Leipzig University
Institute of Computer Science
Degree Programme Computer Science, B.Sc.

The Impact of Sorting Algorithms on Comment Rankings and Visibility: Analyzing Their Influence in an Online Persuasion Forum

Bachelor's Thesis

Donat Meirmanov

1. Referee: Prof. Dr. Martin Potthast

Submission date: March 25, 2025

Declaration

Unless otherwise indicated in the text or references, this thesis is entirely the product of my own scholarly work.

Leipzig, March 25, 2025

.....
Donat Meirmanov

Abstract

Reddit and its online discussion forum ChangeMyView have a significant influence on public opinion due to their ability to attract large, diverse communities that engage in open debates and exchange ideas. This thesis investigates how Reddit's sorting algorithms affect comment rankings, visibility, and their influence on online persuasion within the ChangeMyView online discussion forum. We collected a dataset of snapshots using both the Reddit API and browser scraping, and analyzed key metrics such as comment prominence and comment score across different sorting methods (Q&A, Best, Controversial, Top, New, and Old). Our statistical analysis including cumulative score evaluation, percentile distribution, and correlation tests shows that sorting algorithms impact comment prominence. In particular, sorts that incorporate comment score (e.g., Best and Top) display stronger associations between visibility and scoring. Furthermore, the Q&A sort, which prioritizes comments that have been responded to by the original poster, exhibits an even higher correlation between visibility and scoring, suggesting that the sorting algorithm drives this increased prominence and may influence reader opinions. These findings provide important insights into algorithmic content curation in online discussions and its effects on public opinion.

Contents

1 Introduction **1**

2 Related Work **3**

3 Reddit **5**
3.1 Reddit Terminology 5
3.2 ChangeMyView Subreddit 6
3.3 Comment Section 10
3.4 Sorting Algorithms 12

4 Key Statistics **15**
4.1 Dataset Metrics 16
4.2 Analysis of Comment Visibility 17
4.3 Analysis of Comment Score 21

5 Node Impression Calculation **23**
5.1 Setup 23
5.2 Procedure 26

6 Evaluation **29**
6.1 Statistics of Node Impression and Comment Score 29
6.2 Percentiles of Node Impressions 32
6.3 Comment Occurrence 34
6.4 Correlation of Node Impression and Comment Score 36
6.5 Statistical Significance of Node Impression and Comment Score 38
6.6 Distribution of Node Impressions 40

7 Conclusion **41**

8 Limitations **43**

Bibliography **45**

A Supplementary Figures and Tables **49**
A.1 Node Impression Distributions 49
A.2 Statistics of Node Impression and Comment Score 53

B Data Collection **57**
B.1 Reddit API 57
B.2 Snapshots 59
B.3 Data Storage and Cleaning 59

Chapter 1

Introduction

Social media platforms play a crucial role in information exchange and opinion formation. Among these, Reddit stands out due to its massive user base, making it one of the most visited websites globally with dynamic user-generated content and diverse discussion forums. In particular, the ChangeMyView (CMV) section on Reddit provides an ideal setting for examining debates and persuasive interactions. We focus on CMV because it is explicitly designed around structured argumentation with clear markers of persuasion success, making it uniquely suitable for analyzing the effectiveness of different persuasive strategies. The ranking and display of comments on Reddit are determined by proprietary sorting algorithms that affect which contributions are most visible. By prioritizing factors such as comment score and engagement, these algorithms inherently influence how discussions unfold and how persuasive arguments are perceived. This study specifically focuses on ChangeMyView, where these sorting mechanisms are employed to curate content. We examine the impact of these sorting algorithms on comment rankings and visibility, focusing on two key metrics: node impression, which measures the cumulative visibility of a comment over time, and comment score, which represents the net sum of upvotes minus downvotes cast by users as a direct expression of community approval or disapproval. We address several research questions:

- How do different sorting algorithms (Q&A, Best, Controversial, Top, New, and Old) affect the visibility and ranking of comments on ChangeMyView?
- What is the relationship between comment prominence and its score across these different sorts?
- How do variations in user behavior influence comment prominence?

We collected data from Reddit using both its API and browser snapshots to capture complete submission states. The API provided JSON snapshots of

posts and comments, while browser scraping ensured that the comment order matched what users see. A snapshot was taken every five minutes during the first 48 hours after submission creation, forming a dataset for our investigation.

To answer the first research question on how different sorting algorithms affect the visibility and ranking of comments on ChangeMyView, we compare key metrics across sorts. For the second research question regarding the relationship between a comment's node impression and its comment score, we perform correlation analyses and cumulative score evaluations. Finally, to address the third research question on how variations in user behavior influence comment prominence, we simulate distinct user types and model their reading patterns. This approach shows the impact of algorithmic content curation on online discussions and public opinion.

Our study begins with a review of existing research on online persuasion and algorithmic content curation in Chapter 2, establishing the theoretical groundwork for our investigation. Building on this foundation, we examine Reddit and the unique characteristics of the ChangeMyView subreddit in Chapter 3, providing context about the platform's structure and the community dynamics that shape persuasive discussions. We then detail our methodological approach in Chapter 4, explaining how we collected data through Reddit's API and browser scraping techniques, along with our strategies for data storage and preprocessing. The effects of sorting algorithms on comment visibility and scoring distributions are analyzed in Chapter 4, where we present key statistical results showing patterns in user engagement across different sorting methods. Chapter 5 introduces our methodology for calculating "node impression" and presents mathematical models that simulate various user behaviors. Our experimental results in Chapter 6 employ correlation analysis and statistical tests to demonstrate the relationship between comment prominence and user engagement. Finally, we synthesize our findings and discuss their implications in Chapter 7 and Chapter 8, acknowledging the study's limitations and suggesting directions for future research in this field.

Chapter 2

Related Work

The literature on persuasion in online discussions explores how linguistic, interactional, and personal factors influence persuasive outcomes. Early work on ChangeMyView shows that early participation and careful word choices can make a significant difference in persuasion [Tan et al., 2016]. Researchers demonstrate that effective debaters tend to write longer comments and adopt specific framing strategies [Wiegmann et al., 2022], and Khanna's master's thesis reveals that persuasive strategies evolve over time across multiple linguistic dimensions [Khanna, 2021]. Other studies shift focus toward the internal structure of arguments. A detailed annotation of semantic roles in argumentative units provides a granular view of how individual elements contribute to persuasion [Egawa et al., 2019]. In addition, analyses of semantic types of claims and premises in online forums [Hidey et al., 2017] and examinations of coherent argument structure [Li et al., 2020] enrich our understanding of persuasive discourse.

Significant advances are made in the assessment of argument quality, which is relevant to our investigation of how sorting algorithms may prioritize certain comments. Studies investigate what features of online comments elicit reactions from audiences across various discussion platforms [Jaech et al., 2015]. This research identifies key linguistic and structural factors that predict comment popularity, finding that simple features like comment length and timing significantly influence audience engagement. The analysis of reaction patterns across different communities reveals platform-specific preferences, suggesting that algorithmic content presentation may interact with these preferences to shape discussion visibility, which aligns with our investigation of sorting algorithms on Reddit. Researchers develop a comprehensive framework for computational argumentation quality assessment in natural language, proposing a hierarchical model that evaluates logical, rhetorical, and dialectical dimensions of arguments [Wachsmuth et al., 2017]. The approach offers potential metrics

for determining whether higher-ranked comments contain objectively stronger arguments, which could complement our analysis of algorithmic ranking effects. Similarly, a frame-oriented summarization of argumentative discussions, which identifies diverse perspectives within debates and could help evaluate whether comment sorting algorithms surface a representative range of viewpoints, is introduced [Syed et al., 2023]. These frameworks provide methodological tools that could be integrated with our approach to determine if comment rankings align with argument quality.

Further research delves into user behavior and social dynamics on ChangeMyView. For instance, one study characterizes user susceptibility to persuasive content [Mensah et al., 2019], while another models the cumulative influence of opinion holders over time [Guo et al., 2020]. Insights into how explanatory pointers reinforce key aspects of arguments are also reported [Atkinson et al., 2019]. Additional investigations address comment ranking and debater characteristics. By combining textual features, social signals, and measures of argument quality, researchers improve methods for ranking persuasive comments [Wei et al., 2016]. Other work represents comments as sequences of argumentative units to uncover common strategies [Mirzakhmedova et al., 2023], and personal characteristics derived from Reddit activity are used to predict persuasiveness more accurately [Khatib et al., 2020].

While these studies offer valuable insights into what makes arguments persuasive and the role of debater behavior, they primarily focus on micro-level features and individual dynamics. In contrast, our study investigates how Reddit's content sorting algorithms affect the visibility and popularity of arguments on ChangeMyView. We aim to bridge the gap between algorithmic content curation and persuasive communication by analyzing how sorting influences the displayed order of arguments, as measured by comment scores, and consequently the perceived strength and authority of opinions.

Chapter 3

Reddit

Understanding the structure of Reddit and ChangeMyView provides essential context for our analysis of comment sorting algorithms. The chapter opens with an overview of Reddit's massive global reach and its organization into specialized topic-based communities called subreddits. A detailed explanation of key terminology follows, including the mechanics of posts, comments, and the voting system that influences content visibility. Special attention is given to the ChangeMyView subreddit and its unique delta reward system designed to acknowledge opinion changes. Readers will discover how Reddit's comment section forms a hierarchical tree structure, with important distinctions between visible and collapsed comments. The final sections explore the six different sorting algorithms (Q&A, Best, Controversial, Top, New, and Old) and examine how various viewing modes affect the user experience.

3.1 Reddit Terminology

Reddit is an online social media platform and forum with user-generated content. It has evolved into one of the largest social media platforms globally, with an active user base exceeding 1.2 billion monthly users by 2024. Its extensive user activity places it among the top 20 most visited sites, which underscores its significant influence. Reddit is renowned for the openness and richness of its data. Unlike other platforms, users interact with content by following topics rather than individual accounts.

Reddit is organized into a system of subreddits, which are forums dedicated to specific topics. Each subreddit is defined by its own content guidelines and moderation structure that helps maintain discussion quality. With over 3 million subreddits, Reddit offers a wide variety of discussions covering both niche interests and trending global topics. Within subreddits, users post content known as submissions or posts that can be upvoted or downvoted by others

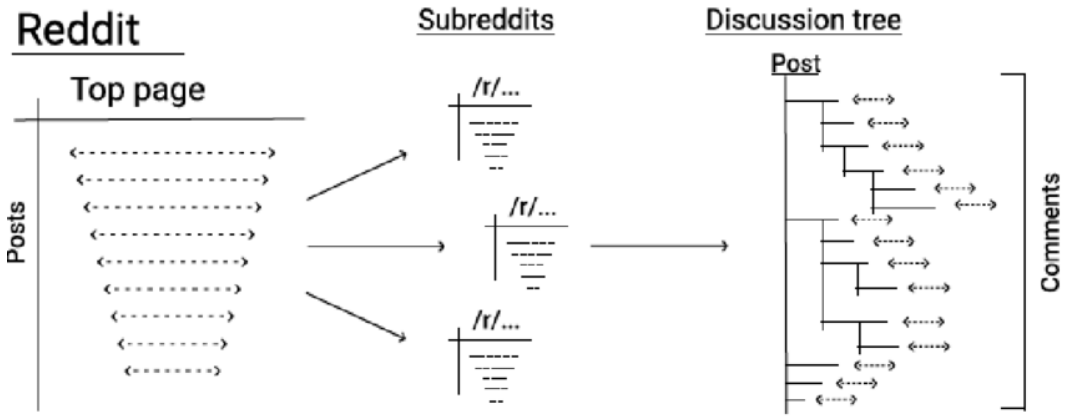


Figure 3.1: Reddit structure [Medvedev et al., 2018]

to express approval or disapproval. An upvote adds one point to the score of a post or comment, while a downvote subtracts one point; this aggregate value, known as the comment score, represents the net total of user votes and directly influences content visibility and positioning. Higher-scoring content is promoted to the top, enhancing its visibility and credibility, while lower-scoring content may be hidden or require an extra click to view. For definitions and additional details on Reddit terminology, we refer to the Reddit Help Center[1]. In our work, the original poster (OP) is the user who submits a post, and comments are the responses that form a discussion tree beneath the submission.

3.2 ChangeMyView Subreddit

ChangeMyView[2] is a subreddit that encourages open-minded discussions and the exchange of diverse perspectives, as illustrated in Figure 3.3. Users post their personal beliefs or viewpoints, inviting others to challenge them with constructive counterarguments. In our research, we focus on this particular subreddit due to its unique structure and explicit goal of changing opinions through rational discourse.

ChangeMyView employs a reward system that allows users to indicate a change of opinion. This reward, called 'Delta' and represented by the Greek letter 'Δ', signifies change or difference. When a user's opinion is changed, they respond with the delta symbol to the user who influenced them, which serves as an award. The ChangeMyView bot, called DeltaBot and operated by the moderators, recognizes the delta symbol and updates the awarded user's

[1] `https://support.reddithelp.com/hc/en-us`

[2] Visit `https://www.reddit.com/r/changemyview` for more information.

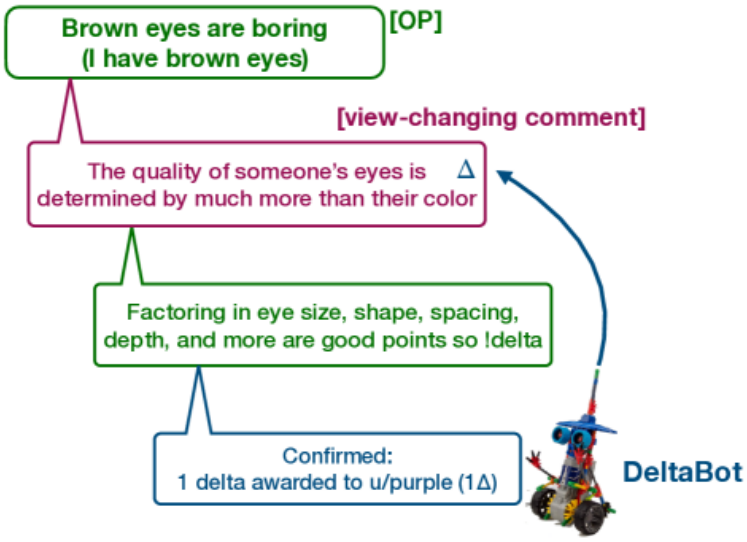


Figure 3.2: The awarding flow on ChangeMyView [Peguero and Watanabe, 2024]

delta score. All posts and comments on ChangeMyView must adhere to strict community rules[3], which are designed to maintain respectful and focused discussions. These rules ensure that debates remain constructive and that all contributions are made in an open-minded and civil manner. They apply equally to original posts and subsequent comments, providing a framework that supports meaningful exchanges. Unlike most subreddits which default to Best sorting, ChangeMyView uses Q&A as the default comment sorting method. For more information on the different sorting algorithms, see Section 3.4. It should be noted that, specific to CMV, comment scores are not visible on any comments during the first 24 hours after a submission is created. Users can still vote during this period, and the sorting algorithms continue to use these votes to determine comment ordering, but the actual score numbers remain hidden.

The delta score is displayed next to the username and is visible to all users. A leaderboard showing delta scores is maintained for various time periods, ranging from daily to all time. A chronological and topic-based history of Delta awards is also available. This system highlights influential opinions and encourages meaningful discussion.

[3]CMV rules are stated on the ChangeMyView wiki: `https://www.reddit.com/r/changemyview/wiki/rules`

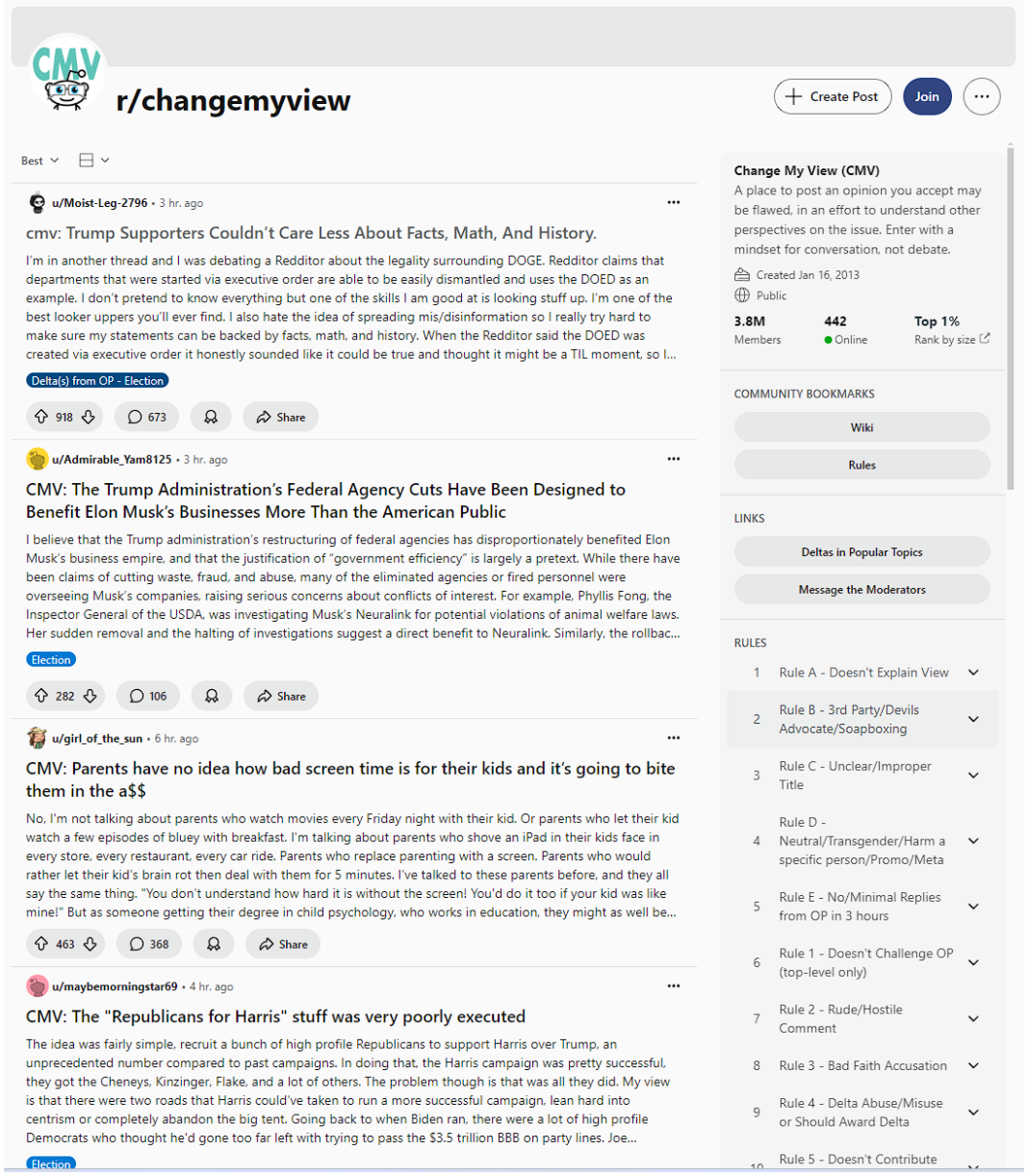


Figure 3.3: ChangeMyView Page with a list of submissions

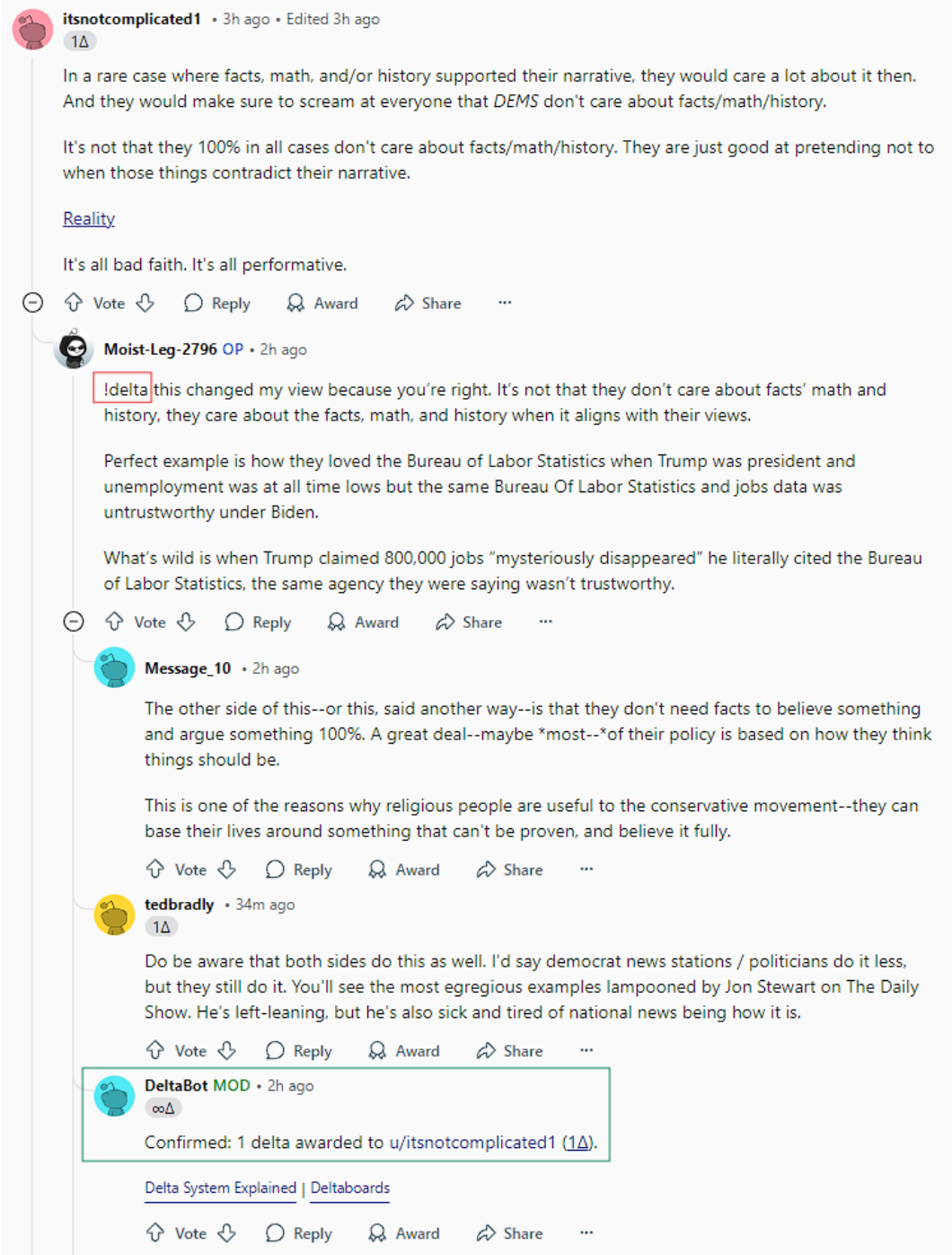


Figure 3.4: Delta awarded. DeltaBot responds to the command (red) and confirms the award (green).

3.3 Comment Section

The comment section of a Reddit submission, as shown in Figure 3.5, can be visualized as a hierarchical tree structure. The root of the tree is the original post (submission) from which the entire comment thread originates. Each top-level comment is a direct response to the submission, forming the first level of branches. Replies to these top-level comments branch out as child nodes. Each reply may have its own set of responses, creating a nested chain of comments that can continue indefinitely.

To manage the potentially extensive comment trees, Reddit implements a comment collapsing feature that controls content visibility in threads. When a comment is in a 'collapsed' state, only the username, vote score, and an expansion symbol are visible, while the comment text and its reply branches remain hidden. Reddit automatically collapses comments based on several factors including depth in the comment tree, vote score thresholds, or moderator actions. Users can manually collapse visible comments by clicking the collapse icon, which reduces the displayed content in the discussion. To reveal a collapsed comment and its replies, users click the '[+]' symbol or the comment header, which expands the hidden content and displays it in the thread.

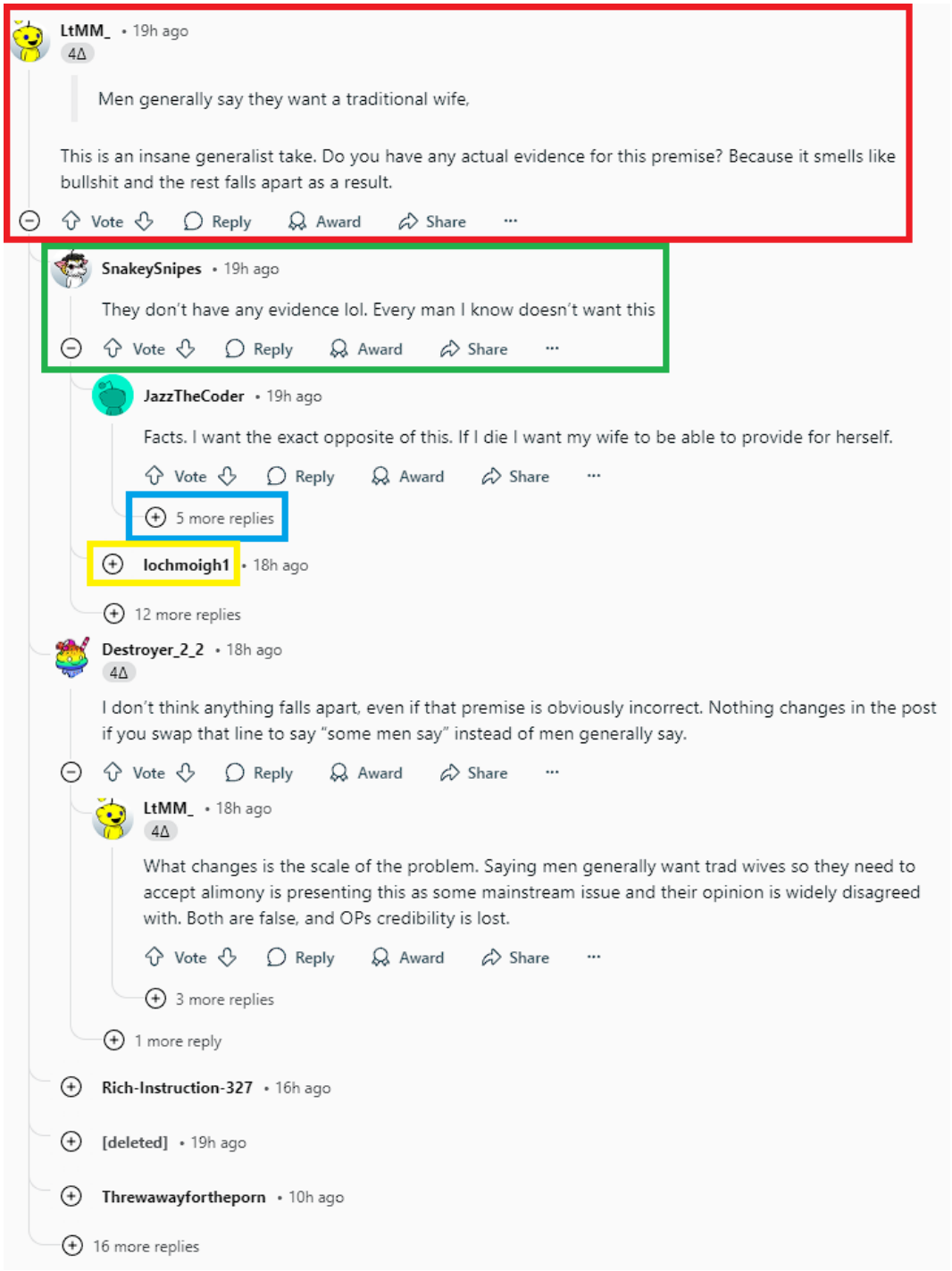


Figure 3.5: Top-level comment (red), child comment (green), collapsed comment section (blue), collapsed comment (yellow).

3.4 Sorting Algorithms

Reddit provides multiple sorting options for posts and comments, as shown in Figure 3.6, each designed to cater to different user preferences and requirements. The sorting of the comments is more important for our research, because the debates on ChangeMyView take place in the comments section.

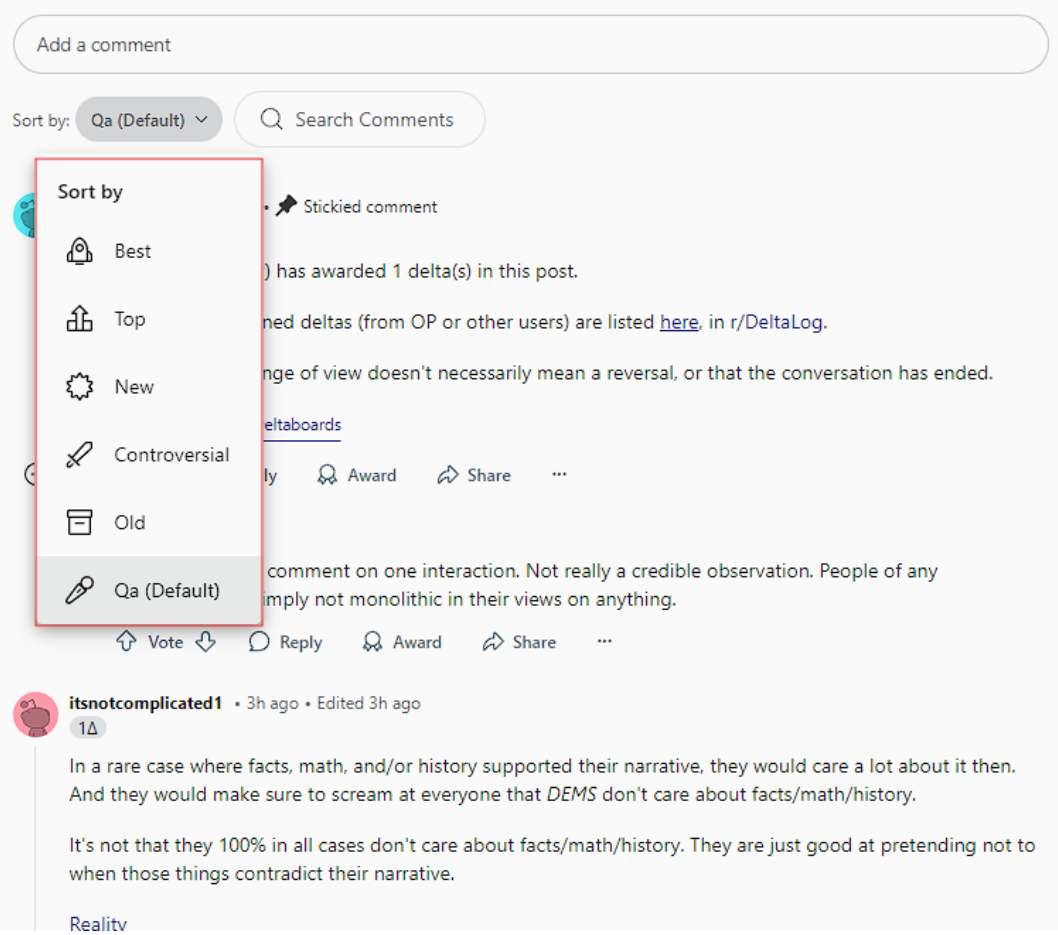


Figure 3.6: The list of available sorting algorithms in the comment section.

Reddit's comment sorting algorithms are proprietary and not fully accessible for external analysis. Nevertheless, we provide a brief overview of the main sorting options based on publicly available information.

Top: Shows comments with the highest net score (upvotes minus downvotes) first.

Best: Uses an algorithm that ranks comments by taking into account both the ratio of upvotes to downvotes and the comment's creation time. This approach is designed to prioritize content that is not only highly rated but also

timely and engaging, ensuring that the most relevant and dynamic discussions are surfaced. It is internally called 'Confidence'.

New: Displays comments in chronological order, with the most recent comments appearing at the top.

Old: Lists comments in reverse chronological order, with the oldest comments appearing first.

Controversial: Prioritizes comments that have a balanced number of upvotes and downvotes, thereby highlighting discussions that have sparked divided opinions.

Q&A: Designed to facilitate interaction between users and the post author by starting with comments that the author has replied to. When multiple comments have received replies from the original poster, they are further sorted by comment score, with higher-scored comments appearing first. This is the default sorting algorithm on ChangeMyView.

3.4.1 'View More Comments' Button

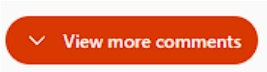


Figure 3.7: 'View More Comments' Button.

Upon opening a submission, Reddit preloads a set number of comments. When the user scrolls to the bottom of the page, a loading spinner appears, indicating that additional comments are being loaded. Subsequently, a button labeled 'View More Comments' (see Figure 3.7) becomes visible at the bottom of the page. To load further comments, the user must click this button. Premium users, however, see more comments by default.

3.4.2 View Modes

There are multiple methods to access and view Reddit. Each mode offers a distinct experience. In our study, we selected the perspective of a Reddit.com user with an active account who does not have a premium subscription.

- **Reddit.com logged-in user**: Logged-in users have access to the full range of platform features and personalized content, forming the primary subject group of this study.
- **Reddit.com anonymous**: Anonymous users, who are not authenticated, experience restricted content access, including a limited number of displayed comments.

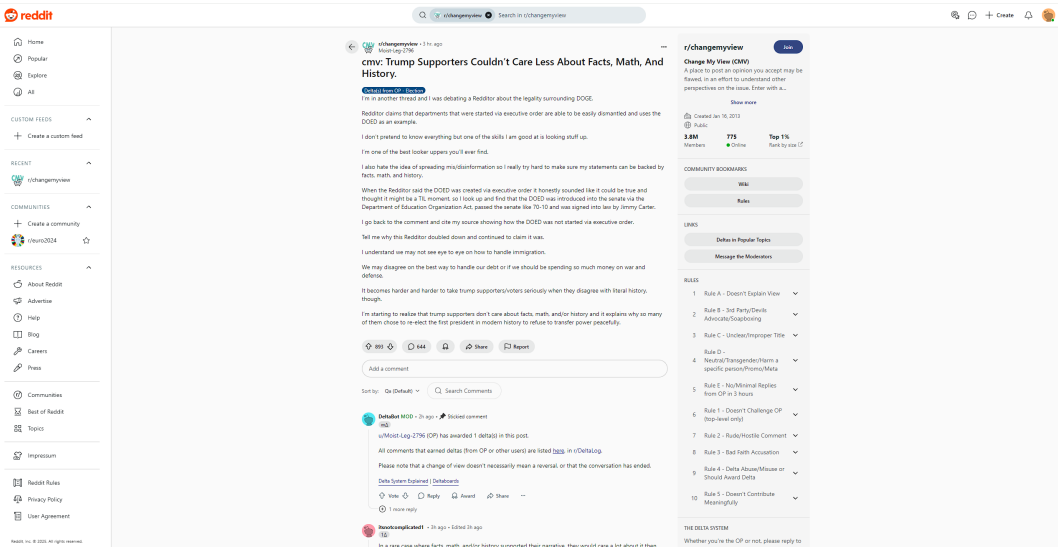


Figure 3.8: The UI of a logged-in user on Reddit.com.

- **Reddit Premium**: The Reddit Premium subscription provides an ad-free interface and extended access to comments and content.
- **Reddit mobile app**: The official mobile application is optimized for touch interaction and smaller screens.
- **new.reddit.com**: The new.reddit.com interface represents a redesigned version of the platform with updated navigational structures and content presentation, although it was discontinued on December 11, 2024.
- **old.reddit.com**: The old.reddit.com interface preserves the original layout and design, offering a consistent structure for users accustomed to the legacy format.
- **3rd Party over API**: Third-party applications utilize the Reddit API to retrieve and present content through customized interfaces.

Chapter 4

Key Statistics

Statistical analysis of our dataset reveals important patterns in comment visibility and scoring across different sorting algorithms. The first section outlines our comprehensive data collection approach using both the Reddit API and browser scraping to capture accurate comment ordering. Core metrics from the dataset are presented next, covering 400,459 state snapshots from 149 discussions with 28,163 comments by 7,421 users. Analysis reveals that nearly half of all comments remain hidden by default, with most collapsed comments rarely becoming visible again throughout a discussion's lifecycle. Particularly noteworthy is the cumulative score distribution, where early comments capture a disproportionate share of total engagement. Statistical correlations between initial comment activity and overall engagement patterns round out the chapter, establishing baseline metrics for deeper analysis.

Table 4.1: Dataset Metrics

Metric	**Value**
Snapshots	400459
Discussions	149
Comments	28163
Users	7421
Deltas	139
Score	92041

4.1 Dataset Metrics

Reddit data was collected using the Python Reddit API Wrapper (PRAW)[1] with moderator privileges to access hidden comment scores. The collection script captured submissions from ChangeMyView subreddit, taking snapshots every 5 minutes for 48 hours across six different comment sorting methods (Best, Q&A, Top, New, Old, Controversial). Due to discrepancies between API responses and browser display, HTML snapshots were captured using Playwright[2] to extract accurate comment order and visibility information. The JSON data was stored in MongoDB[3] to accommodate its complex structure with nested objects and optional fields. MongoDB's flexibility allowed for efficient querying despite the volume of data. Submissions with insufficient visibility data (less than 50%) or missing snapshots for any sorting algorithm were excluded from further analysis, as detailed in Appendix B.

The dataset described in Table 4.1 consists of 400459 state snapshots collected from 149 discussions within 48 hours of their posting. These discussions contain a total of 28163 comments contributed by 7421 unique users, excluding moderators and DeltaBot. The presence of 139 deltas indicates instances where the OP changed their opinion. The total score sums to 92041, comprising both downvotes and upvotes, so the actual vote count is higher but not directly derivable from the Reddit API.

Table 4.2 presents a statistical overview of key metrics. Most users comment sparingly, with a few contributing disproportionately high numbers of comments. Comment scores range widely, reflecting both strong downvotes and significant upvotes, though the average remains modest because a score of 1 is the default value in the API. Authors' engagement levels differ greatly, ranging from no comments to frequent participation, with the median indicating relatively active involvement. Deltas rarely appear in most discussions, with the median suggesting that most submissions had no delta.

[1]PRAW documentation: `https://praw.readthedocs.io/en/latest/`
[2]Playwright documentation: `https://playwright.dev/python/docs/intro`
[3]MongoDB documentation: `https://www.mongodb.com/docs/`

Table 4.2: Statistical Analysis of Dataset Metrics

Metric	Min	Max	Mean	Median
User Comment Frequency	1	159	3.59	1
Comment Score	-68	1075	3.51	1
Author Engagement	0	159	20.54	14
Deltas per Submission	0	9	0.93	0

4.2 Analysis of Comment Visibility

4.2.1 Visibility Analysis

The visibility analysis of different sorting algorithms shows minor differences. These discrepancies arise primarily from technical limitations in capturing snapshots simultaneously and from variations in the number of successful snapshots for each sorting method. Tables 4.3 to 4.8 provide an analysis of comment visibility for different sorting algorithms based on cumulative data collected from the submission time up to 6 hours, 12 hours, 18 hours, and 24 hours. 'Visible Count' represents the total number of comments that are directly visible within each cumulative period. 'Collapsed Count' indicates the number of comments that remain hidden and require an extra click to view. 'Avg Visible/Sub' and 'Avg Collapsed/Sub' show the average number of visible and collapsed comments per submission, respectively. 'Visibility Rate' is the percentage of comments that are visible relative to the total number of comments. The tables show that Reddit hides nearly half of the comments, which requires users to click on them to view the full content.

Table 4.3: Analysis of Comment Visibility for Sort Q&A

Q&A	6h	12h	18h	24h
Visible Count	4021	5345	6194	7094
Collapsed Count	2845	4043	4921	5961
Avg Visible/Sub	41.89	46.89	48.39	50.67
Avg Collapsed/Sub	29.64	35.46	38.45	42.58
Visibility Rate	58.56%	56.93%	55.73%	54.34%

Table 4.4: Analysis of Comment Visibility for Sort Best

Best	6h	12h	18h	24h
Visible Count	4060	5426	6458	7224
Collapsed Count	2835	3987	5183	5804
Avg Visible/Sub	42.29	47.60	50.45	51.60
Avg Collapsed/Sub	29.53	34.97	40.49	41.46
Visibility Rate	58.88%	57.64%	55.48%	55.45%

Table 4.5: Analysis of Comment Visibility for Sort Controversial

Controversial	6h	12h	18h	24h
Visible Count	3946	5306	6200	7058
Collapsed Count	3003	4251	5499	6276
Avg Visible/Sub	41.10	46.54	48.44	50.41
Avg Collapsed/Sub	31.28	37.29	42.96	44.83
Visibility Rate	56.79%	55.52%	53.00%	52.93%

Table 4.6: Analysis of Comment Visibility for Sort Top

Top	6h	12h	18h	24h
Visible Count	4118	5528	6468	7378
Collapsed Count	2746	3949	5081	5932
Avg Visible/Sub	42.90	48.49	50.53	52.70
Avg Collapsed/Sub	28.60	34.64	39.70	42.37
Visibility Rate	60.00%	58.33%	56.00%	55.43%

Table 4.7: Analysis of Comment Visibility for Sort New

New	6h	12h	18h	24h
Visible Count	4080	5456	6523	7455
Collapsed Count	3173	4543	5879	6878
Avg Visible/Sub	42.50	47.86	50.96	53.25
Avg Collapsed/Sub	33.05	39.85	45.93	49.13
Visibility Rate	56.25%	54.57%	52.60%	52.01%

Table 4.8: Analysis of Comment Visibility for Sort Old

Old	6h	12h	18h	24h
Visible Count	3835	5136	6090	6950
Collapsed Count	3302	4665	5960	6787
Avg Visible/Sub	39.95	45.05	47.58	49.64
Avg Collapsed/Sub	34.40	40.92	46.56	48.48
Visibility Rate	53.73%	52.40%	50.54%	50.59%

4.2.2 Development of Comment Visibility

We examined several metrics to better understand comment visibility by aggregating all snapshots of each comment. These metrics capture different aspects of comment behavior over time by analyzing how long a comment remains hidden and how quickly it becomes collapsed after posting. 'Time Collapsed' represents the percentage of the total duration during which a comment is in a collapsed state. 'Time-to-Collapse' measures the elapsed time from when a comment is first posted until it is first observed to be collapsed. 'Reactivation Ratio' indicates whether a comment that has been collapsed later becomes visible again. Low values here suggest that once a comment is hidden, it rarely recovers. For the input data, we only consider comments that have experienced a collapsed state at some point.

Tables 4.9 to 4.14 reveal that many comments are collapsed right after submission, suggesting that Reddit considers some users not sufficiently trustworthy based on its internal metrics. Other comments are collapsed due to downvotes by the community. Although it is generally possible for a collapsed comment to become visible again, it rarely recovers and tends to remain collapsed for the rest of its existence.

Table 4.9: Comment Development of Sorting Algorithm Q&A

Q&A	Mean	Median
Time Collapsed	96.39%	100%
Time-to-Collapse	38.12 min	0
Reactivation Ratio	0.03	-

Table 4.10: Comment Development of Sorting Algorithm Best

Best	Mean	Median
Time Collapsed	96.24%	100%
Time-to-Collapse	39.80 min	0
Reactivation Ratio	0.04	-

Table 4.11: Comment Development of Sorting Algorithm Controversial

Controversial	Mean	Median
Time Collapsed	96.63%	100%
Time-to-Collapse	54.13 min	0
Reactivation Ratio	0.04	-

Table 4.12: Comment Development of Sorting Algorithm Top

Top	Mean	Median
Time Collapsed	97.02%	100%
Time-to-Collapse	37.93 min	0
Reactivation Ratio	0.03	-

Table 4.13: Comment Development of Sorting Algorithm New

New	Mean	Median
Time Collapsed	96.75%	100%
Time-to-Collapse	26.49 min	0
Reactivation Ratio	0.03	-

Table 4.14: Comment Development of Sorting Algorithm Old

Old	Mean	Median
Time Collapsed	97.00%	100%
Time-to-Collapse	41.77 min	0
Reactivation Ratio	0.04	-

4.3 Analysis of Comment Score

4.3.1 Cumulative Score Analysis

In Table 4.15, we examine the accumulation of scores from the first n comments. The table reports two key metrics: Cumulative Score Share (S) and Total Comment Coverage (C) as percentages for each sorting algorithm. In this context, Top_n denotes the number of initial comments used for the calculation. Values of Top_n range from 5 to 50 in increments of 5. Cumulative Score Share represents the fraction of the total comment score contributed by these Top_n comments. Total Comment Coverage indicates the proportion of these Top_n comments relative to the overall comment count. There is no minimum or maximum threshold that restricts the number of comments in a submission. The table shows that the New sort yields the lowest score shares. This result provides a useful consistency check on the statistics. In contrast, the Q&A and Top sorts perform similarly. This observation suggests that the default Q&A sort on ChangeMyView promotes discussions where the original poster is actively engaged. The data also reveals that early comments capture a significant portion of the overall score. For example, in the Q&A sort, the first 10 comments account for a high share of the total score relative to their coverage. This pattern indicates that early comments are more likely to engage users and secure higher scores.

Table 4.15: Cumulative score analysis. 'Cumulative Score Share' (S) and 'Total Comment Coverage' (C) shown as decimal proportions (0-1), with leading zeros omitted for space efficiency. 'Contr.' is Controversial

	Q&A		Best		Contr.		Top		New		Old	
Top_n	S	C	S	C	S	C	S	C	S	C	S	C
5	.47	.16	.46	.16	.26	.16	.48	.16	.12	.16	.38	.17
10	.57	.26	.56	.25	.47	.26	.56	.25	.25	.25	.53	.26
15	.61	.35	.60	.35	.56	.35	.60	.34	.38	.35	.60	.36
20	.64	.43	.64	.42	.64	.43	.64	.42	.48	.42	.65	.44
25	.68	.50	.67	.49	.68	.50	.67	.49	.54	.49	.69	.52
30	.72	.57	.71	.56	.73	.57	.71	.56	.61	.56	.73	.58
35	.74	.61	.74	.61	.75	.61	.74	.61	.67	.61	.76	.63
40	.75	.64	.75	.65	.78	.66	.75	.63	.67	.61	.78	.66
45	.77	.67	.76	.66	.80	.69	.76	.66	.69	.65	.78	.67
50	.80	.73	.79	.72	.83	.74	.78	.70	.71	.68	.81	.71

4.3.2 Cumulative Score Correlation

The tables Table 4.16 and Table 4.17 present the Pearson correlation coefficients (PCC) calculated for different aspects of the initial comment contributions. Table 4.16 shows the correlation between the Cumulative Score Share (CSS) and Total Comment Coverage (TCC). These coefficients, computed using the Pearson correlation formula on the CSS and TCC values obtained from the Top_n comments for each sorting algorithm, as defined in subsection 4.3.1, illustrate a strong connection between the percentage of scores that these initial comments receive and their share among all comments, indicating that the initial comment sample significantly influences the overall score. In contrast, Table 4.17 displays the correlation between CSS and the raw count of initial comments (denoted as Top_n), evaluating how the absolute number of comments in the initial sample influences the CSS. A lower coefficient in this case suggests that the relative proportion of scores is more important than the number of comments in the initial sample.

These findings imply that the timing and ordering of comments in the initial sample play a crucial role in shaping engagement. Notably, the Controversial sorting algorithm exhibits higher coefficients, indicating a stronger link between initial activity and score accumulation in threads with divided opinions. Similarly, the New sorting algorithm shows a high correlation between score share and comment share, although its association with the number of initial comments remains moderate. Together, these patterns underscore that while the initial comment sample captures a significant portion of the total score, it is their relative contribution rather than their absolute count that most strongly drives overall engagement and perceived comment value on ChangeMyView.

Table 4.16: Correlation Coefficients for CSS Share and TCC.

Sort	PCC
Q&A	0.82
Best	0.83
Controversial	0.86
Top	0.82
New	0.90
Old	0.84

Table 4.17: Correlation Coefficients for CSS and Top_n comments.

Sort	PCC
Q&A	0.46
Best	0.45
Controversial	0.61
Top	0.43
New	0.54
Old	0.54

Chapter 5

Node Impression Calculation

Measuring comment prominence requires a sophisticated methodological framework, which this chapter develops through the concept of node impression. Opening with precise definitions of comment score metrics, the discussion progresses to a three-part model for calculating visibility scores. Collapsed state weight, content height weight, and comment depth each capture a different aspect of how comments appear to users. Mathematical models simulate three distinct types of Reddit users: Power Users who thoroughly read everything, Interested Users who selectively engage, and Fast Users who primarily skim top-level content. The chapter culminates in a detailed explanation of the node impression calculation procedure, complete with algorithmic steps that form the foundation for subsequent evaluation.

To find the correlation between comment ranking and scores, we conduct an experiment that aims to measure node impression, a metric indicating how prominently a node appears to different user types under specific sorting algorithms. To achieve this, we first determine the visibility of each node at each time point. The node impression metric is obtained by summing the visibility values across the entire discussion duration. This metric allows us to analyze how comment scores correlate with node impression for specific user types and sorting algorithms.

5.1 Setup

We calculate values from our mathematical model for each comment snapshot and aggregate these values to produce a cumulative assessment for each comment. We refer to comments as nodes because they are parts of the comment tree.

5.1.1 Comment Score and Normalized Comment Score

The comment score on Reddit is defined as the net score obtained by subtracting the number of downvotes from the number of upvotes a comment receives. In this experiment, we take the absolute value of comment scores so that all scores are positive, representing the total number of votes. Normalized comment score is computed by dividing the score of an individual comment by the sum of scores of all comments within the same submission.

5.1.2 Visibility Score

The visibility score of a comment snapshot is determined by multiplying three weights: the collapsed state weight C, the content height weight H, and the depth weight D. Each weight is computed based on the specific behavior of the user type. Values closer to 1 indicate higher prominence under the given sorting algorithm and user type assumptions.

- C indicates whether a comment is expanded or collapsed, with its value reflecting the user's tendency to uncollapse or ignore collapsed comments.
- H measures the relative content height of a comment within the overall snapshot, computed as

$$H(i) = 1 - \frac{h_i}{\sum_j h_j},$$

  where h_i is the height of the comment and $\sum_j h_j$ is the total content height, and the content height of a comment is calculated by dividing content length by an assumed average line length of 110 characters on the desktop version of Reddit.com.
- D represents the comment's position in the discussion tree, for example, assigning lower values to comments that are deeper in the hierarchy.

5.1.3 Modelling of User Behavior

A user initiates reading a CMV discussion by selecting a specific sorting algorithm. The discussion is presented starting at the root, with comments organized according to the chosen algorithm. Certain nodes remain hidden until the user expands them. Not all nodes may be read, and the user may eventually drop out of the discussion.

Little prior research has explored the behavior of Reddit users. Consequently, we decided to simulate user behavior through mathematical modeling. We define three distinct types of users to capture the varied ways in which

readers engage with ChangeMyView discussions. Power Users are very thorough: they expand all comments and pay equal attention to every part of the conversation, regardless of how deeply a comment is nested. In contrast, Interested Users show moderate engagement: they generally read beyond the top comments but tend to give less attention to hidden or deeply nested replies, reflecting a balanced yet selective approach. Fast Users, on the other hand, are highly selective and efficient: they primarily focus on the most visible, top-level comments and often skip over collapsed or deeply nested ones, quickly scanning the discussion for key points. Each user type applies a different weighting scheme for C, H, and D.

We assume that all user types begin reading at the root node and proceed downward through the tree view. We also assume that votes are cast immediately rather than after reviewing all possible votes to make an objective selection. Furthermore, we analyze the first 25 nodes to maintain a relatively consistent window size for visibility analysis. This window is chosen to ensure that the majority of top comments, which receive significant evaluations and attention, are captured. In practice, most of the relevant and influential comments appear within these 25 nodes, thereby providing a robust basis for our analysis.

For Power Users, the model assumes that all comments are treated as expanded and that depth is irrelevant, so they assign $C = 1$ and $D = 1$, while H is computed as

$$H(i) = 1 - \frac{h_i}{\sum_j h_j}.$$

Interested Users incur a modest penalty for collapsed comments by setting $C = 0.8$ and reduce the impact of deeper comments through an exponential decay in D, computed as

$$D(i) = e^{-0.1\, d_i},$$

where d_i is the depth of the comment, with H calculated as

$$H(i) = 1 - \frac{h_i}{\sum_j h_j}.$$

Fast Users are more selective, completely ignoring collapsed comments by assigning $C = 0$ when a comment is collapsed, and they consider only top-level comments by setting $D = 1$ for comments with minimal depth and $D = 0$ for deeper ones; for non-collapsed comments, H is computed as

$$H(i) = 1 - \frac{h_i}{\sum_j h_j},$$

while a default height of 1 is used for collapsed nodes.

5.1.4 Node Impression Definition

Node impression is a metric that quantifies the cumulative prominence of a comment across multiple snapshots. It is calculated by summing the visibility scores of a comment over time, each multiplied by a scale factor, defined by the change in overall comment scores relative to the final snapshot.

For each comment i and each snapshot s, we compute a visibility score $V_{i,s}$ defined as:

$$V_{i,s} = C_{i,s} \times H_{i,s} \times D_{i,s},$$

where:

- $C_{i,s}$ is the collapsed state weight, reflecting whether the comment is expanded or collapsed.
- $H_{i,s}$ is the content height weight, where $\sum_j h_j$ is the total content height of all comments in the snapshot.
- $D_{i,s}$ is the depth weight, which accounts for the comment's position in the discussion tree.

The node impression for comment i is then calculated by aggregating the visibility scores over all snapshots:

$$\text{NodeImpression}(i) = \sum_s \left(V_{i,s} \times \frac{\Delta S_s}{S_{\text{final}}} \right),$$

where ΔS_s is the absolute change in the cumulative comment score between snapshot $s-1$ and snapshot s (with $\Delta S_1 = 1$ for the first snapshot), and S_{final} is the total comment score in the final snapshot.

5.2 Procedure

For each submission, we first retrieve all available snapshots. Then, for each user type (PowerUser, InterestedUser, FastUser) and for each comment in each snapshot, we compute a visibility score $V_{i,s}$ as follows:

$$V_{i,s} = C_{i,s} \times H_{i,s} \times D_{i,s},$$

where $C_{i,s}$ is the collapsed state weight, $H_{i,s}$ is the content height weight, and $D_{i,s}$ is the depth weight; these values are determined by the specific behavior model of the user type. Next, we scale the computed visibility score by the ratio $\Delta S_s / S_{\text{final}}$, where ΔS_s represents the absolute change in the cumulative comment score between the previous snapshot and the current snapshot (with

ΔS_1 set to 1), and S_{final} is the total comment score in the final snapshot. This scaled score is then accumulated into a cumulative node impression for each comment. Finally, after processing all snapshots, we normalize the node impression for each comment by dividing by the total node impression across all comments for the same sort and user type.

5.2.1 Node Impression Calculation Example

Consider a submission with 5 comments, where we track comment C_3 across 3 snapshots.

For **Snapshot 1**, C_3 has depth=2, height=3, is expanded, with total height=12 across all comments.

PowerUser: $V_{3,1} = 1 \times (1 - \frac{3}{12}) \times 1 = 0.75$, scaled by $\frac{\Delta S_1}{S_{\text{final}}} = \frac{1}{100} = 0.01$ to get 0.0075.

InterestedUser: $V_{3,1} = 1 \times (1 - \frac{3}{12}) \times e^{-0.1 \times 2} = 0.614$, scaled to 0.00614.

FastUser: $V_{3,1} = 1 \times (1 - \frac{3}{12}) \times 0 = 0$ due to depth.

For **Snapshot 2**, C_3 has height=4, total height=15, with $\Delta S_2 = 30$ and scaling factor 0.3.

PowerUser: $V_{3,2} = 1 \times (1 - \frac{4}{15}) \times 1 = 0.733$, scaled to 0.22.

InterestedUser: $V_{3,2} = 1 \times (1 - \frac{4}{15}) \times 0.819 = 0.6$, scaled to 0.18.

FastUser: $V_{3,2} = 1 \times (1 - \frac{4}{15}) \times 0 = 0$ due to depth.

For **Snapshot 3**, C_3 is collapsed, with $\Delta S_3 = 69$ and scaling factor 0.69.

PowerUser: $V_{3,3} = 1 \times (1 - \frac{1}{18}) \times 1 = 0.944$, scaled to 0.651.

InterestedUser: $V_{3,3} = 0.8 \times (1 - \frac{1}{18}) \times 0.819 = 0.618$, scaled to 0.426.

FastUser: $V_{3,3} = 0$ for collapsed comments.

Final node impression:

PowerUser $= 0.0075 + 0.22 + 0.651 = 0.879$

InterestedUser $= 0.00614 + 0.18 + 0.426 = 0.612$

FastUser $= 0 + 0 + 0 = 0$

Normalized node impression (assuming totals of 4.5, 3.8, and 5.0):

PowerUser $= \frac{0.879}{4.5} = 0.195$ or 19.5%

InterestedUser $= \frac{0.612}{3.8} = 0.161$ or 16.1%

FastUser $= \frac{0}{5.0} = 0\%$

Algorithm 5.1: Compute Normalized Node Impression

Input: A set of submissions, each with snapshots for various comment sorts; user types {PowerUser, InterestedUser, FastUser}

Output: Normalized Node Impression for each comment

foreach *submission in Submissions* **do**
  Retrieve all snapshots for the submission;
  foreach *snapshot in snapshots* **do**
    foreach *userType in {PowerUser, InterestedUser, FastUser}* **do**
      foreach *comment in snapshot* **do**
        Compute weights $C_{i,s}$, $H_{i,s}$, and $D_{i,s}$ using the functions specific to userType;
        Compute visibility score:

$$V_{i,s} \leftarrow C_{i,s} \times H_{i,s} \times D_{i,s};$$

        Scale the visibility score by the ratio $\Delta S_s / S_{\text{final}}$ (where S_{final} is the total score in the final snapshot):

$$V_{i,s} \leftarrow V_{i,s} \times \frac{\Delta S_s}{S_{\text{final}}};$$

        Accumulate $V_{i,s}$ into the cumulative node impression for the comment;
      end
    end
  end
  foreach *userType in {PowerUser, InterestedUser, FastUser}* **do**
    Normalize the node impression for each comment by dividing by the total node impression for that sort and userType;
  end
end

Chapter 6

Evaluation

Experimental results reveal patterns in how sorting algorithms influence comment visibility and scoring distribution. Statistical analysis across different sorts and user types shows a highly right-skewed distribution where few comments receive disproportionately high attention. Percentile breakdowns quantify this visibility inequality, with upper percentiles showing dramatically higher node impressions compared to median values. Correlation testing between node impression and comment score reveals that the Q&A sort demonstrates the strongest relationship, followed by Best and Top sorts. Wilcoxon rank tests confirm the statistical significance of these findings, with low p-values for most sort methods. Visual analysis of comment occurrence patterns validates the methodology, showing expected behaviors like the predominance of older comments in Old sort and newer comments in New sort.

6.1 Statistics of Node Impression and Comment Score

We have analyzed the statistical measures of node impression and comment score across different sorts and user types. Tables A.1 to A.18 provide summary statistics including mean, median, minimum, and maximum values for each sort and user type, offering a view of the distribution of node impressions and comment scores. The most representative are Tables 6.1 to 6.3. The normalized values of 1.00 represent the maximum possible visibility or impact within the dataset, while 0.00 indicates complete absence of visibility or engagement, effectively standardizing the metrics for easier comparison across different user types and sorting algorithms. The table entries reveal disparities in node impression across user categories, with Power Users achieving a mean of 69.54 compared to Fast Users' 20.86, suggesting that user behavior

influences comment visibility in online discussions. Despite these differences in visibility metrics, all user types share identical score statistics (mean 3.54, median 1.00), with the median score of 1.00 reflecting Reddit's default value for comments that have received no votes, highlighting how the majority of comments receive little to no engagement from the community.

The summary statistics reveal that the mean node impression is significantly higher than the median, suggesting a right-skewed distribution with a few comments receiving disproportionately high attention. Similarly, the comment score exhibits a skewed distribution where most comments receive low scores while a select few accumulate high scores. The percentile analysis further confirms that only a small fraction of comments reach high visibility levels, as indicated by the steep increase from lower to upper percentiles. This trend is particularly evident in New, where the majority of comments appear in only a few snapshots. We have found that the highest values of node impression occur in the Q&A sort, as per, which indicates that some comments can accumulate the most prominence in this sorting algorithm. Because Q&A is the default sort on ChangeMyView, its influence on discussion dynamics is especially important.

Table 6.1: Stats for Power User and sort Q&A

Metric	Mean	Median	Min	Max
Node Impression	69.54	5.71	0.00	5033.66
Normalized Node Impression	0.55	0.60	0.00	1.00
Normalized Score	0.03	0.02	0.00	1.00
Score	3.54	1.00	1.00	640.00

Table 6.2: Stats for Interested User and sort Q&A

Metric	Mean	Median	Min	Max
Node Impression	46.03	3.54	0.00	3820.64
Normalized Node Impression	0.34	0.35	0.00	0.50
Normalized Score	0.03	0.02	0.00	1.00
Score	3.54	1.00	1.00	640.00

Table 6.3: Stats for Fast User and sort Q&A

Metric	Mean	Median	Min	Max
Node Impression	20.86	0.00	0.00	4981.79
Normalized Node Impression	0.10	0.00	0.00	0.50
Normalized Score	0.03	0.02	0.00	1.00
Score	3.54	1.00	1.00	640.00

6.2 Percentiles of Node Impressions

We have analyzed the percentiles of node impression to assess the distribution of comment visibility. The percentiles indicate the node impression value below which a given percentage of comments fall. They were calculated by sorting the node impression scores for each sorting method and extracting the corresponding threshold values. Tables 6.4 to 6.6 show that the lower percentiles are extremely small, while the upper percentiles (e.g., the 90th and 100th) are substantially larger. This skewed distribution indicates that only a few comments receive a high number of impressions, while the majority remain with minimal visibility. Such findings confirm the selective nature of comment prominence in the discussion. Notably, the Fast User exhibits extremely low values across all the sorts until the percentile of 60, which happens due to the strict weighting scheme of this user type.

Table 6.4: Percentiles of Node Impressions for Power User

Percentile	Q&A	Best	Controversial	Top	New	Old
10	0.08	0.23	0.4	0.22	0.34	0.13
20	0.37	1.25	1.804	0.86	1.518	2.12
30	1.18	3.838	5.8	2.642	4.764	6.61
40	2.824	9.052	11.31	5.746	9.29	10.5
50	5.71	18.68	22.95	10.63	19.41	21.03
60	9.62	33.106	35.532	16.26	35.964	30.92
70	15.57	46.138	50.58	21.826	60.29	52.6
80	29.278	79.734	73.21	38.75	103.176	82.38
90	62.092	184.492	152.81	90.056	176.116	187.68
100	5033.66	5950.68	6534.26	5759.26	4617.29	6215.42

Table 6.5: Percentiles of Node Impressions for Interested User

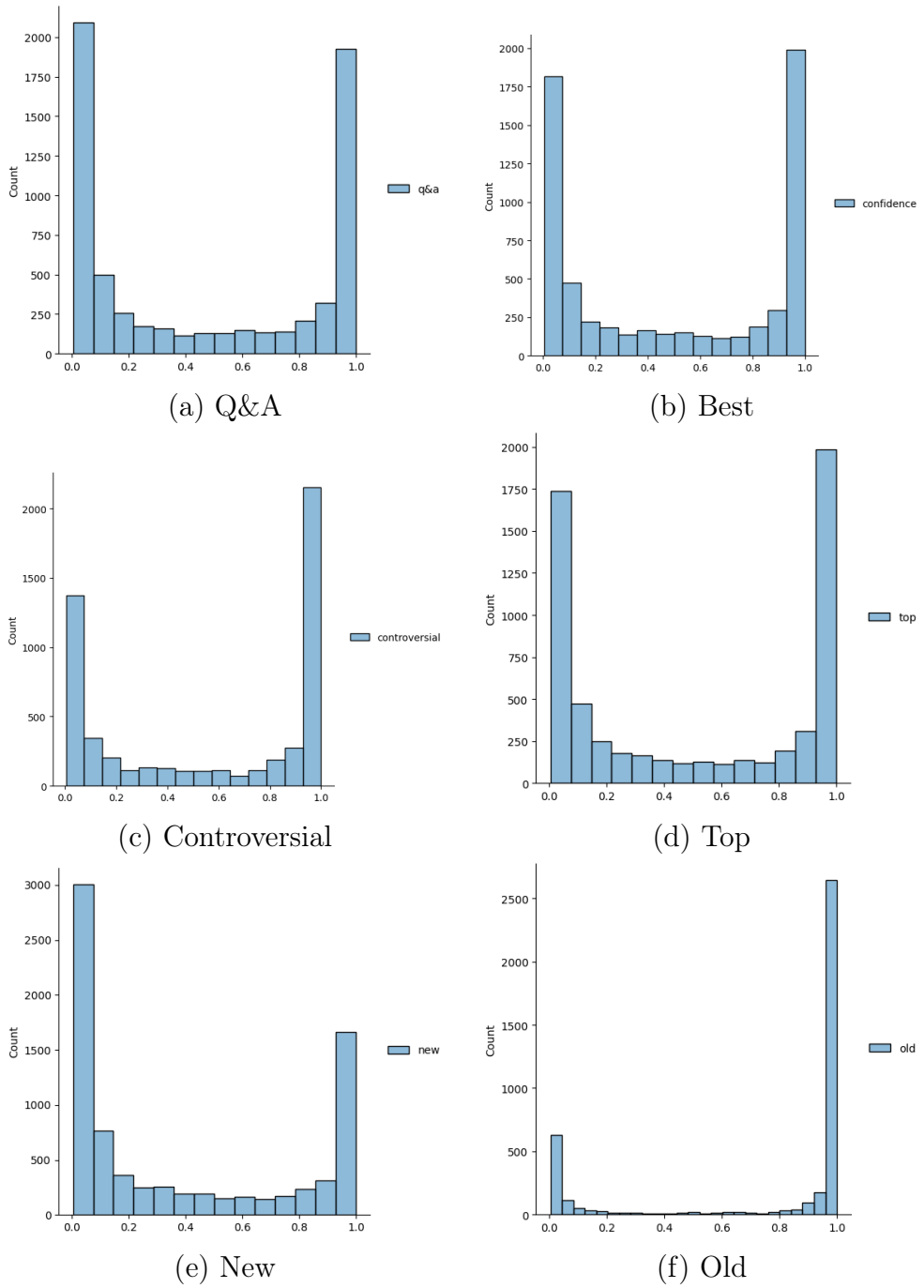
Percentile	Q&A	Best	Controversial	Top	New	Old
10	0.05	0.14	0.232	0.14	0.22	0.09
20	0.23	0.74	1.06	0.53	1.03	1.22
30	0.7	2.37	3.466	1.602	3.24	4.12
40	1.724	5.67	7.006	3.51	6.666	7.04
50	3.54	11.39	14.32	6.35	13.545	12.72
60	6.1	20.742	22.306	10.114	26.79	19.89
70	9.72	30.276	33.794	14.304	43.523	31.76
80	18.382	49.836	51.876	24.186	76.77	51.63
90	40.7	122.87	103.738	58.028	136.949	113.28
100	3820.64	4397.17	5227.37	4293.64	3693.81	4972.32

Table 6.6: Percentiles of Node Impressions for Fast User

Percentile	Q&A	Best	Controversial	Top	New	Old
10	0	0	0	0	0	0
20	0	0	0	0	0	0
30	0	0	0	0	0	0
40	0	0	0	0	0	0
50	0	0	0	0	0	0
60	0	0	0	0	0.21	0
70	0	0	0	0	5.553	0
80	0.828	5.88	6.876	1.32	27.172	3.11
90	11.284	42.394	46.286	17.912	89.649	26.8
100	4981.79	5883.32	6493.84	5682.53	4614.05	6204.85

6.3 Comment Occurrence

We examine the comment occurrence, computed by counting occurrence of comment snapshots in a submission and depicted by Figure 6.1. It is observed that many comments appear almost always while many others appear almost never. This pattern arises because we process only the first 25 comments in reading order from each snapshot, assuming that users will read only these comments. Thus, a node impression is calculated only when a comment is read. The diagrams serve as a valuable sanity check for the algorithm: for example, Old shows very few new comments, likely only due to older comments being deleted, whereas New displays many new comments. The other sort diagrams are nearly identical, revealing a consistent pattern of comments that are almost always visible and comments that are almost never visible. This could suggest that the ranking mechanism distinguishes between highly engaging comments and those that are rarely viewed, possibly reflecting inherent user engagement biases. Furthermore, the overall count of comments is higher in New, which is expected given that new comments appear in nearly every snapshot.



(a) Q&A (b) Best

(c) Controversial (d) Top

(e) New (f) Old

Figure 6.1: Comment occurrence for different sorts

6.4 Correlation of Node Impression and Comment Score

We compute the Pearson correlation coefficient (PCC) between node impression and comment score to examine the relationship between these metrics, as seen in Tables 6.7 to 6.9. The analysis shows a modest positive correlation, suggesting that comments with higher visual prominence tend to have higher scores. We observe that the Pearson correlation between node impression and comment score varies considerably depending on the sort method. In the New and Old sorts, the correlation is low because these sorts are not primarily based on comment score, which naturally leads to a weaker linear relationship. On the contrary, the Best and Top sorts exhibit higher correlation levels, as these methods explicitly incorporate comment score into the ranking algorithm. Notably, the Q&A sort shows an even larger correlation, which indicates a higher density of scores in threads where the OP is actively engaged. Such a pattern may imply that the discussions are dominated by a few high-scoring comments, potentially reducing the variety of opinions readily visible to readers. Furthermore, the strong influence of these prominent comments, as reflected by their vote counts, may affect the reader's opinion by turning their perspective toward the views expressed in these top comments.

Across all user types and sorting algorithms, the p-values are extremely low for the Q&A, Best, and Top sorts, providing evidence against the null hypothesis of no correlation, whereas the p-values for the New and Controversial sorts are comparatively higher, reflecting a less consistent relationship. Overall, these results confirm that the correlations between node impression and comment score are statistically significant, particularly in sorting methods that emphasize comment score. The correlation between node impression and comment score for the Controversial sort is not statistically significant. This weak correlation likely occurs because the Controversial sort specifically prioritizes comments with mixed voting patterns (similar numbers of upvotes and downvotes), meaning that comment score itself is not a strong predictor of visibility in this sorting method.

Table 6.7: Pearson correlation coefficient (PCC) and p-value between Node Impression and score for Power User

	Q&A	Best	Controversial	Top	New	Old
PCC	0.230	0.174	0.017	0.192	-0.028	0.052
p-value	7.43e-78	7.06e-43	0.223	3.89e-51	0.014	0.001

Table 6.8: Pearson correlation coefficient (PCC) and p-value between Node Impression and score for Interested User

	Q&A	Best	Controversial	Top	New	Old
PCC	0.259	0.203	0.021	0.216	-0.028	0.062
p-value	7.49e-99	7.80e-58	0.115	1.80e-64	0.0149	8.01e-05

Table 6.9: Pearson correlation coefficient (PCC) and p-value between Node Impression and score for Fast User

	Q&A	Best	Controversial	Top	New	Old
PCC	0.335	0.272	0.033	0.267	-0.014	0.089
p-value	1.61e-168	2.65e-104	0.015	5.62e-99	0.221	1.77e-08

6.5 Statistical Significance of Node Impression and Comment Score

The normalized score of a comment, which we use in this experiment for Wilcoxon signed-rank test, is calculated by dividing the absolute value of its score by the total score of all comments in the snapshot. If the total score is zero, a default value of 1 is used to prevent division by zero. The resulting value is rounded to two decimal places, providing a measure of the comment's relative contribution to the overall score in that snapshot. As an input for test, we define different selection thresholds, top_3, top_5, and top_{10}, which represent the groups of comments with the n highest and n lowest normalized scores based on their node impressions. For each sorting algorithm, comments are first ordered by their calculated node impression before selecting the top and bottom n comments, where n is 3, 5, or 10. We perform the test by comparing the normalized scores between the top and bottom n comments for each submission and then aggregating these results per user type and sorting algorithm.

All p-values reported in our analysis were rounded to two decimal places, as shown in Tables 6.10 to 6.12. In nearly all cases, the p-values are 0, indicating a very high level of statistical significance. For each user type and sort, most tests produced p-values well below the 0.05 threshold, confirming that the differences in node impression across sorts are unlikely to have arisen by chance. We chose a significance level of $\alpha = 0.05$ as it offers a standard balance between avoiding false positives and maintaining statistical power. Given that a p-value less than α indicates statistical significance, our results are strongly significant. In fact, since most of our p-values are very small, we could have selected an even smaller α value and still achieved significance. However, some exceptions were observed in New, for example, for Power Users the p-value is 0.35, and for Interested Users the p-values are 0.91 and 0.02 for different thresholds. These higher p-values suggest that, in these specific cases, the statistical evidence is less overwhelming. Overall, the significant p-values support our conclusions regarding the differences in comment prominence across the various sorting algorithms of comments on Reddit.

Table 6.10: Statistical significance (P-values) for Power User across different threshold settings

Sort	Top 3	Top 5	Top 10
Q&A	0	0	0
Best	0	0	0
Controversial	0	0	0
Top	0	0	0
New	0.35	0	0
Old	0	0	0

Table 6.11: Statistical significance (P-values) for Interested User across different threshold settings

Sort	Top 3	Top 5	Top 10
Q&A	0	0	0
Best	0	0	0
Controversial	0	0	0
Top	0	0	0
New	0.91	0.02	0
Old	0	0	0

Table 6.12: Statistical significance (P-values) for Fast User across different threshold settings

Sort	Top 3	Top 5	Top 10
Q&A	0	0	0
Best	0	0	0
Controversial	0	0	0
Top	0	0	0
New	0	0	0
Old	0	0	0

6.6 Distribution of Node Impressions

We have studied the distribution of node impressions for different user types and sorts. Notably, the diagrams are similar, with many comments exhibiting very little prominence. In Figure 6.2 a representative node impression distribution for the Power User (Q&A) is shown. For the complete set of similar figures, please refer to Figure A.1, Figure A.2 and Figure A.3 in section A.1.

Especially in New, most comments appear in only a few snapshots. Figure A.3 shows that fast users give particularly low values to comments, often 0, as they tend to disregard collapsed or deep comments. Both interested and power users assign nonzero impressions to a larger number of comments, but few comments achieve decisive prominence throughout the discussion's lifespan. This suggests that the top part of the comment section is dominated by only a few comments.

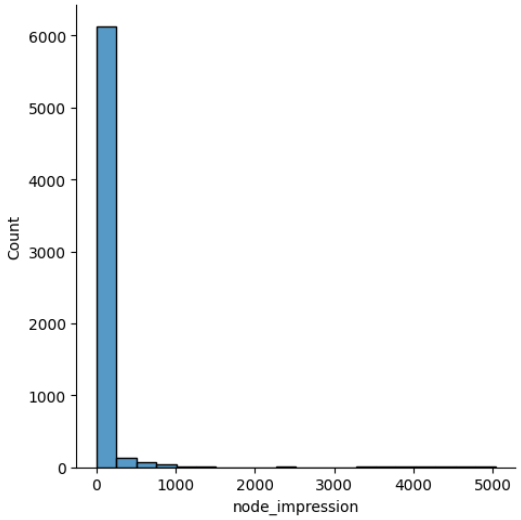


Figure 6.2: Node Impression Distribution for Power User (Q&A)

Chapter 7

Conclusion

This work examines the effect of Reddit's sorting algorithms on comment rankings and visibility within the ChangeMyView subreddit. Our analysis shows that the choice of sorting method influences comment prominence, as measured by node impression, and is reflected in comment scores.

Addressing the first research question on how different sorting algorithms affect visibility and ranking of comments, we found that sorting methods that incorporate comment score into their ranking, such as Best and Top, demonstrate a stronger correlation between visibility and scoring. Moreover, the Q&A sort, which involves active participation from the original poster, shows an even higher correlation, suggesting that the OP's engagement increases comment prominence and may shape reader opinions toward a narrower range of views. The Q&A algorithm, being the default sorting algorithm for ChangeMyView, influences users most strongly. This effect is consistent across all user types modeled in our study. Our findings suggest that many users likely rely on the Q&A algorithm without changing to alternative sorting methods, thereby experiencing a potentially narrower range of persuasive arguments. Our statistical significance testing using the Wilcoxon test established relevance for most user types and sorting algorithms, with Q&A, Best, and Top sorts showing consistently significant results across all user categories. The New sort, however, displayed a weaker statistical significance in some cases, particularly for Power Users and Interested Users, indicating a weaker statistical relationship between comment prominence and score for this sorting method. This pattern reinforces a feedback loop where already-popular comments gain even more visibility.

Regarding the second research question on the relationship between comment prominence and score across different sorts, we found that in contrast to Q&A, Best, and Top sorts, the New and Old sorts, which rely less on comment score, exhibit weaker correlations. These results highlight the influence

of algorithmic design on user engagement and the persuasive impact of online discussions.

For the third research question on how variations in user behavior influence comment prominence, our simulation of different user types (Power Users, Interested Users, and Fast Users) reveals that despite their different reading patterns, all are influenced by the sorting algorithm. Fast Users, who primarily focus on top-level comments, are particularly affected by the algorithm's prioritization, while even thorough Power Users show similar patterns of engagement with highly-scored content.

However, limitations such as processing only the first 25 comments and timing discrepancies between snapshots should be considered when interpreting these results. Future research should focus on adaptive comment selection mechanisms and further refinement of user behavior models to better capture the complexity of online interactions. An important extension would be to incorporate argument quality assessment frameworks, such as those proposed by Syed et al. [2023] and Wachsmuth et al. [2017], to evaluate whether higher-ranked comments genuinely contain more relevant or higher-quality arguments. This would help determine if the algorithmic bias we observed effectively surfaces valuable content or potentially suppresses worthy contributions based on comment score and relevance criteria. Additionally, studies with larger samples across different subreddits would help determine whether these patterns generalize beyond the ChangeMyView community. Overall, this study provides important insights into the relationship between algorithmic content curation and its influence on public opinion. The findings suggest that sorting algorithms may create echo chambers by amplifying certain perspectives, particularly in platforms designed for opinion change like ChangeMyView.

Chapter 8

Limitations

No Assessment of Comment Quality

Our study does not evaluate the quality or relevance of comments in relation to their ranking. We cannot determine whether the algorithmic bias we observed is beneficial or detrimental to discourse quality. It remains unclear whether higher-ranked comments contain more relevant or higher-quality arguments than lower-ranked ones. This presents an opportunity for future research to incorporate quality assessment metrics into our analysis framework. Syed et al. [2023] and Wachsmuth et al. [2017] offer potential methodologies for such analysis. Future work could explore correlations between comment scores and various relevance criteria to determine if sorting algorithms effectively surface higher-quality content or if valuable contributions are being overlooked.

Losses in Browser Snapshot Collection

Due to technical limitations, the number of snapshots per submission varies. Difficulties in gathering data using web scraping tools that mimic browser behavior have led to gaps in our dataset. Some submissions have only their first snapshot captured, for example, an hour after creation. In some cases, snapshots could not be captured due to rate limiting and other network issues. Submissions missing snapshots for one or more sorting algorithms are deemed invalid for our research and are filtered out.

The number of snapshots for a submission may differ between sorting algorithms. To obtain more accurate results, the algorithm assumes that the number of snapshots is identical across all sorting algorithms. To address this inconsistency, we implemented an interpolation function. Due to Reddit's rate limiting, browser loading times, and other factors, there may be a difference of up to one second between snapshots for the six sorting algorithms. Ideally, a

snapshot for each sort would be taken simultaneously. Since this was not possible, we fill the snapshot list for a sorting algorithm with the latest available snapshot of the submission for that sort. However, these data losses might slightly affect the accuracy of our results.

Lack of Prior Research on Reddit User Behavior

There is little research on how Reddit users browse and interact with comments. This gap makes it hard to verify our assumptions about user behavior. To overcome this, we model user behavior using simple assumptions. We define user types such as Power Users, Interested Users, and Fast Users. These models use factors like whether a comment is collapsed, its content height, and its depth in the discussion. This approach allows us to estimate comment visibility and ranking despite the lack of direct user data. However, it could be misleading if it fails to capture the true dynamics of user behavior.

We process only the first 25 comments in reading order from each snapshot for users to read. However, it remains unclear whether 25 is an optimal value because the comment section's tree structure can be either deeply nested or relatively flat, leading to different outputs depending on the comment limit. It is therefore essential to establish an effective mechanism for limiting the number of inputted comments.

Limited Data Gathering Period

We collected data for only one week, due to the intensive nature of capturing snapshots at 5-minute intervals for many submissions simultaneously. This limited timeframe may not capture longer-term trends or seasonal variations in user behavior and content dynamics. Future work should consider extending the data collection period to obtain a more comprehensive view of the observed phenomena.

Focus on ChangeMyView

This study concentrates exclusively on the ChangeMyView subreddit. This focus may limit the generalizability of the findings to the broader Reddit community or other social media platforms. Future studies should consider including additional subreddits to better understand the impact of sorting algorithms across diverse online communities.

Bibliography

David Atkinson, Kumar Bhargav Srinivasan, and Chenhao Tan. What gets echoed? understanding the "pointers" in explanations of persuasive arguments. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing, EMNLP-IJCNLP 2019, Hong Kong, China, November 3-7, 2019*, pages 2911–2921. Association for Computational Linguistics, 2019. doi: 10.18653/V1/D19-1289. URL `https://doi.org/10.18653/v1/D19-1289`.

Ryo Egawa, Gaku Morio, and Katsuhide Fujita. Annotating and analyzing semantic role of elementary units and relations in online persuasive arguments. In Fernando Alva-Manchego, Eunsol Choi, and Daniel Khashabi, editors, *Proceedings of the 57th Conference of the Association for Computational Linguistics, ACL 2019, Florence, Italy, July 28 - August 2, 2019, Volume 2: Student Research Workshop*, pages 422–428. Association for Computational Linguistics, 2019. doi: 10.18653/V1/P19-2059. URL `https://doi.org/10.18653/v1/p19-2059`.

Zhen Guo, Zhe Zhang, and Munindar P. Singh. In opinion holders' shoes: Modeling cumulative influence for view change in online argumentation. In Yennun Huang, Irwin King, Tie-Yan Liu, and Maarten van Steen, editors, *WWW '20: The Web Conference 2020, Taipei, Taiwan, April 20-24, 2020*, pages 2388–2399. ACM / IW3C2, 2020. doi: 10.1145/3366423.3380302. URL `https://doi.org/10.1145/3366423.3380302`.

Christopher Hidey, Elena Musi, Alyssa Hwang, Smaranda Muresan, and Kathy McKeown. Analyzing the semantic types of claims and premises in an online persuasive forum. In Ivan Habernal, Iryna Gurevych, Kevin D. Ashley, Claire Cardie, Nancy L. Green, Diane J. Litman, Georgios Petasis, Chris Reed, Noam Slonim, and Vern R. Walker, editors, *Proceedings of the 4th Workshop on Argument Mining, ArgMining@EMNLP 2017, Copenhagen, Denmark, September 8, 2017*, pages 11–21. Association for Computational

Linguistics, 2017. doi: 10.18653/V1/W17-5102. URL https://doi.org/10.18653/v1/w17-5102.

Aaron Jaech, Victoria Zayats, Hao Fang, Mari Ostendorf, and Hannaneh Hajishirzi. Talking to the crowd: What do people react to in online discussions? In Lluís Màrquez, Chris Callison-Burch, and Jian Su, editors, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 2026–2031, Lisbon, Portugal, September 2015. Association for Computational Linguistics. doi: 10.18653/v1/D15-1239. URL https://aclanthology.org/D15-1239.

Vishal Khanna. Analyzing debaters' persuasiveness on change my view. Master's thesis, Bauhaus-Universität Weimar, 2021. Master's Thesis.

Khalid Al Khatib, Michael Völske, Shahbaz Syed, Nikolay Kolyada, and Benno Stein. Exploiting personal characteristics of debaters for predicting persuasiveness. In Dan Jurafsky, Joyce Chai, Natalie Schluter, and Joel R. Tetreault, editors, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7067–7072. Association for Computational Linguistics, 2020. doi: 10.18653/V1/2020.ACL-MAIN.632. URL https://doi.org/10.18653/v1/2020.acl-main.632.

Jialu Li, Esin Durmus, and Claire Cardie. Exploring the role of argument structure in online debate persuasion. In Bonnie Webber, Trevor Cohn, Yulan He, and Yang Liu, editors, *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing, EMNLP 2020, Online, November 16-20, 2020*, pages 8905–8912. Association for Computational Linguistics, 2020. doi: 10.18653/V1/2020.EMNLP-MAIN.716. URL https://doi.org/10.18653/v1/2020.emnlp-main.716.

Alexey N. Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. The anatomy of reddit: An overview of academic research. *CoRR*, abs/1810.10881, 2018. URL http://arxiv.org/abs/1810.10881.

Humphrey Mensah, Lu Xiao, and Sucheta Soundarajan. Characterizing susceptible users on reddit's changemyview. In *Proceedings of the 10th International Conference on Social Media and Society, SMSociety 2019, Toronto, ON, Canada, July 19-21, 2019*, pages 102–107. ACM, 2019. doi: 10.1145/3328529.3328550. URL https://doi.org/10.1145/3328529.3328550.

Nailia Mirzakhmedova, Johannes Kiesel, Khalid Al Khatib, and Benno Stein. Unveiling the power of argument arrangement in online persuasive discussions. In Houda Bouamor, Juan Pino, and Kalika Bali, editors, *Findings*

of the Association for Computational Linguistics: EMNLP 2023, Singapore, December 6-10, 2023, pages 15659–15671. Association for Computational Linguistics, 2023. doi: 10.18653/V1/2023.FINDINGS-EMNLP.1048. URL https://doi.org/10.18653/v1/2023.findings-emnlp.1048.

Arturo Martínez Peguero and Taro Watanabe. Change my frame: Reframing in the wild in r/changemyview, 2024. URL https://arxiv.org/abs/2407.02637.

Shahbaz Syed, Timon Ziegenbein, Philipp Heinisch, Henning Wachsmuth, and Martin Potthast. Frame-oriented summarization of argumentative discussions. In David Schlangen, Svetlana Stoyanchev, Shafiq Joty, Ondrej Dusek, Casey Kennington, and Malihe Alikhani, editors, *Proceedings of the 24th Meeting of the Special Interest Group on Discourse and Dialogue, SIGDIAL 2023, Prague, Czechia, September 11 - 15, 2023*, pages 114–129. Association for Computational Linguistics, 2023. doi: 10.18653/V1/2023.SIGDIAL-1.10. URL https://doi.org/10.18653/v1/2023.sigdial-1.10.

Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In Jacqueline Bourdeau, Jim Hendler, Roger Nkambou, Ian Horrocks, and Ben Y. Zhao, editors, *Proceedings of the 25th International Conference on World Wide Web, WWW 2016, Montreal, Canada, April 11 - 15, 2016*, pages 613–624. ACM, 2016. doi: 10.1145/2872427.2883081. URL https://doi.org/10.1145/2872427.2883081.

Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. Computational argumentation quality assessment in natural language. In Mirella Lapata, Phil Blunsom, and Alexander Koller, editors, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2017, Valencia, Spain, April 3-7, 2017, Volume 1: Long Papers*, pages 176–187. Association for Computational Linguistics, 2017. doi: 10.18653/V1/E17-1017. URL https://doi.org/10.18653/v1/e17-1017.

Zhongyu Wei, Yang Liu, and Yi Li. Is this post persuasive? ranking argumentative comments in online forum. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 2: Short Papers*. The Association for Computer Linguistics, 2016. doi: 10.18653/V1/P16-2032. URL https://doi.org/10.18653/v1/p16-2032.

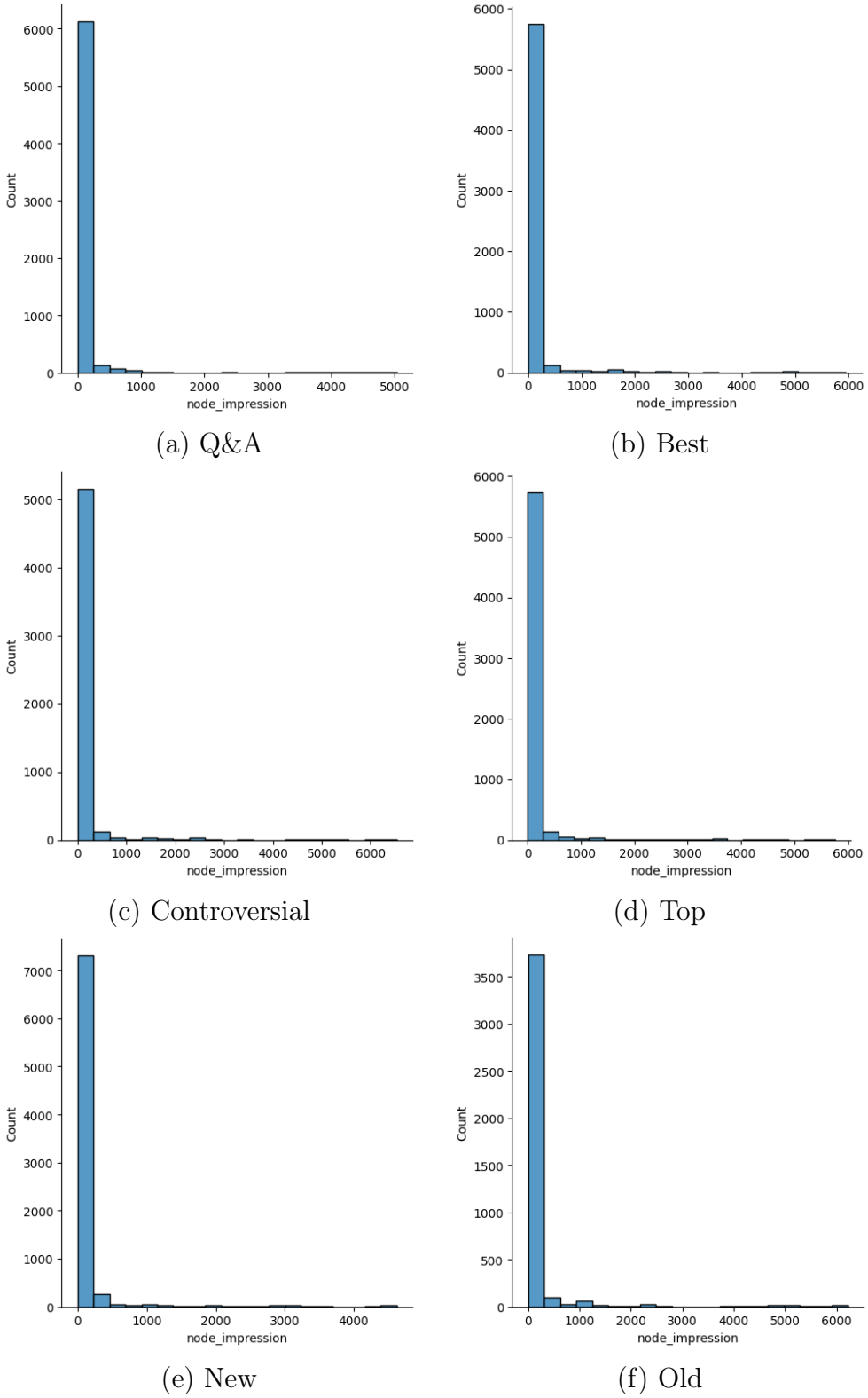
Matti Wiegmann, Khalid Al Khatib, Vishal Khanna, and Benno Stein. Analyzing persuasion strategies of debaters on social media. In Nicoletta Calzolari,

Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics, COLING 2022, Gyeongju, Republic of Korea, October 12-17, 2022*, pages 6897–6905. International Committee on Computational Linguistics, 2022. URL `https://aclanthology.org/2022.coling-1.600`.

Appendix A

Supplementary Figures and Tables

A.1 Node Impression Distributions

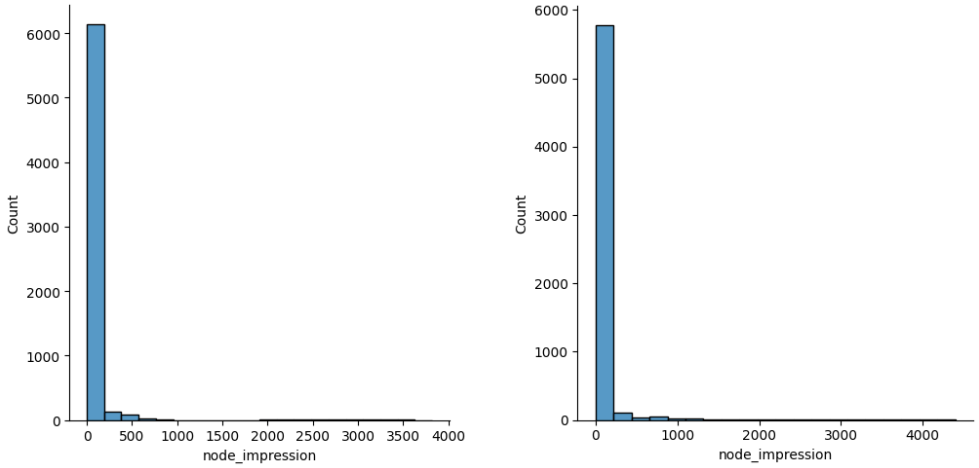


(a) Q&A (b) Best

(c) Controversial (d) Top

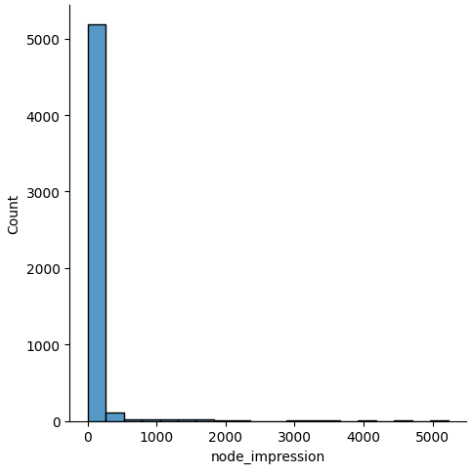
(e) New (f) Old

Figure A.1: Node Impression Distribution for Power User

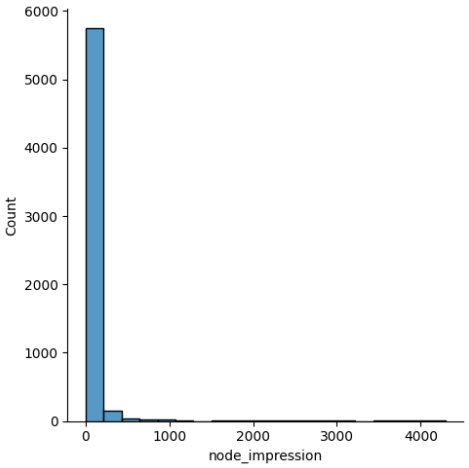


(a) Q&A

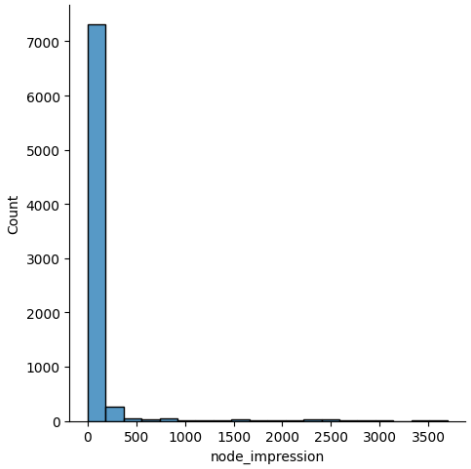
(b) Best



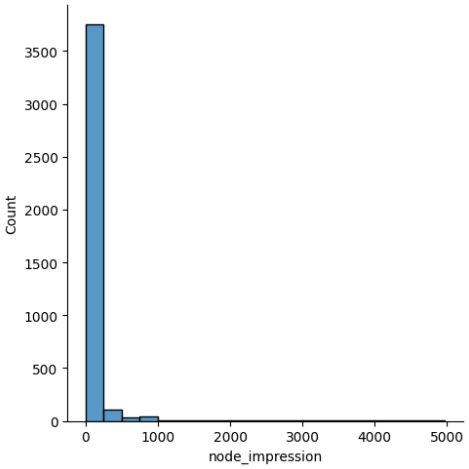
(c) Controversial



(d) Top

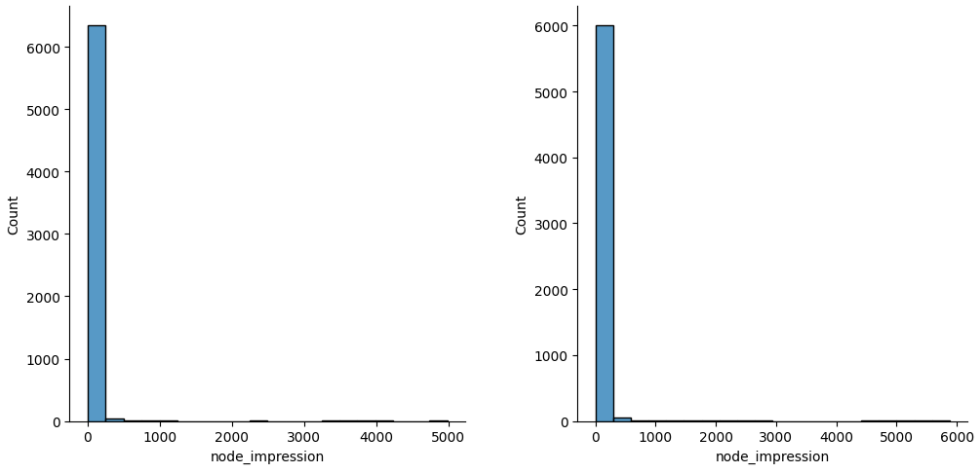


(e) New



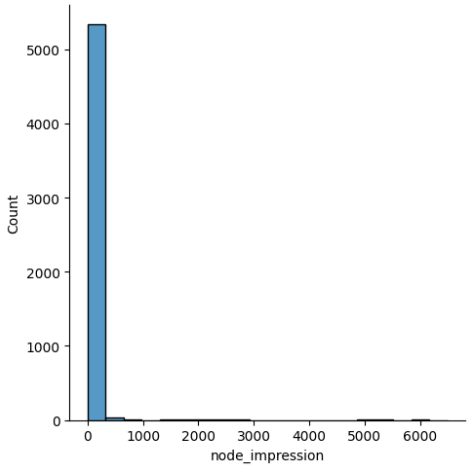
(f) Old

Figure A.2: Node Impression Distribution for Interested User

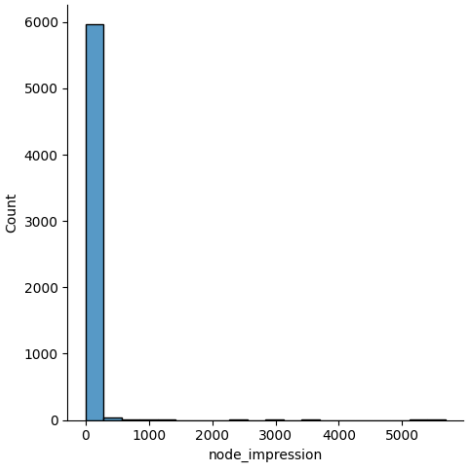


(a) Q&A

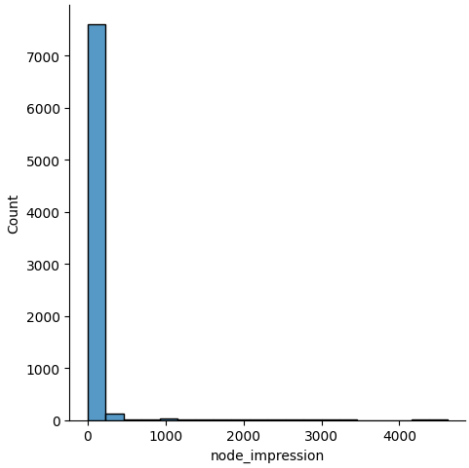
(b) Best



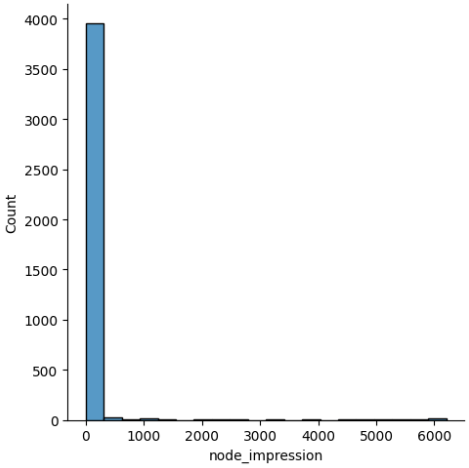
(c) Controversial



(d) Top



(e) New



(f) Old

Figure A.3: Node Impression Distribution for Fast User

A.2 Statistics of Node Impression and Comment Score

Table A.1: Stats for Power User and sort Q&A

Metric	Mean	Median	Min	Max
Node Impression	69.54	5.71	0.00	5033.66
Normalized Node Impression	0.55	0.60	0.00	1.00
Normalized Score	0.03	0.02	0.00	1.00
Score	3.54	1.00	1.00	640.00

Table A.2: Stats for Power User and sort Best

Metric	Mean	Median	Min	Max
Node Impression	143.37	18.68	0.00	5950.68
Normalized Node Impression	0.55	0.60	0.00	1.00
Normalized Score	0.04	0.02	0.00	1.00
Score	3.80	1.00	1.00	640.00

Table A.3: Stats for Power User and sort Controversial

Metric	Mean	Median	Min	Max
Node Impression	113.83	22.95	0.00	6534.26
Normalized Node Impression	0.55	0.60	0.00	1.00
Normalized Score	0.04	0.02	0.00	1.00
Score	2.57	1.00	1.00	634.00

Table A.4: Stats for Power User and sort Top

Metric	Mean	Median	Min	Max
Node Impression	92.80	10.63	0.00	5759.26
Normalized Node Impression	0.56	0.60	0.00	1.00
Normalized Score	0.04	0.02	0.00	1.00
Score	3.82	1.00	1.00	633.00

Table A.5: Stats for Power User and sort New

Metric	Mean	Median	Min	Max
Node Impression	121.92	19.41	0.00	4617.29
Normalized Node Impression	0.50	0.56	0.00	1.00
Normalized Score	0.04	0.03	0.01	1.00
Score	1.33	1.00	1.00	52.00

Table A.6: Stats for Power User and sort Old

Metric	Mean	Median	Min	Max
Node Impression	182.54	21.03	0.00	6215.42
Normalized Node Impression	0.55	0.60	0.00	1.00
Normalized Score	0.05	0.03	0.00	1.00
Score	3.26	1.00	1.00	455.00

Table A.7: Stats for Interested User and sort Q&A

Metric	Mean	Median	Min	Max
Node Impression	46.03	3.54	0.00	3820.64
Normalized Node Impression	0.34	0.35	0.00	0.50
Normalized Score	0.03	0.02	0.00	1.00
Score	3.54	1.00	1.00	640.00

Table A.8: Stats for Interested User and sort Best

Metric	Mean	Median	Min	Max
Node Impression	94.39	11.39	0.00	4397.17
Normalized Node Impression	0.34	0.35	0.00	0.50
Normalized Score	0.04	0.02	0.00	1.00
Score	3.80	1.00	1.00	640.00

Table A.9: Stats for Interested User and sort Controversial

Metric	Mean	Median	Min	Max
Node Impression	76.57	14.32	0.00	5227.37
Normalized Node Impression	0.34	0.35	0.00	0.50
Normalized Score	0.04	0.02	0.00	1.00
Score	2.57	1.00	1.00	634.00

Table A.10: Stats for Interested User and sort Top

Metric	Mean	Median	Min	Max
Node Impression	60.24	6.35	0.00	4293.64
Normalized Node Impression	0.34	0.35	0.00	0.50
Normalized Score	0.04	0.02	0.00	1.00
Score	3.82	1.00	1.00	633.00

Table A.11: Stats for Interested User and sort New

Metric	Mean	Median	Min	Max
Node Impression	91.07	13.55	0.00	3693.81
Normalized Node Impression	0.35	0.35	0.00	0.50
Normalized Score	0.04	0.03	0.01	1.00
Score	1.33	1.00	1.00	52.00

Table A.12: Stats for Interested User and sort Old

Metric	Mean	Median	Min	Max
Node Impression	116.34	12.72	0.00	4972.32
Normalized Node Impression	0.34	0.35	0.00	0.50
Normalized Score	0.05	0.03	0.00	1.00
Score	3.26	1.00	1.00	455.00

Table A.13: Stats for Fast User and sort Q&A

Metric	Mean	Median	Min	Max
Node Impression	20.86	0.00	0.00	4981.79
Normalized Node Impression	0.10	0.00	0.00	0.50
Normalized Score	0.03	0.02	0.00	1.00
Score	3.54	1.00	1.00	640.00

Table A.14: Stats for Fast User and sort Best

Metric	Mean	Median	Min	Max
Node Impression	46.23	0.00	0.00	5883.32
Normalized Node Impression	0.10	0.00	0.00	0.50
Normalized Score	0.04	0.02	0.00	1.00
Score	3.80	1.00	1.00	640.00

Table A.15: Stats for Fast User and sort Controversial

Metric	Mean	Median	Min	Max
Node Impression	38.43	0.00	0.00	6493.84
Normalized Node Impression	0.10	0.00	0.00	0.43
Normalized Score	0.04	0.02	0.00	1.00
Score	2.57	1.00	1.00	634.00

Table A.16: Stats for Fast User and sort Top

Metric	Mean	Median	Min	Max
Node Impression	25.72	0.00	0.00	5682.53
Normalized Node Impression	0.09	0.00	0.00	0.50
Normalized Score	0.04	0.02	0.00	1.00
Score	3.82	1.00	1.00	633.00

Table A.17: Stats for Fast User and sort New

Metric	Mean	Median	Min	Max
Node Impression	53.53	0.00	0.00	4614.05
Normalized Node Impression	0.15	0.00	0.00	0.50
Normalized Score	0.04	0.03	0.01	1.00
Score	1.33	1.00	1.00	52.00

Table A.18: Stats for Fast User and sort Old

Metric	Mean	Median	Min	Max
Node Impression	50.29	0.00	0.00	6204.85
Normalized Node Impression	0.10	0.00	0.00	0.50
Normalized Score	0.05	0.03	0.00	1.00
Score	3.26	1.00	1.00	455.00

Appendix B

Data Collection

To address our research questions, we collected data from Reddit. In this chapter, we describe the methodology used for data collection and discuss the challenges encountered during the process.

B.1 Reddit API

Reddit offers an API that allows users to retrieve data from its servers in a structured and controlled manner. To obtain the necessary credentials for API access, our script was registered under our developer profile on Reddit. After familiarizing ourselves with Reddit's API[1] and its rate limiting guidelines, we sought a library to simplify data querying. We found the Python Reddit API Wrapper (PRAW)[2] to be sufficiently customizable for our needs. This library provides classes for various Reddit entities (such as submissions, comments, and subreddits) and abstracts low-level queries by offering convenient class methods to interact with the data. In addition, PRAW automatically manages Reddit's rate limits by pacing API calls. Since the API is not the bottleneck in our application, the default limit of 600 calls per 10 minutes has proven to be adequate.

B.1.1 API Consumption Script

The script begins with an OAuth2 authorization using our Reddit developer credentials. Once authorized, it enters a loop to query data. First, the script checks for submissions that are already tracked by their submission IDs and then searches for new posts on ChangeMyView that are not yet tracked. For

[1]Reddit API documentation: `https://www.reddit.com/dev/api/`
[2]PRAW documentation: `https://praw.readthedocs.io/en/latest/`

both tracked and new submissions, the script takes a snapshot by querying the full state of each submission once per comment sort (Best, Q&A, Top, New, Old, Controversial).

Each snapshot includes data for the submission, its author, and all associated comments along with their authors. The collected data covers the submission text, title, creation timestamp, score (including upvotes and downvotes), author username and flair (which displays the delta score on CMV), status (approved, deleted, or banned), awards, and references to replies. Snapshots are stored as individual lines in a JSON Lines file named after the submission ID; each line corresponds to a snapshot for one of the selected comment sorts along with its creation timestamp. This process repeats every 5 minutes to track changes, primarily in deltas and scores. After 48 hours from the submission's creation, the script stops taking snapshots for it.

After 48 hours from the submission's creation, the script stops taking snapshots and compresses the corresponding JSON Lines file to save space and accelerate file transfer. Using 7z compression, the files are reduced to up to one-tenth of their original size due to the repetitive nature of parts of the data. The script is deployed in a pod on the Kubernetes server of the webis group. A Persistent Volume Claim is assigned to the pod to ensure sufficient disk space for the script's high disk write activity without disrupting other users.

B.1.2 Asynchronous API Consumption

To minimize discrepancies between snapshots taken with different sorting algorithms, it was crucial to reduce the time interval between API calls. Since PRAW is normally synchronous, its default behavior is to process API calls sequentially, which can introduce inconsistencies if even a second of delay occurs. It is important, because snapshots for all six sorting methods must be captured concurrently for each submission. To overcome this limitation, we employed an asynchronous version[3] of PRAW that enables multiple API calls to be executed simultaneously. A test confirmed that using this asynchronous approach, snapshots taken at the same timestamp across different sorting algorithms yielded identical score values.

B.1.3 Hidden Score

One challenge that required us to adapt our approach and re-collect the data was that Reddit conceals the score of comments during the first 24 hours after they are posted. Since the comment score is a crucial metric for our analysis, it was essential to obtain these scores without delay. We discovered that scores

[3]Async PRAW documentation: `https://asyncpraw.readthedocs.io/en/stable/`

are visible to subreddit moderators. Therefore, we contacted the ChangeMyView admin team to request temporary moderator rights. Upon receiving approval, we updated our developer credentials with the new moderator account, which enabled us to access the hidden scores in the API responses.

B.2 Snapshots

During data collection, we observed that the order of comments provided by the Reddit API differed from the order presented in the browser on Reddit.com. Since the display order of comments is crucial for analyzing sorting algorithms, relying solely on API responses was insufficient. To obtain the correct comment order, we scraped HTML snapshots immediately after retrieving the corresponding JSON API responses. We then executed a script to extract the comment order and visibility data from the HTML files. Additionally, we captured PNG screenshots to manually verify the accuracy of the extracted order.

B.2.1 Framework

Our initial approach used the Selenium[4] framework. However, Selenium's synchronous nature made it difficult to integrate into our asynchronous application. Moreover, managing multiple user profiles required running several Selenium instances concurrently. We migrated to Playwright[5] because it resolved both issues. Playwright's asynchronous interface allowed us to open a submission page immediately after sending an API request without blocking other operations. It also supports browser contexts, which enabled us to manage multiple Reddit user accounts with different cookies within a single driver instance. This improvement increased the efficiency of opening, scraping, and closing web pages, allowing us to handle more submissions within the 5-minute window.

B.3 Data Storage and Cleaning

API calls to Reddit return JSON data representing the state of a submission at a specific moment. Multiple states are grouped by submission ID into a single JSONL file. Due to the complexity of Reddit's API responses, which include over 100 fields and nested objects (many of which are optional), we

[4]Selenium documentation: `https://www.selenium.dev/documentation/`
[5]Playwright documentation: `https://playwright.dev/python/docs/intro`

opted for MongoDB[6] rather than SQL. Each JSON object from the JSONL file was imported separately into a MongoDB collection as a snapshot. When snapshots contained too many comments and exceeded the recommended document size, each comment was imported as a separate document in another collection and linked to its corresponding snapshot. Optimized indices were created to efficiently query a database comprising approximately 420,000 snapshots and 65,000,000 documents of comment snapshots. We then extended the MongoDB documents with data extracted from HTML snapshots. For each browser snapshot, we extracted the reading order and visibility data of comments and appended this information to the corresponding snapshot document. However, obtaining browser snapshots proved less reliable than retrieving API responses. Some API calls failed to capture data from the browser, and these instances were marked in the JSON for proper handling during evaluation.

VDP	**Sub-Nbrs.**
0–10	1
10–20	2
20–30	3
30–40	4
40–50	3
50–60	0
60–70	1
70–80	1
80–90	14
90–100	78

Table B.1: Distribution of Submissions by Percentage of Visibility Data Presence (VDP). VDP is defined as the percentage of snapshots that contain visibility data. "Sub-Nbrs." denotes the number of submissions in each category.

Some submissions in the dataset had too many missing snapshots, making them unsuitable for further analysis. These large gaps in the data would hinder reliable interpolation. Based on the data in Table B.1, submissions with less than 50% visibility data presence were excluded from further analysis. Visibility data consists of the reading order of comments and the collapsed state, which are extracted from the browser snapshots. Submissions that entirely lacked snapshots for one of the sorting algorithms were also considered invalid. Such submissions were moved to a separate collection of invalid snapshots and were excluded from evaluation.

[6]MongoDB documentation: `https://www.mongodb.com/docs/`