Polarized Groups in Discussion Forums: the Case of Reddit

Anonymous

Abstract

In this paper, we investigate the formation of polarized groups in Reddit. We consider the formation of two types of polarized groups, namely, groups that barely communicate with each other, and groups that communicate to express disagreement. We call the former *unsigned polarization* and the latter *signed polarization*. We study the formation of polarized groups both within communities, i.e., subreddits, termed *intra-polarization*, and across communities, termed *inter-polarization*, that discuss the same topic. We also investigate whether controversial posts are more prone to polarization than non-controversial ones. We adopt an approach that exploits the structure of the user interactions. We focus our study on four discussion topics spurred by controversial events. Our results show that in most cases: (i) there is evidence of unsigned inter-polarization but no evidence of unsigned intra-polarization, and (ii) there is no significant signed inter-polarization even between communities with different stances on the event, instead there is some degree of signed intra-polarization. Controversy seems to increase unsigned polarization but does not affect signed polarization.

Introduction

Nowadays, millions of individuals use online social media platforms on a daily basis to interact with family, friends, co-workers and even strangers to express their views on a variety of important topics, from politics and the economy to sports and religion. Different points of view, extreme opinions, diverse backgrounds and personalities may lead to negative social phenomena such as aggressive behavior, bullying, conflicts and polarization.

In this paper, we focus on *polarization*, where people are divided into two opposing groups. In particular, we study the formation of polarized groups in Reddit, an online platform where people meet to discuss. Reddit is a large community made up of thousands of smaller subcommunities known as "subreddits" where users post submissions and comment on submissions. We consider the formation of two types of polarized groups, namely, groups that barely communicate with each other (Garimella et al. 2018), and groups that communicate but only to express disagreement (Bonchi et al. 2019). We call the former *unsigned polarization* and the latter *signed polarization*. A novelty of our approach, is that we study two types of polarization, namely *intra-polarization* and *inter-polarization*, In intra-polarization, opposing groups are formed within a single subreddit (community), while in inter-polarization, the members of the opposing groups belong to different communities.

We examine polarity around specific discussion topics by collecting all related submissions and their comments. We follow an approach that exploits the network of interactions between the users participating in the discussion. For unsigned polarization, we build on previous work that quantifies the levels of polarity in a network through structural properties such as, connectivity and cut edges (Garimella et al. 2018) and boundary nodes (Guerra et al. 2013). For signed polarization, we annotate pair-wise user interactions as positive (agreement) or negative (disagreement) using the Reddit score of a comment. This score combines that downvotes and upvotes that a comment has received from the users. To quantify signed polarization, we use an approach based on properties of the adjacency matrix of the network (Bonchi et al. 2019).

We have collected submissions and related comments posted in a number of Reddit communities about four discussion topics that were spurred by controversial events. Our results show evidence of unsigned inter-polarization but no evidence of unsigned intra-polarization especially among random users. This indicates little communication across communities but no fragmentation inside communities. For signed polarization, there is no significant inter-polarization, even between communities with different stances on the topic. Instead, there is some degree of signed intra-polarization.

Often, polarity is related to *controversy*. Controversy refers to disagreement and the emergence of more that one aggressive side about a topic, usually because the topic affects or is important to many people (Popescu and Pennacchiotti 2010), (Mendoza, Parra, and Soto 2020). We use the Reddit characterization of submissions as controversial, or, non-controversial. Controversy seems to increase unsigned polarization but does not affect signed polarization for both

intra- and inter-polarization. In a sense, this indicates that, in our experiments, controversy decreases the level of communication between opposing groups but does not necessarily increase the level of disagreement between them.

Problem Definition

Our main goal is to explore whether there is polarization in Reddit discussions. In particular, we ask whether polarized groups of users are formed for a specific discussion topic. We consider two types of polarity:

- *Unsigned Polarity:* There is unsigned polarization if the individuals discussing a common topic are divided into two disjoint subgroups with very low interaction between them.
- *Signed Polarity.* There is signed polarization, if the individuals discussing a common topic are divided into two disjoint groups such that they agree with the individuals belonging to their own group and disagree with the individuals belonging to the other group.

We study two different types of polarization, intra- and inter-polarization.

- *Intra-polarization* looks for the formation of polarized groups among the users that discuss a given topic within a single community, i.e., subreddit.
- *Inter-polarization* looks for the formation of polarized groups between two communities A and B that discuss a given topic, such that (the majority of) the members of one of the two groups belong to community A, while (the majority of) the members of the other group belong to B.

The first research question that we explore in this work is:

- **RQ1:** Is there intra- or inter- signed or unsigned polarization in Reddit?

Often polarity is related to controversy. A discussion is controversial if there is disagreement between users and there is more than one aggressive view on the issue being discussed (Popescu and Pennacchiotti 2010), (Datta and Adar 2019). We also investigate the following research question:

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

Methodology

We study polarity generated around a specific discussion topic T. Our approach works in three steps. In the first step, we collect all posts (submissions) on the specific topic T posted in the communities (subreddits) for which we want to compute intra- or inter-polarization. In the second step, we build appropriate *user conversation graphs* to capture the communication between the users that participated in the discussions on topic T. Finally, in the third step, we analyze the graphs by applying various polarity metrics. Figure 1 depicts our 3-step methodology. We detail next the graph construction and the polarity metrics.

Generating the Graphs

The construction of our graphs follows a three-stage process.

Conversation Tree. Our building block is a per-post (submission) *conversation tree (CT)*. The *conversation tree* CT_p for a Reddit submission p is a directed, node-signed tree $CT_p(V_p, E_p, L_p)$ where there is a node c in V_p for each comment in submission p and there is a directed edge $(c_i, c_j, t) \in E_p$ from c_i to c_j, if c_j is a reply to c_i and t the timestamp of this reply. Function $L_p : V_p \rightarrow \{+, -\}$ assigns a label (sign) to each node (comment) in V.

To determine the sign of a comment c, we exploit the score, $score(c)$ assigned to c by Reddit, where $score(c)$ is the number of upvotes minus the number of downvotes that c has received. Specifically, L_p assigns a '+' label to a comment c, if $score(c)$ is positive, and a '-' label, if $score(c)$ is negative.

User Conversation Graph. We use the conversation trees to create *user conversation graphs* (UCG) graphs for a set of submissions P to capture the interactions between the users that participated in these submissions. The *user conversation graph* $UCG_P = (V, E, L)$ for a set $P = \{p_1, \ldots p_m\}$ of submissions p_l, $1 \leq l \leq m$, is a directed edge-signed multigraph where there is a node v in V for each user that participated in any of the submissions in P and there is a directed edge (v_i, v_j, t, p_l) in E from v_i to v_j for each edge (c, c', t) in E_{p_l} where c, c' are comments of u_i and u_j respectively. Function $L : E \rightarrow \{+, -\}$ assigns a label (sign) to each edge.

The sign of an edge between users v_i and v_j expresses the agreement ('+'), or disagreement ('-') between u_i and u_j when the two users interacted in submission p by exchanging the corresponding comments c and c'. Thus, we use the score of c and c' in CT_p to deduce the sign of this interaction. Specifically, we assume that there is agreement and L assigns a positive score '+' to the edge, if both $L_p(c')$ and $L_p(c)$ are either positive or negative. We assume that there is disagreement and L assigns a negative score '-' to the edge, if one of the $L_p(c')$ and $L_p(c)$ is positive and the other one is negative.

The underlying assumption is that the score and thus the sign that a comment receives is an indication of the stance of the majority of the users reading the submission. When both signs are positive (negative), this is an indication that both comments agree (resp. disagree) with the opinion of the majority and thus agree with each other. Similarly, difference in signs indicates disagreement. We tested our assumptions by manually inspecting several submissions. Our inspection confirmed our intuition with the exception of the case of both signs being negative. In this case, we found some examples where the two users did not agree with each other. The main reason is that the negative initial comment caused an extreme reaction of the replying user. The comment of this user although in agreement with the majority was down voted for being inflammatory. We present representative examples in Table 1. The topic of discussion is the convention of Hagia Sophia museum in Istanbul into a mosque. There very few cases of both signs being negative. In our experi-

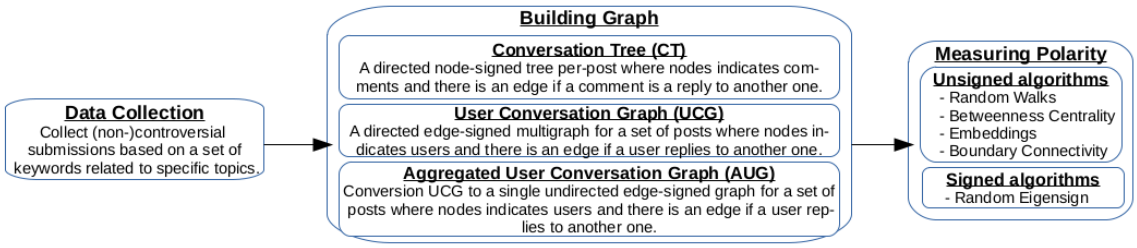


Figure 1: Our 3-step methodology.

ments, we tried different assignments for them and they did not affect our results.

Aggregated User Conversation Graph. From the multigraph UCG_P, we build a simplified *aggregated user conversation graph* AUG_P by replacing all edges between two users u_i and u_j with a single undirected edge. The sign of this edge is positive only if all edges between the two nodes in UCG_P are positive and negative if at least one of them is negative. Thus, we assume that two users disagree with each other, if they have disagreed in at least one conversation with each other. Note also that by replacing the multiple edges between two users with a single one, we ignore the volume of interaction between the two users. The reason is that we do not want our results to be dominated by over-active pairs of users, but, instead we treat all pair-wise interactions equally.

Finally, to measure the different types of polarization, we select appropriate sets P for building various AUG_P graphs. For measuring intra-polarization for a topic T inside a subreddit B, P includes all submissions about T in B. To measure inter-polarization for a topic T in two subreddit B_1 and B_2, P includes all submissions about T in both B_1 and B_2. Similarly, to test polarization of controversial submissions about T, P includes only the controversial submissions.

Measuring Usingned Polarity

To quatify unsigned *intra-* and *inter- polarization*, we use algorithms based on *random walks*, *edge betweenness*, *node distance* (Garimella et al. 2018) and *boundary connectivity* (Guerra et al. 2013).

All algorithms take as input two disjoint sets (groups) of nodes X and Y of a graph and measure the polarity between them. For determining the two groups X and Y, in the case of *intra-polarization*, we use Metis (Karypis and Kumar 1998). Metis was shown in (Garimella et al. 2018) to work well as a graph partitioning technique for detecting polarization. For *inter-polarization* between two subreddits (communities) B_1 and B_2, we also compute the polarity between the actual subreddits. In particular, group X contains the users who participated in submissions in B_1, while group Y the users who participated in B_2.

Random Walks. The *RW* polarization measure uses the notion of random walks:

$$RW = P_{XX}P_{YY} - P_{XY}P_{YX} \tag{1}$$

where $P_{AB}, A, B \in \{X, Y\}$, is the probability of a random walk that starts from partition A and ends in partition B. The more polarized (well-separated) the two partitions X and Y are, the lower the probability of not crossing the two polarized partitions, and therefore, the higher the polarization.

Betweenness. Let $C \subseteq E$ be the set of cut edges between X and Y. The betweenness centrality $bc(e)$ of an edge $e \in E$, is:

$$bc(e) = \sum_{s \neq t \in V} \frac{\sigma_{s,t}(e)}{\sigma_{s,t}},\ and\ bc(e) \in [0, 1] \tag{2}$$

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. The $Betweenness$ polarization measure is:

$$Betweenness = 1 - e^{-KL} \tag{3}$$

where KL is the Kullback-Leibler divergence of the betweenness centrality distributions between the cut and the rest of the edges. The higher the betweenness centrality of the cut edges, the more polarized the network, because paths across the two partitions use only edges from the cut set.

Embeddings. The $Embeddings$ polarization measure is:

$$Embeddings = 1 - \frac{d_X + d_Y}{2d_{XY}} \tag{4}$$

where d_X and d_Y is the average embedded distance between pairs of nodes in X and Y respectively and d_{XY} is the average embedded distance between pairs of nodes across X and Y. The network is polarized if the average embedded distance between pairs of nodes across the two partitions X and Y is large, while within it is small. Distance is calculated using two-dimensional embedding by the Gephi's ForceAtlas2 algorithm (Jacomy et al. 2014).

Boundary Nodes. A node in X or Y is a boundary node if it is connected with at least one node from the opposite group, and at least one of its neighbor is not connected with any member of the opposite group. The $Boundary\ Connectivity\ (BC)$ polarization measure is:

$$BC = \frac{1}{|B|} \sum_{u \in B} [\frac{d_i(u)}{d_b(u) + d_i(u)} - 0.5] \tag{5}$$

where $d_i(u)$ is the number of edges between node u and internal nodes I, while $d_b(u)$ is the number of edges between node u and boundary nodes B. When the number of edges between high-degree boundary and non-boundary

Discussion Dialogue	Edge Sign
A: They didn't say anything for now but in Islam you can't have human figures in a mosque so they need to cover them. The question is how they will cover those priceless mosaics. Shame on Erdogan. (Score: 23) **B**: Yes, I'm Muslim but I felt as if it could've stayed a Museum. (Score: 26)	"+"
A: Yes it is confirmed it will also still be opened for tourists to enter. (Score: 37) **B**: That's good at least. (Score: 11)	"+"
A: It is Islamist act which is a shame for Turkish people. (Score: 21) **B**: No it is not and much like the Mosque of Cordoba is still being kept as Cathedral this one will be preserved as a mosque from now on. (Score: -1)	"-"
A: You do know it was used as a mosque before? (Score: 24) **B**: It was both church and mosque for along time it should be a museum for both Christians and Muslims. (Score: -3)	"-"
A: Best news of today! (Score: 16) **B**: Worst. It's sad that Erdogan did this. (Score: -17)	"-"
A: I hate it,personally, and fear they will remove the ancient Christian art work and icons. It's clear to me, it's meant to antagonise Christians. (Score: -6) **B**: Removing such art work would be a tragedy. (Score: -2)	"+"
A: I'm sorry,these are all straw man arguments and red herrings. Your opinion doesn't matter. (Score: -3) **B**: I checked your profile and not really surprised from what i have seen, you're just another brain-dead Muslim here. (Score: -2)	"-"

Table 1: Annotation of edge sign either "+" or "-" from short conversations between two users, A and B. The score of each comment is mention at the end of each one.

Algorithm 1 Methodology for measuring unsigned polarity

Require: Unsigned $AUG(V, E)$
Ensure: Polarity score (PS)
1: Initialize X and Y polarized groups by applying clustering techniques
2: Measure **PS** $\leftarrow$ Apply *Random Walks*, *Betweenness*, *Embeddings* and *Boundary* algorithms

Algorithm 2 Methodology for measuring signed polarity

Require: Signed $AUG(V, E)$
Ensure: Polarity score (PS), $S1$ and $S2$ polarized groups
1: Detect $S1$ and $S2$ and measure **PS** $\leftarrow$ Apply *Random Eigensign* signed algorithm
2: Analyze polarized subgroups $S1$ and $S2$ i.e., size of polarized groups in relation to AUG size, percentage of positive and negative edges within and across polarized groups etc.

nodes is high then polarity is high, because boundary nodes are strongly connected with users from their group.

The general methodology of measuring unsigned *intra-* and *inter- polarization* is summarized in Algorithm 1.

Measuring Signed Polarization

We use the approach in (Bonchi et al. 2019), where the problem is: given a signed network $G = (V, E)$ where edges $E^+ \subset E$ have positive sign '+' and edges $E^- \subset E$ negative sign '-', find the two subgroups S1 and S2, with the largest possible number of negative edges between them and the largest possible number of positive edges within each of them. We apply the *random eigensign* algorithm that computes the first eigenvector v corresponding to the largest eigenvalue λ_1 of the adjacency matrix.

The general methodology of measuring signed *intra-* and *inter- polarization* is summarized in Algorithm 2.

Data Collection and Graph Generation

In this section, we describe the data collection and the aggregated user graphs (AUGs) that we build.

Crawling data from Reddit

We focus our study of polarity on four discussion topics. Following previous research, we study conflicting topics caused by specific historical events. The first topic is the conversion of Hagia−Sophia in Istanbul from a museum to a mosque on July 24^{th} 2020. The second topic is the Nagorno−Karabakh conflict between Azerbaijan and Armenia on September 27^{th} 2020. The third topic is the Black Lives Matter movement that protests against incidents of police brutality and racially-motivated violence against people of color related to the May 25^{th} 2020 killing of an American hip-hop artist, George Floyd. Finally, our fourth topic is the COVID-19 pandemic on September 22^{nd} 2019, an ongoing pandemic of coronavirus disease 2019 (COVID-19) caused by severe acute respiratory syndrome SARS-CoV-2.

We collect submissions based on a set of keywords specific to each topic using the Reddit API[1]. The data collection for each of the four topics lasted for a period two months before the relevant event and up until the end of 2020. For selecting the communities (subreddits) for each topic: (1) we collected 100 submissions per month from 2007 to 2020 that contain the relevant to each community keywords, (2) we extracted the top-10 subreddits which use these keywords most frequently, and (3) we selected from these top-10 subreddits, 3 subreddits that we know (or, at least it is commonly accepted) that there is a conflict between them. In Table 2, we summarize the topics, the keywords and the selected subreddits.

For each topic of discussion and for each subreddit, we collected all posts and their comments. For each comment,

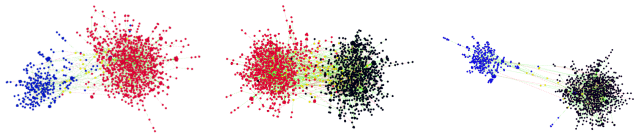
[1]https://www.reddit.com/dev/api/

Topic	Keywords	Subreddits
Hagia Sophia	Hagia Sophia, Ayasofya, $\alpha\gamma\iota\alpha\ \sigma o\varphi\iota\alpha$	Turkey, Greece, Islam
Nagorno-Karabakh	Nagorno, Karabakh, Nagorno-Karabakh	Armenia, Azerbaijan, Turkey
Police violence	George Floyd, Derek Chauvin, police violence, Black lives matter	Unpopularopinion, Bad_Cop_No_Donut BlackLivesMatter
COVID-19	Coronavirus, covid, covid-19 vaccines, vaccination, vaccines	China_Flu, Coronavirus

Table 2: Data collection: topics of discussion, selected keywords and subreddits.

Subreddits	Abbr.	All	NC	C
Turkey (HS)	TurHS	172	95	77
Greece	Gr.	47	37	10
Islam	Isl.	90	78	12
Armenia	Arm.	845	659	186
Azerbaijan	Azer.	852	792	60
Turkey (NK)	TurHK	71	53	18
China_Flu	CF	25,794	21,328	4,466
Coronavirus	Coron.	60,993	57,493	3,500
Unpopularopinion	Unpop.	3,742	3,379	363
Bad_Cop_No_Donut	BCND	1,237	1,165	72
BlackLivesMatter	BLM	3,855	3,186	669

Table 3: Number of collected posts for all, non-controversial (NC) and controversial (C) posts.



(a) Greece & Turkey (b) Turkey & Islam (c) Greece & Islam

Figure 2: Graph visualization of 2-subreddits AUGs networks for Hagia-Sophia topic.

we collect a set of features such as, the author of the comment, the unix timestamp of the reply, the parent comment, and its score (number of upvotes minus downvotes). We annotate a post as non-controversial or controversial based on the annotation by Reddit API. Reddit labels a submission as controversial if it receives an equal amount of upvotes and downvotes. We notice that the average upvote ratio score (ratio of votes counting both upvotes and downvotes) for non-controversial and controversial posts is 89.63% and 44.36% respectively. The number of collected posts is reported in Table 3.

Generating AUG graphs

For measuring *intra-polarization*, we generate three AUGs for each subreddit, one including all posts in the corresponding community, one including only the controversial posts and one including only the non-controversial posts. That is, we consider a total of 33 AUGs. The size of the generated AUGs is shown in Table 4(a).

For measuring *inter-polarization*, we create for each topic T, AUGs for all pairs of related subreddits containing the posts in these two subreddits. For instance, for the topic Hagia−Sophia, AUGs are created for Turkey (HS) & Greece, Turkey (HS) & Islam and Greece & Islam. Again, we created separate AUGs for all, controversial and non-controversial graphs. The size of the 30 generated graphs is shown in Table 4(b). A graphical visualization of 2-subreddits AUGs for the Hagia-Sophia topic is depicted in Figure 2.

Results and Discussion

In this section, we present our results regarding unsigned and signed polarity in Reddit. We discuss both *intra-* and *inter- polarization*, as well as whether controversy increases polarity.

What are the levels of unsigned polarization?

To measure unsigned polarity, we apply methods based on *random walks*, *betweenness*, *embeddings* and *boundary connectivity*. In particular, for the random walk (RW) method, we exploit three types of walks based on the set of starting and ending nodes of the walks. Specifically, ***RW(rr)*** starts from and ends at random nodes, ***RW(rp)*** starts from random nodes and ends at popular (i.e., high degree) nodes and ***RW(pp)*** starts from and ends at popular nodes. In the case of popular nodes, we take the top-10 nodes with the highest degree, and in the case of random nodes, we randomly select 10% of the nodes. In each case, we repeat the experiment 1000 times and report average values.

As said, for determining the two groups X and Y, in the case of *intra-polarization*, we use Metis (Karypis and Kumar 1998), while for *inter-polarization* between two subreddits B_1 and B_2, we compute polarity between the actual subreddits with X containing the users who participated in submissions in B_1, while subgroup Y the users who participated in B_2. *Common* users, that is, users who participate both in posts from B_1 and B_2 are placed in one of the two subgroups X or Y randomly.

Values for *intra-polarization* are shown in Table 5 and values for *inter-polarization* (between the actual subreddits) in Table 6. We report three polarity values, namely, for all, non-controversial and controversial posts. For the *random walks*, *betweenness* and *embeddings* methods, polarity values are in the $[0.0, 1.0]$ range, with values over 0.5 indicating polarization. For the *boundary* method, polarity values are in the $[-0.5, 0.5]$ range, where positive values indicate polarization. Note also that, the polarity values produced by different methods can not be compared with each other, but can only be seen relatively. We discuss first polarity for all

Subreddits	All		Non-Controversial		Controversial	
	#V	#E	#V	#E	#V	#E
TurHS	985	1,802	781	1,292	351	517
Gr.	330	525	285	439	64	83
Isl.	965	1,801	941	1,740	30	30
Arm.	2,918	26,448	2,675	24,867	824	2,193
Azer.	1,818	7,902	1,745	7,473	287	472
TurNK	433	652	377	532	97	120
Unpop.	16,347	25,716	14,880	22,827	2,038	2,816
BCND	12,251	16,950	12,134	16,768	121	131
BLM	2,661	3,284	2,285	2,800	325	368
CF	39,697	219,026	37,469	196,299	9,083	23,963
Coron.	256,677	1,177,128	243,241	1,148,857	13,436	28,271

(a)

2-Subreddits	All		Non-Controversial		Controversial	
	#V	#E	#V	#E	#V	#E
TurHS & Gr.	1,288	2,331	1,044	1,731	414	600
TurHS & Isl.	1,883	3,603	1,669	3,033	404	579
Gr. & Isl.	1,288	2,330	1,222	2,183	64	83
Arm. & Azer.	4,133	34,272	3,886	32,266	1,054	2,664
Arm. & TurNK	3,251	27,100	2,977	25,399	910	2,314
Azer. & TurNK	2,069	8,540	1,970	7,993	369	591
Unpop. & BCND	28,259	42,695	26,705	39,619	2,159	2,950
Unpop. & BLM	18,975	29,017	17,133	25,634	2,366	3,190
BCND & BLM	14,844	20,274	14,345	19,598	325	368
CF & Coron.	279,692	1,394,768	275,763	1,346,856	20,840	52,292

(b)

Table 4: Size of generated AUGs, number of nodes ($\#V$) and edges ($\#E$) (a) per subreddit and (b) 2-subreddits for all, non-controversial and controversial posts.

Subreddit	*RW(rr)*			*RW(pp)*			*RW(rp)*			*Betweenness*			*Embeddings*			*Boundary*		
	All	NC	C	All	NC	C	All	NC	C	All	NC	C	All	NC	C	All	NC	C
Gr.	0.25	0.29	0.26	0.35	0.36	**0.86**	**0.66**	0.79	**1.0**	**0.52**	0.59	**0.65**	0.15	0.16	**0.48**	**0.16**	0.17	**0.30**
TurHS	0.20	0.24	**0.28**	0.04	0.14	**0.31**	0.42	0.51	**0.70**	**0.60**	0.67	0.62	0.07	0.03	0.03	**0.09**	0.14	**0.19**
Isl.	0.22	0.24	-	0.09	0.15	**0.18**	**0.51**	0.53	-	**0.68**	0.54	0.13	0.22	0.27	**0.39**	**0.11**	0.13	0.03
Arm.	0.10	0.09	**0.19**	0.0	0.0	**0.20**	0.12	0.12	**0.33**	**0.68**	0.72	0.66	0.10	0.06	0.04	-0.19	-0.18	**0.07**
Azer.	0.12	0.12	**0.25**	0.0	0.04	**0.28**	0.25	0.28	**0.68**	**0.53**	0.53	**0.59**	0.07	0.06	**0.13**	-0.07	-0.06	**0.15**
TurNK	0.28	0.29	0.25	0.32	0.30	**0.67**	**0.71**	0.77	**0.92**	**0.56**	0.65	0.57	0.29	0.06	**0.29**	**0.14**	0.15	**0.20**
Unpop.	**0.50**	0.49	0.45	0.06	0.13	0.11	**0.55**	0.58	0.39	**0.91**	0.90	0.83	0.21	0.08	0.09	**0.18**	0.17	**0.20**
BCND	**0.57**	0.60	**0.75**	0.20	0.20	**0.77**	**0.60**	0.61	**0.99**	**0.82**	0.83	0.71	0.36	0.21	**0.47**	**0.19**	0.18	0.10
BLM	0.43	0.48	**0.56**	0.03	0.22	**0.69**	**0.65**	0.71	**0.92**	**0.92**	0.90	0.90	0.10	0.18	**0.18**	**0.18**	0.20	**0.22**
CF	0.19	0.22	**0.23**	0.0	0.0	**0.08**	0.13	0.14	**0.20**	**0.60**	0.62	0.50	0.29	0.29	**0.30**	-0.03	-0.02	**0.09**
Coron.	0.16	0.17	**0.26**	0.0	0.0	**0.05**	0.05	0.06	**0.26**	**0.70**	0.72	0.69	0.03	0.10	**0.11**	-0.05	-0.03	**0.16**

Table 5: Unsigned *intra-polarization* scores for all, non-controversial (NC) and controversial (C) posts. Symbol '-' indicates that a polarity score cannot be calculated because there are no users in one of the two groups.

posts and focus on the effect of controversy at the end of this section.

Intra-polarization. Overall as shown in Table 5, most polarity measures of all posts (for all methods except from *betweeness*) are low indicating lack of a high-degree of unsigned polarization within subreddits. In the case of RW methods, *RR(rr)* and *RR(pp)* values are very low, confirming assortative in that within communities, nodes tend to communicate with nodes of similar degree. *RR(rp)* values are higher indicating that random users from one group do not easily reach popular users in the other group, which may mean that popular users may have their own cycle of people they communicate with. *Boundary* nodes show similar behavior as *RW(pp)*. The higher un-signed *intra-polarization* is observed for the "Police violence" topic. This is the more general topic and it seems to create some slight separation in discussions.

Betweenness always reports high levels of *intra-polarization*. Our interpretation is that the objective of Metis is to generate a balanced graph partition into two subgroups minimizing the number of cut edges thus the betweenness centrality of the cut and the rest of the edges is expected to be large resulting in high levels of polarity regardless of the community.

Inter-polarization. As shown, in Table 6, we detect high-degrees of unsigned *inter-polarization* for the "Hagia Sophia" and the "Police Violence" topics. On the contrary, there is no unsigned polarization between the two subreddirs discussing "Covid-19". For the "Nagorno-Karabakh" topic, somehow surprising, there is no substantial unsigned polarity between the two opposing subreddits, Armenia and Azerbaijan, or Azerbaijan and Turkey. There is some polarity between Armenia and Turkey. Note also that for inter-polarity, there is no substantial difference between the three RW variants. Again, *betweenness* produces high values for all pairs of subreddits.

Besides using the actual subreddits, we also applied Metis for determining the two groups. Metis splits the graph into balanced groups and the groups generated in our case did not related to the actual subreddits and showed small polarity.

In Table 7, we report additional information about the communication between the two communities. The percentage of common nodes (i.e., nodes that post in both subreddits) is rather small, especially in the case of high *inter-polarization*. Similar to (Datta and Adar 2019) and (Zhang et al. 2020) we see that users in two subreddits who discuss the same topic through different ideologies have very low author overlap. The same holds for cross edges in the case of *intra-polarization*. Note the high percentage of cross edges for the "Nagorno-Karabakh". Finally, the number of boundary nodes for the non-polarized communities is high, indicating that users from one community comment/interact with users from the other community.

Polarity and controversy. We now compare polarity values

2-Subreddits	*RW(rr)*			*RW(pp)*			*RW(rp)*			*Betweenness*			*Embeddings*			*Boundary*		
	All	**NC**	**C**	**All**	**NC**	**C**	**All**	**NC**	**C**	**All**	**NC**	**C**	**All**	**NC**	**C**	**All**	**NC**	**C**
Gr. & TurHS	**0.61**	0.60	**0.91**	0.45	0.43	**0.98**	0.45	0.57	**0.96**	**0.60**	0.63	0.50	**0.56**	0.54	**0.79**	**0.14**	0.10	**0.18**
TurHS & Isl.	**0.54**	0.58	**0.61**	0.44	0.50	**0.77**	**0.66**	0.68	0.60	**0.77**	0.66	0.46	**0.53**	0.54	0.21	**0.14**	0.16	0.06
Gr. & Isl.	**0.85**	0.87	-	**0.90**	0.92	-	**0.94**	0.95	-	**0.73**	0.73	-	**0.68**	0.69	-	**0.16**	0.16	-
Arm. & Azer.	0.18	0.20	**0.39**	0.0	0.0	**0.38**	0.02	0.02	**0.37**	**0.78**	0.78	0.60	0.35	0.38	**0.43**	-0.06	-0.08	**0.13**
Arm. & TurNK	**0.57**	0.67	0.55	0.21	0.34	**0.70**	0.42	0.44	**0.63**	**0.58**	0.55	0.45	**0.50**	0.53	**0.51**	**0.20**	0.18	**0.22**
Azer. & TurNK	0.14	0.12	**0.35**	0.00	0.00	**0.52**	0.13	0.09	**0.58**	**0.57**	0.67	0.43	0.25	0.26	**0.33**	**0.07**	0.10	**0.14**
Unpop. & BCND	**0.77**	0.78	**0.85**	**0.51**	0.52	**0.88**	**0.73**	0.75	0.47	**0.86**	0.86	0.38	**0.54**	0.55	**0.60**	**0.16**	0.15	**0.18**
Unpop. & BLM	**0.86**	0.86	**0.90**	**0.69**	0.78	**0.89**	**0.75**	0.80	0.69	**0.72**	0.66	**0.73**	**0.65**	0.65	**0.73**	**0.16**	0.17	**0.21**
BCND & BLM	**0.72**	0.69	-	0.39	0.54	-	**0.50**	0.36	-	**0.87**	0.77	-	**0.53**	0.50	-	**0.12**	0.13	0.0
CF & Coron.	0.11	0.12	**0.27**	0.0	0.0	**0.05**	0.02	0.0	**0.04**	**0.74**	0.76	0.70	0.29	0.29	0.25	**0.05**	0.06	0.06

Table 6: Unsigned *inter-polarization* scores for all, non-controversial (NC) and controversial (C) posts. Symbol '-' indicates that a polarity score cannot be calculated because there are no users in one of the two groups.

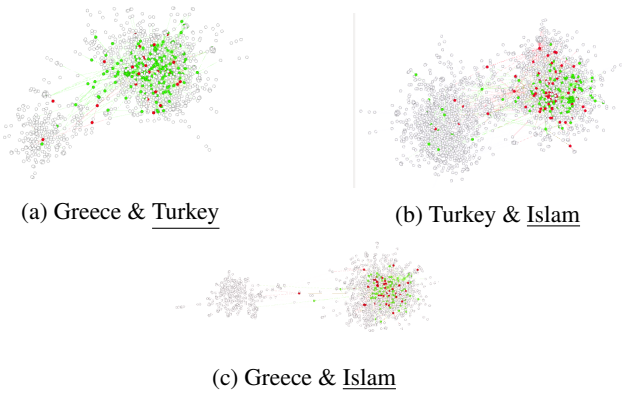
for non-controversial (NC) and controversial posts (C) for both *intra-* and *inter- polarization*. As shown in Tables 5 and 6, the polarity measures for controversial posts are in the vast majority of cases higher that those for the non-controversial posts. This is a strong indication of increased polarization between users that participate in controversial submissions when compared to users participating in non-controversial submissions, for both *intra-* and *inter- polarization*. The limited communication between users participating in controversial submissions for inter-polarity is also confirmed in Table 7, where both common nodes, cross edges and boundary nodes between the communities are much smaller.

What are the levels of signed polarization?

We now look into signed *intra-* and *inter- polarization*. In particular, we search for two groups $S1$ and $S2$ such that the individuals within each group agree with each other (most edges in $S1$ and $S2$ are positive) and disagree with the members of the other group (most cross edges are negative). Note that the polarity score (PS) is strictly related to the size of the graph and no comparison can be made between PS scores computed on different graphs. PS values are in $[0.0, \infty]$.

Intra-polarization. Signed *intra-polarization* scores are shown in Table 8. We see that two polarized groups are detected in every subreddit. The two groups are not balanced. We notice that the degree centrality of the users in the larger group is high, while in the smaller group low. This phenomenon may be interpreted as the case where there is a group of popular users who agree with each other and a smaller group consisting of unpopular users who disagree with them. Note that, in the case of the China_Flu and Coronavirus, only half of the cross edges are positive. These two subreddits had also very low unsigned polarity (Table 5).

Inter-polarization. For signed *inter-polarization*, we look first into the sign of the cross-edges between the two subreddits shown in Table 9. We see that the percentage of positive cross edges is comparable to the percentage of positive edges within each subreddit. So, in contrast to (Kumar et al. 2018) which shows that users in Reddit are mobilized by negative sentiment to comment in another community, we observe that users who interact with individuals from the opposite community seem to agree with them. However, we confirm that the number of common users across subreddits is small similar to (Kumar et al. 2018) (Table 7).



(a) Greece & Turkey (b) Turkey & Islam

(c) Greece & Islam

Figure 3: Visulization of two most polarized subgroups S1 and S2 for 2-subreddits AUG graphs and for all posts. S1 and S2 are colored red and green respectively. The underline declares the most polarized subreddit.

We report polarity scores in Table 10(a). Polarity scores, group sizes and percentage of positive cross edges are comparable to that in the case of *intra-polarization*. We also notice that the majority of users in both of the two groups belong to the same subreddit. This hold for all pairs of subreddits except from Armenia and Azerbaijan, as shown in Table 10(b). For example, for the "Hagia-Sophia" topic, individuals in Turkey are more polarized than Greeks, whereas users from Islam are more polarized than users from both Turkey and Greece. In the case of the "Nagorno-Karabakh" conflict, we notice that between Armenia and Azerbaijan, common users make up the majority of the members of the second polarized group. Furthermore, in the case of "COVID-19", we notice that the Coronavirus subreddit is more polarized than China_Flu while (Zhang et al. 2020) concludes that users in China_Flu care more about China related topics and generate more racist comments. Figure 3 shows a visualization of the two polarized groups for the "Hagia-Sophia" topic.

Polarity and controversy. We computed the polarized scores for non-controversial and controversial posts. The

2-Subreddits	Common nodes			Cross edges			Boundery nodes		
	All	NC	C	All	NC	C	All	NC	C
Gr. & TurHS	4%	2%	0.24%	13.38%	12.42%	1.5%	7%	7%	1%
TurHS & Isl.	5%	4%	1%	19.51%	14.54%	5.52%	10%	10%	3%
Gr. & Isl.	2%	1%	0%	3.86%	3.43%	0%	2%	1%	0%
Arm. & Azer.	16%	14%	6%	63.23%	60%	20%	24%	23%	18%
Azer. & TurNK	9%	8%	4%	36.53%	32%	14%	11%	9%	6%
Arm. & TurNK	9%	3%	1%	36.53%	5.08%	4.45%	25%	23%	10%
Unpop. & BCND	2%	1.25%	0.27%	7.47%	7.35%	1.59%	3%	3%	0.88%
Unpop. & BLM	0.35%	0.29%	0.29%	2.59%	2.28%	2.75%	1%	0.84%	0.46%
BCND & BLM	1%	0.86%	0.30%	10.21%	10.08 %	9.51%	4%	3%	2%
CF & Coron.	7%	6%	8%	21.51%	19.83%	23.36%	18%	16%	20%

Table 7: Percentage of common users, cross edges and boundary nodes for all, non−controversial (NC) and controversial (C) posts.

Subs.	PS	Nodes	PE	EA	PEA	PE
		S1, S2	S1, S2			
Gr.	2.44	2.72%, 20%	NO, 97%	19%	0%	66%
TurHS	3.49	15%, 3%	95%, 0%	17.32%	1.61%	69%
Isl.	2.22	5.90%, 10.77%	41%, 94%	43.02%	16.66%	54%
Arm.	41.79	12%, 1.09%	85%, 0%	3.57%	21%	77%
Azer.	16.34	15%, 1.48%	95%, 0%	7.48%	23%	81%
TurNK	2.41	15%, 0.69%	98%, NO	4.65%	0%	73%
Unpop.	2.47	8.15%, 0.97%	97%, 54%	14.19%	12%	70%
BCND	2.04	4.80%, 0.85%	99%, NO	17.73%	3.84%	70%
BLM	2.06	1.01%, 6.87%	99%, 99%	9.79%	37%	84%
CF	13.18	1.13%, 13%	93%, 85%	8.16%	56%	80%
Coron.	11.55	0.96%, 5.40%	88%, 84%	14.25%	50%	77%

Table 8: Statistics for signed *intra-polarization* and polarized subgroups S1 and S2. Signed polarity score (PS), percentage of nodes (Nodes) in S1 and S2, percentage of positive edges (PE) within S1 and S2, percentage of edges across S1 and S2 (EA), percentage of positive edges across S1 and S2 (PEA) and percentage of positive edges within subreddit (PE). Value NO indicates that there are no edges.

results are similar with those for all posts, for both the case of *intra-* and *inter- polarization*. Also again, for *inter-polarization*, the polarized groups were detected within one of the two subreddits (with the exception of Armenia-Azerbaijan). Table 9 also shows that there is no significant difference in the percentage of positive edges both within each subreddit and across subreddits in the case of non controversial and in the case of controversial posts. Thus, our results show no direct relationship between controversy and signed polarity.

Related Work

The focus of our work is on the formation of polarized groups in Reddit both inside and across communities. We also investigate the relation of controversy and polarization.

Conflicts in Reddit. There has been some previous work on inter-community conflicts among subreddits in Reddit. The authors in (Kumar et al. 2018) show that users in Reddit are mobilized by negative sentiment to comment in another community. They propose a model that combines graph embeddings, user, community and text features to create early-warning systems for community moderators to prevent conflicts. We did not notice significant negative behavior across subreddits. (Datta and Adar 2019) extract community-to-community conflicts in Reddit by aggregating users who behave differently depending on the community they interact with. They also noticed very low overlap between two subreddits where users discuss the same topic through different ideologies. A different approach to detect inter-community conflict in r/place subreddit was proposed by (Vachher et al. 2020). The r/place subreddit is modeled as a canvas where Reddit users can recolor pixels. Their findings show that multiple communities are involved in many conflicts and there is no one-on-one conflicts. In this paper, we focus on pairs of subreddits and leave the study of opposing groups involving multiple communities for future work.

A recent large scale characterization of Reddit users response to the COVID-19 pandemic in (Zhang et al. 2020) looks at the China_Flu and Coronavirus subreddits. By comparing user activity, overlapping posts and language usage, the authors concluded 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. We notice no significant difference in terms of the formation of polarized groups between these two communities.

A case study of Brexit discussions in (Largeron, Mardale, and Rizoiu 2021). investigates 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. Their results show notably that the opinions of the users involved in the same diffusions as the users is a better prediction of their opinion than the content they exchange.

Controversy and polarization. The flow of information between different groups may lead to conflicts (Zachary 1977). The study by (Adamic and Glance 2005) 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. We showed similar behavior in Reddit. Similar conclusions have been drawn in (Garimella et al. 2018) where the authors claim that controversy is related to the level of interaction between influencers with opposite views. A case

A & B Subreddits	PE						PEA		
	A			B					
	All	NC	C	All	NC	C	All	NC	C
Gr. & TurHS	66%	66%	71%	70%	73%	62%	76%	78%	88%
TurHS & Isl.	70%	73%	62%	56%	56%	62%	60%	57%	65%
Gr. & Isl.	66%	66%	71%	56%	56%	-	63%	64%	-
Arm. & Azer.	77%	78%	74%	77%	77%	68%	74%	74%	66%
Arm. & TurNK	77%	78%	75%	74%	74%	74%	64%	64%	62%
Azer. & TurNK	81%	82%	68%	78%	79%	74%	82%	83%	77%
Unpop. & BCND	70%	72%	57%	70%	70%	52%	88%	67%	70%
BCND. & BLM	70%	70%	NO	85%	86%	73%	76%	75%	91%
Unpop. & BLM	70%	72%	57%	84%	86%	73%	83%	87%	74%
CF & Coron.	79%	80%	74%	77%	77%	72%	78%	78%	74%

Table 9: Statistics of actual 2-subreddits (A & B). Percentage of positive edges (PE) within A and B and percentage of positive edges across A and B (PEA) for all, non-controversial (NC) and controversial (C) posts.

A & B Subreddits	PS	Nodes	PE	EA	PEA
		S1, S2	S1, S2		
Gr. & **TurHS**.	3.25	1.78%, 11.87%	66%, 95%	15%	0%
TurHS & **Isl**.	2.20	3.55%, 5.94%	70%, 95%	49%	17%
Gr. & **Isl**.	2.44	3.95%, 6.83%	83%, 98%	42%	19%
Arm. & **Azer**.	42.0	8.92%, 0.84%	85%, 93%	4%	22%
Arm. & TurNK	42.99	11.65%, 1.07%	85%, 0%	3%	19%
Azer. & TurNK	15.30	14%, 2%	94%, 33%	7%	17%
Unpop. & BCND	2.46	5%, 0.53%	97%, 33%	13%	13%
BCND & BLM	2.04	0.76%, 4%	0%, 99%	16%	1.6%
Unpop. & BLM	2.54	7%, 0.72%	97%, 70%	14%	15%
CF & **Coron**.	11.44	0.97%, 5.08%	49%, 84%	16%	54%

(a)

A & B Subreddits	S1			S2		
	A	B	CM	A	B	CM
Gr. & **TurHS**	9%	78%	13%	5%	90%	5%
TurHS & **Isl**.	6%	82%	12%	9%	84%	7%
Gr. & **Isl**.	0%	98%	2%	0%	99%	1%
Arm. & **Azer**.	71%	1%	28%	40%	0%	60%
Arm. & TurNK	99%	0%	1%	94%	0%	6%
Azer. & TurNK	80%	1%	19%	88%	6%	6%
Unpop. & BCND	98%	0.5%	1.5%	96%	0%	4%
BCND & BLM	99%	0%	1%	99%	0.5%	0.5%
Unpop. & BLM	99%	0.1%	0.9%	99%	0%	1%
CF & **Coron**.	4%	84%	12%	4%	83%	13%

(b)

Table 10: Statistics of S1 and S2 polarized subgroups of 2-subreddits (A & B) graphs for all posts. (a) Signed polarity score (PS), percentage of nodes (Nodes) in S1 and S2, percentage of positive edges (PE) within S1 and S2, percentage of edges across S1 and S2 (EA), percentage of positive edges across S1 and S2 (PEA). (b) Percentage of users who make up S1 and S2 subgroups. CM indicates common users. Bold words indicates the most polarized subreddit.

study of Venezuela (Morales et al. 2015) confirms that just a small group of influential individuals propagating their opinions through a social network suffices to produce polarization.

Signed graphs have been used to model interactions in social networks, which can be either friendly and in accordance (positive sign) or antagonistic and in dissent (negative sign). For example, (Akoglu 2014) 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 (Xiao, Ordozgoiti, and Gionis 2020) finds polarized communities that are related to specific input small sets of seed nodes which constitute the two sides of a polarized structure. (Bonchi et al. 2019) discover two polarized subsets of vertices in a signed network where there are mostly positive edges within subsets while there are mostly negative edges across. We use the second approach in this paper. In future work, we could also investigate the first approach by defining appropriate seeds.

Note that in this paper, we use the voting score from posts and comments from Reddit to express agreement and disagreement between users who discuss the same topic aiming to measure polarity in the network and detect the two most polarized sub communities. Some previous work approaches the problem of detecting controversial topics in social media either analyzing the stance (for/against) (Krejzl, Hourová, and Steinberger 2017) and sentiment (language) of the text (Mohammad 2016), (Popescu and Pennacchiotti 2010) or mining structure motifs (Coletto et al. 2017) and (Hessel and Lee 2019) for early warning detection of controversial discussions. Also, there are works which combine some of the above techniques (Mendoza, Parra, and Soto 2020). (Largeron, Mardale, and Rizoiu 2021) to detect controversy and quantify polarity in the network.

Summary

In this work, we investigate the formation of polarized groups of users in Reddit. We study both unsigned polarization, where we look for polarized groups with very lit-

tle communication between them and signed polarization, where we look for polarized groups where the members of each group disagree with the members of the other group. A novelty of our approach is that we exploit the votes of the comments to deduce agreement, or, disagreement between users. We also measure *intra-* (within a single sureddit) and *inter-* (between two subreddits) polarization. We use the discussion trees to construct appropriate aggregated user conversation graphs. We conclude that in most cases there is significant unsigned *intra-polarization*, while there are indications of *inter-polarization*. On the contrary, there is some degree of signed *intra-polarization*, while *inter-polarization* in not significant even between subreddits with opposing views. Finally, controversy seems to increase unsigned polarization but does not affect signed polarization.

References

Adamic, L. A.; and Glance, N. 2005. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, 36–43.

Akoglu, L. 2014. Quantifying political polarity based on bipartite opinion networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 8.

Bonchi, F.; Galimberti, E.; Gionis, A.; Ordozgoiti, B.; and Ruffo, G. 2019. Discovering polarized communities in signed networks. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, 961–970.

Coletto, M.; Garimella, K.; Gionis, A.; and Lucchese, C. 2017. Automatic controversy detection in social media: A content-independent motif-based approach. *Online Social Networks and Media* 3-4: 22–31. ISSN 2468-6964.

Datta, S.; and Adar, E. 2019. Extracting inter-community conflicts in reddit. In *Proceedings of the international AAAI conference on Web and Social Media*, volume 13, 146–157.

Garimella, K.; Morales, G. D. F.; Gionis, A.; and Mathioudakis, M. 2018. Quantifying controversy on social media. *ACM Transactions on Social Computing* 1(1): 1–27.

Guerra, P.; Meira Jr., W.; Cardie, C.; and Kleinberg, R. 2013. A Measure of Polarization on Social Media Networks Based on Community Boundaries. *Proceedings of the International AAAI Conference on Web and Social Media* 7(1).

Hessel, J.; and Lee, L. 2019. Something's Brewing! Early Prediction of Controversy-causing Posts from Discussion Features. *arXiv preprint arXiv:1904.07372* .

Jacomy, M.; Venturini, T.; Heymann, S.; and Bastian, M. 2014. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PloS one* 9(6): e98679.

Karypis, G.; and Kumar, V. 1998. A parallel algorithm for multilevel graph partitioning and sparse matrix ordering. *Journal of Parallel and Distributed Computing* 48(1): 71–95.

Krejzl, P.; Hourová, B.; and Steinberger, J. 2017. Stance detection in online discussions. *arXiv preprint arXiv:1701.00504* .

Kumar, S.; Hamilton, W. L.; Leskovec, J.; and Jurafsky, D. 2018. Community interaction and conflict on the web. In *Proceedings of the 2018 world wide web conference*, 933–943.

Largeron, C.; Mardale, A.; and Rizoiu, M.-A. 2021. Linking the Dynamics of User Stance to the Structure of Online Discussions. *arXiv e-prints* arXiv:2101.09852.

Mendoza, M.; Parra, D.; and Soto, Á. 2020. GENE: Graph generation conditioned on named entities for polarity and controversy detection in social media. *Information Processing & Management* 57(6): 102366.

Mohammad, S. M. 2016. 9 - Sentiment Analysis: Detecting Valence, Emotions, and Other Affectual States from Text. In Meiselman, H. L., ed., *Emotion Measurement*, 201–237. Woodhead Publishing.

Morales, A. J.; Borondo, J.; Losada, J. C.; and Benito, R. M. 2015. Measuring political polarization: Twitter shows the two sides of Venezuela. *Chaos: An Interdisciplinary Journal of Nonlinear Science* 25(3): 033114.

Popescu, A.-M.; and Pennacchiotti, M. 2010. Detecting Controversial Events from Twitter. In *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, CIKM '10, 1873–1876. New York, NY, USA: Association for Computing Machinery.

Vachher, P.; Levonian, Z.; Cheng, H.-F.; and Yarosh, S. 2020. Understanding Community-Level Conflicts Through Reddit r/place. In *Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing*, 401–405.

Xiao, H.; Ordozgoiti, B.; and Gionis, A. 2020. Searching for polarization in signed graphs: a local spectral approach. In *Proceedings of The Web Conference 2020*, 362–372.

Zachary, W. W. 1977. An information flow model for conflict and fission in small groups. *Journal of anthropological research* 33(4): 452–473.

Zhang, J. S.; Keegan, B. C.; Lv, Q.; and Tan, C. 2020. A tale of two communities: Characterizing reddit response to covid-19 through/r/china flu and/r/coronavirus. *arXiv preprint arXiv:2006.04816* .