chapter 2, there are several publications that approach the problem from different perspectives. These perspectives either look at the structure of the network, or at the view and stance of the individual, or at the content analysis or at patterns. Therefore, our approach has been inspired by all these various techniques. Specifically, our approach is a 2−stage pipeline. In the first stage, we generate an aggregated user conversation graph (AUG) either per community (subreddit) or between more communities (subreddits) in which users discuss the same topic. In AUG graphs vertices indicate users and edges indicate connections between them. The connection either can be signed or unsigned based on the type of polarization we study.

The next (second) stage involves algorithms which are applied to AUG graphs and quantify polarization in a network or detect the two most polarized subcommunities in a network. Especially, we apply five methods to quantify unsigned polarization. The methods quantify the polarity in a network using the structure of the network. The techniques are related to *Random Walks* [1], *Betweenness* centrality of edges [1], *Embeddings* [1], *Boundary Connectivity* [12] and *Dipole Moment* [21]. Furthermore, we apply five methods to quantify signed polarization and detect the two most polarized groups in network where within groups there are mostly positive edges while across groups there are mostly negative edges. The methods we work with have been proposed by [2] and they look at the properties of signed adjacency matrices. Experiments are performed with deterministic and random methods namely, *Eigensign*, *Random Eigensign*, *Greedy*, *Bansal* and *LocalSearch*. Since works has shown that *Random Eigensign* method is more promising on political debates, our study makes an in−depth analysis of this method. Afterwards, as we also study motif detection in graphs of controversial and non−controversial posts, we apply relevant algorithms [16] to AUG graphs. Clearly, a more detailed description of graph generation and all the above algorithms is mentioned in the following Chapters 4, 5 and 6.

Chapter 4

Generating graphs

One of the most crucial parts of the thesis is the construction of appropriate graphs from a set of data. Data for us considered the submissions (posts) in Reddit and the users who participate in these. Two types of graphs are generated based on the purpose of this current work that is, quantification of intra and iter either signed or unsigned polarization in Reddit social platform. The first type of graph is a directed conversation graph (CG) which can be generated per post. It is a representation of the conversation in the form of a tree. The second type of graph is an aggregated user conversation graph (AUG). The construction of an aggregated user conversation graph (AUG) is not so simple. [1] concludes that the most appropriate graph structure for quantifying polarity is an undirected graph $G(V, E)$ where V are the vertices and declare network users and there is an edge between two users if they have retweet posts with similar hashtags. On the other hand, signed graphs [2] are necessary for polarity detection. Therefore, the construction of a user graph using data from a forum website such as, Reddit, is a challenge so far. To make it more clear, suppose a set of discussions from Reddit platform which have the format shown in Figure 4.1. The construction of a CG graph for a post and an AUG graph for a set of posts follows.

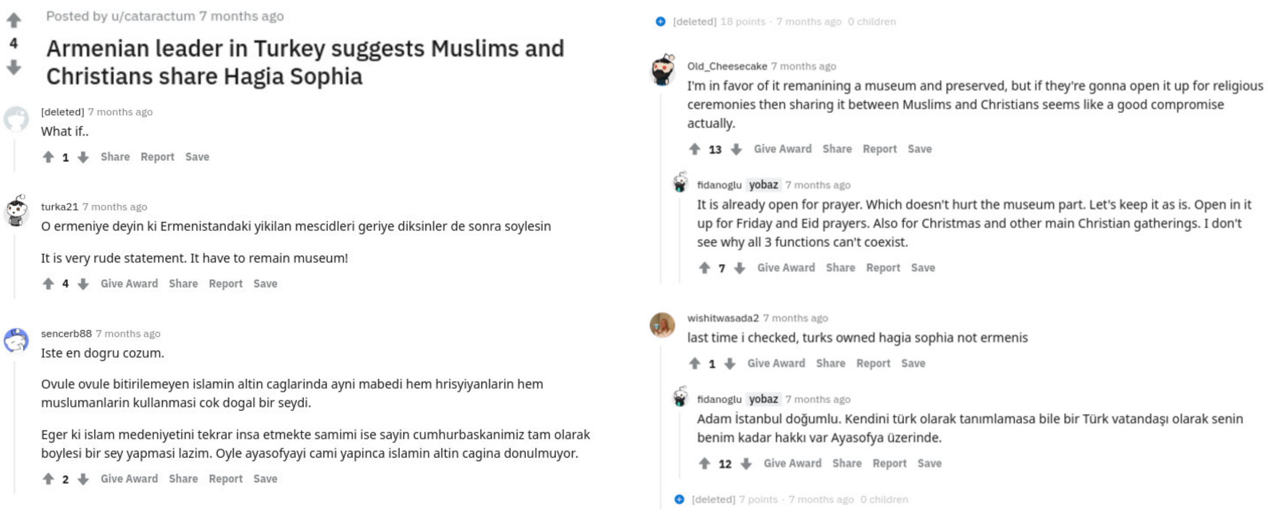


Figure 4.1: Format of a discussion in Reddit.

4.1 Conversation Graph (CG)

$CG(V, E)$ is a directed node − signed conversation graph with V vertices and E edges for a specific post. The nodes declare comments/replies from the post. Each node can have either positive "+" or negative "-" sign. The assignment of the signs is made as follows: each comment has a score value that it, sum of the upvotes and downvotes, if score is higher than zero (more upvotes) then the node sign is "+", otherwise, if the score is less than or equal to zero (the amount of upvotes and downvotes is almost equal or more downvotes exist) then the node sign is "-". Furthermore, there is a directed edge $(v_i, v_j) \in E, i, j \in V$ from node v_i to node v_j, if v_j is an answer to v_i comment. In addition, each edge $(v_i, v_j) \in E$ has a timestamp $t_{(v_i,v_j)}$ that indicates the response time from v_2. The structure of a CG graph is a tree. In Figure 4.2, the directed signed CG graph from the discussion in Figure 4.1 is presenting. The sign of green nodes is "+" and "-" for the red ones.

4.2 Aggregated User Conversation Graph (AUG)

Initially, the general idea of the AUG graphs is to construct graphs from a set of submissions related to a specific topic of discussion. That is, how to generate from a set of CGs a user AUG graph either signed single-edge and unsigned multi-edge. First of all, we need to sort out the meaning of the sign. The annotation of a sign at the edges of the graph deviates from the score of the comments in the discussion and

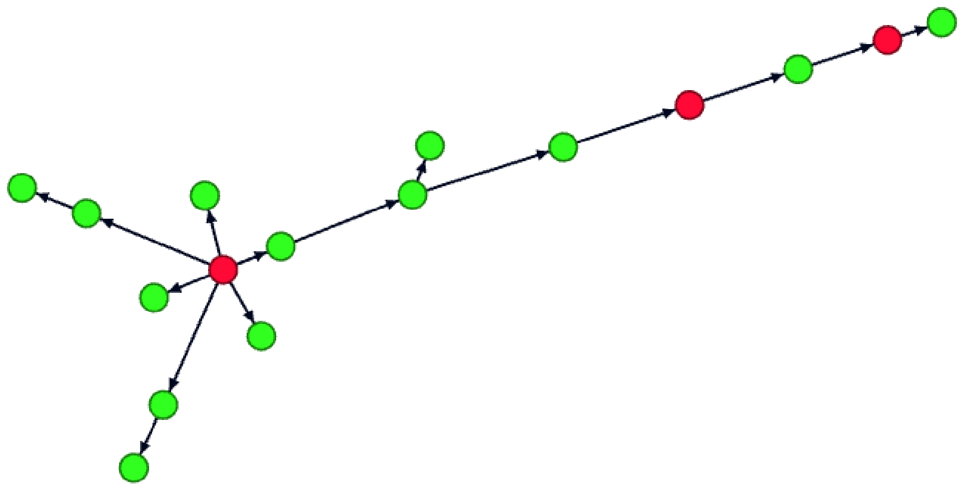


Figure 4.2: Conversation Graph (CG) for a discussion in Reddit. Green color indicates positive "+" comments and red color indicates negative "-" comments.

not from the stance of the comments. For instance, consider the following comment from the discussion in Figure 4.1,

> "*It is already open for prayer. Which doesn't hurt the museum part. Let's keep it as is. Open in it up for Friday and Eid prayers. Also for Christmas and other main Christian gatherings. I don't see why all three functions can't coexist.*" **Score**: 7

The content of the comment can be either positive, negative or neutral based on the topic of the conversation. The votes of the readers can annotate it either as positive, negative or neutral. The comment's score is 7 that is, readers considered it targeted. We assume that this comment is positive and the contributed users agree. However, in the case of the following comment,

> "*I did not read. Apparently, he was the leader of both Turkish and Armenian Catholics.*" **Score**: -5

Similarly, this comment could be either positive, negative or neutral based on the topic of discussion. The readers did not consider it targeted and therefore voted against it. We assume that this comment is negative and the contributed users disagree. Having settled the matter of what we define as positive and negative comment, we continue with the construction of AUG signed (or unsigned) graphs. An AUG graph can be generated either for one (intra$-$polarization) or more (inter$-$polarization) communities. In particular, an AUG graph is constructed using information from a

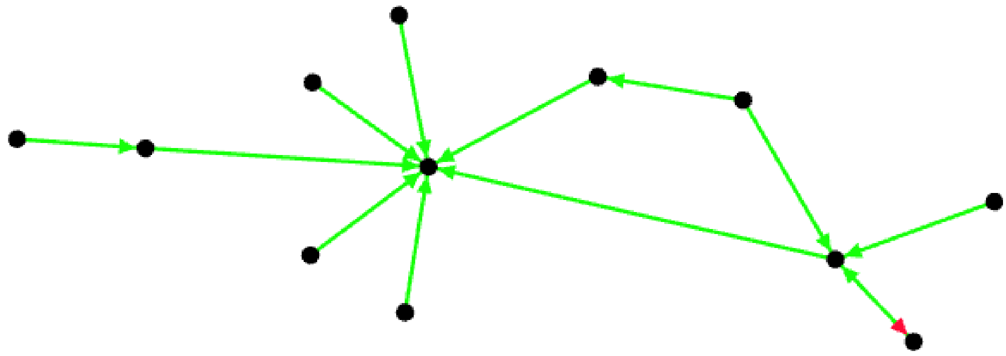


Figure 4.3: Aggregated User Conversation Directed Multi-Edge Graph (AUG_d) for a discussion in Reddit. Green edges indicates agreement and red edges indicates disagreement. (Note: The multi−edge are indistinguishable because of the tool that has been used for ploting.)

set of posts that belong to one or more subreddits. Both directed and non−directed signed (or unsigned) graphs are created for the needs of this thesis. Specifically, directed signed (or unsigned) AUG_d graphs used to extract motifs. Moreover, undirected unsigned AUG_u graphs are used to quantify unsigned polarization and undirected signed AUG_u graphs are used to detect the two most polarized communities.

Initially, let $AUG_d(V, E)$ be a directed signed multi−edge graph with V nodes and E total edges. The nodes declare the users who participate in the discussions of the subreddit(s). There is a directed edge $(i, j) \in E, i, j \in V$ from node v_j to v_i, if v_j replies to v_i. Each edge consists of a set of three features that is, $(t_{ij}, score_{Pushshift}, score_{Praw})$. The t_{ij} feature declares a timestamp that indicates the time at which the response from v_j added in the conversation. Next, $score_{Pushshift}$ and $score_{Praw}$ features declare the score of the comment from user v_j to user v_i for two distinct timestamps, the first one using Pushshift API and the second one using Praw API (more details in Chapter 7). So far, no sign has been assigned to the edges. Our AUG graph construction approach is based on the idea of converting the comment's score into a positive "+" or negative "-" sign. If comment's score is higher than zero then $sign_{Pushshift}$ or $sign_{Praw}$ is "+" (positive) respectively, otherwise, if comment's score is less than or equal to zero then the sign either $sign_{Pushshift}$ or $sign_{Praw}$ is "-" (negative). Consequently, the final set of three features of each edge has the following format $(t_{ij}, sign_{Pushshift}, sign_{Praw})$. In Figure 4.3, we present an example of a directed signed multi−edge graph AUG_d graph.

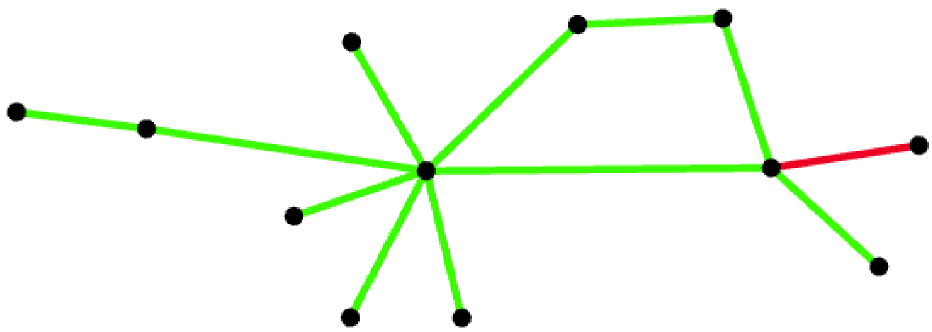


Figure 4.4: Aggregated User Conversation Undirected Graph (AUG_u) for a discussion in Reddit. Green edges indicates agreement and red edges indicates disagreement.

Following the same guidelines, we consider an undirected signed aggregated user graph $AUG_u(V, E)$ with V nodes and E total edges. An AUG_u can be generated either for one, two or more subreddits. Similarly, the nodes in AUG_u graph declare the users who participate in the discussions of the subreddit(s). There is an undirected edge $(v_i, v_j) \in E$ between v_i and v_j, either v_i replied to v_j or v_j replied to v_i. Each edge consists of a set of three features that is, $(t_{ij}, sign_{Pushshift}, sign_{Praw})$ where t_{ij} feature declares a timestamp that indicates the time at which the response from v_j added in the conversation and $sign_{Pushshift}$ and $sign_{Praw}$ features declare the edge sign based on Pushshift and Praw APIs. Since the edges are undirected, the value of features $sign_{Pushshift}$ and $sign_{Praw}$ can be considered as a process of merging directed multi$-$edges from AUG_d into an undirected edge. There are several merging techniques (edge contraction process) in which someone can merge multi$-$edges into an edge such as, random selection between multiple edges or select the first or the last appeared edge from the set of multiple edges or combine the edge weights of multiple edges into a single edge weight etc. This current work assumes the following technique of merging multiple directed edges into an undirected one. If there is even an negative edge (sign "-") between two vertices, then the sign of the undirected edge will be "-". Moreover, timestamp t_{ij} gets the value of the first timestamp of the edge. In Figure 4.4, a representation of an AUG_u graph presented. A summary of all the generated graphs mentioned in Table 4.1.

At this point, we need to mention the difficulties during the AUG graph construction process. In a discussion there are many comments that have been deleted either by the users themselves or by the moderator of the communities. Therefore, we are faced with the manipulation of users who are not characterized by their unique *id* but with a common characterization, that of *deleted*. Thus, "*How should we handle deleted users*?" After experimental analysis, we concluded that removing them from the network does not affect critical changes in our final results and therefore, we end up ignoring them from AUG graphs. Furthermore, a second limitation is that the final AUG graph may not be connected. Therefore, we work with the maximum connected component from AUG_u graph and with the maximum strongly connected component from AUG_d graph.

Table 4.1: Summary of generated graphs

Symbols	**Description of symbols**
Positive node "+"	Node declares comments and indicated by plus sign "+" if comment's score is higher than zero.
Negative node "-"	Node declares comments and indicated by minus sign "-" if comment's score is less than or equal to zero.
$CG(V, V^+, V^-, E)$	Directed signed conversation graph for each post where V are signed vertices (comments) and E are edges. If comment's upvote score is higher than zero then the sign is "+", $v \in V^+$ otherwise, "-" $v \in V^-$. There is an edge between two vertices if it is a response to another comment.
Positive edge "+"	Edge (i, j) declares that either i or j user replies to the other one and indicated by plus sign if the score of the answer is above zero. Agreement is declared.
Negative edge "-"	Edge (i, j) declares that either i or j user replies to the other one and indicated by minus sign if the score of the answer is less than or equal to zero. Disagreement is declared.
$AUG_d(V, E)$	Directed multi-edge aggregated user conversation graph where V are vertices (users) and E signed edges. There is a signed edge between i and j user $(i,\ j,\ t_{ij},\ sign_{Pushshift},\ sign_{Praw})$ if i replies to j at t_{ij}. The sign can be "+" (agreement) or "-" (disagreement) if the comment upvote score is above or below zero respectively.
$AUG_u(V, E)$	Undirected aggregated user conversation graph where V are vertices (users) and E signed edges. There is a signed edge between i and j user $(i,\ j,\ t_{ij},\ sign_{Pushshift},\ sign_{Praw})$ if i replies to j at t_{ij}. The sign can be "+" (agreement) or "-" (disagreement) if the comment upvote score is above or below zero respectively. The sign is "-" if there is at least one negative interaction between them.

Chapter 5

Algorithms

This section describes all the algorithms that have been used to quantify unsigned and signed polarization within a community ($intra-polarization$) and cross communities ($inter-polarization$) in a network. Particularly, we mention five algorithms that focus on the structure of the network and quantify unsigned polarization namely, Random Walks, Betweenness and Embeddings by [1], Boundary Connectivity by [12] and Dipole Moment by [21]. Afterwards, we mention five algorithms that quantify signed polarity in a signed graph, namely Eigensign, Random−Eigensign, Greedy, Bansal and LocalSearch by [2]. Also, the algorithm for counting temporal motifs by [16] is marked out. A $pseudo-code$ for each one of the above methods added.

5.1 Measuring unsigned polarity

This section reports algorithms from the literature that quantify polarization in a network looking only at the structure of the graph. The common denominator of the methods is that there is an undirected graph $G(V, E)$ with $|V|$ nodes and $|E|$ edges. The nodes represent users and there is an edge between two nodes if two users interact with each other. The nodes in G are divided into two disjoint sets (partitions), let them be X and Y. These two partitions can be derived from various clustering algorithms. In [1], the authors claims that METIS clustering [24] (see Appendix A.1) is a quite promising clustering technique that can split the graph in such a way that polarized groups emerge. The use of Real groups that Reddit provides is an another clustering technique that can be applied.

5.1.1 Random Walks

This measure uses the notion of random walks on graphs. It is based on the rationale that, in a controversial discussion, there are authoritative users on both sides, as evidenced by a large degree in the graph. The measure captures the intuition of "*How likely a random user on either side is to be exposed to authoritative content from the opposing side*". The process can be divided into three steps:

1. Select one partition at random.
2. A random walk starts from a random vertex in that partition.
3. The walk terminates when it visits any high−degree vertex (from either side).

The *Random Walk* polarization measure is quantified as

$$Random\ Walk\ (D) = P_{XX}P_{YY} - P_{YX}P_{XY} \tag{5.1}$$

where $P_{AB} \in \{X, Y\}$ is the conditional probability

$$P_{AB} = Prob[start\ in\ partition\ A|\ end\ in\ partition\ B] \tag{5.2}$$

If the two partitions are well−separated (more polarized), then the probability of crossing partitions is low, $P_{XY} \approx 0.0$ or $P_{YX} \approx 0.0$ and the probability crossing within the partition is high, $P_{XX} \approx 1.0$ or $P_{YY} \approx 1.0$. Therefore, the polarization in network is high ≈ 1.0. On the other hand, if the two partitions are not well−separated (less polarized), the probability of crossing partitions is high, $P_{XY} \approx 1.0$ or $P_{YX} \approx 1.0$ and the probability crossing within the partition is low, $P_{XX} \approx 0.0$ or $P_{YY} \approx 0.0$. So, the polarization in network is low ≈ 0.0.

A variation of this algorithm is to make a change on when the random walk ends (Step 3). Instead of ending in a high−degree vertex (from either side), it can terminate at a set of p% of nodes (from either side). The process remains the same, we symbolize the variation as $Random\ Walk\ (P)$. The pseudo − code of *Random Walks* unsigned polarization mentioned in Algorithm 5.1.

Algorithm 5.1 *Random Walk* unsigned polarization

Require: Undirected graph $G(V, E)$, set of nodes of the partition X and Y

Ensure: Polarization score (PS)

1: *// Case Random Walks (D)*
2: $topDegreeNodes_X \leftarrow u \in X$ and u has high degree
3: $topDegreeNodes_Y \leftarrow u \in Y$ and u has high degree
4: *// Case Random Walks (P)*
5: $percentageNodes_X \leftarrow$ randomly selected p% nodes of X
6: $percentageNodes_Y \leftarrow$ randomly selected p% nodes of Y
7: Apply Random Walk process starting from $topDegreeNodes_X$ or $percentageNodes_X$ nodes
8: Apply Random Walk process starting from $topDegreeNodes_Y$ or $percentageNodes_Y$ nodes
9: $AB \leftarrow$ Number of nodes that starts from partition A and ends up in partition B, $A, B \in \{X, Y\}$
10: $e_1 \leftarrow \frac{XX}{XX+YX}$, $e_2 \leftarrow \frac{XY}{XY+YY}$, $e_3 \leftarrow \frac{YX}{XX+YX}$ and $e_4 \leftarrow \frac{YY}{XY+YY}$
11: **PS** $\leftarrow e_1 * e_4 - e_2 * e_3$

5.1.2 Betweenness

This measure uses the notion of edge betweenness and how the betweenness of the cut differs from that of the other edges. Let consider $C \subseteq E$ in the cut defined by the two partitions X and Y. The betweenness centrality $bc(e)$ of an edge $e \in E$, is defined as

$$bc(e) = \sum_{s \neq t \in V} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}},\ and\ bc(e) \in [0, 1] \tag{5.3}$$

where $\sigma_{s,t}$ is the total number of shortest paths between nodes s and t in the graph and $\sigma_{s,t}(e)$ the number of those shortest paths that include edge e. If the partition X and Y are well−separated (more polarized) then betweenness of the edges in C and $bc(e \in C)$ takes values close to one this happens because the shortest paths that connect vertices of the two partitions will pass through the edges in the cut. If the two partitions are not well−separated (less polarized) the numerator ($\sigma_{s,t}(e)$) takes low values (close to zero) and the denominator ($\sigma_{s,t}$) high values. Then, $bc(e \in C)$ ranges at low values and this is happening because the paths that connect vertices across two partitions passes through one of the many edges in the cut. The $Betweenness$ polarization measure is quantifies as

$$Betweenness = 1 - e^{-d_{KL}} \tag{5.4}$$

where d_{KL} is the Kullback−Leibler Divergence of the distributions of size N of edge betweenness on the cut and the rest of the graph. Let p be the distribution of the betweenness centrality of edges on the cut and q the distribution of the betweenness centrality of edges on the rest graph. Then $d_{KL}(p, q)$ is defined as

$$d_{KL}(p, q) = \sum_{i=1}^{N} p(x_i)[log(p(x_i)) - log(q(x_i))] \tag{5.5}$$

By definition the cut edges have higher betweeness centrality therefore, $p(x_i), i \in [1, N]$ takes values close to one and $log(p(x_i)) \approx 0.0$. On the contrary, the rest of the edges (not cut edges) have betweeness centrality $q(x_i), i \in [1, N]$ close to zero therefore, $log(q(x_i)) \approx -\infty$. If distributions p and q are similar then d_{KL} distance takes values close to zero because $p(x_i)$ and $q(x_i)$ ranges close to zero. Then, $e^{-d_{KL}} \approx e^0 = 1$ ranges close to one and finally, $Betweenness \approx 0.0$. Otherwise, if distributions p and q differs the d_{KL} distance takes values close to $+\infty$. Therefore, $e^{-d_{KL}} \approx e^{-\infty} = -\infty$ and $Betweenness \approx 1.0$. The pseudo − code of *Betweenness* unsigned polarization mentioned in Algorithm 5.2.

5.1.3 Embeddings

This measure uses the notion of how short or long is the distance of the nodes based on their node two − dimensional embedding by Gephi's ForceAtlas2 algorithm [29]. Let consider two partitions X and Y of the graph. The $Embeddings$ polarization measure is quantified as

$$Embeddings = 1 - \frac{d_X + d_Y}{2d_{XY}} \tag{5.6}$$

where d_X and d_Y is the average embedded distance among pairs of vertices in the same partition, X and Y respectively and d_{XY} is the average embedded distance

Algorithm 5.2 *Betweenness* unsigned polarization

Require: Undirected graph $G(V, E)$, set of nodes of the partition X and Y

Ensure: Polarization score (PS)

1: $totalEdges \leftarrow E$

2: $cutEdges \leftarrow (u, v) \in E,\ u \in X\ and\ v \in Y\ or\ u \in Y\ and\ v \in X$

3: $restEdges \leftarrow totalEdges - cutEdges$

4: betweenness centrality for set of edges: $betweenness_{totalEdges}$, $betweenness_{cutEdges}$ and $betweenness_{restEdges}$

5: $KL \leftarrow$ Kullback−Leibler distance between $betweenness_{cutEdges}$ and $betweenness_{restEdges}$

6: **PS** $\leftarrow 1 - e^{-KL}$

among pairs of vertices across the two partitions X and Y. If the two partitions are well−separated (more polarized), then the two quantities d_X and d_Y ranges close to zero and d_{XY} close to one. So, the polarity score approaches one, $Embeddings \approx 1.0$. On the other hand, if the two partitions are not well−separated (less polarized), then the two quantities d_X and d_Y ranges close to one and the polarity score approaches zero, $Embeddings \approx 0.0$. The pseudo − code of *Embeddings* unsigned polarization mentioned in Algorithm 5.3.

5.1.4 Boundary Connectivity

This measure uses the notion of the connectivity of boundary nodes in a graph. That is, let consider two partitions X and Y of the graph, then boundary vertices will be more strongly connected to internal vertices than to other boundary vertices of either partition. Let $u \in X$ be a vertex in partition X, u belongs to the boundary of X if and only if 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. Following this definition, let B_X, B_Y be the set of boundary vertices for each partition, and $B = B_X \cup B_Y$ the set of all boundary vertices. By contrast, vertices $I_X = X - B_X$ are said to be the internal vertices of partition X and $I_Y = Y - B_Y$ for partition Y respectively. Also, let $I = I_X \cup I_Y$ be all internal vertices in either partition. The $Boundary\ Connectivity$ ($GMCK$) polarization measure is quantified as

$$GMCK = \frac{1}{|B|} \sum_{u \in B} [\frac{d_i(u)}{d_b(u) + d_i(u)} - 0.5] \tag{5.7}$$

Algorithm 5.3 *Embeddings* unsigned polarization

Require: Undirected graph $G(V, E)$, set of nodes of the partition X and Y

Ensure: Polarization score (PS)

```
Apply Forceatlas2 embeddings to G
Let e_u = [x_u, y_u] be the coordinates of u ∈ V and x_u, y_u ∈ R
Dist_X ← 0, Dist_Y ← 0, Dist_XY ← 0
for u ∈ X(or Y) and v ∈ X(or Y) − {u} do
    Dist_X ← Dist_X + sqrt((x_u − x_v)^2 + (y_u − y_v)^2)
    Dist_Y ← Dist_y + sqrt((x_u − x_v)^2 + (y_u − y_v)^2)
end for
for For u ∈ X and v ∈ Y do
    Dist_XY ← Dist_XY + sqrt((x_u − x_v)^2 + (y_u − y_v)^2)
end for
t_X ← |X|(|X|−1)/2, t_Y ← |Y|(|Y|−1)/2 and t_XY ← |X| * |Y|
d_X ← Dist_X/t_X, d_Y ← Dist_Y/t_Y and d_XY ← Dist_XY/t_XY
PS ← 1 − (d_X + d_Y)/(2 * d_XY)
```

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. If the two partitions are well−separated (more polarized), the quantity $d_i(u)$ takes high values and $d_b(u)$ low values. Therefore, the quantity of the sum will be close to 0.5. So, the polarity $GMCK$ score ranges in high values, $GMCK \approx 1.0$. Otherwise, if the two partitions are not well−separated (less polarized), the quantities $d_i(u)$ and $d_b(u)$ take high and low values respectively. And consequently, the polarity $GMCK$ score ranges in low values, $GMCK \approx 0.0$. The pseudo − code of *GMCK* unsigned polarization mentioned in Algorithm 5.4.

5.1.5 Dipole Moment

This measure uses the notion of dipole moment that has its origin in physics. In physics, the electric dipole moment is a measure of the separation of positive and negative electrical charges within a system, that is, a measure of the system's overall polarity. Let consider two partitions X and Y of the graph and $R(u) \in [-1.0, 1.0]$ be a polarization value assigned to vertex $u \in V$. Intuitively, extreme values of R (close

Algorithm 5.4 *Boundary Connectivity (GMCK)* unsigned polarization

Require: Undirected graph $G(V, E)$, set of nodes of the partition X and Y

Ensure: Polarization score (PS)

$B_X \leftarrow u \in X$ and $\exists(u, v), v \in Y$ and $\exists(u, y), y \in X$ and $\nexists(y, x), x \in Y$
$B_Y \leftarrow u \in Y$ and $\exists(u, v), v \in X$ and $\exists(u, y), y \in Y$ and $\nexists(y, x), x \in X$
$B \leftarrow B_X \cup B_Y$
$I_X \leftarrow X - B_X$, $I_Y \leftarrow Y - B_Y$ and $I \leftarrow I_X \cup I_Y$
$score \leftarrow 0$
for $u \in B$ **do**
    $d_i(u) \leftarrow$ Number of edges $(u, v), v \in I$
    $d_b(u) \leftarrow$ Number of edges $(u, v), v \in B$
    $score \leftarrow score + \frac{d_i(u)}{d_b(u)+d_i(u)} - 0.5$
end for
PS $\leftarrow \frac{score}{|B|}$

to −1.0 or 1.0) correspond to users who belong most clearly to either side of the controversy. Set $R = \pm 1$ for the top 5% highest−degree vertices in each partition X and Y, and set the values for the rest of the vertices by label−propagation. Also, 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 $\Delta A = |\frac{n^+ - n^-}{|V|}|$. Moreover, let gc^+ and gc^- be the average polarization value among vertices n^+ and n^- respectively. Also, set d as half their absolute difference, $d = \frac{|gc^+ - gc^-|}{2}$. The polarization measure $Dipole\ Moment$ $(MBLB)$ is defined as

$$MBLB = (1 - \Delta A)d \tag{5.8}$$

If the two partitions X and Y are well−separated (more polarized), then label propagation will assign different extreme R−values to the two partitions, leading to higher values of the $MBLB$ measure, $MBLB \approx 1.0$. Otherwise, if the two partitions X and Y are not well−separated (less polarized), then $MBLB$ measure ranges close to zero, $MBLB \approx 0.0$. The pseudo − code of ***MBLB*** unsigned polarization mentioned in Algorithm 5.5.

Algorithm 5.5 *Dipole Moment (MBLB)* unsigned polarization

Require: Undirected graph $G(V, E)$, set of nodes of the partition X and Y

Ensure: Polarization score (PS)

```
topDegreeNodes_X ← u ∈ X and u has high degree
topDegreeNodes_Y ← u ∈ Y and u has high degree
topNodes ← topDegreeNodes_X ∪ topDegreeNodes_Y
for each node u ∈ topNodes do
    if u ∈ topDegreeNodes_X then
        opinion_u ← 1
    end if
    if u ∈ topDegreeNodes_Y then
        opinion_u ← −1
    end if
end for
// Apply Label−Propagation method to the nodes u ∈ {V − topNodes}
for u ∈ {V − topNodes} do
    opinion_u ← Average of neighbors opinion
end for
n⁺ ← Number of nodes u ∈ V with positive opinion value
n⁻ ← Number of nodes u ∈ V with negative opinion value
ΔA ← |(n⁺ − n⁻)/|V||
gc⁺ ← Average opinion value among vertices n⁺
gc⁻ ← Average opinion value among vertices n⁻
d ← |gc⁺ − gc⁻|/2
PS ← (1 − ΔA)d
```

5.2 Measuring signed polarity

This section describes the signed polarity measures used in this thesis. The common denominator of the following five methods is that there is an undirected signed graph $G = (V, E, E^+, E^-)$ with $|V|$ nodes, $|E|$ total edges, E^+ edges with positive sign (friendly interaction) and E^- edges with negative sign (antagonistic interaction). The nodes represent users and there is an edge between two nodes if the two users have interacted with each other. The main purpose of the application of the following spectral methods is to discover two polarized communities (subsets of the network vertices) S_1 and S_2 where within communities there are mostly positive edges while across communities there are mostly negative edges. The formal definition of the problem is defined as follows,

> *Given a signed network $G = (V, E^+, E^-)$ with V vertices and signed adjacency matrix A, find a vector $x \in \{-1, 0, 1\}^{|V|}$ that maximizes $\frac{x^T\ A\ x}{x^T\ x}$ and $x^T\ A\ x = cc(S_1, S_2)$ where $cc(S_1, S_2)$ is the maximization function. This metric is called polarity.*

The maximization function $cc(S_1, S_2)$ for $i \in \{1, 2\}$ is defined as

$$cc(S_1, S_2) = \sum_{(u,v) \in S_i\ x\ S_i} \frac{1}{2}[1_{E^+}(u, v) - 1_{E^-}(u, v)] + \sum_{(u,v) \in S_1\ x\ S_2} [1_{E^-}(u, v) - 1_{E^+}(u, v)] \quad (5.9)$$

where 1_S is the indicator function of the set S. The indicator function is a function defined on a set S that indicates membership of an element in a subset $A \in \{E^+, E^-\}$ of S, having the value 1 for all elements of A and the value 0 for all elements of S not in A. That is,

$$1_S(x) := \left\{ \begin{array}{l} 1\ , if\ x \in S \\ 0\ , if\ x \notin S \end{array} \right\} \quad (5.10)$$

Moreover, the adjacency matrix A of the signed network $G = (V, E^+, E^-)$, where positive edges $(i, j) \in E^+$ are indicated by $A_{ij} = 1$, negative edges $(i, j) \in E^-$ are indicated by $A_{ij} = -1$, and non$-$edges are indicated by $A_{ij} = 0$. A partition S_0 (neutral community), S_1 (first polarized community) and S_2 (second polarized community) of V can be represented by a vector $x \in \{-1, 0, 1\}^{|V|}$, whose $i-$th coordinate is $x_i = 0$ if $i \in S_0$, $x_i = 1$ if $i \in S_1$, and $x_i = -1$ if $i \in S_2$. Also, as polarity is penalized with the size of the solution, vertices are only added to one of the two clusters if they contribute significantly to the objective.

Algorithm 5.6 *Eigensign* signed polarization

Require: Adjacency matrix A of signed graph $G(V, E, E^+, E^-)$

Ensure: Polarity score (PS)

Compute v, the eigenvector corresponding to the largest eigenvalue λ_1 of A
Set an empty list $x \leftarrow [\,]$
for $u \in V$ **do**
    Append to list x the sign of v_u
end for
PS $\leftarrow \frac{x^T\, A\, x}{x^T\, x}$

5.2.1 Eigensign

The spectral and deterministic *Eigensign* method works by simply discretizing the entries of the eigenvector of the adjacency matrix corresponding to the largest eigenvalue. The implementation of the algorithm is based on the idea of the Pick−an−edge problem. That is, picking an arbitrary edge, if it is positive, put the endpoints in one cluster, leaving the other cluster empty, if it is negative, put the endpoints in separate clusters. Therefore, given the adjacency matrix A, computes v, the eigenvector corresponding to the largest eigenvalue λ_1 of A. The construction of vector x of size $|V|$ is based on the sign of each element of the eigenvector. That is, for each node $i \in \{1, 2, \ldots, |V|\}$, $x_i = sign(v_i)$. The *Eigensign* algorithm generally outputs a solution comprised of all the vertices in the graph, unless some components of the eigenvector are exactly zero, which is, of course, counter to the motivation of the problem setting. The pseudo − code of *Eigensign* signed polariaztion mentioned in Algorithm 5.6

5.2.2 Random Eigensign

The *Random Eigensign* algorithm is a randomized algorithm that solves the weakness of the *Eigensign* algorithm that is, the vector x is comprised of all the vertices in the graph, unless some components of the eigenvector are exactly zero. More specifically, given the adjacency matrix A, computes the first eigenvector v corresponding to the largest eigenvalue λ_1. Instead of simply discretizing the entries of v, it randomly sets each entry of x to 1 or -1 with probabilities determined by the entries of v. That is, for each $i \in \{1, 2, \ldots, |V|\}$, a Bernoulli experiment runs with success probability $|v_i|$. If it succeeds, then $x_i = sign(v_i)$, otherwise $x_i = 0$. The entries v_i with large magnitude

Algorithm 5.7 *Random Eigensign* signed polarization

Require: Adjacency matrix A of signed graph $G(V, E, E^+, E^-)$

Ensure: Polarity score (PS)

```
Compute v, the eigenvector corresponding to the largest eigenvalue λ_1 of A
Set an empty list x ← [ ]
for u ∈ V do
    Run a Bernoulli experiment with success probability |v_u|
    if experiment succeeds then
        Append to list x the sign of v_u
    else
        Append to list x zero
    end if
end for
PS ← (x^T A x) / (x^T x)
```

$|v_i|$ are more likely to turn into $sign(v_i)$, while entries v_i with small magnitude $|v_i|$ are more likely to turn into 0. The pseudo $-$ code of *Random Eigensign* signed polarization mentioned in Algorithm 5.7

5.2.3 Greedy

The optimization problem of finding a subgraph of maximum density according to the notion of Kannan and Vinay has been studied by [30]. The authors introduce a notion of density for directed graphs that quanties relatively highly connected and is suitable for sparse directed graphs such as the web graph. This study tries to formalize the notion of finding sets of hubs and authorities that are highly connected relative to the rest of the graph. A greedy 2$-$approximation algorithm for undirected densest subgraph is proposed. The main idea is to produce a subgraph of $G(V, E)$ of large average degree. Intuitively, the vertices with low degree throw away in order to produce such a subgraph. In more details, the algorithm maintains a subset $S \subseteq V$ of vertices. Initially, set S contains the vertices of V. In each iteration, the algorithm identies i_{min}, the vertex of minimum degree in the subgraph induced by S. In the sequel, i_{min} is removed from the set S and moves on to the next iteration. This loop process stops when the set S is empty. Of all the sets S constructed during the

Algorithm 5.8 *Greedy* signed polarization

Require: Signed graph $G(V, E, E^+, E^-)$

Ensure: Polarity score (PS)

$totalNodes \leftarrow V$, $subgraph_{best} \leftarrow G$ and $polarity_{best} \leftarrow 0$

while $totalNodes$ not empty **do**

  $v_{min} \leftarrow$ Node that minimize the difference between the number of positive adjacent edges and the number of negative adjacent edges

  $subgraph \leftarrow$ Remove from G node v_{min}

  Remove from $totalNodes$ node v_{min}

  $A_{subgraph} \leftarrow$ Adjacency matrix of $subgraph$

  Compute v, the eigenvector corresponding to the largest eigenvalue λ_1 of $A_{subgraph}$

  Set an empty list $x \leftarrow [\,]$

  for $u \in V_{subgraph}$ **do**

    Append to list x the sign of v_u

  end for

  $polarity_{subgraph} \leftarrow \frac{x^T\ A_{subgraph}\ x}{x^T\ x}$

  if $polarity_{subgraph} > polarity_{best}$ **then**

    $subgraph_{best} \leftarrow subgraph$

    $polarity_{best} \leftarrow polarity_{subgraph}$

  end if

end while

PS $\leftarrow polarity_{best}$

execution of the algorithm, the set S maximizing density $f(S)$ function (i.e. the set of maximum average degree) is returned as the output of the algorithm. The density function $f(S)$ is defined as

$$f(S) = \frac{|E(S)|}{|S|} \tag{5.11}$$

where $E(S)$ is the set of edges included by S, $E(S) = \{ij \in E : i \in S, j \in S\}$. Therefore, the density $f(G)$ of the undirected graph $G(V, E)$ is defined as the optimization problem $f(G) = max\{f(S)\}, S \subseteq V$.

According to this greedy $2-$approximation approach by [30], a variation has been proposed by [2] so that to find two polarized communities in a signed graph. Conse-

Algorithm 5.9 *Bansal* signed polarization

Require: Adjacency matrix A of signed graph $G(V, E, E^+, E^-)$
Ensure: Polarity score (PS)

$polarity_{best} \leftarrow 0$
for $u \in V$ **do**
  $S_1 \leftarrow$ Vertices sharing a positive edge with u
  $S_2 \leftarrow$ Vertices sharing a negative edge with u
  Set an empty list $x \leftarrow [\,]$
  for $v \in V$ **do**
    if $v \in S_1$ or $v = u$ **then**
      Append to list x the value 1
    else
      Append to list x the value -1
    end if
  end for
  $polarity \leftarrow \frac{x^T\,A\,x}{x^T\,x}$
  if $polarity > polarity_{best}$ **then**
    $polarity_{best} \leftarrow polarity$
  end if
end for
PS $\leftarrow polarity_{best}$

quently, *Greedy* algorithm, iteratively removes the vertex minimizing the difference between the number of positive adjacent edges and the number of negative adjacent edges, up to when the graph is empty. At the end, it returns the subgraph having the highest polarity among all subgraphs visited during its execution. The assignment of the vertices to the clusters is guided by the sign of the components of the eigenvector v, corresponding to the largest eigenvalue of A. The pseudo $-$ code of *Greedy* signed polarization mentioned in Algorithm 5.8

5.2.4 Bansal

The method is motivated by Bansal's 3$-$approximation algorithm [31] for finding two polarized communities on a complete signed graph. The correlation clustering operates in a scenario where the relationships between the objects are known. Given a signed graph $G(V, E, E^+, E^-)$ of $|V|$ nodes, $|E|$ edges and E^+, E^- the positive and negative edges respectively. The edge sign ("+" or "-") indicates the similarity between the nodes that are located in the arcs of the edge. If the sign is positive then the two nodes are similar, otherwise the two nodes are unlike. The task of the problem is to find a clustering that either maximizes agreements or minimizes disagreements.

According to this Bansal's $3-$approximation approach, a variation has been proposed by [2] so that to find two polarized communities. Consequently, $Bansal$ method, for each vertex $u \in V$ identifies u together with the vertices sharing a positive edge with u as one cluster, and the vertices sharing a negative edge as the other. Of these $|V|$ possible solutions, the one that maximize polarity is returned. The pseudo $-$ code of *Bansal* signed polarization mentioned in Algorithm 5.9

5.2.5 LocalSearch

This method is based on randomness. That is, let S_r be the initial set of vertices chosen at random. At each iteration, it adds (removes) to (from) the current solution the vertex that maximizes the gain in terms of polarity. Finally, the method terminates when the gain of moving any vertex is lower than 0.2. Also, the assignment of the vertices to the clusters is guided by the signs of v. The pseudo $-$ code of *LocalSearch* signed polarization mentioned in Algorithm 5.10

Algorithm 5.10 *LocalSearch* signed polarization

Require: Adjacency matrix A of signed graph $G(V, E, E^{+}, E^{-})$
Ensure: Polarity score (PS)

$S_1 \leftarrow$ Random vertices from V, $S_2 \leftarrow$ Random vertices from V
Set an empty list $x_{start} \leftarrow [\,]$
for $u \in V$ **do**
  if $u \in S_1$ **then**
    Append to list x_{start} the value 1
  else if $u \in S_2$ **then**
    Append to list x_{start} the value -1
  else
    Append to list x_{start} the value 0
  end if
  $polarity_{best} \leftarrow \frac{x_{start}^{T}\, A\, x_{start}}{x_{start}^{T}\, x_{start}}$
end for
while True **do**
  $S_1' \leftarrow$ Add or remove vertices, $S_2' \leftarrow$ Add or remove vertices
  Set an empty list $x_{new} \leftarrow [\,]$
  for $u \in V$ **do**
    if $u \in S_1'$ **then**
      Append to list x_{new} the value 1
    else if $u \in S_2'$ **then**
      Append to list x_{new} the value -1
    else
      Append to list x_{new} the value 0
    end if
  end for
  $polarity_{new} \leftarrow \frac{x_{new}^{T}\, A\, x_{new}}{x_{new}^{T}\, x_{new}}$
  if $|polarity_{new} - polarity_{best}| > 0.20$ **then**
    $polarity_{best} \leftarrow polarity_{new}$
  else
    Exit While
  end if
end while
PS $\leftarrow polarity_{best}$

5.3 Counting temporal motifs

Let consider a temporal directed graph $G(V, E)$ of $|V|$ nodes and $|E|$ totstartal edges. The nodes represent users and there is one or more edges between two nodes if they interact. A temporal edge is referred as a tuple of three inputs that is, (u_i, v_i, t_i) where $u_i, v_i \in V, i \in |E|$ and t_i is the timestamp of the interaction. Furthermore, temporal motifs are an ordered sequence of timestamped edges conforming to a specified pattern as well as a specified duration of time δ in which the events must occur. So, $\delta-$temporal motif for $k-$node and $l-$edge is defined as a sequence of l edges, $M = (u_1, v_1, t_1), (u_2, v_2, t_2), \ldots, (u_l, v_l, t_l)$ that are time ordered within a δ duration, i.e., $t_1 < t_2 < \cdots < t_l$ and $t_l - t_1 \leq \delta$, such that the induced static graph from the edge is connected and has k nodes.

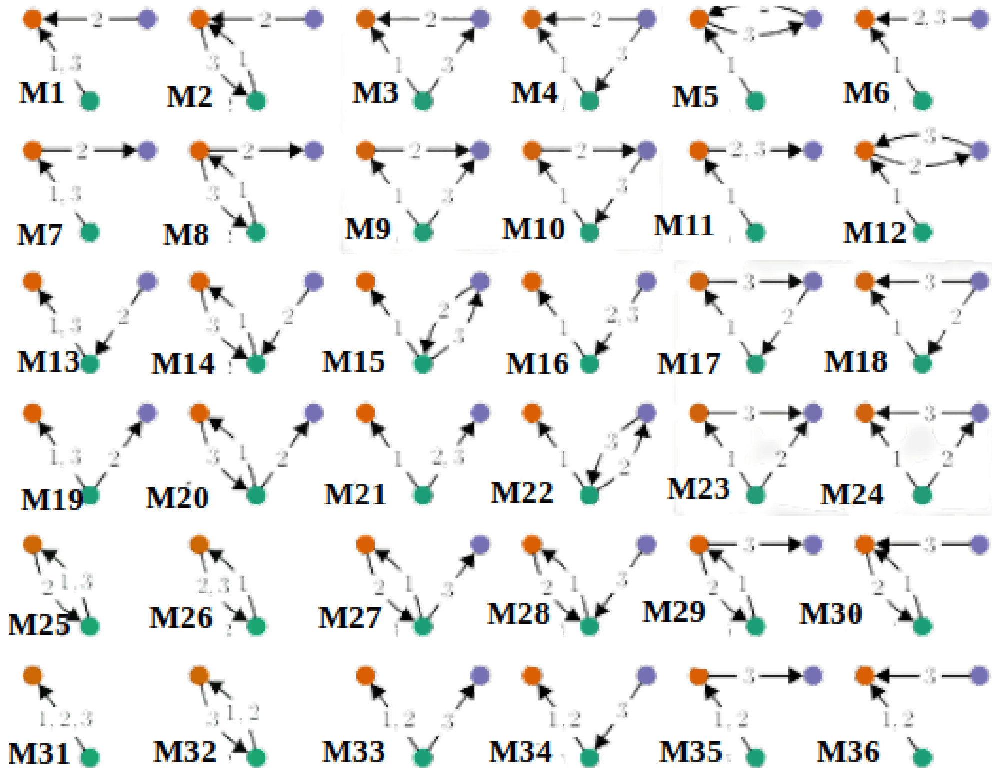


Figure 5.1: All 2−node and 3−node, 3−edge δ−temporal motifs. We index the 36 motifs **Mi**, $i \in [1, 36]$. **M25**, **M26**, **M31** and **M32** are the 2−node motifs and **M3**, **M4**, **M9**, **M10**, **M17**, **M18**, **M23** and **M24** are the eight triangles. The rest motifs are stars. The first edge in each motif is from the green to the orange node. The second edge is the same along each row, and the third edge is the same along each column.

A collection of edges in a given temporal graph is an instance of a $\delta-$temporal motif M if it matches the same edge pattern and all of the edges occur in the

Algorithm 5.11 Count the number of instances of all possible l−edge δ−temporal motifs in an ordered sequence of temporal edges. The keys of counts[·] are accessed in order of length.

Require: Sequence S_0 of edges $(e_1 = (u_1, v_1), t_1), \dots, (e_L = (e_L, t_L), t_L)$ with $t_1 < \dots < t_L$, time window δ

Ensure: Number of instances of each $l-$edge $\delta-$temporal motif M contained in S_0

$start \leftarrow 1$
$counts \leftarrow Counter$, (default $= 0$)
for $end = 1, \dots, L$ **do**
  while $t_{start} + \delta < t_{end}$ **do**
    //DecrementCounts process
    $counts[e_{start}] -= 1$
    for suffix $\in$ counts.keys of length $< l - 1$ **do**
      $counts[concat(e_{start}, suffix)] -= counts[suffix]$
    end for
  end while
  //IncrementCounts process
  for prefix $\in$ counts.keys.reverse() of length $< l$ **do**
    $counts[concat(prefix, e_{end})] += counts[prefix]$
  end for
  $counts[e_{end}] += 1$
end for
Return $counts$

right order within the δ time window. Formally, any time−ordered sequence $S = (w_1, x_1, t'_1), \dots, (w_l, x_l, t'_l)$ of l unique edges is an instance of the motif $M = (u_1, v_1, t_1)$, (u_2, v_2, t_2), ..., (u_l, v_l, t_l) if there exists a bijection f on the vertices such that $f(w_i) = u_i$ and $f(x_i) = v_i$, $i \in 1, 2, \dots, l$ and the edges all occur within δ time i.e., $t'_l - t'_1 \leq \delta$. Therefore, basis of this definition, Algorithm 5.11 by [16] counts the number of ordered subsets of edges from temporal graph G that are instances of a particular motif. In this current thesis, we focus on 2−node and 3−node temporal motifs. All the possible 2−node and 3−node temporal motifs (36 in total) presented in Figure 5.1.

General outline for 2−node motifs. For each pair of nodes u and v for which there is at least one edge, gather and sort the edges in either direction between u and v. Then call Algorithm 5.11 with these edges.

General outline for 3−node star motifs. For each node u in static G graph and for each unique set of $k-1$ neighbors, gather and sort the edges in either direction between u and the neighbors. Then count the number of instances of M using Algorithm 5.11.

General outline for 3−node triangle motifs. Initially, a fast static graph triangle enumeration algorithm used to find all triangles in the static graph G induced by temporal G [32]. For each triangle (u, v, w), merge all temporal edges from each pair of nodes to get a time−sorted list of edges. Then, 5.11 is used to count the number of instances of M.

Chapter 6

General methodology

Chapter 3 reports the general problem that this current thesis tries to investigate. That is, to measure unsigned and signed polarization in one (intra−polarization) or cross more (inter−polarization) communities. Specifically, the target is to answer the next research questions, **RQ1.** Is intra−polarization or inter−polarizatio detected in Reddit? i.e., the members of a community or more communities discuss a common topic and two different opinions emerge, then the members within the community cut off the connections with each other if they disagree or continue to communicate expressing their opposite points of view. **RQ2.** Does controversy increase polarization? i.e., controversial posts are more prone to polarization than non−controversial ones. The last research query which we dealt with is **RQ3.** What are the common motifs of user interaction in the comments of a discussion in the case of controversial and non−controversial posts? i.e., a user responds to a comment from another user then what is the probability of replying back in controversial and non−controversial posts or when a negative comment preceded in a discussion what is the probability of following negative comments etc.

Therefore, Chapter 6 describes the general methodology for approaching the above questions namely, **RQ1**, **RQ2** and **RQ3.** Particularly, the algorithm for measuring

Algorithm 6.1 General methodology for measuring unsigned polarization

Require: AUG_u unsigned graph $G(V, E)$ for one, two or more subreddits
Ensure: Polarization score (PS)
1: *//Initialize two polarized groups*
2: **if** One input subreddit (intra−polarization) **then**
3: $A, B \leftarrow$ Apply METIS partition
4: **else if** Two input subreddits (inter−polarization) **then**
5: $A, B \leftarrow$ Apply METIS partition
6: $X, Y \leftarrow$ Apply Real group partition
7: **else**
8: Generate all 2-subreddits AUG_u unsighed graphs among potential subreddits
9: Go to Step 3 for each 2-subreddits AUG_u combination of graphs
10: **end if**
11: *//Apply unsigned algorithms using METIS and Real groups clustering*
12: PS$_{METIS}$ $\leftarrow$ Apply Algorithms *Random Walks (P)*, *Random Walks (D)*, *Betweenness*, *Embeddings*, *GMCK* and *MBLB* using METIS clustering (apply A and B polarized groups)
13: PS$_{Real}$ $\leftarrow$ Apply Algorithms *Random Walks (P)*, *Random Walks (D)*, *Betweenness*, *Embeddings*, *GMCK* and *MBLB* using Real group clustering (apply X and Y polarized groups)

unsigned polarization (intra and inter), the algorithm for detecting the two most polarized subsets of users in a signed graph that is, signed polarization (intra and inter) and the algorithm for counting temporal motifs described below.

6.1 Measuring unsigned polarity

The process of quantifying unsigned polarization either in one community (intra − polarization) or cross more communities (inter − polarization) uses a 2−step pipeline. The pseudo − code of the methodology displayed in Algorithm 6.1. The first stage of the pipeline (Steps 1-9) is the partitioning of AUG undirected graph into two subgroups. If the input AUG_u graph concerns one subreddit (Steps 1-2) then only METIS clustering can be applied. The Real groups clustering is essentially all the

Algorithm 6.2 General methodology for measuring signed polarization

Require: AUG_u signed graph $G(V, E, E^+, E^-)$ for one, two or more subreddits

Ensure: Polarization score (PS), Two polarized communities S_1 and S_2

1: **PS**, $S_1, S_2 \leftarrow$ Apply Algorithms *Eigensign*, *Random Eigensign*, *Greedy*, *Bansal* and *LocalSearch*.

2: *// Analysis of polarized communities* S_1 *and* S_2

3: Relation of the size of S_1 and S_2 in relation to AUG_u size

4: Percentage of positive and negative edges within and cross S_1 and S_2

5: Percentage of vertices from each subreddit in S_1 and S_2

6: Percentage of vertices that participate in both input subreddits

members of the input subreddit as collected by Reddit. That is, if the two input subreddits are *Armenia* & *Azerbaijan* then the Real group of Armenia are all those users who participate in conversations in Armenia subreddit and Real group of Azerbaijan consists of the users who participate in submissions from group Azerbaijan. In the case of $2-$subreddits input AUG_u graph (Steps 3-5) we apply both METIS and Real groups clustering. Also, in the case of more than $2-$subreddits that is, $k-$subreddits (Steps 6-8) then all $2-$subreddits AUG_u graphs are generated among potential subreddits. Therefore, METIS and Real groups clustering is applied in each one of the generated $2-$subreddits AUG_u graphs. The second stage of the pipeline is the implementation of unsigned algorithms 5.1 in the produced graphs. Two unsigned polarization scores namely, $\mathbf{PS}_{METIS}$ and $\mathbf{PS}_{Real}$ are calculated (Steps 10-11) for each one of the unsigned algorithms.

6.2 Measuring signed polarity

The process of measuring signed polarization and detecting the two most polarized communities either in one community (intra $-$ polarization) or cross more communities (inter $-$ polarization) uses a $2-$step pipeline. The pseudo $-$ code of the methodology displayed in Algorithm 6.2. The first stage of the pipeline (Step 1) is the detection of two polarized communities S_1 and S_2 by applying one of the proposed algorithms in Section 5.2 *Eigensign*, *Random Eigensign*, *Greedy*, *Bansal* and *LocalSearch*. Thereinafter, in the second phase (Steps 3-6) we analyze the correlation

Algorithm 6.3 General methodology for counting temporal motifs

Require: AUG_d unsigned (or signed) graph $G(V, E, E^+, E^-)$ for one, two or more subreddits

Ensure: Number of each possible 2−node, 3−node star and 3−node triangle motifs

1: $\delta \leftarrow$ Input time window in seconds
2: 2−node motifs $\leftarrow$ Apply Algorithm *Temporal Motifs*
3: 3−node star motifs $\leftarrow$ Apply Algorithm *Temporal Motifs*
4: 3−node triangle motifs $\leftarrow$ Apply Algorithm *Temporal Motifs*

of S_1 and S_2 communities. That is, we explore the relation of the two polarized groups in relation to AUG_u graph size, the percentage of positive (agreement) and negative (disagreement) edges within and cross polarized groups, the percentage of vertices from each subreddits and the percentage of common vertices (users who participate in discussions in both input communities).

6.3 Counting temporal motifs

Ongoing, **RQ3** query counts the number of 2−node and 3−node temporal motifs during δ time window (see Section 5.3) in a directed temporal graph G. The pseudo − code of the methodology displayed in Algorithm 6.3. The process of counting the number of temporal motifs is a 1−stage pipeline. Specifically, we apply Algorithm 5.11 to directed signed (or unsigned) AUG_d graph.

Chapter 7

Experimental setup

In this chapter, we report the stages of data collection from Reddit platform. Specifically, we gather information of submissions (posts) from some subreddits related to a topic of discussion. In addition, for each one of these submissions, the evolution of the conversation (users and comments) is collected too. Furthermore, the hyperparameters' initialization of the algorithms that have been used to quantify unsigned and signed polarization (see Chapter 6) mentioned too.

7.1 Crawling data from Reddit

We collect data from Reddit website using Reddit API. We use both Pushshift Reddit dataset [33] and traditional Reddit API (Praw) [1]. But why use two types of APIs? Firstly, Praw API collects all this information currently available in Reddit. Therefore, information that has been deleted either by the users themselves or by the moderators can not be collected. To address this issue (deleted information) we use Pushshift API. Pushshift API is updated in real−time, and includes historical data back to Reddit's

[1]https://www.reddit.com/dev/api/

Table 7.1: Data collection information, topic of discussion, selected keywords, selected subreddits and collection dates.

Topic	Keywords	Subreddits	Collection Dates
Hagia Sophia converted into a mosque (*24 July 2020*)	Hagia Sophia Ayasofya Αγιά Σοφιά	Turkey Greece Islam	1 May 2020 11 December 2020
Nagorno–Karabakh conflict (*27 September 2020*)	Nagorno Karabakh Nagorno–Karabakh	Armenia Azerbaijan Turkey	1 July 2020 11 December 2020
COVID–19 pandemic (*22 September 2019*)	Coronavirus covid covid–19 vaccines vaccination vaccines	China_Flu Coronavirus	1 July 2019 22 December 2020
Killing of American hip-hop artist (George Floyd) (*25 May 2020*)	George Floyd Derek Chauvin Police violence Black lives matter	Unpopularopinion Bad_Cop_No_Donut BlackLivesMatter	1 March 2020 25 January 2020

inception. Also, Pushshift makes it much easier to query and retrieve historical Reddit data, provides extended functionality by providing fulltext search against comments and submissions, and has larger single query limits. The disadvantage of Pushshift is that we rely on values of scores such as, submissions' scores or comments' scores, in a discussion, perhaps the time of their collection is too short and thus does not have time to develop a controversy. Therefore, to address this problem from Pushshift API, we also use Praw API to recollect some scores. Therefore, we have data for two distinct timestamps one from Pushshift (first timestamp) and the second one from Praw (second timestamp).

We collect posts (submissions) based on specific keywords and topics of discussion. The keywords are related to real events that have taken place in recent times. In the literature, most studies use data related to politics, climate change, religion or even historical events. Therefore, our direction of choosing topics to study follows this

Table 7.2: Probability multipliers (score) a user from community A to write in community B (A → B) for each of the topics.

Subreddits	**Score**
Turkey → Greece	24.32
Turkey → Islam	13.87
Greece → Turkey	23.48
Greece → Islam	0.71
Islam → Turkey	11.23
Islam → Greece	0.60

(a) Hagia Sophia

Subreddits	**Score**
Armenia → Azerbaijan	534.76
Armenia → Turkey	129.06
Azerbaijan → Armenia	499.18
Azerbaijan → Turkey	270.76
Turkey → Armenia	118.93
Turkey → Azerbaijan	267.31

(b) Nagorno–Karabakh

Subreddits	**Score**
Unpopularopinion → Bad_Cop_No_Donut	1.93
Unpopularopinion → BlackLivesMatter	–
Bad_Cop_No_Donut → Unpopularopinion	1.69
Bad_Cop_No_Donut → BlackLivesMatter	–
BlackLivesMatter → Unpopularopinion	0.35
BlackLivesMatter → Bad_Cop_No_Donut	25.27

(c) Police violence

Subreddits	**Score**
China_Flu → Coronavirus	37.68
Coronavirus → China_Flu	18.65

(d) COVID–19

similar logic, i.e., events that bring confrontation or polarization between the members. We conclude to study four topics of discussions these are, the conversion of Hagia–Sophia into a mosque, the Nagorno–Karabakh conflict between Azerbaijan, supported by Turkey, and the self–proclaimed Republic of Artsakh together with Armenia, in the disputed region of Nagorno–Karabakh and surrounding territories, the COVID–19 pandemic, also known as the coronavirus pandemic, an ongoing pandemic of coronavirus disease 2019 (COVID–19) caused by severe acute respiratory syndrome SARS-CoV–2 and the Black Lives Matter movement that protest against incidents of police brutality and all racially motivated violence against black people. In Table 7.1, the selected keywords, the respectively the topics of discussion and the subreddits from which they come from presented. The time of the collection is based on dates that the event became widely known.

Why did we choose these subreddits and not others? Initially, our goal is to

Table 7.3: Post and comment features.

Post features	Description
author_id	Original creator of the post
created_utc	Unix timestamp
post_id	Unique post identifier
num_comments	Number of total comments in the post
score	Sum of upvotes and downvotes in the post
upvote_ratio Pushshift	Ratio of votes in the post counting both positive and negative votes using Pushshift
upvote_ratio Praw	Ratio of votes in the post counting both positive and negative votes using Praw
comments	List of comments' unique ids
title	Title of the post
url	URL address of the post
Comment features	**Description**
author_id	Original creator of the comment
created_utc	Unix timestamp
comment_id	Unique comment identifier
parent_id	Parent comment identifier in which the comment is a response
score Pushshift	Sum of upvotes and downvotes of the comment using Pushshift
score Praw	Sum of upvotes and downvotes of the comment using Praw
text	Body of the comment

select communities for which we know (or at least it is commonly accepted) that there is a conflict between them. For instance, if a topic of discussion is about Nagorno–Karabakh conflict then we suppose that communities such as Azerbaijan and Armenia will have diametrically opposed views and there will be intense polarization or even controversy. Therefore, to identify the appropriate communities (subreddits) from which we should collect data, we performed a brief analysis of the keywords we chose for each topic of discussion. That is, we collected 100 submissions that contain these specific keywords from 2007 to 2020 per month. So, we tested which subreddits use these keywords most often. Having extracted these subreddits, the second criterion we consider is that we expect there is interaction of users belonging to different groups. That is, there are users who participate in discussions in both subreddits or whether there is user overlap in the subreddits we have selected. The probability (fraction) of users of subreddit A that also write in subreddit B for each one of the four topics presented in Table 7.2.

Table 7.4: Number of collected posts and the average Pushshift upvote ratio of per non−controversial and controversial posts.

Subreddit	**Total posts**	**Upvote ratio**	**Non−Cont/sial posts**	**Upvote ratio**	**Cont/sial posts**	**Upvote ratio**
Turkey (HS)	172	91%	151	98%	21	41%
Greece	47	98%	46	99%	1	50%
Islam	90	94%	86	97%	4	42%
Armenia	845	98%	834	99%	11	49%
Azerbaijan	852	99%	847	99%	5	45%
Turkey (NK)	71	99%	71	99%	–	–
China_Flu	25,794	98%	25,420	99%	374	47%
Coronavirus	60,993	98%	60,475	98%	518	48%
Unpopularopinion	3,742	98%	3,660	99%	82	47%
Bad_Cop_No_Donut	1,237	96%	1,209	97%	28	44%
BlackLivesMatter	3,855	97%	3,737	98%	118	42%

At this point, we report the features that have been collected. More specifically, for each topic of discussion and for each subreddit, we collect a number of posts. In Table 7.3, we present the collected features for each post (top part). For each post, we collect information of the discussion that is, the comments from the users. Therefore, for each comment we collect a set of features too, these are presented in the second part of Table 7.3. Consequently, the number of total collected posts is presented in Table 7.4 for Pushshift score and Table 7.5 for Praw score. We annotate a post as non−controversial if the upvote ratio score from Pushshift or Praw APIs is higher than 0.62, otherwise, the post annotated as controversial. The value 0.62 comes from experimental analysis using information from the official post annotation by Reddit API. We observe that the period of time in which the data have been collected has a decisive role in the manner of category assignment either as non−controversial or controversial posts. Particularly, using Pushshift score, the type of posts have not yet

Table 7.5: Number of collected posts and the average Praw upvote ratio of per non−controversial and controversial posts.

Subreddit	**Total posts**	**Upvote ratio**	**Non−Cont/sial posts**	**Upvote ratio**	**Cont/sial posts**	**Upvote ratio**
Turkey (HS)	172	66%	95	88%	77	39%
Greece	47	77%	37	87%	10	38%
Islam	90	84%	78	90%	12	48%
Armenia	845	75%	659	83%	186	47%
Azerbaijan	852	86%	792	89%	60	46%
Turkey (NK)	71	77%	53	89%	18	41%
China_Flu	25,794	80%	21,328	87%	4,466	46%
Coronavirus	60,993	89%	57,493	92%	3,500	47%
Unpopularopinion	3,742	92%	3,379	96%	363	47%
Bad_Cop_No_Donut	1237	88%	1165	91%	72	42%
BlackLivesMatter	3855	86%	3,186	94%	669	47%

been formed (non−controversial or controversial) due to their fast collection time; and we can confirm this by the fact that when we use the Praw score (collected at the present time) the number of controversial posts increases.

7.2 Generating Graphs

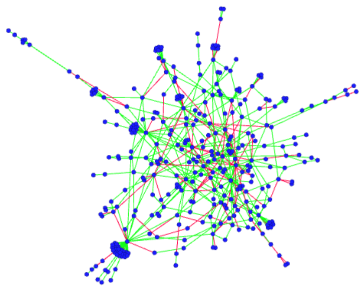
We generate both AUG_d and AUG_u for each subreddit separately that is, 11 AUG_u unsigned graphs and 11 AUG_u signed graphs for measuring polarity and 11 AUG_d graphs for counting temporal motifs. The number of vertices and edges per generated AUG_d and AUG_u graph presented in Table 7.6. Also, a graphic representation of them presented in Figure 7.1.

Subsequently, we create all the possible 2−subreddits graphs for each topic of discussion. Videlicet, an AUG_u (signed and unsigned) and AUG_d between $Turkey$ & $Greece$, $Turkey$ & $Islam$ and $Greece$ & $Islam$ subreddits from topic about Hagia − Sophia etc. Moreover, we create all 2−subreddits possible graphs from non − controversial and controversial posts respectively. Therefore, we create 30 AUG_u unsigned

Table 7.6: Size of AUG_d and AUG_u graphs per subreddit.

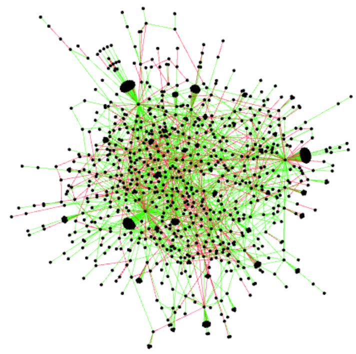
Subreddits	**Nodes** (AUG_d)	**Edges** (AUG_d)	**Nodes** (AUG_u)	**Edges** (AUG_u)	% of "+" edges (Pushshift, Praw)
Turkey (HS)	411	1,797	985	1,802	(92%, 80%)
Greece	148	604	330	525	(91%, 76%)
Islam	527	3,067	965	1,801	(74%, 68%)
Armenia	2,084	77,554	2,918	26,448	(92%, 87%)
Azerbaijan	1,156	14,753	1,818	7,902	(93%, 88%)
Turkey (NK)	121	465	433	652	(98%, 83%)
China_Flu	24,550	292,045	39,697	219,026	(95%, 86%)
Coronavirus	135,897	1,465,407	256,677	1,177,128	(91%, 84%)
Unpopularopinion	6,206	31,395	16,347	25,716	(85%, 77%)
Bad_Cop_No_Donut	3,093	11,922	12,251	16,950	(85%, 79%)
BlackLivesMatter	401	1,329	2,661	3,284	(95%, 89%)

graphs, 30 AUG_u signed graphs and 30 AUG_d graphs. The number of vertices and edges per 2−subreddits generated AUG_d and AUG_u graph presented in Tables B.1, B.2 and B.3 (see Appendix B). Besides, the graphic representation of the graphs depicted in Figures 7.2, 7.3, 7.4 and 7.5 for Hagia−Sophia, Nagorno−Karabakh, Police Violence and COVID−19 respectively. Finally, we generate all the 3−subreddits graphs for every topic of discussion either for non−controversial or controversial posts. We work with 9 additional graphs. In Table B.4 (see Appendix B), the size of the graphs presented. In addition, the graph representation of 3−subreddits graphs mentioned in Figures 7.6. Moreover, the percentage of users per subreddit and the percentage of common users in 2−subreddits and 3−subreddits AUG_u graphs presented in Tables B.5, B.6, B.7 and B.8. Summarizing, we generate 53 AUG_u and 53 AUG_d graphs in total.

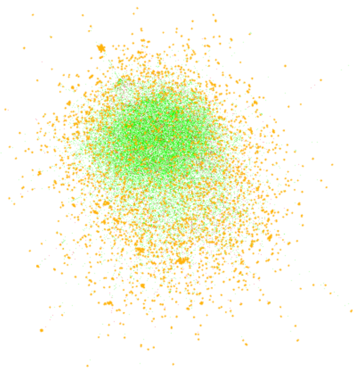


(a) Greece

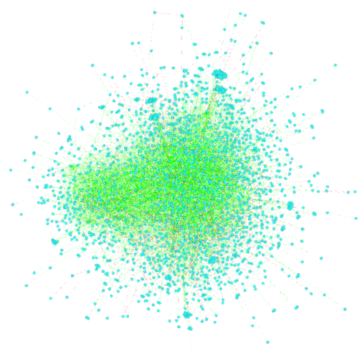
(b) Turkey (HS)



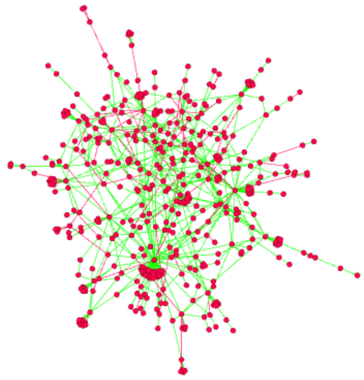
(c) Islam



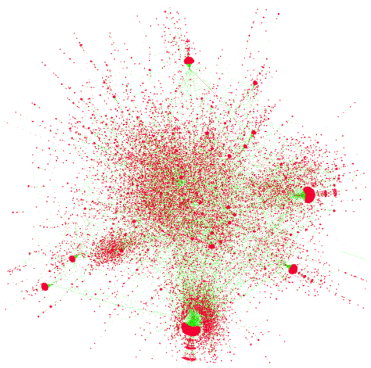
(d) Armenia



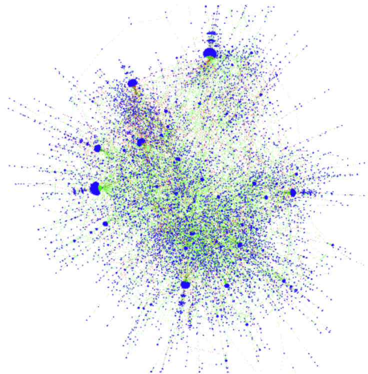
(e) Azerbaijan



(f) Turkey (NK)

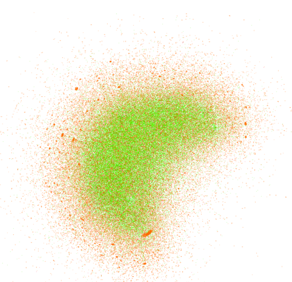


(g) Unpopularopinion

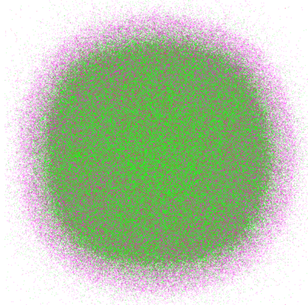


(h) Bad_Cop_No_Donut

(i) BlackLivesMatter

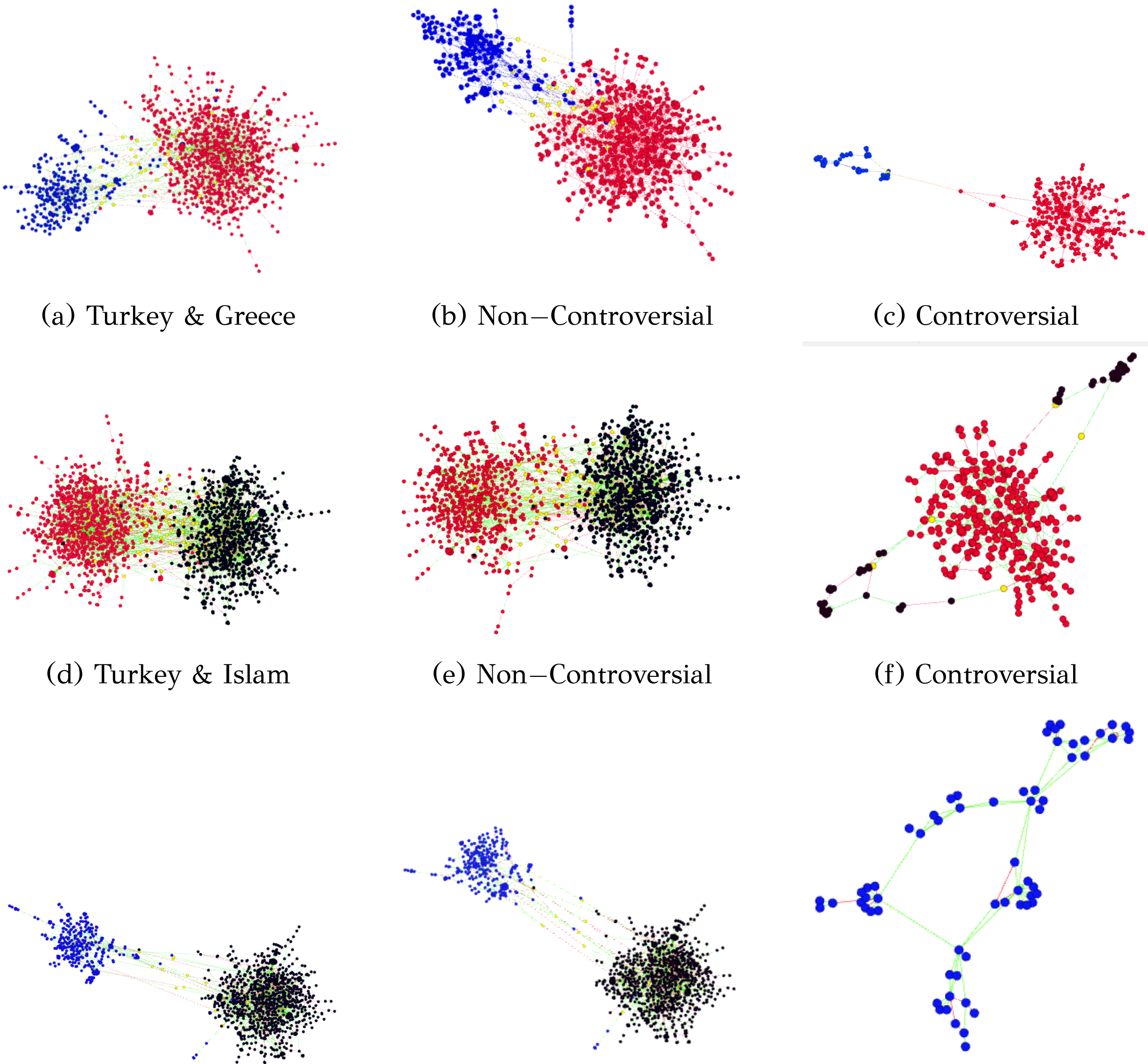


(j) China_Flu



(k) Coronavirus

Figure 7.1: Graph representation for each individual subreddit.

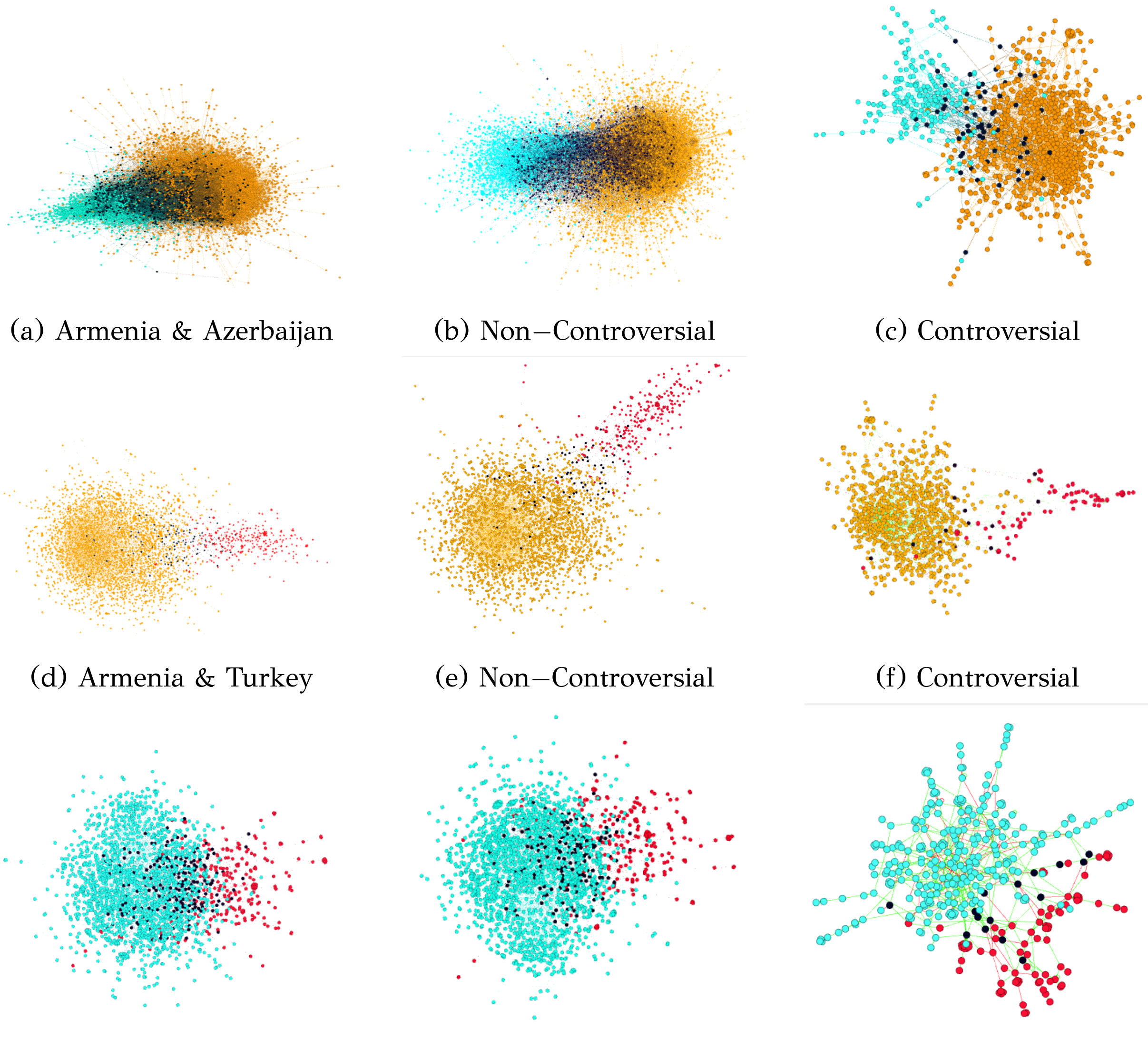


(a) Turkey & Greece (b) Non−Controversial (c) Controversial

(d) Turkey & Islam (e) Non−Controversial (f) Controversial

(g) Greece & Islam (h) Non−Controversial (i) Controversial

Figure 7.2: Graph representation of 2−subreddits for Hagia−Sophia topic. Red color declares Turkey, blue color Greece and black color Islam.

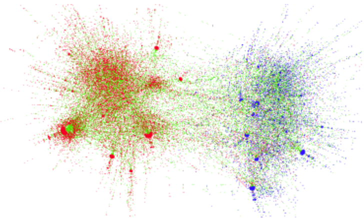


(a) Armenia & Azerbaijan (b) Non−Controversial (c) Controversial

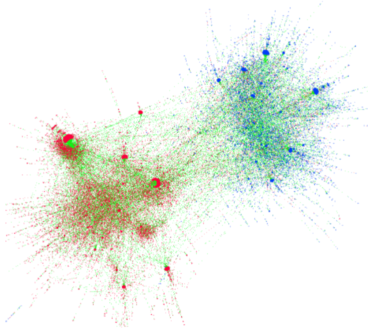
(d) Armenia & Turkey (e) Non−Controversial (f) Controversial

(g) Azerbaijan & Turkey (h) Non−Controversial (i) Controversial

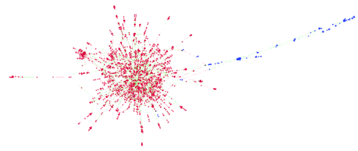
Figure 7.3: Graph representation of 2−subreddits for Nagorno−Karabakh topic. Orange color declares Armenia, cyan color Azerbaijan and red color Turkey.



(a) Unpopularopinion & Bad_Cop_No_Donut

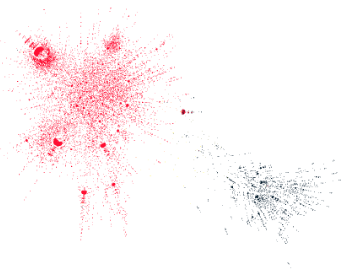


(b) Non−Controversial

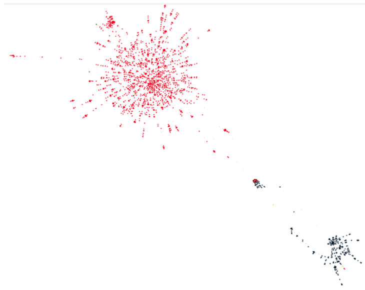


(c) Controversial

(d) Unpopularopinion & BlackLivesMatter

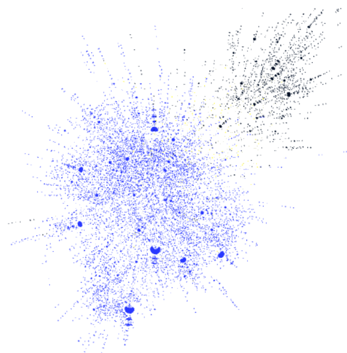


(e) Non−Controversial

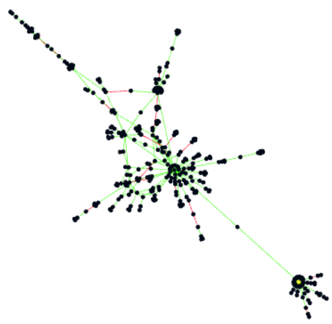


(f) Controversial

(g) Bad_Cop_No_Donut & BlackLivesMatter

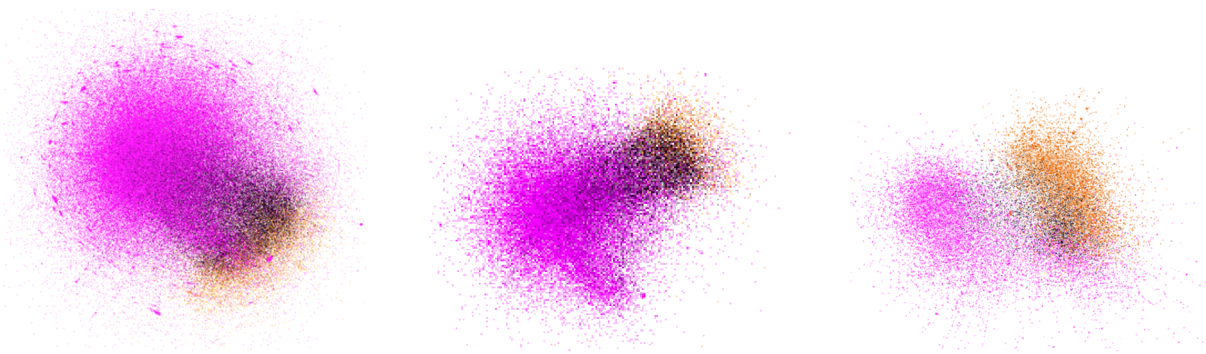


(h) Non−Controversial



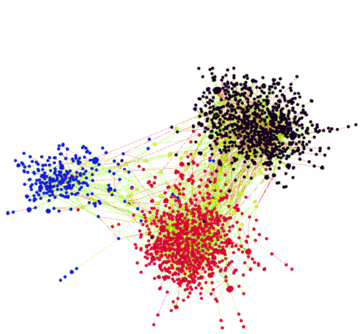
(i) Controversial

Figure 7.4: Graph representation of 2−subreddits for Police Violence topic. Red color declares Unpopularopinion, blue color Bad_Cop_No_Donut and black color BlackLivesMatter.

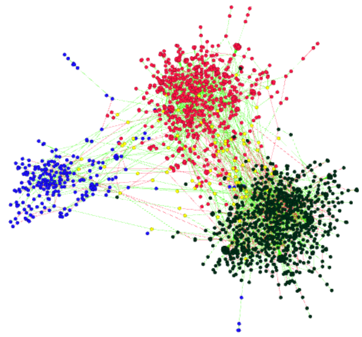


(a) COVID−19 (b) Non−Controversial (c) Controversial

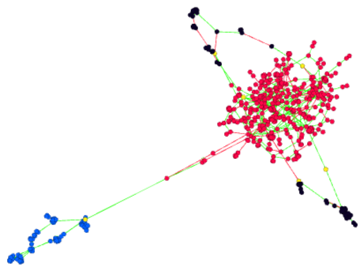
Figure 7.5: Graph representation of 2−subreddits for COVID−19 (China_Flu & Coronavirus) topic. Purple color regards Coronavirus and orange color regards China_Flu.



(a) Hagia–Sophia



(b) Non–Controversial

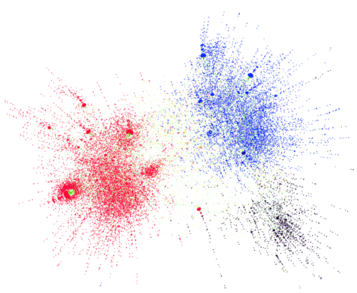


(c) Controversial

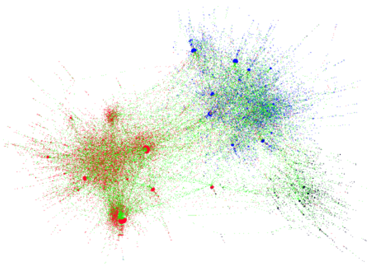
(d) Nagorno–Karabakh

(e) Non–Controversial

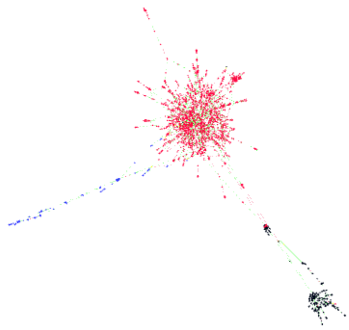
(f) Controversial



(g) Police Violence



(h) Non–Controversial



(i) Controversial

Figure 7.6: Graph representation of 3–subreddits for Hagia–Sophia (Turkey (red) & Greece (blue) & Islam (black)), Nagorno–Karabakh (Armenia (orange) & Azerbaijan (cyan) & Turkey (red)) and Police Violence (Unpopularopinion (red) & Bad_Cop_No_Donut (blue) & BlackLivesMatter (black)) topics both for total, non–controversial and controversial posts.

7.3 Initialization of Hyperparameters

For some algorithms it may be necessary to initialize some of their hyperparameters. A model hyperparameter is a configuration that is external to the model and whose value cannot be estimated from data. Therefore, we need to initialize some of these hyperparameters. Especially, algorithms which quantify unsigned polarization namely, *Random Walks (D)*, *Random Walks (P)*, *Betweenness*, *Embeddings*, *GMCK* and *MBLB*, need an initialization. Initially, due to their randomness, we run Algorithm 6.1 $1k$ iterations and set as final unsigned polarization score the average score from all these iterations. Also, we performed experiments of top 10 high−degree nodes for *Random Walks (D)* method and 10% of random nodes per community for *Random Walks (P)* method. Furthermore, we test two types of nodes partitioning, METIS and Real groups clustering.

For algorithms which detect the two most polarized subgroups in one or more communities and measure signed polarity, no hyperparameters initialization was observed.

Chapter 8

Results and Discussion

This chapter presents the results of all the experiments and an extensive discussion of them. We organize the results based on the research questions we investigate. Initially, we report results for intra−polarization and after that, results for inter−polarization. The analysis of signed polarity in networks focuses on *Random Eigensign* method because it is proposed as the most efficient for political debates. Also, we make a discussion which concerns controversial and non−controversial posts. Finally, the findings about temporal motifs follow. Due to the large number of results, some of them mentioned within Sections and the rest of them placed in Appendix C.

8.1 Is unsigned and signed intra polarity detected in Reddit?

The quantification of unsigned and signed polarity in a community in which individuals discuss a same topic extract interesting conclusions for the stability of the

Table 8.1: Unsigned and signed polarity score per subreddit about Hagia–Sophia conversations.

Algorithms (unsigned)	**Greece**	**Turkey**	**Islam**
Random walks (P)	0.25	0.20	0.22
Random walks (D)	0.35	0.04	0.09
Betweenness	**0.52**	**0.60**	**0.68**
Embeddings	0.15	0.07	0.22
GMCK	0.11	0.06	0.09
MBLB	0.15	0.14	0.12
Algorithms (signed)	**Polarity (Pushshift, Praw)**		
Eigensign	$(4.77, 3.0)$	$(6.14, 5.33)$	$(3.47, 3.56)$
Random eigensign	$(3.12, 2.08)$	$(4.22, 3.23)$	$(2.39, 2.29)$
Greedy	$(5.0, 3.6)$	$(6.71, 5.71)$	$(3.07, 3.15)$
Bansal	$(3.42, 3.42)$	$(3.36, 3.2)$	$(3.0, 3.2)$
LocalSearch	$(1.22, 1.12)$	$(1.71, 1.36)$	$(1.44, 0.85)$

community. For instance, if individuals of Greece community have a common point of view about Hagia–Sophia conversion into a mosque or if Turkish individuals agree or disagree with each other about this conversion. Equally interesting is if users from Armenia and Azerbaijan subreddits are bifurcated and are divided in two subgroups of users. Similar conclusions can be drawn in the case of Police Violence and COVID–19 discussions.

Initially, we investigate the quantification of unsigned polarity in every community separately that is, intra–polarization. We observe that almost all the proposed methods such as, *Random Walks*, *Embeddings*, *GMCK* and *MBLB* does not detect polarity (polarization score less than threshold 0.50) in subreddits. Tables 8.1, 8.2, 8.3 and 8.4 present the unsigned polarization scores for each one of the topics. That is, the structure of the network is not well divided into two subcommunities. In contrast, the users interact with the rest of them converting subreddits into a well connected ones.

We need to note that, the algorithms use information who replies to whom and not information about the score of the replies. This result might is expected because Reddit is a forum website and the main interaction between them is the replies

Table 8.2: Unsigned and signed polarity score per subreddit about Nagorno–Karabakh conversations.

Algorithms (unsigned)	**Armenia**	**Azerbaijan**	**Turkey**
Random walks (P)	0.10	0.12	0.28
Random walks (D)	0.0	0.0	0.32
Betweenness	**0.68**	**0.53**	**0.56**
Embeddings	0.10	0.07	0.29
GMCK	0.0	0.0	0.11
MBLB	0.11	0.09	0.19
Algorithms (signed)	**Polarity (Pushshift, Praw)**		
Eigensign	$(80.29, 74.09)$	$(27.3, 26.58)$	$(3.80, 3.18)$
Random eigensign	$(54.35, 51.43)$	$(17.08, 16.13)$	$(3.05, 2.18)$
Greedy	$(80.43, 74.14)$	$(27.53, 26.83)$	$(4.94, 4.20)$
Bansal	$(50.15, 42.23)$	$(17.60, 17.54)$	$(3.0, 2.5)$
LocalSearch	$(8.14, 7.21)$	$(4.56, 4.22)$	$(1.35, 1.47)$

from one user to the other one. So, we except that each subreddit is well connected and the proposed algorithms about controversy confirm our hypothesis. Unlike, only *Betweenness* method detect the existence of polarization in all subreddits. Specifically, the polarization score is higher than threshold 0.50 for every community and as a consequence, they are annotated as polarized communities. As already mentioned in Section 5.1.2, this metric is based on the similarity of betweenness centrality of cut and the rest of the edges. So, we conclude that a few cut edges exist cross two partitions and if the vertices from both partitions necessarily use the cut edges of the network to cross to the other partition; and as a result the betweenness centrality of cut edges is higher that the rest of them.

As of now, we inquiry the quantification of unsigned polarization in each one of the subreddits separately (intra–polarization). The next query explores the existence of the two most polarized groups (subsets) of the network including additional information from the conversations that is, score of the comments. The main goal of measuring signed polarity in a subreddit is to detect if there are two polarized subcommunities within a subreddit where the the individuals within each subcommunity agree and cross subcommunities disagree. Looking at the results of *Random Eigensign*

Table 8.3: Unsigned and signed polarity score per subreddit about Police Violence conversations.

Algorithms (unsigned)	Unpopularopinion	Bad_Cop_No_Donut	BlackLivesMatter
Random walks (P)	**0.50**	**0.57**	0.43
Random walks (D)	0.06	0.20	0.03
Betweenness	**0.91**	**0.82**	**0.92**
Embeddings	0.21	0.36	0.10
GMCK	0.16	0.16	0.16
MBLB	0.13	0.10	0.10
Algorithms (signed)	**Polarity (Pushshift, Praw)**		
Eigensign	$(6.64, 6.46)$	$(3.38, 3.50)$	$(4.23, 4.20)$
Random eigensign	$(2.39, 2.41)$	$(2.13, 2.05)$	$(2.29, 2.33)$
Greedy	$(7.63, 7.71)$	$(5.9, 4.75)$	$(4.72, 4.62)$
Bansal	$(4.93, 4.58)$	$(3.71, 3.27)$	$(2.84, 2.84)$
LocalSearch	$(1.30, 1.17)$	$(1.33, 1.07)$	$(1.19, 1.11)$

Table 8.4: Unsigned and signed polarity score per subreddit about COVID−19 conversations.

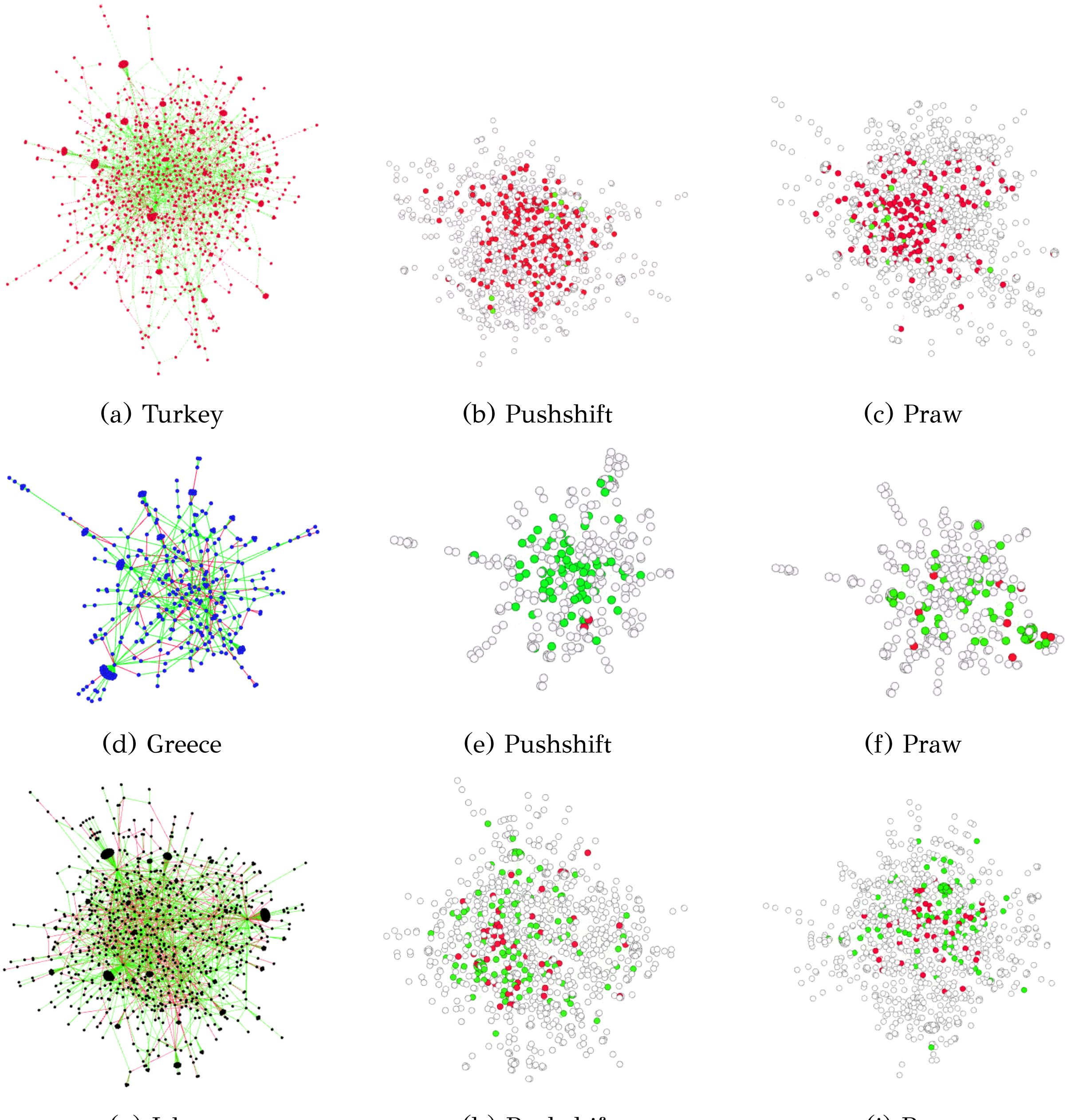
Algorithms (unsigned)	China_Flu	Coronavirus
Random walks (P)	0.19	0.16
Random walks (D)	0.0	0.0
Betweenness	**0.66**	**0.53**
Embeddings	0.29	0.03
GMCK	0.0	0.0
MBLB	0.12	0.10
Algorithms (signed)	**Polarity (Pushshift, Praw)**	
Eigensign	$(50.60, 38.79)$	$(87.03, 74.16)$
Random eigensign	$(25.24, 18.95)$	$(26.97, 19.52)$
Greedy	$(52.07, 41.47)$	$(88.40, 76.13)$
Bansal	$(15.25, 8.96)$	$(19.69, 15.72)$
LocalSearch	$(5.17, 3.61)$	$(3.67, 2.94)$

Table 8.5: Percentage of agreement (positive edges, for short PE) in each community separately and within (G_1 PE, G_2 PE) and cross (for short CE) polarized partitions, G_1 and G_2 for Random Eigensign method and for two time snapshots, Pushshift and Praw.

Subreddit	**PE**	G_1 **PE**	G_2 **PE**	**CE**
	(Pushshift, Praw)			
Turkey (HS)	92%, 80%	97%, 94%	100%	13%, 18%
Greece	91%, 76%	–	98%, 94%	25%, 0%
Islam	74%, 68%	80%, 81%	94%, 97%	27%, 25%
Armenia	92%, 87%	93%, 90%	–	0%, 33%
Azerbaijan	93%, 88%	96%, 95%	–, 100%	33%, 29%
Turkey (NK)	98%, 83%	100%	–, 100%	–, 40%
China_Flu	95%, 86%	–, 76%	95%,88%	50%, 56%
Coronavirus	91%, 84%	90%, 86%	63%, 69%	58%, 57%
Unpopularopinion	85%, 77%	98%, 97%	88%, 71%	34%, 27%
Bad_Cop_No_Donut	95%, 79%	99%	–, 88%	0%, 7%
BlackLivesMatter	95%, 89%	–	99%	0%, 11%

method in Tables 8.1, 8.2, 8.3 and 8.4, we observe that there is an amount of polarity in each one community; excluding some of them with higher values that the others. Such as, Turkey for Hagia–Sophia topic, Armenia for Nagorno–Karabakh topic, Unpopularopinion for Police Violence topic and Coronavirus for COVID–19 topic. Note that the polarity score values are strictly related to the size of the graphs thus unable to do comparisons between them.

Moreover, it is also confirmed that within groups the number of positive edges is higher than negative edges and the number of negative connections cross groups is higher than positive connections (see Table 8.5). Furthermore, we notice that the polarity in the first timestamp (Pushshift) is higher than this one in the second timestamp (Praw). Therefore, the polarity decreases which means that as the discussions progressed, the participants became more flexible and may have strayed from their extreme points of views. The graph representation of the two polarized partitions for each community is presented in Figures 8.1, 8.2, 8.3 and 8.4. Green and red colors declare the two polarized groups.

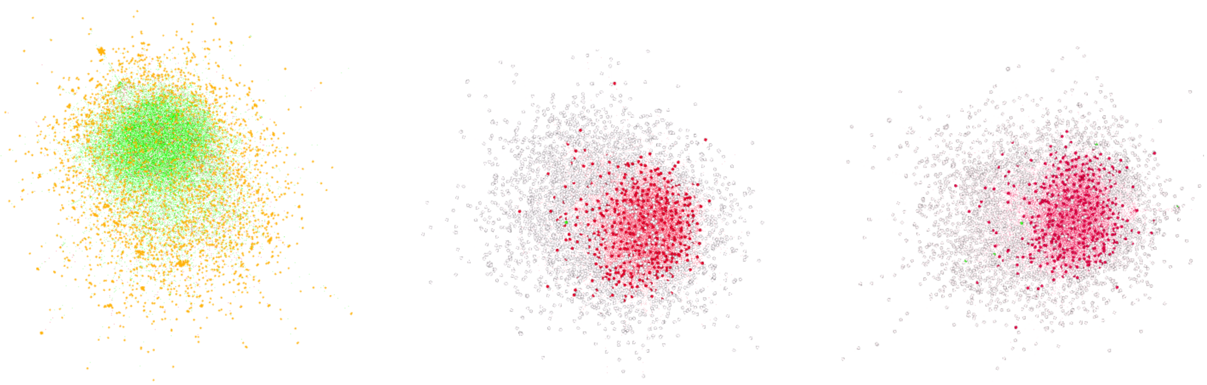


(a) Turkey (b) Pushshift (c) Praw

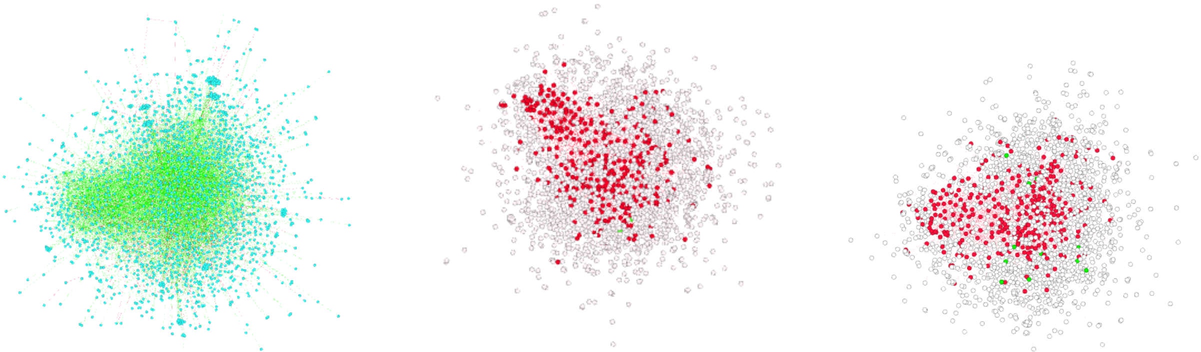
(d) Greece (e) Pushshift (f) Praw

(g) Islam (h) Pushshift (i) Praw

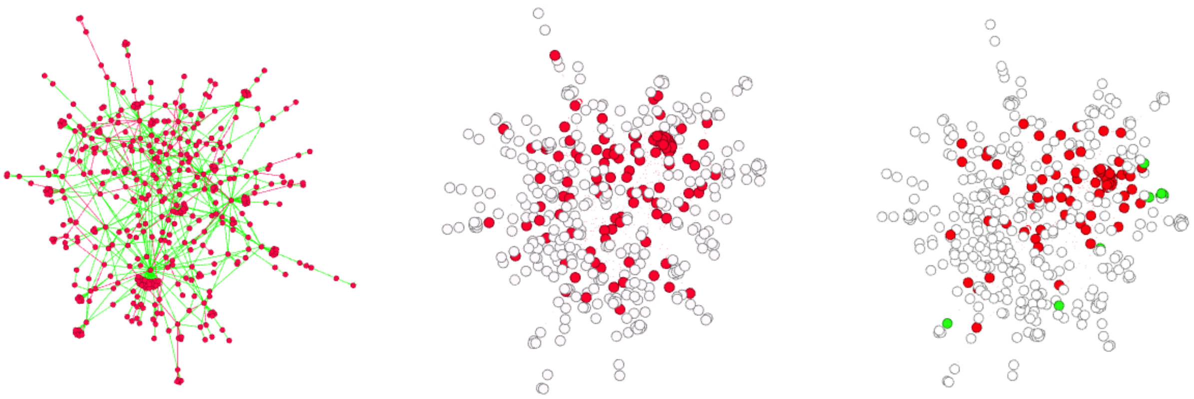
Figure 8.1: Graph representation of two polarized groups per subreddit for two timestamps, Pushshift and Praw, for Hagia–Sophia topic.



(a) Armenia (b) Pushshift (c) Praw

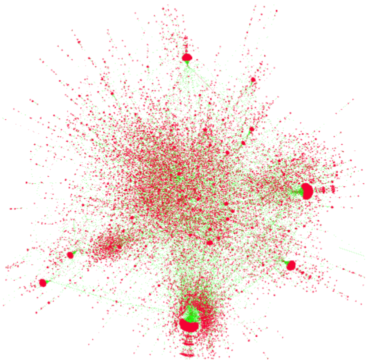


(d) Azerbaijan (e) Pushshift (f) Praw

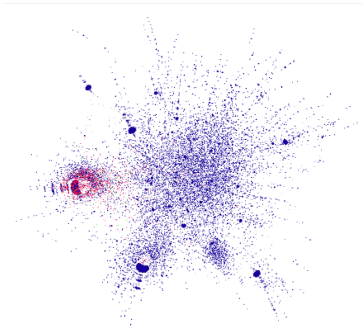


(g) Turkey (h) Pushshift (i) Praw

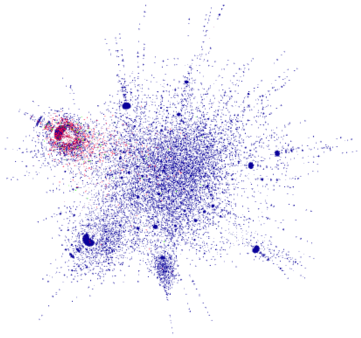
Figure 8.2: Graph representation of two polarized groups per subreddit for two timestamps, Pushshift and Praw, for Nagorno−Karabakh topic.



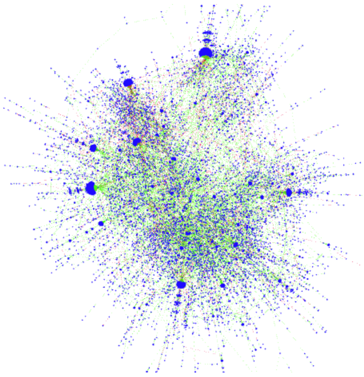
(a) Unpopularopinion



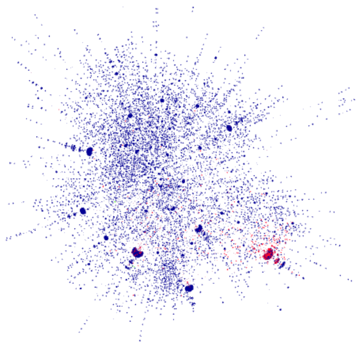
(b) Pushshift



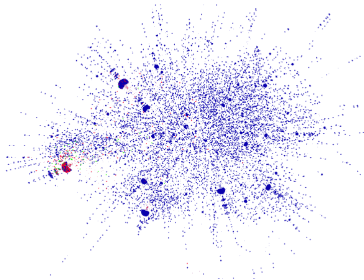
(c) Praw



(d) Bad_Cop_No_Donut

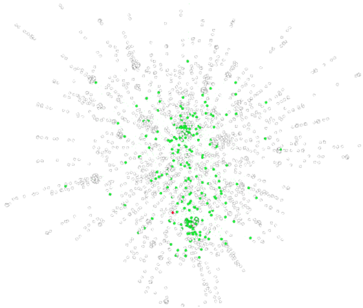


(e) Pushshift



(f) Praw

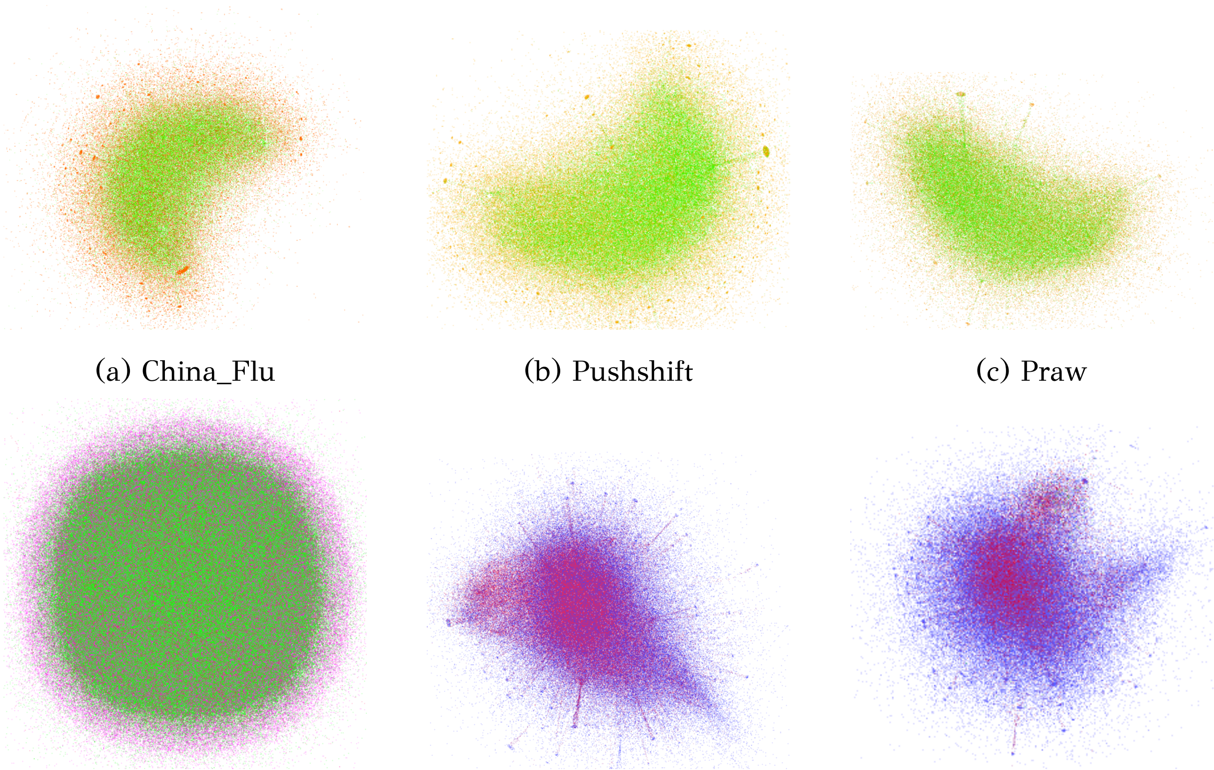
(g) BlackLivesMatter



(h) Pushshift

(i) Praw

Figure 8.3: Graph representation of two polarized groups per subreddit for two timestamps, Pushshift and Praw, for Police Violence topic.



(a) China_Flu (b) Pushshift (c) Praw

(d) Coronavirus (e) Pushshift (f) Praw

Figure 8.4: Graph representation of two polarized groups per subreddit for two timestamps, Pushshift and Praw, for COVID−19 topic.

Conclusions. The layout of Reddit platform tends the individuals in each subreddit (community) to form a bond with the members of the subreddit and as a consequence, they interact quite tightly either they agree or disagree. Therefore, the subreddit is not well−separated into two subcommunities and as a result, the unsigned polarity detection methods successfully identify that there is no polarization focusing on the structure of each community. In other words, no unsigned intra−polarization has been detected in Reddit. On the other hand, a quantity of signed polarization is detected and two polarized subcommunities are engendered in which the members within each one of the two communities agree and cross communities disagree. With different meaning, signed intra−polarization has been detected in Reddit.

8.2 Is unsigned and signed inter−polarity detected in Reddit?

By definition of controversy and polarity, there is controversy in a network if there are extreme points of views and as a consequence, the members of the network are well−separated into two groups who does not communicate with the members of the other group and are strictly connected with the members of their group or there is connection which declares controversy. The next query that we investigate is the quantification of unsigned and signed polarity in a network that is constructed by two or more subreddits that is, inter−polarization. Also, we study the influence of two clustering techniques, METIS and Real groups. However, how does this query can be translated into the events we have selected to study? We examine each topic separately.

Hagia−Sophia. Our assumptions about Hagia−Sophia topic of discussion is that Turkey & Greece and Greece & Islam communities may disagree about this conversion and the interactions between individuals of the communities either will not be conspicuous or will be visible but it will convey controversy. On the other hand, we assume that users from Turkey & Islam may agree and have common views and thus communicate will be noticed cross these people. Lets look though, if our affairs will be confirmed from the experimental analysis.

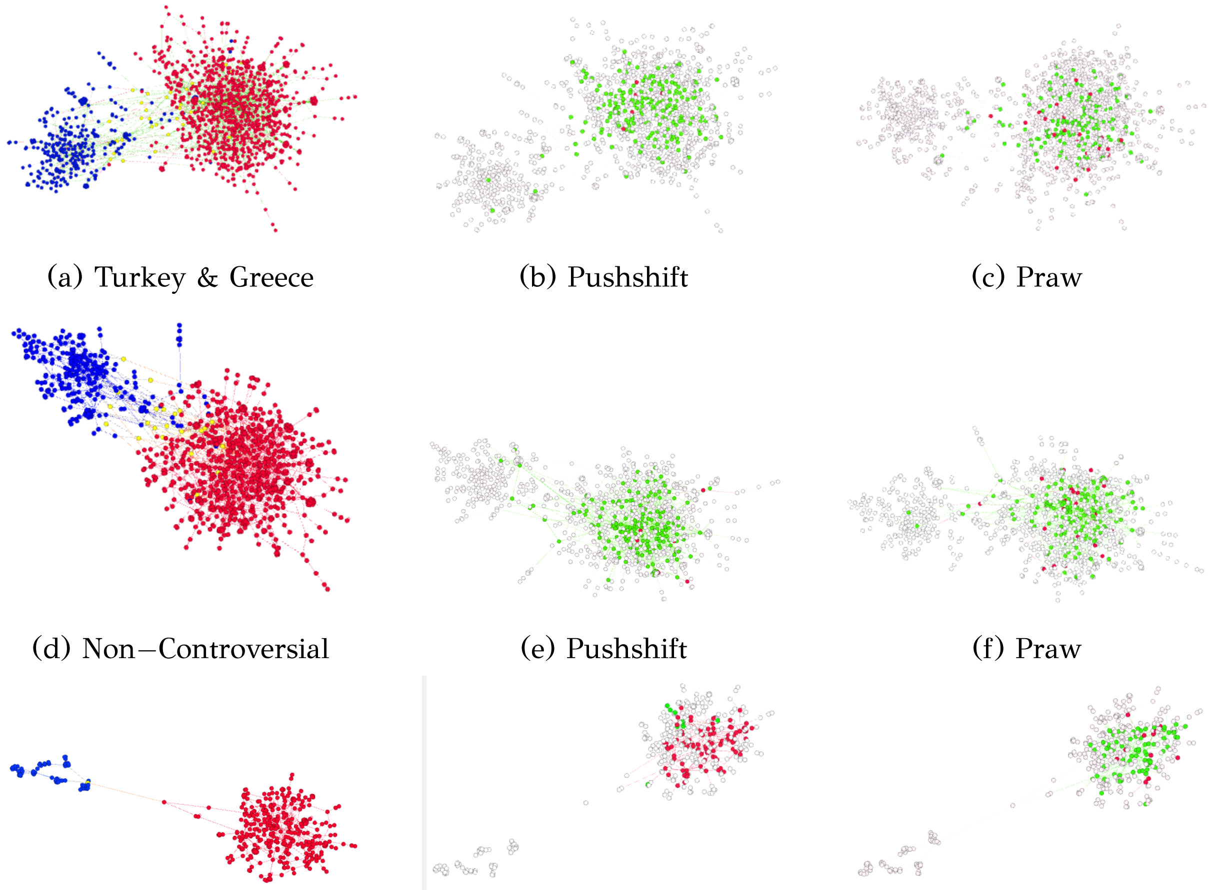
Initially, we observe that unsigned polarity is identified for all 2−subreddits using Real groups clustering technique. Particularly, *Random Walks (P)*, *Betweenness* and

Table 8.6: Unsigned and signed polarity score for Turkey & Greece and for total, non−controversial and controversial posts applying either METIS or Real groups clustering and for two distinct timestamps Pushshift and Praw.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	$0.38, 0.42, 0.40$	$\mathbf{0.61}$, $\mathbf{0.60}$, $\mathbf{0.91}$
Random walks (D)	$0.33, 0.31, 0.30$	$0.45, 0.43, \mathbf{0.98}$
Betweenness	$\mathbf{0.64}, \mathbf{0.72}, \mathbf{0.69}$	$\mathbf{0.60}, \mathbf{0.63}, \mathbf{0.50}$
Embeddings	$0.31, 0.31, 0.30$	$\mathbf{0.56}, \mathbf{0.54}, \mathbf{0.79}$
GMCK	$0.09, 0.09, 0.15$	$0.05, 0.15, 0.22$
MBLB	$0.16, 0.15, 0.15$	$0.09, 0.10, 0.07$
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	$6.10, 5.30, 3.77$	$5.24, 5.12, 3.0$
Random eigensign	$4.12, 3.61, 2.83$	$3.17, 3.12, 2.25$
Greedy	$6.8, 5.85, 4.14$	$5.72, 5.11, 3.30$
Bansal	$3.42, 3.38, 2.8$	$3.42, 3.18, 2.8$
LocalSearch	$1.27, 1.47, 1.22$	$1.22, 1.19, 0.81$

Embeddings detect unsigned inter−polarization between communities (see Table 8.6 for Turkey & Greece subreddits), The polarity score for Turkey & Islam and Greece & Islam subreddits placed in Appendix C and the Tables are C.1 and C.2. Note, that unsigned polarity score between Turkey & Islam is close to threshold 0.5 with which we distinguish the existence of polarity or not in network. Moreover, *GMCK* and *MBLB* methods do not detect polarization in any case. Furthermore, looking at the clustering technique, we notice that using METIS clustering only *Betweenness* method recognize polarity cross communities.

Afterwards, we notice that clustering technique is quite important for the most accurate detection of polarization in the case of unsigned polarity. Specifically, METIS clustering does not detect inter−polarization. However, Real groups clustering works better and quantify polarity cross subreddits. Furthermore, we observe that unsigned algorithms in combination with controversial posts quantify higher polarity score in contrast to non−controversial posts i.e., *Random Walks (P)* unsigned polarity score



(a) Turkey & Greece (b) Pushshift (c) Praw

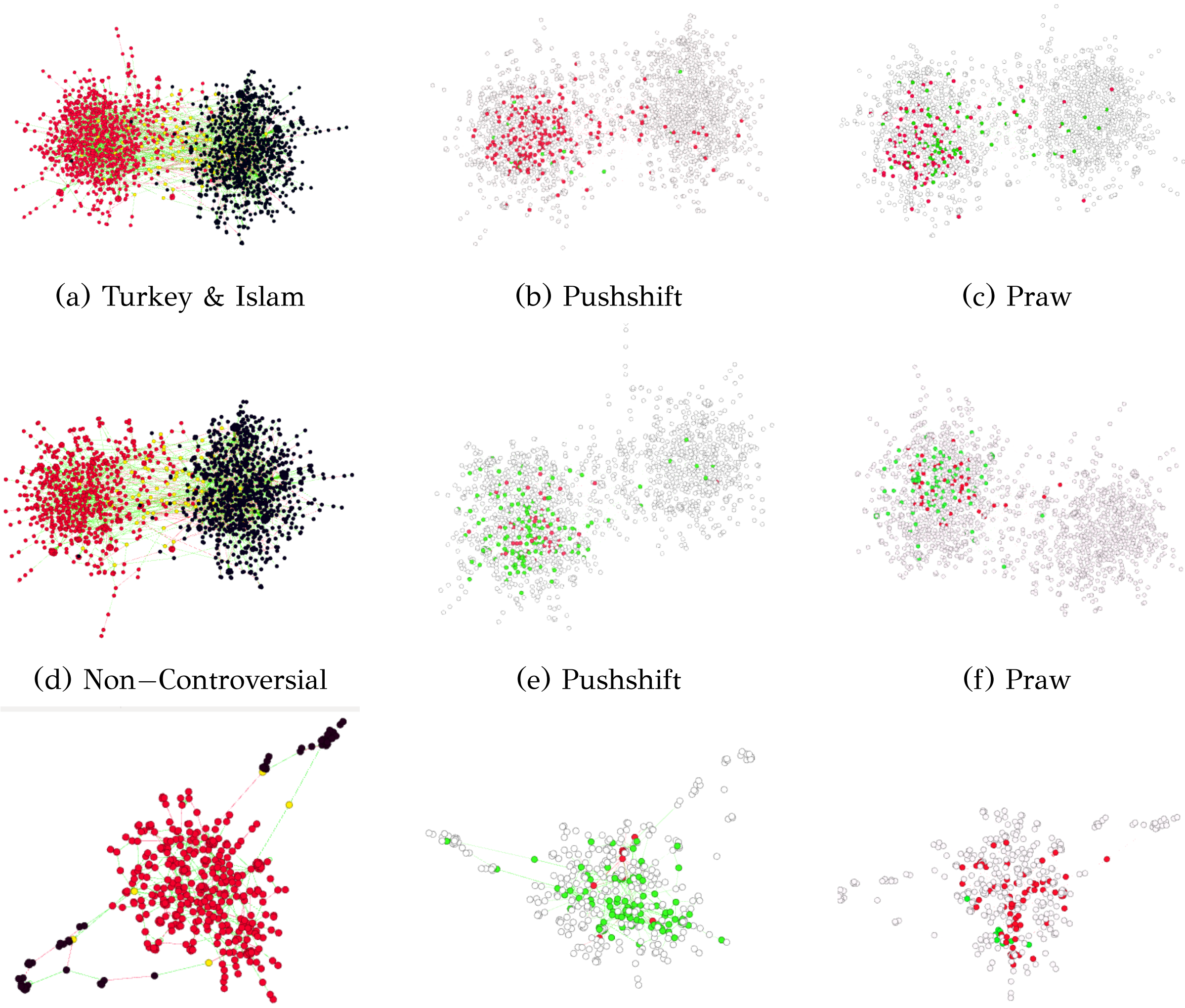
(d) Non−Controversial (e) Pushshift (f) Praw

(g) Controversial (h) Pushshift (i) Praw

Figure 8.5: Turkey & Greece graph representation of two polarized groups.

for non−controversial posts is equals to 0.60 while for controversial posts is 0.91.

Continuing the study, we focus on the detection of the two most polarized subgroups. The analysis focuses on the results from *Random Eigensign* method. We remark that the highest polarity is detected within one of the two communities. To be specific, between Turkey & Greece the most polarized groups are detected interior Turkey community (see Figure 8.5). In particular, members within the polarized subsets either come from Turkey or are users who intermediate members of the two subreddits (see Table 8.8). Next, cross Turkey & Islam, the polarized groups located within Islam (see Figure 8.6) while in the case of Greece & Islam, Islam is more polarized than Greece in non − controversial conversations (see Figure D.1 (a)-(c) in Appendix D). Also, looking at the Table 8.7 we verify that the percentage of agreement inside subgroups is positive (sign of edges is "+") while the external either does not exist or if so, it is negative (sign of edges is "-"). Finally, as the discussion evolves over time, we notice that the existence of polarization continues to appear, however



(a) Turkey & Islam (b) Pushshift (c) Praw

(d) Non−Controversial (e) Pushshift (f) Praw

(g) Controversial (h) Pushshift (i) Praw

Figure 8.6: Turkey & Islam graph representation of two polarized groups.

with fewer participants.

In case of 3−subreddits, Turkey & Greece & Islam, similar to the cases of 2 − subreddits, the most polarized subsets of individuals detected inside Islam subreddit. In Figure D.2 the graph representation of polarized groups mentioned. Also, in Table C.3 signed polarization score presented.

Hagia−Sophia Conclusions. Accordingly, we conclude that unsigned algorithms detect inter−polarization cross all 2−subreddits. Furthermore, controversial posts are more prone to polarity than non − controversial posts. In addition, the two most polarized partitions of the network do not detected cross communities instead, within one of them. In this case, Turkey is more polarized than Greece and Islam is more polarized than Greece and Turkey. Moreover, users from different groups will not go

Table 8.7: Hagia−Sophia. Percentage of agreement (positive edges, for short PE) within and cross (for short CE) polarized partitions G_1 and G_2 for Praw Random Eigensign. NC and C declare non − controversial and controversial posts.

Subreddits	G_1 **&** G_2 **nodes**	G_1 **PE**	G_2 **PE**	**CE**
Turkey & Greece	2% & 12%	100%	95%	23%
Turkey & Greece (NC)	2% & 14%	–	97%	8%
Turkey & Greece (C)	2% & 17%	100%	92%	13%
Turkey & Islam	6% & 3%	93%	83%	22%
Turkey & Islam (NC)	3% & 6%	80%	97%	26%
Turkey & Islam (C)	16% & 2%	95%	100%	0.0%
Greece & Islam	4% & 8%	77%	94%	28%
Greece & Islam (NC)	4% & 9%	76%	96%	25%
Greece & Islam (C)	6% & 31%	–	96%	0.0%

to the opposite community to express negative views. Finally, the intensity of signed polarization diminishes as time passes.

Table 8.8: Hagia−Sophia. Percentage of users from A & B subreddits and common users in polarized groups G_1 and G_2 for Praw Random Eigensign. NC and C declare non − controversial and controversial posts.

	G_1			G_2		
Subreddits A & B	A	B	Common	A	B	Common
Turkey & Greece	91%	–	9%	85%	7%	8%
Turkey & Greece (NC)	95%	5%	–	89%	7%	4%
Turkey & Greece (C)	100%	–	–	100%	–	–
Turkey & Islam	6%	90%	4%	15%	74%	11%
Turkey & Islam (NC)	4%	86%	10%	2.5%	95%	1.5%
Turkey & Islam (C)	95%	2%	3%	100%	–	–
Greece & Islam	–	98%	2%	–	99%	1%
Greece & Islam (NC)	2%	98%	–	1.92%	97%	0.96%
Greece & Islam (C)	100%	–	–	100%	–	–

Nagorno–Karabahk. Afterwards, we look into Nagorno–Karabakh conflict where the main opposite communities are Armenia & Azerbaijan. We assume that the confrontation between these subreddits will be intense as well as between Armenia & Turkey. Thus, we assume that will be no polarization between Azerbaijan & Turkey subreddits based on the official announcement of Turkey that it is in favor of Azerbaijan. Lets check, if our affairs will be confirmed from the experimental analysis.

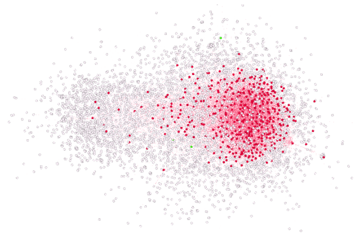
Initially, looking at the unsigned polarization scores (Table 8.9 and Tables C.4, C.5 in Appendix C), we observe that non of the unsigned methods detect inter–polarization cross cross Armenia & Azerbaijan and Azerbaijan & Turkey. Specifically, all unsigned scores either using METIS or Real groups clustering are less than threshold 0.50. This

Table 8.9: Unsigned and signed polarity score for Armenia & Azerbaijan and for total, non–controversial and controversial posts applying either METIS or Real groups clustering for two timestamps Pushshift and Praw.

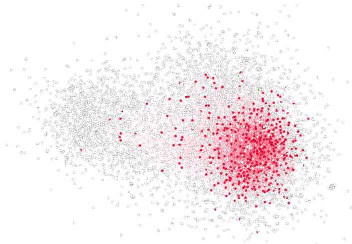
Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	$0.26, 0.28, 0.32$	$0.18, 0.20, 0.39$
Random walks (D)	$0.24, 0.21, 0.39$	$0.0, 0.0, 0.38$
Betweenness	$\mathbf{0.70}, \mathbf{0.69}, \mathbf{0.66}$	$\mathbf{0.78}, \mathbf{0.78}, \mathbf{0.60}$
Embeddings	$0.32, 0.34, 0.24$	$0.35, 0.38, 0.43$
GMCK	$0.0, 0.0, 0.07$	$0.0, 0.0, 0.03$
MBLB	$0.12, 0.12, 0.13$	$0.12, 0.12, 0.15$
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	$80.37, 79.86, 13.10$	$74.08, 73.71, 12.32$
Random eigensign	$54.16, 53.62, 9.61$	$51.02, 50.61, 8.06$
Greedy	$80.48, 79.94, 13.24$	$74.14, 73.72, 12.49$
Bansal	$50.32, 49.68, 7.65$	$42.36, 42.75, 6.55$
LocalSearch	$7.27, 7.30, 2.06$	$6.53, 6.06, 1.60$

means that users both from Armenia & Azerbaijan and Azerbaijan & Turkey have tight connections and the structure of their interactions are not well – separated into two disjoint subgroups. Note, that both opposite subreddits interact with each other not knowing the type of interaction but only the occurrence of interaction. However,

(a) Armenia & Azerbaijan

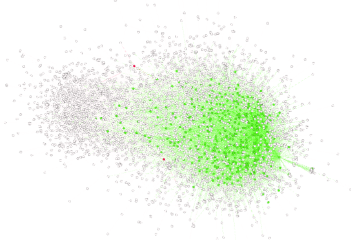


(b) Pushshift

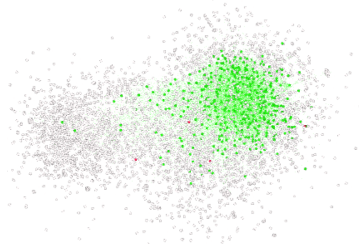


(c) Praw

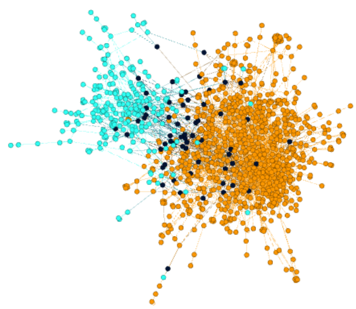
(d) Non−Controversial



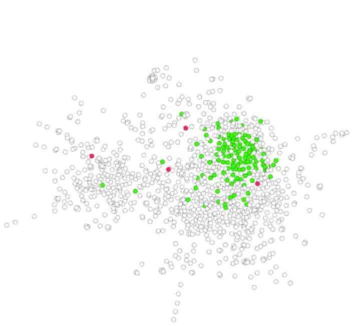
(e) Pushshift



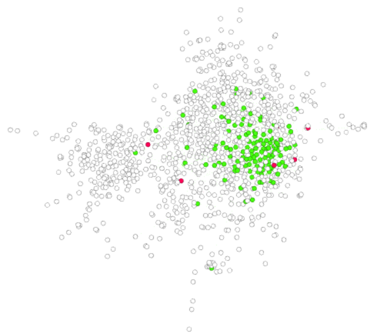
(f) Praw



(g) Controversial



(h) Pushshift

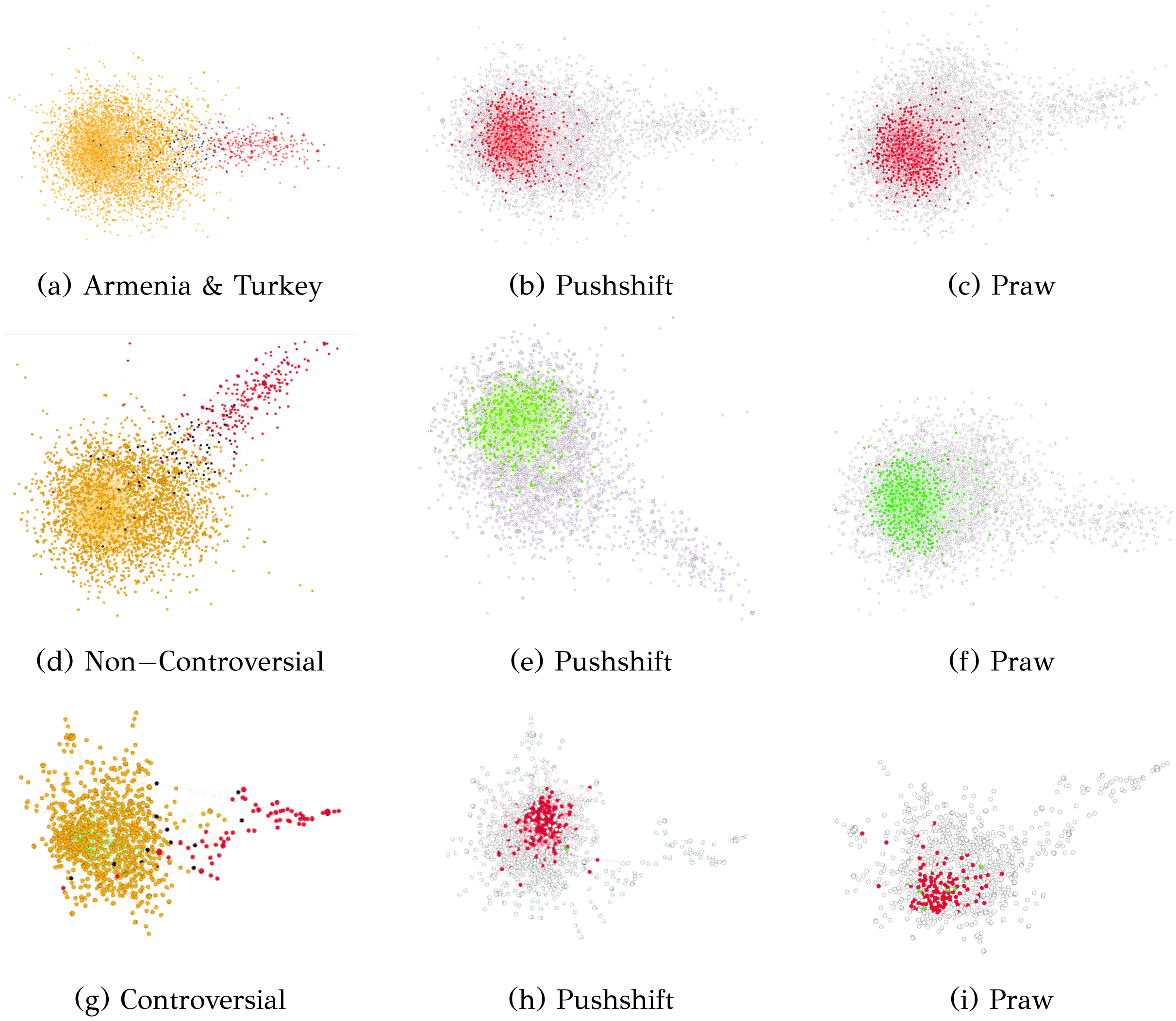


(i) Praw

Figure 8.7: Armenia & Azerbaijan graph representation of two polarized groups.

only *Betweenness* metric quantify unsigned inter − polarization higher than threshold 0.50 in all 2−subreddits options. Note that the unsigned polarization *Betweenness* score cross Azerbaijan & Turkey ranges is low values that is, 0.57 score. Furthermore, we observe that unsigned algorithms in combination with controversial posts quantify higher polarity score in contrast to non−controversial posts i.e., *Random Walks (P)* unsigned polarity score for non−controversial posts is equals to 0.20 while for controversial posts is 0.39. Similarly, *Random Walks (D)*, *Embeddings*, *GMCK* and *MBLB* measure slightly higher inter − polarity.

Continuing, the analysis focuses on *Random Eigensign* signed inter − polarization and the two most polarized subgroups. We remark that the highest polarity is detected within one of the two communities (see Figures 8.7, 8.8 and Figure D.3 in Appendix D). To be specific, between Armenia & Azerbaijan and Armenia & Turkey,



(a) Armenia & Turkey (b) Pushshift (c) Praw

(d) Non−Controversial (e) Pushshift (f) Praw

(g) Controversial (h) Pushshift (i) Praw

Figure 8.8: Armenia & Turkey graph representation of two polarized groups.

the most polarized groups are detected interior Armenia community, while in the case of Azerbaijan & Turkey, Azerbaijan is more polarized than Turkey. Also, looking at the Table 8.10, we observe that the percentage of agreement inside subgroups is positive (sign of edges is "+") while the external either does not exist or if so, it is negative (sign of edges is "-"). Moreover, in Table 8.11 the percentage of users from each subreddit that belongs in each one of the polarized groups referred. In particular, cross Armenia & Turkey and Azerbaijan & Turkey the polarized sub − communities consist of Armenian and common users who participate in both communities. In addition, we notice that as the discussion evolves over time, the existence of inter − polarization continues to exist, however with fewer participants.

Table 8.10: Nagorno–Karabahk. Percentage of agreement (positive edges, for short PE) within and cross (for short CE) polarized partitions G_1 and G_2 for Praw Random Eigensign. NC and C are for non – controversial and controversial posts.

Subreddits	G_1 **&** G_2 **nodes**	G_1 **PE**	G_2 **PE**	**CE**
Armenia & Azerbaijan	9%, 0.04%	90%	–	0.0%
Armenia & Azerbaijan (NC)	0.10%, 10%	–	90%	0.0%
Armenia & Azerbaijan (C)	0.56%, 12%	–	93%	18%
Armenia & Turkey	12%, 0.06%	90%	–	0.0%
Armenia & Turkey (NC)	0.10%, 12%	–	90%	42%
Armenia & Turkey (C)	11%, 0.76%	93%	–	46%
Azerbaijan & Turkey	14%, 0.43%	96%	100%	25%
Azerbaijan & Turkey (NC)	16%, 0.65%	96%	–	42
Azerbaijan & Turkey (C)	17%, 2%	97%	0.0%	12.5%

Table 8.11: Nagorno–Karabahk. Percentage of users from A and B subreddits (A & B subreddits) and from common users in each polarized group G_1 and G_2 using Praw Random Eigensign. NC and C are for non – controversial and controversial posts respectively.

	G_1			G_2		
Subreddits	A	B	Common	A	B	Common
Armenia & Azerbaijan	66%	2%	32%	100%	–	–
Armenia & Azerbaijan (NC)	75%	25%	–	68%	2%	30%
Armenia & Azerbaijan (C)	66%	34%	–	93%	–	7%
Armenia & Turkey	98%	–	2%	100%	–	–
Armenia & Turkey (NC)	100%	–	–	98%	–	2%
Armenia & Turkey (C)	99%	–	1%	100%	–	–
Azerbaijan & Turkey	82%	2%	16%	88%	–	12%
Azerbaijan & Turkey (NC)	82%	1%	17%	77%	8%	15%
Azerbaijan & Turkey (C)	84%	10%	6%	100%	–	–

Furthermore, in case of Armenia & Azerbaijan & Turkey (3−subreddits), similar to the cases of 2−subreddits, the most polarized subsets of individuals detected inside Armenia subreddit. Figure D.4 in Appendix D displays the two most polarized groups in 3−subreddits graph and Table C in Appendix C mentions the signed polarity score.

Nagorno−Karabahk conclusions. We conclude that unsigned algorithms does not detect inter−polarization cross Armenia & Azerbaijan and Azerbaijan & Turkey except *Betweenness* polarity score. Furthermore, controversial posts are more prone to polarity than non − controversial posts. In addition, the two most polarized partitions of the network do not detected cross communities instead, within one of them. In this case, Armenia is more polarized than Azerbaijan and Turkey and Azerbaijan is more polarized than Turkey. Moreover, users from different groups will not go to the opposite community to express negative views. Finally, the intensity of signed polarization diminishes as time evolves.

Police Violence. The next topic of discussion is about Police Violence and the main three subreddits are Unpopularopinion, Bad_Cop_No_Donut and BlackLivesMatter. We assume that Bad_Cop_No_Donut and BlackLivesMatter consist of users who are against police violence according to the description in Reddit and therefore there will be no conflict between them. Contrary to the community Unpopularopinion, our main assumption is that there will be intense disagreement and polarization with the members of the other two subreddits. At this point, we will check whether our assumptions will be confirmed or not through experimental analysis.

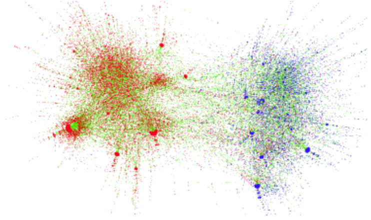
Initially, we look at the unsigned inter − polarization results. Table 8.12 indicates the relevant results between Unpopularopinion & Bad_Cop_No_Donut (see Tables C.7 and C.8 for Unpopularopinion & BlackLivesMatter and Bad_Cop_No_Donut & BlackLivesMatter in Appendix C). We observe that *Random Walks (P)* and *Betweenness* using METIS clustering detect high polarization cross subreddits. On the other hand, *Random Walks (P)*, *Random Walks (D)*, *Betweenness* and *Embeddings* using Real groups clustering quantify polarization cross subreddits. Furthermore, we notice that using Real groups clustering, the polarity in controversial conversations is more prone than non − controversial conversations. That is, *Random Walks (D)* score is equals to 0.52 for non − controversial conversations while it is equals to 0.88 for controversial conversations. Also, our assumption that no inter − polarization exists cross

Table 8.12: Unsigned and signed polarity score for Unpopularopinion & Bad_Cop_No_Donut for total, non−controversial and controversial posts applying either METIS or Real groups clustering and for two timestamps Pushshift and Praw.

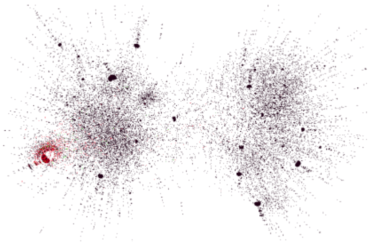
Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	$\mathbf{0.74}, \mathbf{0.76}, 0.49$	$\mathbf{0.77}, \mathbf{0.78}, \mathbf{0.85}$
Random walks (D)	$0.32, 0.35, 0.19$	$\mathbf{0.51}, \mathbf{0.52}, \mathbf{0.88}$
Betweenness	$\mathbf{0.87}, \mathbf{0.87}, \mathbf{0.84}$	$\mathbf{0.86}, \mathbf{0.86}, 0.38$
Embeddings	$0.44, 0.47, 0.04$	$\mathbf{0.54}, \mathbf{0.55}, \mathbf{0.60}$
GMCK	$0.19, 0.19, 0.18$	$0.08, 0.07, 0.13$
MBLB	$0.12, 0.11, 0.15$	$0.10, 0.10, 0.06$
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	$6.64, 6.67, 3.0$	$6.46, 6.52, 2.61$
Random eigensign	$2.49, 2.50, 1.91$	$2.39, 2.36, 1.89$
Greedy	$7.61, 7.61, 3.33$	$7.71, 7.68, 2.85$
Bansal	$4.93, 4.93, 3.27$	$4.58, 4.58, 2.8$
LocalSearch	$1.13, 1.43, 1.17$	$1.25, 1.14, 1.04$

Bad_Cop_No_Donut & BlackLivesMatter does not confirmed.

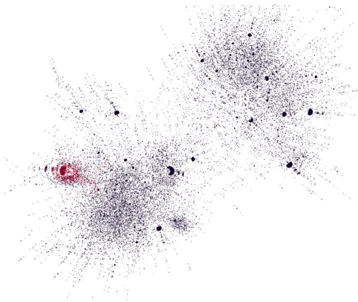
Continuing with signed inter − polarization analysis, we observe similar results to those from Hagia − Sophia discussions and the Nagorno − Karabahk conflict. That is, the highest polarity is detected within one of the two communities (see Figures 8.9 and 8.10 for Unpopularopinion & Bad_Cop_No_Donut and Unpopularopinion & BlackLivesMatter respectively, Figure D.5 for Bad_Cop_No_Donut & BlackLivesMatter in Appendix D). To be specific, between Unpopualopinion & Bad_Cop_No_Donut and Unpopualopinion & BlackLivesMatter, the most polarized groups are detected interior Unpopualopinion community, while in the case of Bad_Cop_No_Donut & BlackLivesMatter, Bad_Cop_No_Donut is more polarized than BlackLivesMatter. Also, looking at the Table 8.13, we observe that the percentage of agreement inside subgroups is high (sign of edges is "+") while the external either does not exist or if so, it is negative (sign of edges is "-"). Moreover, in Table 8.14 the percentage of users from each subreddit that belongs in each one of the polarized groups mention. Finally, we



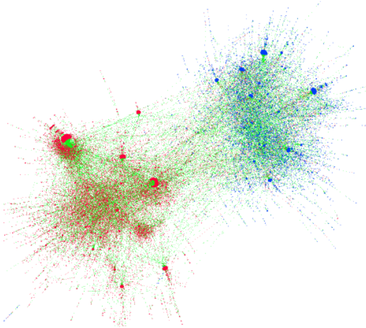
(a) Unpopularopinion & Bad_Cop_No_Donut



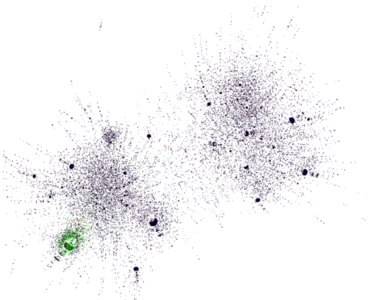
(b) Pushshift



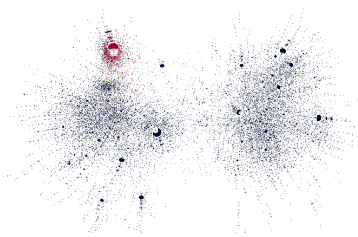
(c) Praw



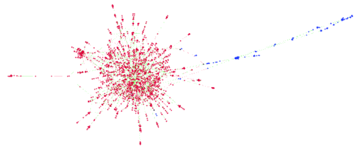
(d) Non−Controversial



(e) Pushshift



(f) Praw



(g) Controversial

(h) Pushshift

(i) Praw

Figure 8.9: Unpopularopinion & Bad_Cop_No_Donut graph representation of two polarized groups.

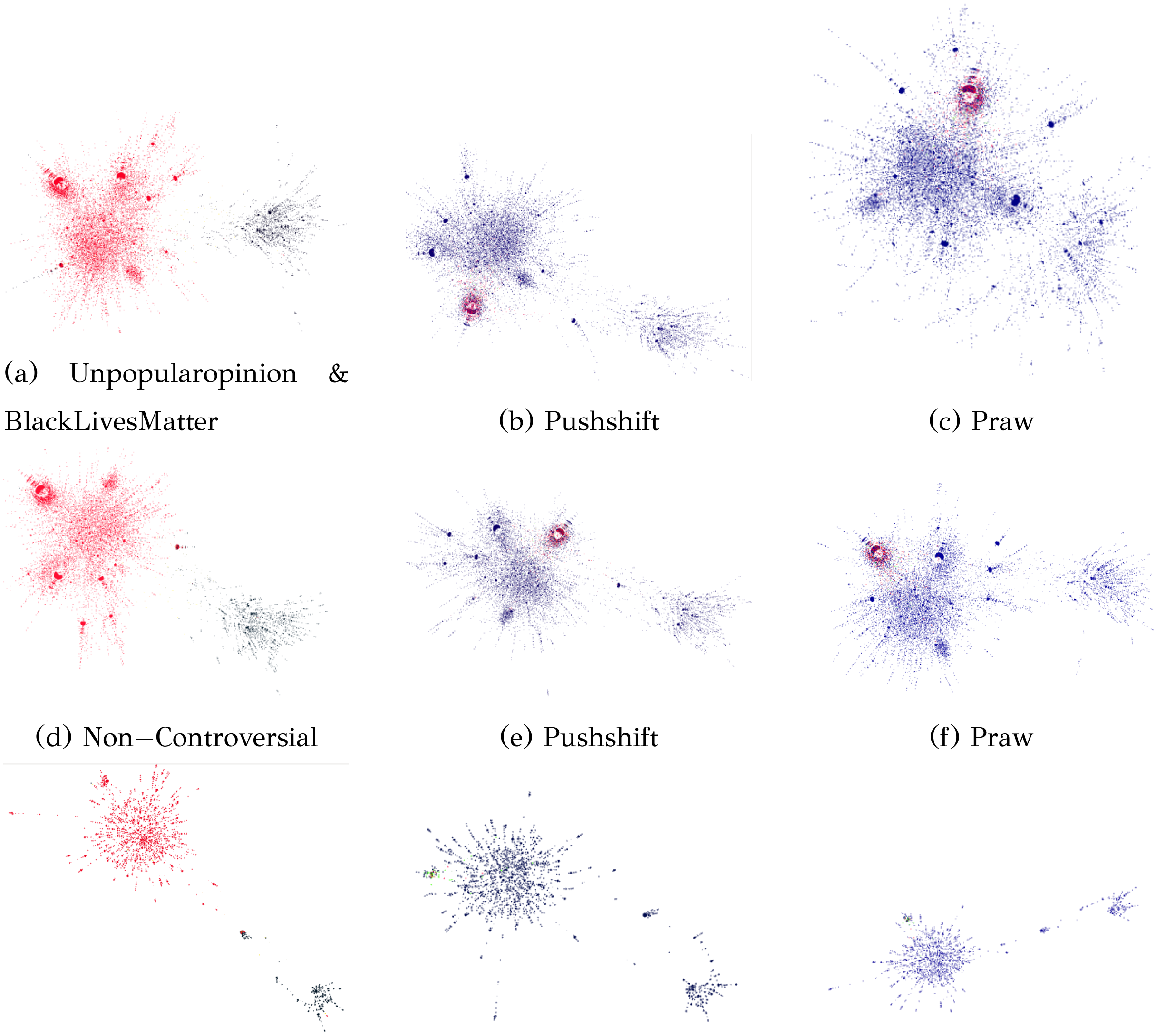
notice that as the discussion evolves over time, the existence of inter − polarization continues to exist, however with fewer participants.

Police Violence conclusions. We conclude that *Random Walk (P)*, *Random Walk (D)*, *Betweenness* and *Embeddings* detect unsigned inter − polarity cross all 2 − subreddits using Real groups clustering. Furthermore, polarity increases as more controversy exists in the network. Additionally, the two most polarized partitions of the network do not detected cross communities instead, within one of them. In this case, Unpopularopinion is more polarized than Bad_Cop_No_Donut and BlackLivesMatter while and Bad_Cop_No_Donut is more polarized than BlackLivesMatter. Moreover, users

Table 8.13: Police Violence. Percentage of agreement (positive edges, for short PE) within and cross (for short CE) polarized partitions G_1 and G_2 for Praw Random Eigensign. NC and C are for non − controversial and controversial posts.

Subreddits	G_1 **&** G_2 **nodes**	G_1 **PE**	G_2 **PE**	**CE**
Unpopualopinion & Bad_Cop_No_Donut	5%, 0.49%	99%	66%	23%
Unpopualopinion & Bad_Cop_No_Donut (NC)	0.51%, 5%	50%	97%	31%
Unpopualopinion & Bad_Cop_No_Donut (C)	1.29%, 1.29%	100%	90%	17%
Unpopualopinion & BlackLivesMatter	7%, 0.73%	98%	40%	24%
Unpopualopinion & BlackLivesMatter (NC)	7%, 0.85%	98%	50%	29%
Unpopualopinion & BlackLivesMatter (C)	1%, 1%	100%	90%	10%
Bad_Cop_No_Donut & BlackLivesMatter	0.68%, 4%	100%	99%	4%
Bad_Cop_No_Donut & BlackLivesMatter (NC)	4.17%, 0.75%	99%	100%	3%
Bad_Cop_No_Donut & BlackLivesMatter (C)	15%, 1%	98%	–	–

from different groups will not go to the opposite community to express negative views. Finally, the intensity of signed polarization diminishes as time evolves.



(a) Unpopularopinion & BlackLivesMatter

(b) Pushshift

(c) Praw

(d) Non−Controversial

(e) Pushshift

(f) Praw

(g) Controversial

(h) Pushshift

(i) Praw

Figure 8.10: Unpopularopinion & BlackLivesMatter graph representation of two polarized groups.

Table 8.14: Police Violence. Percentage of users from A and B subreddits (A & B subreddits) and from common users in each polarized group G_1 and G_2 using Praw Random Eigensign. NC and C are for non − controversial and controversial posts respectively.

	G_1			G_2		
Subreddits (A & B)	A	B	Common	A	B	Common
Unpopualopinion & Bad_Cop_No_Donut	99%	0.07%	1.13%	94%	2%	4%
Unpopualopinion & Bad_Cop_No_Donut (NC)	95%	2%	3%	98%	–	2%
Unpopualopinion & Bad_Cop_No_Donut (C)	100%	–	–	100%	–	–
Unpopualopinion & BlackLivesMatter	99%	0.50%	0.50%	100%	–	–
Unpopualopinion & BlackLivesMatter (NC)	99.5%	–	0.5%	100%	–	–
Unpopualopinion & BlackLivesMatter (C)	100%	–	–	100%	–	–
Bad_Cop_No_Donut & BlackLivesMatter	99%	1%	–	99%	–	–
Bad_Cop_No_Donut & BlackLivesMatter (NC)	99%	0.5%	0.5%	99%	–	1%
Bad_Cop_No_Donut & BlackLivesMatter (C)	–	100%	–	–	100%	–

COVID−19. The final topic we investigate is about Coronavirus pandemic namely, COVID−19. The communities we worked with are China_Flu and Coronavirus. [17] claims that initially these two communities consist of shared users who initially they had similar views but later the users started not to overlap and China_Flu users become more aggressive than Coronavirus. Therefore, based on this assumption we assume that users between these communities will be polarized and will not communicate with each other. Let us see if our assumption will be verified through experimental analysis.

Firstly, we look at the unsigned inter − polarization (see Table 8.15). We observe that none of the proposed algorithm do not detect polarity cross China_Flu & Coronavirus. This is happening because users interact with each other and therefore keep these two communities connected. However, as we have observed in the other topics of discussion, Hagia − Sophia, Nagorno − Karabakh conflict and Police Violence, *Betweenness* unsigned polarization score quantify high polarity cross subreddits. Fur-

Table 8.15: Unsigned and signed polarity score for China_Flu & Coronavirus for total, non−controversial and controversial posts applying either METIS or Real groups clustering for two timestamps Pushshift and Praw.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	0.19, 0.19, 0.42	0.11, 0.12, 0.27
Random walks (D)	0.0, 0.0, 0.13	0.0, 0.0, 0.05
Betweenness	**0.63**, **0.65**, **0.76**	**0.74**, **0.76**, **0.84**
Embeddings	0.13, 0.15, 0.38	0.29, 0.29, 0.25
GMCK	0.09, 0.10, 0.15	0.0, 0.0, 0.0
MBLB	0.10, 0.14, 0.13	0.10, 0.13, 0.15
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	87.24, 85.25, 10.96	73.47, 72.19, 7.87
Random eigensign	29.13, 27.91, 6.22	20.21, 19.05, 4.18
Greedy	89.23, 87.25, 13.53	76.58, 75.20, 10.48
Bansal	19.76, 19.94, 4.80	15.85, 15.99, 3.52
LocalSearch	4.14, 4.00, 2.21	3.18, 3.24, 1.67

Table 8.16: COVID−19. Percentage of agreement (positive edges, for short PE) within and cross (for short CE) polarized partitions G_1 and G_2 for Praw Random Eigensign. NC and C are for non − controversial and controversial posts.

Subreddits	G_1 **& G_2 nodes**	G_1 **PE**	G_2 **PE**	**CE**
China_Flu & Coronavirus	0.57%, 6.33%	67%	86%	42%
China_Flu & Coronavirus (NC)	0.58%, 6.32%	70%	86%	43%
China_Flu & Coronavirus (C)	0.53%, 6.12%	86%	89%	48%

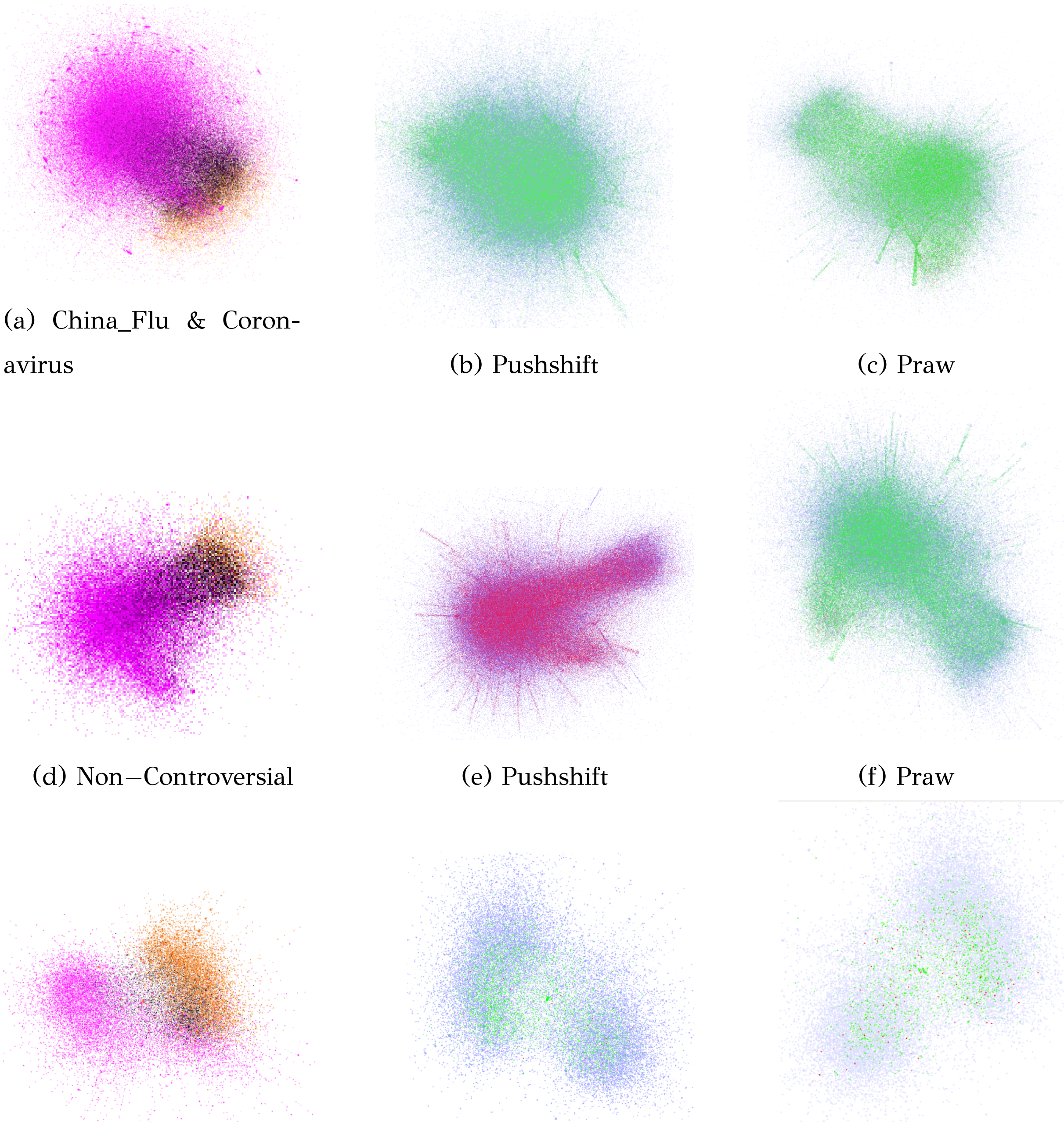
Table 8.17: COVID−19. Percentage of users from A & B subreddits and from common users in each polarized group G_1 and G_2 using Praw Random Eigensign. NC and C are for non − controversial and controversial posts respectively.

	G_1			G_2		
Subreddits (A & B)	A	B	Common	A	B	Common
China_Flu & Coronavirus	3%	90%	7%	4%	81%	15%
China_Flu & Coronavirus (NC)	3%	89%	8%	3%	82%	15%
China_Flu & Coronavirus (C)	36%	47%	17%	36%	42%	22%

thermore, we notice that in the case of controversial posts almost all the proposed algorithms measure higher polarity than non − controversial posts.

After that, we analyze signed inter − polarization between China_Flu & Coronavirus focusing on *Random Eigensign* method. Initially, Figure 8.11 shows the graph representation of two polarized groups applying the signed *Random Eigensign* algorithm. Looking at Tables 8.16 and 8.17, we observe that Coronavirus subreddit is more polarized than China_Flu. Particularly, the percentage of positive edges within each polarized group is high while cross them the percentage of negative edges is higher. Finally, we notice that inter − polarization decreases as time passes.

COVID−19 conclusions. We conclude that only *Betweenness* detect signed inter − polarity. Moreover, polarity increases as more controversy exists in the network. Also, Coronavirus subreddit is more polarized than China_Flu.



(a) China_Flu & Coronavirus (b) Pushshift (c) Praw

(d) Non−Controversial (e) Pushshift (f) Praw

(g) Controversial (h) Pushshift (i) Praw

Figure 8.11: China_Flu & Coronavirus graph representation of two polarized groups.

Overall conclusions. Initially, the unsigned *Betweenness* metric, regardless of the clustering technique (METIS and Real groups) and the topic of discussion, always detects inter − polarization between communities. On the contrary, *Random Walks (P)*, *Random Walks (D)* and *Embeddings* unsigned metrics in combination with Real group clustering detect inter − polarization in both non − controversial and controversial conversations. At the same time, we conclude that *Random Walks (P)*, *Random Walks (D)* and *Embeddings* metrics quantify unsigned polarization higher in the case of controversial submissions than non − controversial ones. In other words, controversial discussions increase inter − polarization of communities.

Furthermore, studying the signed inter − polarization, we conclude that polarization is not detected between actual communities, on the contrary, the most polarized subgroups are within one of the two communities. We have also noticed that the communication between the members of the opposite communities is quite a large percentage positive, i.e., there is an agreement between them. This leads to the conclusion that users who discuss a common topic and are in two communities who theoretically have opposing sights will not join to the opposite group to create controversy. Moreover, polarization decreases over time. That is, the users who consist the two most polarized groups stop joining them and the polarized subgroups consist of fewer participants.

8.3 What are the common motifs of user interaction in the case of non $-$ controversial and controversial posts?

The extraction of motifs either in static or temporal networks leads to conclusions by which we can discover hidden properties of the network structure, i.e. motifs that concern the interaction of network users when they refer to a particular type of discussion such as, politics, sports, education etc. or once more user interaction motifs either in controversial or non$-$controversial discussions. Also, basis of definition of temporal motifs, may find motifs that change as time progresses and thereby understand in depth the structure of the networks and the relation between members of the network. Accurately, we inspect temporal motifs in all generated AUG_d graphs either per community individually or per 2$-$subreddits and 3$-$subreddits separately and studying both controversial or non$-$controversial posts. We work on $\delta-$duration motifs where $\delta = 1$ month. All the possible 2$-$node and 3$-$node temporal motifs (36 in total) presented in Figure 8.12.

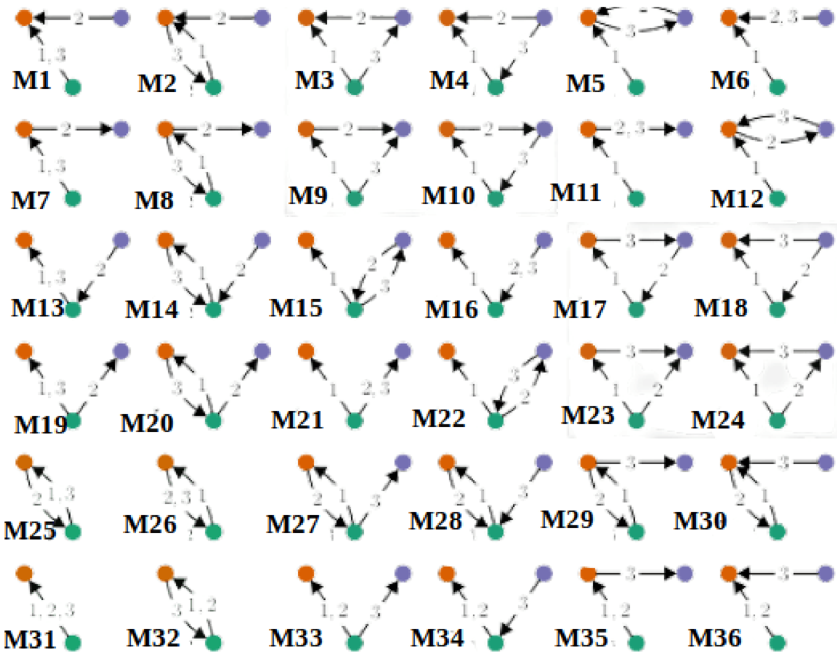


Figure 8.12: All 2$-$node and 3$-$node, 3$-$edge δ$-$temporal motifs. We index the 36 motifs **Mi**, $i \in [1, 36]$. **M25**, **M26**, **M31** and **M32** are the 2$-$node motifs and **M3**, **M4**, **M9**, **M10**, **M17**, **M18**, **M23** and **M24** are the eight triangles. The rest motifs are stars. The first edge in each motif is from the green to the orange node. The second edge is the same along each row, and the third edge is the same along each column.

At first, we investigate motifs between non$-$controversial and controversial posts

Table 8.18: Proportion of 4−top most frequent temporal motifs per 2−subreddits graphs for non−controversial posts.

Subreddits	Temporal motifs (M)			
Turkey & Greece	M15, 6.96%	M5, 6.76%	M30, 6.70%	M29, 6.16%
Turkey & Islam	M5, 7.06%	M30, 6.95%	M15, 6.44%	M12, 5.92%
Greece & Islam	M5, 6.97%	M30, 6.97%	M15, 6.40%	M29, 5.99%
Armenia & Azerbaijan	M36, 22.54%	M6, 16.86%	M1, 13.35%	M34, 3.84%
Armenia & Turkey	M36, 22.87%	M6, 17.07%	M1, 13.54%	M34, 3.76%
Azerbaijan & Turkey	M36, 12.31%	M6, 10.24%	M1, 8.57%	M34, 5.55%
China_Flu & Coronavirus	M1, 28.11%	M34, 20.84%	M6, 7.88%	M36, 5.21%
Unpopularopinion & Bad_Cop_No_Donut	M6, 12.13%	M36, 11.30%	M15, 5.68%	M5, 5.59%
Unpopularopinion & BlackLivesMatter	M6, 13.50%	M36, 12.91%	M1, 5.48%	M30, 5.27%
Bad_Cop_No_Donut & BlackLivesMatter	M15, 7.59%	M21, 6.96%	M22, 6.89%	M5, 6.40%

for all 2−subreddits graphs. In Tables 8.18 and 8.19, the proportion of 4−top most frequent temporal motifs mentioned. The first observation is that no motif stands out among the best of them. That is, the rate of occurrence of the most frequent motif from each topic of discussion is not particularly high so as to distinguish it from the rest of them whether we are referring to non−controversial or controversial case. Nevertheless, **M6** and **M36** were observed to appear in the 2−top most frequent motifs of non−controversial posts in discussions about Nagorno−Karabakh conflict and Police Violence. The communication between $3 - node$ users has the tendency two of the three nodes reply to the 3rd node without him giving a response back.

Figure 8.13: Most frequent motifs for 2−subreddits graphs for non−controversial posts.

Table 8.19: Proportion of 4−top most frequent temporal motifs per 2−subreddits graphs for controversial posts.

Subreddits	**Temporal motifs (M)**			
Turkey & Greece	M30, 7.27%	M5, 7.15%	M15, 7.10%	M29, 6.79%
Turkey & Islam	M5, 7.31%	M15, 7.24%	M30, 7.08%	M29, 6.62%
Greece & Islam	M15, 12.70%	M5, 9.77%	M30, 9.44%	M22, 7.16%
Armenia & Azerbaijan	M15, 6.66%	M22, 6.32%	M5, 6.31%	M12, 5.81%
Armenia & Turkey	M15, 6.77%	M22, 6.37%	M5, 6.19%	M27, 5.74%
Azerbaijan & Turkey	M5, 7.34%	M12, 6.57%	M15, 6.48%	M30, 6.28%
China_Flu & Coronavirus	M36, 7.50%	M6, 7.29%	M30, 6.20%	M28, 6.13%
Unpopularopinion & Bad_Cop_No_Donut	M5, 5.30%	M25, 5.02%	M12, 4.80%	M32, 4.73%
Unpopularopinion & BlackLivesMatter	M5, 5.29%	M25, 4.97%	M12, 4.77%	M32, 4.69%
Bad_Cop_No_Donut & BlackLivesMatter	M5, 7.64%	M12, 6.90%	M15, 6.57%	M22, 5.96%

About controversial posts, we perceive that motif **M5** appears in 3−top most frequent temporal motifs in 9 out of 10 in total 2−subreddits graphs that arise from all four topics of discussion. It should be noted that its incidence rate does not differ from the others but it happens to be in 3−top most frequent. The **M5** motif a communication trend between 3−users in which initially two users respond to the third and then the third user responds to the user who last contacted him. That is, in contrast to non−controversial posts, in this case there is an answer back to someone who contacted me. Finally, it should be noted that the most common motifs in 2−subreddits and 3−subreddits are a combination of the motifs that appear in each subreddit individually (see Tables in Appendix E.1, E.2 and E.3).

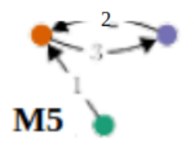


Figure 8.14: Most frequent motif for 2−subreddits graphs for controversial posts.

Conclusions. Overall, no motif was found with a significant incidence rate from the other motifs. However, we notice that in non−controversial posts, users lead not reply back to someone who replied to them. Finally, we observe that in controversial posts, users reply back to users who have contacted them.

8.4 What are the common motifs of comments in a discussion?

The existence of motifs within a network can bring interesting results for study. We have already studied the detection of temporal motifs between users interaction using information from aggregate user conversation AUG_u and AUG_d graphs. However, interesting motifs can also emerge by studying the structure of the discussions, in Reddit. Accurately, we inspect path temporal motifs in all CG graphs. Our goal is to compare the possible 3−node path motifs between non−controversial and controversial posts. By discovering such motifs, we may be able to understand the way in which the comments evolve in a discussion, i.e. a negative comment is followed by a positive comment or vice versa. All the possible 3−node temporal path motifs (8 in total) presented in Figure 8.15.

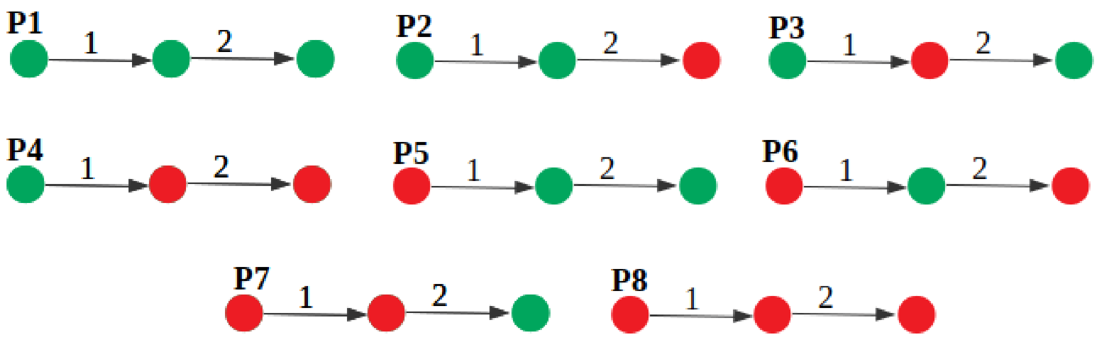


Figure 8.15: All 2−node, 2−edge motifs. We index the 8 motifs **Pi**, $i \in [1, 8]$. The red node has below or equal to zero upvote score and green node has over zero upvote score.

At the begging, we focus on non−controversial posts. In Table 8.20, we notice that the most likely to occur path motif is **P1** regardless of the time of collection of the score value, Pushshift or Praw. However, the probability of **P1** appearing is much lower with the use of Praw comment upvote score compared to the use of Pushshsift score. The explanation of this motif in a real discussion is interpreted as follows, in non−controversial conversations, it is more likely that users agree as the definition of what controversial post means.

Afterwards, investigating controversial posts, we observe that path motifs depend to a large extent on the time of collection of comment score (see Table 8.21). Especially, we notice that at the beginning of the discussion no dispute has yet developed between the participants. However, over time and as the discussions are more compre-

Table 8.20: Proportion of 2−top most frequent temporal path motifs per subreddit for non−controversial posts.

	Path motifs (P)			
Subreddits	**Pushshift**	**Praw**	**Pushshift**	**Praw**
Turkey	P1, 88%	P1, 61%	P3, 3%	P3, 17%
Greece	P1, 89%	P1, 59%	P3, 4%	P6, 11%
Islam	P1, 68%	P1, 54%	P3, 14%	P3, 19%
Armenia	P1, 87%	P1, 75%	P3, 4%	P3, 9%
Azerbaijan	P1, 97%	P1, 55%	P6, 0.68%	P3, 21%
Turkey	P1, 89%	P1, 80%	P3, 5%	P3, 9%
China_Flu	P1, 92%	P1, 76%	P3, 3%	P3, 9%
Coronavirus	P1, 83%	P1, 70%	P3, 6%	P3, 12%
Unpopularopinion	P1, 77%	P1, 57%	P3, 9%	P3, 16%
Bad_Cop_No_Donut	P1, 91%	P1, 71%	P3, 4%	P3, 15%
BlackLivesMatter	P1, 87%	P1, 80%	P3, 6%	P3, 10%

hensive we find that in controversial issues the most common pattern of discussion is the **P5** that is, after a negative comment (disagreement) two positive comments follow. Also, it is very interesting that the second most frequent path motif is **P6**. In this case, we observe that a positive comment a positive comment is between two negative ones.

Conclusions. Overall, the evolution of a discussion on a non − controversial topic of discussion follows a common motif that is, positive comments are followed by positive comments. In contrast to the controversial issues, in case a negative comment is detected then the probability of a positive response is quite high. Furthermore, if a disagreement arises then the most likely flow of the discussion is to alternate positive with negative views.

Table 8.21: Proportion of 2−top most frequent temporal path motifs per subreddit for controversial posts.

	Path motifs (P)			
Subreddits	**Pushshift**	**Praw**	**Pushshift**	**Praw**
Turkey	P1, 78%	P5, 50%	P5, 10%	P6, 24%
Greece	P1, 73%	P1, 49%	P5, 12%	P6, 20%
Islam	P1, 33%	P5, 60%	P5, 25%	P7, 15%
Armenia	P1, 80%	P5, 55%	P3, 6%	P1, 18%
Azerbaijan	P1, 78%	P5, 53%	P2, 10%	P1, 20%
Turkey	P1, 100%	P5, 63%	–	P6, 18%
China_Flu	P1, 87%	P5, 63%	P2, 4%	P6, 13%
Coronavirus	P1, 76%	P5, 57%	P5, 9%	P6, 13%
Unpopularopinion	P1, 63%	P5, 45%	P5, 12%	P6, 20%
Bad_Cop_No_Donut	P1, 85%	P5, 48%	P5, 10%	P6, 30%
BlackLivesMatter	P1, 65%	P5, 68%	P5, 18%	P6, 17%

8.5 Case study of content analysis

Another way of approaching our questions could be by analyzing the meaning of the comment. As already mentioned, we define comment positive if the number of upvotes is more than the number of downvotes. The limitation of this assumption is that a downvote does not provide a global quality assessment of a comment. Rather, a downvoted comment within a subreddit signifies that this particular subreddit perceives the comment as low quality. This is a localized definition of quality defined by the subreddit and it is consistent with Brunton's model of spam. This case study is an analysis of comments within and cross the two most polarized communities that *Random Eigensign* detects.

The analysis concerns the content of the comments about the conversion of Hagia – Sophia into a mosque topic. Some of the most positive (higher score) and negative (lower score) comments between the two polarized groups are reported below.

> *Imagine being a mathematics teacher and instead of doing something useful, like dedicating your life to improving the quality of education in Turkey and the quality of schools so the nation can have a smarter, more knowledgeable future generation (something actually encouraged by Islam), you dedicate your efforts to pestering the authorities to turn a bunch of former mosques back into mosques again.* **Score 4**

> *In Ottoman times when a Census was taken women were not counted. But Cattle were counted. Life was hundreds of years behind the West. Ataturk correctly worked out that Islam was the cause and sent people to be educated in the West. He modernized Turkey up to and beyond some European standards at the time on the basis of secularism. Now the backwards Islamists are coming back to power all rights are being stripped, freedoms stripped, education crumbling. It's all due to primitive Islamists like those in this video.* **Score 20**

> *That house has no spiritual, cultural, or historical value. Kemalism is not to worship Atatürk, but to follow the path it shows and to protect the existence of the country and the nation, to work for its future.* **Score 7**

> *This was a secular decision. Why would anyone from r/Islam or r/Pakistan complain when today was a victory for islam in Turkey? Cry more.* **Score -31**

It was a mosque for half a millennia, the real criticism should be why Ataturk had the audacity to oppose the will of the people and change it to a museum in the first place. Because of your worship of Ataturk you don't dare ask this basic question. Secondly, this change does not mean non Muslims can't visit it or pray in it as well. **Score -4**

Just to let you know, it's not just "Turkey" that has that knowing the contact or being related to get the job, it's happening everywhere in the world. **Score -8**

Looking at the positive comments cross polarized subgroups in Turkey subreddit, we notice that positive comments either refer to historical evidence or are in favor of preserving Hagia – Sophia as a museum. Additionally, analysis negative comments, we observe that the content either is ironic or expresses quite extreme views. In order for text analysis to be more valid, a stance or sentiment analysis tool must be applied.

Chapter 9

Conclusions and Future Work

This thesis investigate the problem of polarization in social media platforms during controversial discussions in Reddit. We define that the network is polarized if it is divided into two disjoint where either there is no communication between them or when communication exists, it expresses disagreement. Also, we define two types of polarization, unsigned and signed polarization. There is unsigned polarization within one (intra − polarization) or more (inter − polarization) communities if members are divided into two groups with opposing views on a specific topic of discussion and there is no communication between them. We define signed polarization within one (intra − polarization) or more (inter − polarization) communities if members within each community agree between them and disagree with members of the opposite community.

After that, we set the main research queries. Particularly, we investigate the unsigned and signed inter and intra polarization in four topics of discussions. Also, we investigate if controversial posts increase polarization in Reddit and finally, we explore common motifs of user interaction in the comments of a discussion in the case of controversial and non−controversial posts. To address these questions, we construct aggregated user conversation graphs (AUG), whose nodes are users and there is an edge between two users if the corresponding users responded to each other. We study both unsigned AUGs and signed AUGs where the sign of an edge is induced by the controversy score of the comments (a plus sign on an edge express agreement and a minus signed disagreement between the two users connected by the edge).

We conclude that, no unsigned intra−polarization has been detected in Reddit. On the contrary, signed intra−polarization exists and the members within each one of the two polarized communities agree and cross communities disagree. Furthermore, metrics based on *Random Walks*, *Betweenness* centrality of edges and *Embeddings* detect unsigned inter − polarization between communities. Also, controversial discussions increase inter − polarization In Reddit. Moreover, the two most polarized subgroups cross more subreddits are detected within one of the two subreddits that is, users who discuss a common topic and are in two communities who theoretically have opposing sights will not join to the opposite group to create controversy. In addition, polarization decreases over time. About temporal motifs, no motif was found with a significant incidence rate from the other motifs.

Finally, some limitations during our implementation have been observed. Specifically, we define a comment positive if the number of upvotes is more than downvotes. This assumptions is not strongly valid because a downvote does not provide a global quality assessment of a comment. Rather, a downvoted comment within a subreddit signifies that this particular subreddit perceives the comment as low quality. Furthermore, one more limitation is the way we select the topics we studied. Their choice was handpicked and and we focused mainly on controversial discussion something that can be considered a disadvantage. Consequently, some future extensions of this current dissertation is to apply tools for stance and sentiment comment analysis. So as to determine whether we can identify a negative comment both with its significance and with the number of votes it receives. Finally an additional extension is related to the temporal motifs in Reddit. Specifically, we would like to try to add sign to the motifs between users. In this way, we could find some correlation between user motifs and path motifs.

Bibliography

[1] K. Garimella, G. D. F. Morales, A. Gionis, and M. Mathioudakis, "Quantifying controversy on social media," *Trans. Soc. Comput.*, vol. 1, no. 1, Jan. 2018. [Online]. Available: https://doi.org/10.1145/3140565

[2] F. Bonchi, E. Galimberti, A. Gionis, B. Ordozgoiti, and G. Ruffo, "Discovering polarized communities in signed networks," in *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, ser. CIKM '19. New York, NY, USA: Association for Computing Machinery, 2019, pp. 961–970. [Online]. Available: https://doi.org/10.1145/3357384.3357977

[3] A.-M. Popescu and M. Pennacchiotti, "Detecting controversial events from twitter," in *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ser. CIKM '10. New York, NY, USA: Association for Computing Machinery, 2010, pp. 1873–1876. [Online]. Available: https://doi.org/10.1145/1871437.1871751

[4] S. Datta and E. Adar, "Extracting inter-community conflicts in reddit," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, no. 01, pp. 146–157, Jul. 2019. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/3217

[5] A. C. E. Lima, L. N. de Castro, and J. M. Corchado, "A polarity analysis framework for twitter messages," *Applied Mathematics and Computation*, vol. 270, pp. 756–767, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0096300315011145

[6] A. AlDayel and W. Magdy, "Stance detection on social media: State of the art and trends," 2020.

[7] K. Dey, R. Shrivastava, and S. Kaushik, "Twitter stance detection — a subjectivity and sentiment polarity inspired two-phase approach," in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, 2017, pp. 365–372.

[8] S. M. Mohammad, "9 - sentiment analysis: Detecting valence, emotions, and other affectual states from text," in *Emotion Measurement*, H. L. Meiselman, Ed. Woodhead Publishing, 2016, pp. 201–237. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081005088000096

[9] C. Largeron, A. Mardale, and M.-A. Rizoiu, "Linking user opinion dynamics and online discussions," 2021.

[10] S. M. Mohammad, P. Sobhani, and S. Kiritchenko, "Stance and sentiment in tweets," *ACM Trans. Internet Technol.*, vol. 17, no. 3, Jun. 2017. [Online]. Available: https://doi.org/10.1145/3003433

[11] P. Krejzl, B. Hourová, and J. Steinberger, "Stance detection in online discussions," 2017.

[12] P. Guerra, W. Meira Jr., C. Cardie, and R. Kleinberg, "A measure of polarization on social media networks based on community boundaries," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 7, no. 1, Jun. 2013. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14421

[13] H. Xiao, B. Ordozgoiti, and A. Gionis, *Searching for Polarization in Signed Graphs: A Local Spectral Approach*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 362–372. [Online]. Available: https://doi.org/10.1145/3366423.3380121

[14] P. Vachher, Z. Levonian, H.-F. Cheng, and S. Yarosh, *Understanding Community-Level Conflicts Through Reddit r/Place*. New York, NY, USA: Association for Computing Machinery, 2020, pp. 401–405. [Online]. Available: https://doi.org/10.1145/3406865.3418311

[15] M. Coletto, K. Garimella, A. Gionis, and C. Lucchese, "Automatic controversy detection in social media: A content-independent motif-based approach," *Online Social Networks and Media*, vol. 3-4, pp. 22–31, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2468696417300721

[16] A. Paranjape, A. R. Benson, and J. Leskovec, "Motifs in temporal networks," in *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, ser. WSDM '17. New York, NY, USA: Association for Computing Machinery, 2017, pp. 601–610. [Online]. Available: https://doi.org/10.1145/3018661.3018731

[17] J. S. Zhang, B. C. Keegan, Q. Lv, and C. Tan, "A tale of two communities: Characterizing reddit response to covid-19 through /r/china_flu and /r/coronavirus," 2020.

[18] S. Kumar, W. L. Hamilton, J. Leskovec, and D. Jurafsky, *Community Interaction and Conflict on the Web*. Republic and Canton of Geneva, CHE: International World Wide Web Conferences Steering Committee, 2018, pp. 933–943. [Online]. Available: https://doi.org/10.1145/3178876.3186141

[19] M. Mendoza, D. Parra, and Álvaro Soto, "Gene: Graph generation conditioned on named entities for polarity and controversy detection in social media," *Information Processing & Management*, vol. 57, no. 6, p. 102366, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S030645732030861X

[20] J. Hessel and L. Lee, "Something's brewing! early prediction of controversy-causing posts from discussion features," 2019.

[21] A. J. Morales, J. Borondo, J. C. Losada, and R. M. Benito, "Measuring political polarization: Twitter shows the two sides of venezuela," *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 25, no. 3, p. 033114, 2015. [Online]. Available: https://doi.org/10.1063/1.4913758

[22] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of Anthropological Research*, vol. 33, no. 4, pp. 452–473, 1977. [Online]. Available: https://doi.org/10.1086/jar.33.4.3629752

[23] L. A. Adamic and N. Glance, "The political blogosphere and the 2004 u.s. election: Divided they blog," in *Proceedings of the 3rd International Workshop on Link Discovery*, ser. LinkKDD '05. New York, NY, USA: Association for Computing Machinery, 2005, pp. 36–43. [Online]. Available: https://doi.org/10.1145/1134271.1134277

[24] G. Karypis and V. Kumar, “A parallel algorithm for multilevel graph partitioning and sparse matrix ordering,” *Journal of Parallel and Distributed Computing*, vol. 48, no. 1, pp. 71–95, 1998. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0743731597914039

[25] L. Akoglu, “Quantifying political polarity based on bipartite opinion networks,” *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 8, no. 1, May 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14524

[26] S. M. Mohammad, “9 - sentiment analysis: Detecting valence, emotions, and other affectual states from text,” in *Emotion Measurement*, H. L. Meiselman, Ed. Woodhead Publishing, 2016, pp. 201–237. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780081005088000096

[27] P. Krejzl, B. Hourová, and J. Steinberger, “Stance detection in online discussions,” 2017.

[28] A. C. E. Lima, L. N. de Castro, and J. M. Corchado, “A polarity analysis framework for twitter messages,” *Applied Mathematics and Computation*, vol. 270, pp. 756–767, 2015. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0096300315011145

[29] M. Jacomy, T. Venturini, S. Heymann, and M. Bastian, “Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software,” *PLOS ONE*, vol. 9, no. 6, pp. 1–12, 06 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0098679

[30] M. Charikar, “Greedy approximation algorithms for finding dense components in a graph,” in *Approximation Algorithms for Combinatorial Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 84–95.

[31] N. Bansal, A. Blum, and S. Chawla, “Correlation clustering,” *Machine Learning*, vol. 56, no. 1, pp. 89–113, Jul 2004.

[32] M. Latapy, “Main-memory triangle computations for very large (sparse (power-law)) graphs,” *Theoretical Computer Science*, vol. 407, no. 1, pp. 458–473, 2008. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0304397508005392

[33] J. Baumgartner, S. Zannettou, B. Keegan, M. Squire, and J. Blackburn, "The pushshift reddit dataset," *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 14, no. 1, pp. 830–839, May 2020. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/7347

Appendix A

Algorithms

Algorithm A.1 Metis Clustering

Require: Undirected graph $G(V, E)$

Ensure: Two polarized partitions X and Y of the graph G

1: *// Phase 1: Coarsening phase*

2: Partition V nodes into k subsets, $V_0, V_1, \ldots, V_k$ such that $V_i \cap V_j = \emptyset$ for $i \neq j$ and $|V_i| = \frac{|V|}{k}$ subject to the sum of the edge−weights of E whose incident vertices belong to different subsets is minimized.

3: Generate a sequence of graphs $G_0, G_1, \ldots, G_k$, where $|V_i| > |V_j|$ and $0 \leq i \leq j \leq k$.

4: *// Phase 2: Partitioning phase*

5: A partition P_k of the graph $G_k = (V_k, E_k)$ is computed based on the Kernighan & Lin algorithm. The size of each patrition is same.

6: *// Phase 3: Uncoarsening phase*

7: The partition P_k of G_k is projected back to G_0 by going through intermediate partitions $P_{k-1}, P_{k-2}, \ldots, P_1, P_0$. If X and Y are the two parts of the bisection, a refinement algorithm selects $X' \subset X$ and $Y' \subset Y$ such that $\{X - X'\} \cup Y'$, and $\{Y - Y'\} \cup X'$ is a bisection with a smaller edge−cut.

8: Return two partitions of the graph, X' and Y'

Appendix B

Size of graphs

Table B.1: Size of 2−subreddits AUG_d and AUG_u graphs.

Subreddits	**Nodes** (AUG_d)	**Edges** (AUG_d)	**Nodes** (AUG_u)	**Edges** (AUG_u)
Turkey & Greece	560	2,435	1,288	2,331
Turkey & Islam	926	4,943	1,883	3,603
Greece & Islam	673	3,676	1,288	2,330
Armenia & Azerbaijan	2,876	92,614	4,133	34,272
Armenia & Turkey	2,204	78,151	3,251	27,100
Azerbaijan & Turkey	1,267	15,502	2,069	8,540
China_Flu & Coronavirus	180,554	1,757,452	279,692	1,394,768
Unpopularopinion & Bad_Cop_No_Donut	9,396	43,917	28,259	42,695
Unpopularopinion & BlackLivesMatter	6,624	32,801	18,975	29,017
Bad_Cop_No_Donut & BlackLivesMatter	3,580	13,507	14,844	20,274

Table B.2: Size of 2−subreddits AUG_d and AUG_u graphs for non−controversial posts.

Subreddits	**Nodes (AUG_d)**	**Edges (AUG_d)**	**Nodes (AUG_u)**	**Edges (AUG_u)**
Turkey & Greece	411	1,656	1,044	1,731
Turkey & Islam	803	4,190	1,669	3,033
Greece & Islam	638	3,487	1,222	2,183
Armenia & Azerbaijan	2,698	87,420	3,886	32,266
Armenia & Turkey	2,021	73,911	2,977	25,399
Azerbaijan & Turkey	1,196	14,274	1,970	7,993
China_Flu & Coronavirus	160,447	1,704,536	275,763	1,346,856
Unpopularopinion & Bad_Cop_No_Donut	8,549	37,856	26,705	39,619
Unpopularopinion & BlackLivesMatter	5,739	26,747	17,133	25,634
Bad_Cop_No_Donut & BlackLivesMatter	3,485	13,030	14,345	19,598

Table B.3: Size of 2−subreddits AUG_d and AUG_u graphs for controversial posts.

Subreddits	**Nodes (AUG_d)**	**Edges (AUG_d)**	**Nodes (AUG_u)**	**Edges (AUG_u)**
Turkey & Greece	106	410	414	600
Turkey & Islam	112	435	404	579
Greece & Islam	16	43	64	83
Armenia & Azerbaijan	633	4,494	1,054	2,664
Armenia & Turkey	507	3,630	910	2,314
Azerbaijan & Turkey	183	994	369	591
China_Flu & Coronavirus	9,144	52,916	20,840	52,292
Unpopularopinion & Bad_Cop_No_Donut	837	4,506	2,159	2,950
Unpopularopinion & BlackLivesMatter	832	4,447	2,366	3,190
Bad_Cop_No_Donut & BlackLivesMatter	27	100	325	368

Table B.4: Size of 3−subreddits AUG_d and AUG_u graphs from total, non−controversial (NC) and controversial (C) posts.

Subreddits	**Nodes** (AUG_d)	**Edges** (AUG_d)	**Nodes** (AUG_u)	**Edges** (AUG_u)
Turkey & Greece & Islam	1,072	5,582	2,180	4,132
Turkey & Greece & Islam (NC)	925	4,716	1,932	3,476
Turkey & Greece & Islam (C)	112	435	467	662
Armenia & Azerbaijan & Turkey	2,980	93,404	4,350	34,910
Armenia & Azerbaijan & Turkey (NC)	2,785	88,033	4,085	32,786
Armenia & Azerbaijan & Turkey (C)	657	4,609	1,128	2,784
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter	9,897	45,592	30,816	46,026
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter (NC)	8,972	39,237	28,890	42,455
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter (C)	837	4,506	2,487	3,324

Table B.5: Percentage of users (first line) and percentage of positive edges using Pushshift and Praw sign (second line) per subreddit in 2−subreddits AUG_u graphs (A & B). Common are the users who participate in both subreddits.

Subreddit	Subreddit A	Subreddit B	Common users
Turkey & Greece	73.99% (92%, 80%)	23.52% (91%, 76%)	2.48% (96%, 82%)
Turkey & Islam	48.59% (92%, 80%)	47.58% (75%, 68%)	3.82% (86%, 74%)
Greece & Islam	25.07% (92%, 76%)	73.99% (74%, 67%)	0.93% (84%, 75%)
Armenia & Azerbaijan	55.89% (92%, 86%)	29.32% (91%, 86%)	14.78% (90%, 85%)
Armenia & Turkey	86.65% (92%, 87%)	10.24% (97%, 82%)	3.10% (89%, 77%)
Azerbaijan & Turkey	79.02% (93%, 88%)	12.03% (97%, 85%)	8.94% (95%, 88%)
China_Flu & Coronavirus	8.21% (95%, 85%)	85.79% (91%, 84%)	5.99% (92%, 83%)
Unpopularopinion & Bad_Cop_No_Donut	56.56% (85%, 77%)	42.11% (95%, 79%)	1.32% (91%, 74%)
Unpopularopinion & BlackLivesMatter	85.82% (85%, 77%)	13.81% (95%, 89%)	0.35% (94%, 89%)
Bad_Cop_No_Donut & BlackLivesMatter	81.65% (95%, 79%)	17.46% (95%, 89%)	0.88% (98%, 78%)

Table B.6: Percentage of users (first line) and percentage of positive edges using Pushshift and Praw sign (second line) per subreddit in 2−subreddits AUG_u graphs (A & B) for non−controversial. Common are the users who participate in both subreddits.

Subreddit	**Subreddit A**	**Subreddit B**	**Common users**
Turkey & Greece	72.70% (92%, 81%)	25.19% (92%, 76%)	2.10% (95%, 82%)
Turkey & Islam	43.43% (92%, 81%)	53.08% (74%, 67%)	3.47% (85%, 72%)
Greece & Islam	22.99% (92%, 76%)	76.26% (74%, 67%)	0.73% (84%, 75%)
Armenia & Azerbaijan	55.01% (92%, 87%)	31.16% (92%, 87%)	13.81% (91%, 85%)
Armenia & Turkey	87.30% (92%, 87%)	10.14% (97%, 84%)	2.55% (91%, 78%)
Azerbaijan & Turkey	80.81% (94%, 89%)	11.31% (97%, 86%)	7.86% (96%, 89%)
China_Flu & Coronavirus	7.86% (95%, 86%)	86.36% (91%, 84%)	5.76% (91%, 81%)
Unpopularopinion & Bad_Cop_No_Donut	54.48% (86%, 78%)	44.25% (95%, 79%)	1.25% (91%, 74%)
Unpopularopinion & BlackLivesMatter	86.57% (86%, 78%)	13.13% (95%, 90%)	0.29% (95%, 92%)
Bad_Cop_No_Donut & BlackLivesMatter	83.72% (95%, 79%)	15.41% (95%, 90%)	0.86% (98%, 77%)

Table B.7: Percentage of users (first line) and percentage of positive edges using Pushshift and Praw sign (second line) per subreddit in 2−subreddits AUG_u graphs (A & B) for controversial. Common are the users who participate in both subreddits.

Subreddit	Subreddit A	Subreddit B	Common users
Turkey & Greece	84.54% (91%, 78%)	15.21% (87%, 81%)	0.24% 100%
Turkey & Islam	85.64% (91%, 78%)	13.11% (83%, 87%)	1.23% (90%, 90%)
Greece & Islam	100% (87%, 81%)	– –	– –
Armenia & Azerbaijan	72.58% (90%, 83%)	21.72% (84%, 77%)	5.69% (86%, 78%)
Armenia & Turkey	89.12% (90%, 83%)	9.45% (100%, 77%)	1.42% (87%, 76%)
Azerbaijan & Turkey	73.71% (84%, 77%)	22.22% (100%, 77%)	4.06% (97%, 83%)
China_Flu & Coronavirus	35.19% (91%, 82%)	56.31% (92%, 82%)	8.49% (92%, 81%)
Unpopularopinion & Bad_Cop_No_Donut	94.11% (82%, 70%)	5.60% (96%, 74%)	0.27% (85%, 78%)
Unpopularopinion & BlackLivesMatter	85.96% (82%, 70%)	13.73% (92%, 88%)	0.29% (88%, 86%)
Bad_Cop_No_Donut & BlackLivesMatter	0.0% –	99.69% (92%, 88%)	0.30% 100%

Table B.8: Percentage of users per subreddit in 3−subreddits AUG_u graphs (A & B & C) for total, non−controversial and controversial posts. Common are the users who participate in both subreddits, A & B, A & C and B & C respectively.

Subreddit	Sub A	Sub B	Sub C	Common users
Turkey & Greece & Islam	40.77%	13.62%	40.82%	1.46%, 3.30%, 0.55%
Turkey & Greece & Islam (NC)	36.59%	13.61%	45.60%	1.13%, 3%, 0.46%
Turkey & Greece & Islam (C)	73.87%	13.49%	11.34%	0.21%, 1.07%, 0.0%
Armenia & Azerbaijan & Turkey	52.32%	25.19%	4.94%	14.04%, 2.32%, 4.25%
Armenia & Azerbaijan & Turkey (NC)	51.70%	27.12%	4.82%	13.14%, 1.86%, 3.79%
Armenia & Azerbaijan & Turkey (C)	66.93%	19.23%	6.56%	5.31%, 1.15%, 1.32%
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter	51.69%	38.22%	8.28%	1.21%, 0.21%, 0.42%
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter (NC)	50.23%	40.50%	7.55%	1.15%, 0.17%, 0.42%
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter (C)	81.58%	4.86%	13.06%	0.24%, 0.28%, 0.04%

Appendix C

Unsigned and signed polarity scores

Table C.1: Unsigned and signed polarity score for Turkey & Islam and for total, non–controversial and controversial posts applying either METIS or Real groups clustering and for two distinct timestamps Pushshift and Praw.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	**0.58**, **0.57**, 0.33	**0.54**, **0.58**, **0.61**
Random walks (D)	**0.53**, **0.60**, 0.32	0.44, **0.50**, **0.77**
Betweenness	**0.67**, **0.66**, **0.60**	**0.77**, **0.66**, 0.46
Embeddings	**0.55**, **0.51**, 0.16	**0.53**, **0.54**, 0.21
GMCK	0.19, 0.16, 0.15	0.03, 0.09, 0.00
MBLB	0.14, 0.13, 0.14	0.14, 0.13, 0.20
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	6.07, 3.47, 3.81	3.5, 3.57, 3.0
Random eigensign	3.89, 2.35, 2.59	2.24, 2.26, 2.43
Greedy	6.72, 4.73, 4.14	5.69, 4.54, 3.30
Bansal	3.36, 3.36, 2.8	3.2, 3.18, 2.5
LocalSearch	1.41, 1.31, 1.32	1.32, 1.08, 0.80

Table C.2: Unsigned and signed polarity score for Greece & Islam and for total, non−controversial and controversial posts applying either METIS or Real groups clustering and for two distinct timestamps Pushshift and Praw.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	$0.45, 0.45, 0.25$	$\mathbf{0.85}, \mathbf{0.87}, -$
Random walks (D)	$0.33, 0.43, \mathbf{0.85}$	$\mathbf{0.90}, \mathbf{0.92}, -$
Betweenness	$\mathbf{0.85}, \mathbf{0.77}, \mathbf{0.65}$	$\mathbf{0.73}, \mathbf{0.73}, -$
Embeddings	$0.29, 0.32, 0.48$	$\mathbf{0.68}, \mathbf{0.69}, -$
GMCK	$0.04, 0.10, 0.30$	$0.17, 0.20, -$
MBLB	$0.12, 0.12, 0.30$	$0.09, 0.09, -$
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	3.47, 3.47, 3.16	3.55, 3.55, 3.16
Random eigensign	2.31, 2.44, 2.60	2.19, 2.25, 2.41
Greedy	2.55, 3.6, 3.15	3.77, 2.66, 3.15
Bansal	3.42, 3.38, 2.8	3.42, 3.2, 2.8
LocalSearch	1.03, 1.15, 1.8	1.09, 1.35, 1.54

Table C.3: Signed polarity score for Turkey & Greece & Islam for two timestamps Pushshift and Praw and for total, non−controversial and controversial posts.

Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	$6.0, 3.47, 3.81$	$5.10, 3.57, 3.0$
Random eigensign	$3.89, 2.28, 2.75$	$3.07, 2.32, 2.38$
Greedy	$6.86, 5.06, 4.14$	$5.75, 4.22, 3.30$
Bansal	$3.42, 3.38, 2.80$	$3.42, 3.18, 2.80$
LocalSearch	$1.18, 1.11, 1.06$	$1.21, 1.08, 1.10$

Table C.4: Unsigned and signed polarity score for Armenia & Turkey and for total, non−controversial and controversial posts applying either METIS or Real groups clustering for two timestamps Pushshift and Praw.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	$0.18, 0.20, 0.23$	$\mathbf{0.57}$, $\mathbf{0.67}, \mathbf{0.55}$
Random walks (D)	$0.00, 0.00, 0.20$	$0.21, 0.34, \mathbf{0.70}$
Betweenness	$\mathbf{0.82}, \mathbf{0.83}, \mathbf{0.73}$	$\mathbf{0.58}, \mathbf{0.55}, 0.45$
Embeddings	$0.12, 0.11, 0.12$	$\mathbf{0.50}, \mathbf{0.53}, \mathbf{0.51}$
GMCK	$0.00, 0.00, 0.05$	$0.17, 0.22, 0.15$
MBLB	$0.12, 0.13, 0.11$	$0.18, 0.19, 0.15$
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	80.29, 79.85, 13.06	74.09, 73.70, 12.32
Random eigensign	53.26, 54.81, 8.93	49.75, 50.55, 8.50
Greedy	80.43, 79.90, 13.24	74.14, 73.71, 12.49
Bansal	50.15, 49.51, 7.65	42.23, 42.61, 6.55
LocalSearch	6.38, 8.69, 1.64	7.14, 7.68, 1.51

Table C.5: Unsigned and signed polarity score for Azerbaijan & Turkey and for total, non − controversial and controversial posts applying either METIS or Real groups clustering.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	$0.13, 0.12, 0.25$	$0.14, 0.12, 0.35$
Random walks (D)	$0.004, 0.07, 0.22$	$0.00, 0.00, \mathbf{0.52}$
Betweenness	$\mathbf{0.57}, \mathbf{0.56}, \mathbf{0.62}$	$\mathbf{0.57}, \mathbf{0.67}, 0.43$
Embeddings	$0.10, 0.05, 0.11$	$0.25, 0.26, 0.33$
GMCK	$0.00, 0.00, 0.12$	$0.00, 0.00, 0.10$
MBLB	$0.12, 0.14, 0.14$	$0.07, 0.07, 0.10$
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	27.47, 27.38, 4.45	26.62, 26.45, 4.13
Random eigensign	16.10, 16.54, 3.02	16.37, 15.82, 2.98
Greedy	27.69, 27.57, 4.60	26.88, 26.69, 4.23
Bansal	17.17, 17.17, 3.14	17.54, 17.49, 3.38
LocalSearch	4.31, 3.49, 1.40	4.04, 3.47, 1.11

Table C.6: Signed polarity score for Armenia & Azerbaijan & Turkey for two timestamps Pushshift and Praw and for total, non−controversial and controversial posts.

Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	$80.37, 79.86, 13.06$	$74.08, 73.71, 12.32$
Random eigensign	$55.66, 54.67, 9.05$	$50.90, 50.75, 8.28$
Greedy	$80.48, 79.94, 13.24$	$74.14, 73.72, 12.49$
Bansal	$50.32, 49.68, 7.65$	$42.36, 42.75, 6.55$
LocalSearch	$7.34, 7.13, 1.98$	$6.02, 6.45, 1.85$

Table C.7: Unsigned and signed polarity score for Unpopularopinion & BlackLivesMatter for total, non−controversial and controversial posts applying either METIS or Real groups clustering for two timestamps Pushshift and Praw.

Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	**0.57**, **0.58**, **0.56**	**0.86**, **0.86**, **0.90**
Random walks (D)	0.17, 0.13, 0.20	**0.69**, **0.78**, **0.89**
Betweenness	**0.91**, **0.91**, **0.94**	**0.72**, **0.66**, **0.73**
Embeddings	0.14, 0.11, 0.14	**0.65**, **0.65**, **0.73**
GMCK	0.14, 0.15, 0.18	0.08, 0.06, 0.08
MBLB	0.11, 0.10, 0.14	0.06, 0.06, 0.06
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	6.64, 6.67, 3.0	6.46, 6.52, 2.61
Random eigensign	2.41, 2.56, 1.72	2.45, 2.43, 1.92
Greedy	7.61, 7.57, 4.0	7.72, 7.66, 2.25
Bansal	4.93, 4.93, 3.27	4.58, 4.58, 2.8
LocalSearch	1.42, 1.24, 1.15	1.18, 1.41, 0.94

Table C.8: Unsigned and signed polarity score for Bad_Cop_No_Donut & BlackLives-Matter and for total, non−controversial and controversial posts applying either METIS or Real groups clustering for two timestamps Pushshift and Praw.

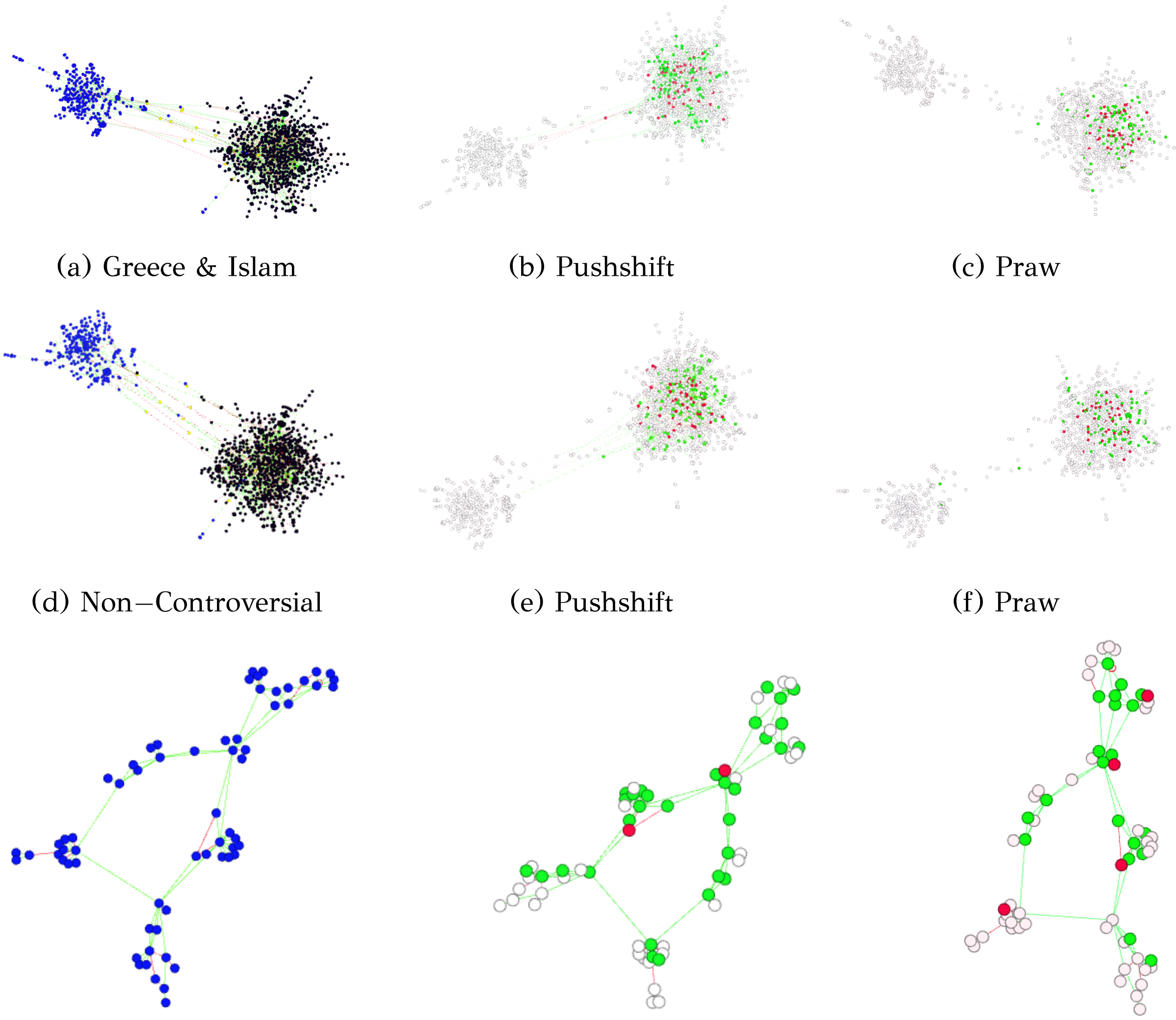
Algorithms (unsigned)	**Polarity METIS**	**Polarity Real**
Random walks (P)	**0.56**, **0.58**, **0.56**	**0.72**, **0.69**, –
Random walks (D)	0.13, 0.17, **0.53**	0.39, **0.54**, –
Betweenness	**0.89**, **0.85**, **0.90**	**0.87**, **0.77**, –
Embeddings	0.12, 0.34, 0.19	**0.53**, **0.50**, –
GMCK	0.16, 0.16, 0.22	0.03, 0.05, –
MBLB	0.11, 0.11, 0.33	0.13, 0.13, –
Algorithms (signed)	**Polarity Pushshift**	**Polarity Praw**
Eigensign	3.83, 3.83, 2.35	3.50, 3.50, 2.21
Random eigensign	2.24, 2.07, 1.92	2.03, 2.05, 1.66
Greedy	5.90, 5.84, 2.8	4.91, 4.91, 2.8
Bansal	3.71, 3.71, 2.0	3.27, 3.27, 2.0
LocalSearch	1.36, 1.47, 1.40	1.16, 0.94, 1.20

Table C.9: Signed polarity score for Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter for two timestamps Pushshift and Praw and for total, non−controversial and controversial posts.

Algorithms (signed)	**Polarity (Pushshift)**	**Polarity (Praw)**
Eigensign	6.64, 6.67, 3.0	6.46, 6.52, 2.61
Random eigensign	2.48, 2.51, 2.10	2.50, 2.49, 1.76
Greedy	7.63, 7.63, 3.2	7.68, 7.68, 2.29
Bansal	4.93, 4.93, 1.39	4.58, 4.58, 2.8
LocalSearch	1.09, 1.39, 1.16	1.23, 1.10, 1.02

Appendix D

Graphic representation of graphs

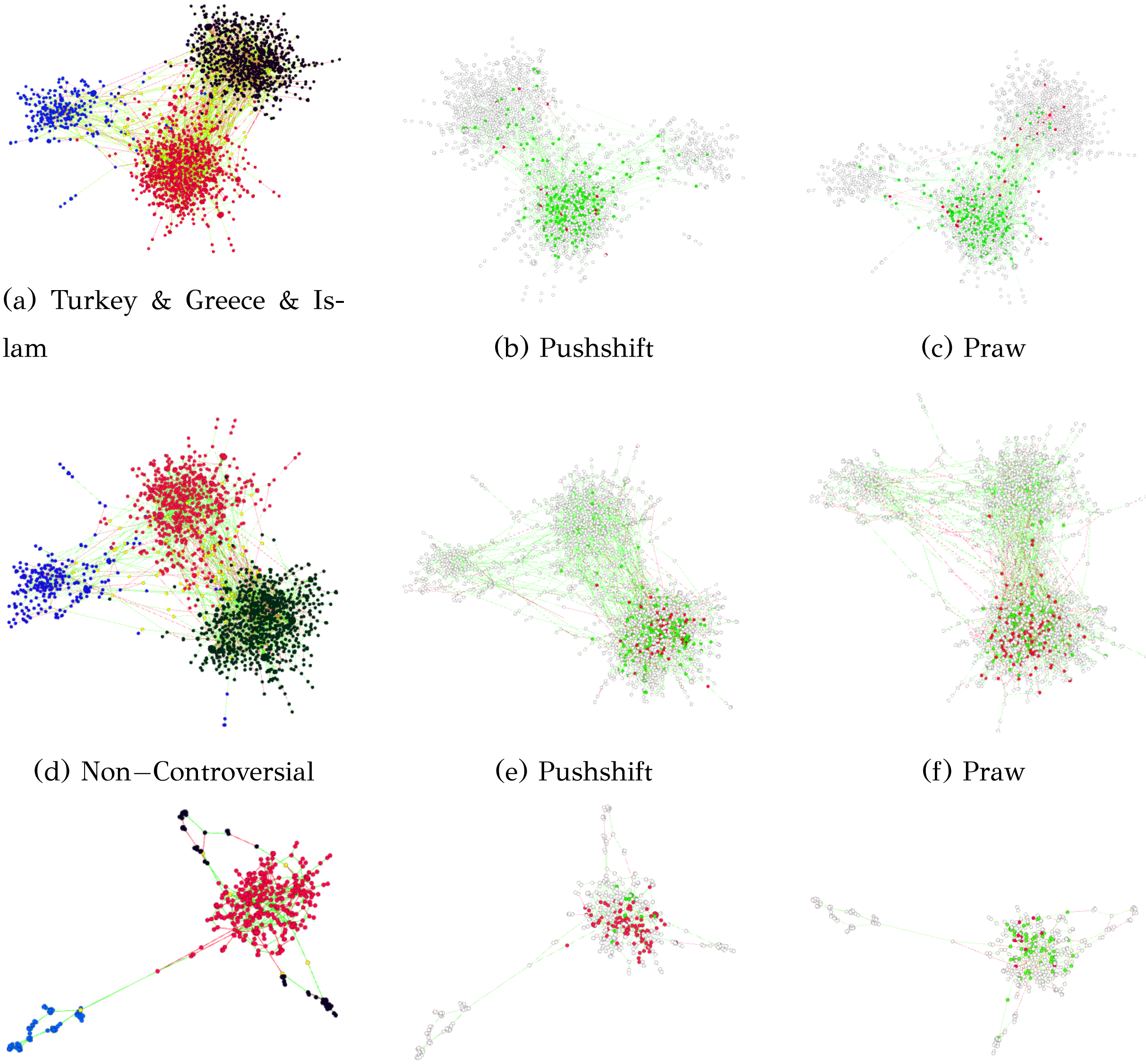


(a) Greece & Islam (b) Pushshift (c) Praw

(d) Non–Controversial (e) Pushshift (f) Praw

(g) Controversial (h) Pushshift (i) Praw

Figure D.1: Greece & Islam graph representation of two polarized groups.



(a) Turkey & Greece & Islam

(b) Pushshift

(c) Praw

(d) Non−Controversial

(e) Pushshift

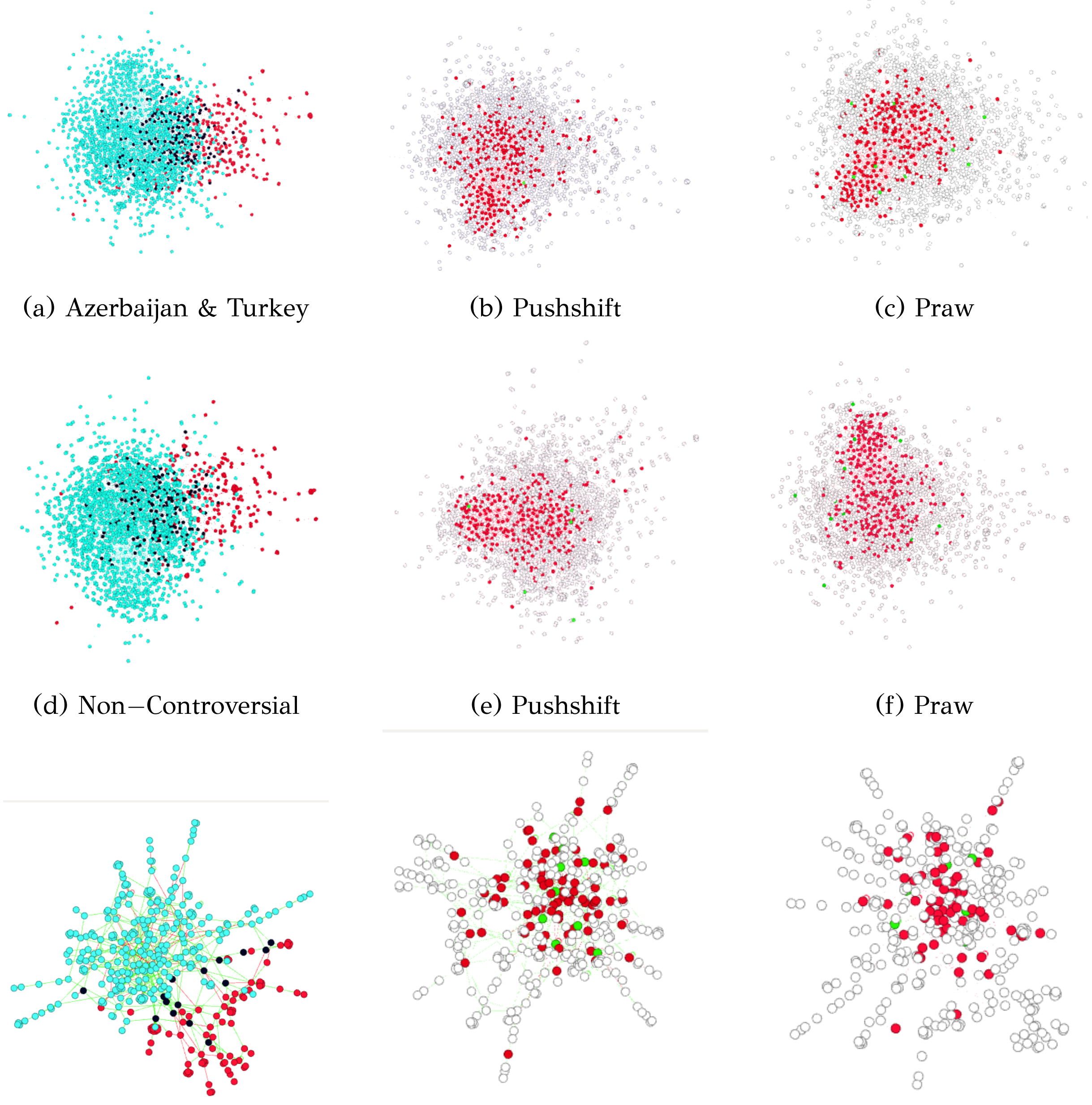
(f) Praw

(g) Controversial

(h) Pushshift

(i) Praw

Figure D.2: Turkey & Greece & Islam graph representation of two polarized groups.

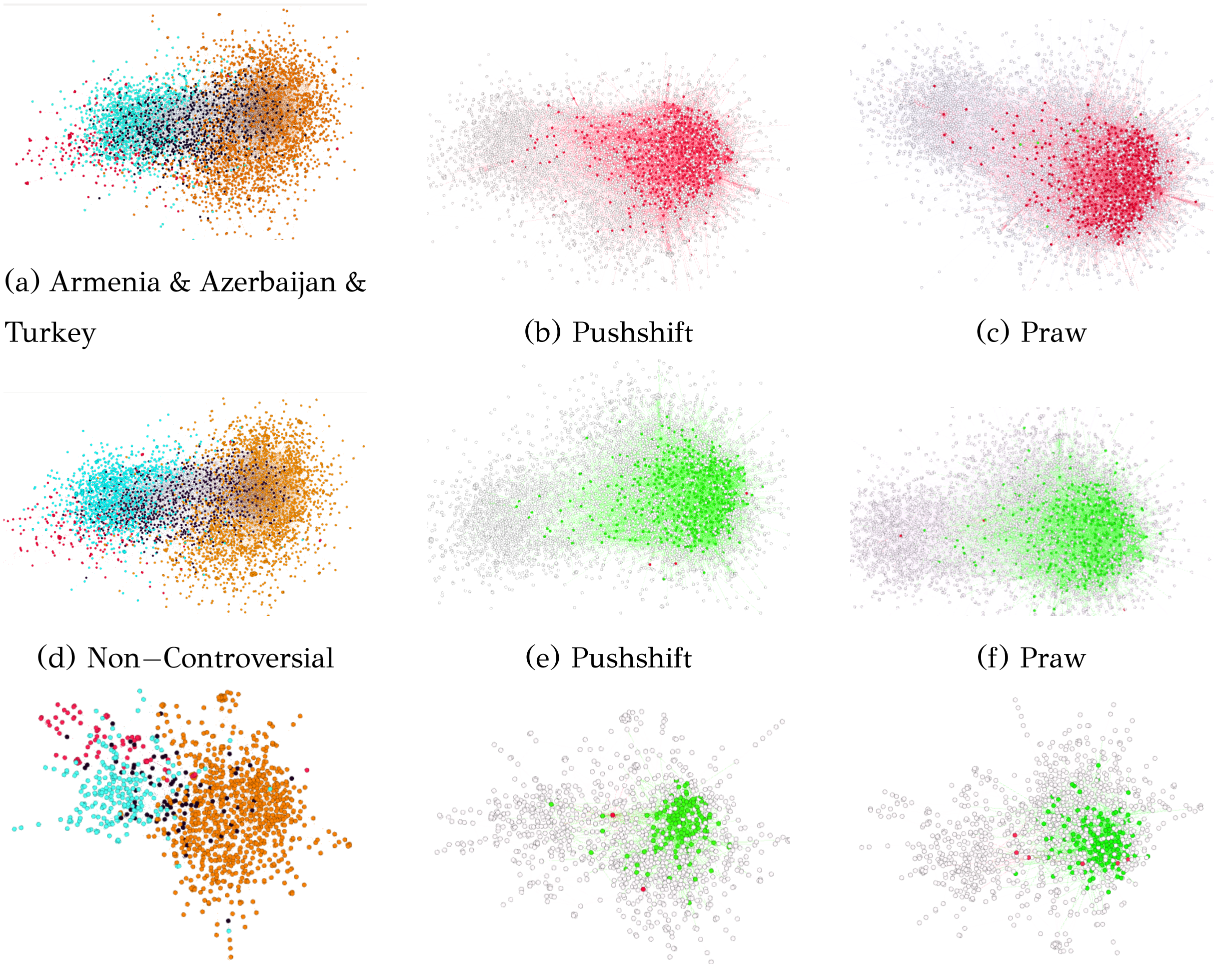


(a) Azerbaijan & Turkey (b) Pushshift (c) Praw

(d) Non−Controversial (e) Pushshift (f) Praw

(g) Controversial (h) Pushshift (i) Praw

Figure D.3: Azerbaijan & Turkey graph representation of two polarized groups.



(a) Armenia & Azerbaijan & Turkey

(b) Pushshift

(c) Praw

(d) Non−Controversial

(e) Pushshift

(f) Praw

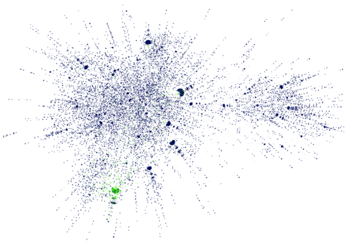
(g) Controversial

(h) Pushshift

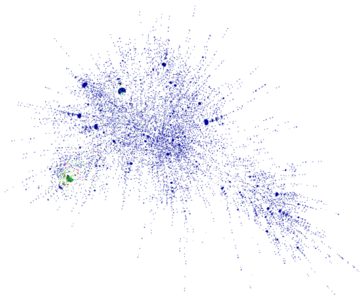
(i) Praw

Figure D.4: Armenia & Azerbaijan & Turkey graph representation of two polarized groups.

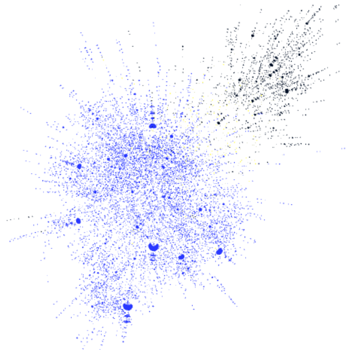
(a) Bad_Cop_No_Donut & BlackLivesMatter



(b) Pushshift

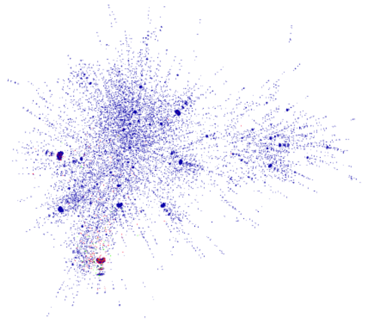


(c) Praw

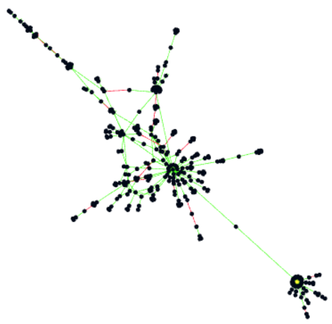


(d) Non−Controversial

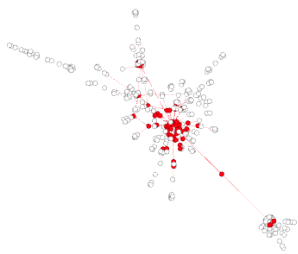
(e) Pushshift



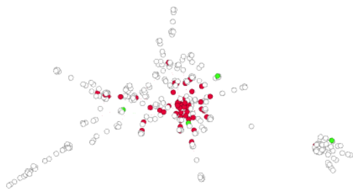
(f) Praw



(g) Controversial

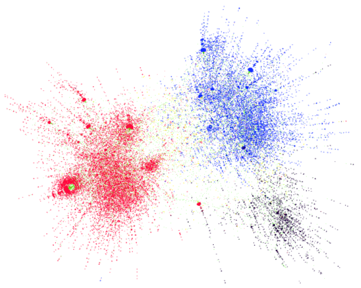


(h) Pushshift

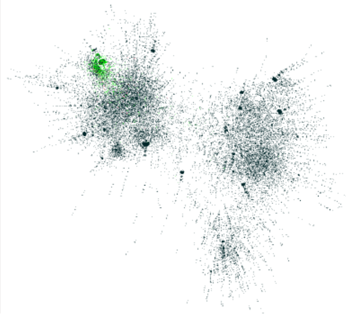


(i) Praw

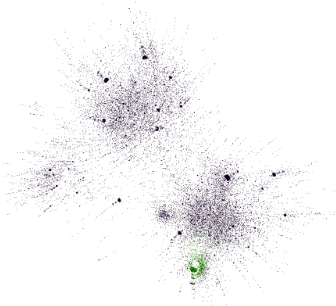
Figure D.5: Bad_Cop_No_Donut & BlackLivesMatter graph representation of two polarized groups.



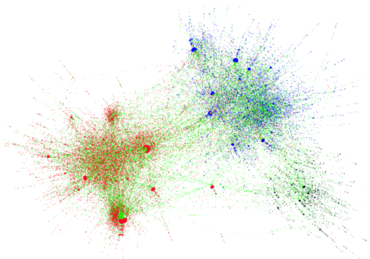
(a) Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter



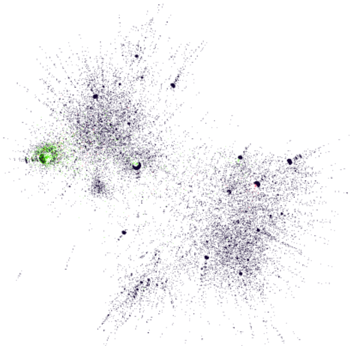
(b) Pushshift



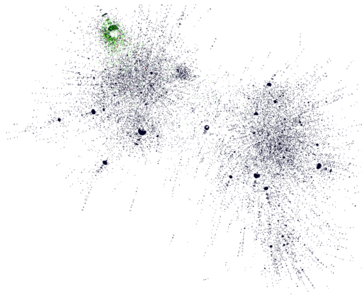
(c) Praw



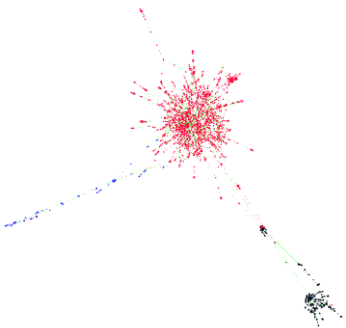
(d) Non−Controversial



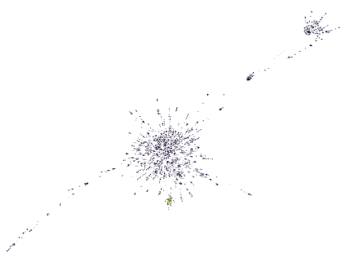
(e) Pushshift



(f) Praw



(g) Controversial



(h) Pushshift

(i) Praw

Figure D.6: Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter graph representation of two polarized groups.

Appendix E

Temporal Motifs

Table E.1: Proportion of 4−top most frequent temporal motifs per subreddit independently.

Subreddits	Temporal motifs (M)			
Turkey	M5, 7.44%	M15, 6.95%	M30, 6.72%	M12, 6.49%
Greece	M29, 7.70%	M30, 7%	M15, 6.77%	M27, 6.48%
Islam	M5, 7.02%	M30, 6.97%	M15, 6.42%	M29, 5.98%
Armenia	M36, 22.50%	M6, 16.81%	M1, 13.36%	M34, 3.81%
Azerbaijan	M36, 12.08%	M6, 10.14%	M1, 8.35%	M34, 5.48%
Turkey	M30, 9.50%	M28, 7.48%	M29, 7.06%	M15, 6.94%
China_Flu	M6, 36.50%	M36, 29.53%	M1, 21.58%	M21, 4.32%
Coronavirus	M1, 25.75%	M34, 23.04%	M16, 21.86%	M35, 14.07%
Unpopularopinion	M6, 11.38%	M36, 10.68%	M5, 5.31%	M30, 5.19%
Bad_Cop_No_Donut	M15, 7.98%	M22, 7.42%	M21, 6.92%	M5, 6.69%
BlackLivesMatter	M31, 27.19%	M21, 10.16%	M34, 6.81%	M33, 6.06%

Table E.2: Proportion of 4−top most frequent temporal motifs per 2−subreddits graphs.

Subreddits	Temporal motifs (M)			
Turkey & Greece	M5, 7.02%	M15, 6.90%	M30, 6.80%	M29, 6.33%
Turkey & Islam	M5, 7.11%	M30, 6.93%	M15, 6.49%	M12, 6.03%
Greece & Islam	M30, 6.98%	M5, 6.93%	M15, 6.43%	M29, 6.07%
Armenia & Azerbaijan	M36, 22.17%	M6, 16.60%	M1, 13.17%	M34, 3.87%
Armenia & Turkey	M36, 22.50%	M6, 16.81%	M1, 13.36%	M34, 3.81%
Azerbaijan & Turkey	M36, 12.06%	M6, 10.13%	M1, 8.34%	M34, 5.48%
China_Flu & Coronavirus	M1, 27.01%	M34, 19.84%	M6, 9.64%	M36, 6.66%
Unpopularopinion & Bad_Cop_No_Donut	M6, 10.34%	M36, 9.42%	M15, 5.77%	M5, 5.70%
Unpopularopinion & BlackLivesMatter	M6, 11.23%	M36, 10.56%	M5, 5.27%	M30, 5.15%
Bad_Cop_No_Donut & BlackLivesMatter	M15, 7.67%	M22, 7.18%	M21, 7.15%	M5, 6.42%

Table E.3: Proportion of 4−top most frequent temporal motifs per 3−subreddits graphs for total, non−controversial (NC) and controversial (C) posts.

Subreddits	**Temporal motifs (M)**			
Turkey & Greece & Islam	M5, 7.03%	M30, 6.95%	M15, 6.50%	M29, 6.04%
Turkey & Greece & Islam (NC)	M5, 6.99%	M30, 6.95%	M15, 6.46%	M29, 5.96%
Turkey & Greece & Islam (C)	M5, 7.31%	M15, 7.24%	M30, 7.08%	M29, 6.62%
Armenia & Azerbaijan & Turkey	M36, 22.17%	M6, 16.60%	M1, 13.17%	M34, 3.87%
Armenia & Azerbaijan & Turkey (NC)	M36, 22.55%	M6, 16.86%	M1, 13.35%	M34, 3.84%
Armenia & Azerbaijan & Turkey (C)	M15, 6.69%	M22, 6.34%	M5, 6.30%	M12, 5.80%
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter	M6, 10.23%	M36, 9.34%	M15, 5.75%	M5, 5.67%
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter (NC)	M6, 11.95%	M36, 11.17%	M15, 5.65%	M5, 5.55%
Unpopularopinion & Bad_Cop_No_Donut & BlackLivesMatter (C)	M5, 5.30%	M25, 5.02%	M12, 4.80%	M32, 4.73%