Application of Lotka's Law to Study the Retweets on #Research

Jenita Shiloh V
Research Scholar
Department of Library and Information Science,
University of Madras

Fazlunnisa H
Assistant Professor
Department of Library and Information Science,
University of Madras, Chennai,
E-mail: fazul_8383@yahoo.com

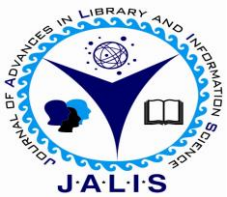
Abstract

Twitter is an extraordinary microblogging tool used by millions of users regularly. Through its primary purpose is socially connecting users and helping them share their thoughts, it has become an excellent tool for aggregating research by researchers and academicians globally. The present study collates data from Twitter using #research shared on a particular date using data mining techniques. The collected data is analysed based on retweets for the country of origin, accompanying hashtags, and most popular tweets, and the content of the tweets is also analysed. It is astonishing that though there are 1584 unique tweets, only 425 are retweeted. There are about 274 retweets that are retweeted at least once, while there are only 4 retweets that are repeatedly shared by other users at least more than 10 times. The study investigates the usage of #research on Twitter to glean insight into the content of the majorly shared tweets.

.

Keywords
Twitter; hashtags; research; retweets

Electronic access
The journal is available at www.jalis.in
DOI: **10.5281/zenodo.14877550**



Journal of Advances in Library and Information Science
ISSN: 2277-2219 Vol. 14. No.1. 2025. pp.12-19

1. Introduction

A population of 350 million is actively using Twitter daily. Like the bird in the logo representing the short chirping of birds, tweets are meant to be short conversations. Nevertheless, highly entropic and transmits rich information to the user community. Originally the text content of tweets was designed for a character limit of 140 (Boyd et al., 2010). Since November 2017, the character limit has been increased to 240 (Rosenberg et al., 2020) to prevent cramming while tweeting. Twitter started as a short messaging service to connect with like-minded people. It is one of the most popular social networking tools as it keeps its user interface incredibly basic. This medium has helped the user community connect with a global network of enthusiastic people ready to share information – including new information, updates, and resharing in all aspects of human life and behaviour. Twitter enhances social networks' potential to reach out, broaden awareness, and deepen knowledge spheres while participating in and becoming a part of different communities (Java et al., 2007). Using Twitter allows the user to follow the latest news, sail the trendline, build a brand, upskill and monetize, establish a reputation, initiate revolutions, and much more. However, the unique format has kept Twitter ranked as the best microblogging site. Twitter's competitors have still not found a way to keep up with the game. The best example is the use of the hashtag symbol (#). No other network works as well with hashtags as Twitter does (Skaza & Blais, 2017), of categorising tweets using keywords or phrases. This extraordinary tool is built with an exclusive algorithm that displays content about the user's interests rather than feeding random content. Speaking in the sense of information retrieval, increasing the precision than recall ratio in any query. Concisely, it keeps the user feed engaging as per likes, views, follows, and shares of the user. Thus, making Twitter the real winner among social networking sites. It is also noted that Twitter's actual visitors are three times more than its registered users. It has a 78% engagement rate growth – almost 10 times more than Instagram and 40 times more than Facebook. Twitter's user base is the second most educated on social networking tools, topped only by LinkedIn's users. Despite the impressive statistics, Twitter is useful for other reasons outside the numbers.

Twitter has some extraordinary features, such as other social media features, that allow users to tweet, retweet, mention, comment, follow, pin, and

hashtag. Some users while tweeting, create and share original content – in any form as a text, image, audio, or video post; some prefer to retweet. When a user posts and someone replies on Twitter, it is referred to as a tweet. Forwarding of tweets is known as retweets (Kwak et al., 2010). The retweet system helps pass on the message and gain strength on the specific tweet. Followers and non-followers can see tweets related to their interests on their feeds. The '#' symbol is followed by a word called 'hashtag.' Twitter users can also search and retrieve tweets using #hashtags. Hashtags also help highlight the topics trending for the day. Researchers who prefer to stay updated on a particular topic can simply follow their interest using a hashtag. They will be intimidated regarding the relevant tweets when any user tweets using the particular hashtag. Similar to the # symbol which is all about creating user-defined keywords, the '@' symbol has a beautiful feature that can be used to 'mention' a particular page or another Twitter account, thereby linking that account to the current page. Again, this is a very useful tool as it helps in connecting people and build collaborations among like-minded users. Thus, Twitter can be a very valuable social networking tool that helps researchers create new content, partake in constructive conversations, build collaborations, and share research-related information – publication/funding /training and other opportunities.

Twitter and research are inseparable as it help connect people to share their viewpoints. Twitter has a good potential to spread research and help in receiving an impact. In terms of helping researchers, Twitter has provided Application Programming Interfaces (APIs) to enable data collection. To access the API, a user needs to apply for a Twitter developer account. Once the application is approved, the user can access four keys: 1. Consumer key, 2. Consumer secret, 3. Access token, and 4. Access secret. These keys aim to authenticate the user to access Twitter data such as tweets and profile information. Twitter's API is the most powerful tool available for collecting data generated by Twitter user interactions. Twitter data is a diverse and important data source for researchers and policymakers. Twitter is seen to be an extraordinary tool to predict the future popularity of a topic by exploring the previous and current usage of a hashtag. In this study, the application of content analysis helps to determine the concept behind tweets.

Review of literature

Apart from being widely used as the fastest medium of information transfer, Twitter is also a viable place to carry out exploratory studies. Various academicians and researchers have investigated Twitter's potential (Ruiz-soler, 2017). Twitter is home to many health-related professionals (Lee et al., 2014) and is a suitable research tool for many researchers (Ovadia, 2009). On average, 6000 tweets per second, which makes 500 million tweets per day (Skaza & Blais, 2017). A user can tweet about 2400 tweets per day including the retweet count. Despite the 280-character limit on Twitter, it offers basic but effective ways to link tweets to more general topics, individuals, and organisations (Murthy, 2013). Twitter was mostly utilized in 2009 to disseminate factual information about the H1N1 virus, but it was also a source of viewpoints and first-hand reports. The study by Chew & Eysenbach, (2010) shows the possibility of using social media for "Infodemiology" research on public health. The content analysis shows that Twitter is not just a tool for dissemination but serves as a source for understanding user's opinions in real time. In a related study, researchers examined the keywords "depressed", "#depressed", "depression," or "#depression" on Twitter;

The study is a boon to many mental health professionals to identify their target audience and reach them at ease as it is observed that many who have tweeted using these keywords were identified as individuals in young age through the demographics available (Cavazos-Rehg et al., 2016). Likewise, 149 tweets were analyzed in a study carried out during the pandemic to find that many tweeters with arthritis condition were identified with anxiety while trying to get connections from peers for medicine (Berkovic et al., 2020). Another study carried out to understand the freshman experience suggests that Twitter provides a valuable source of real-time data for monitoring academic and social experiences (Sam Liu et al., 2018). In a similar content analysis, the study analyzed tweets regarding the 'dabbing' of marijuana and its consequences. In January 2015, Five 5,000 tweets were chosen at random. It was discovered that 71% of them had to do with dabbing marijuana concentrates, among which strong highs, and overdabbing were common themes. There have been reports of physiological and psychological repercussions, the most common being breathing problems and passing out. The most frequently reported respiratory effect was coughing, and several users even indicated that they intended to pass out while dabbing. While the research adds to our understanding of the effects of marijuana

concentrates, further work is necessary to determine the extent of the health risks connected with them (Cavazos-Rehg, et al., 2016).

The studied literature indicates that there is a need to explore the use of Twitter not just as a platform to share information, but as a universal database to get a gleaning insight into the minds and emotions of people sharing the information.

2. Significance and scope of the study

the scope and coverage of this paper are mentioned to limit the study to better understand the working of social media tools in the research and development arena. The scope and its significance are discussed in detail in the following section:

1. Microblogging Site: Twitter

Web 2.0 provides universal access to the world of scientific information, leading to increased academic productivity. Recognizing social media as a channel of dissemination and an enhancer of academic productivity is key to unlocking the potential of this channel. While many social networking sites let users express their opinions in paragraphs or pages, Twitter has a character limit per tweet, which promotes brief communication and makes it easy for users to exchange ideas and information rapidly. Twitter helps users to stay informed in real-time with rapid access to global news and trends. Aiding connections with a multicultural community worldwide, participating in conversations, and exchanging viewpoints on a variety of subjects (Curlin et al., 2019). By utilizing Twitter API one can access large datasets to study the trends, sentiments, public opinions by various analyses. Hence this study, in line with many more as projected in the literature review aims to focus and highlight the impact of the microblogging site Twitter on the research arena.

2. Hashtag / Keyword used: #research

Though there have been studies to understand a particular hashtag or a word such as #depression (Cavazos-Rehg et al., 2016), #rheumatology (Abbasi-Perez et al., 2021), covid-19 vaccine (Siru Liu & Liu, 2021), HINI Virus (Chew & Eysenbach, 2010), there is no such found on # research. The academic wold revolves around research and development studies. It is pertinent that # research is chosen to understand and map the extent and reach of SMT in promoting research activities.

There have been numerous studies that indicate Twitter to be a popular social media tool for information dissemination (Antoniadis et al., 2015), with an average of 6000 tweets per second, which sums up to 500 million tweets per day (Sayce, 2022). It is understood that the world of social media is constantly generating information at a tremendous rate and speed, data was extracted to quantify and visualize the pattern of information sharing on this media.

3. Objectives of the study

The major objectives of the study is to:

- Quantify the number of tweets on *#research* and determine the quantum of retweets for the chosen period.
- Analysing the information-sharing behaviour using retweets
- Spatial mapping of the retweets to identify the most productive countries.

4. Methodology

The data for the present study was collected from Twitter using the Twitter API. All tweets on #research were retrieved. For the study purpose and to achieve the mentioned objectives, the data extracted is limited to the following criteria:

Keywords / Hashtags used	: #research
Language	: English
Type of Tweet	: Retweets

The data collection and curation is represented in the following chart for easy understanding. The following section presents the data collected using appropriate tables and charts and discusses the findings.

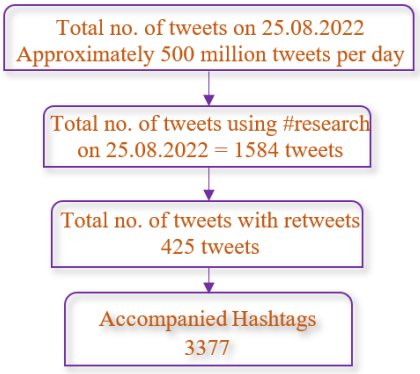


Figure 1: Data collection steps

The retrieved data was incorporated into an MS Excel for further analysis.

5. Analysis and Interpretation of Data:

The search using #research on Twitter yielded 1584 tweets as of August 25th, 2022 out of which other users retweeted 425 tweets. The study proceeds further with an analysis of the 425 retweeted tweets. Table 1 indicates the number of times a tweet is retweeted. It is observed that 4 unique tweets are retweeted by 26, 22, 19, and 13 Twitter account holders, other than the ones who posted it originally, on a single day. It is also seen that 274 tweets are retweeted just once. Retweets symbolize a conscious process by the user/Twitter account holders' views, opinions, sentiments, and beliefs to the information shared through the tweet.

Table 1: Tweets - Retweets count relativity

S. No.	Unique Individual tweets	Maximum number of times retweeted
1	1	26
2	1	22
3	1	19
4	1	13
5	2	9
6	2	7
7	4	8
8	7	6
9	7	5
10	11	4
11	41	3
12	73	2
13	274	1

Alfred Lotka's law on author productivity was used to determine the relationship between the number of retweets. Accordign to Lotka's law more number of individual authors produce less number of articles compared to less number of individual author who contribute to more research papers. Similarly it is found using the below correlation that one 1 tweet is retweeted more number of times on # research indicating that it is more relavant. Also it is noted that 274 individual tweets were only retweeted once.

5.1. Retweets and Lotka's law - correlational study

Let's denote:

- T_1 as the number of tweets retweeted once,
- T_{26} as the number of tweets retweeted 26 times.

Given:

- $T_1 = 274$
- $T_{26} = 1$

Lotka's law suggests that $Tn \propto 1/n^2$. For n=26 and n=1:

$$T_{26} \propto 1/26^2$$
$$T_1 \propto 1/1^2$$

To find the proportionality constant k:

$$T_{26} = k.\ 1/26^2$$
$$T_1 = k.1/1^2$$

Solve k using the value T_1:

$$274 = k.\ 1$$
$$K = 274$$

Calculating T_{26} by using the value of k.

$$T_{26} = 274.1/26^2$$
$$T_{26} = 274.1/676$$
$$T_{26} = 0.405$$

This theoretical value from Lotka's law ($T26 \approx 0.405$) is much less than the observed value ($T26=1$).

In summary:

- Lotka's law predicts fewer tweets being retweeted 26 times (0.405) compared to the observed data (1).
- The observed data indicates a slightly heavier tail than Lotka's law, meaning more tweets are retweeted 26 times than predicted.

Figure 2 shows the tweets that have received the most retweets under the hashtag #research. It is clear from this figure that the tweets are from company accounts rather than individual profiles. The tweet with 26 retweets originated from a verified account belonging to the Country USA, named NOAA Research, with 92K+ followers. The tweet talks about the ozone layer and the depletion of harmful chemicals. This is followed by a tweet about registration for a European research event with 22 retweets from a verified account with 1 Lakh+ followers. The subsequent tweet is about a mapping method posted by a company focused on genome and mining with 19 retweets and 538 followers in the account.

Another tweet from FlagMedia account is on partner benefits with 13 retweets and 645 followers originating from USA.

Screen Name	Full Name	Tweet Text	Retweets	Favorites	App	Followers	Follows	Verified	User Since	Location	Bio
@NOAAR	NOAA Rese	Good news for the ozone layer! An annual analysis tracking the decline of ozone-depleting chemicals shows the levels of these harmful chemicals have significantly dropped, reaching a milestone in the recovery of the #ozonelayer. #noaa #research	26	81	Twitter W	92923	531	☑	9/14/2011	US	NOAA Research provides the research foundation for understanding our planet and technological innovation and scientific advances that improve our lives.
@EUScier	EUScience	📢 Did you know that the registration for the #RiDaysEU is already open? Do not miss Europe's main #research event on 🗓 28 - 29 September. 👀 Join policymakers, researchers & entrepreneurs from all over Europe 🇪🇺 and let's shape the future together! 💡 👍	22	29	Twitter W	119735	750	☑	9/21/2010	European L	Official account of DG Research & Innovation @EU_Commission, managing @HorizonEU & implementing Commissioner @GabrielMariya strategy. Follow also @JEPaquetEU
@Mining(	GenomeMir	Comparison of read mapping and variant calling tools #retweet #OpenScience #DataScience #100DaysOfCode #WomenWhoCode #BigData #MachineLearning #serverless #Linux #Python #coding #SciComm #phdchat #Rstats #Bioinformatics #science #research #genomics #edu	19	10	Twitter W	538	284		7/23/2020		Interested in #genomics #bioinformatics #plants #metabolism
@FlagMe	Flag Media	PARTNER BENEFITS from @flagmediagroup 🌎 -- #FlagMedia #Online #NewsSource #Research #Inspo #Design	13	47	Loomly	645	54		7/20/2021	Frisco, TX	We own dozens of digital publications + PR Daily Wire • Full service #PR #SocialMedia #Marketing agency • #credibility #authority #reputation #engagement

Figure 2 : Tweets with a higher number of retweets



Figure 3: Frequency of accompanied hashtags in a word cloud

As shown in Figure 3, a word cloud based on frequency represents each hashtag. A total of 3077 accompanying hashtags were discovered in the tweets which were collected using the #research hashtag. Applying Zipf's law of word occurrence to the hashtags in this study reveals that the hashtags commonly and repetitively used are assignments, data science, python, programming, and thesis. Zipf's law helps to bring the relation between word frequency and rank.

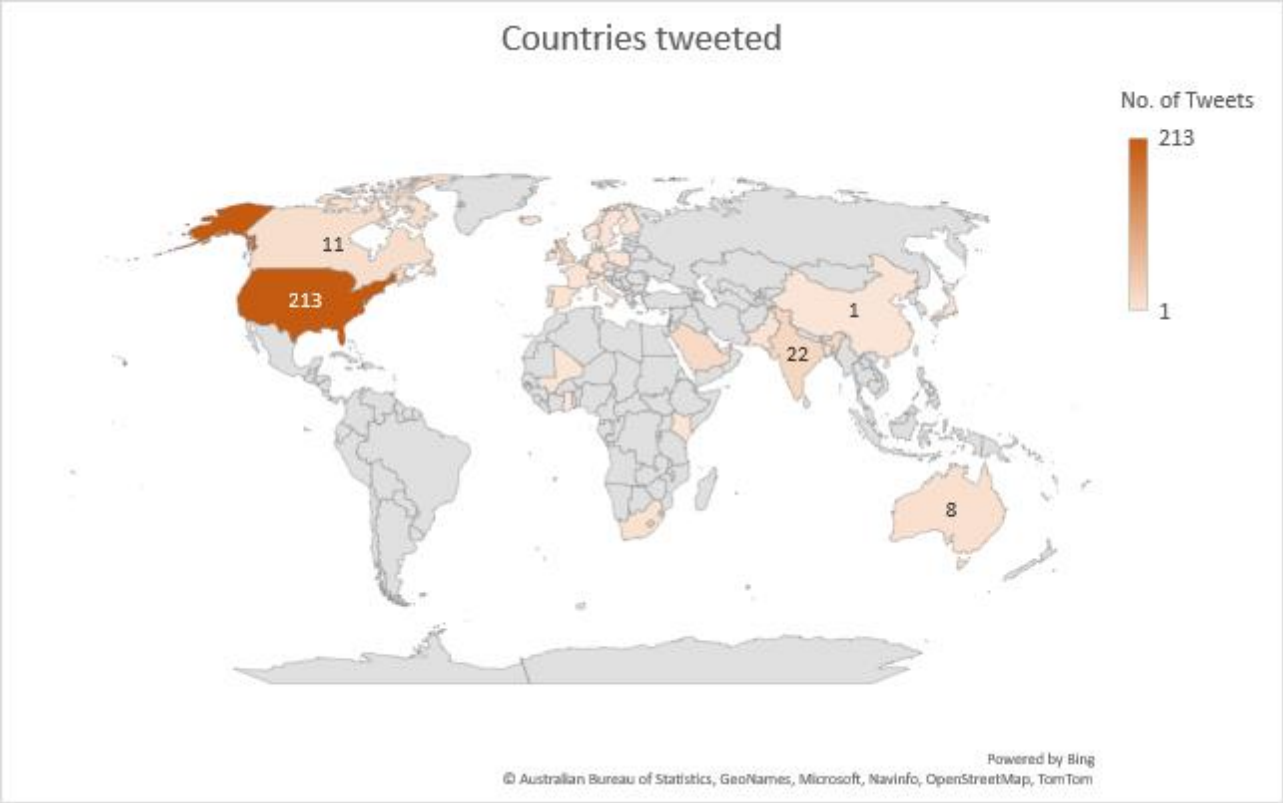


Figure 4: Global representation of the tweeting nations

Figure 4 represents where the tweets originated from. Among 425 retweets analyzed, only 357 tweets have the location identity. It is found that most of the tweets related to #research originated from the United States of America with 213 tweets, followed by the United Kingdom with 29 tweets, and India with 22 tweets. Whereas, countries like China, Japan, Italy, Finland, Scotland and few others are identified with a single tweet on *#research* as shown in Table 2.

Table 2: Rank-wise distribution of countries based on the number of tweets

Rank	Country Name	No. of Tweets
1.	United States of America	213
2.	United Kingdom	29
3.	India	22
4.	Saudi Arabia	16
5.	Canada	11
6.	Ireland	10
7.	Australia	8
8.	South Africa	6
9.	Belgium	4
10.	Denmark	3
11.	Spain	3
12.	Austria	2
13.	England	2
14.	France	2
15.	Germany	2
16.	Pakistan	2
17.	Poland	2
18.	Portugal	2
19.	Rwanda	2
20.	Sweden	2
21.	Switzerland	2
22.	Bangladesh	1
23.	China	1
24.	Finland	1
25.	Ghana	1
26.	Iceland	1
27.	Italy	1
28.	Japan	1
29.	Kenya	1
30.	Mali	1
31.	Norway	1
32.	Scotland	1
33.	United Arab Emirates	1

Among the 213 tweets by the country USA, 151 tweets were identified as tweets from companies who are solely helping the research community.

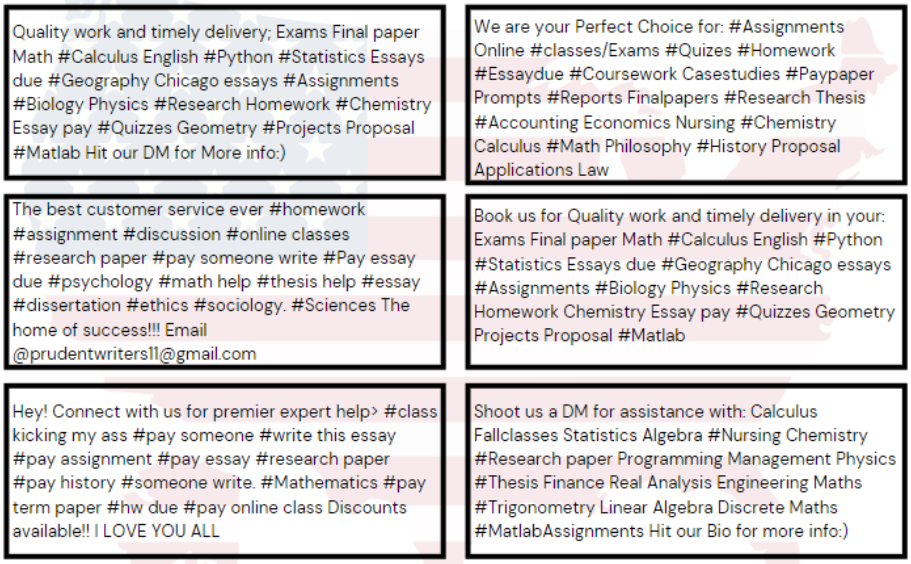


Figure 5: Company tweets from USA

A few examples of tweets from different companies with the United States of America as the location are represented in Figure 5. The company tweets are about helping or assisting the research community by writing essays, and articles, or finishing their assignment. It is to be noted that all these services by the company are paid services. It is also observed that the company offers help to every field of study, and aids assistance to complete assignments to doctoral thesis.

6. Discussion and Conclusion

The present study has been an eye-opener to study the use of the social media platform Twitter for sharing and promoting research activities. It is observed that though a negligible number of tweets are shared on #research, 20% have been retweeted, which indicates that academicians and researchers are actively involved on the Twitter platform. Also, Lotka's inverse square law indicates specific dynamics of the Twitter platform, including viral trends, and social networking effects which are unknown in the academic scenario earlier. Nonetheless, the general inverse-square relationship still provides a rough approximation of the retweet distribution. It is observed that since the study focuses on #research tweets in the academic scenario, tweets not using the chosen hashtag could have been neglected, this amounts to the limitation of the study. The researchers propose to make #research into a trending and viral hashtag to assimilate and study the phenomenon on Twitter. It is to be noted that the businesses promote the use of #research in all of their tweets. It is no surprise that the USA and the UK are the leading retweeters indicating wide social media presence, it is surprising to see that India is in the 16th position in the Research Leadership Index, according to the report on Research Fronts by Clarivate Analytics (Hoffman, n.d.), is the 3rd largest retweeting nation. This shows that there is considerable social media presence that can be tapped to gain more research visibility for the Indian researchers.

References

1. Abbasi-Perez, A., Alvarez-Mon, M. A., Donat-Vargas, C., Ortega, M. A., Monserrat, J., Perez-Gomez, A., Sanz, I., & Alvarez-Mon, M. (2021). Analysis of tweets containing information related to rheumatological diseases on Twitter. *International Journal of Environmental Research and Public Health*, *18*(17), 18179094. https://doi.org/10.3390/ijerph18179094
2. Antoniadis, K., Zafiropoulos, K., & Vrana, V. (2015). Communities of Followers in Tourism Twitter Accounts of EuropeanCountries. *European Journal of Tourism Hospitality and Recreation*, *6*(1), 11–26.
3. Berkovic, D., Ackerman, I. N., Briggs, A. M., & Ayton, D. (2020). Tweets by people with arthritis during the COVID-19

pandemic: Content and sentiment analysis. *Journal of Medical Internet Research*, *22*(12), 1–17. https://doi.org/10.2196/24550

4. Boyd, D., Golder, S., & Lotan, G. (2010). Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. *2010 43rd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2010.412
5. Cavazos-Rehg, P. A., Sowles, S. J., Krauss, M. J. ., Agbonavbare, V., Grucza, R., & Bierut, L. (2016). A content analysis of tweets about high-potency marijuana. *Drug and Alcohol Dependence*, *166*, 100–108. https://doi.org/https://doi.org/10.1016/j.drugalcdep.2016.06.034
6. Cavazos-Rehg, P. A., Krauss, M. J., Sowles, S., Connolly, S., Rosas, C., Bharadwaj, M., & Bierut, L. J. (2016). A content analysis of depression-related tweets. *Computers in Human Behavior*, *54*, 351–357. https://doi.org/10.1016/j.chb.2015.08.023
7. Chew, C., & Eysenbach, G. (2010). Pandemics in the age of Twitter: Content analysis of tweets during the 2009 H1N1 outbreak. *PLoS ONE*, *5*(11), 1–13. https://doi.org/10.1371/journal.pone.0014118
8. Curlin, T., Jaković, B., & Milološa, I. (2019). Twitter usage in Tourism: Literature Review. *Business Systems Research*, *10*(1), 102–119. https://doi.org/10.2478/bsrj-2019-0008
9. Java, A., Song, X., Finin, T., & Tseng, B. (2007). Why We Twitter: Understanding Microblogging Usage and Communities. *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis*, 56–65. https://doi.org/10.1145/1348549.1348556
10. Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 591–600. https://doi.org/10.1145/1772690.1772751
11. Lee, J. L., DeCamp, M., Dredze, M., Chisolm, M. S., & Berger, Z. D. (2014). What are health-related users tweeting? A qualitative content analysis of health-related users and their messages on Twitter. *Journal of Medical Internet Research*, *16*(10), e237. https://doi.org/10.2196/jmir.3765
12. Liu, Sam, Zhu, M., & Young, S. D. (2018). Monitoring freshman college experience through content analysis of tweets: Observational study. *JMIR Public Health and Surveillance*, *4*(1), 1–8. https://doi.org/10.2196/publichealth.7444
13. Murthy, D. (2013). *Twitter: Social Communication in the Twitter Age*. Polity press.
14. Ovadia, S. (2009). Exploring the Potential of Twitter as a Research Tool. *Behavioral & Social Sciences Librarian*, *28*(4), 202–205. https://doi.org/10.1080/01639260903280888
15. Rosenberg, H., Syed, S., & Rezaie, S. (2020). The Twitter pandemic: The critical role of Twitter in the dissemination of medical information and misinformation during the COVID-19 pandemic. *Canadian Journal of Emergency Medicine*, *22*(4), 418–421. https://doi.org/10.1017/cem.2020.361
16. Sayce, D. (2022). *The Number of tweets per day in 2022*. https://www.dsayce.com/social-media/tweets-day/
17. Skaza, J., & Blais, B. (2017). Modeling the infectiousness of Twitter hashtags. *Physica A: Statistical Mechanics and Its Applications*, *465*, 289–296. https://doi.org/https://doi.org/10.1016/j.physa.2016.08.038