University of Technology Graz

Sentiment, Fake News, and News Sharing Behaviour

Report

written as part of the course
706.025 Foundations of Computational Social Systems

submitted by:
Group 5
Lejla Kovać, Nao Kameshima, Vanessa Draxler

Graz, on 10.02.2026

GitHub: https://github.com/vanessa-draxler/FCSS-Group-Project

Table of contents

1 Introduction 1
2 Methodology 2
2.1 Research question 1 3
2.2 Research question 2 4
2.3 Research question 3 5
3 Conclusion 6
4 Critique 7
5 References 8

List of tables and figures

Table 1: Distribution of articles over sentiments 3

Figure 1: Number of shares by sentiment 4
Figure 2: Number of shares by veracity 5
Figure 3: Interaction of sentiment x veracity on number of shares 5

1 Introduction

First and foremost, before briefly explaining each part of our project, it is important to give you an insight into our motivation which led us to setting the research questions, doing the analysis and interpreting the results. In the past few years, the occurrence of fake news on social media platforms has risen significantly, which we all witnessed. As social networks became one of the primary sources of information for most people, the speed and volume at which news need to be produced and shared to satisfy users have increased dramatically. As a consequence, it became difficult to distinguish between facts, e.g. factual reporting and fake news, e.g. misleading content (Abraham et al., 2025). This blurred boundary sets some serious challenges, not only for the individual media literacy, but also for the stability of public discourse. Therefore, the importance of detecting and understanding fake news in order to reduce social polarization is undeniable. What do we mean by that? When fake news spread fast and widely, reinforcement of extreme viewpoints as well as deepening divisions between different social groups is more than possible. That leads to undermining the trust in institutions and traditionally known media, which is why it is necessary to further examine the mechanisms that drive the spread of fake news.

The previous mentioned challenges are often discussed in the context of so-called "post-truth" age, a term that describes a social environment in which objective facts, or objective side of the fact is less relevant in shaping public opinion than personal beliefs and emotions (Abraham et al., 2025). In this environment, emotionally charged content is what matters, not the information itself, even if it is verified. Prior observations on this topic imply that especially negative messages have a tendency to be shared or retweeted more, in comparison to positive or neutral ones (Tsugawa & Ohsaki, 2017). The fact that emotional tone may play a key role in determining how the message or news is shared led us to this project and formulating our first independent variable – sentiment of a news article. Going further, researches showed that fake news, in comparison to the real ones, were spreading faster and reaching more people (Vosoughi et al., 2018). Investigating this, we set our second independent variable – veracity of a news article.

Finally, this study seeks to investigate which aspects of a news article post influences its number of shares – our dependent variable – and what their interaction makes. It focuses on three research questions:

- RQ1: Does the sentiment of a news article post (negative/neutral/positive) affect its number of shares?
- RQ2: Does the veracity of a news article post (fake/true) affect its number of shares?
- RQ3: Does the combination of sentiment and misinformation influence the number of shares?

2 Methodology

This project operated with openly available dataset on Kaggle (Yadav, 2025), which generally provides structured datasets for research purposes. Our dataset includes approximately 4,000 news articles coming from both reliable and unreliable sources. Reliable sources included media such as CNN and *The New York Times*, while the unreliable sources incorporated websites known for spreading misinformation. Three main features used for the quantitative analysis, whose explanation we would provide in later paragraphs, were:

- Number of shares: How often was the article shared? (dependent variable)
- Sentiment score: What is the emotional tone of the shared article? (independent variable)[1]
- Veracity of the news: Is the shared article true or fake? (independent variable)[2]

When it comes to data processing, first steps involved cleaning and preparing the dataset for further analysis. This consisted of checking for missing or inconsistent values and making sure that all relevant variables were in a usable format. Even though the used dataset provided a continuous sentiment score, it wasn't appropriate for categorical comparison, which required conversion to three discrete categories: positive, neutral and negative. Specifically, following the previous study by Abraham et al. (2025), if the score>0.01, the polarity of the article is positive, if the score<-0.01, the polarity of the article is negative, and otherwise it is neutral. This transformation allowed us to do a group comparison and aligned the data structure with previously defined research questions.

Following this process, exploratory data analysis was performed as a next step before main statistical analysis in order to understand the basics of the analyzed dataset. In this step we examined distributions of the number of shares, compared behavior by sentiment category and by veracity. The goal of this phase was to detect possible outliers or any other anomalies before applying other statistical tests. Overall speaking, this methodological path ensured that the data was systematically clean, transformed and ready for the further analysis.

To give an answer to the research questions in terms of numbers, Analysis of Variance (ANOVA) tests were conducted. A one-way ANOVA was used for RQ1 (effect of sentiment on number of shares) and RQ2 (effect of veracity on number of shares). A two-way ANOVA for RQ3 (the interaction effect between sentiment and veracity on the number of shares). This approach enabled the identification of both individual and interaction effects of the independent variables on news sharing behavior.

Last but not least, graphical visualizations were created using Seaborn and Matplotlib to illustrate our results. Boxplots, distribution plots, and interaction graphs were used to picture

[1] Emotional tone is expressed as a numerical value
[2] News have label (Yes/No), indicating the veracity of the article

differences and relationships between groups, which made them easier to interpret and to support the numbers with visual evidence and make our explanation easier to understand.

The first table gives us a descriptive overview of the relationship between sentiment category and news veracity. Even if it is only descriptive, it gives as a good starting point for ANOVA because we can clearly see how our variables are distributed before we examine our main and interaction effects. It shows the percentage distribution of fake and real news articles within each sentiment group (negative, neutral, and positive). The "Sum" column confirms that each sentiment category represents 100% of the articles within that group, meaning the percentages of fake and real news are calculated relative to the total number of articles with the same sentiment.

Table 1: Distribution of articles over sentiments

	Fake	Real	Sum
Negative	50.63%	49.37%	100%
Neutral	46.38%	53.62%	100%
Positive	50.82%	49.18%	100%

Source: Own illustration

From Table 1, we can observe that for negative sentiment articles, nearly 51% are classified as fake news, while 49% are real. A very similar pattern appears for positive sentiment articles, where, again, about 51% are fake and 49% are real. In contrast, neutral sentiment articles show a slightly different distribution: 46% are fake and 54% are real, but still quite equally distributed.

What is important for us to conclude out of this table is that the differences between fake and real news across sentiment categories are relatively small. There is no strong imbalance in any category that suggests strong association of any sentiment with fake or real news. This descriptive result already hints that the common assumption — that negative content is more likely to be fake — is not supported by this dataset. Instead, fake and real news appear to be distributed quite evenly across the sentiments, with only a minor tendency for neutral articles to be more often real.

2.1 Research question 1

The result of one-way ANOVA implicates that the main effect of sentiment on the number of shares is only slightly significant at the 10% significance level ($p = 0.067$), meaning that sentiment shows low tendency to have an impact on news sharing behavior, but not statistically significant, which is an answer to our first research question.

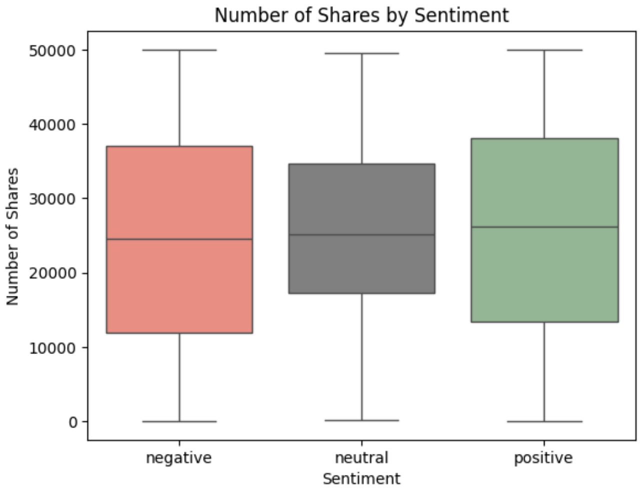

Figure 1: Number of shares by sentiment (Source: Own illustration)

The visualization uses a boxplot, which helps to compare the distribution of shares among the three sentiment categories. On the x-axis, the three sentiment groups—negative, neutral, and positive—are displayed. The y-axis indicates the number of shares. Each boxplot summarizes the statistical distribution within a group. The horizontal line inside each box marks the median number of shares. The box itself represents the interquartile range (IQR), which means the middle 50% of the data. The whiskers extend toward the minimum and maximum values, excluding extreme outliers.

From the visualization, we can notice a couple of things such as:

- The median of shares is slightly higher for positive articles compared to negative and neutral ones.
- Neutral articles have lower median compared to the other two groups.
- The ranges of all three boxplots overlap considerably, which visually suggests that the differences between sentiment groups are not very strong.

2.2 Research question 2

Again, one-way ANOVA showed us the main effect of veracity on the number of shares – statistically insignificant (p=0.148). Since this value is well above the 5% significant level, we can say that veracity itself does not significantly influence news sharing behavior. The data examined in this study does not clearly demonstrate an independent influence of veracity on the number of shares, despite prior research frequently suggesting that fake news spread more broadly.

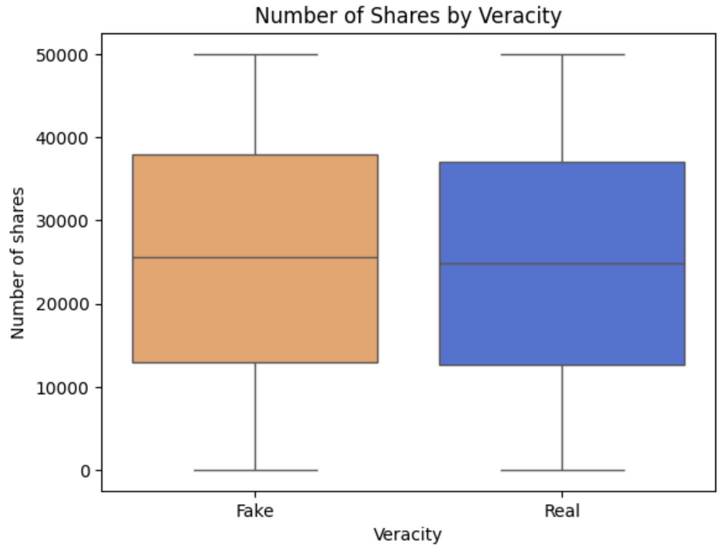

Figure 2: Number of shares by veracity (Source: Own illustration)

The two veracity categories—Fake and Real—are displayed on the x-axis. The number of shares is shown on the y-axis. The statistical distribution within each group is summarized once again by each boxplot.

Visually speaking, the two boxplots are similar, as well as in the first case:

- Fake and real news articles have about the same median.
- Both groups' data spreads are similar, with wide ranges suggesting a significant degree of sharing behavior variability.
- The boxes and whiskers show a high degree of overlap, indicating that there is little difference in the frequency of sharing between fake and real news.

2.3 Research question 3

In this case, we analyzed the interaction effect through two-way ANOVA which showed that the interaction effect between sentiment and veracity is statistically significant at the 5% level ($p = 0.015$). This means that, although sentiment and veracity separately did not show strong independent effects, their combination does significantly influence sharing behavior.

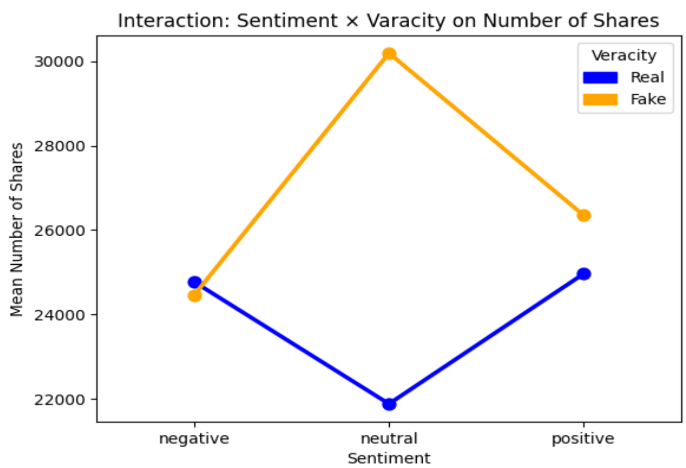

Figure 3: Interaction of sentiment x veracity on number of shares (Source: Own illustration)

Unlike the previous boxplots that examined our independent variables separately, this visualization is an interaction plot, which is specifically designed to show how two independent variables together have an impact/effect on our dependent variable. On the x-axis, the three sentiment categories are displayed: negative, neutral, and positive. The y-axis represents the mean number of shares. Two separate lines are drawn: one for real news articles and one for fake news articles. Each point on the lines corresponds to the average number of shares for that particular combination of sentiment and veracity.
The two presented lines are not parallel, which is a key visual indicator of an interaction effect. This means that the impact of sentiment on news sharing behavior changes depending on whether the article is fake or real. For real news, the number of shares is lowest for neutral sentiment and slightly higher for negative and positive sentiment. For fake news, the pattern is almost the opposite: neutral fake news articles receive the highest average number of shares, clearly exceeding both negative and positive fake news. The difference between fake and real news is particularly pronounced in the neutral category, where fake news shows a strong peak while real news shows a dip.
To answer research question 3: This finding implicates that news sharing behavior may not be explained by emotional tone or veracity in isolation, but pairing these factors shape reasonable explanation. What is interesting is that neutral fake news receive the highest number of shares. This conflicts our main assumption that highly emotional or negative misinformation is always the most viral.

3 Conclusion

The goal of this project was to provide explanations of which features of news articles have influence on their number of shares. After detailed literature research, we focused ourselves on sentiment, veracity and their interaction effect. Our results showed that news sharing patterns cannot be explained by exclusively looking at only one of these factors, but by analyzing their interaction it is possible to give some reasonable answers.
Regarding the individual effects of our independent variables, sentiment showed only a marginal main effect on the number of shares, which suggests that this variable (isolatley) has slight influence on the number of shares, but not strong enough to confirm a significant impact. On the other side, veracity alone did not show a statistically significant effect, meaning that there is no difference between the engagement of real and fake news articles.
The most important outcome of our statistical analysis was the significant interaction effect of sentiment and veracity: the influence of sentiment on number of shares depends on whether a news article is true or fake. Specifically speaking, the results showed that fake news articles with neutral sentiment had a tendency to receive the highest number of shares. If we translate this to understandable speech, it means that, despite the expectation that emotionally extreme

misinformation (regardless of the direction) is always the most shared and wanted one, neutral fake informations is especially effective in spreading widely between people.
Talking about reliability, behind our conclusions stands consistent pattern across both visualisations and statistical tests, especially the interaction got from the two-way ANOVA, which was significant. Even though, the results should be interpreted and read with moderate confidence rather than absolute certainty because of potential limitations, which we will further explain in the next paragraph.

4 Critique

As every study, this one has also several limitations which we would like to name because they minimize the generalizability and validity of its findings and open the possibility for some alternative explanations and basis for further research papers.
First of all, the dataset does not fully represent all news articles or social media platforms. It is limited to a specific Kaggle dataset of 4,000 articles and probably does not involve the diversity of global media, cultural differences, or platform-specific algorithms that have deep impacts on visibility and news sharing behaviour. As a result, the conclusions have a risk of not being universally applicable.
Second but not less important, the analysis focused only on three variables — number of shares, sentiment, and veracity — while other factors that can possibly have an impact, such as the credibility of the source, topic sensitivity, political orientation, posting time, audience size, or platform recommendation algorithms, were not taken into consideration. The absence of these variables means that some of the observed effects could possibly be explained by unmeasured external influences rather than sentiment or veracity alone.
After those two comes the facts that the assumption which was our starting point (negative posts are more likely to be fake) was not supported by the statistical analysis. Fake and real news were distributed relatively evenly across sentiment categories. Again, this decreases generalizability and connects to the previously mentioned fact – there are some unconsidered factors which could explain the behavior of users.
Finally, the statistical results for individual effects were either weak or non-significant, which leaves us only with the interaction effect as basis of our assumptions and explanations. Is this actually enough? Even though this interaction was statistically significant, we still need to come back to one of our limitations – the dataset. We believe that future studies with larger and more diverse datasets, additional and potentially relevant and significant variables and meta-analysis of all previous findings could be a key to some strong conclusions about this topic.

5 References

Abraham, T. M., Wen, T., Wu, T. & Chen, Y. (2025). Leveraging data analytics for detection and impact evaluation of fake news and deepfakes in social networks. *Humanities and Social Science Communications Vol. 12, article no. 1040.*

Tsugawa, S. & Ohsaki, H. (2017). On the relation between message sentiment and its virality on social media. *Social Network Analysis and Mining, Volume 7, article no. 19.*

Vosoughi, S., Roy, D. & Aral, S. (2018). The spread of true and false news online. *Science Vol. 359, Issue 6380, p. 1146–1151.*

Yadav, K. (2025). Fake News Detection [Data set]. Kaggle. https://www.kaggle.com/datasets/khushikyad001/fake-newsdetection/data