

# Measuring Conversational Contentiousness on Reddit with Sub-Graph Motifs

*Keywords: Conversations, Discussion, Motifs, Conflict, Community*

## Extended Abstract

**Introduction** Social media platforms bring users into conversation with people that they might not otherwise interact with, creating the opportunity for stimulating exchanges or heated disagreement. The conversations that take place reflect both the propensity for disagreement on a particular topic as well as the emotional tenor of the subject at a specific moment. Previous work has highlighted user-level personal and demographic characteristics that predict participation in conflictual online discussion and noted that such discussions are more likely to include outsiders with less connection to the group [1]. Other research has focused on the quality of discussion on Yahoo News, emphasizing the importance of language in constructive online discussions [2].

We expand on this literature by developing sub-graph motifs for Reddit and using them to measure conversational conflict across communities and over-time [3]. Reddit is a social network made up of communities created and run by users. Users make posts, which others can respond to in a nested comment-tree structure and then up-vote or down-vote those posts and comments. Voting awards *karma*, providing a psychic reward to users while also positively or negatively affecting a post or comment’s place in Reddit’s ranking algorithms. All else equal, a post or comment with more up-votes will be seen by more users than a post or comment that receives fewer up-votes or more down-votes.

**Data.** We construct sub-graph motifs between two users where User A makes a post or comment, User B replies to that post or comment and then User A up-votes or down-votes User B’s reply. We are interested in the ratio of motifs that end in a down-vote to those that end in an up-vote, which we’ll call *contentiousness*. User A’s vote conveys the value they ascribe to user B’s comment, with an up-vote conveying positive reinforcement for commenting and that other users should read the comment—a down-vote conveys the opposite.

Using posts from the first two months of 2023 and limiting our analysis to the 9,131 subreddits that generated 1,000 or more motifs in this period and which have more than 1,000 members, we compute an overall contentiousness score by subreddit. These scores ranged from 2.8 to approximately 0. The 10 most contentious subreddits are all dedicated to discussion of news or politics, with the single most contentious subreddit being, perhaps unsurprisingly, *r/israel\_palestine*. The least contentious subreddits are overwhelmingly of a sexual nature, but also feature private discussion groups of women expecting babies the same month or are built around collaboration such as in *r/counting*, where members collaborate on esoteric counting tasks. Using a dataset of subreddit topic labels for 1,252 of the largest non-age-restricted subreddits, the 5 most contentious topics were football (soccer), U.S. politics, science, economics and sports while the least contentious were computing, dogs, celebrities, cycling and gaming.

**Results.** Using survey data from Weld, Zhang & Althoff that asked users to rate subreddits they were members of, we regressed contentiousness against user assessments of subreddits along their 31 surveyed dimensions [4]. Controlling the family wise error rate at the .05 level, contentiousness was a significant predictor of time on subreddit, where a 1 standard-deviation

increase in contentiousness ( $\sigma = 0.18$ ) predicted a 1.5 minute increase in daily time spent on the subreddit. Conversely, the same increase in contentiousness led to an approximately .4 point (on a 10-point likert scale) drop in participants' assessments of the quality of content on the subreddit and a .45 point drop in their assessment of how "included and able to contribute...new and existing members" feel.

Finally, we explore how the contentiousness of a subreddit changes by following the dynamics of the 2021 Formula 1 Season in r/formula1. The 2021 season is infamous for repeated controversies and a close championship fight between leading drivers Max Verstappen and Lewis Hamilton. We calculated the daily contentiousness of the subreddit throughout the season, highlighting the days after a race, which are 26% more contentious ( $p < .05$ ) than other days. Figure 1 shows r/formula1's contentiousness by day, with race days highlighted in red and a loess smoother fit separately to the two sets of days. The five most controversial moments of the 2021 season as listed by First Sportz are highlighted with dashed lines. The controversies, and especially the first crash between the two drivers at Silverstone, are outliers, indicating that the contentiousness of discussions in the subreddit spiked when controversial events happened in the sport.

**Discussion.** The ability to detect changing dynamics with minimal computational effort provides moderators and community leaders the opportunity to monitor and respond to changes in the level of conflict in their communities. It could also provide users and recommendation algorithms with an opportunity to tailor users' experiences to their preferred conversational tenor. Users that prefer easy discussions might opt into image-based subreddits or subreddits based around providing mutual support, while other users might opt for more intense discussions about sports, politics or science.

It might seem intuitive that users might always prefer lower-contentiousness environments, but low-contentiousness can also manifest as an echo chamber. r/superstonk, where users egg each other on to make risky investments, often in meme stocks, has a lower contentiousness (.12) than r/stocks (.46), but the quality of financial advice is typically higher in the latter subreddit. Within a particular topic, the optimal amount of contentiousness is often not zero, as communities that fail to offer negative feedback can be overrun by an unquestioned consensus. In analyzing communities, one must bear in mind that some disagreement keeps a community vital and honest.

## References

- [1] S. Levy, R. E. Kraut, J. A. Yu, K. M. Altenburger, and Y.-C. Wang, "Understanding conflicts in online conversations," in *Proceedings of the ACM Web Conference 2022*, pp. 2592–2602, 2022.
- [2] C. Napoles, J. Tetreault, A. Pappu, E. Rosato, and B. Provenza, "Finding good conversations online: The yahoo news annotated comments corpus," in *Proceedings of the 11th linguistic annotation workshop*, pp. 13–23, 2017.
- [3] M. Coletto, K. Garimella, A. Gionis, and C. Lucchese, "Automatic controversy detection in social media: A content-independent motif-based approach," *Online Social Networks and Media*, vol. 3, pp. 22–31, 2017.
- [4] G. Weld, A. X. Zhang, and T. Althoff, "What makes online communities 'better'? measuring values, consensus, and conflict across thousands of subreddits," 2021.

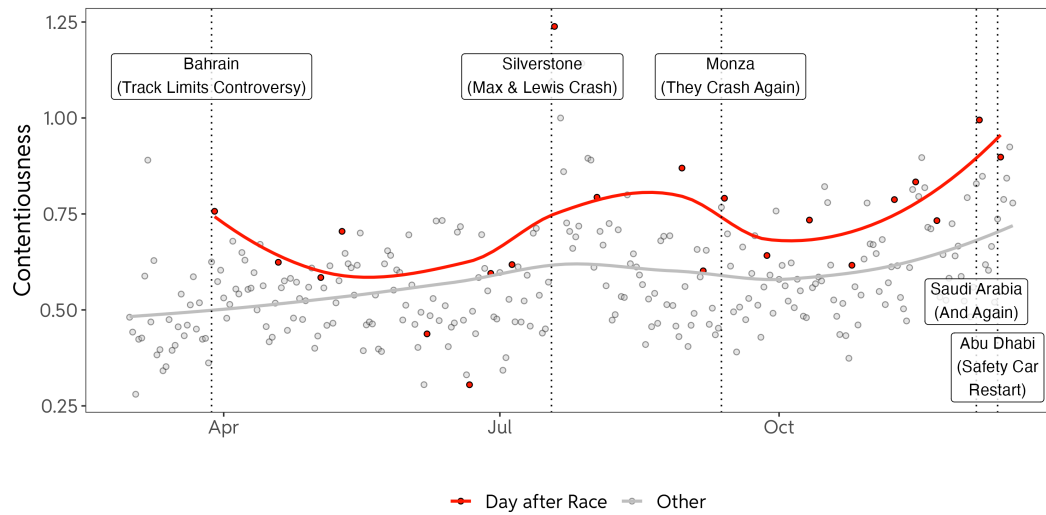


Figure 1: Contentiousness by day in 2021 for r/formula1