2023 IEEE 24th International Conference on Information Reuse and Integration for Data Science (IRI) | 979-8-3503-3458-6/23/$31.00 ©2023 IEEE | DOI: 10.1109/IRI58017.2023.00041

Meme It Up: Patterns of Emoji Usage on Twitter

Mohammad Shiri
Department of Computer Science
Old Dominion University
Norfolk, VA, USA
mshir001@odu.edu

Oleksii Dubovyk
Department of Biological Sciences
Old Dominion University
Norfolk, VA, USA
odubo001@odu.edu

Golbarg Roghaniaraghi
Department of Economics
Old Dominion University
Norfolk, VA, USA
grogh001@odu.edu

Sampath Jayarathna
Department of Computer Science
Old Dominion University
Norfolk, VA, USA
sampath@cs.odu.edu

***Abstract*—As emojis have grown in their popularity in social media over the last decades, they not only enrich messaging with emotional connotations but offer a convenient system for studying the effects of memes on ideas' survival. In this work, we treat emojis like standardized memes to test the impact of their usage on different facets of success within social media. Specifically, we extracted random individual tweets from Twitter to construct a list of emojis used within each tweet. With this dataset, we aimed to address three distinct questions: (1) whether there are specific patterns of emoji usage that increase tweet popularity; (2) whether emojis usage on tweeter can be a good predictor of the stock market trading volume; and (3) whether there is a specific subset of emojis associated with low-quality tweets (e.g., spam). We found no evidence of the positive effects of emoji usage on tweet popularity. However, there was a reason to claim that negative emojis may trigger an intensive response from the audience. For some companies, we were able to accurately predict stock patterns based on emoji usage. Finally, there clearly was a specific subset of emojis used in low-quality tweets. This work may serve as a starting point for a deep investigation of the emoji-meme system, as this topic seems to be relatively fresh in the literature.**

***Index Terms*—Twitter Analysis, Popularity Trend, Memes, Emoji**

I. Introduction

Richard Dawkins, a famous evolutionary biologist, coined the term ”meme” denoting a gene of information existing in the pool of ideas: ”Just as genes propagate themselves in the gene pool by leaping from body to body via sperms or eggs, so memes propagate themselves in the meme pool by leaping from brain to brain” [1]. A meme can be defined as an idea, behavior, style, or usage that spreads from person to person within a culture, according to the Merriam-Webster dictionary. Anything can be a meme: a rumor, a funny picture, or a political ideology. Similarly to biological genes, not all memes survive the process of natural selection, and it is important for them to have a set of characteristics that fit their environment to ensure long-term survival.

In this work, we focus on a particular subset of memes – emojis – that are, like words, used to transmit ideas and moods in an environment of social media. We will use Twitter as such an environment and tweets as carriers of emojis, similar to biological organisms in an environment carrying specific genes.

Emojis can perform variable functions in communication, they can either be used instead of nouns or verbs, or adjectives substituting those in otherwise grammatically complete sentences or enriching such sentences via additional emotional meaning [2]. Also, the meaning of an emoji might vary among societies, age groups, and social groups, and be perceived differently among them [3]. It has been shown that patterns of emoji usage vary from country to country and the distribution of different emoji usage is similar to one that can be found in biological communities of organisms [4]. In the study presented here, using Twitter API, we test whether there is an effect of emoji usage on tweet popularity metrics, such as the number of tweet likes, retweets, and replies.

Using Twitter data offers a big data framework and real-time processing that helps us determine the relationship between stock and Twitter [5]. Therefore, in this research, we aim to investigate how emojis can be used as an indicator of sentiment to predict stock market fluctuations. Emojis have become a popular form of communication on social media, and recent studies have shown that they can effectively convey sentiment [6], [7]. By analyzing the frequency and sentiment of emojis used in tweets related to specific stocks, we can determine if they have any predictive power for the stock market.

Our aim in this regard is to evaluate the impact of emojis on the stock market through the analysis of categorized tweets related to six major companies: Google, Tesla, Advanced Micro Devices, Amazon, Apple, and Microsoft. To collect the tweets, we retrieved those that contained cashtags - a clickable symbol created by adding a dollar sign in front of a stock ticker symbol (e.g., $AAPL). Within the categorized tweets, we counted the number of emojis and investigated the correlation between changes in the daily number of emojis and the stock market trading volume for each company relative to the respective cashtags. Overall, Twitter data can provide valuable insights into the relationship between news, sentiment, and the stock market. By incorporating emojis as an indicator of

979-8-3503-3458-6/23/$31.00 ©2023 IEEE
DOI 10.1109/IRI58017.2023.00041

sentiment, we aim to broaden our understanding of how social media can be used to predict stock market fluctuations.

Additionally, low-quality content detection methods for spam and phishing are widely used in social media, aiming to use machine learning techniques, but we are not aware of any studies emphasizing the role of emojis in low-quality content. We present a detection system for low-quality content that uses emojis along with other features extracted from users and content to see whether there are any patterns of emoji usage in spam and/or phishing. Our main contributions can be summarized as follows:

1. Investigates the impact of emojis on tweet popularity metrics such as likes, retweets, and replies.
2. Explores the correlation between emoji usage in tweets and the stock market trading volume for specific companies.
3. Presents a detection system that utilizes emojis to identify low-quality content, such as spam and phishing.

II. Related Work

A. Memes and Ideas Popularity

In the modern world, the flow of information one is constantly exposed to is enormous compared to the pre-Internet era [8]. Among such an intense flow, an individual should prioritize what information to pay attention to, therefore, all ideas within the information space are in constant competition with each other, spreading similarly to infections among the susceptible population if there is any feature that may increase the fitness of such ideas [9]. It has been shown that the mechanisms of informational selection obey similar laws to natural selection [10], exhibiting complex dynamics over large periods of time [11], affecting the behavior of individuals [12], [13] and manipulating public opinions [14]. It is yet unclear what features make memes more successful compared to others, and emojis provide a convenient standardized system to address this question [15]. As it has been shown that emojis vary in perceived sentimental value [16], we could also expect that emojis deferentially affect tweets' popularity.

B. Emojis and the Stock Market

The news media plays a crucial role in disseminating information to the public, particularly when a company declares its intent to purchase another company [17]. The efficient market hypothesis (EMH) suggests that stock market prices are primarily influenced by new information, such as news, rather than present and past prices [18]. Therefore, the stock market has always been a complex and unpredictable environment, making it difficult for investors to make informed decisions about when to buy or sell their shares. With the rise of social media platforms, researchers have started exploring the use of sentiment analysis techniques to predict stock market trends based on the emotions expressed on social media platforms [19].

The increasing shift of news consumption to online platforms has been a trend observed in recent years. Studies have shown that 77% of all social media users keep up with the news at least once a day, and among Twitter users, 81% keep up with the news daily [20]. Furthermore, Twitter has become the preferred medium for breaking news, consistently leading Facebook or Google Plus [21]. This trend suggests that Twitter data may offer valuable insights into the impact of news on the stock market.

However, news alone may not be the only factor that affects stock market values. Public mood or attitude can also play a significant role. Psychological studies have shown that emotions, as well as information, influence decision-making [22], [23]. In fact, the stock market and social media are constantly evolving, and the changing trends of Twitter can have a significant impact on the stock market [24]. Bollen, Mao, and Zeng (2011) were among the first to explore the link between Twitter mood and stock market trends. They used a mood index based on 9.8 million tweets from 2.7 million users to predict daily changes in the Dow Jones Industrial Average (DJIA) with an accuracy of 87.6% [25]. Other studies have also found tweets to be a strong predictor of the stock market, and these predictions can be used to identify short-term trading opportunities [19], [26].

C. Low-Quality Content

Machine learning techniques utilizing statistical features have been recently used in the majority of studies involving the detection of low-quality tweets. These features may be extracted from the tweet itself or the account that published it. In order to detect low-quality content, Wang et al. [27] examined four types of features from Twitter data: user features, content features, n-grams, and sentiment features. The most discriminating set of features was user-based, but the most time-consuming features were sentiment and content-based. Aggarwal et al. [28] created a browser extension that instantly flags tweets that are phishing while users are browsing Twitter. Twenty-two features must be extracted from a particular tweet in four categories (URL, WHOIs, Tweet, and network-based characteristics). After being trained and tested on these features, three machine learning classifiers, Naive Bayes, Decision Tree, and Random Forest, provided the best results. In order to identify between low-quality tweets and authentic tweets, Gupta et al., [29] used a variety of user- and tweet-based features, including account age, number of followers, number of accounts followed, and number of tweets, in addition to the bag of word representations of the textual data. Chen et al., [30] provided a thorough analysis of low-quality information from the viewpoint of social network users. In addition, the authors developed a real-time system that extracts 32 features from two categories: direct features, which can be quickly computed and retrieved, and indirect features, which take longer to extract but are more important for categorizing low-quality tweets. This system achieved remarkably high accuracy.

III. Methodology

In this study, we considered Twitter as a source of data for public emotions and used emojis as an indicator of people's sentiments. To avoid possible bias caused by the dynamics of Twitter popularity among users, we stratified the search by each calendar day between January 2007 and September 2022. We used different endpoint URLs along with different sets of query parameters (Tables I and II) to extract tweets from Twitter API V2 according to date (from May 2014, when Twitter started supporting emojis, to May 2022), tweet ID (labeled low and high-quality tweets identified in recent work), and cashtags (identifier of tweets discussing a company's stock market) to collect the three desired Twitter datasets that are used in this study.

The following variables were included: author ID, date when the tweet was created, conversation ID, tweet ID, language, count of retweets, replies, likes, and quotes, tweet source (i.e., using which platform the tweet was published), the text of the tweet, date when the author's profile was created, username, whether an account is verified by Twitter, count of followers, following, and tweets published by the account.

We acknowledge the fact that a major proportion of the tweets might be considered invisible by Twitter users, meaning that they would not be exposed to assessment by users and will not experience any selection pressure. We assume that these invisible tweets would most likely be published by accounts with a low count of followers, therefore, we dropped all individual tweets published by accounts with less than 1000 followers. We empirically choose this threshold as some tweets might become disproportionately popular despite a low followers count of the authoring profile, but with that threshold, we expect to eliminate truly invisible tweets at the cost of losing relatively few valid observations. After accounting for the "invisible" tweets, the resulting training dataset contained 286,505 unique tweet records.

A. Testing Effect of Emojis on Tweet Popularity

For each tweet, the list of emojis used was extracted from the tweet text using matches with a predetermined list of emoji characters based on the dataset provided by Kralj Novak et al. [16]. This allowed us to estimate, (i) whether a tweet contained any emojis and (ii) what was the ratio of emojis to words in the text. Moreover, negative, neutral, and positive sentimental scores were available [16] for the emojis considered, which allowed us to estimate the mean weighted sentimental value of emojis used in each tweet varying from 0 (i.e., negative) to 1 (i.e., positive), where 0.5 would refer to the neutral mean sentimental value of the emojis used [31], [32].

We tested whether binary usage of emojis affects tweets popularity (quantified as logarithmically transformed likes count, retweets count, or replies count) using a permutational hypothesis testing whether the difference of means between two populations is different from a difference between means of two populations randomly drawn from a merged pool of observations (1,000 permutations), due to the high skewness of the response variables. The effect of the weighted sentimental value of emojis on tweet popularity metrics within the tweet was tested using a generalized linear model with a Poisson link function [33] and mean sentimental value with the ratio of emojis per word as predictors.

B. Effects of Emojis on Stock Market

To analyze the stock market trading volume, we collected 100 daily stock trading volume and price data from Yahoo Finance between May 2014 and May 2022 for six companies: Google, Tesla, Advanced Micro Devices, Amazon, Apple, and Microsoft. We chose these companies due to their high number of tweets about their stock market and their active stock market over the time frame we were testing for. We then categorized tweets for each company by retrieving tweets that included the specific cashtag, which is a dollar sign followed by the company's ticker symbol. This allowed us to filter content on Twitter and collect data about each company efficiently. Afterward, we counted the number of emojis in the daily categorized tweets and tested for any relationship between the stock market and the daily number of emojis per day for each company from May 2014 to May 2022.

To determine the impact of emoji usage on Twitter on stock market trading volume, we first employed an Ordinary Least Squares (OLS) method to identify if there is any relationship between the number of emojis in tweets and the stock market trading volume for each company. The initial OLS regression model is specified as follows:

$$v_i = \alpha + \beta e_i + \delta p_i + \epsilon_i \tag{1}$$

where v_i is the dependent variable representing the stock market trading volume, e_i is the independent variable representing the number of emojis in tweets, p_i represents the control variable of price, α is the intercept, and lastly β and δ indicate the coefficients.

We then applied a Vector Autoregressive (VAR) method [34] to predict stock market trading volume for each company, using the series of the number of emojis in tweets that included the cashtag of each company in addition to their stock prices series. To do this, we first checked for the stationarity of the series using the Augmented Dickey Fuller (ADF) test [35], and then used the Akaike information criterion (AIC) [36] as a model selection criterion to select the best order for the models for each company. Finally, we trained our stock data by inputting the selected variables into our model from May 2014 to March 2022, and predicted the stock market trading volume from April 2022 to May 2022.

C. Patterns of Emoji Usage in Low-Quality Content

We extracted a set of 16 features from tweets and the accounts that posted them, along with crafted features such as a bag of words of tweets, black list words count using a bag of words, follow rate (follower count/following count), and emoji count in tweets and account description, in the obtained Twitter dataset. Referenced tweets, retweet count, reply count, like count, and quote count were the features associated with tweets provided by Twitter data. Verification status, follower count, following count, tweet count, listed count, geographical location, and description were the features associated with accounts publishing the tweets provided by Twitter data. We used this set of features to train machine learning classifiers on detecting low-quality tweets.

TABLE I

Parameters for Twitter API Query based on the keyword (to extract all tweets from May 2014 to May 2022, and tweets included companies cashtag from the same time frame)

Parameter	**Value**
url	https://api.twitter.com/2/tweets/search/all
query	keyword(everything for first dataset and cashtags of selected companies for second dataset)
start_time	05/01/2014
end_time	05/01/2022
max_results	500
expansions	author_id,in_reply_to_user_id,geo.place_id
tweet.fields	id,text,author_id,in_reply_to_user_id,geo,conversation_id,created_at,lang,public_metrics,referenced_tweets,reply_settings,source
user.fields	id,name,username,created_at,description,public_metrics,verified
place.fields	full_name,id,country,country_code,geo,name,place_type

TABLE II

Parameters for Twitter API Query based on tweet ID (to extract labeled low- and high-quality tweets

Parameter	**Value**
url	https://api.twitter.com/2/tweets?
ids	list of tweet IDs
expansions	author_id,in_reply_to_user_id,geo.place_id
tweet.fields	id,text,author_id,in_reply_to_user_id,geo,conversation_id,created_at,lang,public_metrics,referenced_tweets,reply_settings,source
user.fields	id,name,username,created_at,description,public_metrics,verified
place.fields	full_name,id,country,country_code,geo,name,place_type

We trained a model on the low-quality labeled tweets obtained from Chen et al. [30] and build a classifier model to predict the quality of tweets in our original Twitter dataset. The provided training data contains tweet IDs along with labels indicating low-quality and high-quality tweets. We pulled the tweets by their IDs using Twitter API, but unfortunately, almost half of the tweets are not accessible anymore because these tweets were collected in [30] and most of the low-quality tweets might be deleted by now. Among the existing tweets, 41,796 are labeled as low-quality and 3,550 as high-quality tweets. We split the dataset to train and validate (80/20), and trained classifier models on them.

IV. Results

Tweets that contained any number of emojis had a lower like count (log-difference = $-0.401, p < 0.001$), retweet count (log-difference = $-0.193, p < 0.001$), and reply count (log-difference = $-0.043, p < 0.001$). Both mean sentimental value and ratio of emojis per word negatively affected like count ($Likes \sim 3.881 - 1.531 \cdot Sentiment - 2.318 \cdot (Emoji/word)$), retweet count ($Retweets \sim 8.014 - 0.819 \cdot Sentiment - 0.329 \cdot (Emoji/word)$), and reply count ($Replies \sim -0.144 - 0.778 \cdot Sentiment - 1.184 \cdot (Emoji/word)$).

A. Effects of Emoji Usage on Stock Market

The results of our OLS model reveal a statistically significant relationship between the number of emojis in tweets that include the specific cashtag for each company and the company's stock trading volume, as revealed by the OLS model.

To further explore the relationship between emoji usage and stock market trading volume, we applied the VAR method to predict the volume of stock trading for each company. We evaluated the accuracy of our predictions using the mean absolute percentage error (MAPE) metric. Our results demonstrate that the predicted stock market trading volumes for Microsoft, Amazon, Tesla, Advanced Micro Devices, and Apple have accuracies of 80.07%, 70.00%, 69.41%, 68.89%, and 63.82%, respectively. However, the prediction accuracy for Google was relatively low at 22.33%.

To better understand our results, we plotted our predicted values against the actual values for Microsoft, Amazon, and Tesla, (see Figures 1-3) which had the highest prediction accuracies.

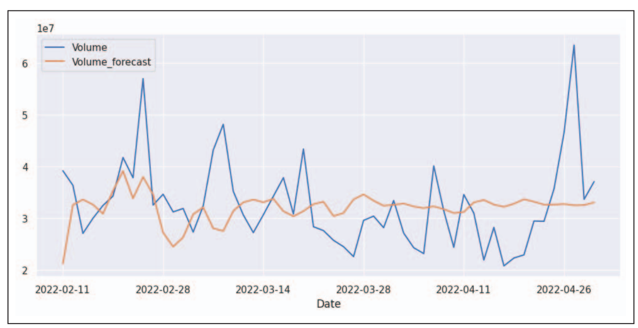

Fig. 1. Predicting Microsoft stock trading volume by emoji usage in Twitter and Compare the Prediction with the Actual Values of Trading Volume. MAPE is 80.07% and the accuracy for the first month is higher.

B. Emoji-Based Detection of Low-Quality Content

We trained multiple classifiers on the existing half of the labeled tweets. Overall, we build a classifier model that outperforms the state-of-the-art model in [30] with an F1 score of 0.8469 compared to 0.8379. Evaluating different classifiers, it was the Extra Trees classifier (see Table III) that achieved the highest performance for predicting low-quality tweets.

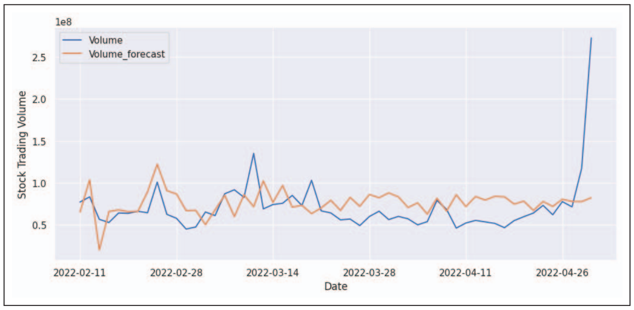

Fig. 2. Predicting Amazon stock trading volume by emoji usage in Twitter and Compare the Prediction with the Actual Values of Trading Volume. MAPE is 70.00% and the prediction is more reliable for the first month.

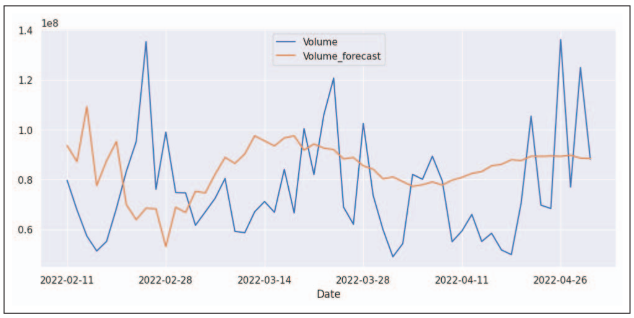

Fig. 3. Predicting Tesla stock trading volume by emoji usage in Twitter and Comparing the Prediction with the Actual Values of Trading Volume. MAPE is 69.41% and the prediction doesn't predict the exact ups and downs of stock Trading Volume, however it shows the market's trend.

Finally, we predicted the quality of tweets in our original Twitter dataset by pre-trained an Extra Trees ensemble model on the labeled tweets leveraged in Chen et al. [30]. The model detected 2,702 tweets as low-quality among 286,505 tweets. In Figure 4, associated emojis with all tweets and low-quality tweets are shown. Figure 4(A) shows the most common emojis in our Twitter dataset and Figure 4(B) shows the emojis that are highly likely to exist in low-quality tweets.

TABLE III
PERFORMANCE OF CLASSIFIER MODELS ON PREDICTION OF LOW-QUALITY CONTENT

Model	Accuracy	F1
ExtraTreesClassifier	0.96	0.84
BaggingClassifier	0.96	0.83
RandomForestClassifier	0.96	0.83
XGBClassifier	0.95	0.83
LGBMClassifier	0.95	0.83
KNeighborsClassifier	0.95	0.81

V. DISCUSSION

Our findings on the effects of emoji usage on tweet popularity metrics might seem surprising and counter-intuitive, but we draw some inference: excessive use of emoji might make a tweet less popular since a user needs to put an additional reader to distinguish the information in the tweet and perceive relevant information, which may make such tweets less interesting compared to the clear and concise statement. Moreover, the results of the constructed regression models suggest that an excess of emojis and their prevalence compared to conventional text may even make this negative effect more apparent. Simultaneously, users might find positively connotated emojis less interesting compared to negative ones, as negative information is more demanded in the media [37].

In terms of predicting stock market trends, our results showed varying degrees of accuracy, with MAPE ranging from 19.93% for Microsoft to 77.67% for Google. While the use of emojis as an input series has shown promise, it may not be the most reliable predictor on its own. To further improve the accuracy of predictions, we suggest exploring the sentiment of the emojis used in tweets and developing a model that incorporates this information. By taking into account the positive or negative connotations associated with the emojis, we can potentially create a more nuanced and effective model for stock market prediction.

The most frequent emojis used in low-quality tweets are well aligned with our intuitive prediction and experience surfing social media, as some emojis like ”kiss mark” and ”heart with arrow” are very popular in that kind of tweet. Our model performs very well on a poor training dataset. However, the number of emojis in tweets and account descriptions as features may not be as important as the type or category that emojis fall into.

VI. CONCLUSION

Overall, it is challenging to make an accurate and statistically justified inference about the effect of emojis on tweet popularity, but, to some extent, we could argue that the negative sentiment of used emojis can help promote a tweet as negative news and ideas often have a powerful response from the audience. This topic has great potential, as we show that emojis and tweets can serve as a convenient system resembling the evolution and success of ideas. The usage of emojis may be related to stock market success, as well as emoji usage in social media can affect public opinion about a company, ultimately affecting stock market volume. A prediction system can help decision-makers find better solutions for advertising campaigns.

Finally, we show that some emojis can be used as an early-warning indicator of low-quality content. Provided that emoji can be readily recognized within a text, this conclusion can be useful for social media users for avoiding potentially harmful content.

REFERENCES

[1] R. Dawkins, *The Selfish Gene*. Oxford University Press, Oxford, UK, 1976.
[2] N. Na'aman, H. Provenza, and O. Montoya, “Varying linguistic purposes of emoji in (Twitter) context,” in *Proceedings of ACL 2017, Student Research Workshop*, Vancouver, Canada, Jul. 2017, pp. 136–141.

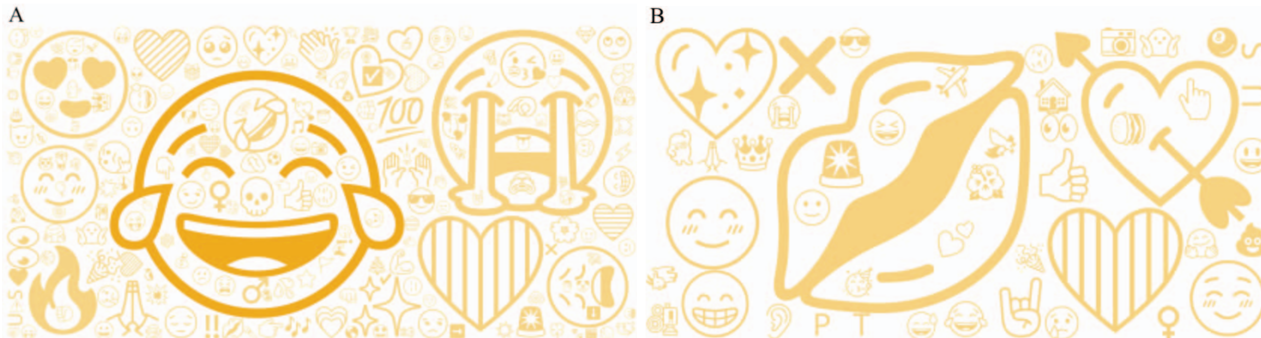

Fig. 4. Emoji clouds of, (A) all tweets in our dataset, and (B) detected low-quality tweets by the trained ML model representing emojis with high-Frequency usage. The size of emojis corresponds to their frequency, with more frequently used emojis displayed in larger sizes.

[3] F. Barbieri, L. E. Anke, and H. Saggion, "Revealing patterns of twitter emoji usage in barcelona and madrid," in *International Conference of the Catalan Association for Artificial Intelligence*, 2016.

[4] M. Kejriwal, Q. Wang, H. Li, and L. Wang, "An empirical study of emoji usage on twitter in linguistic and national contexts," *Online Social Networks and Media*, vol. 24, p. 100149, 2021.

[5] H. Alostad and H. Davulcu, "Directional prediction of stock prices using breaking news on twitter," in *2015 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)*, vol. 1. IEEE, 2015, pp. 523–530.

[6] P. Kralj Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PloS one*, vol. 10, no. 12, p. e0144296, 2015.

[7] F. Barbieri, M. Ballesteros, and H. Saggion, "Are emojis predictable?" *arXiv preprint arXiv:1702.07285*, 2017.

[8] C. Fuchs, *Internet and Society*. Routledge, Dec. 2007.

[9] L. Weng, A. Flammini, A. Vespignani, and F. Menczer, "Competition among memes in a world with limited attention," *Scientific Reports*, vol. 2, no. 1, Mar. 2012.

[10] C. Heath, C. Bell, and E. Sternberg, "Emotional selection in memes: The case of urban legends." *Journal of Personality and Social Psychology*, vol. 81, no. 6, pp. 1028–1041, Dec. 2001.

[11] J. Brody, "How ideas work: memes and institutional material in the first 100 years of the neighborhood unit," *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, vol. 9, no. 4, pp. 329–352, Sep. 2015.

[12] S. Jayarathna and F. Shipman, "Rationale and architecture for incorporating human oculomotor plant features in user interest modeling," in *Proceedings of the 2016 ACM on Conference on Human Information Interaction and Retrieval*, 2016, pp. 281–284.

[13] S. Jayarathna, A. Patra, and F. Shipman, "Unified relevance feedback for multi-application user interest modeling," in *Proceedings of the 15th ACM/IEEE-CS Joint Conference on Digital Libraries*, 2015, pp. 129–138.

[14] G. Brown, "Web culture: using memes to spread and manipulate ideas on a massive scale," *Interface: The Journal of Education, Community and Values*, vol. 13, pp. 187–193, 2013.

[15] T. Highfield and T. Leaver, "Instagrammatics and digital methods: studying visual social media, from selfies and GIFs to memes and emoji," *Communication Research and Practice*, vol. 2, no. 1, pp. 47–62, Jan. 2016.

[16] P. K. Novak, J. Smailović, B. Sluban, and I. Mozetič, "Sentiment of emojis," *PLOS ONE*, vol. 10, no. 12, p. e0144296, Dec. 2015.

[17] R. Karlemstrand and E. Leckström, "Using twitter attribute information to predict stock prices," *arXiv preprint arXiv:2105.01402*, 2021.

[18] G. LĂZĂROIU, "The role of social media as a news provider." *Review of Contemporary Philosophy*, vol. 13, pp. 78 – 83, 2014.

[19] V. S. Pagolu, K. N. Reddy, G. Panda, and B. Majhi, "Sentiment analysis of twitter data for predicting stock market movements," in *2016 international conference on signal processing, communication, power and embedded system (SCOPES)*. IEEE, 2016, pp. 1345–1350.

[20] T. Rosenstiel, J. Sonderman, K. Loker, M. Ivancin, and N. Kjarval, "Twitter and the news: How people use the social network to learn about the world," *Online at www. americanpressinstitute. org*, 2015.

[21] M. Osborne and M. Dredze, "Facebook, twitter and google plus for breaking news: Is there a winner?" in *Eighth International AAAI Conference on Weblogs and Social Media*, 2014.

[22] S. Jayarathna, A. Patra, and F. Shipman, "Mining user interest from search tasks and annotations," in *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, 2013, pp. 1849–1852.

[23] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 2013, pp. 99–127.

[24] M. Song and M. C. Kim, "Rtˆ 2m: Real-time twitter trend mining system," in *2013 International conference on social intelligence and technology*. IEEE, 2013, pp. 64–71.

[25] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *Journal of computational science*, vol. 2, no. 1, pp. 1–8, 2011.

[26] Y. Mao, W. Wei, and B. Wang, "Twitter volume spikes: analysis and application in stock trading," in *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, 2013, pp. 1–9.

[27] B. Wang, A. Zubiaga, M. Liakata, and R. Procter, "Making the most of tweet-inherent features for social spam detection on twitter," *arXiv preprint arXiv:1503.07405*, 2015.

[28] A. Aggarwal, A. Rajadesingan, and P. Kumaraguru, "Phishari: Automatic realtime phishing detection on twitter," in *2012 eCrime Researchers Summit*. IEEE, 2012, pp. 1–12.

[29] H. Gupta, M. S. Jamal, S. Madisetty, and M. S. Desarkar, "A framework for real-time spam detection in twitter," in *2018 10th international conference on communication systems & networks (COMSNETS)*. IEEE, 2018, pp. 380–383.

[30] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "A study on real-time low-quality content detection on twitter from the users' perspective," *PloS one*, vol. 12, no. 8, p. e0182487, 2017.

[31] S. Sazzed and S. Jayarathna, "A sentiment classification in bengali and machine translated english corpus," in *2019 IEEE 20th international conference on information reuse and integration for data science (IRI)*. IEEE, 2019, pp. 107–114.

[32] S. Sazzed and S. Jayarathna, "Ssentia: a self-supervised sentiment analyzer for classification from unlabeled data," *Machine Learning with Applications*, vol. 4, p. 100026, 2021.

[33] I. D. Currie, "On fitting generalized linear and non-linear models of mortality," *Scandinavian Actuarial Journal*, vol. 2016, no. 4, pp. 356–383, 2016.

[34] R. A. Davis, P. Zang, and T. Zheng, "Sparse vector autoregressive modeling," *Journal of Computational and Graphical Statistics*, vol. 25, no. 4, pp. 1077–1096, 2016.

[35] R. Mushtaq, "Augmented dickey fuller test," 2011.

[36] S. I. Vrieze, "Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic)." *Psychological methods*, vol. 17, no. 2, p. 228, 2012.

[37] M. Trussler and S. Soroka, "Consumer demand for cynical and negative news frames," *The International Journal of Press/Politics*, vol. 19, no. 3, pp. 360–379, Mar. 2014.