

Eximia journal

www.eximiajournal.ro

Vol. 14/2025

PLUS
COMMUNICATION P



International
Communication & PR

Exploring the Societal Attitudes: A Sentiment Analysis of the Public Views on the LGBT Community Using Natural Language Processing

Charis Pasion¹, Kevin Jay Daniel¹, Joseph M. A. Huelgas^{1,2}, Angelo Morancil¹,
Eddie De Paula^{1,2}, Niel Bunda^{3,*}

¹ Graduate School, University of St. La Salle, Bacolod City, 6100, Philippines

² Computer Science Department, University of St. La Salle, Bacolod City, 6100, Philippines

³ Information Technology Department, University of St. La Salle, Bacolod City, 6100, Philippines

Corresponding author: *niel.bunda@usls.edu.ph

ABSTRACT

This study investigates global attitudes toward the LGBTQ community through sentiment and thematic analysis of Twitter data. Utilizing the OSEMN framework, the research analyzes a dataset of 29,890 tweets collected over a six-day period. After data preprocessing using natural language processing techniques, exploratory analysis revealed key linguistic patterns and commonly discussed topics. Latent Dirichlet Allocation (LDA) topic modeling uncovered four dominant themes within the discourse: (1) promoting sexual health through political engagement, (2) balancing freedom of expression and inclusivity in digital spaces, (3) empowerment and transformation through representation, and (4) addressing prejudice and discrimination in political leadership. Sentiment analysis results show that 47.73% of the tweets expressed positive sentiment, often reflecting support, celebration, and advocacy for LGBTQ rights and identity. In contrast, 26.25% were negative, with many tweets containing hate speech, misinformation, or stigmatizing language tied to religious or political ideologies. The remaining 26.02% were classified as neutral, reflecting a mix of informational or non-committal content. Visualizations such as word clouds, bi-grams, tri-grams, and a network graph, supported the interpretation of topic clusters and revealed the nuanced dynamics of online LGBTQ discourse. The findings suggest that Twitter functions both as a platform for visibility and affirmation, and as a space where prejudice persists. This disparity emphasizes the ongoing need of awareness and advocacy in online spaces and the intricacy of digital environments in influencing public perceptions of marginalized groups.

Keywords — *Sentiment Analysis; thematic analysis; LDA; LGBT; NLP; public views; societal attitudes*

1. INTRODUCTION

In contemporary society, the discussion surrounding the Lesbian, Gay, Bisexual, and Transgender (LGBT) community has become increasingly prominent, reflecting a diverse range of opinions and attitudes. Often known as LGBTQ+, GLBT, gay, or queer, the LGBT community is a loosely defined collection of people bound together by common cultural traits and involvement in social activities [1]. The community helps to challenge society norms including heterosexism, homophobia, biphobia, transphobia, sexualism, and conformist pressures by celebrating pride, diversity, uniqueness, and sexuality.

This heightened visibility of the LGBT community within public discourse has led to a polarization of opinions. On the one hand, there is a growing acceptance and support for the rights and dignity of LGBT individuals, while on the other hand, opposition and skepticism persist. The discussion has spilled over onto social media platforms, where a substantial volume of opinions, sentiments, and narratives are shared, shaping the broader societal dialogue. Social media platforms have become powerful arenas for public expression and opinion-sharing. The diversity of voices on these platforms mirrors the complexities inherent in societal attitudes toward the LGBT community. Debates, discussions, and narratives unfold in real-time, offering a rich source of data that can be harnessed to better understand the sentiments surrounding the LGBT community.

Research on sentiment analysis of LGBT-related content has made valuable contributions, but a predominant limitation is the focus on specific geographic regions. For example, Aldinata et al. [2] study explored sentiment

analysis on LGBT-related tweets, concentrating exclusively on the diverse states in America. In addition, Doğan et al. [3] conducted a sentiment analysis to evaluate hate speech present in Turkish LGBT-related tweets. Kusuma et al. [4] explore the sentiment of Indonesian Twitter users during the time when The International Day Against Homophobia, Biphobia, and Transphobia (IDAHOBIT) was observed in the country. While insightful for understanding sentiments within countries such as the United States, Turkey, or Indonesia, such regional studies may not capture the global spectrum of opinions and emotions toward the LGBT community.

This study adopts a more expansive approach, utilizing a diverse dataset comprising tweets from various regions worldwide. While most research studies concentrate on specific countries or regions, this study aims to fill this void by providing a comprehensive analysis of sentiments towards the LGBT community on a global scale. The utilization of a dataset encompassing tweets from diverse global locations enables a broader scope, allowing for cross-cultural comparisons to examine attitudes toward the community comprehensively. By undertaking this global perspective, our study contributes valuable insights that extend beyond geographical boundaries, fostering acceptance, inclusivity, and interventions to protect the rights and dignity of LGBT individuals on a global scale.

2. RESEARCH METHODOLOGY

In this study, the OSEMN framework is used as a guiding methodology for the data science process. This model outlines five essential steps: obtaining the data (O), cleaning and preparing it (S), exploring it to uncover patterns and insights (E), building analytical models (M), and finally, interpreting the results to draw meaningful conclusions (N). Each phase of this framework shaped how the data was handled and analyzed throughout the research [5].

As part of the Obtain phase in the OSEMN framework, we sourced the dataset from Kaggle, a well-known platform for machine learning and data science communities that offers a wide range of open-source datasets [6]. Specifically, a dataset containing approximately 32,456 tweets related to the LGBT community. These tweets were collected over a six-day period, from August 21 to August 26, 2022. The timeframe likely captures online discussions influenced by significant global events related to LGBT issues, which may provide relevant context for the data collected.

The raw Twitter data underwent a thorough cleaning process to ensure its quality and usability for analysis. Out of the original dataset, 29,890 tweets were retained after initial filtering. Data cleaning was performed using the R programming language within the RStudio environment [7], which is widely used for statistical computing and data visualization. The cleaning process involved multiple steps commonly applied to social media data. First, tweets were converted to lowercase to maintain consistency [8]. Then, unnecessary elements such as URLs, user mentions (e.g., @username), hashtags, retweet indicators (e.g., "RT"), and special characters were removed [9]. Punctuation and numerical digits that did not contribute meaningfully to the textual analysis were also stripped [10]. In addition, common stop words (such as "the," "and," "is") were filtered out to reduce noise in the data [11], [12]. The remaining text was tokenized, and where necessary, stemming or lemmatization was applied to group words with similar roots [13]. These steps ensured that the dataset was cleaned and standardized, making it suitable for subsequent exploration and modeling [8].

Latent Dirichlet Allocation (LDA) were then applied [14] extract patterns and thematic structures from the cleaned dataset. LDA enabled us to uncover underlying topics within the tweets by identifying clusters of words that frequently appear together. A variety of visual modeling tools were used to analyze the dataset further like word cloud to depict the most frequently mentioned words related to LGBT topics, n-grams to highlight the most common word pairings and sequences and network graph to visualize the connections between words across different topics, illustrating how themes overlap and interact.

3. RESULTS AND DISCUSSION

After a thorough data cleaning process, the dataset was refined and prepared for analysis to extract meaningful insights using through sentiment analysis and thematic modeling. One of the first notable observations was the average tweet length, which was 87.97 characters. This reflects the nature of Twitter as a platform designed for brief, rapid communication [15]. Despite the character limitations, users were often able to express clear and impactful messages, ranging from expressions of personal identity to commentary on social and political matters. This brevity, while constraining in form, encourages concise and emotionally charged language, making tweets a rich source of public sentiment and discourse, particularly around complex and sensitive topics such as those concerning the LGBT community.

statements about identity or opinion. Meanwhile, phrases like “part lgbt” and “lgbt flag” reflect themes of identity affiliation and symbolic representation.

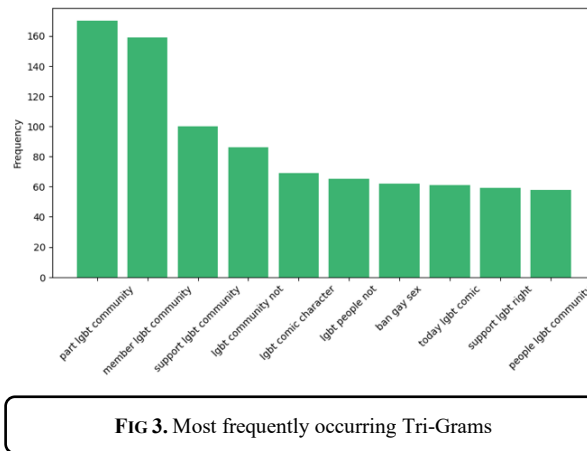


FIG 3. Most frequently occurring Tri-Grams

Fig. 3 illustrates the most frequent trigrams which is three-word combinations that are extracted from tweets related to LGBTQ discourse. The most common trigram, “part lgbt community,” followed closely by “member lgbt community,” highlights a recurring theme of inclusion and identity within the LGBTQ space. These phrases suggest that users often emphasize their personal affiliation or involvement with the community, reinforcing a sense of belonging. The third most frequent trigram, “support lgbt community,” reflects affirming discourse, where expressions of solidarity and advocacy are prominent. Additionally, “lgbt community not” may reflect contrasting or clarifying stances within ongoing debates or criticisms, perhaps linked to political or ideological commentary. Other trigrams such as “lgbt comic character” and “today lgbt comic” suggest a focus on representation in media, particularly in comic books and popular culture. Meanwhile, phrases like “ban gay sex” indicate discussions around policy, censorship, or legal issues affecting the LGBTQ population, potentially in specific regions. Trigrams like “support lgbt right” and “people lgbt community” further reinforce the overarching themes of advocacy, rights, and collective identity. The distribution of these trigrams reflects a multifaceted conversation that balances pride and visibility with ongoing political struggles and the assertion of rights.

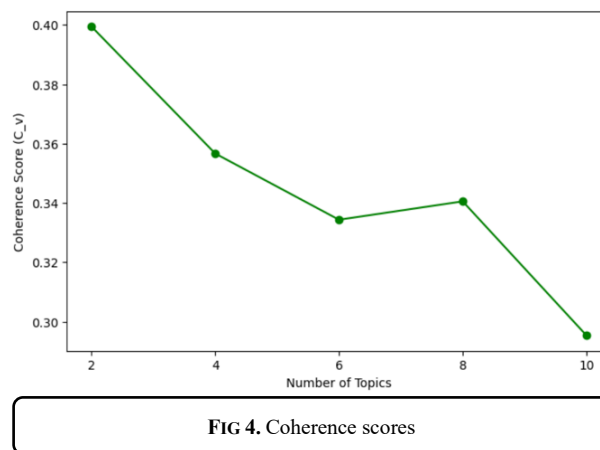


FIG 4. Coherence scores

Fig. 4 presents a line graph depicting the coherence scores (C_v) corresponding to various topic counts in the LDA model applied to LGBTQ-related tweets. Before applying LDA, it is essential to determine the optimal number of topics to ensure that the resulting model produces interpretable and meaningful groupings. The coherence score is a quantitative metric used to evaluate the interpretability of topics, with higher scores indicating greater semantic consistency among the top words in each topic [17]. As illustrated, the coherence score is highest at 2 topics, indicating that a model with two topics provides the most interpretable groupings of terms within this dataset. However, since a two-topic model may be overly simplistic for capturing the diversity of conversations within LGBTQ discourse, a balance must be found between interpretability and thematic depth. The score drops to 0.3658 for 4 topics, which still offers acceptable coherence while allowing for a more nuanced analysis of

themes. Beyond four topics, the coherence score continues to decline slightly and fluctuates—reaching its lowest point at 10 topics. This downward trend suggests that models with a higher number of topics may result in more fragmented or overlapping themes, reducing clarity and interpretability. Based on this graph, a 4-topic model appears to offer an optimal balance—retaining a relatively high coherence score while allowing the LDA model to capture meaningful variation across different dimensions of LGBTQ discourse. Therefore, selecting four topics offers a practical trade-off between maintaining coherence and capturing the thematic diversity of LGBTQ discourse.

TABLE 1. Top Weighted Words for Each Topic

Topic 1	Topic 2	Topic 3	Topic 4
lgbt	lgbt	lgbt	lgbt
not	people	not	people
would	like	community	make
im	right	gay	country
love	dont	many	need
person	one	thats	issue
black	not	im	etc
men	community	sex	lot
white	want	never	school
much	woman	get	medium
hate	even	first	nothing
law	trans	book	youre
also	think	could	child
ppl	support	stop	u
conservative	life	queer	kid

TABLE 2. Dominant tweets per Topic

Topic	Contribution	Most Dominant Tweet
1	78.58%	“shares:Tory conference: LGBT group unveils politics-themed condoms: LGBT+ Conservatives unveil politics-themed condoms to promote sexual health at Birmingham gathering. Thank you”
2	86.01%	“i dont really care how people mod their games but taking it down is also discriminatory against the people who dont really care for lgbt i dont care who or what you identify. its just a flag. taking that down that mod is technically anti American”
3	75.16%	“Midnight on Olympus - Aoife & Gen. Germanicus face murder, magic & mutiny. Aoife flees to Greece; returns to the arena as gladiatrix Amazonia Nyx & unites w/ the Sisters of Light in an epic war against the Brotherhood of the Sun. #SFFPit #A #LF #HF #MR #LGBT 4320 Aoife & Gen. Germanicus face murder, magic & mutiny. Aoife flees to Greece; returns to the arena as gladiatrix Amazonia Nyx & unites w/ the Sisters of Light in a war against the Brotherhood of the Sun in Midnight on Olympus. #SFFPit #A #LF #HF #MR #LGBT”
4	80.12%	“Japan : A history of sexist, homophobic and transphobic remarks by vice-minister Mio Sugita. Ms Sugita compared #equalmarriage to a human marrying a machine, or pets, and diminished claims of sexual assault, and victim-blamed sexual assault survivors #LGBT”

Table 1 displays the top weighted words associated with each topic generated by the LDA model. These word groupings provide insight into the underlying themes present in the LGBTQ-related tweets. Notably, the word “lgbt” appears prominently across all four topics, reflecting its centrality and broad relevance to the discourse. However, the co-occurring terms for each topic differ, helping to distinguish the unique focus of each cluster. In

Topic 1, the presence of words such as “love,” “im,” “person,” “black,” “men,” and “white” suggests a focus on individual identity, emotional expression, and possibly discussions around intersectionality, particularly race and gender within the LGBTQ context. This topic appears to center on personal narratives and identity-related issues. Topic 2 is characterized by terms like “people,” “like,” “right,” “dont,” “woman,” “trans,” and “support,” indicating a focus on civil rights, gender inclusivity, and advocacy. The inclusion of “trans” and “support” implies conversations about gender identity and the push for equitable treatment and recognition, aligning with ongoing social movements. In Topic 3, words such as “community,” “gay,” “many,” “never,” “first,” “queer,” and “book” suggest discourse related to cultural identity and representation, including experiences shared within the LGBTQ community and how those experiences are reflected in literature or media. The presence of “stop” and “could” might indicate discussions around change, resistance, or calls to action. Topic 4 includes terms such as “make,” “country,” “issue,” “school,” “child,” and “kid,” which point toward broader societal or political discussions. These likely encompass educational policy, family dynamics, and national-level challenges related to LGBTQ rights or visibility. The recurring mention of younger demographics suggests that part of the conversation may focus on LGBTQ youth, schooling environments, and protective legislation.

Table 2 presents the most dominant or representative tweet for each of the four LDA-derived topics, selected based on the highest contribution (topic probability) from the model. These tweets exemplify the core themes identified in Table 1, which lists the most heavily weighted words per topic. Together, these two tables provide a comprehensive view of both the abstract topic structure and the real-world expressions of those themes within the dataset. In Topic 1, the dominant tweet discusses a conservative political group unveiling politics-themed condoms to promote sexual health. This aligns closely with high-weight terms in Table 1 such as “love,” “person,” “law,” “conservative,” and “white”, which collectively point toward political engagement, health awareness, and identity issues. This topic is further reflected in Fig. 1, where terms like “support,” “right,” and “community” stand out. Additionally, bigrams such as “support lgbt” (Fig. 2) reinforce the same advocacy-driven tone. For Topic 2, the representative tweet expresses frustration over the removal of an LGBT-themed game modification, citing it as discriminatory and anti-American. The vocabulary in Table 1 under Topic 1 includes “people,” “right,” “support,” “trans,” and “dont,” reinforcing a focus on digital inclusivity, freedom of expression, and the policing of LGBTQ content online. This sentiment is mirrored in Fig. 3, where phrases like “support lgbt community” and “battle lgbt right” highlight a contentious but active online environment around representation and rights in digital spaces. Topic 3 features a tweet promoting a fantasy novel with LGBTQ representation. This reflects Topic 3’s key terms like “community,” “gay,” “queer,” “book,” and “first” from Table 1, pointing toward themes of cultural expression, identity, and transformation. The presence of media-related trigrams in Fig. 3—such as “lgbt comic character”—aligns with this narrative of LGBTQ visibility and representation in literature and storytelling. The dominant tweet for Topic 4 discusses homophobic and transphobic remarks made by a Japanese political figure. The words associated with Topic 4, “country,” “issue,” “child,” “school,” “need,” and “make”, suggest structural and institutional discourse. The tweet also directly aligns with terms in Fig. 1 such as “rights,” “people,” and “community,” and connects to bigrams like “anti lgbt” (Fig. 2), pointing to the social and political challenges faced by LGBTQ individuals.

TABLE 3. Mapping of topics to latent themes

	Topic 1	Topic 2	Topic 3	Topic 4
	lgbt	lgbt	lgbt	lgbt
	not	people	not	people
	would	like	community	make
	im	right	gay	country
	love	dont	many	need
	person	one	thats	issue
	black	not	im	etc
	men	community	sex	lot
	white	want	never	school
	much	women	get	medium
	hate	even	first	nothing
	law	trans	book	youre
	also	think	could	child
	ppl	support	stop	u

	conservative	life	queer	kid
Underlying Themes	Promoting Sexual Health Through Political Engagement and Awareness	Balancing Freedom of Expression and Inclusivity in Digital Spaces	Empowerment and Transformation	Addressing Prejudice and Discrimination in Political Leadership

Table 3 shows how the topics generated by the LDA model were mapped to broader latent themes within the dataset. Each topic was represented by its most relevant terms, which helped in identifying an overarching theme that captures the core meaning behind the group of words. Each of the four underlying themes are discussed in context, using both the key terms and supporting literature, to arrive at meaningful interpretation that reflect different yet interconnected aspects of LGBTQ discourse across social, political, and cultural domains.

Underlying Theme 1 focuses on how political engagement plays a vital role in advocating for comprehensive sexual health education and services by raising public awareness through political platforms, which helps destigmatize sexual health issues and fosters informed decision-making. Additionally, community mobilization and political involvement empower individuals to demand improved sexual health resources and representation while actively participating in political processes helps address disparities in sexual health services, ensuring equitable care across all populations [18], [19].

Underlying Theme 2 is about promoting freedom of expression in digital spaces involves establishing guidelines that protect open dialogue while preventing the proliferation of harmful speech [20]. To achieve true inclusivity, it is essential to create environments where marginalized groups feel safe and respected, even amidst the exchange of differing opinions. According to Kossowska et al. [21], this means not only safeguarding the right to express diverse viewpoints but also ensuring that these digital spaces foster an atmosphere of respect and safety for all users. By balancing the protection of open dialogue with the need for a respectful and inclusive environment, we can support both free expression and the well-being of all participants in digital communities.

Underlying Theme 3 highlights empowerment and transformation within the LGBTQ+ community, emphasizing how inclusive policies and supportive environments can drive social change. This theme explores how institutions, governments, and organizations play a role in enabling LGBTQ+ individuals to assert their rights and identities more confidently. For example, Francis [22], describes how South African NGOs work alongside schools to empower LGBT+ youth by promoting gender and sexuality diversity in educational spaces. Similarly, in 2021, France’s legalization of Assisted Reproductive Technology (ART) for female couples and single women marked a transformative step in expanding family rights and access to parenthood for LGBTQ+ individuals, reflecting a broader shift toward civil rights in the Western world [23]. Nedeljko and Kaučič [24] further emphasize the importance of creating safe, inclusive environments for LGBTQ+ older adults—a growing and often overlooked group—who continue to face discrimination and stigma. Together, these examples illustrate how empowerment through legal reform, education, and inclusive practices can transform individual lives and contribute to broader societal acceptance.

Underlying Theme 4 canters on addressing prejudice and discrimination in political leadership, particularly as it affects LGBTQ individuals who seek political representation. In recent years, there has been a notable rise in the number of LGBTQ candidates running for public office, reflecting increasing visibility and a slow but steady shift in societal acceptance [25]. However, despite this progress, many still encounter persistent discrimination rooted in entrenched gender norms, religious conservatism, and widespread homophobia or transphobia [26]. A major source of opposition often comes from conservative and religious voting blocs, who are generally less supportive of LGBTQ candidates and more likely to hold negative attitudes toward their involvement in politics. In response to these barriers, LGBTQ activists and allies frequently mobilize to resist discriminatory actions and advocate for equal rights. As Blackington [27] notes, these activist efforts are critical in countering harm and fostering a more equitable political landscape for LGBTQ individuals.

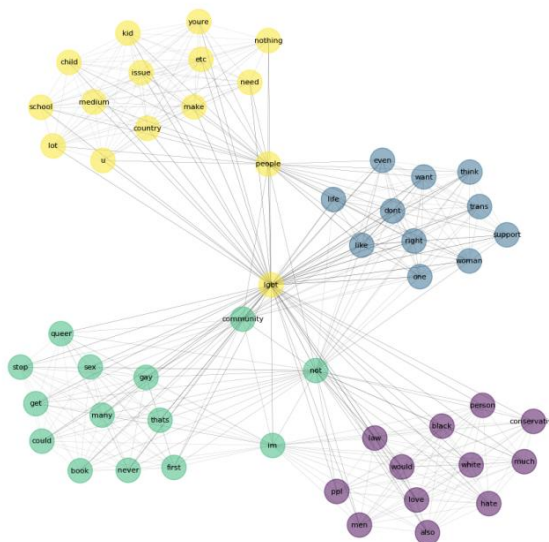
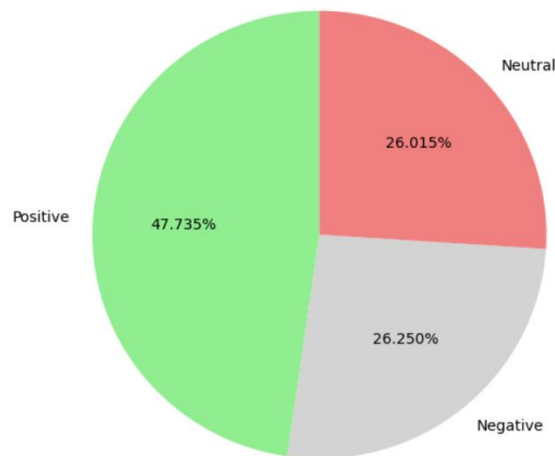


FIG 5. Network graph of words across LDA-generated topics

Fig. 5 visualizes the network graph of words across topics generated from the LDA model. Each node represents a word, and edges indicate co-occurrence relationships within the same topic. The color-coding of nodes corresponds to different topic clusters, while the central placement of the word “lgbt” highlights its role as a key connector across all themes, reinforcing its prominence as observed in Tables 1 and 2. The structure of the graph visually supports the topical groupings identified earlier. The yellow cluster in the upper left corresponds to Topic 4, featuring terms such as “child,” “school,” “issue,” “kid,” “country,” and “need.” This aligns with the structural and institutional focus noted in Table 1 and the representative tweet in Table 2, which discussed political and societal discrimination, particularly in the context of education and public discourse. This grouping reinforces Theme 4: Addressing Prejudice and Discrimination in Political Leadership. On the right side, the blue cluster reflects Topic 2, which includes words like “people,” “right,” “trans,” “support,” and “woman.” These terms relate to gender identity, advocacy, and civil rights, echoing both the dominant words in Table 1 and the tweet in Table 2 that critiques digital censorship, supporting Theme 2: Balancing Freedom of Expression and Inclusivity in Digital Spaces. The green cluster on the left represents Topic 3, characterized by words such as “community,” “gay,” “queer,” “book,” and “many.” This cluster highlights discussions of LGBTQ representation in media and culture, aligning with the tweet about a fantasy novel featuring LGBTQ characters and reflecting Theme 3: Empowerment and Transformation. Meanwhile, the purple cluster at the bottom relates to Topic 1, which includes terms like “love,” “conservative,” “law,” “black,” “white,” “men,” and “hate.” These words point to complex intersections of politics, race, identity, and emotional expression. The tweet referencing a politically themed health campaign by LGBT+ Conservatives is a strong fit for this grouping, supporting Theme 1: Promoting Sexual Health Through Political Engagement and Awareness.

In addition, Fig. 1, 2, and 3 further reinforce these interpretations. Fig. 1 highlights central terms like “community,” “right,” “support,” and “people,” emphasizing their significance within the network. Fig. 2 features commonly occurring phrases such as “lgbt community” and “support lgbt,” which are consistent with the network’s structure and the high connectivity of these terms. Fig. 3 captures phrase-level patterns like “support lgbt community,” “member lgbt community,” and “battle lgbt right,” which reflect similar relationships found within the graph’s overall structure. Fig. 5 confirms the dominant words across each topic are not only thematically cohesive but also interconnected, capturing the nuanced and multifaceted nature of LGBTQ discourse on Twitter. It strengthens the coherence of previous analyses, linking the LDA outputs to real-world expressions of identity, advocacy, and social commentary.

**FIG 6.** Sentiment distribution

Sentiment analysis, otherwise known as opinion mining, utilizes Natural Language Processing (NLP) to determine whether a user or an individual's attitude toward a certain service, product, or topic is positive, neutral, or negative [28]. Results of the sentiment analysis depicted that 47.735% of the dataset, or 14,268 tweets showed a positive attitude towards LGBT. Members of the community tend to seek support from social media platforms due to the struggles they experience in their respective offline environments. Social media platforms allow them to establish online interactions and build genuine relationships with open-minded individuals and fellow members of the community [29]. With this, there is highly likely a lower internalized stigma towards LGBT that gives them a sense of security to openly express their sexual identity, tackle relevant social issues, and sharing of information affirming their identity [30]. On the other hand, 7,846 tweets, or a total of 26.250% of the dataset, expressed negative opinions towards LGBT. The presence of the LGBT community in social media is like a double-edged sword in a way that it can be both beneficial and threatening to the members of LGBT. Despite the growing social acceptance of users towards the community, homophobic and transphobic content in social media, such as hate speech, discrimination, or cyberbullying, are still prevalent in the present time [31]. Such acts targeted towards vulnerable individuals are deeply rooted in personal prejudice, religious beliefs, strict gender roles, and culture. In terms of neutrality, 26.015% or 7,776 tweets illustrate that there are users who opt to stay in the middle ground instead of actively choosing a certain side. Users who have ambiguous stances about LGBT rights commonly do not voice out their opinions about the matter to avoid offending anyone or simply do not want to involve themselves in the narratives of the community.

Analyzing the top ten dominant negative tweets is pertinent to determining the sources of negativity towards the LGBT community by understanding why prejudice and targeted hate speech exist. The top tweet that appeared in this category mentioned "pedophilia" as a slur to demean the members of the community by saying that they tolerate pedophilic acts. According to Nematy et al. [32], haters of persons with pedophilia (PWP) have a tendency to extend their hostility to other protected groups or individuals they perceive as different by displaying racist, ableist, and homophobic insults. The presence of slurs such as "pedophilia" can also be attributed to the Parental Rights in Education bill or the "Don't Say Gay or Trans" bill passed by the Florida Senate, which took effect on July 1, 2022. This law bans lectures on sexual orientation and gender identity in grade levels from kindergarten up to third grade. A month after the bill took effect, there has been an increase of 406% in the usage of slurs to oppose the LGBT community, as per a 2022 report conducted by the Human Rights Campaign and Center for Countering Digital Hate. Twitter users openly raised their opinions on this issue, which is evident in the fourth top negative tweet. The second leading tweet talks about their concern about the amount of violence in the form of hateful speech that queer individuals experience. The anonymity that social media platforms provide to users is one of the factors that greatly contributes to hate speech [33]. Twitter, in particular, became a great source of offensive remarks since users can freely post their opinions without exposing their real identities. The third top tweet focuses on the religious context. LGBT pride uses the rainbow as its symbol, angered some of the religious groups since it is also a symbol of the covenant of God with mankind as mentioned in the Old Testament scriptures. These groups who are conventionally against homosexuals view it as a mockery of a biblical symbol.

To have a deeper understanding of the sentiments of Twitter users on LGBT it is also important to take into perspective how it is being positively perceived. The top ten positive tweet suggests that LGBT individuals are open to searching for potential relationships in dating applications. In recent times, the prospects of dating or making friends through LGBT dating apps such as Grindr, OkCupid, and Badoo are widely used. Byron et al. [34] suggest that apps are virtual spaces that can provide a sense of security and community that is safer compared to finding queer intimacies in public spaces that can be prone to unwelcoming environments or hostility from other people. On the other hand, the second top tweet focuses on politics and that the user would vote for the best candidate no matter what gender they identify with. In a study by Everith & Horvath [35], a political candidate's gender identity or sexual orientation may not be a big deal since it is becoming evident that society is not significantly impacted by their preferences and openness with their identity. In addition, the public is gradually becoming more accepting and less threatened by candidates who are members of the LGBT community running for political positions. Most of the top tweets under this category are about LGBT characters or representation in forms of entertainment like television, film, and music. At the present time, there has been a gradual increase in the inclusion of LGBT characters in the aforementioned media forms as society becomes more accepting of queer individuals [36]. This implies that exposure and including queer characters in books, movies, and songs may eventually lead to the normalization of LGBT relationships in popular culture.

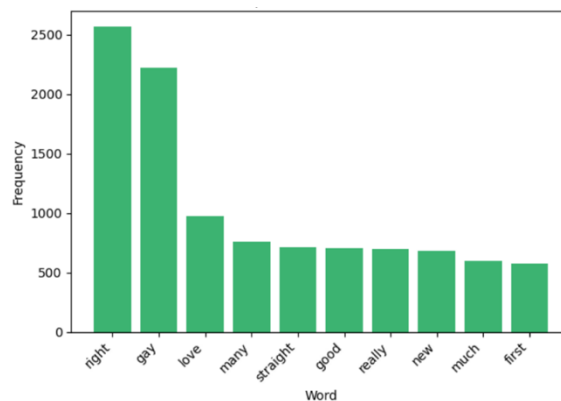


FIG 7. Top 10 positive words

The top positive word, as shown in Fig. 7, was “right” with 2,568 occurrences. Its prevalence can be attributed to the advocacies and movements concerning the rights of the LGBT community. Twitter parties were spearheaded using hashtags such as “#LGBTRights”, “#GayRights”, and “#EqualRights” to spread awareness and advocate for the discussion and acceptance of legal and social rights of queer identities. The second most dominant word is “gay” with 2,218 occurrences which is likely present in discourses tackling homosexuals. Furthermore, popular tweets and campaigns containing the word “gay” are utilized, like “#GayPride”, “#GayRights”, and “#GayLove”. The next positive word is “love” with 972 occurrences which is widely used, particularly during the celebration of Pride month and other significant LGBT events. It mainly talks about the concept of love being universal in the context of relationships and unions among LGBT members. Also, it is used to encourage people to spread kindness and positivity towards the members of the community. “Straight” is another noteworthy term that relates to the straight supporters showing their support and open-mindedness for LGBT rights.

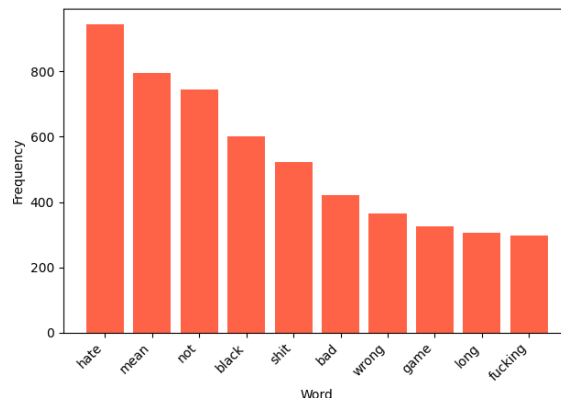


FIG 8. Top 10 negative words

In addition to looking at the context of the positive sentiments, it is important to look at the top 10 most negative words, as shown in Fig. 8, used in tweets in order to understand the challenges and negative opinions about the LGBT community. "Hate" is the first word to surface in this category with 944 occurrences and is commonly used in discussions on homophobia, transphobia, and other discriminatory behaviors reflecting hostility and negative sentiments against queers. "Mean" with 794 occurrences is the second word to appear in this category and can be used to describe hateful, offensive, and discriminatory language or behaviors faced by LGBT individuals. Moreover, words such as "not" and "wrong" reflect rejection, denial, or disagreement with issues relating to LGBT. Users may have utilized such words to express their opposition by deeming LGBT movements as "not supportive", "not accepted", "wrong to be gay", or "wrong behavior". The word "black" is also frequently used and could be related to discussions about the intersection of issues on gender and racism faced by black LGBT individuals.

4. CONCLUSIONS

This study examined global societal attitudes toward the LGBTQ community by analyzing a large collection of tweets through NLP and topic modeling techniques using the OSEMN methodology. The results give valuable insights into how individuals on Twitter show support, criticism, or neutrality toward LGBTQ issues, hence reflecting real-time public opinion across several cultural and social settings. One of the most significant outcomes of this research is the clear indication of growing global acceptance and support for LGBTQ individuals. Nearly half (47.7%) of the tweets in the dataset expressed positive sentiments, emphasizing themes of love, rights, pride, community, and solidarity. This change toward positivity indicates that the public discourse—especially on Twitter—is progressively turning into an outlet for validation and visibility. These sentiments help to normalize LGBTQ identities and support equal rights, hence helping to create inclusive communities. The study also showed that a significant percentage of tweets which is 26.25%, expressed negative feelings usually mixed with hate speech, religious condemnation, or false information. These results are just as important since they draw attention to the continuous difficulties LGBTQ people experience in online environments. Understanding the nature and origins of negative sentiment can inform targeted strategies to counteract stigma, discrimination, and online harassment. By identifying and mapping dominant themes, such as advocacy for sexual health, the tension between free speech and safe digital spaces, empowerment through representation, and the political struggle against prejudice, this research not only documents current social attitudes but also uncovers the underlying narratives that shape them. These insights can inform policymakers, educators, advocacy groups, and social media platforms in developing more effective communication strategies, educational campaigns, and protective measures.

For the LGBTQ community, the impact of this research is twofold. First, it validates the presence and visibility of LGBTQ voices online, showing that many users actively support and defend their rights. Second, it provides a roadmap for identifying where digital support is strong, where it is lacking, and where intervention is most needed. By shedding light on both the supportive and hostile corners of online discourse, this study empowers LGBTQ individuals and allies with knowledge they can use to advocate for safer, more inclusive online and offline spaces. Finally, this paper contributes to the growing body of work that seeks to understand how digital platforms shape and reflect societal values. It underscores the importance of continued vigilance, advocacy, and community-

building in the digital age, and offers a foundation for future research to explore evolving LGBTQ discourse across cultures and over time.

ACKNOWLEDGMENT

We would like to express our deepest gratitude to all those who contributed to the completion of this study. Special thanks go to our colleagues for their insightful guidance and feedback throughout the research process.

REFERENSI

- [1] Y. Cheng, X. Zhou, and K. Yao, "LGBT-Inclusive Representation in Entertainment Products and Its Market Response: Evidence from Field and Lab," *J. Bus. Ethics*, vol. 183, no. 4, pp. 1189–1209, Apr. 2023, doi: 10.1007/s10551-022-05075-4.
- [2] Aldinata, A. M. Soesanto, V. C. Chandra, and D. Suhartono, "Sentiments comparison on Twitter about LGBT," *Procedia Comput. Sci.*, vol. 216, pp. 765–773, Jan. 2023, doi: 10.1016/j.procs.2022.12.194.
- [3] M. B. Doğan, Oban, Volkan, and G. and Dikeç, "Qualitative and Artificial Intelligence-based Sentiment Analyses of Anti-LGBTI+ Hate Speech on Twitter in Turkey," *Issues Ment. Health Nurs.*, vol. 44, no. 2, pp. 112–120, Feb. 2023, doi: 10.1080/01612840.2022.2158407.
- [4] J. Kusuma, H. Hartono, and B. Hayadi, "Sentiment Analysis of IDAHOBIT Celebrations using Naïve Bayes and Decision Tree Algorithms," *Int. Conf. Inf. Sci. Technol. Innov. ICoSTEC*, vol. 2, pp. 52–58, Mar. 2023, doi: 10.35842/icostec.v2i1.37.
- [5] I.-S. Jeon, K. Shin-Yu, and K. Seong-Joo, "Developing Standards for Educational Datasets by School Level: A Framework for Sustainable K-12 Education," *Sustainability*, vol. 16, no. 12, pp. 4954–4954, 2024, doi: 10.3390/su16124954.
- [6] P. Fan *et al.*, "Unlocking the Potential of Competition-Based Learning: A Case Study of Kaggle in Big Data Analytics Education," presented at the 2024 10th International Conference on Smart Computing and Communication, ICSCC 2024, 2024, pp. 589–595. doi: 10.1109/ICSCC62041.2024.10690416.
- [7] A. I. Kabir, K. Ahmed, and R. Karim, "Word Cloud and Sentiment Analysis of Amazon Earphones Reviews with R Programming Language," *Inform. Econ.*, vol. 24, pp. 55–71, Dec. 2020, doi: 10.24818/issn14531305/24.4.2020.05.
- [8] C. P. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, no. 3, pp. 509–553, 2023, doi: 10.1017/S1351324922000213.
- [9] M. Kunilovskaya and A. Plum, "Text Preprocessing and its Implications in a Digital Humanities Project," presented at the International Conference Recent Advances in Natural Language Processing, RANLP, 2021, pp. 85–93. doi: 10.26615/issn.2603-2821.2021_013.
- [10] T. B. D. Lima *et al.*, "Sequence Labeling Algorithms for Punctuation Restoration in Brazilian Portuguese Texts," *Lect. Notes Comput. Sci. Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinforma.*, vol. 13654 LNAI, pp. 616–630, 2022, doi: 10.1007/978-3-031-21689-3_43.
- [11] R. J. Rupa, J. F. Sohana, and M. Rahman, "Comparative analysis of bengali stop word detection using different approaches," presented at the 2021 International Conference on Automation, Control and Mechatronics for Industry 4.0, ACMI 2021, 2021. doi: 10.1109/ACMI53878.2021.9528279.
- [12] A. A. Bichi, R. Samsudin, and R. Hassan, "Automatic construction of generic stop words list for hausa text," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 25, no. 3, pp. 1501–1507, 2022, doi: 10.11591/ijeecs.v25.i3.pp1501-1507.

- [13] R. Hafeez *et al.*, “Contextual Urdu Lemmatization Using Recurrent Neural Network Models,” *Mathematics*, vol. 11, no. 2, 2023, doi: 10.3390/math11020435.
- [14] C. Sharma, S. Sharma, and Sakshi, “Latent DIRICHLET allocation (LDA) based information modelling on BLOCKCHAIN technology: a review of trends and research patterns used in integration,” *Multimed. Tools Appl.*, vol. 81, no. 25, pp. 36805–36831, Oct. 2022, doi: 10.1007/s11042-022-13500-z.
- [15] S. Nazeer, P. Yayavaram, and L. J. Anbarasi, “Advanced Algorithmic Techniques for Topic Prediction and Recommendation—An Analysis,” *Lect. Notes Electr. Eng.*, vol. 967, pp. 125–132, 2023, doi: 10.1007/978-981-19-7169-3_12.
- [16] T. Y. Koonce *et al.*, “Using a Natural Language Processing Approach to Support Rapid Knowledge Acquisition,” *JMIR Med. Inform.*, vol. 12, no. 1, p. e53516, Jan. 2024, doi: 10.2196/53516.
- [17] E.-B. Braune *et al.*, “Identification of a Notch transcriptomic signature for breast cancer,” *Breast Cancer Res.*, vol. 26, no. 1, p. 4, Jan. 2024, doi: 10.1186/s13058-023-01757-7.
- [18] S. Kadariya *et al.*, “Community Organising Frameworks, Models, and Processes to Improve Health: A Systematic Scoping Review,” *Int. J. Environ. Res. Public Health*, vol. 20, no. 7, Art. no. 7, Jan. 2023, doi: 10.3390/ijerph20075341.
- [19] W. Ekezie *et al.*, “Health and social care experience and research perception of different ethnic minority populations in the East Midlands, United Kingdom (REPRESENT study),” *Health Expect.*, vol. 27, no. 1, p. e13944, 2024, doi: 10.1111/hex.13944.
- [20] Á. Díaz, “Online Racialization and the Myth of Colorblind Content Policy,” *Boston Univ. Law Rev.*, vol. 103, no. 7, pp. 1929–1983, Dec. 2023, Accessed: Mar. 27, 2025. [Online]. Available: <https://www.proquest.com/docview/2916711995/abstract/5D24FEE5FCD649C7PQ/1>
- [21] M. Kossowska *et al.*, “Internet-based micro-identities as a driver of societal disintegration,” *Humanit. Soc. Sci. Commun.*, vol. 10, no. 1, pp. 1–10, Dec. 2023, doi: 10.1057/s41599-023-02441-z.
- [22] D. A. Francis, “Queer and trans activism in South African schools: Non-Governmental Organisations (NGOs) disrupting LGBT+ inequalities,” *Teach. Teach. Educ.*, vol. 148, p. 104705, Oct. 2024, doi: 10.1016/j.tate.2024.104705.
- [23] M. Agopiantz *et al.*, “Assisted reproductive technology in France: The reproductive rights of LGBT people,” *J. Gynecol. Obstet. Hum. Reprod.*, vol. 52, no. 10, p. 102690, Dec. 2023, doi: 10.1016/j.jogoh.2023.102690.
- [24] M. Nedeljko and B. M. Kaučič, “Creating a safe and inclusive living environment for LGBTQ+ older adults: A review of research, policy and literature,” *IFAC-Pap.*, vol. 56, no. 2, pp. 9966–9971, Jan. 2023, doi: 10.1016/j.ifacol.2023.10.698.
- [25] J. Everitt and M. Tremblay, “Are Openly LGBTQ2+ the New Sacrificial Lambs? Campaign Contexts and the Gendered Implications for LGBTQ2+ Candidates,” *Can. J. Polit. Sci. Can. Sci. Polit.*, vol. 56, no. 2, pp. 300–324, Jun. 2023, doi: 10.1017/S0008423923000161.
- [26] G. Magni and A. Reynolds, “Voter Preferences and the Political Underrepresentation of Minority Groups: Lesbian, Gay, and Transgender Candidates in Advanced Democracies,” *J. Polit.*, vol. 83, Oct. 2020, doi: 10.1086/712142.
- [27] C. A. Blackington, “Polarizing Protests: Democratic Backsliding and Mobilization,” Ph.D., The University of North Carolina at Chapel Hill, United States -- North Carolina, 2024. Accessed: Mar. 27, 2025. [Online]. Available: <https://www.proquest.com/docview/3056960856/abstract/2B829C8413F44D76PQ/1>

- [28] S. Pandya and P. Mehta, *A Review On Sentiment Analysis Methodologies, Practices And Applications*. 2020.
- [29] R. C. H. Chan, “Benefits and risks of LGBT social media use for sexual and gender minority individuals: An investigation of psychosocial mechanisms of LGBT social media use and well-being,” *Comput. Hum. Behav.*, vol. 139, p. 107531, Feb. 2023, doi: 10.1016/j.chb.2022.107531.
- [30] S. A. Brandt and C. L. Carmichael, “Does online support matter? The relationship between online identity-related support, mattering, and well-being in sexual minority men,” *Comput. Hum. Behav.*, vol. 111, p. 106429, Oct. 2020, doi: 10.1016/j.chb.2020.106429.
- [31] B. R. Chakravarthi, “Detection of homophobia and transphobia in YouTube comments,” *Int. J. Data Sci. Anal.*, vol. 18, no. 1, pp. 49–68, Jun. 2024, doi: 10.1007/s41060-023-00400-0.
- [32] A. Nematy, S. Flynn, and S. McCarthy-Jones, “YouTube Commenters’ Discourse of Paedophilia: A Qualitative Social Media Analysis,” *Sex. Cult.*, vol. 28, no. 1, pp. 71–95, Feb. 2024, doi: 10.1007/s12119-023-10117-8.
- [33] C. Macias, M. Soto, T. Alcántara, and H. Calvo, “Impact of Text Preprocessing and Feature Selection on Hate Speech Detection in Online Messages Towards the LGBTQ+ Community in Mexico,” presented at the IberLEF@SEPLN, 2023. Accessed: Mar. 27, 2025. [Online]. Available: <https://www.semanticscholar.org/paper/Impact-of-Text-Preprocessing-and-Feature-Selection-Macias-Soto/964d9b807bff6f7ed4d0c4f8f2ced8f18862a95a>
- [34] P. Byron, K. Albury, and T. Pym, “Hooking up with friends: LGBTQ+ young people, dating apps, friendship and safety,” *Media Cult. Soc.*, vol. 43, no. 3, pp. 497–514, Apr. 2021, doi: 10.1177/0163443720972312.
- [35] J. Everitt and L. Horvath, “Public Attitudes and Private Prejudices: Assessing Voters’ Willingness to Vote for Out Lesbian and Gay Candidates,” *Front. Polit. Sci.*, vol. 3, Jun. 2021, doi: 10.3389/fpos.2021.662095.
- [36] K. Thomson, “An Analysis of LGBTQ Representation in Television and Film,” *Bridg. Undergrad. J. Contemp. Connect.*, vol. 5, no. 1, Jul. 2021, [Online]. Available: https://scholars.wlu.ca/bridges_contemporary_connections/vol5/iss1/7