TWITTER HATE COMMENT ANALYSIS

Project report in partial fulfillment of the requirement for the award of the degree of
Bachelor of Technology
In
Computer Science and Engineering

Submitted By

Name – Apratim Ghosh

Enrollment Number - 12020009001228

Section - A

Roll Number – 15

Under the guidance of

Prof.Amartya Chakraborty

&

Prof. Sumit Anand

Department of Computer Science and Engineering

UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA

University Area, Plot No. III – B/5, New Town, Action Area – III, Kolkata – 700160

CERTIFICATE

This is to certify that the project titled **Twitter Hate Comment Analysis** submitted by **Shaswata Karan (12020009001183)** student of UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA, in partial fulfillment of the requirement for the degree of Bachelor of Computer Science & Engineering, is a bonafide work carried out by him under the supervision and guidance of Prof. Amartya Chakraborty & Prof. Sumit Anand during the 6th Semester of the academic session of 2021 - 2022. The content of this report has not been submitted to any other university or institute. I am glad to inform that the work is entirely original and its performance is found to be quite satisfactory.

_____ _____

Prof. Amartya Chakroborty

Assistant Professor

Department of Computer Science & Engineering UEM, Kolkata

Prof. Sumit Anand

Assistant Professor

Department of Computer Science & Engineering UEM, Kolkata

_____

Prof. Sukalyan Goswami

HOD, Department of Computer Science and Engineering UEM, Kolkata

ACKNOWLEDGEMENT

I would like to thank everyone whose cooperation and encouragement throughout this project remain invaluable to me.

I am sincerely grateful to my guides Prof. Amartya Chakroborty and Prof. Sumit Anand of the Department of Computer Science & Engineering, UEM, Kolkata, for their wisdom, guidance, and inspiration that helped me to go through with this project and take it to where it stands now.

Last but not least, I would like to extend my warm regards to my family and peers who have kept supporting me and always had faith in my work.

- Apratim Ghosh

TABLE OF CONTENTS

ABSTRACT

Hate Speech is one of the most dangerous problems which is disturbing social and religious harmony. Social media which plays a vital role in our lives has now become a responsible element in spreading hate speech. This project report presents a comprehensive analysis of hate speech on Twitter, using Tweepy, an open-source Twitter API, to fetch relevant data. Our study sought to investigate the behavioral patterns of users responsible for spreading hateful comments online. After collecting a sample of tweets, we conducted research on the data, leveraging sentiment analysis techniques to segregate the tweets into positive and negative sentiments. Using Affinn lexicon and TextBlob, we determined the polarity and subjectivity of the tweets and performed k-means clustering to group the data into positive and negative clusters. We then plotted a graph to visualize the clustering results and analyzed the modularity of the data. Our findings provide valuable insights into the prevalence and nature of hate speech on social media, and offer suggestions for addressing this critical issue.

Keywords: hate speech, Twitter, troll, bully, gephi, tweepy, centrality

INTRODUCTION

In the past 10 years, the number of people using social networks and online forums has increased exponentially. Every 60 seconds, there are 510,000 comments generated on Facebook and around 350,000 tweets generated on Twitter. Social media is a very popular way for people to express their opinions publicly and to interact with others online. The people interacting on these forums or social networks come from different cultures and educational backgrounds. Unfortunately, any user who interacts online, whether on social media, forums, or blogs, runs the risk of being singled out or harassed by others using abusive language or expressing hate in the form of racism or sexism, which could have an adverse effect on both his or her online experience and the community at large. The existence of social networking services creates the need for detecting user-generated hateful messages prior to publication. Any published text that is used to express hatred towards some group with the intention to humiliate its members is considered a hateful message. This can lower the self-esteem of people, leading to mental illness and a negative impact on society. Furthermore, toxic language can take various forms, such as troll, cyber- bullying, harassment which was one of the major reasons behind suicide. Although hate speech is protected under the free speech provisions in some countries, e.g., the United States, there are other countries, such as Canada, France, United Kingdom, and Germany, where there are laws prohibiting it as being promoting violence or social disorder. Social media services such as Facebook and Twitter have been criticized for not having done enough to prohibit the use of their services for attacking people belonging to some specific race, minority etc. They have announced though that they would seek to battle against racism and xenophobia. Nevertheless, the current solutions deployed by, e.g., Facebook and Twitter have so far been to address the problem with manual effort, relying on users to report offensive comments. This not only requires a huge effort by human annotators, but it also has the risk of applying discrimination under subjective judgment. Moreover, a non-automated task by human annotators would have a strong impact on system response time, since a computer-based solution can accomplish this task much faster than humans. The massive rise in the user-generated content in the above social media services, with manual filtering not being scalable, high-lights the need for automating the process of on-line hate-speech detection.

In this paper, introducing social network analysis and its importance, then discussing Twitter as a rich resource for hate speech analysis. In the following sections, classifying tweets of various users based on polarity and subjectivity of their tweets and hence clustering them into two groups- positive and negative respectively.

LITERATURE SURVEY

Including user information in methods for detecting hate speech is an under-researched area. However, related to hate speech detection are studies of the people that post hateful content online, including characteristics and behavioral traits that are typical of the authors behind aggressive behavior, hate speech or trolling.

Chen et al. (2012) proposed a Lexical Syntactic Feature architecture to bridge the gap between detecting offensive content and potentially offensive users in social media, arguing that although existing methods treat messages as independent instances, the focus should be on the source of the content.

Waseem and Hovy (2016) stated that among various extra-linguistic features, only gender brought improvements to hate speech detection.

Papegnies et al. (2017) mention a plan to use context-based features for abuse detection, especially those based on the networks of user interactions. Several authors share this intention but face the challenge that user information often is limited or unavailable.

Cheng et al. (2015) characterized forms of antisocial behavior in online discussion communities, comparing the activity of users that have been permanently banned from a community to those that are not banned. The study found the banned users to use less positive words and more profanity, and concentrate their efforts on a small number of threads. They also receive more replies and responses than other users.

Hardaker (2010) defined a troller as a user who appears to sincerely wish to be part of a group, including professing or conveying pseudo-sincere intentions, but whose real intentions are to cause disruption or to trigger conflict for the purposes of their own amusement.

Buckels et al. (2014) studied the characteristic traits of Internet trolls by looking at commenting styles and personality inventories, and found strong positive relations among commenting frequency, trolling enjoyment and trolling behavior and identity.

Dan Jurafsky discussed sentiment tokenization issue, logical negation and extracting features for sentiment classification.

Gaurav Dubey et al. emphasized predefined dictionaries or sentiment tools can not contain the proper score of every word in the context to a sentence, hence forming a cluster of the results from both the tools' score, where grouping of 'definitely' positive and 'definitely' negative tweets was done.

Romero, et al. discussed the need to classify data and study of prediction parameters for any data analysis project
Zeerak Waseem et al. proposed normalizing the data by removing stop words, with the exception of "not", special markers such as "RT" (Retweet) and screen names, and punctuation.

Dana Warmsley explores the use of machine learning algorithms in understanding and detecting hate speech, hate speakers and polarized groups in online social media and highlights the major differences between different abusive language detection strategies, different annotation strategies are appropriate depending on the type of abuse.

PROBLEM STATEMENT

Social media has become an indispensable part of our daily life. Various social media platforms like WhatsApp, Facebook, Instagram, Twitter, etc. are giving people a chance to connect with each other across distances. Despite having so many perks, social media is often said to be one of the most harmful elements in society. Social media can have tragic consequences if not monitored. It has been observed that Twitter is most extensively used for spreading offensive comments and explicit content. Online hate content has become a major issue due to the exponential increase in the use of the internet by people of different cultures and educational backgrounds. Spiteful content often causes irreparable harm to the subjects. Many popular actors, politicians, singers, and media personnel raised their voices against hatred, while on the flip side, some people keep irrationally spreading hate content all over the internet. It is against Twitter's policy to spread violent threats, victimize others, and express hate. However, there is no simple way to identify all such users from Twitter's massive database so that appropriate action can be taken.

We tried to analyze various users' behavioral patterns by analyzing their tweets on Twitter and detecting the positive and negative tweets by means of clustering.

PROPOSED SOLUTION

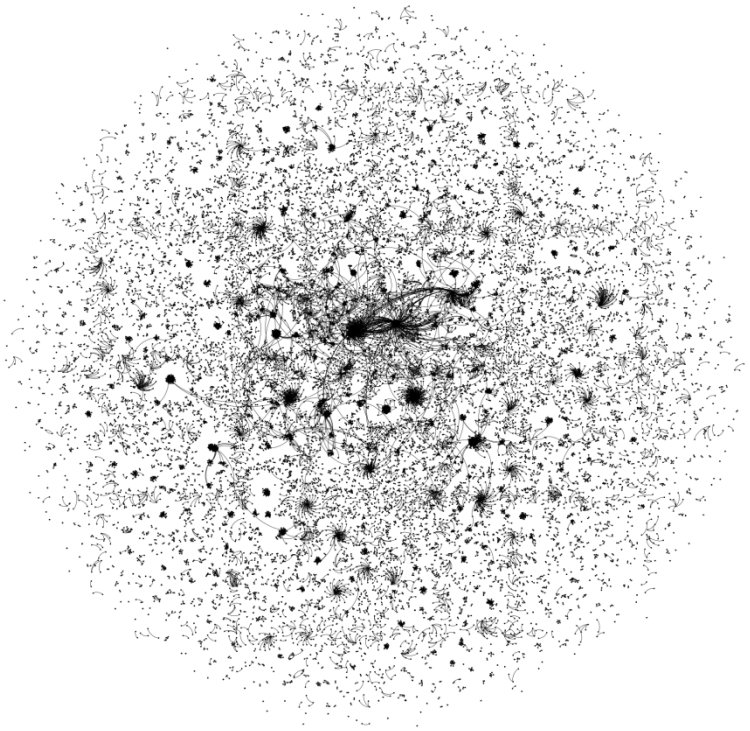
To address the issue of hate speech on Twitter, several proposed solutions can be implemented. One solution is to develop a hybrid approach to detect hate speech by combining rule-based and machine learning-based methods. This can improve the accuracy of detecting offensive language. Another solution is to implement multilingual support to accurately identify and classify hate speech in different regions and languages.

Transparency and user feedback can also be increased to build trust and credibility with users. Investing in user education and awareness can empower users to identify and report hate speech, creating a safer and more positive online environment. Stricter content moderation policies can also deter users from

spreading hate speech. Additionally, community-based reporting can help identify new trends in hate speech and improve the performance of automated detection systems.

EXPERIMENTAL SETUP

The project has been started by presenting the existing problem with freedom of speech on the Internet and the misuse of social media platforms like Twitter. These problems have become an integral part of the motivation. A comprehensive analysis work has been conducted by referring to the existing works in this field and coming up with a proposed solution for the problem. To identify hate speech in tweets, Python programming has been used. In this case, tweepy, an open-source Python library is used that provides a convenient approach to access the Twitter API.

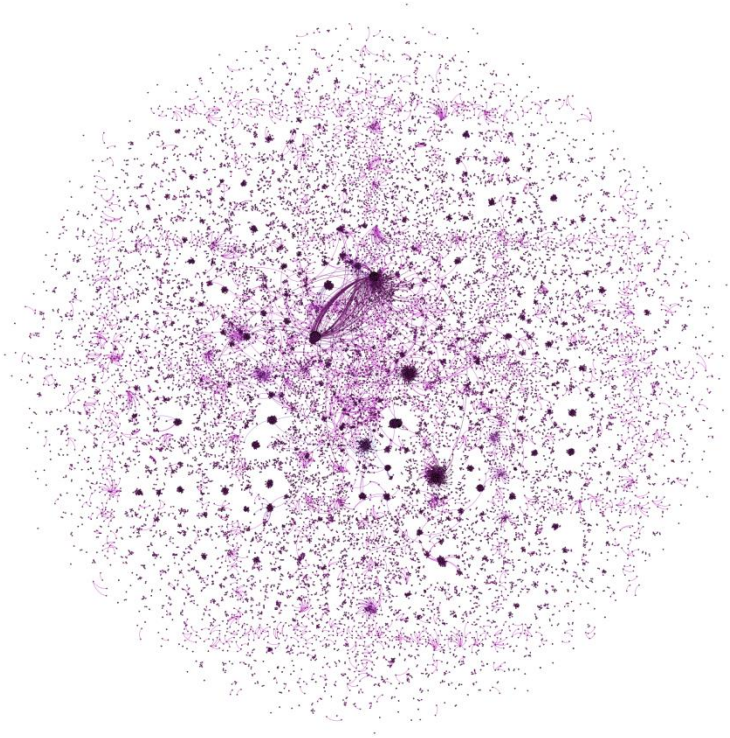


An adjacency list has been generated, after extracting the username and the retweet mentions from the dataset that was acquired using our algorithm. Tests were conducted using the Twitter dataset by exporting the values from the adjacency list to a .csv file into gephi. Gephi is an open-source software for network visualization as shown in Fig.1 and analysis. It helps to intuitively reveal patterns and trends, highlight outliers and tell stories with the data. To continue with the analysis, the number of factors using the Data Laboratory feature in Gephi were estimated, including Degree centrality, Eigenvector centrality, Betweenness centrality, and Pagerank centrality.

To determine the true sense(meaning what a user wants to convey) of a tweet whether it is negative or positive, “NOT” has been appended to each tweet. Then using affin lexicon, each tweet’s polarity and subjectivity has been calculated. An adjacency list has been created. Finally using the K-means clustering algorithm, the list has been divided into different clusters. Using

the software Gephi, modularity has been calculated and a graph has been generated for analyzing the behavior of the twitter users.

RESULT ANALYSIS



Degree Centrality

In graph theory, the degree (or valency) of a vertex of a graph is the number of edges incident to the vertex, with loops counted twice.

Degree Centrality is defined as the number of links incident upon a node (i.e., the number of ties that a node has). The higher the degree, the more central the node is. This can be an effective measure since many nodes with high degrees also have high centrality by other measures.

The mathematical formula to calculate

$$C_D(x) = \frac{\sum_{y=1}^{N} a_{xy}}{N-1}$$

degree centrality:
where N is the number of nodes on the graph and a has a value of either 0 or 1, depending on whether or not the nodes x and y share an edge.

ID	LABEL	IN-DEGREE	OUT-DEGREE	DEGREE
AOC	AOC	308	0	308
SirKazamJeevi	SirKazamJeevi	268	1	269
elonmusk	elonmusk	266	0	266
theestallion	theestallion	264	0	264
slang_100	slang_100	105	137	242

imyogi_26	imyogi_26	187	0	187
ThatCivilTweet	ThatCivilTweet	167	0	167
SirHairyPoppins	SirHairyPoppins	155	0	155
kundu_koushani	kundu_koushani	130	3	133
bennyjohnson	bennyjohnson	106	0	106

Table showing top user's Degree-Centrality

The table shows that AOC has a maximum of In-Degrees and a minimum of Out-Degrees. Therefore, we may conclude that American politician and activist @Alexandria Ocasio-Cortez, often known by her initials AOC, has the greatest effect among hateful terms like slang and offensive.

Additionally, @SirKazamJeevi is quite influential in terms of hatred. The individual has a sizable following. The table makes it obvious that this individual is very relevant on Twitter because of the large amount of out-degrees he has.

@Elon Musk has a maximum of In-Degrees and a minimum of Out-Degrees.Elon Musk is an influential face of twitter.

In the Table we see that theestallion has a maximum of In-Degrees and a minimum of Out-Degrees.After analyzing we may conclude that theestallion formally known as TINA SNOW, a american rapper has the greatest effect among hateful terms like troll,slang.

Furthermore @slang_100 has a mix of In-Degrees and Out-Degrees. After looking closely we may conclude that @Slang_100 is a rapper from South Africa,who has the greatest effect among hateful terms like slang.

@Imyogi_26 has the highest number of In-Degrees and a minimum of Out-Degrees. After a little bit of research we conclude that Imyogi_26 is a cricket lover,and has the greatest effect among hateful terms like revenge.

@ThatCivilTweet has a number of In-Degrees and a minimum of Out-Degrees. It is a fanbase account, and this account uses hate terms like kill,revenge.
@SirHairyPoppins has a number of In-Degrees and a minimum of Out-Degrees. After research we conclude that he is a US professional term like kill.

@Kundu_koushani has a number of In-Degrees and a minimum of Out-Degrees.From our analysis we conclude that she is a famous Indian lawyer and she is searching for the death of sushant singh rajput and she wants justice. She uses hate terms like narcissism.

@Bennyjohnson has a number of In-Degrees and a minimum of Out-Degrees.after research we conclude that American politician,journalist "Benny" Johnson , often known by Bennyjohnson currently serving as

chief creative officer at conservative organization Turning Point USA.he use hate terms like narcissism,troll, narcissism, kill, revenge.

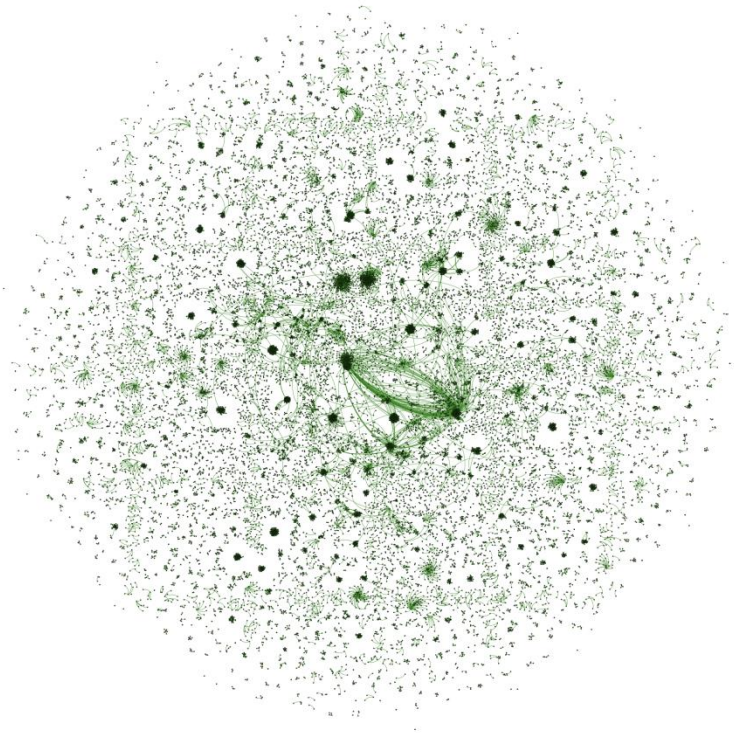
Degree centrality is a good measure of the total connections a node has, but will not necessarily indicate the importance of a node in connecting others or how central it is to the main group.

Betweenness centrality

In graph theory, betweenness centrality is a measure of centrality in a graph based on shortest paths. For every pair of vertices in a connected graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through (for unweighted graphs) or the sum of the weights of the edges (for weighted graphs) is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex.

The betweenness centrality of a node v is given by the expression:

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



Where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v (not where v is an endpoint).

Note that the betweenness centrality of a node scales with the number of pairs of nodes as suggested by the summation indices. Therefore, the calculation may be rescaled by dividing through by the number of pairs of nodes not including v, so that $g \in [0,1]$. The division is done by (N-1)(N-2) for directed graphs and (N-1)(N-2)/2 for undirected graphs, where N is the number of nodes in the giant component. Note that this scales for the highest possible value, where one node is crossed by every single shortest path. This is often not the case, and normalization can be performed without a loss of precision

$$\text{normal}(g(v)) = \frac{g(v) - \min(g)}{\max(g) - \min(g)}$$

Which results in:
max(normal) = 1
min(normal) = 0
Note that this will always be scaling from a smaller range into a larger range, so no precision is lost.

ID	LABEL	BETWEENNESS-CENTRALITY
slang_100	slang_100	14248.0
kundu_koushani	kundu_koushani	390.0
PRGuy17	PRGuy17	190.0
HawazGirmay	HawazGirmay	114.0
FutureUncleBae	FutureUncleBae	103.0
Vukile_Vee_	Vukile_Vee_	103.0
TimEastCoast	TimEastCoast	92.0
ke_troll	ke_troll	81.0
Venkate98225655	Venkate98225655	79.0
UN	UN	65.0

Table showing top user's Betweenness-Centrality

@Slang_100: He has maximum betweenness centrality. He is a Rapper from Durban, South Africa. He has appeared most often on the shortest path between nodes in the network.

@Kundu_koushani: She is known as koushani kundu. She has the most closeness centrality.

@PRGuy17: PRGuy is the Australian political commentator Jeremy Maluta. He has maximum betweenness centrality as maximum people reach out to them.

@HawazGirmay: He has the most closeness centrality. His tweets are mostly offensive.

@FutureUncleBae: He is from South Africa. He posts about cartoons and funny memes. His node shows the maximum eccentricity.

@Vukile_Vee_: He is from Durban, South Africa. He uses slang. His node shows the maximum eccentricity.

@TimEastCoast: He is an East Coaster in Toronto. His posts are mostly related to bullying, misogyny and economic inequity. He has the maximum betweenness centrality.

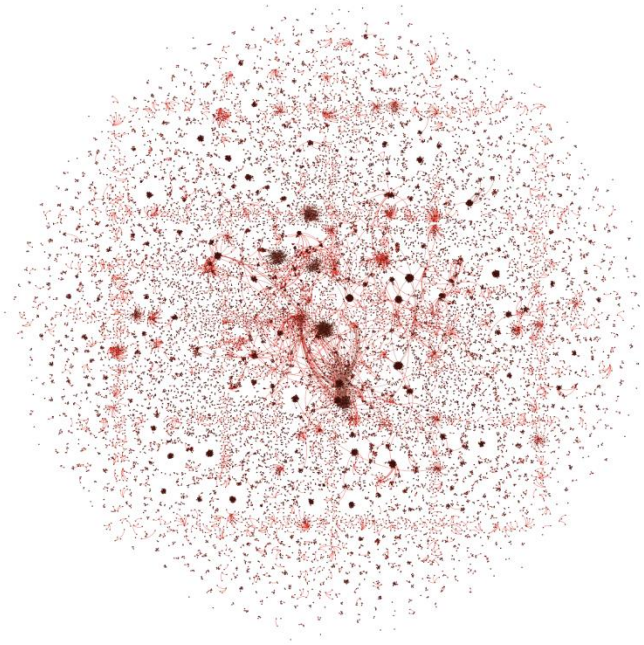
@ke_toll: It is a parody from Nairobi, Kenya. He has the shortest path between nodes in the network.

@Venkate98225655: He tweets about political issues. He tweets about narcissism. He has the maximum betweenness centrality.

@UN: UN is the official twitter account of the United Nations whose headquarters is situated in New York. It is an organization which has been working to create a better world where all people are able to enjoy peace, prosperity & human rights. It works towards equality so it has maximum reach.

Betweenness centrality is a widely used measure that captures a person's role in allowing information to pass from one part of the network to the other either in a good or bad way.

PageRank Centrality



PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites.

PageRank Centrality algorithm gives a bigger rank to pages that have large indegrees. Twitter, on the other hand, looks at follower relations and looks for accounts with large outdegrees and to determine popularity.

PageRank calculation is where it gets tricky. The PR of each page depends on the PR of the

$$PR(u) = \sum_{v \in B_u} \frac{PR(v)}{L(v)}$$

pages pointing to it. But we won't know what PR those pages have until the pages pointing to them have their PR calculated and so on… And when you consider that page links can form circles it seems impossible to do this calculation! What that means to us is that we can just go ahead and calculate a page's PR without knowing the final value of the PR of the other pages. That seems strange but, basically, each time we run the calculation we're getting a closer estimate of the final value.

The PageRank value for a page u is dependent on the PageRank values for each page v contained in the set Bu (the set containing all pages linking to page u), divided by the number L(v) of links from page v. The algorithm involves a damping factor for the calculation of the PageRank. It is like the income tax which the govt extracts from one despite paying him itself.

ID	LABEL	PAGERANK-CENTRALITY
SirKazamJeevi	SirKazamJeevi	0.0052746053080764636

theestallion	theestallion	0.004677721661867698
AOC	AOC	0.0039672050551522895
imyogi_26	imyogi_26	0.0036681982913691663
ThatCivilTweet	ThatCivilTweet	0.0032894909255948364
SirHairyPoppins	SirHairyPoppins	0.003069542979533763
slang_100	slang_100	0.002884374217474962
elonmusk	elonmusk	0.0027630135856196946
kundu_koushani	kundu_koushani	0.002543652401733602
turtlebreezee	turtlebreezee	0.002126908924986305

Table showing top user's PageRank-Centrality

@SirKazamJeevi: The fact that SirKazamJeevi has the highest PageRank centrality in the table indicates that he is well-known on Twitter. He is a satirist from India who posts mostly hateful posts about influential businessmen, politicians, and religion.

@Theestallion: In terms of PageRank centrality, she is second. She is a rapper from the US

@AOC: This list places Alexandria Ocasio-Cortez (AOC) third. So, we can say that she has many connections.She is an activist and a politician.She has been praised for her substantial social media presence in comparison to that of her fellow members of Congress.

@imyogi_26: The fact that Imyogi_26 is ranked fourth on the PageRankcentrality list indicates how influential he is when it comes to sharing hateful posts. Imyogi_26 has a passion for cricket and is a huge Virat Kohli fan.

@ThatCivilTweet: ThatCivilTweet has a PageRank of 5. She always attacks Muslims to spread hate via twitter..

@SirHairyPoppins: SirHairyPoppins has a PageRank of 6. He has been spreading a lot of negativity about cow meat consumption.

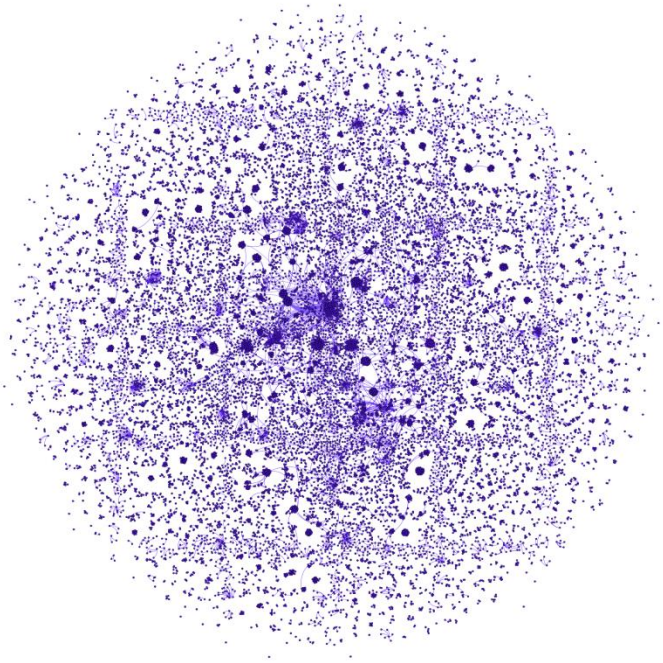
@slang_100: PageRank for slang_100 is 7. Although he is a new Twitter user, his followers are steadily increasing. He consistently spreads hatred toward the community.

@elonmusk: Elon Musk has 114.5 million followers and a PageRank of 8.We all know who Elon Musk so it is quite typical for him to receive vitriol, the majority of which is based on rumors.

@kundu_koushani: She has a PageRank of 9.She doesn't spread many rumors, but she is involved in a lot of them.

@Turtlebreezee: The PageRank of turtlebrezze is 10. We can find a wide range of bizarre comments in his account. Soon, he will be a sensation for hate speech.You will benefit more if you report this account than if you avoid it.

Eigenvector centrality



Eigenvectors are a special set of vectors associated with a linear system of equations (i.e., a matrix equation) that are sometimes also known as characteristic vectors, proper vectors, or latent vectors.

The eigenvector is a vector that is associated with a set of linear equations. Eigenvector centrality is a measure of the influence of a node in a network. Eigen centrality index calculates the centrality of a Twitter user based not only on their connections but also based on the centrality of that user's connections. Thus, eigenvector centrality is more important than Degree centrality. In a social network graph, edges originating from high-scoring nodes contribute more to the score of a node than connections from low-scoring nodes. A high eigenvector score means that a node is connected to many nodes that themselves have high scores.

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in V} a_{v,t} x_t$$

For a given graph G:=(V, E) with |A|=(av,t) vertices let be the adjacency matrix, i.e. av,t=1 if the vertex is linked to vertex, and av, t =0 otherwise. The relative centrality score, xv, of vertex v can be defined as:

ID	Label	Eigenvector Centrality
Slang_100	Slang_100	1.0
SirKazamjeevi	SirKazamjeevi	0.507736

AOC	AOC	0.444681
Elonmusk	Elonmusk	0.391354
Theestallion	Theestallion	0.384359
Imyogi_26	Imyogi_26	0.269747
ThatCivilTweet	ThatCivilTweet	0.240897
SirHairyPoppins	SirHairyPoppins	0.223587
Kundu_koushani	Kundu_koushani	0.187254
Bennyjohnson	Bennyjohnson	0.153297

Table showing top 10 users Eigenvector-Centrality

@Slang_100: He has maximum eigenvalue centrality means he has done interaction with such users who again have maximum eigenvalues. This denotes he has maximum influence in the network. From our research we found that he is a Rapper from Durban, South Africa.

@SirKazamjeevi: He comes second in eigenvector centrality which depicts his interactions were done among other most influential people of this network. He is a satirist from India who mostly shares hate posts about religion, influential businessman and politician.

@AOC: Alexandria Ocasio-Cortez also known as AOC is the third most influential person in this network. Her higher value of eigenvector suggests that she is very popular in her field. She is a politician and activist.

@Elonmusk: Elon Musk ranks 4^{th} in eigenvector centrality list which fairly denotes how influential he and his connections are in terms of sharing hate posts.

@Theestallion: Similarly Megan Thee Stallion ranks 5^{th} which means she is quite active in sharing explicit content such as her following and followers .

@Imyogi_26: Imyogi_26 ranks 6^{th} in eigenvector centrality list which fairly denotes how influential he is in terms of sharing hate posts.
@ThatCivilTweet: ThatCivilTweet ranks 7^{th} in eigenvector centrality list which fairly denotes how influential he is in terms of sharing hate posts through his fan account.

@SirHairyPoppins: SirHairyPoppins ranks 8th in the list. Sir Hairy Poppins is from the US. He is a professional photographer, artist and doodler. He is a political satire.

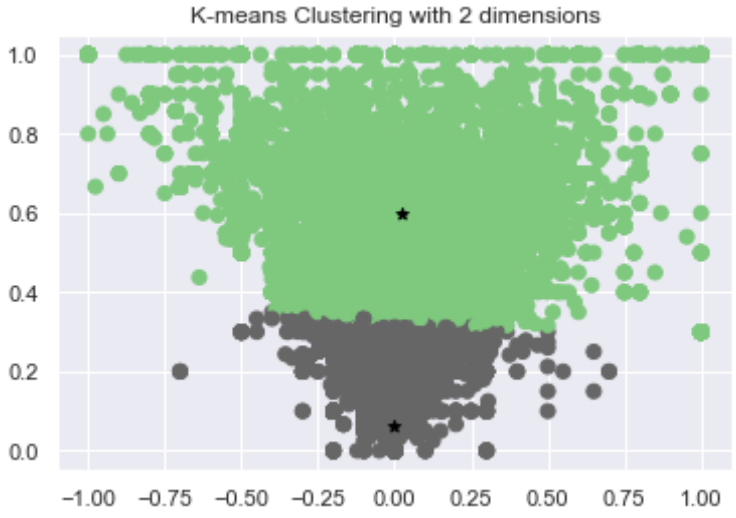
@Kundu_koushani: Kundu_koushani is known as Koushani Kundu. She is a famous lawyer. Kundu_koushani is the 9th most influential person in this network. Her higher value of eigenvector suggests that she is very popular in her field.

@Bennyjohnson: Benny Johnson is the 10th most influential person in this network. His higher value of eigenvector suggests that he is very popular in his field. Benny Johnson, original name Beneful "Benny" Johnson is an American political columnist and a journalist, currently serving as chief creative officer at conservative organization Turning Point USA.

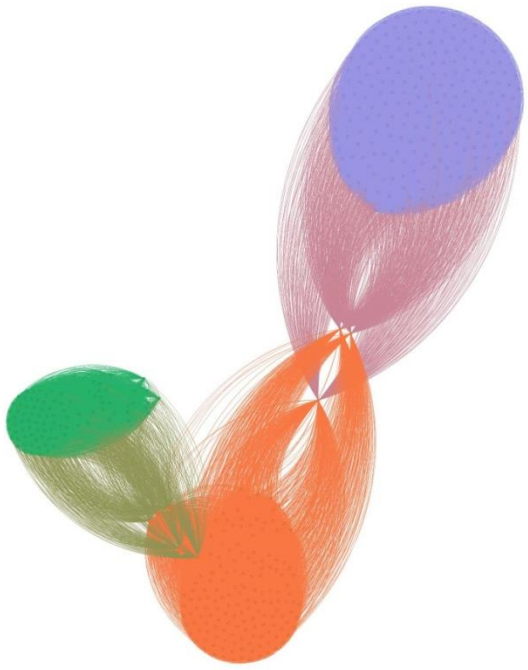
From the chart given above we can conclude that the user Slang_100 has maximum eigenvalue centrality (1.0) that means not only he has maximum influence in this network but his connections are also influential. Similarly SirKazamjeevi, Aoc,Elon Musk have higher importance in this network.

K-Means Algorithm

K-Means clustering is a popular unsupervised machine learning algorithm that is used to group similar data points together. The algorithm works by partitioning the data into k clusters, where k is a predefined number of clusters that the user chooses. The algorithm then iteratively assigns each data point to the nearest cluster centroid and updates the centroid to the mean of the data points assigned to it.



Utilizing the Afinn Lexicon to determine each user's subjectivity and polarity, the k-means algorithm has been utilized to create clusters based on subjectivity and polarity. Depending on the data, clusters are created. After that, a plot is created to display the data.



Modularity

Modularity is a measure of the strength of the division of a network into groups or clusters. It is a quantitative measure that evaluates how well the nodes in a network are grouped into clusters, based on the assumption that nodes within the same

cluster have more connections to each other than to nodes in different clusters.

The modularity of a network is typically represented by a single number between -1 and 1. A higher modularity score indicates a stronger and more distinct division of the network into clusters. A modularity score of 0 indicates that the network is not partitioned into any clusters, while a negative modularity score suggests that the network is not well-clustered or may be random.

The formula for calculating the modularity of a network is as follows:

$Q = 1/2m * \text{sum over } i,j \ (A_{ij} - k_i*k_j/2m) * \text{delta}(c_i, c_j)$

where:
Q is the modularity score
m is the total number of edges in the network
A_{ij} is the weight of the edge between nodes i and j
k_i and k_j are the degrees of nodes i and j, respectively
c_i and c_j are the community assignments of nodes i and j, respectively, with delta(c_i, c_j) being 1 if $c_i = c_j$ and 0 otherwise.
The modularity formula measures the difference between the number of edges within communities and the expected number of edges within communities if edges were placed randomly. The term $A_{ij} - k_i*k_j/2m$ is the difference between the actual and expected number of edges between nodes i and j. The summation over i and j calculates the total difference between the actual and expected number of edges between all pairs of nodes, while the delta function ensures that only nodes within the same community are considered. Finally, the modularity score is normalized by 1/2m.

The resultant network's modularity has been determined. The final result was 0.895.

CONCLUSION & FUTURE SCOPE

The detection of hate speech is challenging due to various factors such as language, cultural nuances, and the sheer volume of content generated. The paper provides empirical study of hate speech detection in tweets. To identify hate speech in tweets, Python programming has been used. In this case, tweepy, an open-source Python library is used that provides a convenient approach to access the Twitter API. Analysis of the data has been done by forming graphs and determining various parameters, classifying tweets of various users based on polarity and subjectivity of their tweets and hence clustering them into two groups- positive and negative respectively. Finally Modularity has been calculated to showcase the strength of the division of a network into groups or clusters.

In the future, the proposed solutions to address hate speech on Twitter can be further improved and expanded upon. For example, the hybrid approach for detecting hate speech can be enhanced by

incorporating more advanced machine learning techniques such as deep learning. Multi-lingual support can be expanded to cover more languages and regions, and community-based reporting can be improved to increase its effectiveness. Overall, there is great potential for further development and improvement in the fight against hate speech on Twitter and other social media platforms.

REFERENCES

1. H. Watanabe, M. Bouazizi and T. Ohtsuki, "Hate Speech on Twitter: A Pragmatic Approach to Collect Hateful and Offensive Expressions and Perform Hate Speech Detection," in *IEEE Access*, vol. 6, pp. 13825-13835, 2018, doi: 10.1109/ACCESS.2018.2806394.
2. Unsvåg, Elise Fehn, and Björn Gambäck. "The effects of user features on Twitter hate speech detection." *Proceedings of the 2nd workshop on abusive language online (ALW2)*. 2018.
3. Tontodimamma, A., Nissi, E., Sarra, A. *et al.* Thirty years of research into hate speech: topics of interest and their evolution. *Scientometrics* 126, 157–179 (2021).
4. Pitsilis, Georgios K., Heri Ramampiaro, and Helge Langseth. "Effective hate-speech detection in Twitter data using recurrent neural networks." *Applied Intelligence* 48.12 (2018): 4730-4742.
5. Koushik, Garima, K. Rajeswari, and Suresh Kannan Muthusamy. "Automated hate speech detection on Twitter." *2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA)*. IEEE, 2019.
6. Pereira-Kohatsu, Juan Carlos, et al. "Detecting and monitoring hate speech in Twitter." *Sensors* 19.21 (2019): 4654.
7. Waseem, Zeerak, and Dirk Hovy. "Hateful symbols or hateful people? predictive features for hate speech detection on twitter." *Proceedings of the NAACL student research workshop*. 2016.
8. Jurafsky, Dan, and Christopher Manning. "Natural language processing." *Instructor* 212, no. 998 (2012): 3482
9. Ahuja S, Dubey G. Clustering and sentiment analysis on Twitter data. In2017 2nd International Conference on Telecommunication and Networks (TEL-NET) 2017 Aug 10 (pp. 1-5). IEEE..
10. C. Romero, S. Ventura, P. de Bra, and C. Castro, “Discovering prediction rules in aha! courses,” Proceedings of the International Conference User Modelling, 25-34.
11. .Waseem Z, Hovy D. Hateful symbols or hateful people? predictive features for hate speech detection on twitter. In Proceedings of the NAACL student research workshop 2016 Jun (pp. 88-93).
12. Davidson T, Warmsley D, Macy M, Weber I. Automated hate speech detection and the problem of offensive language. InProceedings of the international AAAI conference on web and social media 2017 May 3 (Vol. 11, No. 1, pp. 512-515).