Using hashtag and word graph networks to analyze trending social media topics and communication

Andrew Curtis
Utah State University
December 15, 2022

***Abstract*—Social media usage is ubiquitous in current culture. Trends and social commentary arise and die and act as a stage for communication and discourse. These trends not only help to consolidate and transmute ideas but also play a role in influencing change in society outside of social media. Analysis of social media networks can come from many different perspectives and a variety of methods can be employed. This research offers an analysis of using hashtag and word graph networks generated from tweets to analyze trending social media topics and communication. Findings of the research highlight the usefulness of these methods in capturing co-creational terms and communication, as well as the challenges of analyzing the public's interaction with social media. Several guidelines and suggestions are given for future research on the topic.**

***Index Terms*—Social Media, Graphs, Networks, Co-Creational**

I. Introduction

The social media platform of Twitter utilizes hashtags as a method of indexing posts, called tweets. These hashtags can sometimes be a more useful method of grouping ideas and consolidating communication than simple liking, commenting and sharing. For this analysis, both hashtags and words contained in the tweets themselves were used to generate graph networks. Both approaches yield different results when applied to differing topics of interest.

A. Related Work

A useful definition and perspective to take when conducting this research is co-creation. Co-creation is creating shared meaning and interpretation of terms between differing groups [1]. In this case it can be any entities on the social media platform. Often it can be useful to further break down these entities into communities of interest. One way of dividing these entities is between organizations, like news or entertainment corporations, and the general public. These two groups share information between them and co-create terms. These terms are sometimes hashtags that then organize further communication. Thoughts and ideas on a topic are prone to rapidly arising and dying but serve to elucidate key themes within viral movements.

B. Chosen Topics of Interest

Three topics of interest were chosen to serve as the background for this analysis. Topics were chosen to be representative of trending and culturally significant events and ideas. They were also selected to be on a variety of subject matter. In practice, the subject matter and different communities of people that interacted with the topics greatly influenced the results of the analysis methods applied. The first topic was on the National Football Leauge (NFL) and concussions. Over the 2022 NFL season, there were several high-profile cases of players suffering from concussions and CTE-related injuries. Controversy ensued over the treatment and protocol of dealing with these injuries. Analysis of this topic could reveal sentiment of the public on the issue and the way sports organizations and news outlets cover the issue.

The second topic was on Elon Musk and Twitter. In 2022, Elon Musk purchased ownership of Twitter for $44 billion dollars. His leadership has so far proven to be divisive and controversial at the least. Many of the public lament his choices and believe it is leading to the downfall of the social media company. Others however, praise his proposed efforts to bring free speech back to the platform.

The third topic deals with Kanye West and specific capture of tweets in the morning following his appearance on the Alex Jones podcast (1 December 2022). On the podcast, a hooded and masked Kanye West made a series of extreme anti-semitic comments and shared approval for Nazis and Adolf Hitler. The event sparked a high amount of related posts and engagement on the Twitter platform. It also serves as an excellent example of rapid co-creation of terms as well as their often short lifespans.

II. Method

A. Data Collection

The methodology for data collection involved pulling 3000 tweets for each topic. Multiple methods were employed on limiting search terms for the tweets. Tweets were pulled if a "search term" word or hashtag appeared in them. One word and one hashtag search term approaches were tested, as well as two search terms. For the two search terms they were pulled if they both appeared in the same tweet. Tweets were pulled for all topics over a date range of the 1st-3rd of December 2022. Efforts were made to reduce the amount of "stop" words and similar words in the analysis in order to highlight terms that enhance the context and themes of the communication occurring.

For graph construction, two methods were used for selecting nodes. The more simple method tested was having nodes be words or hashtags appearing in the same tweet as the search terms. The second method involved having these nodes be bi-grams found in tweets with the search terms appearing.

nodes were words or hashtags appearing the same tweet as the search terms. Bi-grams are where two words or features appear adjacent to each other in a sequence (i.e. two words in a row). This was done in order to attempt to capture a different more complex set of information than singular words. This method could perhaps shed a little more light on co-created terms and ideas that do not appear singularly. Edges were constructed between words, hashtags or bi-grams that appear together in other tweets.

The Kanye West topic results feature two sets of data taken four hours apart for comparison. This method was in an effort to capture the rapid change in trends and co-creational terms being used at the time.

B. Analytical Measures

Several qualitative and quantitative measures were investigated in order to try and characterize the data and communities. From these methods, the research hopes to capture some information on co-creational terms and trends of the topics of interest as well as information on the communities involved and their sentiments on the issues. Several methods were used to create different visual graphical representations of the networks. For the NFL and Elon Musk topics, connected bi-gram graphs were created using the 40 most common bi-grams appearing in the tweets. Two sets of ego-centric word graphs were created for the Kanye West topic from the two sets of four hour separated Twitter tweet pulls.

For quantitative measures, common word graphs were created for each topic. The bar graphs are the simple count of words appearing in the pulled tweets based on the search terms. These tend to capture main subjects of conversation as well as some co-created terms of the communities of interest. Similarly, normalized degree centrality was run on the created graph networks and the results are displayed. These return a similar view of popular terms found in the tweets.

Sentiment analysis of the tweets was conducted using a pre-trained natural language processing model. Histograms were created for each topic. For these histograms, one is considered to be a very positive sentiment, 0 to be neutral and -1 to be a very negative sentiment. The total sentiments from tweets graphs show the total tweets pulled and their results. The polarized sentiments from tweets graphs had the neutral or 0 sentiment bin removed and show only the remainder of the tweets that were polarized. This was done in order to clarify the total sentiment on a topic but also to show if there was a trend or set of outliers for a certain topic.

III. Experimental Results

The following sections are broken down by topic of interest. Qualitative and quantitative measures are displayed and the results are discussed in context for that topic.

A. NFL and Concussions

Many search term methods were tested when looking at NFL and Concussions. Pulling by hashtags proved to be fairly ineffective as they were none that were consistently being used to communicate about the topic. Organizations such as teams and sports journalists were the main communities posting on this topic with reduced amounts of engagement from Twitter accounts representing individuals and the public. These organizations did not appear to be effective at co-creational engagement with the public. Many tweets simply involved player injury updates. For the results below, the search words of NFL and concussion co-appearing in the tweet were used for scraping.

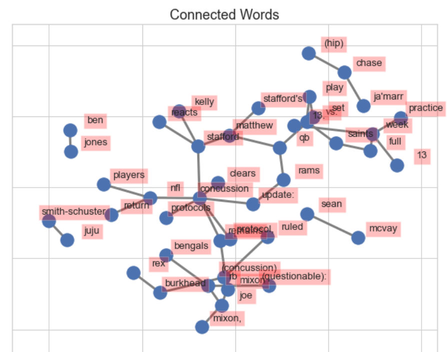

Fig. 1. 40 most connected bigrams within Tweets

It can be seen from the common word graph (Fig.2) and the degree centrality table (Fig.3) that the subject of the tweets are mostly player and team related. The purpose of the tweets was mostly to dispense some information rather than drive any sort of communication or change. Co-created terms are mostly not present.

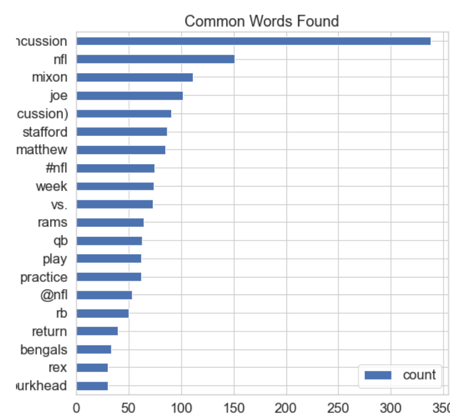

Fig. 2. NFL Most Common words among 3000 Tweets

The total sentiments from tweets analysis on the NFL data shows the large majority of tweets lie in the neutral histogram bin (Fig.4). Among the polarized tweets, there were few extreme negative sentiments but the graph appears to have a right skew overall.

	Words	Normalized Degree Centrality
0	concussion	0.162791
1	stafford	0.093023
2	week	0.093023
3	joe	0.069767
4	mixon	0.069767
5	(concussion)	0.069767
6	matthew	0.069767
7	qb	0.069767
8	nfl	0.069767
9	rb	0.069767
10	rams	0.046512
11	play	0.046512
12	vs.	0.046512
13	burkhead	0.046512
14	set	0.046512
15	update:	0.046512
16	saints	0.046512
17	full	0.046512
18	remains	0.046512
19	chase	0.046512
20	rex	0.023256
21	practice	0.023256
22	kelly	0.023256
23	reacts	0.023256
24	return	0.023256
25	13	0.023256
26	protocol,	0.023256
27	stafford's	0.023256
28	ben	0.023256
29	jones	0.023256
30	ja'marr	0.023256
31	(hip)	0.023256
32	bengals	0.023256
33	clears	0.023256
34	mixon,	0.023256
35	ruled	0.023256
36	juju	0.023256
37	smith-schuster	0.023256
38	protocols	0.023256
39	players	0.023256
40	(questionable):	0.023256
41	13.	0.023256
42	sean	0.023256
43	mcvay	0.023256

Fig. 3. NFL Tweet Degree Centrality

B. *Elon Musk and Twitter*

The selected search terms for the results of this section were if #Elon or #ElonMusk appeared with the word or hashtag of Twitter. The content of words and themes captured in these tweet searches were very temporally-related. A search pulling tweets from a couple hours difference produced substantially unique results. For example most tweets shown in the results relate to rumors of Twitter's possible ban from the Apple app store, whereas hours later the topic of conversation had shifted to commentary about free speech on the platform.

The connected words graph (Fig.5) reveals several themes occurring the conversation around Elon Musk and Twitter. Among these, Tim Cook and the app store frequently appear. The communities present in tweets on this topic seemed to be diverse. More individuals were present in this topic compared to the NFL and concussions. Free speech was also a common thread in the conversations. Interestingly co-creational terms were not as present in these tweets but the issues were still well defined by more traditional words and the issues within

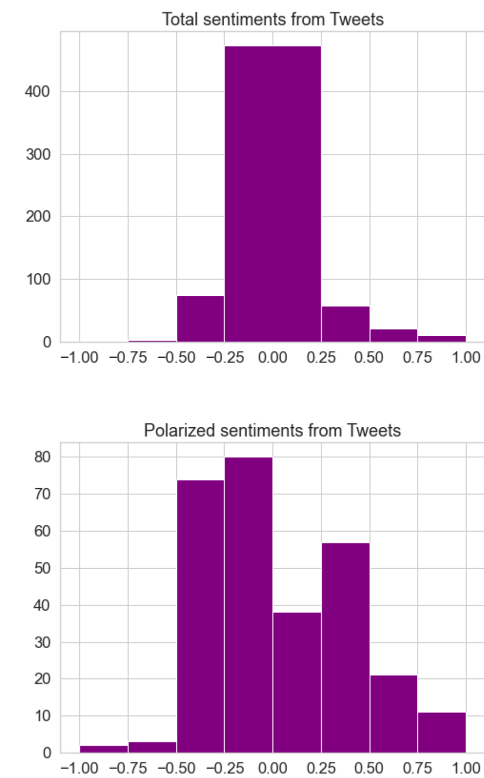

Fig. 4. NFL Tweet Sentiment

the topic, such as free speech and the potential ban from the apple app store.

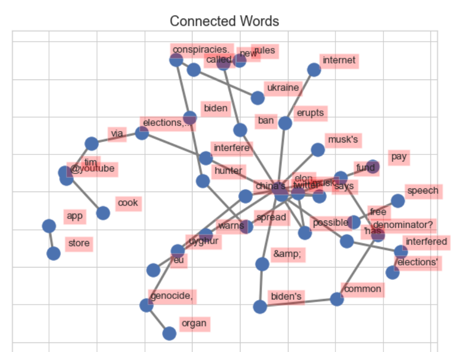

Fig. 5. 40 most connected bigrams within Tweets

The total sentiments histogram (Fig.8) shows again that most tweets are analyzed to be fairly neutral in sentiment. However, we do see that among the polarized tweets it tends to be more positive. This potentially could be a biased result of the pre-trained sentiment analysis which would often see "free speech" something that sounds positive appearing in the tweets. However, there were many tweets observed of

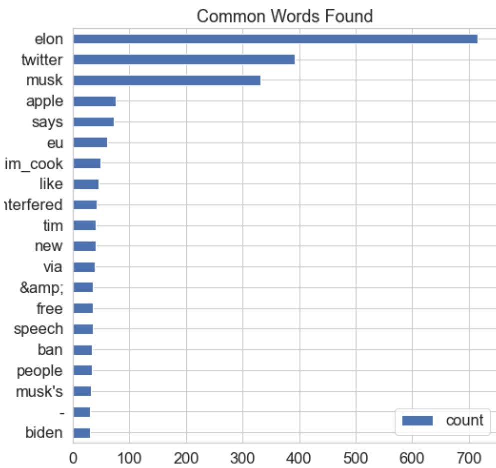

Fig. 6. Elon Musk Most Common words among 3000 Tweets

	Words	Normalized Degree Centrality
0	twitter	0.250000
1	elon	0.208333
2	musk	0.083333
3	says	0.083333
4	warns	0.083333
5	'has	0.083333
6	interfered	0.083333
7	erupts	0.083333
8	via	0.041667
9	@youtube	0.041667
10	musk's	0.041667
11	free	0.041667
12	speech	0.041667
13	tim	0.041667
14	cook	0.041667
15	eu	0.041667
16	elections'	0.041667
17	ban	0.041667
18	interfere	0.041667
19	internet	0.041667
20	possible	0.041667
21	app	0.041667
22	store	0.041667
23	hunter	0.041667
24	biden	0.041667

Fig. 7. Elon Musk Tweet Degree Centrality

individuals that were communicating their support for Elon Musk and his changes to the Twitter business and platform.

C. Kanye West

Results focus on tweets containing the search term #KanyeWest. For this topic, the large majority of tweets came from accounts representing individuals rather than organizations. Related hashtags and words changed rapidly when testing different search terms. This topic was extremely viral and trendy. Co-creational terms were much more prevalent. However, the terms were not being consistently applied and arose and died rapidly.

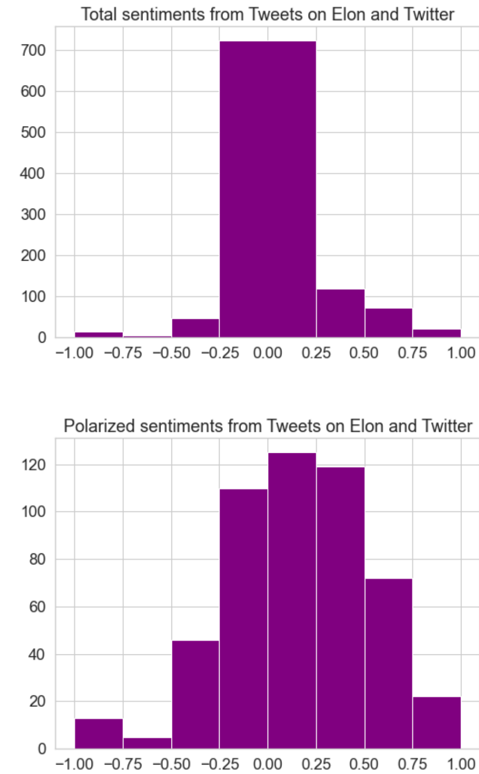

Fig. 8. Elon Musk Tweet Sentiment

The two networks shown in Fig.9 were created from batches of tweets pulled four hours apart as mentioned previously. Interestingly, co-created hashtags were much more prevalent in use for this topic. The hashtags being employed changed rapidly but appeared commonly in the constructed graphs and quantitative measures. The difference between the ego-centric graphs created show the rapid change of terms being used. There were also many more verbs present than in other topics. Opinions rather than strict information were more common among tweets. One theory as to why this is the case is because the public was much more involved and engaged in this topic.

The next set of figures were common words (Fig.10) for one pull as well as the normalized degree centralities (Fig.11) for the two 4 hour separated searches. Significant differences in terms can be observed in the two centrality tables.

The polarized sentiment analysis (Fig.12) reveals a higher amount of polarized tweets present than in the other topics. Sentiments ranged over the values but more extreme negative sentiments are present versus other topics.

IV. Analysis

A. Trends among trending: Difficulty of network analysis and improving methodology

The motivation of this section is to identify and present what was successful with analysis methods, what the challenges

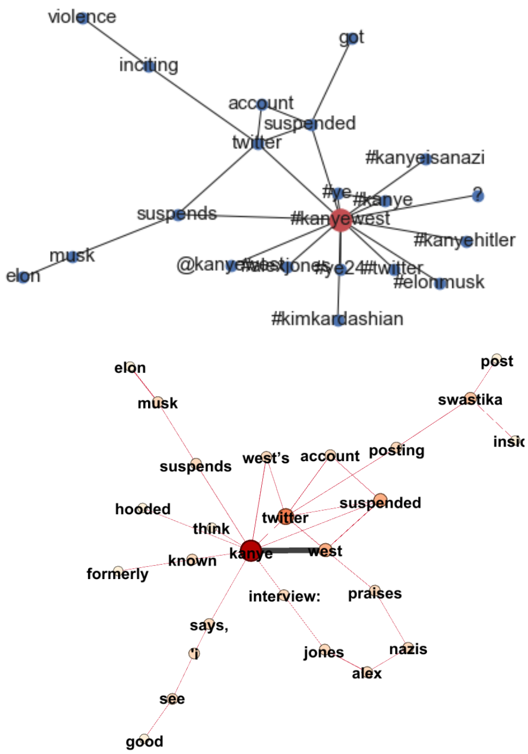

Fig. 9. Ego graph separated by 4 hours time

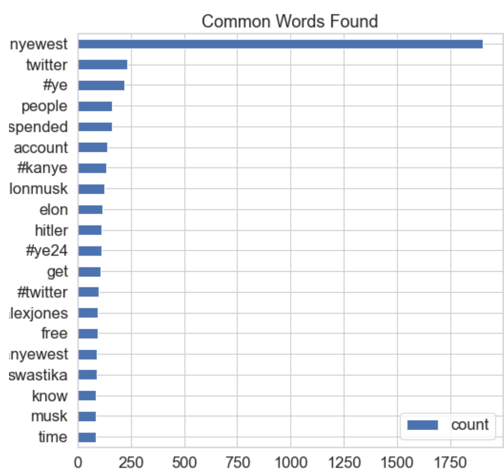

Fig. 10. Kanye West Most Common words among 3000 Tweets

were with topic and public relations analysis using networks and suggestions on revised methodology.

Results show that the process of co-creation of terms and hashtags and their use are dependent on whether they are being primarily driven by posts from organizations or the public. If conditions are met for this process to occur, co-creational terms can ignite engagement with a topic and strongly drive trending themes and discourse. Threads of conversation and

	Words	Normalized Degree Centrality
0	kanye	0.256410
1	twitter	0.153846
2	west	0.102564
3	suspended	0.102564
4	swastika	0.076923
5	musk	0.051282
6	alex	0.051282
7	jones	0.051282
8	west's	0.051282
9	account	0.051282
10	posting	0.051282
11	suspends	0.051282
12	praises	0.051282
13	good	0.051282
14	see	0.051282
15	known	0.051282
16	10	0.051282
17	nazis	0.051282
18	interview:	0.051282
19	'i	0.051282
20	says,	0.051282
21	elon	0.025641
22	kim	0.025641
23	kardashian	0.025641
24	chris	0.025641
25	paul	0.025641
26	things	0.025641
27	think	0.025641
28	free	0.025641
29	speech	0.025641
30	formerly	0.025641
31	top	0.025641
32	inside	0.025641
33	searches	0.025641
34	hooded	0.025641
35	nick	0.025641
36	fuentes	0.025641
37	post	0.025641
38	star	0.025641
39	david	0.025641

	Words	Normalized Degree Centrality
0	#kanyewest	0.350
1	twitter	0.125
2	suspended	0.100
3	suspends	0.075
4	musk	0.050
5	#ye	0.050
6	#kanye	0.050
7	account	0.050
8	speech	0.050
9	inciting	0.050
10	.	0.050
11	elon	0.025
12	alex	0.025
13	jones	0.025
14	chris	0.025
15	paul	0.025
16	#kanyeisanazi	0.025
17	free	0.025
18	#ye24	0.025
19	#alexjones	0.025
20	kim	0.025
21	kardashian	0.025
22	#twitter	0.025
23	#elonmusk	0.025
24	violence	0.025
25	mental	0.025
26	health	0.025
27	#kanyehitler	0.025
28	freedom	0.025
29	star	0.025
30	david	0.025
31	swastika	0.025
32	inside	0.025
33	social	0.025
34	media	0.025
35	?	0.025
36	got	0.025
37	mentally	0.025
38	ill	0.025
39	@kanyewest	0.025
40	#kimkardashian	0.025

Fig. 11. Degree Centralities separated by 4 hours time

engagement tend to last longer when communities agree on a few consistent co-creational terms. Some trendy and "meme" terms can quickly rise to prominence but will die out quickly if this consistency is not reached.

Using these terms as filters for graph networks for social media analysis can have very different effects. Search terms chosen can help to identify communities and themes. However slightly different filtering with these search terms can quickly lead to identifying a whole other set of communities and sentiment. This is illustrative of some of difficulty present in engineering networks that effectively capture entire topics. Resulting analysis on these networks should be given within the context of the topic to reduce misinterpretation. Empirical analysis of social media trends proved difficult in this research. It could also not be shown that filtering by words or hashtags was more effective. Effective choices of search term types and filtering was dependent on the co-creation that is occurring within the topic on the platform.

Many different perspectives can be taken when performing social media network analysis. It was fairly easy to create

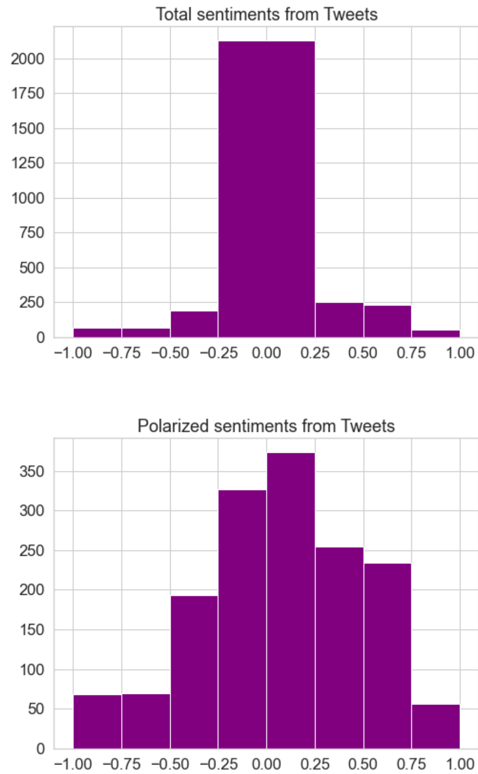

Fig. 12. Kanye West Sentiment Analysis

a variety of different networks with the data available, as well as generate quantitative measures. However, it was also a challenge at times to interpret and extract actionable meaning from the results. One issue was the dynamic nature of themes as well as co-creational terms used to communicate and connect ideas. Snapshots in time provide some insight but combining further analysis on information diffusion could provide interesting results.

From this research, several suggestions and insights are given on methodology for social media network analysis. Focusing on hashtags tends to increase the topic relevance of pulled tweets. Focusing on words increases connections for network analysis and gives more consistent results for capturing themes within a topic. Contextualization of network results is essential for providing analysis on a topic. Trends are not completely quantitative and require explanation for actionable results. Further specifying communities of interest for a study would appear to greatly improve results. A suggested process methodology for analysis is: 1. Identify topic of interest, 2. Network exploration, 3. Identify community of interest, 4. Revise analysis, 5. Contextualize and present results.

B. Future Work

Some future work that would improve results would be introducing dynamic networks to capture information diffusion and better elucidate trending topics and themes. Improved visualization can be leveraged to more effectively convey information. Selecting and limiting study to more specific communities of interest would aid in producing actionable results. Selecting a narrow goal or motivation for research would also help with community selection and perhaps also guide research to more effective selection of quantitative measures.

V. Conclusions

In this article, a study of trending social media topics and communication on the Twitter platform is presented. Specifically, graph networks are created using hashtags and words. Qualitative views of the constructed graphs as well as quantitative measures and sentiment analysis are presented. Three trending topics of interest were chosen for analysis. These were NFL and concussions, Elon Musk and Twitter, and Kanye West. Analysis was conducted in context for each topic and revealed that co-creational terms and interaction on the Twitter platform can vary significantly based on communities and how they engage with that particular topic. Finally, a discussion on the effectiveness of the research as well as the challenges of social media topic analysis is presented. Suggestions for revised methodology and future work is also given.

References

[1] Ying Xiong, Moonhee Cho, Brandon Boatwright, "Hashtag activism and message frames among social movement organizations: Semantic network analysis and thematic analysis of Twitter during the #MeToo movement,"Public Relations Review, Volume 45, Issue 1, 2019, Pages 10-23, ISSN 0363-8111, https://doi.org/10.1016/j.pubrev.2018.10.014.

[2] Xieling Chen, Haoran Xie, Zongxi Li, Gary Cheng,"Topic analysis and development in knowledge graph research: A bibliometric review on three decades,", Neurocomputing, Volume 461, 21 October 2021, Pages 497-515, https://doi.org/10.1016/j.neucom.2021.02.098.

[3] Wu, Lingfei, et al., "Graph Neural Networks for Natural Language Processing: A Survey,", arXiv preprint arXiv:2106.06090 (2021).