Community Based Emotional Behaviour Using Ekman's Emotional Scale

Debadatta Naik, Naveen Babu Gorojanam, and Dharavath Ramesh(✉)

Indian Institute of Technology (ISM), Dhanbad 826004, India
deba.uce03@gmail.com, naveenbabugorojanam@gmail.com,
drramesh@iitism.ac.in

Abstract. In the current era, the analysis of social network data is one of the challenging tasks. Social networks are represented as a graphical structure where the users will be treated as nodes and the edges represent the social tie between the users. Research such as community identification, detection of centrality, detection of fraud, prediction of links and many other social issues are carried out in social network analysis. However, the community plays an important role in solving major issues in a real-world scenario. In a community structure, nodes inside a community are densely connected, whereas nodes between the communities are sparsely connected. Determining the emotional behavior from communities is the major concern because, emotional behavior of community helps us to solve problems like brand reachability, find target audience, build brand awareness and much more. Ekman's emotional scale is a popular categorical model which assumes that there is a finite number of basic and discrete emotions and is used to classify the emotions. In this paper, a novel method is proposed to determine the community-specific emotional behavior of the users related to a particular topic. Communities are formed based on network topology rather than emotions. Girvan Newman algorithm is used to construct the communities of different users, who share their views on twitter media platforms regarding a topic. Then Ekman's emotional scale is used to categorize the emotions of the users of each community. This identifies how the people of different communities react to an incident. The incident can be treated as NEWS/Trending threads on Twitter/Facebook shares. The emotional analysis is done community-specific so that the behavioral analysis of an incident is performed specifically for that community. Further, the comprehensive experimental analysis shows that the proposed methodology constructs influential communities and performs emotional analysis efficiently.

Keywords: Community detection · Emotional analysis · Ekman's emotion scale · Tweet emotions recognition

1 Introduction

The popularity of social media is increasing day by day. This leads to huge research interest and opportunities in the study of communities. The most used

S. S. Rautaray et al. (Eds.): I4CS 2020, CCIS 1139, pp. 63–82, 2020.
https://doi.org/10.1007/978-3-030-37484-6_4

techniques in the study are community detection and sentiment analysis [8]. Community detection finds a group of people who are closely associated with each other, while the sentiment analysis determines the emotional behavior of a person and specifies his perception on a particular topic i.e., states how an individual feels. In social-media-based micro communications, the content individuals produce, and the emotions therein can affect other people's emotional states. Analysis of the social network data for community detection and their behavior has become the fundamental aspect of social network analysis [23]. This gives us an idea about how the users belonging to a particular community interact with each other and how various communities are interacting among themselves too. The analysis of communities is a topic of high research interest due to its wide range of applications. One such important application is to find the communities of interest, then focus only on those communities for marketing [13]. Thus, analyzing, and detecting human emotions based on the behavior exhibited by the communities has a direct effect on social networks [16,21]. In social networks, people are connected to each other via; social media. One person can easily detect the emotions of the persons who are involved in the group chatting on a particular topic. Safety is one of the main concerns in the present scenario. Based on the emotional status of the community, it can be possible to predict and trap the criminal activities of a particular community. In commercial applications, the organization tries to find out the emotions of people of different communities based on the launched products. For example, when launching a new product into the market, find the communities which are interested in the product. Through the advertisement, companies can focus on the communities which have shown high interest in the product. Apart from this, the organization also tries to find out the problems associated with their product from the reviews of the dissatisfied communities. This gives more exposure to the organization to make future marketing strategies. To solve this, in this work, two real-time issues; one from a commercial application and another from a social application are demonstrated with a different set of data sets.

In the current scenario, WhatsApp, Facebook, Instagram, LinkedIn, Twitter, and other social network sites are the main source for the generation of data. The structure of twitter works as "follow" and "following" relationship between the users. The user following can receive the notifications of the public posts of the other user updates [11]. The present research work focuses primarily on detecting people's positive or negative feelings. However, these approaches fail to detect the other dimensions of emotions like anger, disgust, fear, happiness, sadness, and surprise [15,24]. Many researchers have proposed methodologies for the extraction of emotions from an individual or considering all the users as a single community. However, the existing methodologies are ineffective, time-consuming and expensive to use in the applications like; study the behavior of people on a product, study the behavior of communities on social issues and many more. It may lead to a huge loss as compared to the profit for the investors. In addition, organizations are also unable to study the emotional behavior of individuals from different communities. It creates a gap for the researchers to conduct analysis on the communities rather than each individual or on a single group. Moreover, the quantifying approach for extracting the emotions of communities is a fast and cost-effective analysis. The aforementioned limitations motivate us to present a novel methodology that is capable enough to capture the emotions of the communities. So, every community's emotional level is monitored. The emotional level of the user is modeled using Ekman's Emotional Scale [10], which classifies human emotions into six in number; anger, disgust, fear, happiness, sadness, and surprise. Later, we calculate how a user's tweet is influencing the emotional behavior of the others following. The main contributions of the proposed work are summarized as follows:

- We propose an algorithm to detect the emotional behavior of communities, giving emphasis to the topological structure of the user's network.
- The more accurate scale of human emotions is determined using Ekman's emotional scale.
- We also demonstrate the impact of influential person's tweets on other people of a community by using two real-life data sets.

The rest of the paper is structured as follows. Section 2 presents the background topics in community detection and sentiment analysis. Section 3 presents the proposed methodology for sentiment analysis as a community wise. Section 4 contains implementation with a sample data set of the tweets, where emotional behavior is obtained for various communities. Section 5 depicts the experimental results and discussion for two different topics. Threats to validity are mentioned in Sect. 6. Conclusion and future work are presented in Sect. 7.

2 Related Work

Social network analysis has been one of the trending topics due to the increasing number of user's day-by-day. Community detection is a method through which it can be achieved that the users who are closely connected to each other are made into communities. In other words, the users inside a community are strongly correlated whereas users between the communities are loosely correlated [20]. One of the primitive algorithms for finding communities was proposed by Girvan and Newman, which forms the communities based on edge betweenness. In this work, edges are removed between the communities until each community gets isolated [17]. There are other community detection algorithms that work based on modularity, where modularity is the density of edges inside the community against the density of edges outside the community [4].

Sentiment analysis is another important topic that is used to recognize the emotion of a person. This analysis has a high significance in social network analysis because most of the people share their opinions on social networks [18]. The opinions of people are classified by Ekman's emotional scale, which classifies the emotions of humans into six categories: Anger, Sadness, Happiness, Disgust, Surprise, and Fear. This is one of the dominant models for classifying emotions based on the sentences [5]. An overview of sentiment analysis approaches has been described in [2]. One of the primitive methods for sentiment analysis on twitter data was carried out by Go et al., where the sentiments are obtained using machine learning algorithms like Naive Bayes, SVM, and maximum entropy [12]. In this work, the method involves prepossessing and post-processing of tweets. In post-processing, the emotions are extracted from the tweets and any tweet with both negative and positive emotion is removed. The retweets are also removed. As a result, the twitter data which has been used for emotional detection are reduced. The main drawback of this method is that it states only two emotional scales. A significant method was proposed by Barbosa and Feng, where the sentiment analysis is carried out through subjectivity classification and polarity detection [3]. Subjectivity classification is done by removing the disagreement approval, the tweets by the same user, and the top opinion words from the twitter data set. The polarity detection classifies the emotion of the tweet into positive or negative by dividing each category of tweets with the total number of tweets. The main drawback of this approach is that filtering of tweets is performed by choosing the specific tweets which satisfy the subjective condition. As a result, the number of original tweet data is significantly reduced and analysis is performed on less number of tweets. Agarwal et al. proposed the methodology, where the twitter data are classified into positive, negative and neutral [1]. The classification in this model is performed with respect to Tree kernel, unigram, and Feature-based models. Parts of speech tagger has been used to classify the emotions in this model. The Tree kernel model and feature-based model outperformed the unigram model, but the emotions are still restricted to three in number i.e., positive, negative and neutral. Xu et al. [24] proposed a method, where sentiment analysis and community detection are combined together to determine sentimental communities. In this work, the emotions are classified based on the opinion of the user towards a product, service, candidate or a topic. This analysis was used for market research to identify the sentiment analysis of the communities. The main drawback of this approach is that the communities are formed either to portray positive emotion or negative emotion. Due to this, the actual behavior of the users could not be determined because this method classifies happiness or surprise both as positive and sadness or disgust both as negative. Further, the classification of emotions was performed by Kanavos et al. [15], where the users are formed into communities based on the post-impact and influence metric. The post-impact of a tweet determines the impact of the tweet posted by the user and the influence metric of the user is calculated based on the influence metric and impact of the post. Emotions are recognized using Ekman's emotional scale and mapped into six different emotional scales. In this work, influential communities are produced based on the emotional analysis. However, in this method, users from different communities may be grouped in a single community because communities are formed based on emotions and influence rather than the topological structure of the users' network. This is because users are formed into groups based on influence metric rather than the connectivity. Dragoni et al. proposed a new approach where emotions are captured based on OntoSenticNet [9]. OntoSenticNet is a commonsense ontology that works based on a SenticNet. The SenticNet is a semantic network consisting of 10,000 primitive concepts. This work states that many lexicon-based knowledge bases are available for sentiment analysis, but there are no sentiment-based ontologies. The same reason is the limitation of this work. In a fast emerging environment, the number of lexicons is increasing rapidly, whereas the primitive concepts used in this model are limited to 10,000 in number. The approach proposed in [7] by Cambria et al. is similar to the approach used in [9]. However, a slight modification was done. While using the same primitive concepts, this method employs context embedding. Context embedding is the process of automatically discovering new conceptual primitives from text and embedding them to the common sense ontologies. This is an improvement in sentiment detection, but this method could not overcome the symbol grounding problem in sentiment analysis. Due to this limitation, this work could not perform sentiment analysis on language independent and symbol independent text data. Also, this approach uses AI for concept prediction. But, if the same concept library is used for any new text data analysis, it might not be fruitful because the new text data might not produce same concept ontology. Kanavos et al. [14] proposed a similar kind of methodology in which author collected and analyzed the tweet's emotions and based on these tweets, influential communities are derived. Each tweet is evaluated using ekman's scale. Then weighted version of communities are formed based on modularity optimization of the network and emotional state of the retrieved tweets. In this work, we propose a method that overcomes the ambiguity of various types of users coming under the same community. The most influential person can alter the emotions of the other users of the same community, i.e., shift emotions of people from the negative side to the positive side or vice versa. This methodology can be of great use in market research because it allows investors to monitor how emotions are being altered due to the tweet by the most influential person of the community.

3 Proposed Method

A social network can be interpreted as a graph structure $G = (V, E)$, where V and E represents the set of users and links between the users respectively. In this work, follow to follow factor analysis is carried out on the user's profile to identify the connections between the users. The user profiles are then mapped into the network of nodes, where they are connected to each other through edges. Then communities are formed in the network by using the Grivan Newman algorithm which uses edge betweenness to remove the edges iteratively. After this, for each of the communities formed, we calculate the emotional levels of each user in each community based on Ekman's emotional scale. Ekman's emotional scale gives the highest dimensions of emotions. This gives us an idea about how the users are reacting to a topic which is trending on Twitter. After this, we find the most influential person from each community using the two parameters; impact factor and influence metric [14]. These two parameters are calculated for each user by using (2) and (3). A user with the highest influence metric value will be treated as the most influential person of the community from which he or she belongs. Tweet of the influential person has a great impact on the emotions of the other users in the community. It mainly helps to alter the emotions of the people of the community. As a result, the emotional behavior of the whole community will be changed. This provides how influential is the community with respect to the tweets by the most influential person of that community. The details of our approach are presented in Algorithm 1. Algorithm 1 is mainly divided into three major subsections: Community formation, Emotional analysis, and Influential impact. These subsections are illustrated as follows.

Algorithm 1. Emotional Analysis based on Community Detection using Ekman's Scale

Input: Tweets from the users on a particular topic. This twitter data set is mapped into the graph structure $G = (V, E)$ where V is the set of users and E is the set of connections between the users.

Output: Emotions of each community.

1. Network formation based on follow to follow factor.
2. Community formation based on Girvan Newman algorithm.
3. Analysis of emotions level in communities using Ekman's emotional scale.
4. Identification of the most influential person in each community.
5. Repeat the Step 3 to perform the emotional analysis with new tweet data

3.1 Community Formation

In this subsection, the users are grouped into communities based on how strong the connection exists between them. Since, the Girvan Newman algorithm is the simple and standard algorithm used for detecting the non overlapping communities based on the edge betweenness, we have used it for non overlapping community formation [17]. In a graph, edge betweenness centrality represents the number of shortest paths that go through an edge in a graph or network. Each edge in the graph is associated with a betweenness centrality value. An edge with high centrality value represents a bridge between the parts of the graph [6]. Mathematically, betweenness centrality of a node u is defined as:

$$BC(u) = \sum_{u \neq x \neq y} \frac{B_{xuy}}{B_{xy}} \tag{1}$$

where, B_{xuy} is the number of geodesic paths between the nodes x and y that passes through the node u and B_{xy} is the total number of geodesic paths between the node x and y. Normalized betweenness centrality is obtained by dividing the Eq. 1 into $(n-1)(n-2)/2$. Since most of the connections in the twitter are follow to follow type, we use undirected and unweighted edges between the nodes when graphs are constructed. The working principle of the Girvan Newman algorithm for community detection is represented in Algorithm 2. In Algorithm 2, communities are detected by eliminating the edge with high betweenness centrality value iteratively. In a network, edges that are existing between the communities are found to be of high betweenness centrality value. After the removal of an edge having the highest centrality value, betweenness centrality for all the remaining edges is recalculated. The end result of Algorithm 2 shows the dendrogram, representing the different communities.

Algorithm 2. Girvan Newman Algorithm

Input: The network $G = (V, E)$ where $n = |V|$ and $e = |E|$.
Output: Community Structure
1 Edge betweenness is calculated for all the edges present in the network.
2 Edge with the highest betweenness value is removed first.
3 Edge betweenness of affected edges is recalculated after the removal of the edge with the highest value.
4 Step 2 and Step 3 are repeated until no edges remain in the network.

3.2 Emotional Analysis

In this subsection, detailed procedures for the extraction of emotions from the tweets are discussed. Emotions are generated from the tweets of the user using the tool called Emotion recognition tool [19]. The workflow diagram of the Emotion recognition tool is presented in Fig. 1. The tweet is initially divided into

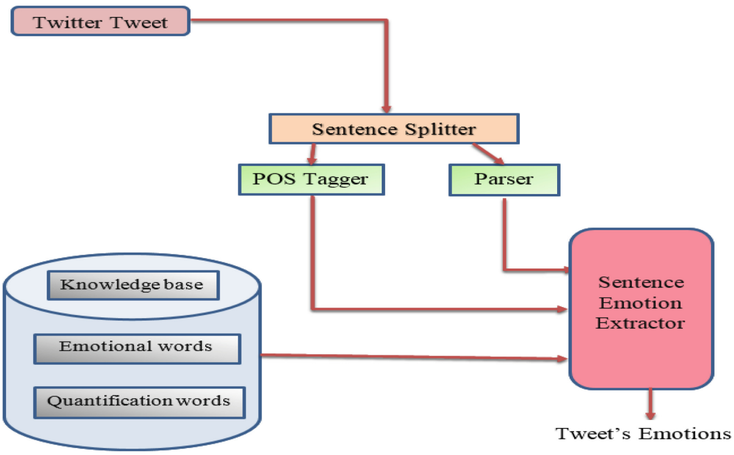

Fig. 1. Emotion recognition tool

words and the words are analyzed using parts of speech tagger and a tree structure is created. After this, stanford parser analyses the words and their relationship between them to create the dependency tree. The tree constructed has triplets namely, the root describing the emotion, the left node representing the governor and the right node representing the dependent. After this step, the triplets are compared with the knowledge base of the tool for further analysis. The knowledge base is a collection of words that portray emotions based on the lexical structure of the words used in the sentences. These emotional words are spotted on WordNet affect source [22]. After this, the sentence emotion extractor outputs the emotional status of the user. The emotions are anger, disgust, fear, happiness, sadness, and surprise according to Ekman's emotional scale.

3.3 Influential Impact

In social media, people's emotional behavior and thoughts are influenced by the people they follow. Often, the users with the highest number of followers term themselves as social media influencers. Anything posted by these people will have an impact on the followers. *Example: if a user with 10,000 followers tweets saying the new X product is totally worth buying. If at least 6000 followers are active and check out the tweet, there are higher chances that at least 30–40% of them might actually buy the product.* So, there is a need to find an influential factor of each user in the community to determine the dominant user. A twitter post has features like a retweet, replies, clicks, favorites, and mentions. Based on these factors, we obtain the influence factor as described in [15] and is shown in (2).

$$post\_impact = ((ret+1)\times(rep+1)\times(fav+1)\times(men+1)\times(cli+1))/directtweets \tag{2}$$

where, ret: Retweets, rep: Replies, fav: Favorites, men: Mentions, cli: Clicks. From (2), we can get how much impact is being generated by the post. To avoid the value becoming zero, a constant value of 1 is added to each feature of the tweet. Similarly, (3) depicts the influencing metric of the users in a community, which mainly depends on the follow to follow count, frequency of posting by the user and the post_impact.

$$Inf\_metric = \log(FtF+1) \times freq \times post\_impact \tag{3}$$

4 Implementation

Experiments are performed on a workstation which is configured with 3.4 GHz clock speed. It has 2 TB hard disk and 8 GB of RAM. A sample data set of 26 users from A to Z is considered and is presented in Table 1. Tweets have been gathered from these users based on a topic called new mobile launch, which is a different topic from case 1. Users have mentioned their opinions about the product in their tweet. Then communities are formed for these users by implementing the Girvan Newman algorithm. In Fig. 2, it is shown that four communities are formed based on the connectivity between the users and these are; Employees, Freelancers, Students, and Critics. Based on the work profile, names are assigned to the communities for better convenience purpose. The connections between the users are taken in the form of GML file. Geography Markup Language is the XML grammar defined by the Open Geospatial Consortium (OGC) to express geographical features. The community structure is shown in Fig. 2.

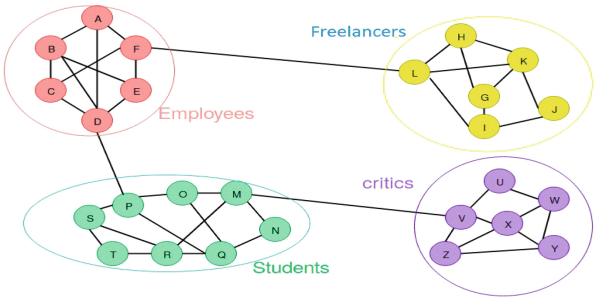

Fig. 2. Community structures

First, users are grouped into their respective communities using Algorithm 2. Each of the above tweets is then analyzed with the emotion recognition tool and the emotion of each tweet is extracted. Emotion levels of these communities are presented in Table 2. It is observed that the opinions of the critics community are mostly positive because they are expressing happiness and surprise mostly. Students are showing the disgust and sadness emotions towards the product. So the students' community should be focused upon in order to improve the sales of the mobile and get a positive impact from the students' community. To achieve this goal, a tweet from the most influential user of a student community must convey happiness emotion. Then the followers will get influenced and they may change their emotion in the retweet. This strategy makes the emotional scale to be shifted towards the positive side. In the students' community, the students R and Q have the highest number of followers, i.e., a good influence in the students' network. If anyone of these users starts tweeting positively about the product, the opinions of the followers will change and the community will head towards the positive side of the emotional scale, i.e., they start expressing surprise or happiness and not disgust, anger or fear.

5 Experimental Results

In this work, we have considered two different cases to validate the proposed methodology. These cases are the *launch of new iphone11* and the *mob lynching* in India. The analysis is done by collecting the tweet data from real users from twitter.

Table 1. Tweets of the users on a mobile launch.

User name	Tweet	Emotions
A	This better have wide lens camera #nophone	Fear
B	Still no bluetooth :-(#nophone #missingbluetooth	Sadness
C	It's gonna be wonderful #nophone	Happiness
D	Can't wait #oneweektogo #nophone	Happiness
E	Wow... the wait is over #nophone	Surprise
F	The camera looks bad #nophone	Sadness
G	No office subscription :-(#nophone #dissapointed	Sadness
H	Except camera, nothing to boast about #nophone	Disgust
I	Same processor #nophone	Angry
J	#nophone ready to be the new flagship killer	Happiness
K	Launching so soon #takenbysurprise #nophone	Surprise
L	One week to go #nophone	Happiness
M	Overpriced :'(#nophone	Sadness
N	#nophone please concentrate on millenials	Fear
O	Uhhh. . . no headphone jack #nophone #nojack	Surprise
P	Please cut down the cost #overpriced #nojack	Fear
Q	Too costly for a phone with no headphone jack #nophone #nojack	Disgust
R	No fast charger included in the box :'(#notfair #nophone	Sadness
S	Is the company accepting kidney as payment #toocostly #nophone	Sadness
T	Same as old #nophone but price +300$	Disgust
U	It should come with amoled display #nophone	Fear
V	Still the old snapdragon chipset #plsupgrade #nophone	Sadness
W	The specs amaze me #surprised #nophone	Surprise
X	Why is it priced twice the previous variant #nophone	Angry
Y	It's surprising that #nophone kept its word regarding camera	Surprise
Z	The best camera about to hit the market #nophone	Happiness

Table 2. Emotional behavior of communities.

Community	Anger	Disgust	Fear	Happiness	Sadness	Surprise
Employees	0%	0%	16.67%	33.34%	33.34%	16.67%
Students	0%	25%	25%	0%	37.50%	12.50%
Freelancers	16.67%	16.67%	0%	33.34%	16.67%	16.67%
Critics	16.67%	0%	16.67%	16.67%	16.67%	33.34%

5.1 Dataset Description

We have taken the data from twitter. These are the real data tweeted by the different users based on two topics. Details of these datasets are given in Table 3. The Girvan Newman algorithm is used for the community formation on the users who have posted the tweet. The users who are involved in posting the tweets related to the topic; Release of the new iPhone 11 are grouped into four different communities and their names are assigned based on the highest number of user's work profile found in these formed communities. These are Youtubers, General Public, News Channels, and Tech Pages. Assignment of names is optional, but for better understanding purpose names are assigned in this work. Similarly, communities of users found in the Mob lynching topic are; Youth, Political parties, News Channels, and Social pages. The emotional analysis is done for these communities using the Emotion recognition tool and Ekman's emotional scale. We then find the influential person in each community of the network to know, how each of these communities is influenced by the tweets of the influential person.

Table 3. Dataset description.

Sl. no.	Tweet topic	No. of tweets analyzed	No. of tweets analyzed after detecting influential node	Source
1	Release of new iPhone 11	193	193	Twitter (Open Source)
2	Mob lynching in India	170	170	Twitter (Open Source)

5.2 Results and Discussions

***Release of the new iPhone 11*:** Initially, emotional levels of 193 users are computed from their tweets, considering all the users as a single community. The analysis categorized the emotions of users into six basic emotions which are shown in Fig. 3. A maximum number of people have shown the surprise emotion

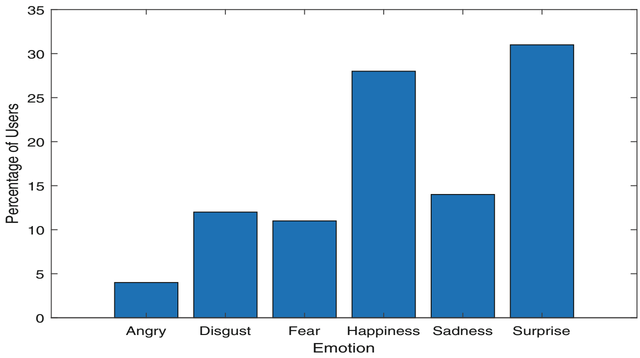

Fig. 3. Emotional levels for *new iPhone 11* launch

Table 4. Communities with emotional levels for *new iPhone 11* launch.

Communities	Angry	Disgust	Fear	Happiness	Sadness	Surprise
Youtubers	4.34%	10.86%	4.34%	32.6%	17.39%	30.43%
General public	7.69%	10.25%	12.82%	33.33%	12.82%	23.07%
News channels	7.89%	13.15%	15.78%	15.78%	15.78%	31.57%
Tech pages	6.06%	15.15%	12.12%	18.18%	18.18%	30.30%

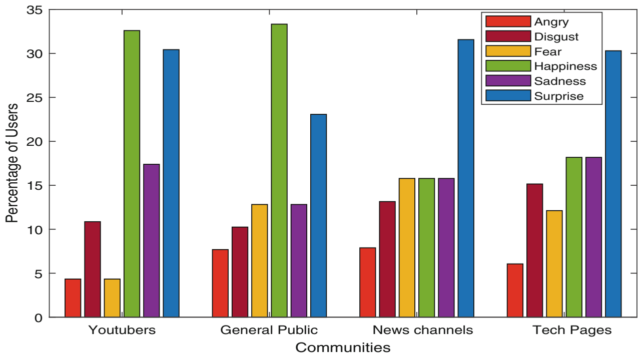

Fig. 4. Emotional levels for *new iPhone 11* launch (Community Specific)

on this topic and it is 31% of the total. The least percentage of people have shown their anger emotion. This analysis shows the emotional levels of all the users without the community-specific and gives us an idea about how the users are reacting to the iPhone 11 launch. However, from this analysis, it is not sure that which category of the people are reacting and how their emotional behavior is being shown. So, no substantial decision can be made based on this analysis. To overcome this problem, the Girvan Newman algorithm is used to categorize the users into different communities. Using this algorithm, four communities named Youtubers, General public, News channels, and Tech pages are formed from the user network. The emotional analysis of these communities gives a clear idea about the emotional behavior of the users. Moreover, the analysis also helps to do good market research and gives an idea of where to focus more. Table 4, shows the general public and News Channels community describes the highest and lowest value of happiness respectively. However, News channel community illustrates the highest anger and Tech pages community shows the highest sadness as compared to other communities. Therefore, investors must pay attention to these communities in marketing strategies. The graphical representation of Table 4 is depicted in Fig. 4. An influential person plays a vital role in changing the emotions of the users of these two communities. A positive tweet (an emotion conveying happiness) of an influential person may change the emotion of users of the same community towards the product. This is illustrated clearly by analyzing a positive tweet from an influential person named Marques Brownlee from You Tuber community. He tweeted about iPhone 11 as: ***"NEW VIDEO: The iPhone 11 Models! Youtu.be/rie69pow668 – RT!"***. This tweet has got 731 retweets, 223 replies, 10135 favorites, 46 mentions, 1204 clicks, 193 direct tweets, 3319628 F2F, and 3.4 frequency. Then according to (2), the post_impact factor can be calculated as:

$post\_impact = 4779.94$ (Value is divided by 10^8 to normalize the value)

And the influence metric is calculated as:

$Inf\_metric = 105961.873$

Using (3), the most influential person can be detected for a community and based on his tweet, followers may change their mindset related to a particular topic. The same process is applied to other communities to determine the most influential person. Tweets are again collected from all the users, after the posting of a tweet by the most influential person of each community. Then the emotional levels are re-evaluated for each user. The significant change in the emotional levels of all the communities after the tweet by an influential person of each community are shown in Table 5. As shown in Table 5, in the Youtubers community, disgust feeling decreased by 8%, sadness decreased by 13%, anger decreased to 0%. However, the positive scales of emotions are increased; like happiness increased by 14% and surprise increased by 10%. In the same way, the emotional levels vary in different communities after a tweet by the most influential person from each community. The graphical representation of the modified emotional levels is depicted in Fig. 5.

Table 5. Emotional levels after the tweet by an influential person of each community

Communities	Angry	Disgust	Fear	Happiness	Sadness	Surprise
Youtubers	0%	2.2%	6.6%	46.6%	4.4%	40%
General public	4.87%	19.51%	14.63%	14.63%	9.75%	36.58%
News channels	0%	10.52%	15.78%	10.52%	31.57%	31.57%
Tech pages	5.88%	11.76%	17.64%	29.41%	11.76%	23.52%

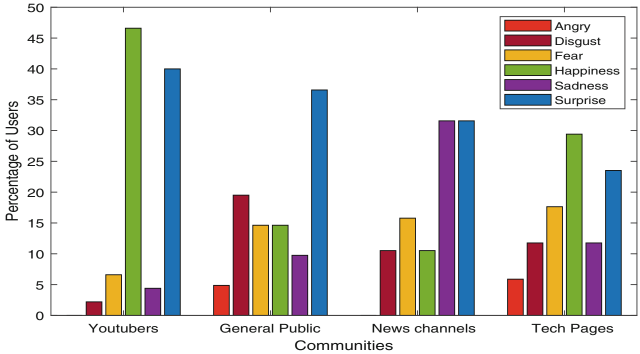

Fig. 5. Emotional levels after the tweet by an influential person of each community

The cumulative analysis of emotional levels is done by considering twitter data of all the users before and after the tweet of the most influential person. The cumulative data presented in Table 6, shows how the emotional levels are actually being portrayed by each community at the present course of time. The graphical representation of these data is depicted in Fig. 6. The analysis shows the significant improvement or deterioration of the emotional levels.

Table 6. Emotional levels for *new iPhone 11* launch (Cumulative)

Communities	Angry	Disgust	Fear	Happiness	Sadness	Surprise
You Tubers	1.48%	8.14%	5.18%	37.77%	13.33%	34.07%
General public	6.72%	13.44%	13.44%	26.89%	11.76%	27.73%
News channels	5.26%	12.28%	15.78%	14.03%	21.05%	31.57%
Tech pages	6%	14%	14%	22%	16%	28%

Mob Lynching: The same procedure is applied to the topic of Mob lynching in India to analyze the emotions of people. The related tweet data are collected to perform the analysis. The normal emotional analysis of the people with respect to the mob lynching is performed using Ekman's emotional scale and is shown in

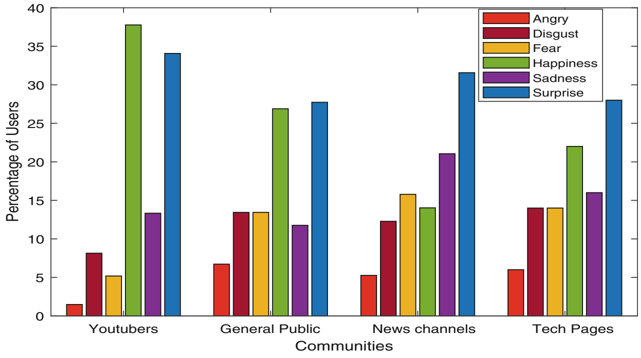

Fig. 6. Emotional levels for *new iPhone 11* launch (Cumulative)

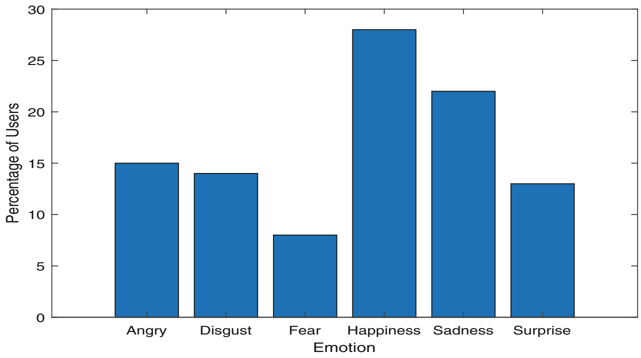

Fig. 7. Emotional levels for *Mob Lynching*

Fig. 7. The main disadvantage of this analysis is that researchers are not able to identify which community of people are reacting and portraying which emotion with respect to the topic. Further, no significant decisions can be made based on this analysis. To overcome these problems, analysis is carried out according to the proposed methodologies. The results shown in Table 7 are illustrating the emotional levels of all the communities.

Using Algorithm 2, four communities have been discovered. These are Youth, Political parties, News channels, and Social pages. Youth and political party's communities show the highest value of happiness emotion, which is never desirable by society. However, the political party's community shows the highest value of sadness emotion. The analysis also states that most of the communities are portraying sadness emotion high as compared to the other emotions. Since this is a controversial topic, most of the user's tweets are displaying such emotions. The graphical representation is depicted in Fig. 8, where emotion levels of all the communities are shown.

Table 7. Communities with emotional levels for *Mob Lynching*

Communities	Angry	Disgust	Fear	Happiness	Sadness	Surprise
Youth	24%	16%	4%	24%	20%	12%
Political parties	12%	4%	4%	24%	48%	8%
News channels	23.52%	17.64%	5.88%	11.76%	17.64%	23.52%
Social pages	10.71%	7.14%	21.42%	14.28%	25%	21.42%

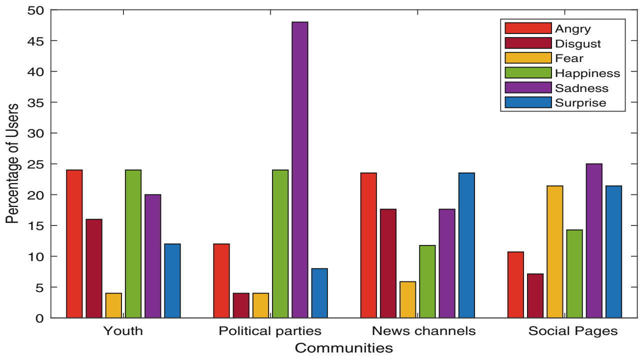

Fig. 8. Emotional levels for *Mob Lynching* (Community Specific)

It is observed that a user named Mr. Ashok Gehlot from political parties posted a tweet that gathered a lot of attention from other users of the community. The tweet is like: ***"The #Rajasthan Protection from Lynching Bill, 2019 was passed by the State Assembly today. The Bill has been brought in to provide for strict punishment to curb such incidents"***.

The above tweet has got 411 retweets, 61 replies, 2088 favorites, 29 mentions, clicks = 2051, 170 direct tweets, 73007 F2F, and 4.16 frequency. Then post_impact factor is calculated as:

$post\_impact$ = 183.154 (Value is divided by 10^8 to normalize the value)

And the influence metric is calculated as:

$Inf\_metric = 4487.717$

The same process is applied to other communities to determine the most influential person. Tweets are again collected from all the users, after the posting of a tweet by the most influential person of each community. Then the emotional levels are re-evaluated for each user. The significant change in the emotional levels of all the communities after the tweet by influential persons are shown in Table 8. As shown in Table 8, the happiness scale has significantly increased from 24% to 54.54% and the anger scale has reduced from 12% to 6.06% in the community of political parties. The emotional scales also varied in other communities. The graphical representation is depicted in Fig. 9, where emotion levels of all the communities are shown.

Table 8. Emotional levels after a tweet by the influential person of a community

Communities	Angry	Disgust	Fear	Happiness	Sadness	Surprise
Youth	17.64%	14.7%	14.7%	29.41%	17.64%	5.88%
Political parties	6.06%	9.09%	0%	54.54%	6.06%	24.24%
News channels	5.88%	29.41%	5.88%	41.17%	11.76%	5.88%
Social pages	13.04%	21.73%	8.69%	13.04%	34.78%	8.69%

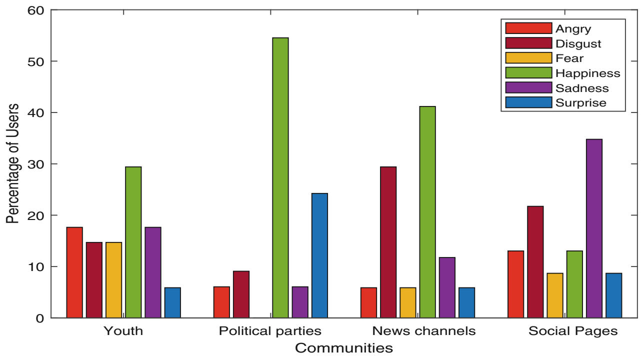

Fig. 9. Emotional levels after a tweet by the influential person of a community

The cumulative analysis of emotions illustrates the overall emotion levels portrayed by each community by considering all the tweets that have been posted by the users, The result of the cumulative analysis is shown in Table 9 and its graphical representation is depicted in Fig. 10. The cumulative analysis is useful for monitoring the emotional levels of each individual community with respect to a topic, whereas the breakdown of tweets before a tweet by the most influential person and after a tweet by the most influential person gives us how the communities are influenced. This can be used as a market research topic by studying the emotional behavior changes in communities in the presence of an influential factor.

Table 9. Emotional levels of each community for *Mob lynching* (Cumulative)

Communities	Angry	Disgust	Fear	Happiness	Sadness	Surprise
Youth	21.42%	15.47%	8.33%	26.19%	19.04%	9.52%
Political parties	8.62%	6.89%	1.72%	41.37%	21.13%	17.24%
News channels	14.7%	23.52%	5.88%	26.47%	14.7%	14.7%
Social pages	11.76%	13.72%	15.68%	13.72%	29.41%	15.68%

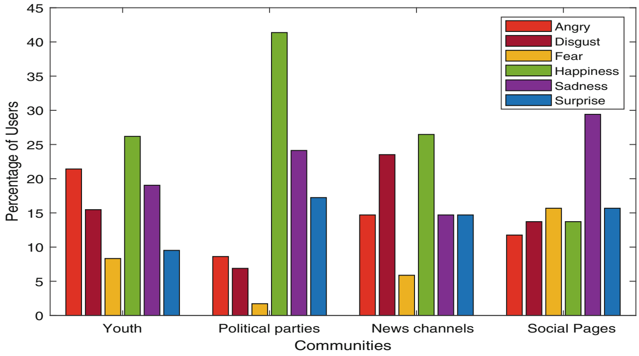

Fig. 10. Emotional levels of each community for *Mob lynching* (Cumulative)

6 Threats to Validity

In this work, a novel approach is presented to analyze the emotion levels of each user in community-specific. A well-known community detection algorithm called Girvan Newman is implemented to discover the communities of users. The method proposed can show how the influence factor can alter the emotions of the communities in the network. However, there are some limitations associated with this approach. These are summarized as follows:

- The Twitter extraction tool considers only the tweets which are specifying the topic name with #tag.
- Communities with small size are ignored in the analysis because Girvan Newman algorithm unable to detect the small size communities.
- The proposed approach unable to identify the actual emotional scale if sarcasm is being portrayed in the tweet.

7 Conclusions and Future Work

In this paper, we have proposed a novel methodology which analyses the community wise emotional behavior of the users in twitter dataset. Initially, the Girvan Newman algorithm is used to detect the communities and then emotions are determined using Ekman's emotional scale. This shows how various communities are reacting to a trending topic on Twitter or any other social media platform. The reactions of the users are analyzed based on their emotional behavior. This analysis is really helpful for market research as it determines how opinions of people are divided for a particular event. To improve the opinion on the product or event, the influential users are identified from each community to post the positive tweet. By doing so, more number of users can be inclined towards the positive emotional scale. The proposed approach overcomes the limitations that are associated with the existing methods. In earlier methods, mainly two emotional scales (positive and negative) were detected and emotions were user-centric rather than community-centric.

As future work, scalability being the major issue needs to be addressed using algorithms which can form communities faster for a larger given data set. Also, a shared-memory approach can be used for the parallel community detection process. Otherwise, techniques like Canopy clustering can be used which gives an estimate of clusters for community detection. In conclusion, we will use faster parallel techniques for community detection in time i.e., Modularity based approaches.

Acknowledgment. This research work is supported by Indian Institute of Technology (ISM), Dhanbad, Govt. of India. The authors wish to express their gratitude and heartiest thanks to the Department of Computer Science & Engineering, Indian Institute of Technology (ISM), Dhanbad, India for providing their research support.

References

1. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., Passonneau, R.: Sentiment analysis of twitter data. In: Proceedings of the Workshop on Language in Social Media (LSM 2011), pp. 30–38 (2011)
2. Aggarwal, C.C., Zhai, C.: Mining Text Data. Springer, Heidelberg (2012). https://doi.org/10.1007/978-1-4614-3223-4
3. Barbosa, L., Feng, J.: Robust sentiment detection on Twitter from biased and noisy data. In: Proceedings of the 23rd International Conference on Computational Linguistics: Posters, pp. 36–44. Association for Computational Linguistics (2010)
4. Blondel, V.D., Guillaume, J.L., Lambiotte, R., Lefebvre, E.: Fast unfolding of communities in large networks. J. Stat. Mech: Theory Exp. **2008**(10), P10008 (2008)
5. Botvinick, M., Jha, A.P., Bylsma, L.M., Fabian, S.A., Solomon, P.E., Prkachin, K.M.: Viewing facial expressions of pain engages cortical areas involved in the direct experience of pain. Neuroimage **25**(1), 312–319 (2005)
6. Brandes, U.: A faster algorithm for betweenness centrality. J. Math. Soc. **25**(2), 163–177 (2001)
7. Cambria, E., Poria, S., Hazarika, D., Kwok, K.: Senticnet 5: discovering conceptual primitives for sentiment analysis by means of context embeddings. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
8. Deitrick, W., Hu, W.: Mutually enhancing community detection and sentiment analysis on Twitter networks. J. Data Anal. Inf. Proc. **1**(03), 19 (2013)
9. Dragoni, M., Poria, S., Cambria, E.: Ontosenticnet: a commonsense ontology for sentiment analysis. IEEE Intell. Syst. **33**(3), 77–85 (2018)
10. Ekman, P.: Basic emotions. Handb. Cogn. Emot. **98**(45–60), 16 (1999)
11. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proc. Natl. Acad. Sci. **99**(12), 7821–7826 (2002)
12. Go, A., Bhayani, R., Huang, L.: Twitter sentiment classification using distant supervision. CS224N Proj. Rep. Stanf. **1**(12), 2009 (2009)
13. Han, M., Yan, M., Cai, Z., Li, Y., Cai, X., Yu, J.: Influence maximization by probing partial communities in dynamic online social networks. Trans. Emerg. Telecommun.Technol. **28**(4), e3054 (2017)

14. Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A.: Emotional community detection in social networks. Comput. Electr. Eng. **65**, 449–460 (2018)
15. Kanavos, A., Perikos, I., Hatzilygeroudis, I., Tsakalidis, A.K.: Integrating user's emotional behavior for community detection in social networks. In: WEBIST (1), pp. 355–362 (2016)
16. Kreibig, S.D.: Autonomic nervous system activity in emotion: a review. Biol. Psychol. **84**(3), 394–421 (2010)
17. Newman, M.E.: Fast algorithm for detecting community structure in networks. Phys. Rev. E **69**(6), 066133 (2004)
18. Pang, B., Lee, L., et al.: Opinion mining and sentiment analysis. Found. Trends Inf. Retr. **2**(1–2), 1–135 (2008)
19. Perikos, I., Hatzilygeroudis, I.: Recognizing emotion presence in natural language sentences. In: Iliadis, L., Papadopoulos, H., Jayne, C. (eds.) EANN 2013. CCIS, vol. 384, pp. 30–39. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-41016-1_4
20. Pizzuti, C.: GA-Net: a genetic algorithm for community detection in social networks. In: Rudolph, G., Jansen, T., Beume, N., Lucas, S., Poloni, C. (eds.) PPSN 2008. LNCS, vol. 5199, pp. 1081–1090. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87700-4_107
21. Silberstein, R.B.: Electroencephalographic attention monitor. US Patent 4,955,388, 11 Sep 1990
22. Strapparava, C., Valitutti, A., et al.: Wordnet affect: an affective extension of wordnet. In: Lrec, vol. 4, p. 40. Citeseer (2004)
23. Thelwall, M., Wilkinson, D., Uppal, S.: Data mining emotion in social network communication: gender differences in myspace. J. Am. Soc. Inf. Sci. Technol. **61**(1), 190–199 (2010)
24. Xu, K., Li, J., Liao, S.S.: Sentiment community detection in social networks. In: 6th Annual Conference on 2011 iConference: Inspiration, Integrity, and Intrepidity, iConference 2011, pp. 804–805 (2011)