

Mapping the social landscape through social media

Journal of Information Science
2020, Vol. 46(6) 776–789
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0165551519865487
journals.sagepub.com/home/jis



Girish Chandra Joshi 

Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, India

Mayuri Paul

Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, India

Bhriku Kumar Kalita

Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, India

Vikram Ranga

Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, India

Jiwan Singh Rawat

Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, India

Pinkesh Singh Rawat

Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, India

Abstract

Being a habitat of the global village, every place has established connections through the strength and power of social media, piercing through the political boundaries. Social media is a digital platform, where people across the world can interact. This has a number of advantages of being universal, anonymous, easy accessibility, indirect interaction, gathering and sharing information when compared with direct interaction. The easy access to social networking sites (SNSs) such as Facebook, Twitter and blogs has brought about unprecedented opportunities for citizens to voice their opinions loaded with emotions/sentiments. Furthermore, social media can influence human thoughts. A recent incident of public importance had presented an opportunity to map the sentiments, involved around it. Sentiments were extracted from tweets for a week. These sentiments were classified as positive, negative and neutral and were mapped in geographic information system (GIS) environment. It was found that the number of tweets diminished by 91% over a week from 25 August 2017 to 31 August 2017. Maximum tweets emerged from places near the origin of the case (Haryana, Delhi and Punjab). The trend of sentiments was found to be – neutral (47.4%), negative (30%) and positive (22.6%). Interestingly, tweets were also coming from unexpected places such as United States, United Kingdom and West Asia. The result can also be used to assess the spatial distribution of digital penetration in India. The highest concentration was found to be around metropolitan cities, that is, Mumbai, Delhi and lowest in North East India and Jammu & Kashmir indicating the penetration of SNSs.

Keywords

Digital literacy; GIS; sentiment mapping; spatio-temporal analysis; text mining

Corresponding author:

Vikram Ranga, Centre of Excellence for NRDMS in Uttarakhand, Department of Geography, Kumaun University, SSJ Campus, Almora 263601, Uttarakhand, India.

Email: vikram.ranga85@gmail.com

1. Introduction

Social media is spreading its footprints rapidly and effectively worldwide. Social media refers to the websites and applications that enable users to create and share content or to participate in social networking (Oxford dictionary). With the pace of modernisation, the boom in Internet users is growing not only in developed world but also in the developing countries. As of 30 June 2017, Asia accounts for covering 49.7% of the total Internet users, followed by Europe with 17% and Australia/Oceania comprises least number of Internet users with 0.7%, while North America was ranked first with a social media penetration rate of 88.1%, followed by Europe with 80.2%, Australia/Oceania with 69.6%, Latin America/Caribbean with 62.4%, Middle East with 58.7%, Asia with 46.7%, and least by Africa with 31.2% against global average of 51.7%. Internet penetration is also increasing at a lightning speed as evident from African experience where growth rate was recorded to be 8503.1% from 2000 to 2017, followed by Middle East with 4374.3%, Latin America/Caribbean with 2137.4%, Asia with 1595.5%, Europe with 527.6%, Oceania/Australia with 269.8% and least by North America with 196.1% against the global average of 976.4% [1]. Being an open and user-friendly platform, users range across wide realm of the globe. Social media gained hike as users can exchange thoughts and ideas through SNSs allowing to disperse published information, pictures, videos, etc., seeking the attention of other fellow users distributed worldwide. Furthermore, the users of SNSs are free to express their views without the fear of being directly judged and do not experience any social barrier [2]. The posts on social media are often coloured in sentiments such as positive, negative, joyous and sad, thus forming a sentimental landscape regarding any topic/issue at hand. Sentimental landscape can be thought as a virtual landscape where the emotions, opinions and sentiments related to any topic or issue are published. Twitter,^a one of the most popular social media platforms, is used by people to express their views or voice their opinions with very little hindrance. These opinions expressed in the form of tweets can form chain as users post similar/contrasting tweets or retweet the same post. The first spectacular case in which human thoughts were influenced by social media was the ‘Jasmine Revolution’, which was a form of civil resistance in Tunisia [3]. The war fought on Twitter around #AllEyesOnISIS, announcing the invasion of Islamic State of Iraq and Syria (ISIS) on northern Iraq in 2014 was deemed as an Internet war [4]. SNSs played an active role in spreading the message of the terrorist group and encouraging particularly young people to travel to Middle East and join the group and/or play a supporting role [5].

These sentiments published in SNSs, therefore, can be studied to construct a sentimental landscape using sentiment analysis (SA) which is a technique by which one can extract the sentiments inscribed within the text [6,7]. The realm of the spread of human imprints in the form of emotions rising from the contraction and expansion of sentiments over the landscape can be collected together to visualise in the form of a map. Mapping the sentiments of humans in spatial scenario depicts that the social ground is an emerging phase of social media study [8–10]. Sentiment mapping (SM) helps in mapping the sentiments flowing over the web geographically [11]. SM in contrast to SA produces its result in the form of maps for visual treat of the end result showing the sentiments varying over the landscape (globe) [9]. In recent years, opinions published in social media have helped in reshaping businesses, product ratings, public awareness on global climatic change and human emotions, which have had profound impact on our social and political landscape [12,13]. The purpose of this study is to analyse the role of social media as a strong conveyer of emotions in a landscape and to depict the results gathered in the form of maps. The research questions for this study were:

1. How can social media platform be used to map sentiments?
2. How do the dynamics of mapped sentiments vary over space and over a time period?

2. Study area

The study area is based on the backdrop of the alleged rape case of the third chief (Baba Ram Rahim) of Dera Sacha Sauda, which is a non-profitable social welfare and spiritual organisation, established in April 1949 by Khema Mal Ji (Shah Mastana Ji Balochistani). Gurmeet Ram Rahim is the present chief of the Dera since 23 September 1990. The main centre of Dera Sacha Sauda (29°32' 1.51" N latitude and 75°1' 3.8" E longitude) is situated at Sirsa, Haryana, India, as depicted in Figure 1. Dera or ashram is a gist of monastic community or a place of religious retreat. In India, people visit Dera or Ashram for mythic value and the charismatic appeal of the Babas/Masters of the ashram. Furthermore, it also provides a sense of security to the visitors as it promotes equality, thus attracting followers from all stratas of the society. The backdrop of sexual allegation on the Head of the Dera was aired after the reopening of the case on 25 August 2017, which was filed on December 2002. The Master of the Dera had an allegation of sexually harassing two sadhavis (female monks). The case gained the nation's attention after the disciples of Ram Rahim resorted to violence near the origin of the case, that is, Sirsa, Haryana. On 28 August 2017, the Dera Sacha Sauda Chief was pronounced guilty and sentenced to 20 years of jail. The case took an ugly phase after 28 August 2017, when the followers against the ruling by a Court

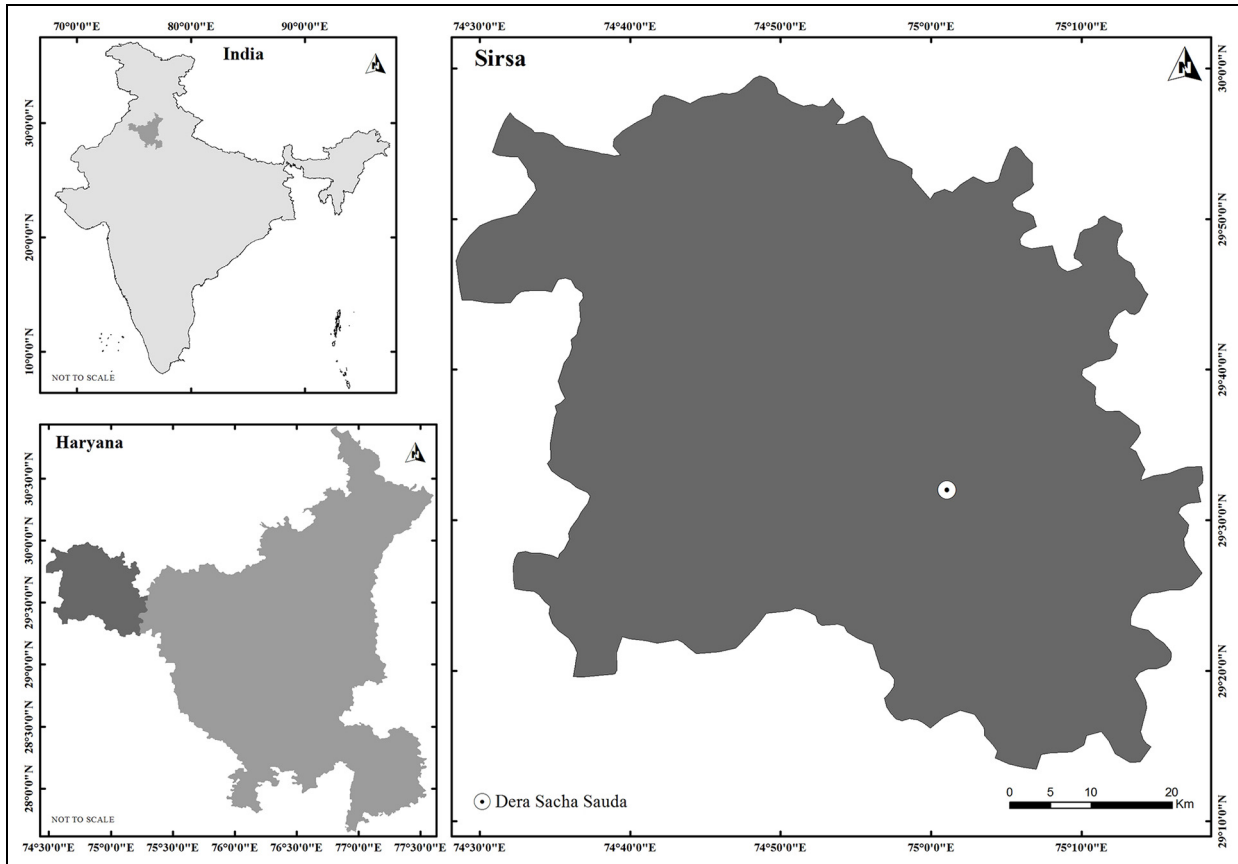


Figure 1. Study area map.

got involved in violence [14]. As a result, the followers of Master residing at New Delhi, Punjab and Haryana turned out to create violence against the decision of the court. The violence left 36 dead and over 250 injured [15]. At a preliminary stage, only India was considered as the study area but later the world was also included as the tweets were recorded from other parts of the world.

3. Dataset and methodology

3.1. Data acquisition

Dataset for the SA and the further SM process was collected from Twitter. R [16] was utilised for gathering and further processing the data. To collect the data, tweets were downloaded over a period of a week from 25 August 2017 to 31 August 2017 through Twitter application protocol interface (API). This particular time period was taken in order to get maximum sentiments related to the case flowing in Twitter as it opened on 25 August 2017 and the judgement came on 28 August 2017. Twitter API is a public domain which was accessed using the R-text mining process. ‘Text mining’ is a part of ‘Data mining’ process. ‘Data mining’ is a process by which interesting knowledge can be discovered from large number of data [17].

For downloading the tweets using the Twitter API, an application was created with Twitter [18]. After that, access tokens were generated in order get authorisation from Twitter for downloading the tweets. ‘ROAuth’ [19] package was used for getting authorisation and streaming tweets from Twitter. ‘RCurl’ [20] was used to create the general network as HyperText Transfer Protocol(s) (HTTP) and File Transfer Protocol (FTP) to establish the client interface in R Studio. The Twitter API allows users to search for a particular word or phrase they are looking for. Tweets containing the term ‘RamRahim’ OR #RamRahim OR ‘Ram Rahim’ OR ‘Ram Raheem’ OR ‘RamRaheem’ were downloaded using ‘searchTwitter’ function under the ‘twitterR’ [21] package. A call of 3000 tweets published in English only was made for each day. The ‘twitterR’ package was used for the whole data acquisition process. One limitation encountered in the data

acquisition process was that the tweets were getting repeated in the dataset. Sometimes, the Twitter API provided 3000 tweets with repeated tweets, and sometimes it provided < 3000 tweets including repeated tweets. Unique tweets were extracted using ‘duplicated’ function under ‘base’ [16] package, and separate data frames were prepared containing unique tweets for each day.

3.2. Retrieval of locations

Another major challenge for the study was to find the location of the extracted tweets because Twitter API does not provide the exact geographic coordinates of its users. However, search radius can be set for specific regions from where tweets can be downloaded. Many users, voluntarily, make their location public. Consequently, locations were retrieved only for 47.02% of the total unique tweets. The list of user information was converted to a data frame containing all information about users. Some false entries in the location field were also found in the profile information dataset like ‘on the planet Earth’, ‘aapkedilmein’ (inside your heart) and ‘in hell’ and locations with multiple addresses like ‘Delhi/New York’. So, these noises were removed from the dataset manually in MS Excel 2007, and a dataset with reliable location information was prepared. The locations in this dataset were geocoded^b with the help of the ‘ggmap’ [23] package, and their corresponding geographic coordinates (latitude/longitude) were retrieved. The locations with nil/NA values were deleted in MS Excel. A dataset was prepared by merging the dataset containing unique tweets and the dataset with geographic coordinates using the common field called ‘screenName’, which contains the handle name. Furthermore, tweets were cleaned in order to make a corpus of text contained within the tweets, which would be clean and appropriate for the sentiment polarity classification process.

3.3. Calculation of sentiment score and polarity classification

For SA, different methods are available for extracting the sentiment automatically. First method is the lexicon-based approach, which involves calculating the orientation of a document from the present semantic orientation of words or phrases published [24]. Second method is the text classification approach, which involves building classes from the given texts or sentences [25], essentially a supervised classification of assignment of sentiment score in context to the polarity of the published texts. The latter approach is also known as the statistical or machine-learning approach [26]. In this study, the second method was used for sentiment classification or the polarity classification process as the first method is usually used in identifying the polarity in a bigger domain such as for a whole document or an article. The tweets were classified to identify the polarity of each extracted tweet. A sentiment score for cleaned tweets was calculated based on the positive/negative words embedded in the tweet. The sentiment scores were calculated by using the dictionaries of positive and negative words provided by Liu et al. [27]. These two dictionaries contain positive and negative words in two text files separately. Each tweet was searched for positive and negative words listed in these dictionaries. Every positive and negative word was given a score of +1 and -1, respectively, and the overall sentiment score for a tweet was calculated by adding these scores. Polarity classification was done based on the overall sentiment score of the tweet, and the polarities were defined as positive (score > 0), negative (score < 0) and neutral (score = 0) using ‘plyr’ [28] and ‘stringr’ [29] packages. Separate datasets were prepared for each polarity for each day for mapping purpose.

3.4. Sentiment Mapping

This step involved mapping of sentiments embedded within the tweets on a geographic map. Arc Map (version 9.3) software was used for mapping purpose. Kernel Density^c (KD) rasters were made for each polarity for each day. These rasters were then reclassified to a uniform scale based on the values of KD rasters for the week. A series of maps were made from these rasters for each polarity, showing the distribution of tweets over a period of week. Separate maps were made for each polarity for India and the world. The output cell size for producing KD rasters was set to 16 and 133 km for India and the world, respectively. The whole set of methodology has also been described in the form of a chart (Figure 2).

4. Results and discussion

The results obtained from the study are shown in the form of graphs and maps. Figure 3 represents the temporal variation in the collection of tweets. It was found that the number of tweets had been decreasing continuously from 25 August 2017 to 31 August 2017 and varied greatly over the week.

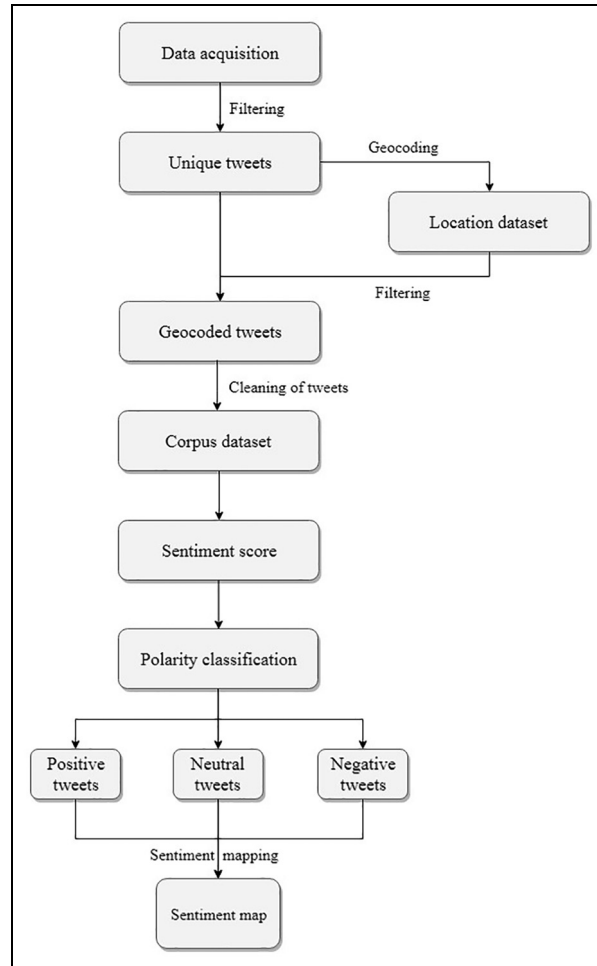


Figure 2. Methodology flowchart.

The total number of unique tweets collected over the study period were 1863, where the maximum number of tweets were 646 (34.7%) on the first day itself (i.e. 25 August 2017) followed by 341 (18.3%) on the next day. A rise in the number of tweets was witnessed on 28 August 2017, raising the number to 373 (20%) from 189 (10%) on the previous day. The sudden increment was due to the fact that the accused was pronounced guilty and was sentenced to imprisonment for 20 years on that day. As the day was marked as the Day of Judgement, people started sharing their thoughts and emotions on the topic over social media, and Twitter witnessed the fact. The tweets started diminishing over time as the concern of the people started decreasing at the rate of 91.6% per week.

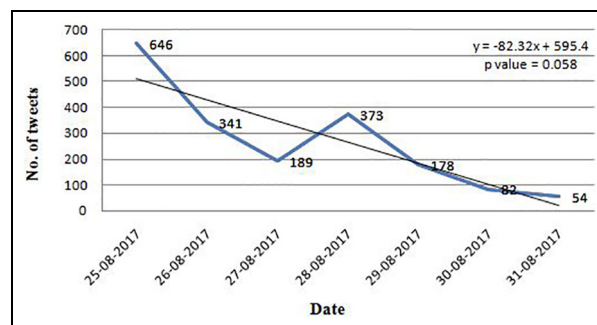


Figure 3. Variability of number of tweets over the week.

Figure 4 shows the relative percentage of positive, negative and neutral sentiments collected on each day. It was found that the sentiments with neutral polarity were most prevalent followed by tweets with negative polarity and then positive polarity over the week. Sentiments with negative polarity were prevailing more than positive in the first half, most probably, because most of the people were criticising the accused (Figure 4). Sentiments with positive polarity were prevailing more in the later half as compared to the first half because the people were supporting the judgement given by the court (Figure 4). The overall trend of sentiments for the whole week was found to be - neutral (47.4%), negative (30%) and positive (22.6%)⁷.

The locations of sentiments of different polarities, that is, negative, positive and neutral were mapped for India (Figures 5–7) and then for the world (Figures 8–10). These figures helped in spatio-temporal analysis of sentiments. Spatio-temporal analysis deals with spatial and temporal aspects of data. It tells us how the data are varying in geographic space over a period of time. This spatio-temporal analysis helped in analysing how the tweets were varying over the space for a certain period of time.

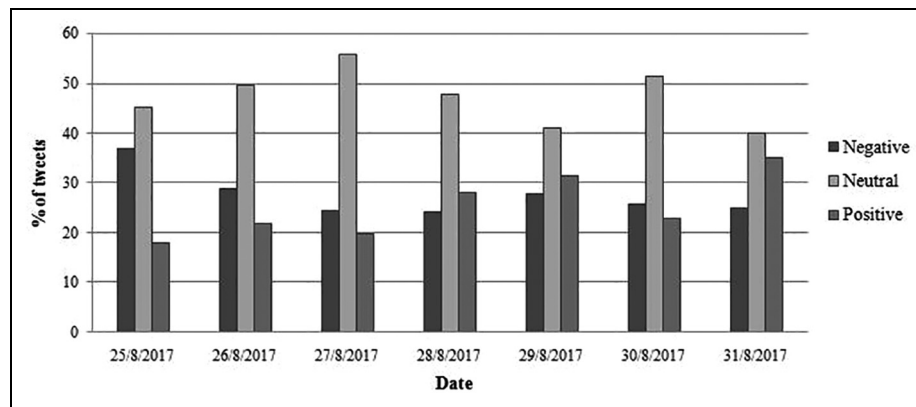


Figure 4. Polarity distribution of tweets per day in percentage.

The maximum tweets were collected from North India followed by Central India and then South India. Maximum tweets were collected on 25 August 2017 because of the fact that it was a ‘hot topic’ of discussion when it came into limelight on this day. The tweets diminished from A to C, but the number of tweets increased on D (i.e. on 28 August 2017) as it was the day when judgement came resulting in the imprisonment of the accused. The tweets again diminished after D up to G gradually (Figures 5–7). Through the whole week, the area with tweets having maximum negative polarity was found to be near the origin of the case and Delhi/NCR (National Capital Region) region. It was found that the density of tweets with positive polarity was more pronounced in southern India after the conviction of the accused. More number of positive tweets were collected from central and southern India compared to negative tweets after the conviction.

It was observed that, by and large, the whole country criticised the accused with negative tweets in the first half of the observational period although positive tweets were also observed. It was found that, in the second half, while the density of negative tweets lied mostly near Delhi/NCR region, positive tweets were more evenly distributed, especially in the north–south direction. Neutral tweets came from all over the country in the first half, but in the later half, they were only observed in northern India. The maps showed the distribution of tweets geographically demarcating the places from where the tweets originated, thus showing the reach or extent of social media. It was found that most tweets originated near the origin and were concentrated in Delhi/NCR region.

At the world level, tweets were received from United States, Australia, New Zealand, England, Western Europe, Indonesia and Arab Countries. Figures 8–10 show the diffusion of sentiments regarding the case over the globe.

Tweets from the rest of the world followed the same trend as that of tweets coming from India. This shows that the globe also keeps an eye over the news and stays active in the SNSs. The tweets become negligible until it ceased to zero over the time period. Very few numbers of tweets were observed in the second half of the observation. Figure 9 shows that very less number of tweets with positive sentiments came from the globe except for the Middle East region. It was also observed that the tweets with positive sentiments were more evenly distributed on 29 August 2017 compared to the first half of the study period (Figure 9). The density of neutral sentiments was low in the beginning and the end but a drastic change was observed on 28 August 2017 when the judgement came (Figure 10).

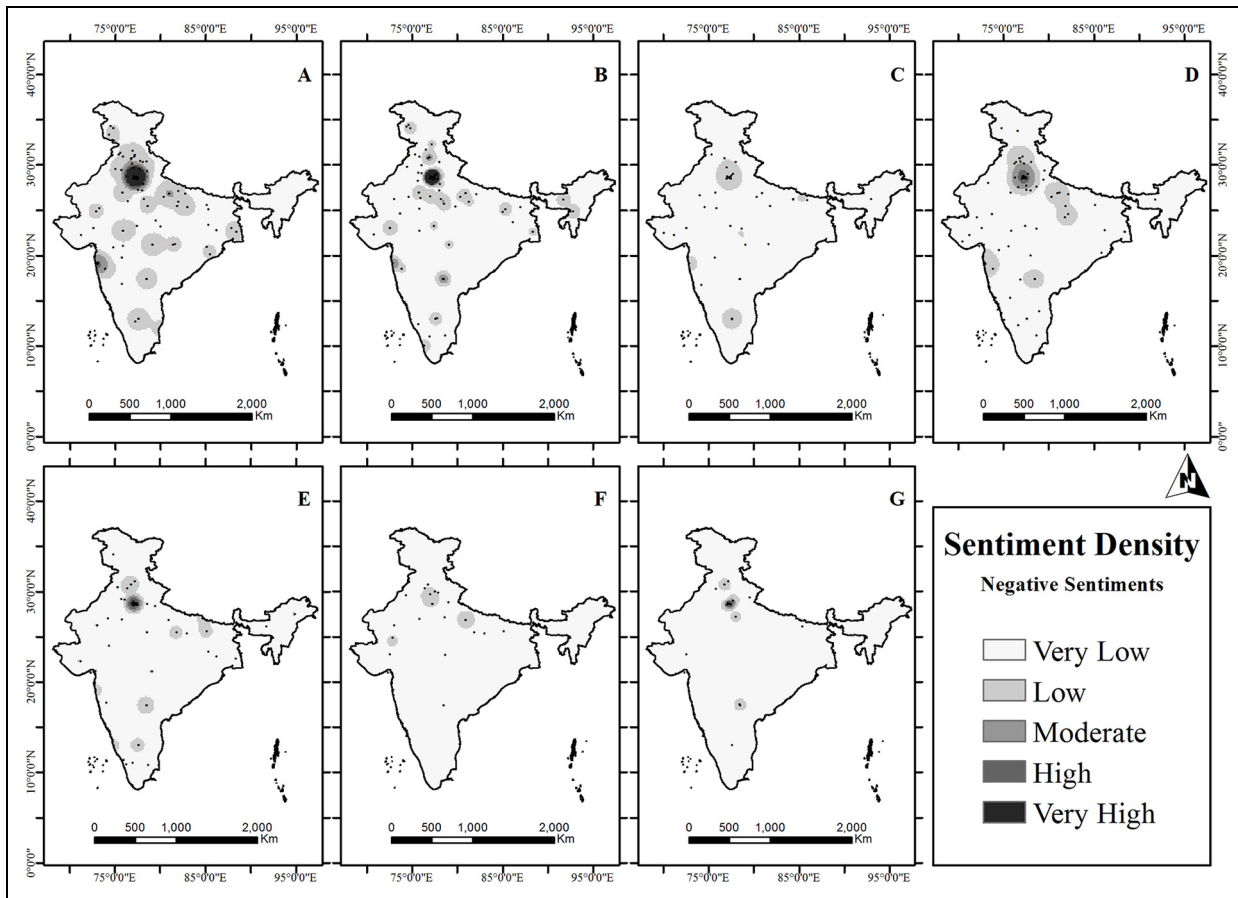


Figure 5. Spatio-temporal distribution of negative sentiments (India) from A to G (from 25 August 2017 to 31 August 2017).

At the world level, it was observed that the negative tweets were prevailing more than positive and neutral tweets. It was also observed that the density of positive tweets was more than negative tweets on 28 August 2017 (i.e. when the judgement came). It was found that the overall tweets diminished over time because people on social media react enormously when the topic is hot or trending but they ultimately forget the issue and move on. This study also showed that how the geography of a place affects the sentiments of people as it was seen that people near the origin of the case were more involved.

The study also highlighted the other aspects of such similar studies, for example, digital literacy, Internet penetration, awareness of people and trend of emotion dynamics. Digital literacy means that a person is familiar with computers, Internet or related technologies. With the help of Figures 5–7, a broad picture of digital literacy in India can be visualised. In the study, it was found that most of the states of India are digitally literate, meaning that there are users in these states who are familiar to Internet and are an active member of social media. Some states including Chhattisgarh and some North-Eastern states (except Assam and Meghalaya) were not active in tweeting relative to the issue. This shows that there is a digital divide between North-east India and the rest of India.

Most of the research work has been done in the field of SA [30–33], but a very little work has been done on SM [8,9]. SM can help us in visualising and analysing the results of SA and thus getting the real picture of the whole scenario. Also, it helps in visualising the dynamics of people’s sentiments/emotions over a period of time. SA only provides the statistics of the sentiments of people but by geocoding tweets, geography can be, effectively, related to these sentiments. Consequently, it can produce dynamic maps showing the polarities of sentiments. Thus, SM can play an important role in deciphering the trend of sentiments fluctuating over time, which can further help in making informed decisions.

Pang and Lee [30] performed SA on documents by identifying the overall polarity in sentences by applying standard machine learning techniques. In contrast to this, this study performed SA on the tweets collected from Twitter. Wilson [31] developed a system for automatically identifying the contextual polarity of a phrase in a two-step process. In the first step, the phrase was classified as polar or neutral, and then, in the second step, the contextual polarity of the phrases

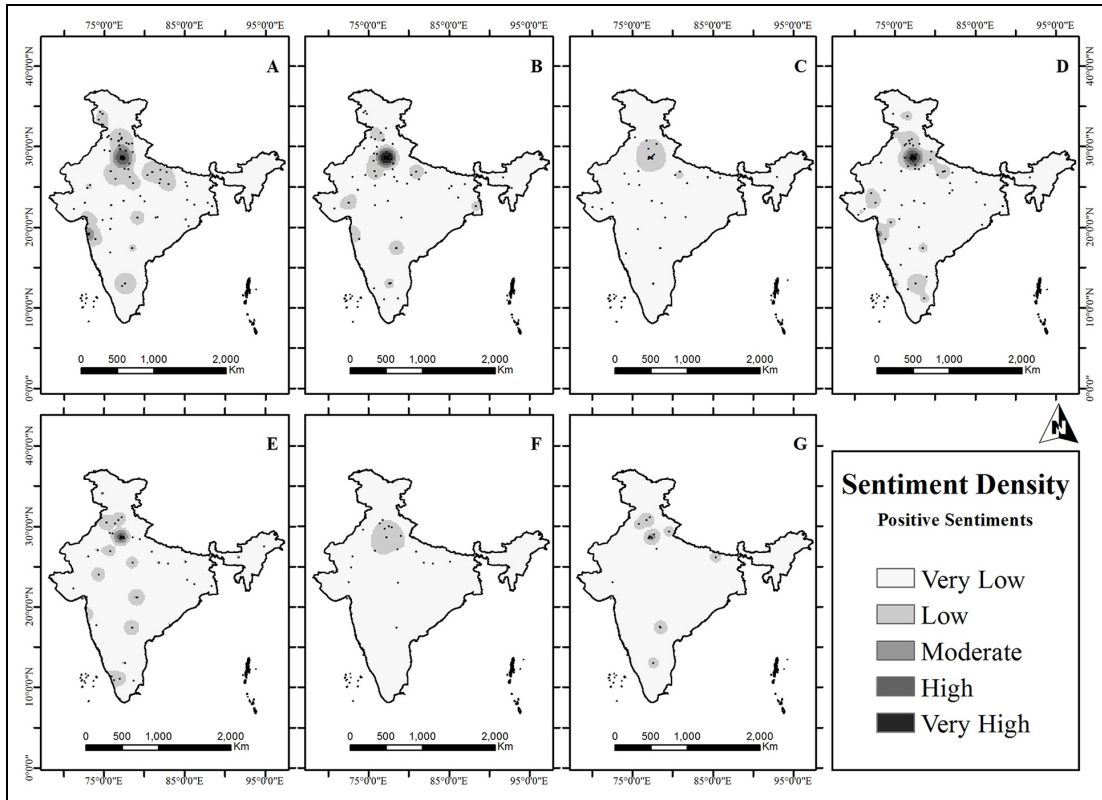


Figure 6. Spatio-temporal distribution of positive sentiments (India) from A to G (from 25 August 2017 to 31 August 2017).

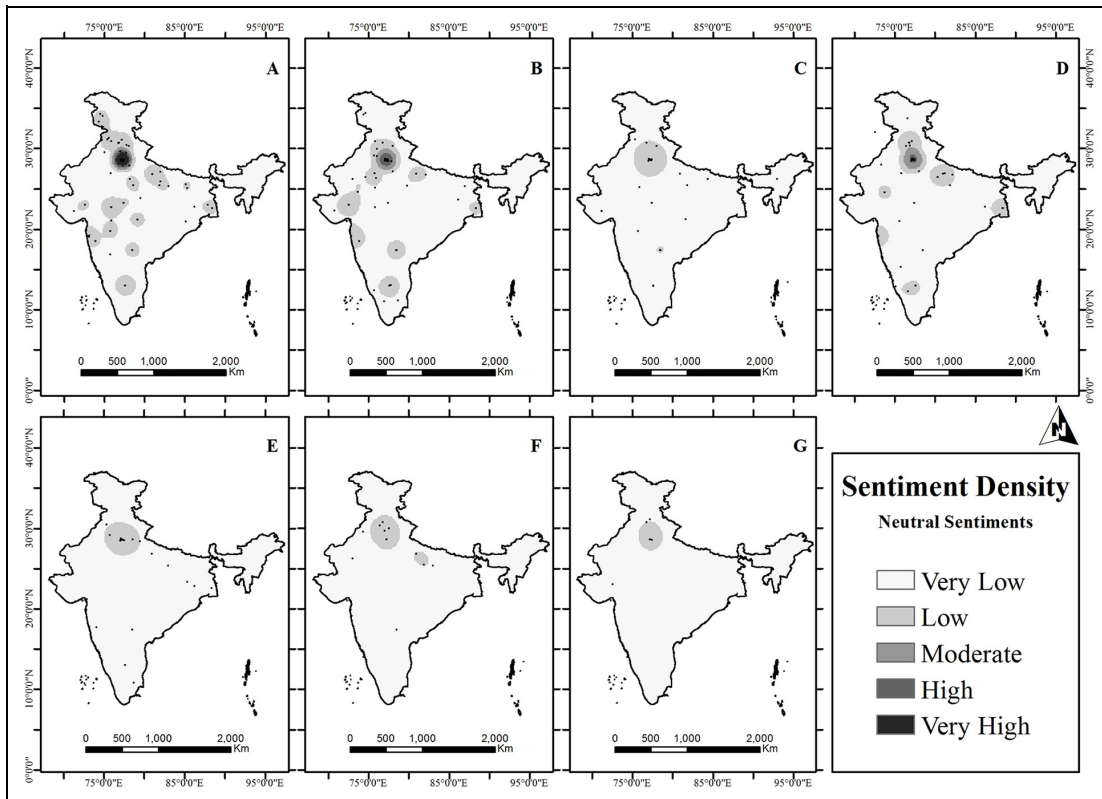


Figure 7. Spatio-temporal distribution of neutral sentiments (India) from A to G (from 25 August 2017 to 31 August 2017).

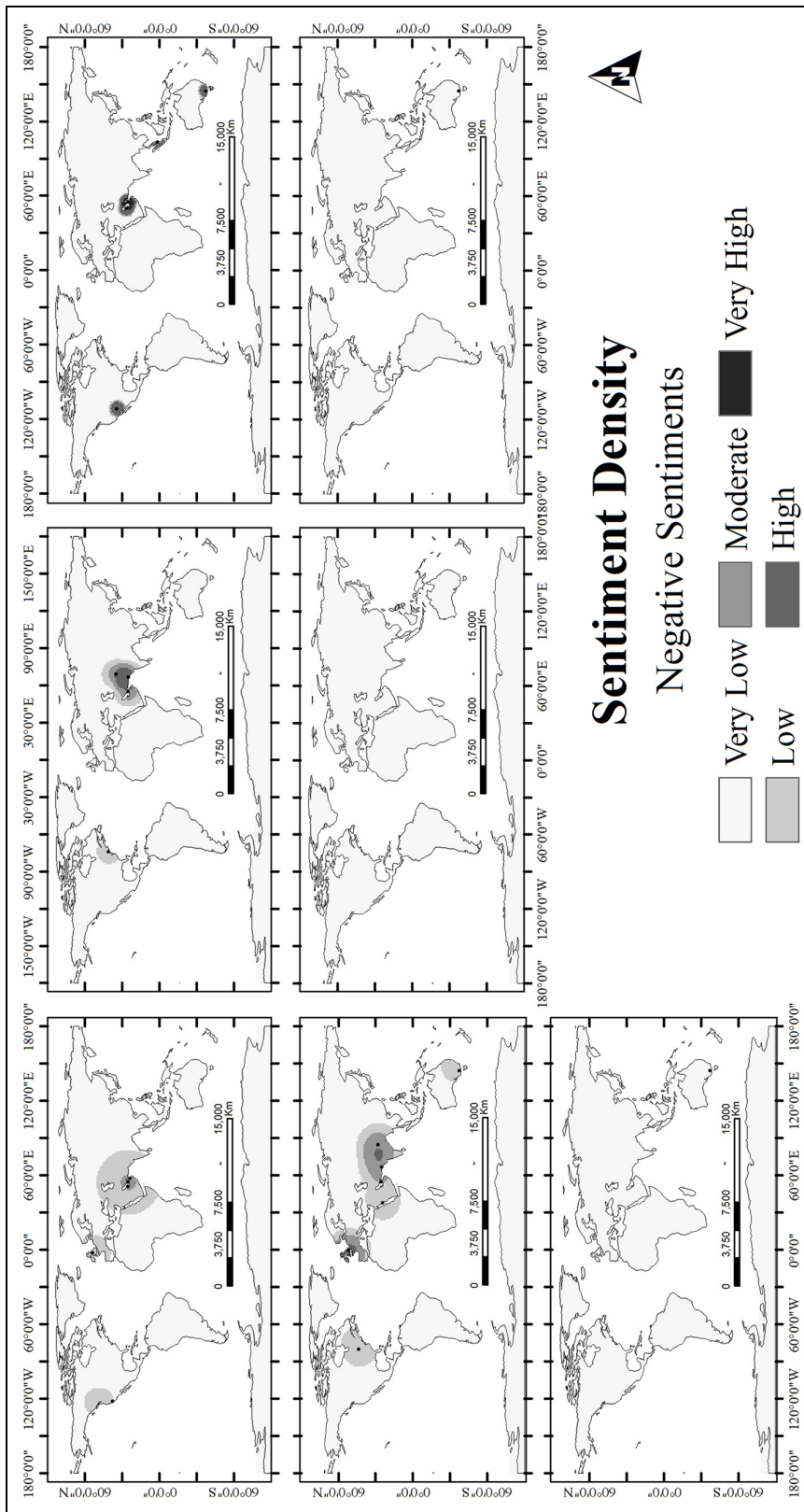


Figure 8. Spatio-temporal distribution of negative sentiments (world) from A to G (from 25 August 2017 to 31 August 2017).

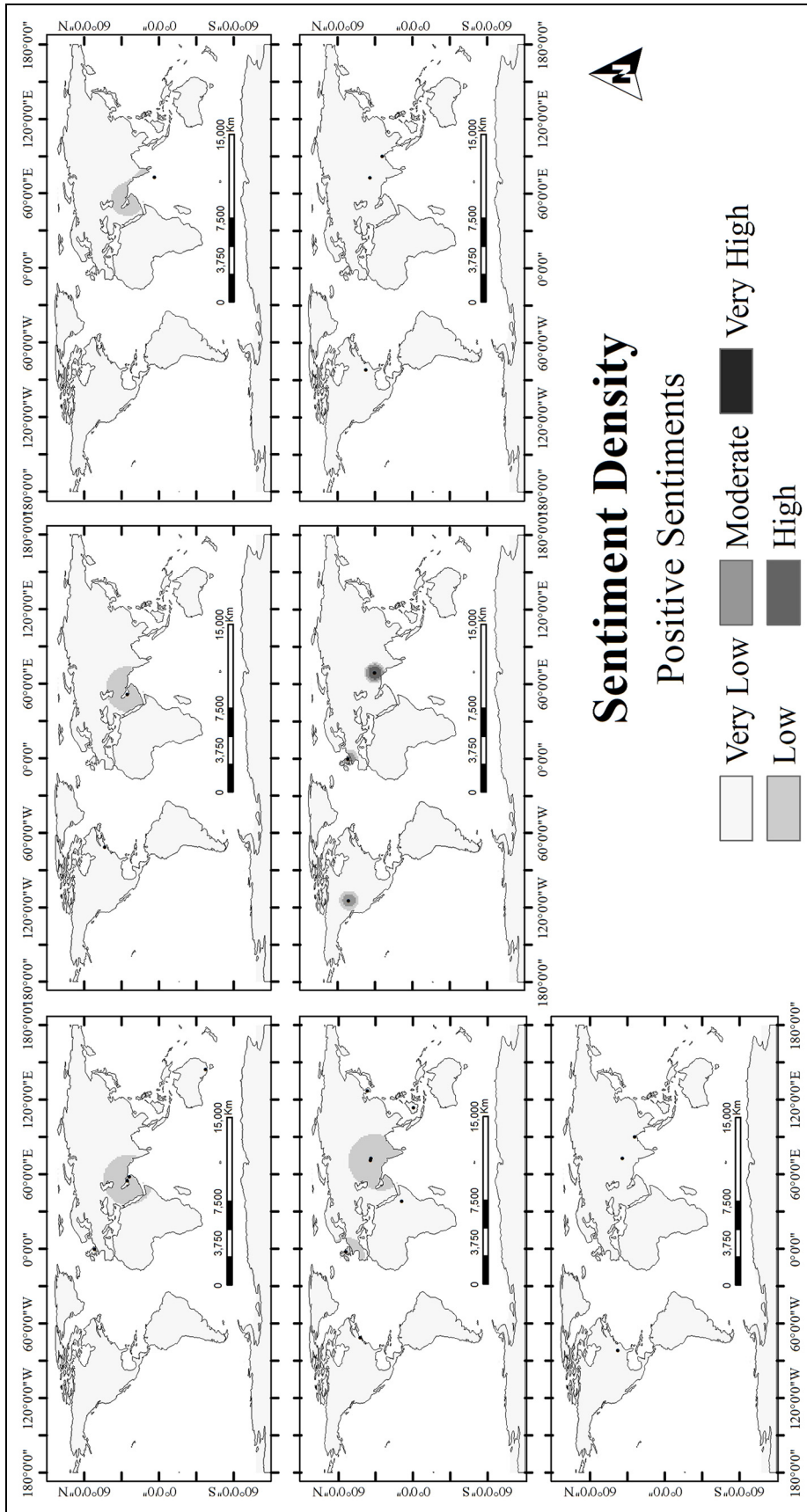


Figure 9. Spatio-temporal distribution of positive sentiments (world) from A to G (from 25 August 2017 to 31 August 2017).

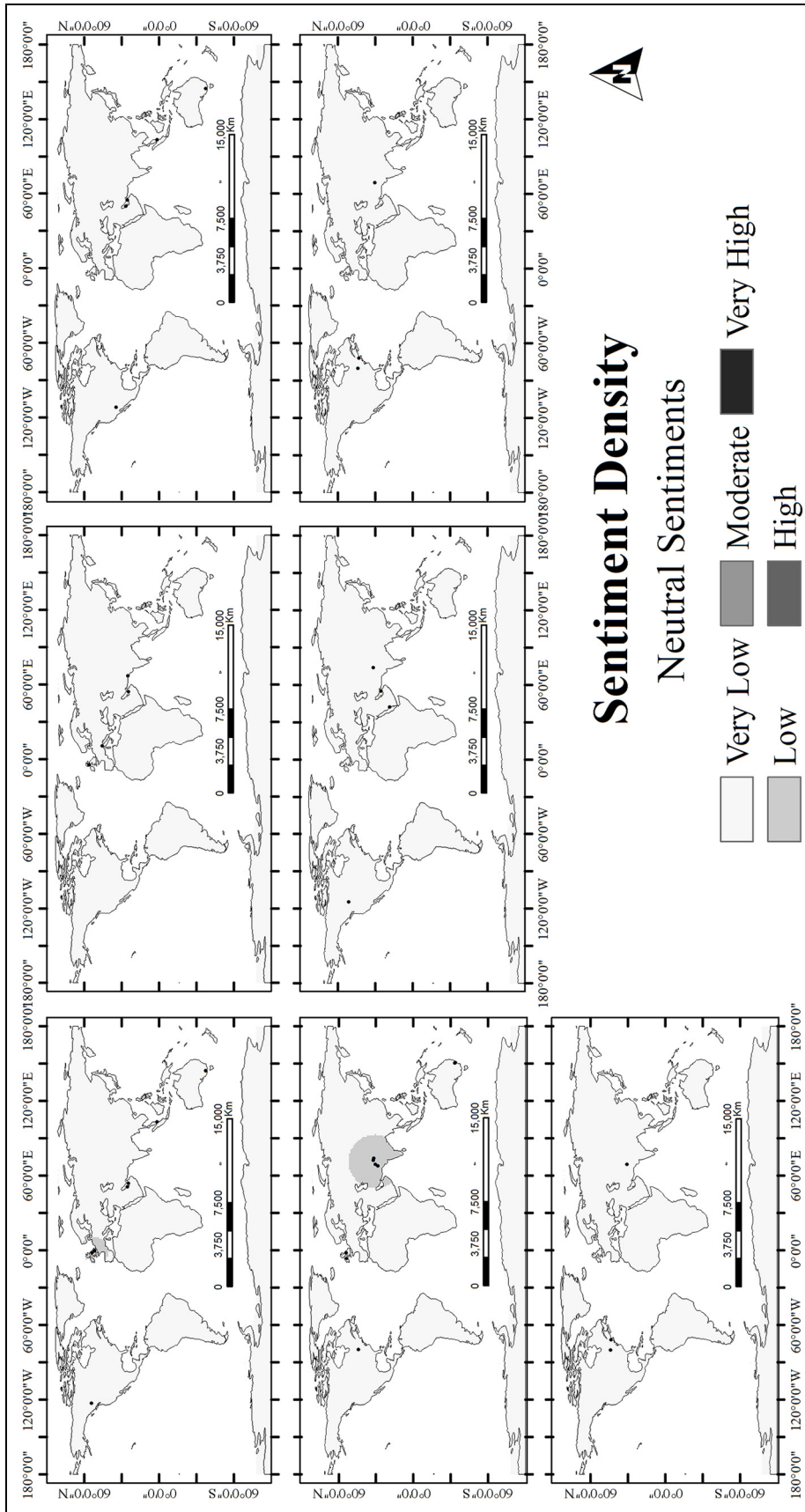


Figure 10. Spatio-temporal distribution of neutral sentiments (world) from A to G (from 25 August 2017 to 31 August 2017).

marked polar in the first step was identified. This article attempted to simplify the classified tweets as positive, negative or neutral by calculating the overall sentiment score of the tweet based on the algebraic sum of the sentiment score of each word which is polar. Taboada et al. [32] presented a lexicon-based approach, which is a word-based approach for extracting the sentiments from the text. They showed that a manually built dictionary can considerably enhance the sentiment classification process. Agarwal et al. [33] introduced two new resources which are ‘emoticon dictionary’ and ‘acronym dictionary’ for the sentiment classification process. This study used a lexicon-based method for sentiment classification by using the dictionary provided by Liu et al. [27]. Tweets were classified based on the sentiments they contain and then mapped to their corresponding geographic locations along with their polarity. One such work, in which sentiments were geo-mapped, was done by Caragea et al. [9]. They geo-mapped the sentiments derived from posts in Twitter during Hurricane Sandy. They also showed how people’s sentiments change relative to the distance from the disaster. They showed the relation between people’s moods and their proximity to the disaster. In this study also, a similar attempt was made to show how the geography of a place reshapes the emotional state of people as people get easily influenced emotionally by the trends in social media.

5. Conclusion

Human being has an inherent capacity to understand emotions of oneself and the surroundings too. Using this ability, one can channelise his/her own and others’ thoughts. Social media ends up being a very rich data source for human emotions. A very clear image of the extent of social media can be seen through the maps produced in this study. Social Media is a blend of micro-blogging sites and applications which makes the publications of thoughts in web essentially a vital factor of social media study. The published texts followed up by images and videos are loaded with human sentiments. SM is a new dimension to SA study given by the growing intelligent minds, which can help us to visualise the reality prevailing in a geographic landscape in the form of maps. Geography is the study related to the spatial context of any phenomenon or features, which can be human emotions too. Mapping the emotions in a specified boundary took shape after the assignment of polarity to the sentiments (grouped as positive, negative and neutral). With the help of SA, a proper analysis of sentiments embedded within the text was performed. In order to map the sentiments, the extraction of the locations of users is the most essential part. In this study, the locations of users publishing text were retrieved and then mapped in spatial context. Furthermore, spatio-temporal maps of sentiments were prepared to analyse the trend of sentiments flowing over Twitter over a period of time. The whole process lends itself to produce a sentiment map, which visually speaks about the details of the work done.

It can be concluded that the geography of a place certainly plays an important role in influencing the emotions of people. It was found that people, who were closer, geographically, to the place of the origin of the case, were more vulnerable and participated enormously in publishing tweets. It was also found that the people of metropolitan cities participated contrastingly higher than other cities.

The journey from SA to SM is an initiative to showcase the power of social media in the present day context binding it with geographic information system (GIS) techniques. This study also intends to show the strength of combination of GIS and R-programming language. GIS played an important role in mapping the sentiments and thus producing sentiment maps, which aided in visualising the ground reality. Finally, as far as the futuristic growth is concerned it can be applied in many fields like in business and marketing by getting the real picture of ground situation. Furthermore, this can help decision-makers to assess the situations prevailing on the ground and in making informed decisions.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.


Notes

^aA user can tweet 2400 tweets in a day, and each tweet is limited to a length of 280 characters, so users try to be as original as possible, or provide relevant information.

^bGeocoding is the process by which one can convert addresses (like a street address) into geographic coordinates, which can be used to mark locations on a map [22].

^cA Kernel Density raster shows the density of points over a region.

ORCID iD

Girish Chandra Joshi  <https://orcid.org/0000-0002-0010-1835>

References

- [1] Statista. Social media: worldwide penetration rate 2019 statistic. *Statista*, <https://www.statista.com/statistics/269615/social-network-penetration-by-region/> (accessed 3 June 2019).
- [2] Hanna R, Rohm A and Crittenden VL. We're all connected: the power of the social media ecosystem. *Bus Horiz* 2011; 54: 265–273.
- [3] Cottle S. Media and the Arab uprisings of 2011: research notes. *Journalism* 2011; 12: 647–659.
- [4] Brooking ET and Singer PW. War goes viral. *The Atlantic*, November 2016, <https://www.theatlantic.com/magazine/archive/2016/11/war-goes-viral/501125/>
- [5] Blaker L. The Islamic state's use of online social media. *Mil Cyber Aff* 2015; 1: 4.
- [6] Jhaveri D, Chaudhari A and Kurup L. Twitter sentiment analysis on E-commerce websites in India. *Int J Comput Appl* 2015; 127: 14–18.
- [7] Pang B and Lee L. Opinion mining and sentiment analysis. *Found Trends® Inf Retr* 2008; 2: 1–135.
- [8] Ratkiewicz J, Conover M, Meiss M et al. *Truthy: mapping the spread of astroturf in microblog streams*. In: *Proceedings of the 20th international conference companion on World wide web*, 28 Mar 2011.
- [9] Caragea C, Squicciarini AC, Stehle S et al. Mapping moods: geo-mapped sentiment analysis during hurricane sandy. In: *Proceedings of the ISCRAM*, University Park, PA, 18–21 May 2014.
- [10] Hochman N and Manovich L. Zooming into an Instagram City: reading the local through social media. *First Monday*, 1 July 2013, <https://firstmonday.org/article/view/4711/3698>
- [11] Clarke K. Sentiment mapping. *Maps Mania*, <http://googlemapsmania.blogspot.com/2014/10/sentiment-mapping.html> (2014, accessed 3 June 2019).
- [12] Hriday SAA, Ekram MT, Islam MS et al. Localized twitter opinion mining using sentiment analysis. *Decis Anal* 2015; 2: 8.
- [13] Sampei Y and Aoyagi-Usui M. Mass-media coverage, its influence on public awareness of climate-change issues, and implications for Japan's national campaign to reduce greenhouse gas emissions. *Glob Environ Change* 2009; 19: 203–212.
- [14] Lama P. Gurmeet Ram Rahim Singh rape case: train set afire, 12 places hit across Delhi. *Hindustan Times*, 26 August, <http://www.hindustantimes.com/>, <http://www.hindustantimes.com/delhi-news/gurmeet-ram-rahim-singh-rape-case-train-set-afire-12-places-hit-across-delhi/story-n7KblTrfPRL415XIY4baP.html> (2017, accessed 28 November 2017).
- [15] Mufti I. Ram Rahim convicted: Chandigarh-Delhi airfare soars, trains cancelled till Aug 28. *Hindustan Times*, 27 August, <http://www.hindustantimes.com/>, <http://www.hindustantimes.com/india-news/ram-rahim-convicted-chandigarh-delhi-airfare-soars-tors-26k-trains-cancelled-till-august-28/story-Ei3mNiVPH7rfXXHydBbiAJ.html> (2017, accessed 28 November 2017).
- [16] R Core Team. *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing, 2017, <https://www.R-project.org/>.
- [17] Han J, Pei J and Kamber M. *Data mining: concepts and techniques*. New York: Elsevier, 2011.
- [18] Twitter Developers, <https://developer.twitter.com/en/apps> (accessed 4 June 2019).
- [19] Gentry J and Lang D. ROAuth: R Interface for OAuth, 2015, <https://CRAN.R-project.org/package=ROAuth>
- [20] Lang D CRAN Team. RCurl: general network (HTTP/FTP/...) client interface for R, 2016, <https://CRAN.R-project.org/package=RCurl>
- [21] Gentry J. twitteR: R based Twitter client, 2015, <https://CRAN.R-project.org/package=twitteR>
- [22] Get Started, Geocoding API, Google Developers, <https://developers.google.com/maps/documentation/geocoding/start> (accessed 4 June 2019).
- [23] Kahle D and Wickham H. ggmap: spatial visualization with ggplot2. *R J* 2013; 5: 144–161.
- [24] Turney PD. Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th annual meeting on association for computational linguistics*, Philadelphia, PA, 6 July 2002, pp. 417–424. Stroudsburg, PA: Association for Computational Linguistics.
- [25] Pang B, Lee L and Vaithyanathan S. Thumbs up? Sentiment classification using machine learning techniques. In: *Proceedings of the ACL-02 conference on empirical methods in natural language processing*, Philadelphia, PA, 6 July 2002, vol. 10, pp. 79–86. Stroudsburg, PA: Association for Computational Linguistics.
- [26] Liu B. *Sentiment analysis and opinion mining*. San Rafael, CA: Morgan Claypool Publishers, 2012.
- [27] Liu B, Hu M and Cheng J. Opinion observer: analyzing and comparing opinions on the web. In: *Proceedings of the 14th international conference on world wide web*, Chiba, Japan, 10–14 May 2005, pp. 342–351. New York: ACM.
- [28] Wickham H. The split-apply-combine strategy for data analysis. *J Stat Softw* 2011; 40: 1–29.
- [29] Wickham H. stringr: simple, consistent wrappers for common string operations, 2017, <https://CRAN.R-project.org/package=stringr>
- [30] Pang B and Lee L. A sentimental education: sentiment analysis using subjectivity summarization based on minimum cuts. In: *Proceedings of the 42nd annual meeting on association for computational linguistics*, Barcelona, 21–26 July 2004, pp. 271–278. Stroudsburg, PA: Association for Computational Linguistics.

-
- [31] Wilson T. Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on human language technology and empirical methods in natural language processing*, Vancouver, BC, Canada, 6–8 October 2005, pp. 347–354. Stroudsburg, PA: Association for Computational Linguistics.
- [32] Taboada M, Brooke J, Tofiloski M et al. Lexicon-based methods for sentiment analysis. *Comput Linguist* 2011; 37: 267–307.
- [33] Agarwal A, Xie B, Vovsha I et al. Sentiment analysis of Twitter data. In: *Proceedings of the workshop on languages in social media*, Portland, OR, 23 June 2011, pp. 30–38. Stroudsburg, PA: Association for Computational Linguistics.