

Controversial information diffuses faster and further in Reddit

Jasser Jasser¹, Ivan Garibay², Steve Schinert², and Alexander Mantzaris³

¹ Department of Computer Science, University of Central Florida, Orlando, FL, United States

² Department of Computer Science, University of Central Florida, Orlando, FL, United States

³ Department of Statistics and Data Science, University of Central Florida, Orlando, FL, United States

{Jasser.Jasser,Ivan.Garibay,Steve.Scheinert,Alexander.Mantzaris}@ucf.edu

Abstract. Online actors seeking to further spreading their message frequently resort to polarizing and radicalizing social media users by creating controversial and divisive content. In some cases, these messages contain false misleading information. Tracking these messages and their influence would help in further understand disinformation on social media and formalize a strategy to control its spread. This work asks whether controversially initiated information cascades have different characteristics than those not controversial in Reddit. From the perspective of a Reddit user, when authoring a comment to contribute, there is a question as to what value can be derived from making the content controversial or not, and what effect should be expected. This question is explored using data collected from Reddit consisting of 36 million comments on posts related to cybersecurity issues. From the results, it can be seen that controversial content travels faster and further from its origin. Being able to monitor this effect over time improves our ability to predict information cascades' characteristics and highlight users that exploit this avenue for disseminating misleading content.

Keywords: Controversiality, Information, Diffusion, Reddit, Polarization

1 Introduction

The emergence of online social networks opened the door for many to amplify content freely. Much of this activity is benign and even positive, but this openness has also formed a platform for generating false narratives and promoting it to wider audiences for every individual with an internet access. There is active research into polarization on online social networks [24, 2]. Controversiality of content is one feature that could draw attention to that content [16]. The spreading and promotion of controversial content support the polarization of social networks, especially in the political sphere [6]. A question naturally arises of whether such behavior is beneficial to the content's promoter and that there

is an incentive for the activity. This work investigates this question and whether controversy within a discussion would bring more user attention to authored content and promote the spread of the content further and faster.

Reddit is a popular online platform that allows users to spread knowledge and share opinions through posts consisting of both textual and visual elements. It allows users to respond to others' posts by either commenting or voting for or against other users' posts and comments, known as up- and down-votes [22]. Social media users display a tendency to join to homogeneous communities that share the same beliefs and are dedicated to the same topic (echo chambers [23, 13]). Reddit provides spaces, known as *subreddits*, where these discussions occur and which facilitate more focused discussion, and the presence of controversy is genuine and not artificial. A study by the Pew Research Center (2013) concluded that 6% of online adults are Reddit users [9], and by the time of writing, Alexa⁴ ranks it as #6 in the United States and #15 globally⁵. Reddit uses up- and down-votes to identify controversial content; with these votes representing agreement or disagreement, respectively, if the comment received a fair amount of polarized votes, then the comment is considered controversial. The identification of controversial content is determined by the Reddit's formula for controversial comments, the final labeling for which is provided through the Reddit API. According to Reddit's definition, a comment is controversial if 1) the sum of up- and down-votes is greater than or equal to 7, and 2) the ratio of up- and down-votes lies between selected upper ($ub = 0.6$) and lower bounds ($lb = 0.4$)⁶ as shown in equation (1).

$$C_i = \begin{cases} True, & lb < c < ub \\ False, & else \end{cases} \quad (1)$$

Here C_i is the comment index and whether it is classified as controversial or not. Additionally, *subreddit* moderators may tag a post or comment as controversial, regardless of the number of votes. We assume that the tagging of a comment to be controversial is the result of users being invested in the discussion and that the polarization in the voting comes as users are swayed by some aspects of the discussion.

This work explores the effect of the introduction of controversial content in Reddit comments by observing the differences between conversations that are labeled as controversial and those that are not. The results show that conversations that contain controversial content during the early stages of their lifetime possess more activity, and users involved in that conversation have a wider influence than conversations that are not labeled as controversial.

⁴ A web traffic analysis company owned by Amazon

⁵ <https://www.alexa.com/siteinfo/reddit.com>

⁶ We found the algorithm responsible for identifying controversial comments in the *reddit-archive* repository in GitHub, <https://github.com/reddit-archive/reddit/blob/753b17407e9a9dca09558526805922de24133d53/r2/r2/models/builder.py>

2 Related Work

The concept of polarization and controversiality has been explored in multiple fields within the literature, from its identification in social networks and blogs [1, 6, 3] to its detection in online news and web pages [15, 5, 7]. There are two main themes of the research, the *identification* and *quantification* of controversy within the discussion. Other literature explores the effect of controversial information such as the work of [12], which studies the effect of collective attention on the evolution of controversial debates. The work of [19] explores the effect of the controversy on the emotions and language within online news.

The identification of controversy in social discussions is an area of on-going research with many different approaches to the identification of controversiality. The work of [5] identifies controversy within a topic by looking at the sentiment of the discussion and the range of sentiment polarity across the users involved in the controversy, as measured through natural language processing and sentiment analysis. Similarly in the work of [20], the authors proposed identifying controversial events by also considering the measurement of polarized opinions in the wake of a widely-viewed event. The work of [21] investigated opinion distribution in social media to identify controversial discussions and the size of their opinion groups.

Case studies that have attempted to quantify controversy and controversiality have focused on techniques and methods to measure the amount of controversy within a network or discussion. The work of [11] quantifies controversy in online echo chambers by considering a topic that is controversial and building the conversation graph of that topic which depicts the opinion alignment between the users. The authors then partition the graph to identify the sides of the controversy where the amount of controversy is measured from the characteristics of that graph. The work of [14], compares two types of social networks, those who are formed from polarized context and those who are not by analyzing the boundary between a pair of communities. The authors then distinguish between the polarized from the non-polarized communities by the concentration of high-degree nodes found in the boundary. The authors found that polarized networks tend to have low concentration "popular (high-degree)" nodes along the boundary between communities.

3 Data and Investigation

The data used for this investigation is collected from Reddit and provided as part of the "Computational Simulation of Online Social Behavior (SocialSim)" DARPA program⁷. This vast dataset consists of more than 36 million comments on posts related to cybersecurity issues. For this study, we considered the posts in our dataset that have at least 100 comment. The number of posts collected is 47,940 where the total comments in those posts exceeded 16 million comments.

⁷ <https://www.darpa.mil/program/computational-simulation-of-online-social-behavior>

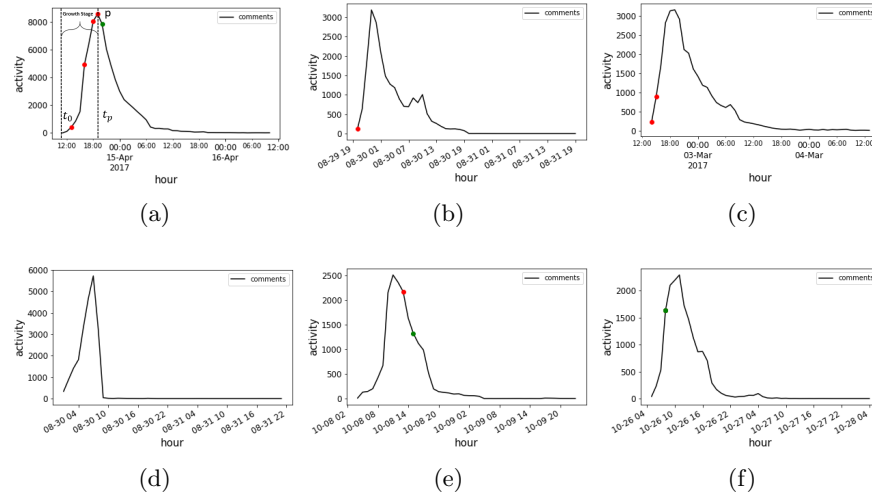


Fig.1: Controversially initiated and non-controversially initiated cascades, (a,b,c) are controversially initiated posts' cascades while (d,e,f) are non-controversial posts' cascades where the red dots represent a comment labeled as controversial by Reddit that is directed to the post's author while a green dot is a comment labeled controversial by Reddit that is directed to another comment.

the posts are then analyzed to be classified either as controversially initiated or non-controversial based on whether there exist any comments that are labeled controversial according to Reddit (as explained above) within the early stages of the posts. The classification yielded 23,101 posts that are labeled as controversially initiated, and 24,839 are labeled non-controversial. The distribution of controversial and non-controversial cascades from the same *subreddits* is almost equal in quantity. As an example, there are 8,528 controversial and 12,349 non-controversial cascades coming from the '*pcmasterrace*' *subreddit*, and 5,450 controversial and 4,149 non-controversial cascades coming from the '*android*' *subreddit*.

In social networks such as Reddit, users publish posts in regard of a specific topic where they share information and opinions. The posts are time stamped and so are the comments from the users that replied to that post. This forms a temporal propagation graph, also known as information cascade [4, 18]. The classification of cascades is based on the enclosure of controversially labeled comments within the posts that form the cascade. Cascades have a growth stage which is the period between the time of the first comment (t_0) and the peak time (t_p) of the post's cascade as shown in Fig1(a). We define a cascade that follows a post's as controversial if the cascade contains a comment, directed to the post's author, that is labeled controversial and occurred within the growth stage.

Assuming that the discussion reached a level of polarity in the opinions within the growth stage of the cascade that drove the discussion. We do not consider controversial comments that happened after the growth stage since at that stage users have lost interest in the topic and the polarization did not reinvigorate the discussion. Fig1 shows posts' cascades where Subfigures 1(a,b,c) are classified as controversial while Subfigures 1(d,e,f) are classified as non-controversial. Notice Subfigures 1(e,f) where in Subfigure 1(e) there are two controversial comments after its growth stage and for that it was classified as non-controversial, and in Subfigure 1(f) there is one controversial comment that is within the growth stage, however, it was not directed to the post's author and so it was not classified as controversial.

The data come with the sentiment polarity and subjectivity of the comments calculated using the "pattern.en" model⁸. The model ranks the polarity of the comment sentiment based on the contained adjectives on a scale from -1 to 1 where -1 means highly negative sentiment while 1 is a highly positive sentiment. The subjectivity is ranked by the model on a scale from 0 to 1 where 0 is very objective and 1 is very subjective. The sentiment polarity and subjectivity of the authors were calculated by taking the average sentiment of all comments written by that author. If an author wrote a comment that got labeled as controversial by Reddit, that author is labeled as controversial. The dataset contains 1,172,886 unique authors. Controversial authors consist of around 0.06% of that total or 69,607 authors to be precise. Our data shows that authors that are labeled controversial tend to have balanced polarity that is normally distributed with a mean slightly above 0, meaning they are more likely to be positive in their comments. Unlike the controversial authors, non-controversial authors mostly have no sort of polarity in their text. In terms of subjectivity, controversial authors tend to have a balance between being objective or subjective about the topic, hence their sentiment subjectivity average are closer to 0.5, while non-controversial authors are mostly objective in their comments.

A social network is a form of complex network [10], a social structure that consists of social actors and the relationships between them. The relationship and connectedness between the actors define and distinguish those that are influencing the network from those who do not. The work of [17], argues that most influencers are found within the k -core of the network. A k -core is the "largest subgraph where vertices have at least k interconnections" [8]. We considered the independent authors that commented on each other comments for every post to measure the connectedness among them and discovering the subset of nodes with the highest *coreness*. This subset of authors defines the nodes responsible for influencing the dissemination of information. The k -core of that graph network is analyzed where k is $\max(\text{coreness})$. We are interested in finding the number of nodes that are labeled controversial and their ratio within these k -cores.

We investigated whether the earlier the resurgence of the controversial comments within the cascade would lead to drastic changes in the characteristics by defining epochs within the growth stage of the cascade. The period is then

⁸ <https://www.clips.uantwerpen.be/pages/pattern-en>

divided into three epochs, epoch 1 for the first quartile, epoch 2 for the second quartile, and epoch 3 for the second half of the growth stage. If one of these epochs contains the majority of the controversial comments, we consider that epoch as the period with the highest concentration of controversy, and we associate the post’s cascade with that period. An epoch 1 cascade means that the cascade’s highest concentration of controversy happened during the first quartile of the growth stage.

4 Results

The results of the analysis show that posts’ cascades that contained controversially labeled messages during its growth stage produce a larger total number of comments and a larger burst, indicating more involvement from users. Another feature is that more unique users are involved in the cascades which is controversially labeled. This clearly shows that its beneficial for a user who seeks to spread a particular message faster and further, to infuse controversy within the context of the message. The plots in Figure 2 show this effect presented on three different measurements. Each plot is a CCDF (complementary cumulative distribution function) where the x-axis is the quantity in question shown in the title and the y-axis is the probability for that measure to exceed that x-axis value. In subfigure 2(a), shows that the posts that attract the largest contributions are dominated by initial signals of controversy. As the cascades grew, a difference emerged; non-controversial cascades did not exceed 25,000 comments whereas the controversial cascades did display posts having 60,000 or more comments. Subfigure 2(b) shows the peak size probability where controversial cascades tend to generally have larger peaks. Subfigure 2(c) shows the probability for cascades to exceed a certain number of unique authors contributing comments to a post. Along with this measure it can be seen how the controversially initiated cascades produce much larger values for the upper tail of the distribution.

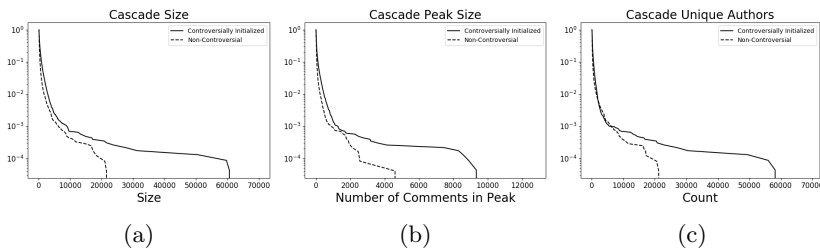


Fig. 2: Descriptive analysis results of posts’ cascade, (a) shows the difference in size between controversially initiated post’s cascade and non-controversial one. (b) shows the difference in peak size, and (c) shows the difference in the number of unique authors

Figure 3 shows the results from conducting a network analysis on the post structure in the Reddit dataset. The results here show a more distinctive differentiation between the two types of cascades. Subfigure 3(a) shows the CCDF for the network size of the controversially and non-controversially initiated cascades where the network size is computed by using the total number of edges produced within the entity of the post. Subfigure 3(b) looks at the number of nodes within the defined k -core in the cascades' network. It is clearly visible that controversial cascades end up containing more nodes within their k -core and so are the controversial nodes within these k -cores as shown in Subfigure 3(c). This shows that network properties of the different cascades are more affected by content which attracts controversial initiations or that they shape the subsequent discussions.

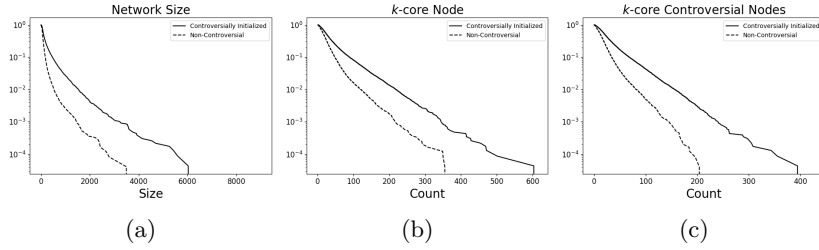


Fig. 3: Network analysis results of posts' cascade, (a) shows the difference in network size (number of edges) between controversially initiated post's cascade and non-controversial one. (b) shows the difference in the number of nodes within the networks' k -core, and (c) shows the difference in the number of controversial nodes within the networks' k -core

Figure 4 shows an evolutionary temporal analysis of the network changes within the first 24 hours. Subfigure 4(a) shows the evolution the number of links in the network. The activity for the controversial cascades is greater and for the time duration observed has this increase from each in the post creation. Subfigure 4(b) shows the k -core evolution where the number of nodes within the controversial k -core increased drastically over the other ones. Subfigure 4(c) compares the evolutionary number of controversial nodes found within the k -core. This reinforces that the cascades which contain controversial activity during their early stages continue to swiftly progress for longer periods of time than the non-controversial cascades.

Figure 5 shows the results of comparing the first three epochs by the cascade size and the peak size of the cascade. Subfigure 5(a) shows that epoch 1 controversial cascades produced the largest number of comments and peak size, followed by epoch 2 then epoch 3 cascades. This indicates that the earlier the resurgence of the controversiality within the cascade the further the information will disseminate within, while later resurgence of controversial comments might not drastically affect the information flow.

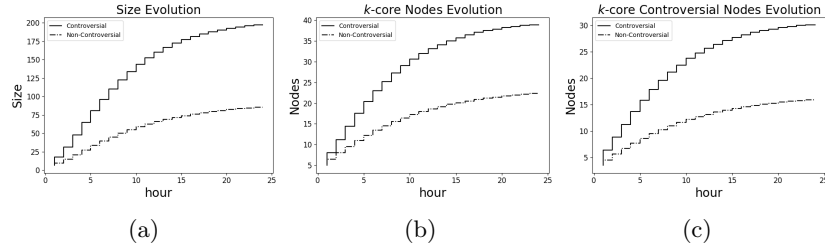


Fig. 4: Network evolution analysis results of posts' cascade, (a) shows the difference in network size (number of edges) hourly growth between controversially initiated post's cascade and non-controversial one. (b) shows the difference in the hourly growth of the number of nodes within the networks' k -core, and (c) shows the difference in the hourly growth of the number of controversial nodes within the networks' k -core

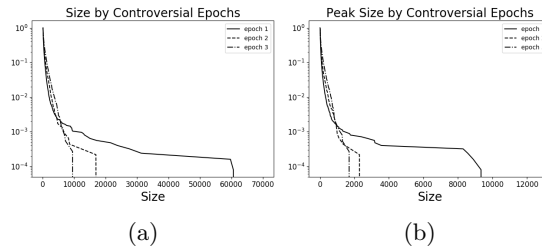


Fig. 5: The posts' cascades size and peaks size based on the epoch that contains the highest concentration of the controversial comments. (a) shows the comparison of posts' cascades size, while (b) shows the comparison of the peak size

5 Conclusion

The work sought to explore whether controversial content will have a greater chance of increased activity amongst other users. The analysis uses descriptive statistics and network analysis for the amount of activity a cascade produced. The k -core analysis conducted to study the structure of the networks' core. The temporal dimension is also explored in an analysis of the evolution of contributors. The results show that content which was controversial is associated with higher degrees of activity. This can shed light on various marketing strategies for competing for attention online. Even if not completely understood as to why this occurs, it can be utilized by accounts on social media seeking to create more influence. It can also help explain the spread of disinformation.

6 Acknowledgments

This work was partially supported by grant FA8650-18-C-7823 from the Defense Advanced Research Projects Agency (DARPA). The views and opinions contained in this article are the authors and should not be construed as official or as reflecting the views of the University of Central Florida, DARPA, or the U.S. Department of Defense.

References

1. Adamic, L.A., Glance, N.: The political blogosphere and the 2004 u.s. election: Divided they blog. In: Proceedings of the 3rd International Workshop on Link Discovery. pp. 36–43. LinkKDD '05, ACM, New York, NY, USA (2005). <https://doi.org/10.1145/1134271.1134277>, <http://doi.acm.org/10.1145/1134271.1134277>
2. Akoglu, L.: Quantifying political polarity based on bipartite opinion networks. In: Eighth International AAAI Conference on Weblogs and Social Media (2014)
3. An, J., Quercia, D., Crowcroft, J.: Partisan sharing: Facebook evidence and societal consequences. In: Proceedings of the Second ACM Conference on Online Social Networks. pp. 13–24. COSN '14, ACM, New York, NY, USA (2014). <https://doi.org/10.1145/2660460.2660469>, <http://doi.acm.org/10.1145/2660460.2660469>
4. Bikhchandani, S., Hirshleifer, D., Welch, I.: A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy* **100**(5), 992–1026 (1992)
5. Choi, Y., Jung, Y., Myaeng, S.H.: Identifying controversial issues and their sub-topics in news articles. In: Pacific-Asia Workshop on Intelligence and Security Informatics. pp. 140–153. Springer (2010)
6. Conover, M.D., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on twitter. In: Fifth international AAAI conference on weblogs and social media (2011)
7. Dori-Hacohen, S., Allan, J.: Detecting controversy on the web. In: Proceedings of the 22nd ACM international conference on Conference on information & knowledge management. pp. 1845–1848. CIKM '13, ACM, New York, NY, USA (2013). <https://doi.org/10.1145/2505515.2507877>, <http://doi.acm.org/10.1145/2505515.2507877>
8. Dorogovtsev, S.N., Goltsev, A.V., Mendes, J.F.F.: K-core organization of complex networks. *Physical review letters* **96**(4), 040601 (2006)
9. Duggan, M., Smith, A.: 6% of online adults are reddit users. *Pew Internet & American Life Project* **3**, 1–10 (2013)
10. Easley, D., Kleinberg, J., et al.: *Networks, crowds, and markets*, vol. 8. Cambridge university press Cambridge (2010)
11. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M.: Quantifying controversy in social media. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining. pp. 33–42. WSDM '16, ACM, New York, NY, USA (2016). <https://doi.org/10.1145/2835776.2835792>, <http://doi.acm.org/10.1145/2835776.2835792>

12. Garimella, K., De Francisci Morales, G., Gionis, A., Mathioudakis, M.: The effect of collective attention on controversial debates on social media. In: Proceedings of the 2017 ACM on Web Science Conference. pp. 43–52. WebSci '17, ACM, New York, NY, USA (2017). <https://doi.org/10.1145/3091478.3091486>, <http://doi.acm.org/10.1145/3091478.3091486>
13. Garrett, R.K.: Echo chambers online?: Politically motivated selective exposure among internet news users. *Journal of Computer-Mediated Communication* **14**(2), 265–285 (2009)
14. Guerra, P.C., Meira Jr, W., Cardie, C., Kleinberg, R.: A measure of polarization on social media networks based on community boundaries. In: Seventh International AAAI Conference on Weblogs and Social Media (2013)
15. Kaplun, K., Leberknight, C., Feldman, A.: Controversy and sentiment: An exploratory study. In: Proceedings of the 10th Hellenic Conference on Artificial Intelligence. pp. 37:1–37:7. SETN '18, ACM, New York, NY, USA (2018). <https://doi.org/10.1145/3200947.3201016>, <http://doi.acm.org/10.1145/3200947.3201016>
16. Kim, J., Wyatt, R.O., Katz, E.: News, talk, opinion, participation: The part played by conversation in deliberative democracy. *Political communication* **16**(4), 361–385 (1999)
17. Kitsak, M., Gallos, L.K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H.E., Makse, H.A.: Identification of influential spreaders in complex networks. *Nature physics* **6**(11), 888 (2010)
18. Leskovec, J., Krause, A., Guestrin, C., Faloutsos, C., Faloutsos, C., VanBriesen, J., Gance, N.: Cost-effective outbreak detection in networks. In: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining. pp. 420–429. ACM (2007)
19. Mejova, Y., Zhang, A.X., Diakopoulos, N., Castillo, C.: Controversy and sentiment in online news. arXiv preprint arXiv:1409.8152 (2014)
20. Popescu, A.M., Pennacchiotti, M.: Detecting controversial events from twitter. In: Proceedings of the 19th ACM International Conference on Information and Knowledge Management. pp. 1873–1876. CIKM '10, ACM, New York, NY, USA (2010). <https://doi.org/10.1145/1871437.1871751>, <http://doi.acm.org/10.1145/1871437.1871751>
21. Qiu, J., Lin, Z., Shuai, Q.: Investigating the opinions distribution in the controversy on social media. *Information Sciences* **489**, 274–288 (2019)
22. Stoddard, G.: Popularity dynamics and intrinsic quality in reddit and hacker news. In: Ninth International AAAI Conference on Web and Social Media (2015)
23. Sunstein, C.R.: *Echo chambers: Bush v. Gore, impeachment, and beyond*. Princeton University Press Princeton, NJ (2001)
24. Taylor, C., Mantzaris, A., Garibay, I.: Exploring how homophily and accessibility can facilitate polarization in social networks. *Information* **9**(12), 325 (2018)