Approaching controversial discussions on Reddit platform

Abstract

The term controversy within a debate is particularly important for the abnormal evolution of a society. This intense presence of controversy in social media discussions leads the people to study and quantify this social behavior by several. Studies have focused on well-known social media sush as Twitter, Facebook, political blogs e.t. and a dozen of metrics have been proposed to measure the conflict within them. So far, an unexplored platform is the Reddit website, in which hundreds of conversations are born every day. This paper try to settle the issue of quantifying the controversy in discussions on Reddit using the structure of a conversation and not its content. Furthermore, the act of users to broadcast news generates a graph which can measure the collision. The process of measuring the controversy relies on random walks, a metric which brings satisfactorily results for some other datasets. Finally, we compare the actual score of controversy which provided by Reddit with our own results for verification. Our conclusions are that a graph from the repost activity of users has better results than a graph that represents the who-replies-to-whom activity. Also, there are more suspended users if the topic is controversial.

1 Introduction

Various temporal and ephemeral social phenomena sush as peer influence (Newcomb and M. 1962; Lewis et al. 2012), bias (Helweg-Larsen and Shepperd 2001), polarization (Polley 1987; Lamm and Myers 1978; Myers and Lamm 1976), extremism (Schmidt, Joff, and Davar 2005; Hogg et al. 2013) and controversy (David and Krantz 1999; Boring and G. 1929) are spreading rapidly and without control nowadays. Such kinds of phenomena become noticeable in the last fifteen years through the social media. Indeed, social media is a critical area and more and more people are using them on a daily basis for social interaction, information seeking, expression of opinion, information sharing and etc. (Anita 2013).

Through this interactions of individuals, the controversy that arises between them over some issues is a primary impetus to start this current work. The dispute ranges between political, economic, ethical issues and etc. (D. 1992). The study of controversy in social media is not new, reading the literature, someone could notice that several studies have been set up for some social networks sush as Facebook (Carlson 2018; Hendriks, Duus, and Ercan 2016; Bessi et al. 2016), Twitter (Javier et al. 2015; Garimella and Weber 2017; Burgess and Fernndez 2016), Youtube (Bessi et al. 2016; Burgess and Fernndez 2016) and Tumblr (Burgess and Fernndez 2016). Reddit platform is an interesting platform for research in which many discussions are held on a daily basis. The researches which are focusing on conflict mining on Reddit website are relatively few (Nora 2019; Buntain and Golbeck 2014; Buyukozturk, Gaulden, and Dowd-Arrow 2018; Hessel and Lee 2019).

The majority of previous studies focus on controversy and polarization without basing on how much controversial or polarized is a topic during a discussion. Many projects have attempted to quantify the polarization and the controversy on social media debates (Guerra et al. 2013; Matakos, Terzi, and Tsaparas 2017). Mathematical approaches have been proposed to quantify the controversy from an online converastion (Garimella et al. 2015; Morales et al. 2015). Measuring this phenomenon is particularly useful for many reasons some of which are that better information could be provided to people with extreme views and the recommendations systems could be more targeted for stubborn people.

This current work is motivated by interest in observing controversies at societal level with a deeper focus on Reddit forum and how intense these disputes are. Similar work has focused on the content of conversations and how quickly users respond and initiate a confrontation (Hessel and Lee 2019). The approach of this work is inspired by (Garimella et al. 2015) who measures the contentiousness of a topic based on the structure of a discussion rather than its content. This paper takes a first study to estimate the disagreement looking at the structure of the exchange of views on Reddit. Some pending fundamental questions inspect such as: *What is the appropriate conversion from a discussion to a graph that indicates a relationship between its participants? Can this generated network structure properly express the conflict within it? Do users of a regular conversation in Reddit have fewer or more users whose accounts are read-only?* In particular, it studies two ways of creating a graph from a deliberation and what is the amount of controversy for each of them comparing this quantity with their actual approach.

The contributions of this work are the following:

- We propose two methods of generating a graph from a discussion based on who-reply-to-whom and who transmits content to another person.
- We divide the graph into two subgroups, where possible there is a dispute. We rely on existing clustering algorithms.
- We express as a number the existing argument using random walks, a technique applied on Twitter data and other blogs.
- We present experimentation with this framework and Reddit data and show how the disagreement performance. We compare our results with a set of true statistics and we find that a repost graph brings better results.

2 Related Work

The analysis of controversy and polarization on online social media has attracted considerable attention. A number of papers have provided very interesting cases of studies. Metrics have been proposed to measure the degree of polarization between two communities. In (Guerra et al. 2013), they suggest a metric based on the analysis of the boundary of a pair of (potentially polarized) communities, which better captures the notions of antagonism and polarization and they have shown that polarized networks tend to exhibit a low concentration of high-degree nodes in the boundary between two communities. The data they needed was from Twitter, Facebook and political blogs.

Furthermore, in (Morales et al. 2015), a methodology to measure the emergence of polarization from social interactions has been recommended and it is based on the notion of dipole moment. They have shown that our methodology is able to detect different degrees of polarization in the conversation, depending on the participants behavior, given by the structure of the network. The tested dataset is from Twitter.

Three more suggestions have been made by Garimella, Morales, Gionis and Mathioudakis (Garimella et al. 2015), which pay particular attention on random walks, betweenness centrality and low-dimensional embedding. Both in the retweet and topic-induced follow graph the controversy can be identified while on the opposite side the simple content-based representations do not work in general. Their simulation was studied in already tested data but also in new ones collected from Twitter.

A recent research using Reddit platform as the main data source (Hessel and Lee 2019), try to predict the ultimate controversiality of posts, leveraging features drawn from both the textual content and the tree structure of the early comments within 15 minutes that initiate the discussion. They demonstrated that early discussion features are predictive of eventual controversiality in several Reddit communities.

Our work is one of the few that tests one of the best techniques to quantify polarization using data from Reddit website. Our findings on this dataset show that the share function and its relevant graph can count the controversy something that has not been studied by them (Hessel and Lee 2019).

3 Methodology

In this section, we outline the methodology followed to simulate the quantification of controversy on Reddit. For better understanding of this work, the problem has been broken down into three main actions: 1. Generate graph, 2. Partition graph and 3. Measure controversy as suggested in (Garimella et al. 2015). Having in mind this pipeline, each one of these three basic steps has its own significance but they are also interconnected. The output of one step is the input of the next one except for the last step which quantifies the disagreement in a debate. A detailed description of each of the main operations follows below.

3.1 Generate a graph from a discussion on Reddit

The purpose of this initial stage is to build a graph G from a conversation on Reddit that represents the activity related to a single or multiple posts from a specific topic. In the graph, a vertex is assigned to each user who contributes in it, and the edges are inserting according to one of the following two approaches, which capture different aspects of the data source. The first proposal is a graph based on the structure of the discussion, this means who replies to whom and the second suggestion is focused on which user relayed a submission through a known subreddit.

Reply graph. In this initial approach, a reply graph (who-comments-to-whom) will be produced for a distinctive submission from a specific category. On Reddit, there is the main submission's author and the people who respond to him or to the other people involved in the same conversation. An undirecetd edge exists between two users u and v, if user v responds to user u. Consider the succeeding colloquy instance,

Author of the submission
User A: Makes a comment
User B: Replies on user A
User C: Replies on user B
User D: Makes a comment

The post is accompanied by a reply graph which has five vertices, let it be $V = \{author, A, B, C, D\}$ and four edges, where these are $E = \{(author, A), (A, B), (B, C), (author, D)\}$. Due to the fact that, there are still comments by deleted user, the use of the largest connected component of the graph is required for smoother implementation of the method.

Share graph. Deepening into our second proposition, a share graph is building looking at more than one posts from a specified matter. The two way connection between user u and user v appears if user v reposts the content of user v in a different subderrit from the original one. Taking as example the following repost structure,

Submission A by author A
Author C repost submission A
Author D repost submission A
Author E repost submission A

Submission B by author B
Author E repost submission B
Author F repost submission B

, the equivalent share graph is composed by six nodes, $V = \{A, B, C, D, E, F\}$ and the inserted edges are five and these are $E = \{(A, C), (A, D), (A, E), (B, E), (B, F)\}$. The maximum connected component of the graph is used and in this case as well.

3.2 Dividing the graph into two communities

In this second step, the conversation graphs which have been produced from the above mentioned stages (have a look at 3.1) are fed into a graph partitioning algorithm to extract precisely two partitions. Ideally, the two groups are strongly interconnected and weakly connected to the rest of the network (Buluc et al. 2013). The best-case scenario from this separation would be that the members from each group have a common position on the particular post. The set of the users in the two disjoint parties belong in different sides of the debate most likely (Conover et al. 2011).

After the execution of the graph building, the procedure of seperating the network is performed using a known multilevel recursive-bisection, multilevel k-way, and multi-constraint partitioning algorithm, the METIS (cut based) algorithm (Karypis 1997). Aware of the plethora of partitioning algorithms, the choice of this procedure was substantiated on (Garimella et al. 2015) framework. While they try many graph partitioning algorithms of other types sush as spectral clustering, label propagation and affiliation-graph-based models, they notice that the difference among these methods is not significant. However from visual inspection METIS generates the most distinct partitions.

3.3 Measuring the controversy on a graph

The third and the last simulation stage carries out the quantification of the controversy of a post or a collection of posts. The goal of this phase is the estimation of the argument score by a decimal value. The routine takes as input two specific and known from the previous scopes parameters, the first one is the structure of the conversation graph (subsection 3.1) and the second one is the information on the members composing each community (subsection 3.2).

The simulation of measuring the fight is relied on one of the three proposed methods by (Garimella et al. 2015), the random walks which effectively manages and quantifies the fight over any structural network. The random walks operation is based on the rationale that, in a controversial discussion, there are authoritative users on both sides and captures the intuition of how likely a random user on either side is to be exposed to authoritative content from the opposing side. Accordingly, the definition of the Random Walk Controversy (RWC) metric is presented below,

$$\text{RWC} = \text{P}_{XX}\text{P}_{YY} - \text{P}_{YX}\text{P}_{XY}$$

where P_{AB}, A, B can be one of the two seperated communities which comes from the second stage and the conditional probability P_{AB} equals to

$$P_{AB} = Pr[start\ in\ partition\ A \mid end\ in\ partition\ B]$$

This Random Walk Controversy score ranges from $[-1.0, +1.0]$. It is close to one when the probability of crossing sides is low, and close to zero when the probability of crossing sides is comparable to that of staying on the same side.

3.4 Verifying our results with the real controversy score

Reaching the point of quantification of the dispute, the next and perhaps most important step is to verify our results with the real ones. Reddit platform lays out the percentage of the upvotes for every submission counting the amount of the downvotes. Having this particular information, the amount $(1 - upvote\ ratio)$ forces out the downvote ratio of a publication. Therefore, this logic could be defined as the actual controversy value.

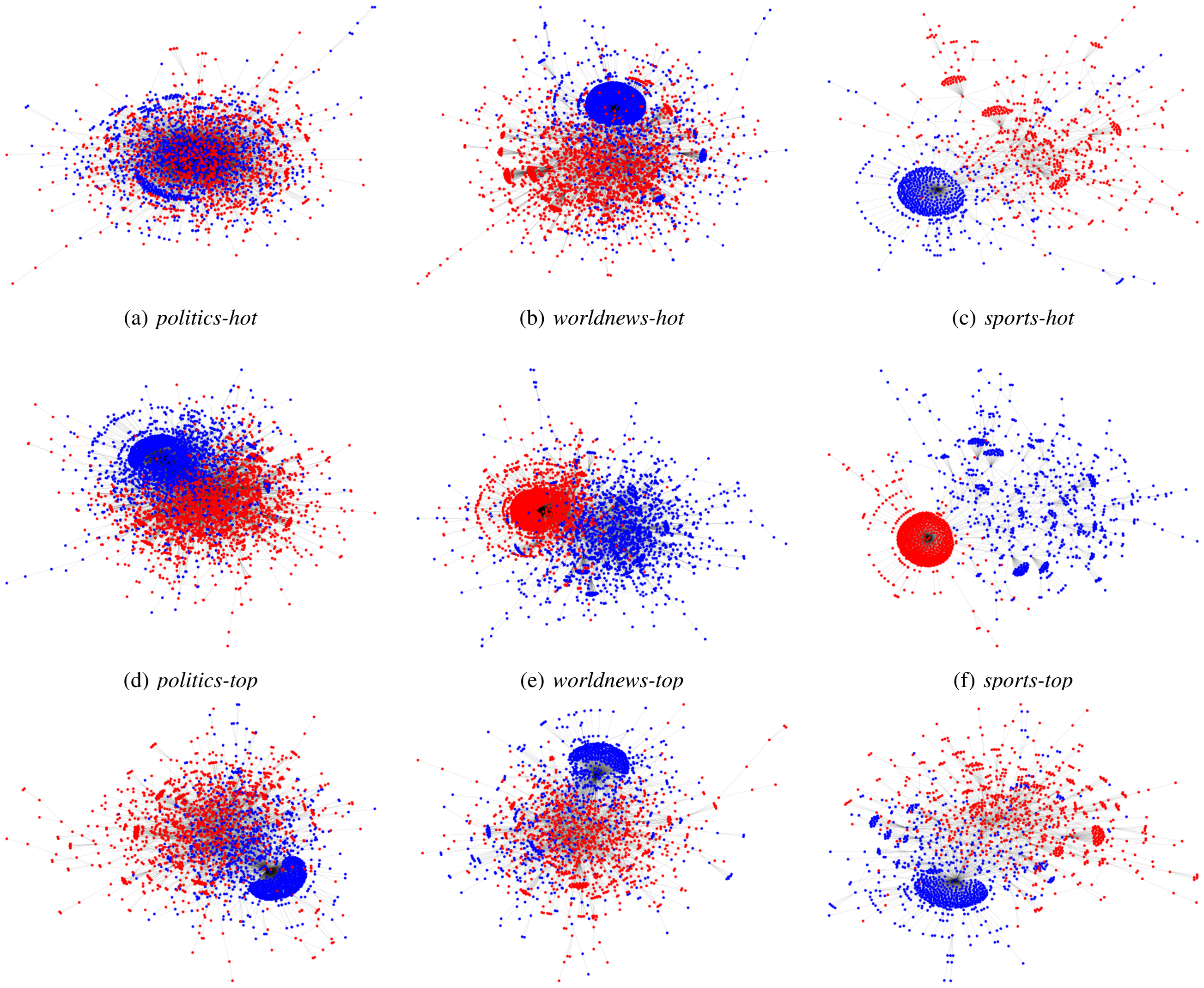
4 Experiments Setup

In this section, more details about the data collection and the values of initialization parameters of the methods in stages of clustering the network (insights into subsection 3.1) and measuring the conflict in the graph (insights into subsection 3.3) are outlined. We start with the data collection by Reddit website, how many and which subreddits have been studied and what is their classification. Keep up with the initialization of the parameters needed to implement random walks sush as how many people will be used for the simulation and since we are referring to random walks, the procedure will be repeated to have reliable results, what is the value of repetitions. Concluding, how we will use the quantity $(1 - upvote\ ratio)$ for each of the two proposed types of graphs, in order to achieve a good approach.

Data crawling. Reddit platform is our main source of data crawling which is a website comprising user-generated content and discussions of this content. Discussions on Reddit are organized into user-created areas of interest called *subreddits*. There are about $\sim 138k$ active subreddits among a total of 1.2 million, as of July 2018 (source from the site https://redditmetrics.com/). The data gathering started since 24 to 29 of January, 2020 for three hand-picked subreddits. Trying to answer the question *What are the top a hundred subreddits with the most subscribers*?, the picked out subreddits are the *politics*, *worldnews* and *sports*. All these three options are over the 56^{th} place from the top one hundred subreddits with the most subscribers.

Afterward, an additional property that Reddit supplies is that the submissions with more up-votes appear towards the top of their subreddit and therefore the following question was developed: *Based on what posts' classification, the submissions will be selected*?. Reddit categories its content with many options sush as best, hot, top, new, rising and controversial order[1]. The three cases where the research is focused are the hot (newer posts are ranked ahead of older posts,

[1] How Reddit ranking algorithms work by Amir Salihefendic in 2015 (https://medium.com/hacking-and-gonzo/how-reddit-ranking-algorithms-work-ef111e33d0d9)



(a) *politics-hot* (b) *worldnews-hot* (c) *sports-hot*

(d) *politics-top* (e) *worldnews-top* (f) *sports-top*

(g) *politics-controversial* (h) *worldnews-controversial* (i) *sports-controversial*

Figure 1: Graph partitioning using METIS algorithm plots for reply graphs using hot order in (a) *politics*, (b) *worldnews*, (c) and *sports* subreddits, top order in (d) *politics*, (e) *worldnews*, and (f) *sports* topics and finally for controversial order in (g) *politics*, (h) *worldnews* and (i) *sports* categories. Blue and red colors declare the members of both communities.

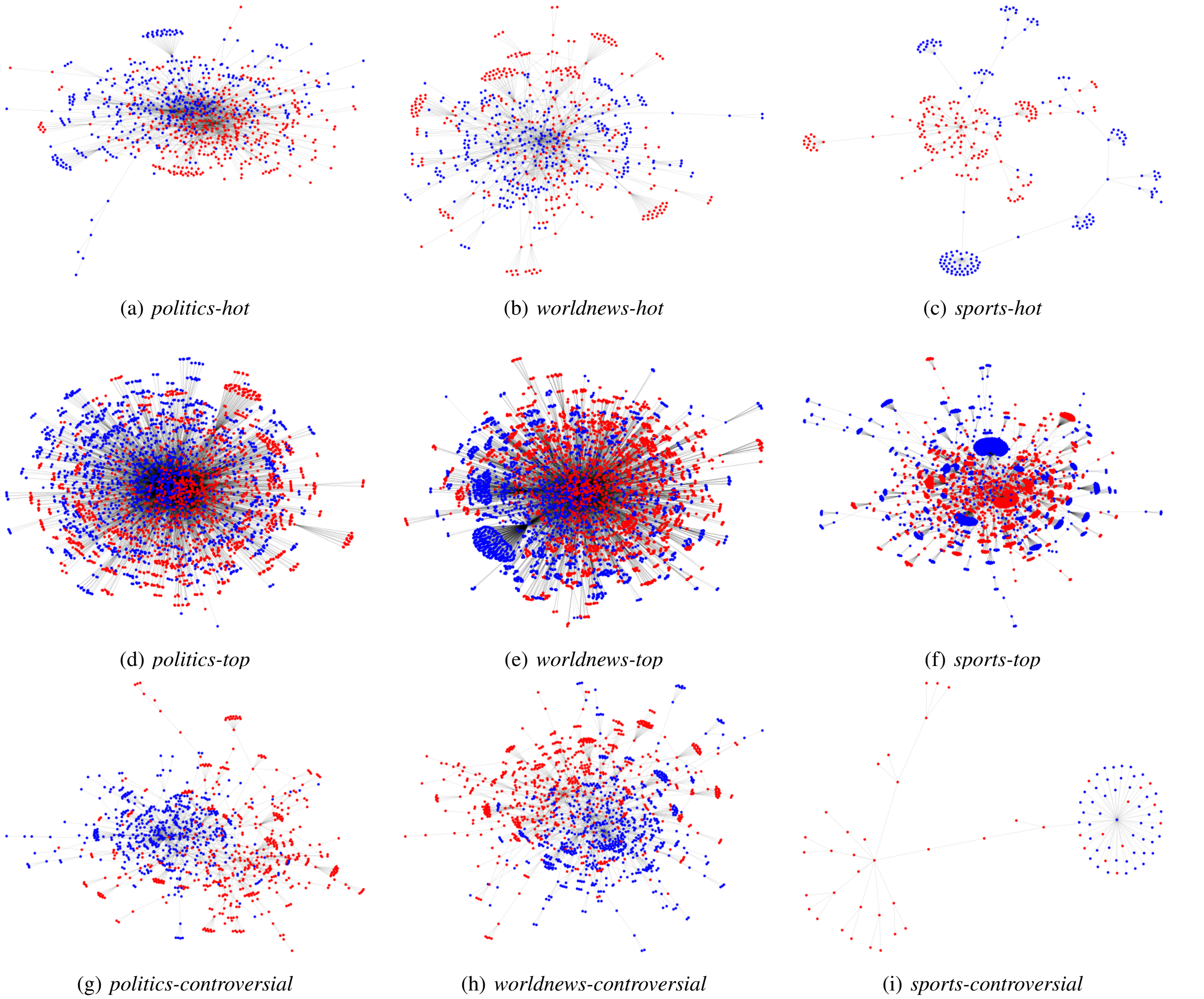
but if a post gets lots of upvotes then it gets pushed ahead in time), top (a submission gets the most upvotes over the set period) and controversial (a submission has roughly the same amount of upvotes and downvotes).

As a consequence of what has been reported so far, for the construction of the who-replies-to-whom graphs, a post is needed for each one of the selected topics of discussion (e.i. politics, worldnews and sports) and for each publications' assortment (hot, top and controversial). Therefore, three submissions have been selected for each category, and consequently nine submissions in total. On the contrary, the generation of a share graph is obliged to gather more than one publication. About $\sim 1k$ submissions are taken into account for every conversation subject. One issue that has been noticed is that for the sports category the number of edges for hot and controversial sorting is less than ~ 350, something that is not in line with the reply graphs whose number of edges exceeds the $\sim 10k$ number of edges.

METIS algorithm. During the execution of the second phase, the implementation of the segmentation algorithm of the graph does not require specific initialization values. The function $nxmetis.partition^2$ has been used and during its execution the generated undirecetd graph (have a brief look in subsection 3.1) and the number of communities are desired. The number of communities is two for the context of this research.

Random Walks Controversy score. The most straightforward way to compute the controversy score is via random walks with restart which is a much more efficient computation. Having as guideline the determinate performance by

[2] https://networkx-metis.readthedocs.io



(a) *politics-hot*

(b) *worldnews-hot*

(c) *sports-hot*

(d) *politics-top*

(e) *worldnews-top*

(f) *sports-top*

(g) *politics-controversial*

(h) *worldnews-controversial*

(i) *sports-controversial*

Figure 2: Graph partitioning using METIS algorithm plots for share graphs using hot order in (a) *politics*, (b) *worldnews*, (c) and *sports* subreddits, top order in (d) *politics*, (e) *worldnews*, and (f) *sports* topics and finally for controversial order in (g) *politics*, (h) *worldnews* and (i) *sports* categories. Blue and red colors declare the members of both communities.

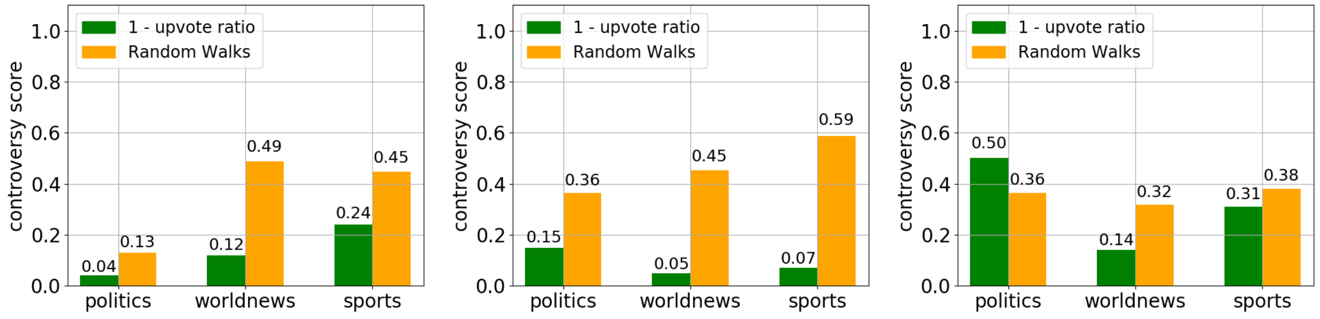
(Garimella et al. 2015)[3], the method running for more accuracy 1000 repetitions for the 0.10 percent of total members who have participated in the discussion or relayed news taken into account for the execution.

Using ground truth. As discussed in subsection 3.4, the $(1 - upvote\ \ ratio)$ is the basic amount on which the results will be verified. Meditating a reply graph (see subsection 3.1 and info about Reply graph), only one submission is taking into account from every category (insights in Section 4 and Data crawling, 1^{st} paragraph) and for every type of return sort (insights in Section 4 and Data crawling, 2^{nd} paragraph), the validation amount is $(1 - upvote\ \ ratio)$. Conversely, for a crosspost graph (see subsection 3.1 and info about Share graph) due to the fact that more than one submission is counting ($\sim\ 1k$ submissions), the final factual controversy score is the average of $(1 - upvote\ ratio)$ from all submissions which are located in the same subreddit (the selected subreddits are the politics, worldnews and sports) and for every classification's type of these (hot, top and controversial).

5 Simulation Results

So far, we provided the necessary information on how the controversy of a conversation can be organized and on what data the simulation will take place. In this section, we present the results from the extensive simulations performed, under different experimental settings used: politics, worldnews and sports subreddits, one submission from hot, top and controversial classification for reply graphs and close to $\sim 1k$ posts respectively for share graphs.

[3] https://github.com/gvrkiran/controversy-detection

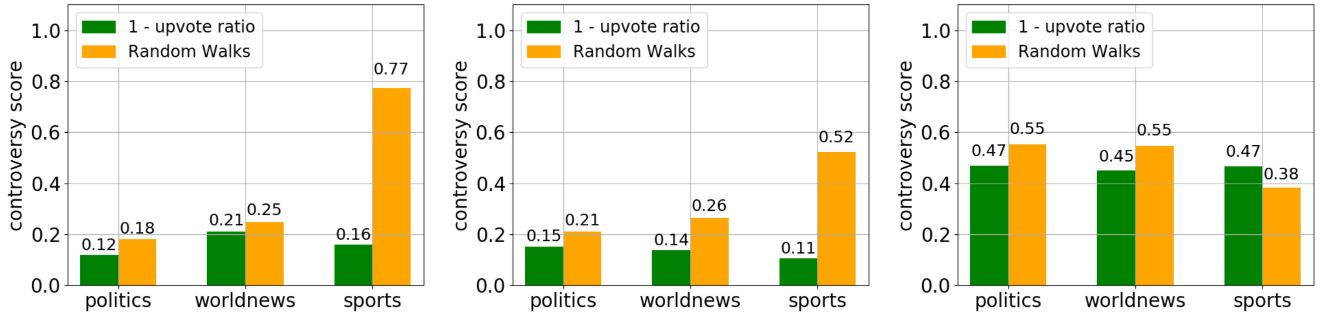


(a) *hot*

(b) *top*

(c) *controversial*

Figure 3: Plots the real controversy score (1 - upvote ratio) for reply graphs versus the RWC score using (a) *hot*, (b) *top*, (c) and *controversial* classification.



(a) *hot*

(b) *top*

(c) *controversial*

Figure 4: Plots the real average controversy score (1 - upvote ratio) for share graphs versus the RWC score using (a) *hot*, (b) *top*, (c) and *controversial* classification.

5.1 Does the graph partition representation predetermine the controversy?

One of the first dilemmas of quantifying the conflict in a disputation is if we could come to any conclusion as to whether a discussion is controversial before we even reach the third stage of simulation (see subsection 3.3).

The answer of this question is positive and Figures 1 and 2 helps us to answer this question. Observing the reply graphs partitioning for hot order (see Figures 1(a)-(b)), top order (see Figures 1(d)-(e)) and controversial order (see Figures 1(g)-(h)) the two non-common communities overlap, so there is no clear separation between them. So, for these two discussion subjects politics and worldnews, we except that the controversy score ranges in low boundaries close to 0.0 and up to 0.40 but on the opposite side and for the sports category perhaps the controversy score makes its presence more noticeable (insights in Figures 1(c), (f) and (i)) and more than 0.40.

About the second proposed type of graph generation which are the share graphs and using hot and top classification, the communities' representation in Figures 2(a), (b), (d), (e) and (f) show that the one group overlaps the other looking at politics and worldnews submissions. Nevertheless, during the controversial classification, the communities' representation in Figures 2(g)-(h) have a distinct network division into two groups without intense overlap of teams and for all three subreddits for politics, worldnews and sports issues. As far as sports broadcasting is concerned, although controversial topics are not relayed by many users, their mode of transmission shows the existence of a dispute (see Figure 2(i)).

Upshots. Having carefully studied the representation of the two communities from a conversation network, we conclude that hot, top and controversial forum threads irrespective of the topic of conversation show a strong imbricate between the two producing groups. Therefore, we can get a first sight at what issues are likely to be controversial.

5.2 Which network structure best simulate a controversial topic of discussion?

The second and perhaps most important question which the current paper try to answer is if the who-replies-to-whom graph or the share graph or both of them could have the most suitable structure which could be derived from a discussion or a set of them and could give a better sense of whether or not a theme is controversial on Reddit.

Initially analyzing the results for reply graphs as can view in Figures 3(a)-(b), it is observed that for hot and top conversations the approximate controversy value for a variety of disputation topics deviates a bit from the actual controversial aggregate. More specifically, we notice that crawling hot and top posts, the predicted controversy score based on random walks exceeds more than 0.29 points (it is about the middle value from the difference between $(1 - upvote\ ratio)$ and RWC score) regardless of the argument theme. While the actual median controversy score and this one from RWC method is 0.095 and 0.45 respectively. From the other side and continuing studying Figure 3(c), a who-reply-to-whom graph produced by controversial grouping the median difference for the selected disputation topics is 0.14. An approach much closer to the actual middle controversy score which receives the price 0.31 and the predicted middle value is 0.36.

Having studied and presented the results for reply graphs above, it remains to consider the behavior of the share graphs. Taking a first and quick look in Figure 4, we would say that the random walks controversy score approaches the real score better in this case. It is worth pointing out again that we are no longer referring to a single publication but to a collection of them for each of the modules that have been selected and have been studied (i.e. politics, worldnews and sports). Analyzing each classification method separately (see Figure 4(a)), and starting with the most upvoted recently posts (hot) notices that for all issues the middle difference of RWC score from the existent one is 0.06, which implies that through a network where the users crosspost other's content the existence of a conflict can be traced. Similar behavior is observed for the other two options of grouping (top and controversial), otherwise speaking in controversial order, the mean deviation from the real values of contention is 0.089.

Upshots. Answering this second query, we find out that the RWC score by (Garimella et al. 2015) has greater and more enduring success feeding it in the entrance with a share graph rather than a reply graph. An exception can be made while collecting controversial data, a who-reply-to-whom graph has a positive impact due to the fact that can estimate the a controversial chat.

5.3 How suspended users are separated in a share graph?

At this point investigation, we are focusing on share graphs and the three different types of order (hot, top and controversial). Looking at each one of the subreddits, we notice that in the top contents, suspended users [4] make their appearance a little more noticeable and when they come out from controversial topics their percentage grows rapidly. Looking at hot submissions (see Figure 5, orange bars), the percentage of suspended users is close to zero. For top submissions (see Figure 5, red bars), the percentage shows a slight increase which is approaching up to 3% for all the three picked up subreddits. Nonetheless, the temporarily prevent from interaction between users do not exceed almost the 5% of the total individuals on the network for the cases where the crossposted submissions are not controversial according Reddit platform. An important observation is that the number of suspended users is high if the submissions are controversial. In Figure 5 and focusing on blue bars, the percentage of suspended users is distinguished and tends to 50% for conversations about sports. Another remark is that the suspended users are uniform distributed in each one of the two communities on the network.

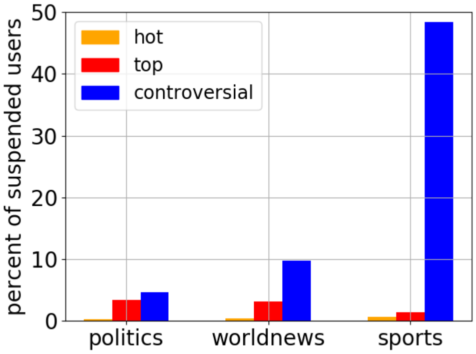


Figure 5: Plots the percentage of the suspended users in the network for the three subreddits (politics, worldnews and controversial) and for the three selected grouping types (hot, top and controversial).

Upshots. The population of the suspended users is high if the issues being relayed are controversial. This amount decreases if the discussing issues have a positive impact on the public.

6 Conclusions

In this paper, we performed an attachment for quantifying controversy by (Garimella et al. 2015) testing their best proposed metric on crawled Reddit data. We test three handpicked and crowded subreddits sush as politics, worldnews and sports and we build for each one of these discussion topics and for three types of classification which the Reddit platform provides (hot, top and controversial) two graph representations. The first produced graph representation is a who-replies-to-whom graph and the second one is a who-repost-from-whom graph. The next step is the process of partitioning the generated graph using METIS algorithm. Finally, the last procedure is to quantify the controversy on them using random walks. Our aim is to evaluate our controversy score with the actual one which the Reddit forum provides. By this comparison, we found out two interesting conclusions, the first one is that the share generated graph for both three grouping types approaches the actual controversy. And the second conclusion is that, if a topic is controversy more suspended users appear on it.

Having that knowledge for some discussions on Reddit, the controversy score can be used to generate more targeted and effective recommendation systems that foster a healthier communication on Reddit and do not restrict the users to their bubbles. It is of paramount importance to understand to what extent a discussion is polarized, so that things do not spiral out of control, or create isolated echo chambers.

[4] Suspended users effectively have their account put into read-only mode. The primary actions they will not be able to perform are: voting, submitting posts, commenting and sending private messages. https://www.reddit.com/r/announcements/comments/3sbrro/account_suspensions_a_transparent_alternative_to/

References

[Anita 2013] Anita, W. 2013. Why people use social media: a uses and gratifications approach. 16(4):362–369.

[Bessi et al. 2016] Bessi, A.; Zollo, F.; Vicario, M. D.; Puliga, M.; Scala, A.; Caldarelli, G.; Uzzi, B.; and Quattrociocchi, W. 2016. Users polarization on facebook and youtube. *PloS one* 11(8):e0159641–e0159641. 27551783[pmid].

[Boring and G. 1929] Boring, and G., E. 1929. The psychology of controversy. *Psychological Review* 36(2):97–121.

[Buluc et al. 2013] Buluc, A.; Meyerhenke, H.; Safro, I.; Sanders, P.; and Schulz, C. 2013. Recent advances in graph partitioning.

[Buntain and Golbeck 2014] Buntain, C., and Golbeck, J. 2014. Identifying social roles in reddit using network structure. In *Proceedings of the 23rd International Conference on World Wide Web*, WWW 14 Companion, 615620. New York, NY, USA: Association for Computing Machinery.

[Burgess and Fernndez 2016] Burgess, J., and Fernndez, A. M. 2016. Mapping sociocultural controversies across digital media platforms: one week of #gamergate on twitter, youtube, and tumblr. *Communication Research and Practice* 2(1):79–96.

[Buyukozturk, Gaulden, and Dowd-Arrow 2018] Buyukozturk, B.; Gaulden, S.; and Dowd-Arrow, B. 2018. Contestation on reddit, gamergate, and movement barriers. *Social Movement Studies* 17(5):592–609.

[Carlson 2018] Carlson, M. 2018. Facebook in the news. *Digital Journalism* 6(1):4–20.

[Conover et al. 2011] Conover, M.; Ratkiewicz, J.; Francisco, M.; Goncalves, B.; Menczer, F.; and Flammini, A. 2011. Political polarization on twitter.

[D. 1992] D., N. 1992. Controversy: Politics of technical decisions (3rd edition). *Sage Publications, Inc.* 8.

[David and Krantz 1999] David, and Krantz. 1999. The null hypothesis testing controversy in psychology. *Journal of the American Statistical Association* 94(448):1372–1381.

[Garimella and Weber 2017] Garimella, V. R. K., and Weber, I. 2017. A long-term analysis of polarization on twitter.

[Garimella et al. 2015] Garimella, K.; Morales, G. D. F.; Gionis, A.; and Mathioudakis, M. 2015. Quantifying controversy in social media.

[Guerra et al. 2013] Guerra, P. C.; Jr, W. M.; Cardie, C.; and Kleinberg, R. 2013. A measure of polarization on social media networks based on community boundaries.

[Helweg-Larsen and Shepperd 2001] Helweg-Larsen, M., and Shepperd, J. A. 2001. Do moderators of the optimistic bias affect personal or target risk estimates? a review of the literature. *Personality and Social Psychology Review* 5(1):74–95.

[Hendriks, Duus, and Ercan 2016] Hendriks, C. M.; Duus, S.; and Ercan, S. A. 2016. Performing politics on social media: The dramaturgy of an environmental controversy on facebook. *Environmental Politics* 25(6):1102–1125.

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

[Hogg et al. 2013] Hogg; A., M.; A., K.; and van den Bos Kees. 2013. Uncertainty and the roots of extremism. *Journal of Social Issues* 69(3):407–418.

[Javier et al. 2015] Javier, B.-H.; Walid, M.; Kareem, D.; and Ingmar, W. 2015. Content and network dynamics behind egyptian political polarization on twitter. 700–711.

[Karypis 1997] Karypis, G. 1997. Metis:unstructured graph partitioning and sparse matrix ordering system. *Technical Report*.

[Lamm and Myers 1978] Lamm, H., and Myers, D. G. 1978. Group-induced polarization of attitudes and behavior. volume 11 of *Advances in Experimental Social Psychology*. Academic Press. 145 – 195.

[Lewis et al. 2012] Lewis; Kevin; Gonzalez; Marco; Kaufman; and Jason. 2012. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences* 109(1):68–72.

[Matakos, Terzi, and Tsaparas 2017] Matakos, A.; Terzi, E.; and Tsaparas, P. 2017. Measuring and moderating opinion polarization in social networks. *Data Mining and Knowledge Discovery* 31(5):1480–1505.

[Morales et al. 2015] 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.

[Myers and Lamm 1976] Myers, D. G., and Lamm, H. 1976. The group polarization phenomenon. *Psychological Bulletin* 83(4):602–627.

[Newcomb and M. 1962] Newcomb, and M., T. 1962. Student peer-group influence. *The American college: A psychological and social interpretation of the higher learning.* 469–488.

[Nora 2019] Nora, L. 2019. In whose honour?: Understanding native american sports mascots, the washington redskins, and the divisive online discourse about them.

[Polley 1987] Polley, R. B. 1987. Exploring polarization in organizational groups. *Group and Organization Studies* 12(4):424–444.

[Schmidt, Joff, and Davar 2005] Schmidt, C.; Joff, G.; and Davar, E. 2005. The psychology of political extremism. *Cambridge Review of International Affairs* 18(1):151–172.