



Lifetime of tweets: a statistical analysis

Yashasvi Koul¹ · Kanishk Mamgain¹ · Ankit Gupta¹

Received: 13 February 2022 / Revised: 5 July 2022 / Accepted: 9 July 2022 / Published online: 4 August 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Austria, part of Springer Nature 2022

Abstract

Social media has become such a large part of people's life that even if little at a time, that influence can accommodate over time and can manipulate or even form new opinions. The authors have gathered data with which it is easily understood that the growth of Twitter, the people within its engagement range and its potential for becoming a portal of information sourcing as well as incidents have grown considerably well over the last decade and are well expected to grow into the next decade as well due to the new generation telecom technologies. This study aims to understand how much time Twitter trends remain 'hot' based on various parameters including but not limited to demography, the incident, time period or the people affected. The main objective is to gather data about different trending topics over different time periods and then analyze the pattern of how tweet volume due to that Twitter trend increased or decreased over a few days. This allows to demonstrate that Twitter can be a powerful tool to manipulate public opinion since this reaches a large number of users in a lot of developed countries. The influence of tweets can be seen from the fact that even a tweet done from a non-influential person's account can garner enough attention to become worldwide phenomenon. Towards the end of the study, the authors used a visual medium to depict how various topics fared over the 5 days that tweets were scraped.

Keywords Twitter · Trend · Graphs · Distribution · Opinions · Patterns of trends

1 Introduction

The Oxford dictionary has defined social media as- "websites and applications that enable users to create and share content or to participate in social networking". But is that all there is to it? In today's world, social media is as much a part of the everyday lives of working people, students, and other people over a particular economic level as eating and sleeping. Why? This is because social media has evolved from a site for information sharing to a platform for people to voice their thoughts, form groups on an ideology, or even build

their own ideology. People currently consume rather than use social media. We chose the word 'consumed' since this is a product that has become standard for checking up on the nearest and dearest on to keep a closer eye on the trendiest bit of news in the most remote region of the planet.

While social media is ubiquitous in America and Europe, Asian countries like Indonesia lead the list of social media usage (<https://digitalmarketinginstitute.com/blog/social-media-what-countries-use-it-most-and-what-are-they-using>). More than 4.5 billion people use social media, as of October 2021 (<https://datareportal.com/social-media-users>). According to Pew Research Center, social media users are generally younger. Almost 90% of respondents between the ages of 18 and 29 utilised some sort of social media. Furthermore, these individuals are more educated and somewhat rich, earning more than \$75,000 each year (<https://www.pewresearch.org/internet/fact-sheet/social-media/>). Given the above facts, it is clear that a large portion of the working individuals and the students who'll become the part of this group in few years have some form of connection to social media. So the effect of any change to working or the distribution of content or the content itself on social media reaches a large populace. Social media user numbers have

Y. Koul, K. Mamgain, A. Gupta: These authors contributed equally to this work.

✉ Yashasvi Koul
y.koul16@gmail.com
Kanishk Mamgain
kanishkmamgain2000@gmail.com
Ankit Gupta
theankit@yahoo.com

¹ Department of Computer Science and Engineering, Chandigarh College of Engineering and Technology, 160019 Chandigarh, India

seen strong growth over the past 12 months too, with 409 million new users joining social media in the year to October 2021. That equates to annualised growth of 9.9 percent, or an average 13 new users every single second (<https://datarportal.com/social-media-users>). This means that relevance of social media has been on constant growth over the past many years and the trend does not seem to be slowing down.

One such social media platform is Twitter that was started in the US in 2006. As of 2013, it was one of the ten most-visited websites and has been described as “the SMS of the Internet” (D’Monte 2009). The sheer no. of people having accounts or engaging on Twitter may not be as large as facebook (<https://www.visualcapitalist.com/ranked-social-networks-worldwide-by-users/>) but the no. of queries or tweets made over the limited no. of characters provided by Twitter shows that it is quite popular for putting out short sentenced content to sum up what the authors want to say. According to Alexa traffic statistics, Twitter is the world’s eleventh most popular website (Alexa 2019). As of second quarter of 2021, Twitter has “206 million monetizable daily active users worldwide” (<https://www.statista.com/statistics/242606/number-of-active-Twitter-users-in-selected-countries/>). This gives an insight into how many people actively participate and make their information and opinions public through the platform. So it was on this basis that we have determined the rise and fall of the no. of tweets, which are the so-called trends, which are associated with a particular topic over a specific period of time. A “trending subject” or simply “trend” is a name, phrase, or topic that is discussed more frequently than others. Trending topics become popular as a result of a deliberate effort by users (such as boosting an election candidate) or as a result of an occurrence that causes people to speak about a certain issue (such as a TV series or earthquake). Since there are no specific constraints on what tweets are tweeted on what topic in which time period we will be considering the trends for only a certain time period after whatever incident or happening had happened. There is currently no clear picture of what causes various issues to become immensely popular, or why some topics last longer in the public spotlight than others. There is evidence that one factor that causes topics to deteriorate over time is novelty (Yang and Leskovec 2011). The influence of network members on content propagation is one of the factors that contribute to the popularity of specific subjects. Some individuals create material that has a deep emotional resonance with their followers, leading it to spread and gain popularity (Romero et al. 2011). The site’s success is based on the capacity to follow and be followed, which results in a dense network of contacts who transmit and respond to messages. This results in a massive and continuous information flow in which even the most insignificant post may quickly become a global hot issue (Murthy 2013).

The paper has been structured in the following way: Sect. 2 presents a comprehensive collection of research literature including recent works on Analysis on Social Media through different techniques. Then, in Sect. 3, a detailed information about Twitter trends is given. Section 4 shows the methodology and flow of work. Then, in Sect. 5, results are discussed. Section 5.1 is dedicated to the Tables that were made during the analysis. Finally, conclusions are portrayed in Sect. 5.3.

2 Motivation and some earlier work

There have been some previous research works towards examining Twitter connections. A lot of them included that most of the opinions that were made or put forth on Twitter or any other social networking site were not of one’s own. They were either manipulated or were changing according to the masses. In simpler terms, the content’s resonance with the social network’s users has a significant impact on how Twitter trends change (Asur et al. 2011; Zubiaga et al. 2009). This is not only in the case of news, but also with products of different brands. Twitter has been seen as a means of spreading word-of-mouth advertising. Various opinions are made regarding a particular brand depending on the type of thread or trend which is active at that particular time span (Jansen et al. 2009). Studies related to hashtags are being conducted to determine their behaviour over time by demonstrating general temporal patterns that hashtag-containing tweets follow by presenting a reliable time series clustering technique. Yang and Leskovec (2011).

There has also been research on social influence and dissemination in the past. Cha et al. (2010) compared three distinct types of influence measures: indegree, retweets, and user mentions. While retweets and mentions were shown to be highly connected, the in-degree of users was found to be unrelated to the other two variables. As a result, they suggested that the number of followers is not necessarily a reliable indicator of influence. Many researches have revealed that extracting tweets regarding a certain trend causes the user demographics and Twitter use habits to reflect the subject’s real-world qualities (Cheong and Lee 2009). In a larger sense, academic study on different facets of Twitter has been rising in the recent years and the Twitter API is playing a major role in it Mayer (2009). Studies are being conducted on the Twitter’s user base growth rates, geographic distribution, and other statistical features (Krishnamurthy et al. 2008). The characteristics of hot topics on Twitter have received some attention. The majority of them concentrate on detecting events and topics (Albakour et al. 2013; Zhao et al. 2012), as well as summarising scheduled events as they happen in real time. Number of tweets in a particular phase may or may not include a lot of contradictory information.

Chen Lin et al. in their research (Chen et al. 2021) presented novel realtime event summarization framework called IAEA (Integrity-Aware Extractive-Abstractive realtime event summarization) in which they integrated an inconsistency detection module into a unified extractive-abstractive framework.

Another strategy that is being followed to understand the trends on Twitter is the bifurcation of topics based on different needs. Some of them are lexical analysis, time to reach, trend reoccurrence, trending time, tweets count, and language (Annamoradnejad and Habibi 2019). One of the strategy adopted in Lee et al. (2011) was to divided trends into 18 groups using two different methodologies. Sports, Music, and Movies had the biggest amount of trending topics, according to their findings, which were based on 768 distinct trending themes. Furthermore, certain services have been created to explain Twitter trends on their own or in a collaborative scenario (Gao et al. 2014). A study was conducted which focussed on Twitter trend analysis and presented a method for detecting trends in streams of tweets (Khan et al. 2021). The suggested method ranked the top phrases and hashtags in real-time Twitter trends and found the hot topics. By collecting tweets on comparable themes into manageable clusters, a research provided a method for automatically detecting hot topics which were addressed mostly on social media (Anandarao and Chellasamy 2021). The authors used modified density peak clustering (MDPC) algorithm for the same. There are studies which involve machine learning algorithms in Social Media Analysis (Balaji et al. 2021).

Social media can be the deciding factor in many events (Fan and Michael. 2014) where people give their opinions on different matters. Singh et. al. in their paper 'Social Media Analysis through Big Data Analytics: A Survey' (Singh et al. 2019) conducted a survey using the various Big Data Analysis techniques such as NLP, news analysis, opinion mining, scraping, and sentiments analysis for unstructured textual data sets. Batrinca et. al. in their paper 'Social media analytics: a survey of techniques, tools and platforms' (Batrinca and Treleaven 2015) have tried to put up categorization of software tools used for data analysis and also the reason for boom in these analyses and the structure for these analyses. Extended and structured literature analyses were conducted through which researches identified challenges faced during the topic discovery, data collection, and data preparation and proposed a solution for the same (Stefan et al. 2018). But also, there is a concern with social bots that imitate humans and manipulate opinions on social media (Aldayel and Magdy 2022). A study was conducted in which a supervised method was introduced which distinguished between the bots and legitimate users with a high accuracy rate (Marcelo et al. 2020).

Along with a large userbase of Twitter, there are some mischievous elements in the society who start a trend of

an event by moulding the original event. Key problem of detecting facts from several conflicting sources is truth discovery, which requires concurrently deciding whether the source is plausible and hence whether the article is accurate. According to Haas and Wearden (2003), this is because as the number of sources grows, the control or 'gate-keeping' function shifts away from the producers and towards the recipients of the information. The phrase 'gatekeeping', formulated by Lewin and K. (1947), refers to the process by which content creators choose what sort of narrative will be closely scrutinised, reported on, and hence what information will be provided to the receivers. Therefore, many things can impact the perceived credibility of online materials (2008). This problem of fake news spreading on Twitter was the most during the coronavirus pandemic and became a topic for study (Gupta et al. 2022). For this reason, in recent years, false information detection (FID) has been a hot study area (Bin et al. 2020; Collins et al. 2021). Kai Shu et. al. provided a fake news data repository (Kai et al. 2020) to facilitate fake news-related research.

3 Twitter trends

A "trending subject" or simply a "trend" on Twitter is a term, phrase, or topic that is referenced at a higher rate than others. Users can make a determined attempt to make a trending subject popular, or an event might encourage others to discuss about a certain topic. These topics help Twitter and their users to understand what is happening in the world and what people's opinions are about it. Many public incidents, such as the wildfires in San Diego and the earthquake in Japan, have demonstrated the impact of Twitter trends (Sakaki et al. 2010).

But trends are not always spontaneous. Sometimes for business, political or some other purpose the platform has been used to garner attention of the masses to a certain topic. It has quite an effect because even if a part of population tweets about it, the attention garnered is not always proportional to the no. of tweets. For, e.g., let's say that a certain star sportsperson who is quite famous worldwide tweets about a certain incident or brand or some issue somewhere in the world, that particular tweet will have more retweets than if a certain salary man tweeted the exact same words. Due to the tweet of that person, the number of people who follow that sportsperson on Twitter will get informed about this and there is a high probability that a lot of them may retweet that tweet or tweet their own opinion. This all goes to say that even though those people retweeted or tweeted themselves because of following a certain person, the tweets were counted towards the incident whose '#' (Users are encouraged to use hashtags to classify their tweets, which

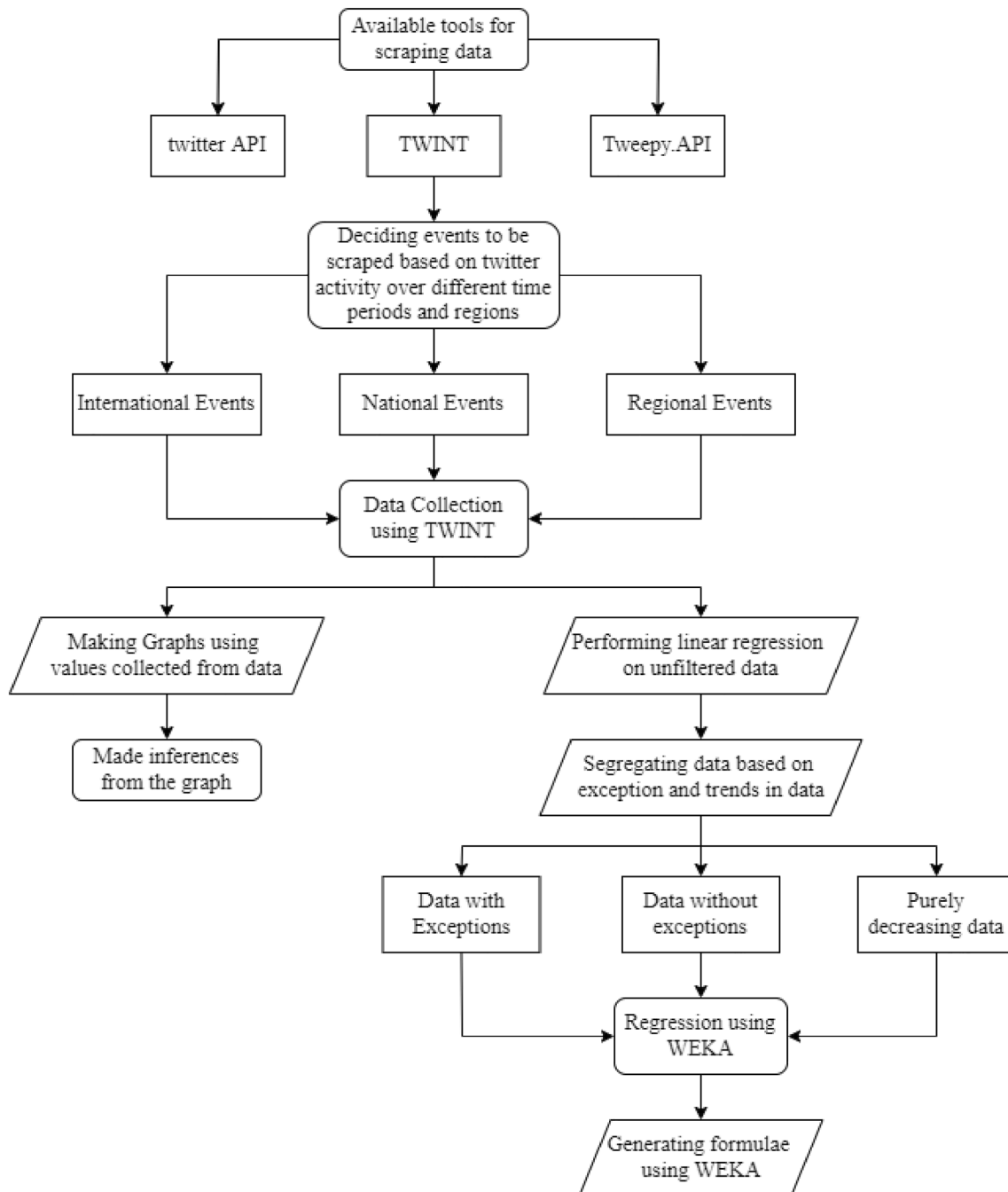


Fig. 1 Flow of Work

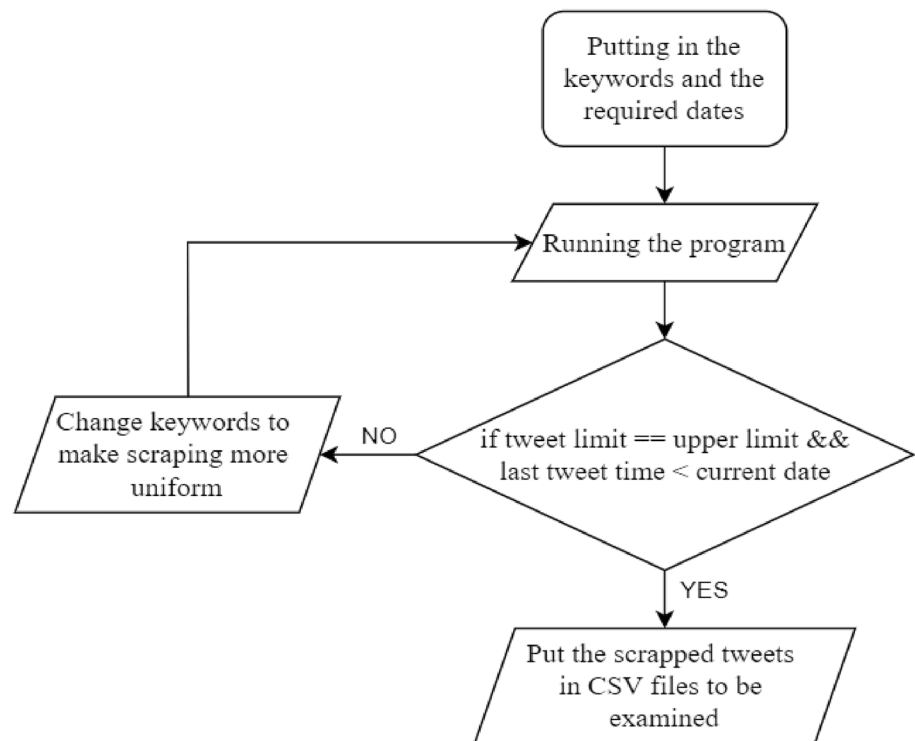
are any keywords preceded by the hash sign '#', for example, #sorcery.) or references were in the tweet.

4 Methodology

4.1 Data collection

The overall flow of work is shown in Fig. 1. We used Twint (<https://github.com/twintproject/twint>), an open-source API framework that allows users to access the Twitter data base to scrape tweets. Though it also makes use of the official Twitter API, it does not follow the same constraints,

Fig. 2 Working of twint



which is why it was very helpful in providing specific constraints regarding start and end date of tweets scraping, the limit of tweets to be scraped and the keywords according to which tweets are to be scraped. Python scripting made it easy to automate this and allow to segregate the data into CSV (A comma-separated values file is a text file with values separated by a comma. A data record is represented by each line in the file. Each record has one or more fields, which are separated by commas. The name for this file format comes from the usage of the comma as a field separator. A CSV file generally carries tabular data (numbers and text) in plain text, with the same number of fields on each line.) files that allowed to prepare visual graphs.

As for the constraints since the topics are ranging from different year, different Twitter audience and regions, the authors decided it would be best to keep an upper limit on to how much tweets are to be scraped so as to allow comparison between similar happening events or topics that occurred in different regions or time periods. Data taken have not been segregated based on the content of the tweets itself so there are chances of redundant tweets and tweets with same keywords but completely unrelated to scraped topic in some cases. For this purpose, the keywords were chosen in such a manner that the tweets scraped would have the highest possibility of being in the range of the topic we intended to be collected. Working of twint is shown in Fig. 2.

4.2 Data classification

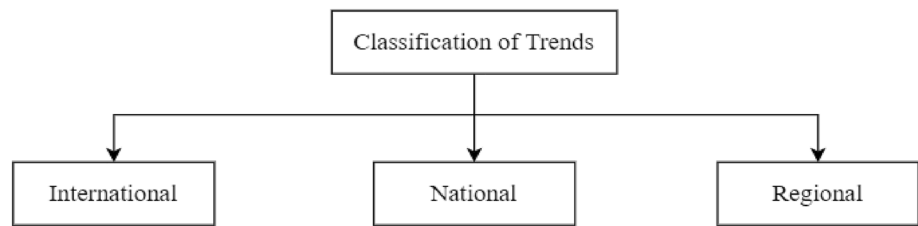
The data which was collected were classified majorly in three categories

- International news
- National news
- Regional news

The basis of classification of trends into the three categories mentioned, i.e. national, international and regional was mainly decided by two factors:

- Firstly, the volume of tweets coming in on a certain event is determined by the severity of the event. Since we had no metric of determining severity or a similar scale, we decide to go with the next determinant that is effect and reach of a certain event from its epicentre.
- Secondly, the reach of an event beyond a certain bubble around its epicentre. Even if an event is not covered by state news channels, if it becomes ‘viral’ on social platforms, there is a high chance that it will get more engagement than it would have gotten without the social platform.

Based on these hypotheses of ours, we determined that the reach of events could roughly be categorised into the mentioned three categories (Fig. 3). The determining factors were of course tweet volume and geographical location, but

Fig. 3 Classification of tweets

we kept the determining factor as the effect and influence of said event or news.

1. The International events would be those in which the said event gets engagement from almost all active and passive social platform users, getting its Twitter volume incredibly high, e.g. The 2011 Tohoku Earthquake.
2. The National events would be those which have most of their Twitter volume coming in from users within the country with some contribution from neighbouring states, e.g. AB Vajpayee death.
3. The regional events would be those in which most of the tweets came from users in the same area, state or surrounding counties, and the volume was considerably less than national and international events.

Although there are some exceptions, like the JNU protest which is listed as regional because the tweet volume for its 5 days from the initial day of event is quite low and most of the volume came from users around Delhi, but as the story unfolded, after some days the tweet volume exploded and it consisted of tweets from users all over the country as well as some other countries as well. So here, if our parameter for checking tweet volume was up to 10 days, then this might have been classified as a National Event.

4.3 Deciding paramaters

After collection of data, the next step of the process was deciding what parameters which would be taken in consideration over a particular time period. So, to gather results of data specifically and make them easily understandable, each topic had tweets scraped for 5 days from the start day of incident up to 5 days later, and graphs were created for each of these files.

This serves two purposes:

- Comparing volume of tweets of different topics over different time periods becomes easy.
- It makes graph for tweets per hour for the day's 24 h, so the graph provides the structure for the smallest details.

4.4 Issues encountered

Collecting all the data, references and inferences mentioned in this paper, there was fair share of issues that were encountered. The reasoning for why these issues were encountered in the first place and what possible solutions they had and what alternatives were taken are given below:

- Since the Twitter API has some limitations in scraping tweets, i.e. the number of tweets as well as other constraints like keyword specification and timeline specification. To make scraping efficient, the authors made use of TWINT open-source tool.
- TWINT also had its fair share of troubles. It being an open-source tool, any PR (Pull request) or patch that updates any existing bug requires changes to the current version, for smooth functioning. Also, the patches for these issues are not immediately found so the time for which scraping data could not be done varied from 2 days to 1 week. Still, providing an upper limit helped for faster and more efficient scraping
- Since the events on which scraping was done have timelines from 2011 to 2021, the volume of tweets corresponds to quite arbitrary number even after similar upper limit constraints were placed on all of them. This made forming a single pattern pertaining to all data quite difficult. So, the data were further segregated as shown in the Calculations section.

5 Results

Before discussing inferences of study, we obtained through this process of scraping and looking over data graphs, we will discuss what results the data showed us in this section. As mentioned above the topics chosen can be classified on basis tweets that were tweeted. The main volume of tweet and the keywords that were contained made classification into three categories, which are not completely exclusive. So, the topics which are region specific will have most of the tweet volume made up of the Twitter users in that region specifically. Similarly, the tweets in incident specific will most probably contain tweets volume of people tweeting

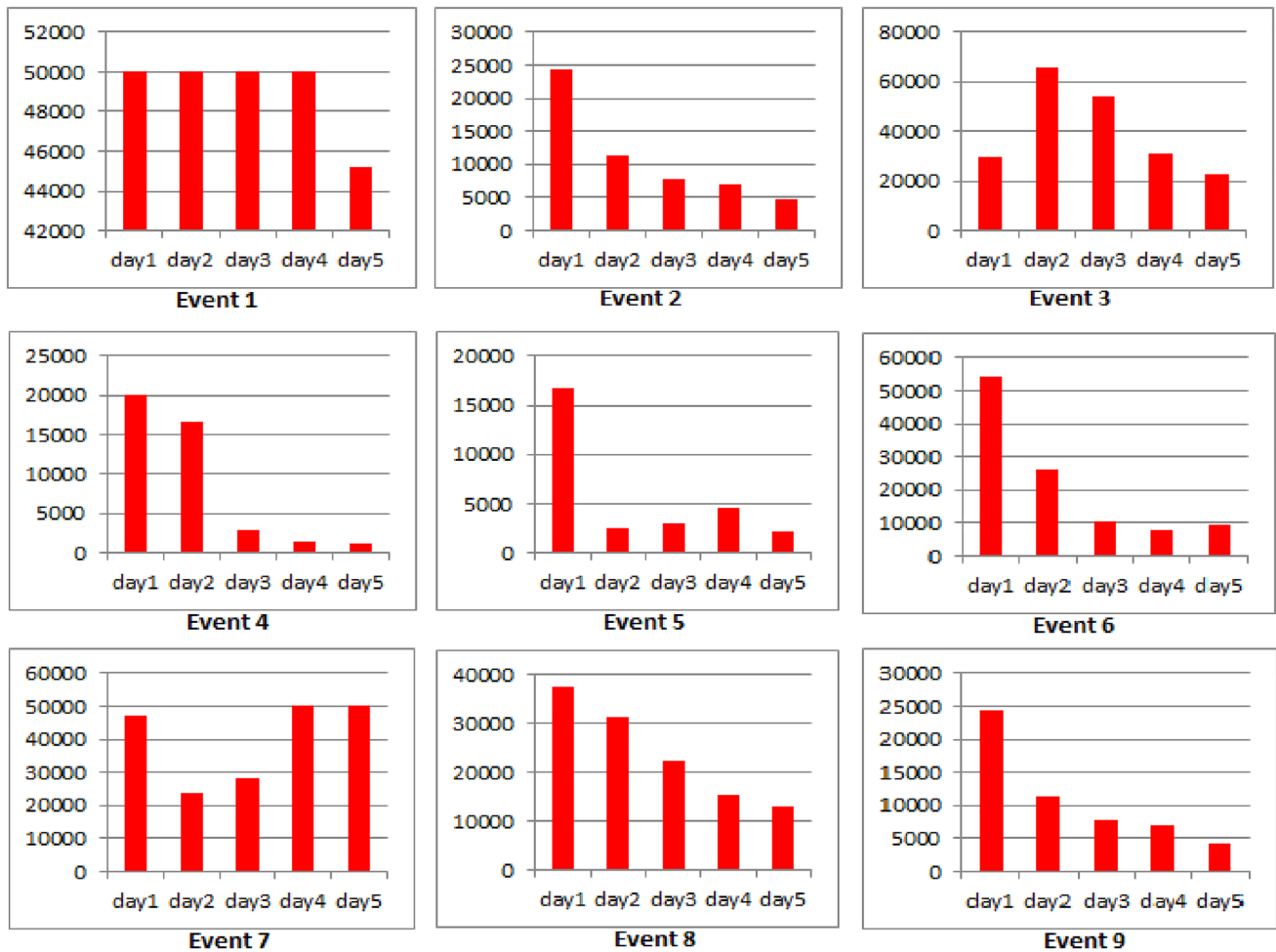


Fig. 4 Graphs of various international events (part 1)

with regard to that incident, and same will be the case in person-specific classification.

A tweet with a tagged keyword from our script may also get collected even if it has unrelated content. For example, if we are scraping for Euro 2016 and choose the keywords ‘euro’ & ‘2016’, a tweet that says ‘today took a inter euro trip, best day of 2016’, may also get collected since it satisfies the criteria. So in situations like these to prevent the file filling up with irrelevant data, we preferred getting more specific with keywords and getting more relevant tweets than getting larger volume.

Let us look on some of the facts which affect the trends of the news:

- Memorable events such as the 26/11 attacks, 9/11 attacks or the Paris attacks, trend on Twitter every year on that particular day on which the event happened (refer Fig. 4, Event 3).
- As the news reaches different time zones, the tweets of a particular event start to increase. As in case of the

Capitol riots, as the news reached around the world, tweets increased (refer Fig. 5, Event 13).

- Any news related to a **famous personality** was considered as the hot topic and was a trending topic in which the tweet flow saw an increasing trend for a long period of time. For example, the news of Sushant Singh Rajput committing suicide instantly gained popularity and as the various proofs and facts related to the incident were being disclosed, the tweet flow increased (refer Fig. 6, Event 25).
- News related to **race, caste or religion** spread really fast, and in various countries where multiple religions are practised and there is no official religion recognised by the government, these news stories gain popularity at a very fast rate and we see a spike in the number of tweets related to that event (refer Fig. 7, Event 27).
- Due to human nature, many times it happens that we get bored from a subject and we need a new subject to debate or discuss upon. So time plays an important role to determining the trending topic. In most of the cases,

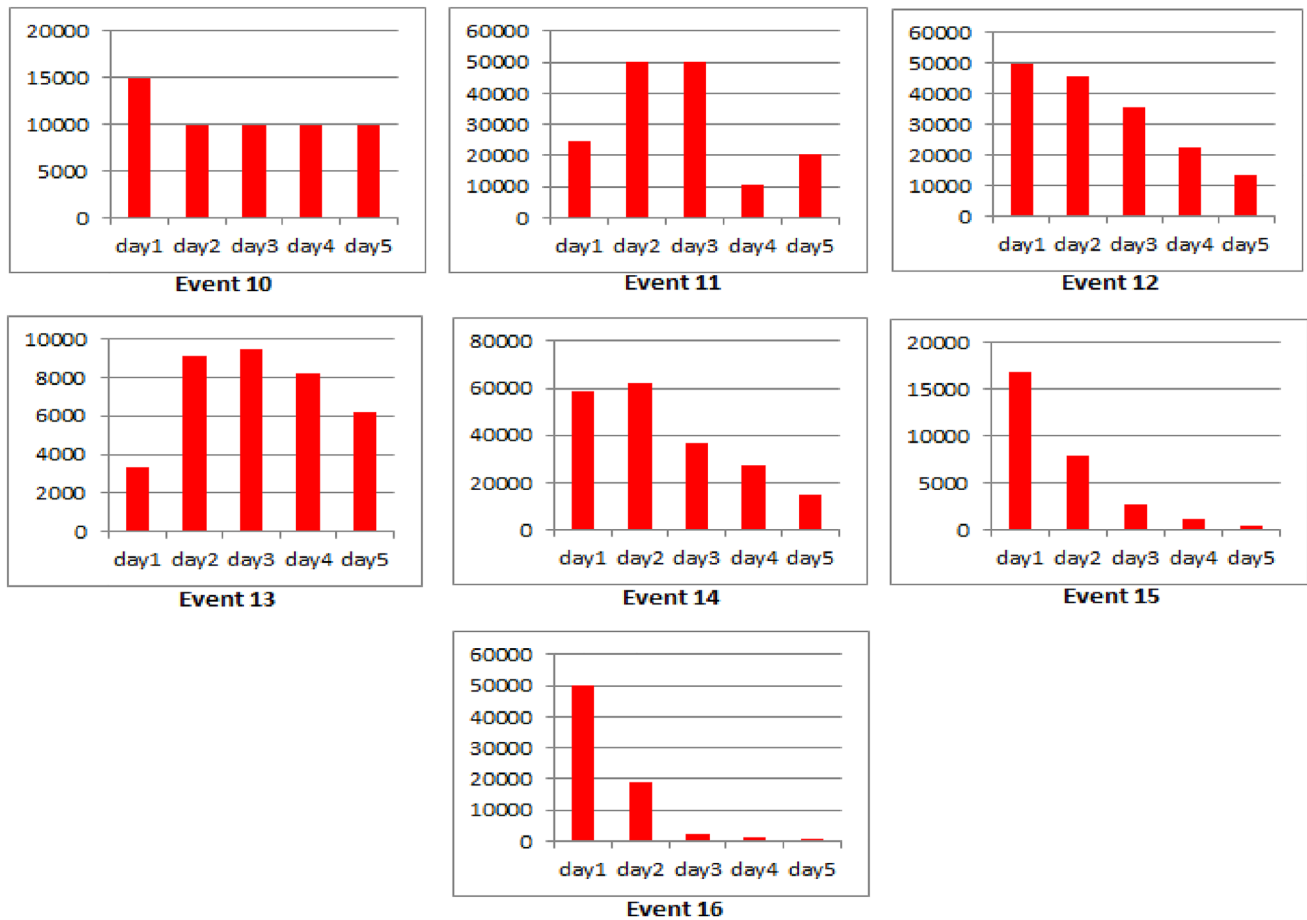


Fig. 5 Graphs of various international events (part 2)

the news trends start to decrease after a week or so for a particular event.

- As the various facts related to the events were being revealed, people around the world started to drop their opinions on the issue which further increased the tweet flow. For example, Delhi riots (refer Fig. 7, Event 29).
- Topics relating to **entertainment industry** like movies or songs, usually trend for 2-3 days before fading.
- Since the majority of the events taken into consideration are international, it can be safely said that after a certain time, most of the events have had tweet volume from worldwide Twitter community.
- Since all the values scraped are quite arbitrary, there is no single correlation between the values, a rough correlation using linear regression on the data scraped is provided in subsection 5.2.

Following graphs were drawn as tweet count vs days for the first 5 days:

5.1 Tables

Below shown are the tables with event number, event name and the next five days and the corresponding tweet count.

5.2 Calculations

In Tables 1, 2, 3 and 4, the authors have categorized the scraped data for 5 days in different types of news. It also contains events which have increasing or non-consistent trends and these are given the name of exceptions.

Data with exceptions: (refer Fig. 8)

The authors performed linear regression on the data for predicting day 5. The correlation was such:

$$\text{day5} = (-737.34) + 0.89 * \text{day4}$$

Data without exceptions: (refer Fig. 9)

The authors performed linear regression on the data without exceptions for predicting day 5. The correlation was such:

$$\text{day5} = (-508.71) + 1.09 * \text{day4} - 0.15 * \text{day3} - 0.07 * \text{day2}$$

Data with purely decreasing trend: (refer Fig. 10)

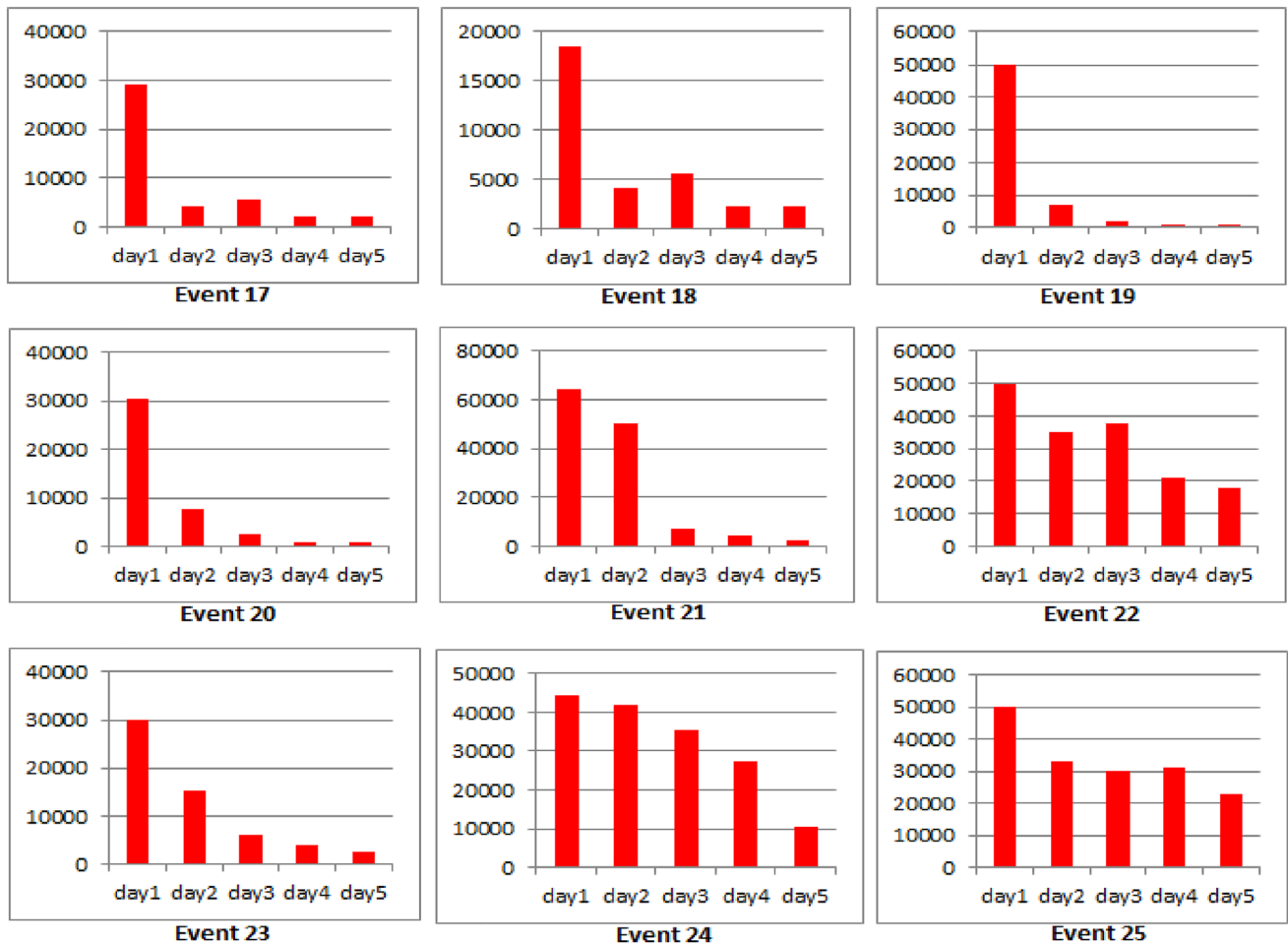


Fig. 6 Graphs of various national events

The authors performed linear regression on the data with purely decreasing trends for predicting day 5. The correlation was such:

$$\text{day5} = (-2047.766) + 0.747 * \text{day3}$$

According to (<https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/>), to gauge how closely two variables are related to one another, correlation coefficients are utilised. The most common correlation coefficient is Pearson’s, and however, there are other varieties as well. The correlation coefficient known as Pearson’s correlation, sometimes known as Pearson’s R, is frequently employed in linear regression. Using the correlation coefficient, we can determine the type of and proportion of relationship between the variables in the formulae used to describe tweet trends. As can be seen in the pictures above, we have used linear regression to determine on what basis the tweets fell day after day and whether there was some particular day after the initial incident in which drop was obviously significant. Using correlation coefficients, we can easily conclude that:

1. In dataset with exceptions, the correlation coefficient came out to be 0.9391, which is quite close to 1 and is positive. This tells us that the relationship between trend on day 5 which we have calculated is strongly related to tweets on day4 as its variable.
2. In data without exceptions, the correlation is 0.955, so we can say that if we remove exceptions from dataset, it is closer to ideal scenario since 0.955 shows that in this dataset, the relationship between tweets on day4 and day 5 is stronger than in data with exceptions since its correlation coefficient is 0.9391.
3. In dataset with purely decreasing trends only data in which no increasing trend can be seen for any event, correlation coefficient is 0.92, indication that the relationship between variables which in this case are day 3 and day 5 is still strong but less that previous 3 cases.

In all 3 points above, all the cases have a positive and above 0.9 correlation coefficient this shows that the results regression on our data has very strong variable relationships

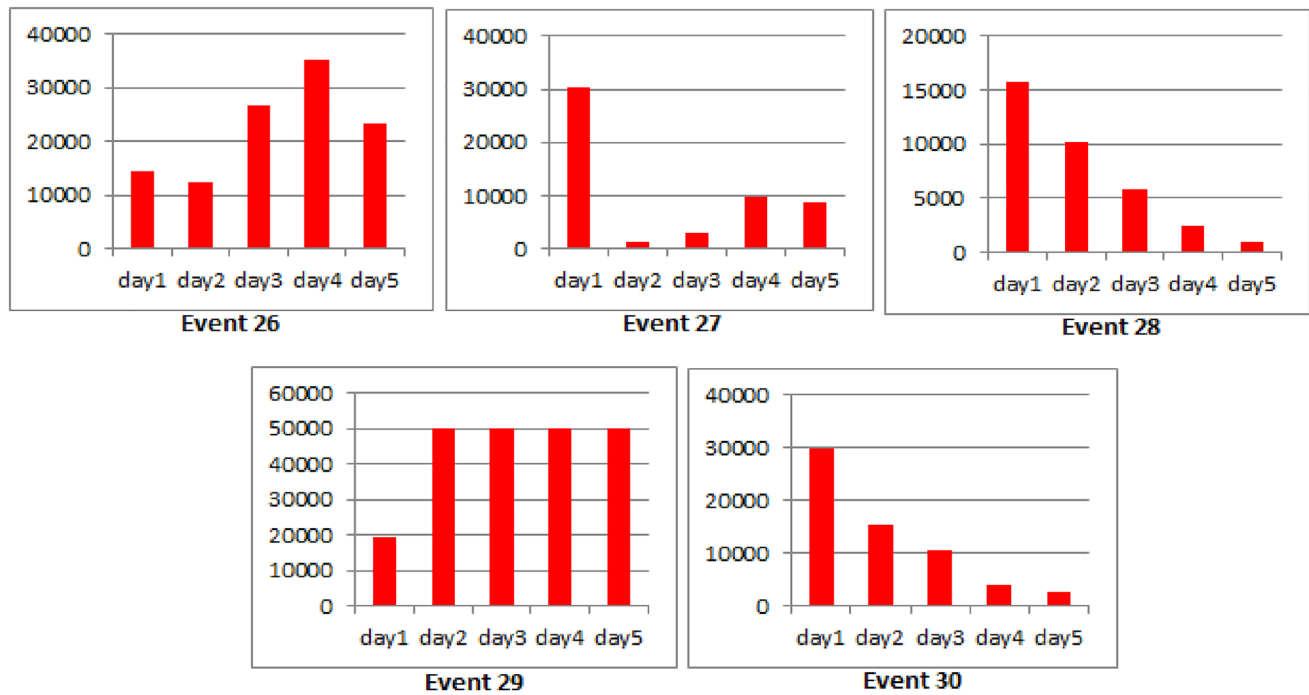


Fig. 7 Graphs of various regional events

Table 1 International news

Event no.	Event name	Day1	Day2	day3	Day4	Day5
1	Japan earthquake (11-03-11)	50003	50002	50007	50009	45173
2	2014 Football worldcup (14-07-14)	24415	11273	7693	7080	4728
3	Paris Attacks (13-11-15)	29406	65594	53903	30980	22654
4	leicester 2016 (02-05-16)	20058	16572	2812	1406	1127
5	PV Sindhu 2016 Silver (19-08-16)	16717	2504	2984	4603	2094
6	Surgical Strikes URI (29-08-16)	54182	26200	10481	8149	9660
7	2016 presidential elections (06-01-17)	47154	23590	28376	50009	50014
8	#metoo Harvey weinstein (15-10-17)	37342	31340	22333	15339	13028
9	2019 cricket world cup (30-05-19)	24417	11273	7695	7082	4278
10	Kobe bryant death (26-01-20)	15012	10014	10006	10005	10016
11	George Floyd (25-05-20)	24585	50000	50017	10758	20400
12	Biden won elections (03-11-20)	50010	45552	35652	22343	13256
13	Capitol Hill Riots (06-01-21)	3294	9161	9493	8194	6224
14	Copa America (11-07-21)	58924	62161	36462	26942	14832
15	Danish Siddiqui (16-07-21)	16821	7794	2666	1178	415
16	Neeraj Chopra Gold (07-08-21)	50016	19029	2243	1381	815

allowing us to determine data on Day5 based on the formulas given and, in all probability, they will be close to the actual scenarios. Of course, there will be exceptions.

5.3 Analysis

As shown above, the results we got were in the form of above graphs. In this section, the authors will explain what they inferred from looking at numerous graphs just like the one above.

Table 2 National news

Event no.	Event name	Day1	Day2	Day3	Day	Day5
17	2014 Lok Sabha Elections (16-05-14)	29300	4129	5593	2333	2284
18	Telangana formed (02-06-14)	18533	4127	5591	2331	2282
19	Mangal Yaan (24-09-14)	50011	6737	1698	833	896
20	Make in India (25-09-14)	30545	7574	2451	900	750
21	AB Vajpayee death (16-08-18)	64449	50016	6976	4527	2631
22	2019 Lok Sabha elections (23-05-19)	50024	35340	37520	21000	18000
23	Chandrayan 2 (06-09-19)	30153	15245	6153	3816	2500
24	Ram Mandir (9-11-19)	44561	42045	35652	27545	10489
25	SSR suicide (14-06-20)	50007	33163	30146	30917	22739

Table 3 Regional news

Event no.	Event name	Day1	Day2	Day3	Day	Day5
26	JNU Protests (9-02-16)	14540	12587	26640	35300	23264
27	Padmavat riots (25-01-18)	30303	1246	2847	9869	8708
28	Kaveri water dispute verdict (16-02-18)	15760	10240	5777	2450	989
29	Delhi riots (23-02-21)	19250	49999	50007	50005	50013
30	Lakhimpur incident (03-10-21)	30017	15475	10475	4147	2456

The significance of [bold] in table 4 is that it depicts the maximum percentage increase in cases where we are depicting exception in the drop of tweet volume.

- First, we would like to start with the fact that a common trend that is seen with different Twitter trends or trends in general over the years is that the volume of tweets of messages or any other form of putting out opinion on the social media has increased each year and it especially exploded just after the restrictions due to COVID pandemic, in which people stayed at a certain location for a long period of time and the amount of time spent on surfing sites like Twitter increased.
- Secondly, we take a look at the change in tweeting pattern over the days after incidents. In most of the data collected the number of tweets usually stayed around same or a bit higher than the 1st day, but there were some exceptions. In cases where the incident or the happening was prolonged or had greater than expected far reaching effect, the tweets volume kept on increasing even until the 4th or 5th day.
- The trends of tweets also depend on the region or demographic area to which it is specific. For example, the leading no. of Twitter users outside U.S. are in Japan and India, respectively, so any incident which occurred in this vicinity or affecting people of this region usually had a greater tweet volume than if it was a vague global incident.
- In some cases, that affected all of humanity or had a reach to all humanity, like the global pandemic or an attack on a city harbouring immigrants from different countries tends to get tweets from people of these countries as well even if they are not affected directly.
- An interesting inference that was obtained, the volume of tweets increased by a large margin after 2016. This was because of the introduction of Jio which offered internet at such prices that people who could earlier not afford also were able to access social networking sites like Twitter. Since India being the country with third largest number of Twitter users this had a significant effect on the tweet volume.

According to Rivier University (<https://www.rivier.edu/academics/blog-posts/an-introduction-to-behavioral-psychology>), behavioural psychology, or behaviourism, is a theory suggesting that environment shapes human behaviour. Conversations people hear and news they read especially in this digital age are a large part of their surroundings and that is what ultimately influences or induces opinions. The results shown above and their analysis done is to show that how opinions of people whether their own or induced from some idea can be influenced based on what is going on around them, i.e. what they hear, what they read, what gets broadcasted and what is considered right objectively. Twitter is a platform used by a lot of people throughout the world. It is also a platform where a lot of people with power, influence and fame put out their thoughts and opinions as statements. That is the reason why today on Twitter one can find politicians battling out public issues on a social media platform instead of the parliament, people supporting or condemning statements. This section of people supporting, condemning or forming a public opinion around an event, incident or

Table 4 Percentage drop (day-wise)

Event no.	Day1-2	Day2-3	Day3-4	Day4-5	
1	0	0	0	9.60000	–
2	53.82756	31.75729	7.96828	33.22033	–
3	– 123.06332	17.82327	42.52639	26.87540	Exception
4	17.37959	83.03161	50.00000	19.84352	–
5	85.02123	– 19.16932	– 54.25603	54.50792	–
6	51.64445	59.99618	22.24978	– 18.54215	–
7	49.97243	– 20.28825	– 76.23696	– 0.00999	Exception
8	16.07305	28.73962	31.31688	15.06617	–
9	53.83134	31.73955	7.96621	39.59333	–
10	33.29336	0.07988	0.00999	– 0.10994	–
11	– 103.37604	– 0.03400	78.49131	– 89.62632	Exception
12	8.91421	21.73340	37.33030	40.67045	–
13	– 178.11171	– 3.62405	13.68376	24.04198	Exception
14	– 5.49351	41.34264	26.10937	44.94840	–
15	53.66506	65.79420	55.81395	64.77079	–
16	61.95417	88.21272	38.43067	40.98479	–
17	85.90000	– 35.40000	58.20000	2.10000	–
18	77.73161	– 35.47370	58.30799	2.10210	–
19	86.52896	74.79590	50.94228	– 7.56302	–
20	75.20379	67.63929	63.28029	16.66666	–
21	22.39445	86.05246	35.10607	41.88204	–
22	29.35391	– 6.16864	44.02985	14.28571	–
23	49.44118	59.63922	37.98147	34.48637	–
24	5.64619	1.41515	0.02412	3.36631	–
25	33.68328	9.09748	– 2.55755	26.45146	–
26	13.43191	– 111.64693	– 32.50750	34.09631	Exception
27	95.88819	– 128.49117	– 246.64559	11.76410	Exception
28	35.02538	43.58398	57.59044	59.63265	–
29	– 159.73506	– 0.01600	0.00399	– 0.01599	Exception
30	48.44588	32.31017	60.41050	40.77646	–

any other occurrence is what we have studied and analysed, that for how long do people keep talking or in this case keep tweeting about. It is important to understand how impactful event one tweet can get if it can generate multiple responses and those responses generate multiple responses which sets off a chain effect. This causes a lot of people who may have little to no knowledge about the conversation to also join in, which in turn creates a scenario in which seems like a narrative is being dictated (American Institute of Physics 2014). Although in the average incident related tweets, the photos and graphs above show that the average tweets per day started to fall off after 5 days. This shows that in most cases, people usually tend to send a tweet or two within a span of five days of the incident, when the hype around that particular event or incident is still recent. When the hype begins to die down, the tweets, the news and conversations about that event or incident also trickles down. This says

something about the human psychology, that it is influenced by what is going around in the surroundings but does not remain influenced for a long time except for certain incidents

6 Conclusion

From the current research, the authors concluded that Social Media usage is increasing at a very high rate and many people rely on Twitter for the latest news. Many researchers are currently working on how the trends on Twitter change with time. This research found that a no matter what the news is related to, it stays in trending for 5–6 days, which can be concluded from the graphs attached, which also indicates the human behaviour/psychology towards a particular topic. Also analyzing the trends in graph keeping in mind the type of incident and its approximate reach to general public, we have also mentioned in results how the news breaks out on

```

Linear Regression Model

day5 =

    0.8992 * day4 +
    -737.7409

Time taken to build model: 0.2 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient           0.9391
Mean absolute error              3269.4629
Root mean squared error          5163.4821
Relative absolute error          27.1231 %
Root relative squared error      33.4455 %
Total Number of Instances       30

```

Fig. 8 Data with Exceptions

```

Linear Regression Model

day5 =

    -0.0762 * day2 +
    -0.1524 * day3 +
    1.0928 * day4 +
    508.7197

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient           0.9553
Mean absolute error              2397.5667
Root mean squared error          3605.3522
Relative absolute error          25.2314 %
Root relative squared error      28.5 %
Total Number of Instances       23

```

Fig. 9 Data without exception

Twitter and how is it expected to grow based on previous incidents seen in the same category of incidents. The authors have also mentioned in result section the various reasons for why some tweets on a similar incident may get explode on Twitter in terms of tweet volume while some see a stable and then decreasing trend. The analysis on the

data we procured also shows that there is a lot of overlap in tweets in incidents where two events having probable effect on each other happen near to each other w.r.t to the time of incident. We would like to mention that the purpose of this study was to understand how much does Twitter affect news instances, by getting them to spread

```

Linear Regression Model

day5 =

    0.747 * day3 +
    -2047.7662

Time taken to build model: 0 seconds

=== Cross-validation ===
=== Summary ===

Correlation coefficient          0.9213
Mean absolute error             2448.9291
Root mean squared error         4096.5106
Relative absolute error         33.0538 %
Root relative squared error     37.0927 %
Total Number of Instances      14

```

Fig. 10 Purely decreasing Data

about and for how long the discussion, debate or deliberation on a certain topic happens on Twitter after the initial mention of an incident.

Funding The authors did not receive support from any organisation for the submitted work.

Declaration

Conflict of interest The authors declare that there has not been any conflict amongst the authors in the work stated.

References

- Albakour MD, Macdonald C, Ounis I (2013) Identifying local events by using microblogs as social sensors. In: OAIR. ACM, pp 173–180
- Aldayel A, Magdy W (2022) Characterizing the role of bots' in polarized stance on social media. *Soc Netw Anal Min* 12:30. <https://doi.org/10.1007/s13278-022-00858-z>
- Alexa (2019) "Alexa Top Sites,"
- American Institute of Physics (AIP) (2014) How Twitter shapes public opinion. *ScienceDaily*. ScienceDaily, 11. <www.sciencedaily.com/releases/2014/03/140311123816.htm>
- Anandarao S, Chellasamy SH (2021) Detection of hot topic in tweets using modified density peak clustering. *Ing Syst Inf*, 26(6):523–531. <https://doi.org/10.18280/isi.260602>
- Annamoradnejad I, Habibi J (2019) A comprehensive analysis of Twitter trending topics. In: 2019 5th international conference on web research (ICWR), pp 22–27. <https://doi.org/10.1109/ICWR.2019.8765252>.
- Asur S, Huberman BA, Szabó Gá, Wang C (2011) Trends in social media : persistence and decay. In: 5th international AAAI conference on weblogs and social media. <https://doi.org/10.2139/ssrn.1755748>
- Balaji TK, Annavarapu CSR, Bablani A (2021) Machine learning algorithms for social media analysis: a survey. *Comput Sci Rev* 40:100395. <https://doi.org/10.1016/j.cosrev.2021.100395> (ISSN 1574–0137)
- Batrinca B, Treleaven PC (2015) Social media analytics: a survey of techniques, tools and platforms. *AI Soc* 30:89–116. <https://doi.org/10.1007/s00146-014-0549-4>
- Botambu C, Hoang DT, Nguyen NT, Dosam H (2021) Trends in combating fake news on social media-a survey. *J Inf Telecommun* 5(2):247–266. <https://doi.org/10.1080/24751839.2020.1847379>
- Cha M, Haddadi H, Benevenuto F, Gummadi KP (2010) Measuring user influence in Twitter: the million follower fallacy. In: Fourth international AAAI conference on weblogs and social media
- Cheong M, Lee V (2009) Integrating web-based intelligence retrieval and decision making from the Twitter trends knowledge base. In: Proc. CIKM 2009 co-located workshops: SWSM 2009, pp 1–8
- D'Monte, L (2009) "Swine Flu's Tweet Tweet Causes Online Flutter". *Business Standard*. Retrieved February 4, 2011. Also known as the 'SMS of the internet', Twitter is a free social networking service
- Fan W, Gordon M (2014) The power of social media analytics. *Commun ACM* 57:74–81. <https://doi.org/10.1145/2602574>
- Gao D, Li W, Cai X, Zhang R, Ouyang Y (2014) Sequential summarization: a full view of Twitter trending topics. *IEEE/ACM Trans Audio, Speech, Language Process* 22(2):293–302
- Guo B, Ding Y, Yao L, Liang Y, Zhiwen Y (2020) The future of false information detection on social media: new perspectives and trends. *ACM Comput Surv* 53(4):36. <https://doi.org/10.1145/3393880>
- Gupta A, Bansal A, Mamgain K, Gupta A (2022) An exploratory analysis on the unfold of fake news during COVID-19 pandemic. In: Somani AK, Mundra A, Doss R, Bhattacharya S (eds) *Smart systems: innovations in computing*. smart innovation,

- systems and technologies, 235th edn. Springer, Singapore. https://doi.org/10.1007/978-981-16-2877-1_24
- Haas C, Wearden ST (2003) E-credibility: building common ground in Web environments. *L1 Educ Stud Lang Lit* 3:169–184
<https://datareportal.com/social-media-users>
<https://digitalmarketinginstitute.com/blog/social-media-what-count-ries-use-it-most-and-what-are-they-using>
<https://github.com/twintproject/twint>
<https://www.pewresearch.org/internet/fact-sheet/social-media/>
<https://www.rivier.edu/academics/blog-posts/an-introduction-to-behavioral-psychology>
<https://www.statista.com/statistics/242606/number-of-active-twitter-users-in-selected-countries/>
<https://www.statisticshowto.com/probability-and-statistics/correlation-coefficient-formula/>
<https://www.visualcapitalist.com/ranked-social-networks-world-wide-by-users/>
- Jansen BJ, Zhang M, Sobel K, Chowdury A (2009) Twitter power: tweets as electronic word of mouth. *J Am Soc Inf Sci* 60(11):2169–2188
- Krishnamurthy B, Gill P, Arlitt M (2008) A few chirps about Twitter. In: Proc. WOSN, pp 19–24
- Lee K, Palsetia D, Narayanan R, Patwary M. A, Agrawal A, Choudhary A (2011) Twitter trending topic classification. pp 251–258
- Lewin K (1947) Frontiers in group dynamics: concept, method and reality in science; social equilibria and social change. *Hum Relat* 1:5–40
- Lin C, Ouyang Z, Wang X, Li H, Huang Z (2021) Preserve integrity in realtime event summarization. *ACM Trans Knowl Discov Data* 15(2):29
- Mayer M (2009) What the trend? Available from <http://www.whattrend.com>
- Mendoza M, Tesconi M, Cresci S (2020) Bots in social and interaction networks: detection and impact estimation. *ACM Trans Inf Syst* 39(1):32. <https://doi.org/10.1145/3419369>
- Metzger MJ, Flanagin AJ (2008) Digital media, youth, and credibility (pp 73–100). Cambridge, MA: The MIT Press
- Murthy D (2013) Twitter: social communication in the Twitter age. Polity Press, Cambridge, UK
- Romero DM, Galuba W, Asur S, Huberman BA (2011) Influence and passivity in social media. In: 20th international world wide web conference (WWW'11)
- Sakaki T, Okazaki M, Matsuo Y (2010) Earthquake shakes Twitter users: real-time event detection by social sensors. In: Proceedings of the 19th international conference on World wide web
- Shu K, Mahudeswaran D, Wang S, Lee D, Liu H (2020) Fakenewsnet: a data repository with news content, social context and spatiotemporal information for studying fake news on social media. *Big Data* 8(3):171–188
- Singh S, Arya P, Patel A, Tiwari AK (2019) Social media analysis through big data analytics: a survey. In: Proceedings of 2nd international conference on advanced computing and software engineering (ICACSE) 2019, Available at SSRN: <https://ssrn.com/abstract=3349561> or <https://doi.org/10.2139/ssrn.3349561>
- Stieglitz S, Mirbabaie M, Ross B, Neuberger C (2018) Social media analytics - Challenges in topic discovery, data collection, and data preparation. *Int J Inf Manage* 39:156–168. <https://doi.org/10.1016/j.ijinfomgt.2017.12.002> (ISSN 0268-4012)
- Ullah KH, Shumaila N, Kishwar N, Danial S, Ahsan M (2021) Twitter trends: a ranking algorithm analysis on real time data. *Expert Syst Appl* 164:113990. <https://doi.org/10.1016/j.eswa.2020.113990> (ISSN 0957-4174)
- Yang J, Leskovec J (2011) Patterns of temporal variation in online media. In: Proceedings of the fourth ACM international conference on Web search and data mining, WSDM '11, pp 177–186
- Zhao W, Shu B, Jiang J, Song Y, Yan H, Li X (2012) Identifying event-related bursts via social media activities. In: Proceedings of the 2012 joint conference on empirical methods in natural language processing and computational natural language learning, pp 1466–1477. Association for Computational Linguistics
- Zubiag A, Martínez-Unanue R, Fernández V (2009) Getting the most out of social annotations for Web page classification. In: DocEng'09 - Proceedings of the 2009 ACM symposium on document engineering. pp 74–83. <https://doi.org/10.1145/1600193.1600211>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.