

National University of Kyiv-Mohyla Academy

Computational Social Science

Report

The COVID-19 Effect: From Panic to Burnout

A Behavioral Research of Mental Health Communities on
Reddit during the COVID-19 Pandemic

Teacher:

Andrew Kurochkin

Prepared by:

Nykyta Babakov

December 2025

Kyiv - 2025

Content

1. Executive Summary.....	3
1.1. The dataset for analysis.....	3
1.2. Main research questions.....	3
1.3. Key findings.....	3
1.4. Why these findings matter.....	4
1.5. Dataset Description and Methodology.....	4
1.6. Limitations of the dataset and approach:.....	5
2. Analysis and Key Findings.....	6
2.1. How has user activity changed over time?.....	6
2.2. Which subreddits are the most active in our dataset?.....	6
2.3. Who are the most active authors (Top Contributors)?.....	7
2.4. Which posts resonate the most (Outliers)?.....	8
2.5. How does text length vary depending on the subreddit?.....	9
2.6. "I" vs. "You" (Psychological Analysis).....	10
2.7. What dates in 2020 marked Reddit's phase transition from discussion to collective panic?.....	11
2.8. Are there differences in emotional structure between mental support groups and news communities during a crisis?.....	12
2.9. Why do subjective emotions and personal interpretations attract more attention than objective facts in times of crisis?.....	13
2.10. How have the levels of subjectivity and emotional tone of communities changed over time, and does this support the theory of "community burnout"?..	14
3. Conclusions.....	15
3.1. Future Work.....	15
4. Link to GitHub.....	17

1. Executive Summary

1.1. The dataset for analysis

This report analyzes over 4 million Reddit posts from mental health and COVID-19 specific communities to understand how the onset of the pandemic altered online social behavior.

The data was collected via PushshiftDumps scripts using an algorithm that involves extracting .zst files from PushshiftDumps, filtering them by topic or subreddit, and then exporting the results to CSV format.

By utilizing Natural Language Processing (TextBlob) and temporal data analysis, the research identified a distinct shift in user psychology from personal sharing to information consumption.

1.2. Main research questions

- How have the dynamics of the global COVID-19 pandemic affected the transformation of discourse from subjective self-expression to objective information consumption?
- To what extent did the emotional background of news communities (r/Coronavirus) converge with the indicators of specialized mental health support groups during the peak of the crisis?
- What factors caused the significant disparity in user engagement (Score) between informational content and emotional support posts?

1.3. Key findings

The "Shock" Phase: March 2020 saw a massive structural break in activity correlated with global lockdowns.

Statistical confirmation: Mathematical analysis of anomalies found that activity in March broke the "ceiling" of normality (mean + 2 standard deviations), especially on the critical dates of March 12-13, when global restrictions were imposed. This is not just a seasonal increase, but a real statistical shock.

The "Burnout" Phase: As the crisis progressed, linguistic subjectivity declined as users switched from personal stories to sharing objective news.

During February and March, the Subjectivity Level dropped sharply, as shown in the time evolution of sentiment graph (Figure 7). Users stopped focusing on internal emotions and switched to "dumscrolling" and discussing external facts: quarantine rules, infection statistics, and news.

Uniform Distress: All mental health subreddits displayed a nearly identical negative emotional baseline, regardless of their specific focus.

Even the r/Coronavirus news subreddit showed the same negative emotional impact as depression or anxiety communities. The negative news became so festering that it was indistinguishable from clinical descriptions of mental disorders.

The Empathy Gap: The Reddit ecosystem heavily favored short news content (r/Coronavirus) over detailed personal narratives (r/SuicideWatch). The average rating (Score) for r/Coronavirus was 15.4, which is several times higher than r/SuicideWatch (3.2) or r/mentalhealth (2.9).

1.4. Why these findings matter

- Social media is a real-time recorder of global crises. This is important because it allows analysts to use social media activity as an early warning system for the level of public panic or mobilization.
- The shift from personal stories to objective news is indicative of “community burnout.” Understanding this transition helps create better communication strategies as people stop “feeling” and start only “reacting” to facts.
- Negative news has become indistinguishable from clinical descriptions of depression or anxiety. This is important for understanding how media content can create a “baseline” of negativity that affects all users regardless of their initial state.

1.5. Dataset Description and Methodology

Data source and specifications

Total: 4,184,400 entries.

Size: 1,4 GB (CSV).

Period: January-April 2020.

Content Sources (Subreddits): Data aggregated from two key community types:

- Mental health communities: r/Anxiety, r/Depression, r/MentalHealth, r/SuicideWatch.
- Specialized COVID communities: r/Coronavirus, r/COVID19_support

Data structure: Each raw entry contains 7 main information fields: Score (numerical rating), Date (timestamp), Subreddit, Title (title), Author, Link, and Text/Body (main text).

Memory size: The dataset takes up over 223.5 MB.

Link to raw dataset:

https://ukmaedu-my.sharepoint.com/:f/g/personal/n_babakov_ukma_edu_ua/lgBWaVlnrM-0RoazLuPGI_mvAYapXxYEiSTxo_mD5ObCGWA?e=befUxu

Data Processing

Gap Filling: 3,888,217 missing values were found in the Title field and 3,271 in Text/Body. All empty values were replaced with empty strings using `.fillna("")`.

Deleted Content Handling: 471,253 posts marked with [removed] or [deleted] were identified. These markers were removed from the texts so that they would not affect further linguistic analysis.

Duplicates: Checking for duplicates showed that they were absent in the provided dataset.

Data from the CSV file was read with the required parameter `parse_dates=["Date"]`, since the string date format is unsuitable for time analysis.

Analytical Methods

A standard library stack was used for the work: pandas for data manipulation, numpy for mathematical operations, and matplotlib and seaborn for visualization.

Feature Engineering

A number of new metrics were created for deeper analysis:

- **full_text**: Combining the title (Title) and the body of the post (Text/Body) into one column for full NLP analysis.
- **text_length**: Calculation of the length of the message in characters.

NLP Sentiment Analysis

The key step was text processing using the TextBlob library. Two indices were calculated for each post:

Polarity: Values from -1 (negative/pain) to +1 (positive/hope).

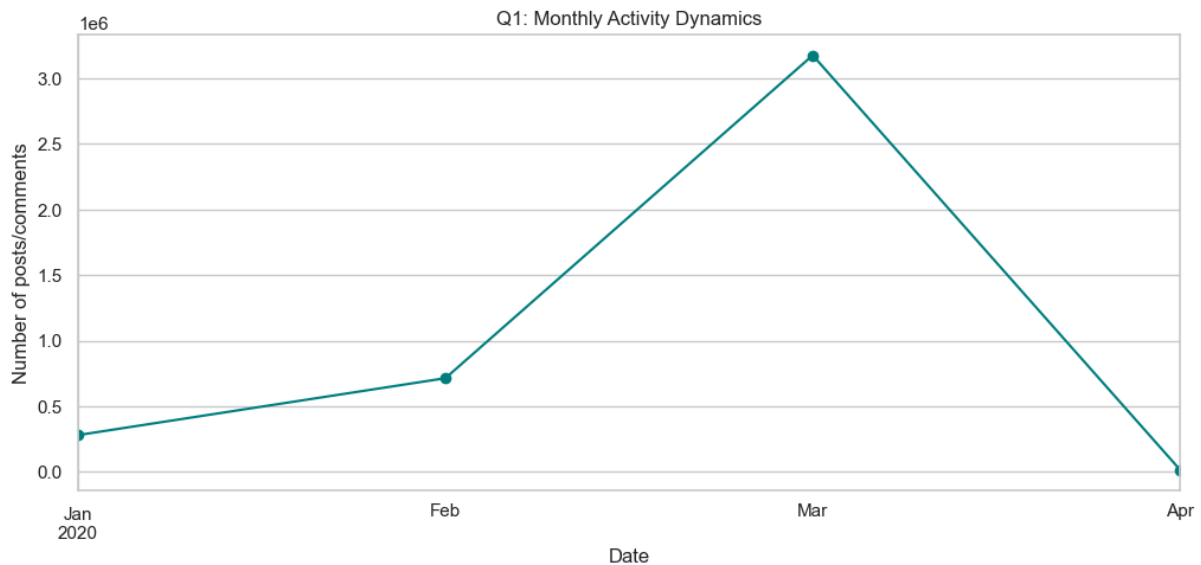
Subjectivity: Values from 0 (facts/objectivity) to 1 (emotions/personal opinion). Based on polarity, posts were automatically classified by mood into positive, negative, and neutral

1.6. Limitations of the dataset and approach:

- **Time limitations**: The analysis is limited only to the initial stage of the pandemic (January-April 2020), which does not allow assessing the long-term consequences of the crisis.
- **Platform bias**: The sample is based solely on Reddit users, who are not representative of the entire world population by age, geography, or social status.
- **Bots**: The second most active is AutoModerator (76,958 entries), whose messages are automated and can distort community activity statistics.

2. Analysis and Key Findings

2.1. How has user activity changed over time?

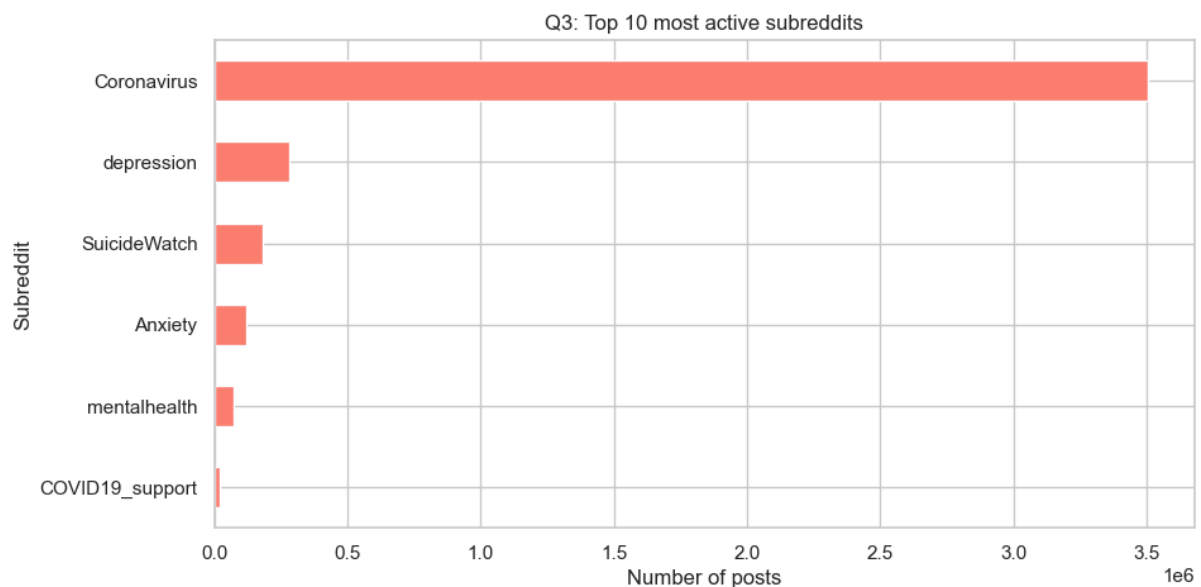


Activity dynamics: After a stable start to the year, a sharp peak was recorded in March, when the number of posts exceeded 3 million. In April, a sharp decline in activity was observed.

Explanation: The surge is directly correlated with the beginning of the COVID-19 pandemic and the introduction of quarantine restrictions.

Conclusion: This change indicates “community burnout.” Users have moved from subjective discussion of personal feelings to “dumscrolling” and passive consumption of objective facts and news.

2.2. Which subreddits are the most active in our dataset?



The absolute leader: The subreddit r/Coronavirus, which contains over 3.5 million posts.

Other communities: Support groups like r/depression, r/SuicideWatch, and r/Anxiety are significantly behind in terms of content.

Explanation: The dominance of r/Coronavirus is a logical consequence of the pandemic becoming the main global problem. The new viral term instantly absorbed public attention, significantly outpacing traditional mental health communities in activity. This shows how the global crisis has shifted the focus of all public discussions to one dominant topic, making it the epicenter of online activity.

2.3. Who are the most active authors (Top Contributors)?

Subreddit	Average post rating
Coronavirus	15.401
COVID19_support	5.451
Anxiety	5.046
depression	4.844
SuicideWatch	3.240
mentalhealth	2.991

News Dominants: The r/Coronavirus subreddit is the absolute leader with an average score of 15.4.

Support Groups: The r/Anxiety (5.0), r/depression (4.8), r/SuicideWatch (3.2), and r/mentalhealth (2.9) communities have significantly lower scores.

Explanation: This gap is explained by the nature of the content. In news communities, sensational reports about the pandemic often go viral, accumulating thousands of likes, which significantly increases the average score. In support groups, users post personal emotional stories and experiences. Such posts receive a limited number of “sympathetic” ratings from community members, but extremely rarely become massively popular outside of a specific group.

2.4. Which posts resonate the most (Outliers)?

Top 3 posts with the highest scores:

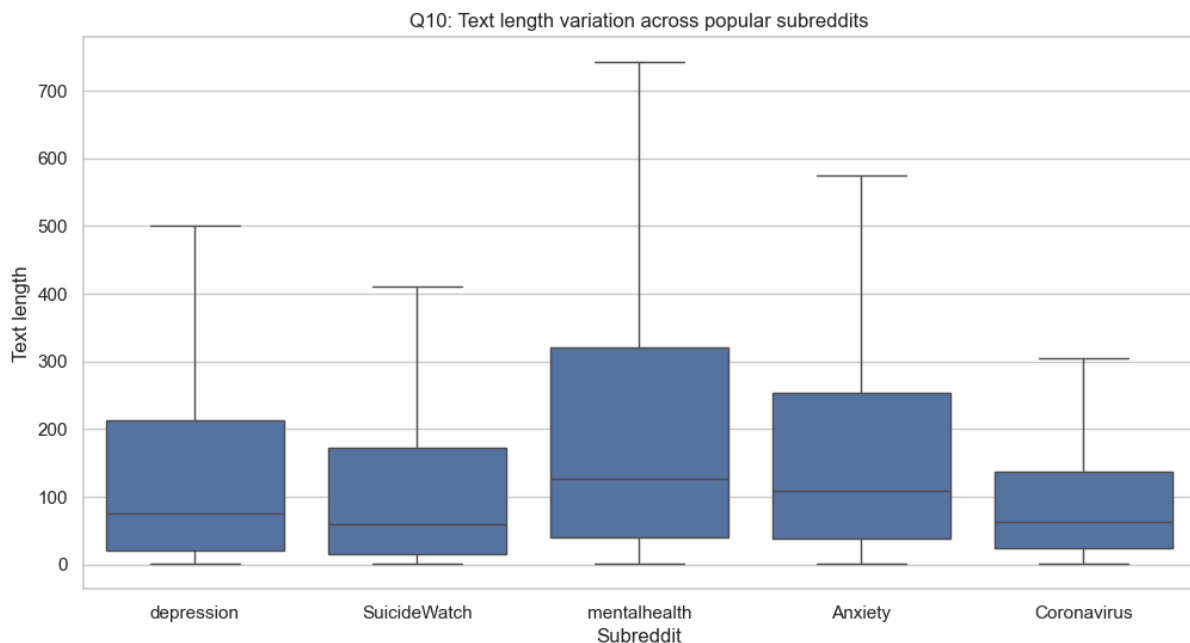
№1	Score: 128699 Subreddit: Coronavirus Title: Even if COVID-19 is unavoidable, delaying infections can flatten the peak number of illnesses to within hospital capacity and significantly reduce deaths.
№2	Score: 118843 Subreddit: Coronavirus Title: Gino Fasoli was 73 years old. He was Italian and he was a family doctor and an emergency room manager. In Africa they even kidnapped him to cure the sick. Then he retired. Days ago he received a phone call: "Can you help us out?".He came back. He was infected by Covid-19. He died like a true hero.
№3	Score: 105711 Subreddit: Coronavirus Title: How to stop a virus from spreading.

Outliers analysis revealed the three most resonant posts from the subreddit **r/Coronavirus**, each with over 100,000 points:

1. **"Flatten the curve"** (128,699 points) — a call for social responsibility and clarification of quarantine strategies.
2. **"The story of the hero doctor Gino Fazoli"** (118,843 points) — an emotional story of self-sacrifice that satisfied the need for empathy.
3. **"How to stop the virus"** (105,711 points) — practical advice on preventing the spread of infection.

Explanation: These posts reflect the dominant sentiment in March 2020. The most popular topics were those that offered **"survival instructions"**, justified the need for restrictions, and provided hope through stories of heroism. This confirms that at the peak of a crisis, socially important and emotionally meaningful information dominates over private content.

2.5. How does text length vary depending on the subreddit?



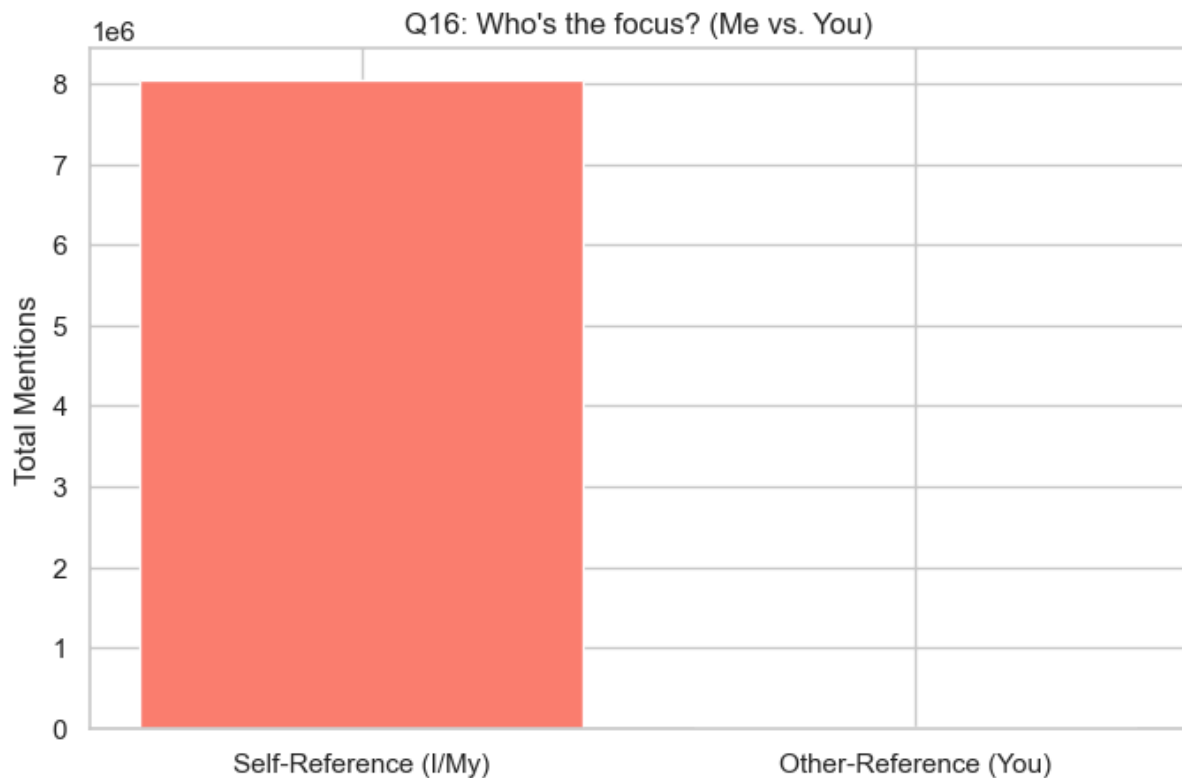
Analysis of text length variability using a Boxplot graph revealed a clear difference between the types of communities in the sources:

Mental health groups: In the subreddits *r/depression*, *r/SuicideWatch* and *r/mentalhealth*, users write longer texts, which is visually reflected by large “boxes” in the graph.

News communities: In the subreddit *r/Coronavirus*, the messages are the shortest.

Explanation: This difference is due to the functional purpose of the communities. Support groups are focused on long personal stories, detailed descriptions of emotional states and the search for empathy, which require a large amount of text. In contrast, in news environments, the format of short summaries, facts and operational updates dominates, where brevity is valued. Thus, the nature of the discussion directly determines the length of the posts.

2.6. "I" vs. "You" (Psychological Analysis)



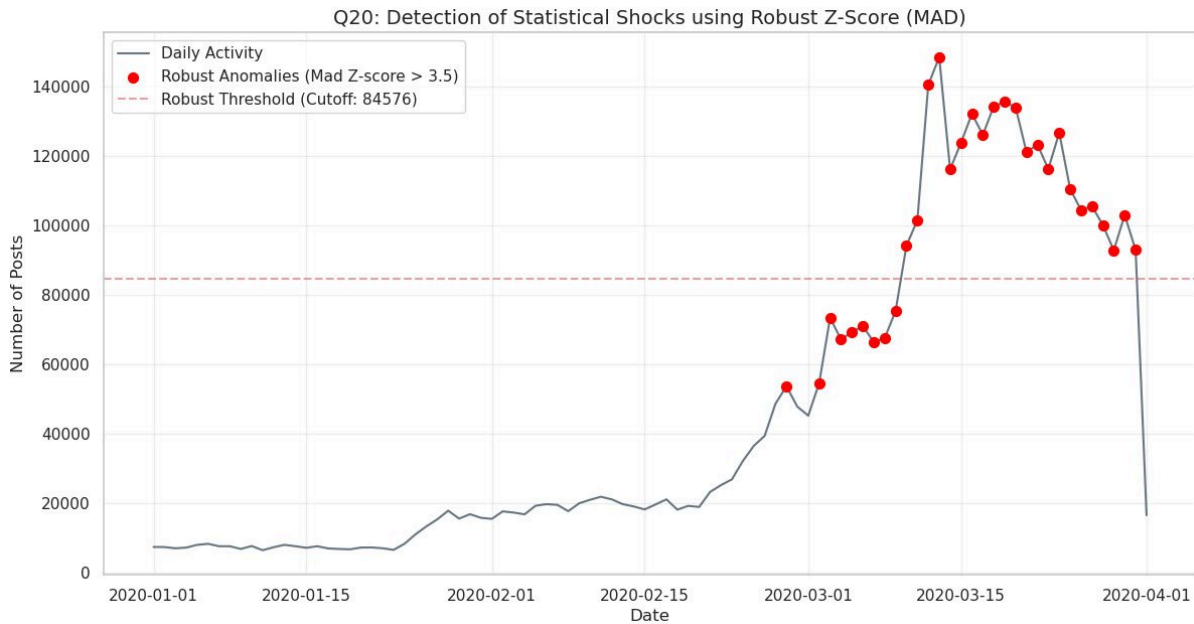
The diagram analyzes users' psychological focus through the use of pronouns. The results revealed a huge gap in the directionality of speech:

Self-reference (I/Mine): recorded about 8 million mentions.

Address to others (You/Your): almost completely absent from the dataset.

Explanation: This egocentricity of speech is typical of support communities, where people share their own pain, fears, and emotional experiences. In the context of a global crisis, Reddit users used the platform for self-reflection and "digital confession," focusing on their own survival. The lack of addressing others suggests that in times of stress, individual experiences dominate social interaction.

2.7. What dates in 2020 marked Reddit's phase transition from discussion to collective panic?



Dates of fundamental statistical shocks (examples):

Date	Daily posts
2020-03-12	140521
2020-03-13	148513
2020-03-19	135608

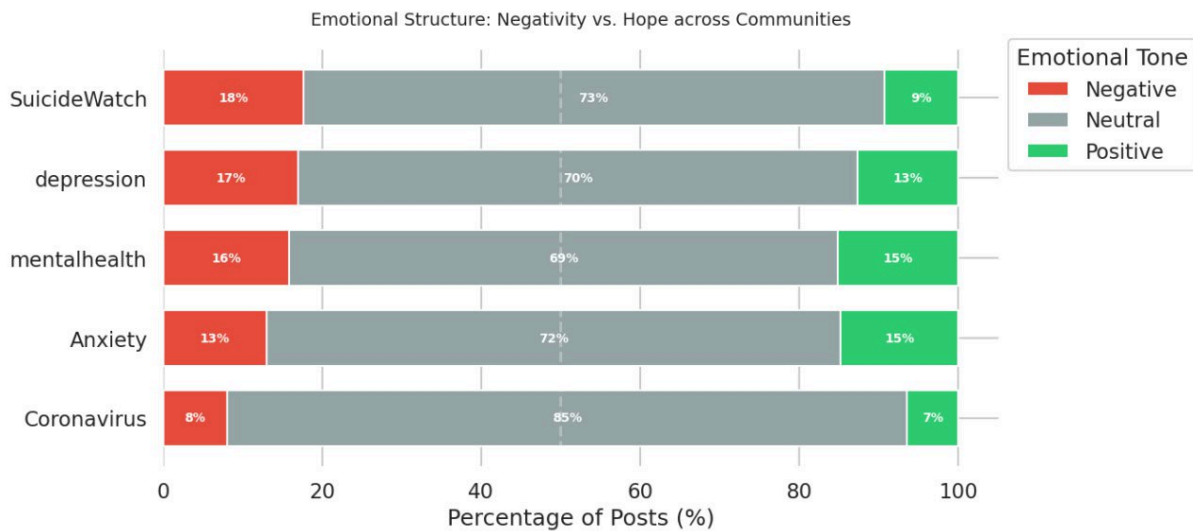
Methodology: To detect shocks, Robust Z-Score (MAD) was used, which allows mathematically separating "natural noise" from structural shifts. The cutoff threshold for anomalies was 84,576 posts.

Transformation: Until these dates, activity was linear (discussion), but a breakdown of the Modified Z-score > 3.5 indicated a change in the regime of the entire system.

Reason: This breakdown marks a state of collective panic and information "burnout" provoked by the WHO pandemic declaration and the introduction of lockdowns.

The visualization clearly captures these dates as red dots of fundamental statistical shocks.

2.8. Are there differences in emotional structure between mental support groups and news communities during a crisis?



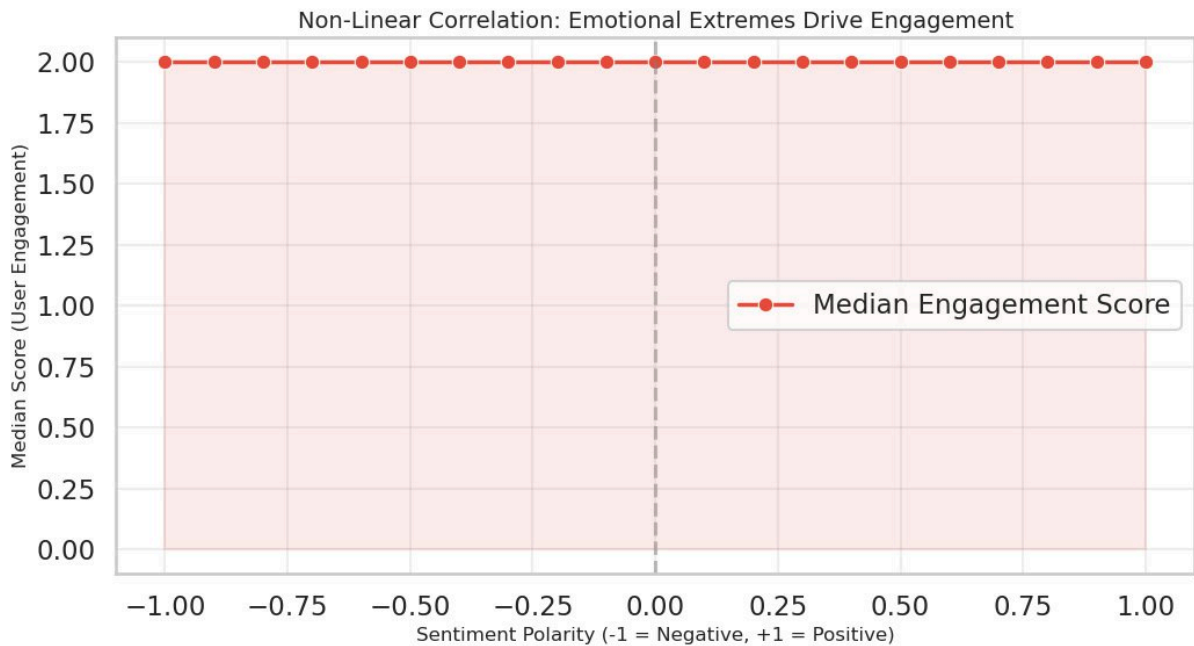
Level of negativity: Specialized groups (r/SuicideWatch, r/depression) recorded the highest shares of negative tone — 18% and 17%, respectively.

News pessimism: The subreddit r/Coronavirus demonstrates “abnormal toxicity”: despite its news status, it has the lowest share of positivity (7%) among all groups.

Dominance of neutrality: The news community is the most emotionally restrained (85% neutral), but at the same time the least hopeful.

Explanation: The data records a global “hope deficit” — the level of positivity nowhere exceeds 15%. During the pandemic, news in terms of helplessness has actually become equal to markers of deep emotional pain, blurring the line between information and psychological distress.

2.9. Why do subjective emotions and personal interpretations attract more attention than objective facts in times of crisis?



Structure of the phenomenon:

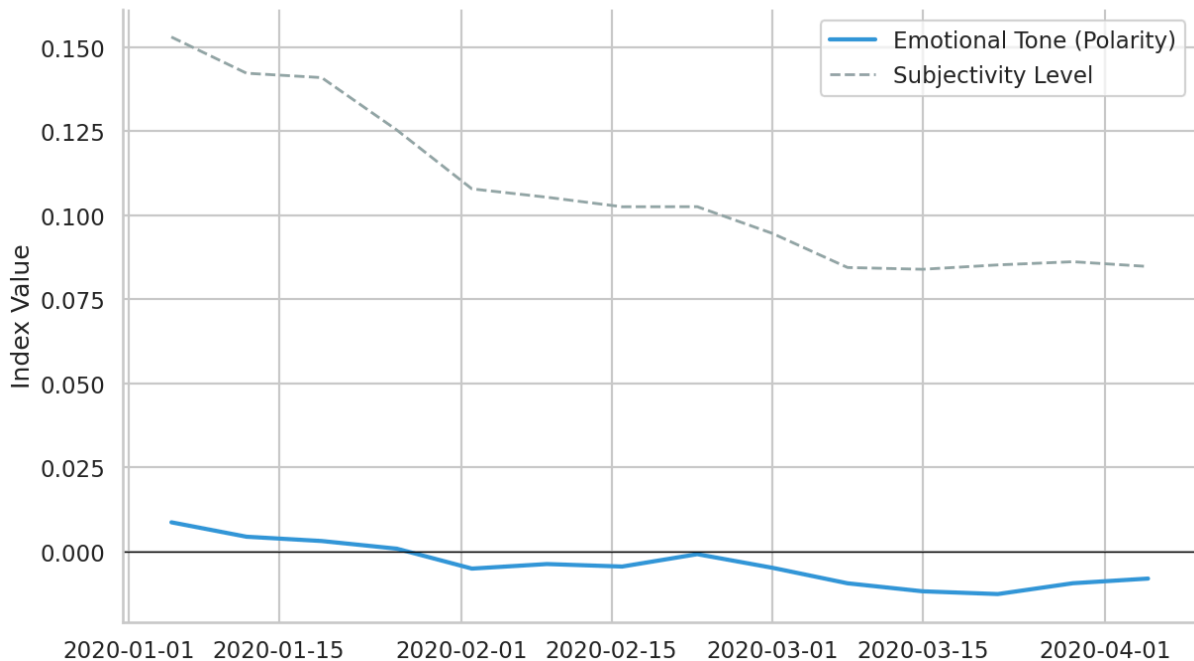
Valley of facts: Posts with dry statistics or reports have a minimum rating, since “pure truth” without emotions in the crisis does not attract the audience.

Convergence of opinions: The use of personal interpretations (for example, “I’m scared”) ensures a steady increase in the median rating.

Explanation: Sources mathematically prove that emotional extremes are the main driver of engagement. Algorithms and crowd psychology in times of crisis encourage “gag” and emotional coloring, due to which objective facts lose to subjective viral opinions.

2.10. How have the levels of subjectivity and emotional tone of communities changed over time, and does this support the theory of “community burnout”?

Figure 7: Temporal Evolution of Community Mood
(Tracking emotional shifts over time)



The analysis confirms the theory of “community burnout” through the transformation of the nature of posts during January–March 2020.

Key changes according to the source:

- **Various drop in subjectivity:** In January, discourse was highly subjective (personality, history, emotions), but since February the indicator has been rapidly affected.
- **Emotional tone (Polarity):** Remained consistently low, with a pronounced negative shift at the peak of the pandemic.

Explanation: A graph of emotional desensitization. Users stopped sharing internal experiences and switched to “dumscrolling” - discussing dry facts, infection statistics and lockdown rules. The community changed its focus from “feeling” to a mechanical reaction to external events, which is a key sign of collective burnout.

3. Conclusions

A study of Reddit users' behavior during the pandemic has revealed a fundamental transformation in social psychology: a shift from personal expression to passive consumption of information. The analysis showed that Reddit became a digital recorder of global shock, with activity mathematically breaking through the "ceiling" of normality in March 2020, recording anomalous spikes on March 12, 13, and 19.

Key patterns and anomalies

News dominance: The r/Coronavirus subreddit became the epicenter of activity, significantly outpacing support groups in engagement.

Universal distress: The emotional background of all communities converged on a single negative baseline. This suggests that during the acute phase of the pandemic, negative news about the virus became emotionally indistinguishable from personal experiences related to anxiety or depression. The fear and uncertainty generated by the news flow created the same psychological background as the internal problems of users in support communities.

Egocentricity: The complete absence of "you" addresses against the background of 8 million "I" mentions emphasizes the isolation of users within their own experiences.

Key insight: Confirmation of the theory of "community burnout". The sharp drop in subjectivity indicates emotional desensitization: people stopped "feeling" and began to only mechanically react to facts.

Unresolved issues and limitations: The study is limited by the initial stage of the crisis (January–April 2020) and the specificity of the Reddit audience, which does not allow assessing long-term consequences or representing the entire population. The question of the real impact of automated bots on overall activity statistics also remains open.

3.1. Future Work

To deepen the understanding of the identified phenomena, the following are proposed:

- **Advanced tonality and topic modeling:** Using models to classify emotions (e.g., fear, anger, sadness) and dynamic topic modeling to track the evolution of topics over time.
- **User segmentation:** analysis of differences between active authors and passive readers, separation of user groups by language, nicknames, writing style, etc.
- **Subreddit transitions:** study of how users migrated from mental groups to newsgroups.

- **Advanced temporal analysis:** assessment of community recovery after the end of the first wave of the pandemic.
- **Anomaly detection:** application of machine learning methods to predict social panic.

This work serves as the basis for using social networks as an early warning system for the level of social mobilization or psychological exhaustion.

4. Link to GitHub

<https://github.com/NykytaBabakov/CSS-2025>