Jurnal Komunikasi: Malaysian Journal of Communication
https://ejournal.ukm.my/mjc
Volume 41(4) 2025: 515-529

A Novel Cross-Audience Analysis for Multi-Shared Content on Social Media

HUSSEIN Y. HAZIMEH*
Al Maaref University, Lebanon

FATIMA MAZLOUM
Lebanese University, Lebanon

ABSTRACT

Social media has gained popularity due to the existing technologies and advancements in internet and smartphone technologies. Recently, various regions have experienced different types of conflicts and wars. People in affected regions need real-time updates on these events. As it is mentioned earlier, social media has become a principal resource for such content. Multiple channels can be created across different platforms. For instance, a certain media organisation is capable of creating an account on different social media channels in a simple way. Users connected to these channels receive the same content but in a different structure and from different platforms. In this context, repetitive and redundant content delivered to users is necessary to deliver important content and check for the right one. This study aims to compare a set of metrics that highlights the interconnection between the audience community and the multiple accounts of media organisations that exist on social media. In particular, this study develops a specific framework that can detect the overlapping between interactive users across a media organisation channel on its pages on different social media platforms. In addition, this paper compares the sentiment analysis on these contents as well as the interactivity level and finally detects the hate speech of these contents. To address its proposition, this paper involved two popular social media platforms that are Facebook and X (Twitter). The results show interesting facts when comparing these metrics on the same posts.

Keywords: *Audience analysis, social media, artificial intelligence, hate speech, source of information.*

INTRODUCTION

With the continuous occurrence of global events and the rise of modern technology and artificial intelligence, people increasingly need efficient ways to stay updated on real-world events. Since several decades ago, the world, with its audience, has been receiving information and news via typical media mediums, such as the radio, the television, and the newspaper. However, with the revolution in the information technology sector after 2000, people and companies have been flooded with information technology products, mainly software, hardware, and services. These include, but are not limited to, computing architectures, smartphones, internet connections, internetworking applications, emails, and more. The usage and incorporation of these software and hardware advancements are a necessity for people to communicate and for companies to evolve. In the last five years, the world has been subjected to enormous events that have affected millions of people in different countries.

*Corresponding author: hussein.hazimeh@mu.edu.lb
E-ISSN: 2289-1528
https://doi.org/10.17576/JKMJC-2025-4104-28
Received: 3 August 2025 | Accepted: 12 November 2025 | Published: 12 December 2025

These events are categorised into political, conflict, social, and health domains. Since COVID-19, people found themselves in need of being connected to systems that deliver continuous, reliable, and informative updates about this event. Smartphones and social media were among the top popular platforms that have validated their proficiency in this event (Bao et al., 2020; Neely et al. 2021; Saima et al., 2022).

Actually, social media is not considered as an original source of information. Social media content is published by both individuals and organizations that provide event updates, subscribe and follow the accounts of the organisations that publish event-related content and news. In the last two years, a lot of social media platforms have emerged, in addition to the existing platforms, such as telegram, WhatsApp channels and X-platform (Twitter). Organisations found themselves obliged to create accounts on all these platforms to match and reach every user in their target community. However, despite the variety of social media channels, the content published remains the same and the content delivered to the users also remains the same. The users found themselves receiving the same information from the same resources, however from different channels. This study aims to apply natural language processing tasks to fulfil three objectives: (1) Find the overlapping between social media users. (2) Detect the hate speech between two popular social media platforms. (3) Detect the hate speech and compare it on both social media platforms. To address its objective, the data collected from two main leading social media platforms Facebook and X-platform. To prove the validity of its approach, the paper collected 5,000 posts from both platforms. For every post, the paper retrieves its comments, and reactions such as likes, and user profile identities and tracks them, comparing the overlapping, the sentiment, the hate speech and the interactivity metrics. These metrics will reflect the behaviour of the community against the social media platform itself, which will give meaningful insights for media organisations to understand the behaviour of their audience on both platforms. This results in better management of content publishing and will enhance the decision-making regarding publishing content on these platforms. The main contributions of this paper are summarized as follows:

(1) Creating a dataset of 5000 posts collected from Facebook and X. The posts are the same and published by the same account.
(2) Applying an exploratory data analysis task to the collected data.
(3) Investigating the overlapping between users' identities on Facebook and X platform
(4) Detecting hate speech and comparing it between Facebook and X.
(5) Analysing the sentiment between the same posts on different channels.

This paper will be structured as follows. Section one has covered the introduction. In section two, several studies listed in the literature. In section three, the methodology explained behind this work. Section four presented the main results of this paper. In section five, a discussion presented on the potential future studies and the conclusion of this work.

LITERATURE REVIEW

This section identifies a list of several research papers that proposed approaches and systems for content analytics on social media platforms. Mainly, the paper considers approaches that investigate Facebook and X platforms for content similarity and dissimilarity. In addition, it considers approaches that analyse content on both Facebook and X platforms to reveal certain metrics such as sentiment analysis, hate speech detection, and other problems.

In Yu et al. (2021), the authors analysed altmetrics data on Facebook and Twitter. They compared in detail the accuracy metric of altmetrics data on both platforms. The results of this study revealed that there are seven different types of errors with respect to the accuracy metric of altmetrics on both Facebook and Twitter. In another research (Al-Mashhadani et al., 2022), the authors have analysed the sentiment on Facebook and X (Twitter) platforms. The main contribution of their paper was to leverage big data solutions to analyse large-scale sentiment on these platforms. The same authors were able to detect sentiment polarity on social media comments and get highly accurate results.

Saroj and Pal (2023) analysed disseminated information engaged with stakeholders and feedback on government initiatives on both Facebook and X (Twitter) where they aimed to compare these metrics on ministers from the Indian government and track them across three years in order to detect their results. However, Sörensen et al. (2023) compared in a study published at the University of Zurich, Switzerland, among content published on Facebook, Twitter and Instagram regarding higher educational institutions. The aim of their study was to reveal the communication increase and decrease on all these platforms. The authors have discovered that certain universities and educational institutions are gaining more popularity on Instagram, however other universities are gaining more popularity on Twitter and Facebook. Notably, the authors have collected data from these platforms across 17 years, resulting in 330,000 posts from 200 accounts (Sörensen et al., 2023).

Some studies, such as Akobiarek (2024), have explored multi-purpose social media analytics with limited emphasis on media content, focusing instead on specific regions like Indonesia (Subekti, 2024). These approaches did not utilize algorithmic methods; rather, they relied on manual techniques such as surveys, interviews, and observations. Data analysis was conducted using basic statistical and manual tools. While these methods are valid, they lack scalability for handling large datasets—an issue this paper aims to address.

Building on this, several researchers have investigated cross-platform behaviour using user overlap, content similarity, and engagement metrics. For example, Anisa et al. (2024) examined the effectiveness of health seminar promotion on Instagram and found that specific stylistic and linguistic features significantly influenced engagement. Although focused on Instagram, the methods of content classification and user behaviour modelling offer transferable insights for studies comparing Facebook and X. Similarly, Aichner et al. (2021) provided a comprehensive review of social media definitions and applications over a 25-year period, demonstrating how user expectations and engagement behaviours have evolved with the platforms.

"The rapid dissemination of information on social media facilitates the propagation of hate speech, which can reach a vast audience with minimal effort" (Amalia, Harani, & Prianto, 2024, p. 339). In the context of hate speech detection, Vidgen et al. (2020) contributed a dynamic dataset generation method that adapts to new forms of online abuse. Their methodology emphasizes robustness in capturing evolving toxic language and is particularly relevant to platforms like Twitter, where anonymity and brevity foster impulsive and sometimes harmful interactions. Their work demonstrated how pre-trained models like BERT can be fine-tuned to capture nuanced hate speech forms, which this paper adopts in its implementation.

Similarly, Sachdeva et al. (2022) proposed a multi-label classification model trained on data from YouTube, Reddit, and Twitter. Their approach tackled the ambiguity in labelling hate speech by applying Rasch measurement theory to account for annotator disagreement. This reflects a broader trend in natural language processing (NLP) where datasets are increasingly evaluated through the lens of human perception and subjectivity—an issue that this paper circumvents by benchmarking multiple models on standardized criteria.

Another line of research worth mentioning is focused on platform-specific audience behaviour. For instance, studies have found that Twitter users are often more politically engaged and aggressive in tone compared to Facebook users, who may be more community-oriented and less confrontational. This aligns with the findings in this paper, where sentiment analysis shows a higher proportion of neutral and negative sentiment on X, while Facebook interactions skew more positively. This reinforces the notion that platform design and community norms significantly shape user-generated content.

Further, Lomborg and Bechmann (2014) critically discussed the use of APIs for collecting data from social media. As these authors indicate, data access restrictions have limited researchers' ability to conduct large-scale analytics, especially after 2018 when many social platforms tightened their API usage policies. This justifies the methodology in the current study, which uses advanced web scraping with proxy rotation to bypass such limitations. Their work is foundational in emphasizing ethical considerations, which are also respected in this paper's design.

Cross-audience analysis has also received attention in political and crisis communication research. For example, during COVID-19, governments and media agencies distributed similar content across multiple platforms to ensure message consistency and coverage. However, user reception, amplification, and feedback varied widely by platform, suggesting that content virality and engagement are highly context dependent. This supports the paper's hypothesis that multi-platform distribution does not ensure homogenous reception.

In terms of computational models, researchers increasingly employ ensemble methods and hybrid frameworks to analyse sentiment and hate speech. This paper aligns with this trend by incorporating multiple pre-trained models and comparing their output to derive more accurate conclusions. The usage of Jaccard similarity for overlapping audience detection is also well-supported in the literature, especially in studies on community detection, recommendation systems, and social graph analysis.

In summary, the reviewed literature highlights the need for robust, scalable, and ethically sound frameworks for analysing cross-platform content on social media. Although existing studies offer valuable insights into sentiment analysis, hate speech detection, and content engagement metrics, few have tackled the problem from a cross-audience, multi-platform perspective. By integrating web scraping, NLP, and audience overlap metrics, this study fills a noticeable gap in the field and provides a replicable methodology for media organizations seeking deeper audience insights.

E-ISSN: 2289-1528
https://doi.org/10.17576/JKMJC-2025-4104-28

METHODOLOGY

This section introduces the methodology for this paper. Essentially, it developed a complete framework to address the objectives listed in this paper. The framework comprises universal components that enable the development of an end-to-end framework capable of fulfilling the objectives addressed in this paper. This framework is completely automated, which means there is no human intervention required to develop the solution. It has leveraged and utilised existing data analytics and machine learning solutions to develop components in the paper's framework. While addressing the application of the framework, the paper will list the challenges that encountered its authors during the implementation and configuration of every component in their methodology.

Currently, there are a lot of open-source tools that help address social media analytics tasks. However, in this particular task, the leading challenge is to select the corresponding and highly-performance tools among the existing ones. The framework prepared will explain how the solutions were selected in order to create a highly capable, reliable, and accurate solution.

The initial task in the methodology is data collection. Collecting data from the web is becoming challenging according to the barriers set up by the owners of websites and web systems. Social media is considered large-scale websites that comprise vast amounts of data that can be employed in many tasks in multiple domains (Aichner et al., 2021). The data on social media has been open for access and analytics until 2018 (Aichner et al., 2021). Most of the social media websites have limited access to their data and cancelled the existing APIs (Parry, 2024). An API is an application programming interface that provides a service for users to request data from the main website, and APIs were open to the public (Lomborg & Bechmann, 2014). However, when the data becomes big, the social media owners stop providing this service to interested users as well as companies. However, thanks to data scraping technologies developed, highly reliable web scrapers can extract any public data from any website including social media. These web scrapers are becoming very popular within the data science and artificial intelligence community. They are becoming adopted large-scale real-world projects that need data, especially from social media. In finding a solution for data collection, this paper decided to benefit from the power of the existing web scrapers to collect public data from the required social media accounts. Nevertheless, these web scrapers have a major limitation which is bypassing the CAPTCHA or human validation step required by a certain website. In order to avoid this problem, as it has been mentioned, this paper decided to apply its research on Facebook and X (Twitter) platforms. So far, these two platforms have not applied human validation steps during browsing activities even by humans or by web scrapers.

Despite the robustness and efficiency of existing web scrapers as solutions for collecting data from social media and the web, these scrapers still suffer from several shortcomings, mainly related to the detection of the social media platform itself, that the agent extracting and collecting the data is robot-based and not a human. And web scrapers fall under the category of robot-based web interactions. Despite this problem, thanks to proxy servers that provide promising solutions to fix this issue, anti-scraping restrictions using proxies are widely adopted in web scraping. A proxy allows you to change IP addresses and send requests as if they were coming from different users. These proxies have three key types. (1) residential proxies: IP is assigned to real users, difficult to detect but expensive. (2): datacenter proxies: are cheaper but more easily blocked. (3) rotating proxies: automatically change IP after each request.

To select the best option, the paper used ScrapeOps with rotation, which allows the authors in this particular project to collect data efficiently. In this case, the paper has solved the issue of anti-scraping restrictions suffered by websites such as social media platforms.

In the second phase, the data collected from social media accounts, mainly posts, will be entered into a pre-processing engine. In this phase, this paper applies a set of necessary pre-processing steps that switch the data into high-quality and clean data. Mainly, the pre-processing steps addressed in this framework are as follows; (1), removing stopwords. (2), ignoring anonymous accounts. (3), skipping replies. (4), data storage. For stopwords, this paper removes all stopwords, such as propositions, from natural text, including posts and comments, in order to enhance the part that analyses the data. For anonymous accounts, some of the users on social media who interact with posts are anonymous. These users will never be involved or participate in any analytic task. For skipping replies, in this work, this paper decided to use single comments during the analytics. Replies often white research issues that are concerned with discussion across people and interactivity with them, which will not be addressed in this work. Finally, all pre-processed data will be stored in a structured relational database management system. Storing the data will simplify retrieving it and referring it for analysis.

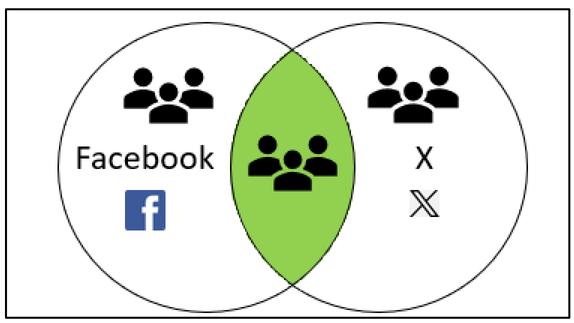

Figure 1: Intersection (overlapping between two social media channels) (author illustration)

a. Overlapping Analysis

To calculate the overlapping (Figure 1) index between two social networks, this paper implemented the Jaccard index similarity. This similarity measures the intersection across two given sets. In the case of this paper, the two sets are Facebook and X (Twitter) platforms. As it has been discussed, for a given page, certain users interact with its posts in terms of likes, reactions, comments, and shares. For pages' accounts on all social media platforms, only users do the interaction. To investigate the shared users across these pages, this paper has collected the users on every platform and then applied Jaccard to find the similarity score between the two platforms. According to Equation 1, Jaccard finds the intersection between the two social networks in terms of users and divides them into the union. This will return a similarity or overlapping score normalised between 0 and 1.

$$J(F,X) = \frac{|F \cap X|}{|F \cup X|}$$ Equation 1 Jaccard coefficient

b. Hate Speech Analysis

To detect hate speech and user-generated content on most platforms, this paper must use models that are built on machine learning and deep learning. In this regard, this paper has not created a model from scratch, which requires a lot of work and data availability. To work in a high-performance and easy manner, this paper uses all existing models from the literature that have high performance on hate speech online detection. Vidgen et al. (2020) prepare one of the top models used in the state-of-the-art for hate speech detection. This model was created in 2020 and tested on 40,000 entries. It is capable of detecting hate speech of different types, mainly *derogation*, *animosity*, *threatening* language, support for hateful entities, and *dehumanisation*. In this vein, this paper targeted three types of hate speech on social media. The first one is animosity, which detects the content that contains expressions of abuse against a group implicitly or subtly. The second type is threatening language.

This one is used to find hate speech that contains expressions, support for or encourages inflicting harm on a group or identifying perversions of a group. And the final category is dehumanisation. In this one, hate speech detects content which it perceives or treats people as less than human, widely animals, or involves describing groups as leeches, germs, or insects. According to the results of this model, the accuracy has reached 91.17% maximum, which is considered excellent for these types of natural language processing tasks.

In order to well evaluate and apply its approach for hate speech detection, this paper also tested another model for benchmarking purposes. This model named hate speech multi-label classification, published in 2022, was also created for detecting hate speech on larger scale online data (Sachdeva et al., 2022). The data set of this model is tested on 50,000 social media comments collected from YouTube, Reddit, and Twitter. However, this model has less advantages over the first one. In terms of types of hate speech, it has more types compared to the first one. However, this model has not released the accuracy results, which makes it under doubt for implementing real-world systems. However, in order to decide the model that is better for its scenario, this paper has also tested this one on its data. The results show, as it will be discussed in the Results section, that this model has lower performance than the first one, despite its several hate types available.

c. Sentiment Analysis

Finally, in this methodology, it addresses the problem of automatically finding the sentiment of the comments published by users on posts of the media channel. The purpose of this study is to compare the sentiment scores and polarities on the same posts published on both platforms and compare the overlapping and similarity between these sentiment polarities. Sentiment analysis is an essential task in natural language processing. It is used to find the sentiment of a given text as a polarity. The polarity is the tone of the feeling detected from the written and natural language content. This polarity has different values, negative, neutral, and positive. Negative refers to the disagreement or bad feelings against this particular post. Neutral means that the comment written by a certain user is neither with nor against the shared content. Finally, positively reflects the total agreement and satisfaction of users against the comments and posts published by the media organisation. In order to calculate the sentiment polarity and find it, this paper uses existing NLD models that are well-performed in analysing and finding the sentiment efficiently.

Sentiment is different from hate speech detection. Hate speech is orientated on finding particular abuse and hate content predefined and must be detected in users' comments to study the percentage and existence of this hate conversation published on social media. However, sentiment is more orientated towards finding the feelings of the user who wrote a comment on a certain post, where hate is always something negative and abusive. However, the sentiment is not the same. To analyse the sentiment as required, this paper collected a sample of 150 posts from a social media channel profile, where posts considered in this examination are the same. However, of course, the comments are different. Thus, it started by finding the sentiment of every comment. Every comment is treated as an input, and the polarity of this comment will be used as an output. The total polarities encountered in this part are the summation of the total comments of every post. As seen in Equation 2, the total comments are collected per platform, Facebook or X.

$$Total\ comments\ (F \mid X) = \sum_{i=1}^{100} \quad C_i^{F,X} \textit{ Equation 2 Total polarities computation}$$

RESULTS

This section represents the results of this work. The main results are categorised as follows. First of all, the paper presented the total likes and shares interactivity across both channels on Facebook and X (Twitter) platforms. Then, the paper displayed the average likes and comments per day and a calendar-wise heatmap representation for both channels, also on both platforms. Then, the paper tracked these values and presented them in the histogram line chart in order to recognise the similarity and dissimilarity across these values. After that, this paper compares the total likes and the overlapping links between Facebook and Twitter in terms of likes and shares for both Channel 1 and Channel 2.

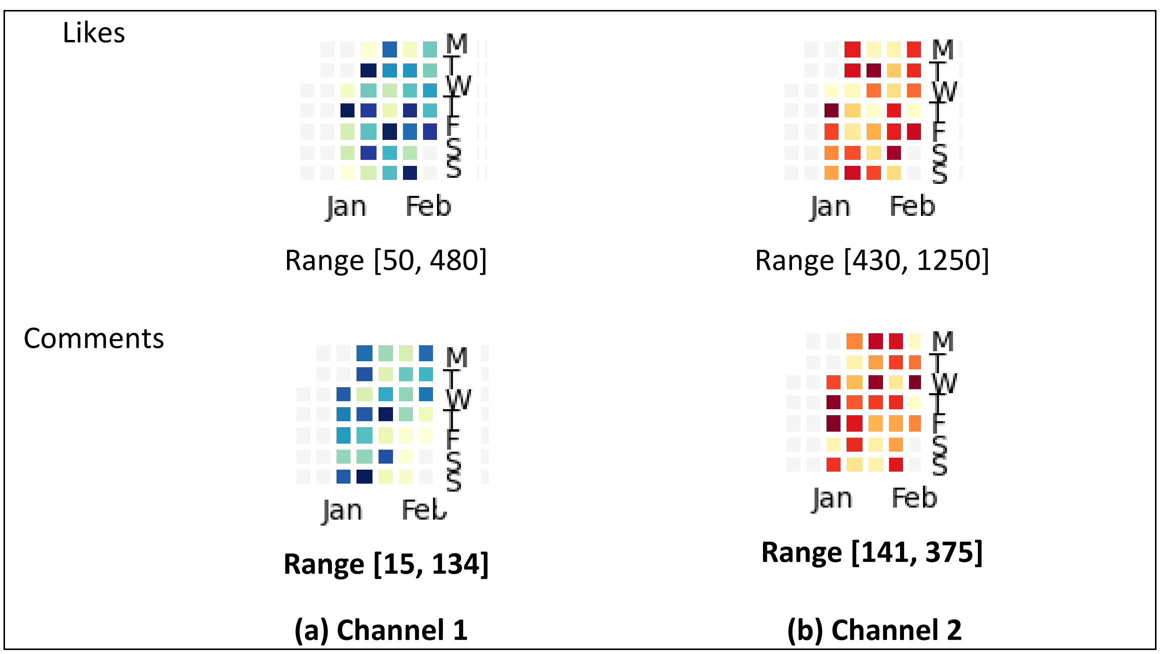

Figure 2: Total likes per day during Jan and Feb on X platform for Channel 1 and Channel 2

Figure 2 shows the results for total likes per day during January and February months on the X platform for Channels 1 and 2. As it can be observed in the table for likes the range of total likes per day was between a minimum of 50 likes and a maximum of 480 likes for Channel 1. For Channel 2 the range was higher where the minimum was 430 likes, and the maximum was 1250 likes. As it can be observed on some days in the heat map calendar, the dark values dominate more than the light values.

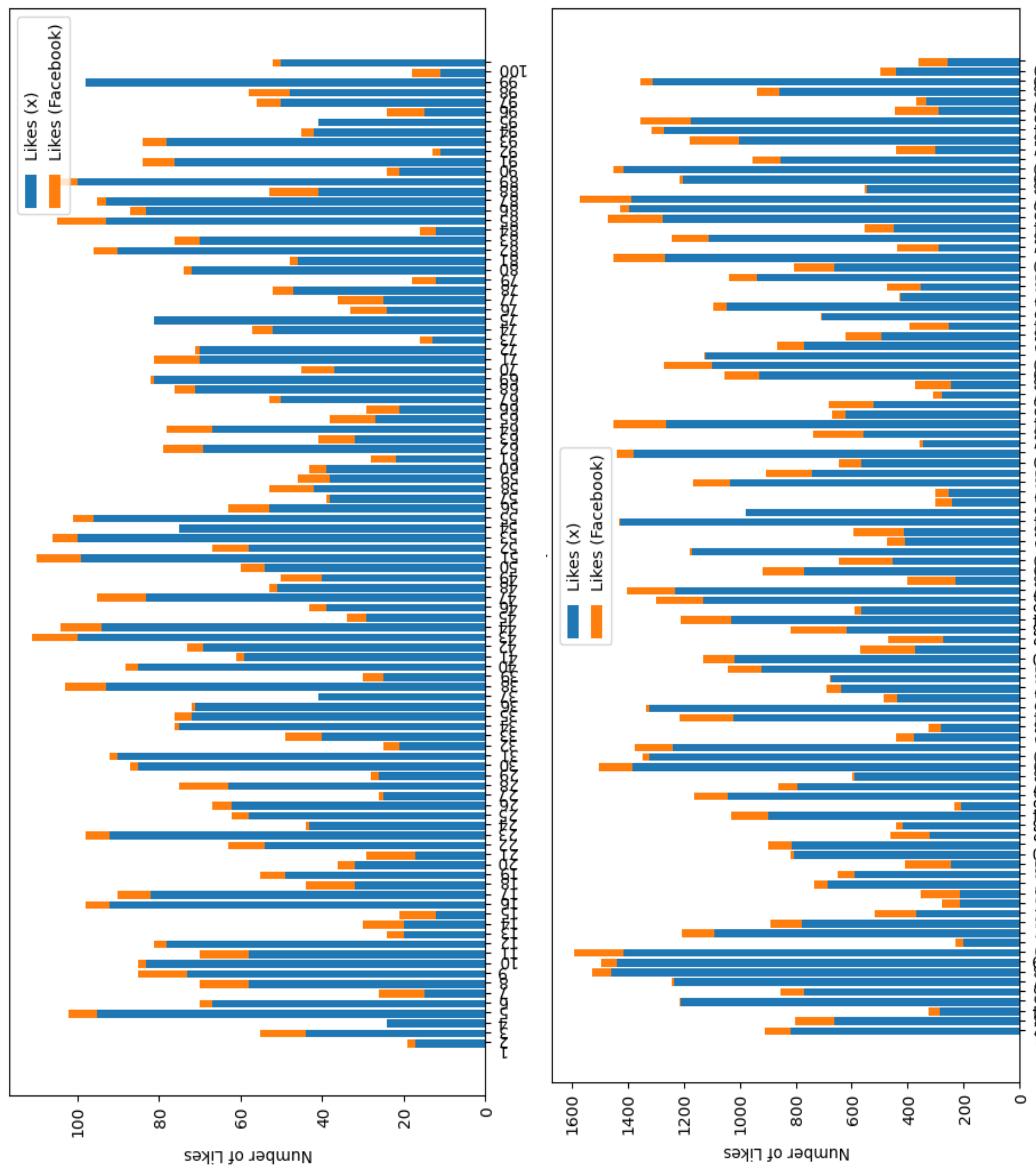

Figure 3: Likes and comments distribution over posts for both Facebook (right) and X(left)

In its analytics, this paper has detected that when users are highly interactive against posts published by the page an event happens on this day. On the other hand, for the comments, it can be observed that for Channel 1, the range was between 15 and 134 whereas for Channel 2 the range was between 141 and 375. The values for comments are dramatically decreased compared to likes and this is normal as a feature on social media where the number of comments is usually lower than the number of likes. However, in comparing the two channels with each other it can be observed that Channel 1 is less than Channel 2. However, on the other hand, the total followers of Channel 1 are 9 million on Facebook whereas the total followers of Channel 2 are 4.8 million on Facebook. This indicates that there is no relationship between the number of likes and the number of comments in comparison with the total followers of the Facebook page itself. In conclusion, user interactivity does not depend on the total followers by the page, and this is very important to address every page open on Facebook or on X platform boosting the page to gain a lot of followers is not a successful strategy where the most important is to get more interaction to the posts themselves.

The figures above will display the results comparing likes and share values across Facebook and X platforms on Channels 1 and 2. This indicator has been tracked across 100 posts on different dates. In part A, it can be observed that the comments for Channel 1 were higher than those of Channel 2. This means that users interact with Channel 1 more than Channel 2, despite having more followers on this channel, which means that likes are independently affected by the number of followers of the channel itself. Similarly, for Figure 2, the likes percentage is also higher on Channel 1 compared to Channel 2, independent of the number of followers. However, comparing likes and comments with each other, it can be observed that comments are lower than likes on both platforms.

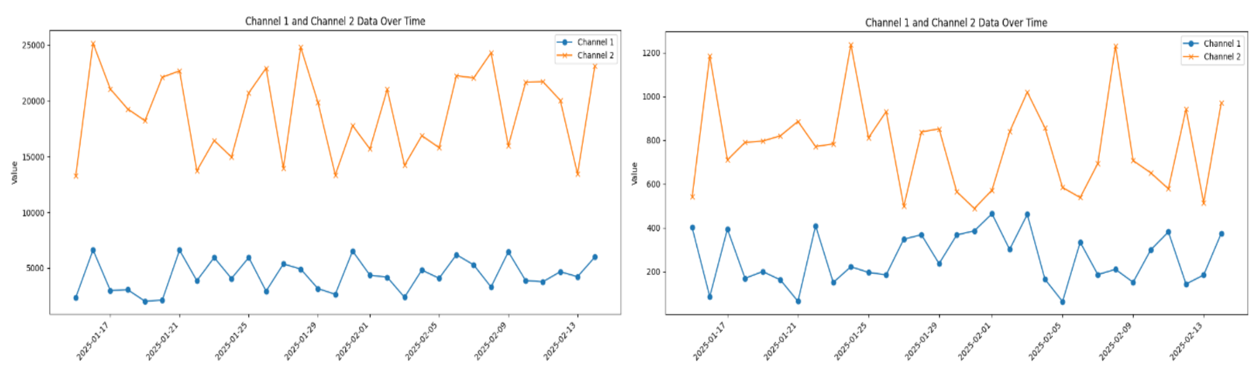

Figure 4: Total likes and comments on Facebook and X platforms for both Channels 1 and 2.

The results in Figure 4 displays the total likes on Facebook and Twitter or Channels 1 and 2. In Figure A, the analytics shows the likes interaction every day during February and January on the X platform for Channels 1 and 2. You can see here that Channel 1 has a much lower likes total per day compared to Channel 2. However, in Figure 4, there are the total commands published on every post on each day on the X platform for Channel 1 and Channel 2. Comparing likes and commands metrics, it can be observed that for different pages, users' likes have higher variance compared to commands that have lower variance. So, users usually have deviated values for the likes metric in comparison with the comments metric. Despite this high difference in the likes factor, sometimes commands could be close to each other. In conclusion, it can be deduced that likes and commands are not highly correlated to each other. So, a higher number of likes does not mean a higher number of commands and vice versa.

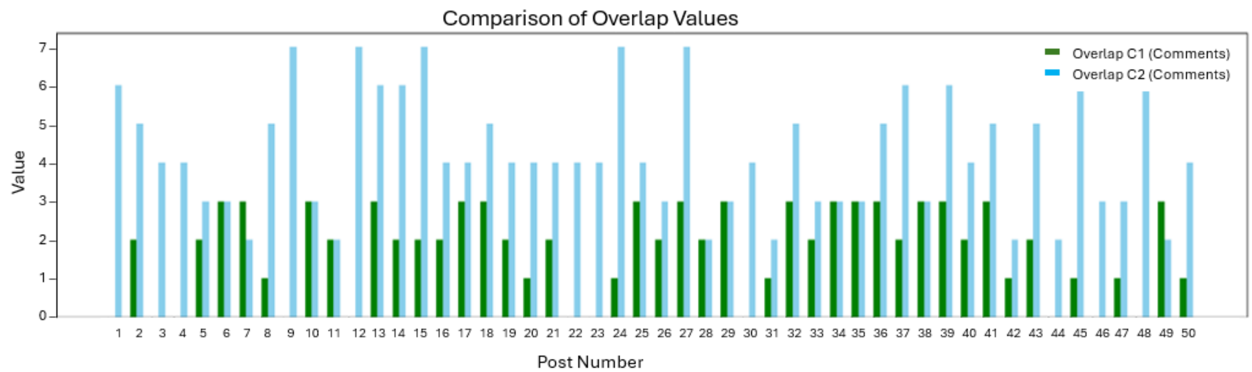

Figure 5: Overlapping index for comments published on Facebook and X platforms for Channel 1 and Channel 2

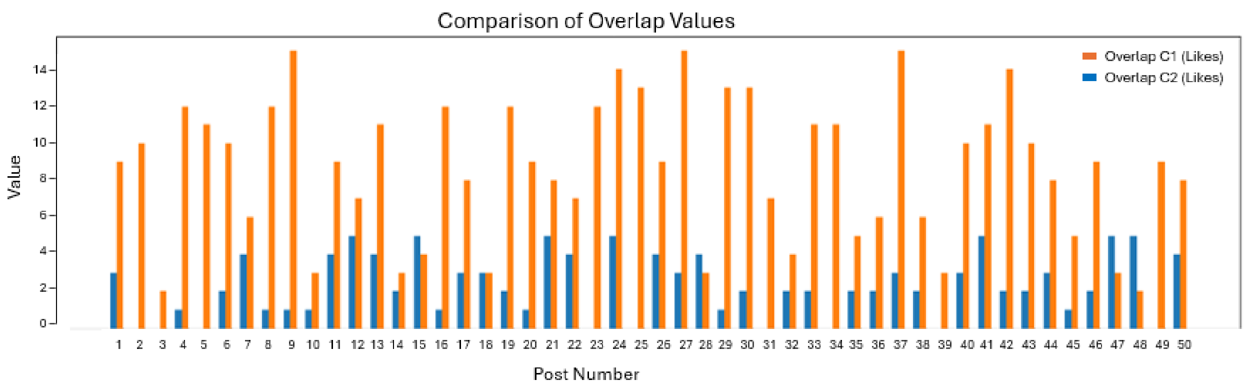

Figure 6: Overlapping index for likes published on Facebook and X platforms for Channel 1 and Channel 2

Figure 5 and Figure 6 illustrate the overlapping index between the two networks, Facebook and X. The overlapping index represents the common users that intersect between Facebook and X platforms on the same posts. For instance, there is a post published on the Facebook page by Channel 1 and the same post is published on X platform also by Channel 1. Here it might have overlapping users across comments and likes. In order to study this overlapping factor, this paper analyses the number of posts published across both platforms on both channels.

Figure 5 and Figure 6 illustrate the first 50 overlapping results. As observed in Figure 5, there is an overlap between comments on Facebook and X platforms. However, for Channel 2, the overlapping was higher than Channel 1. This is because interaction on this channel is higher compared to the other one. On the other hand, in Figure 6, it can also be observed that the overlap is higher in the case of Channel 2. However, in comparison between likes and comments overlapping percentages, it can be concluded that overlapping between likes is higher than overlapping between comments. In conclusion, users have higher overlapping percentages between Facebook and X's platforms given the likes metric compared to the comments metric.

Table 1: The results of the hate-speech detection model

Metric	**X (Twitter)**	**Facebook**
Total Comments Analyzed	5,000	5,000
Threatening (%)	18% (900 comments)	12% (600 comments)
Dehumanization (%)	10% (500 comments)	7% (350 comments)
Animosity (%)	8% (400 comments)	5% (250 comments)
Most Targeted Groups	Political content, fashion religion, crisis	Political content, fashion religion, crisis
Top Hate Speech Categories	Insults, Threats, Misinformation	Stereotyping, Harassment, Defamation

To find the results of sentinel analysis, it is significant to find first all the polarities per every comment, on every post, and their posts published on both platforms. As it is observed in Figure 7, the three polarities considered in this study are positive, neutral and negative.

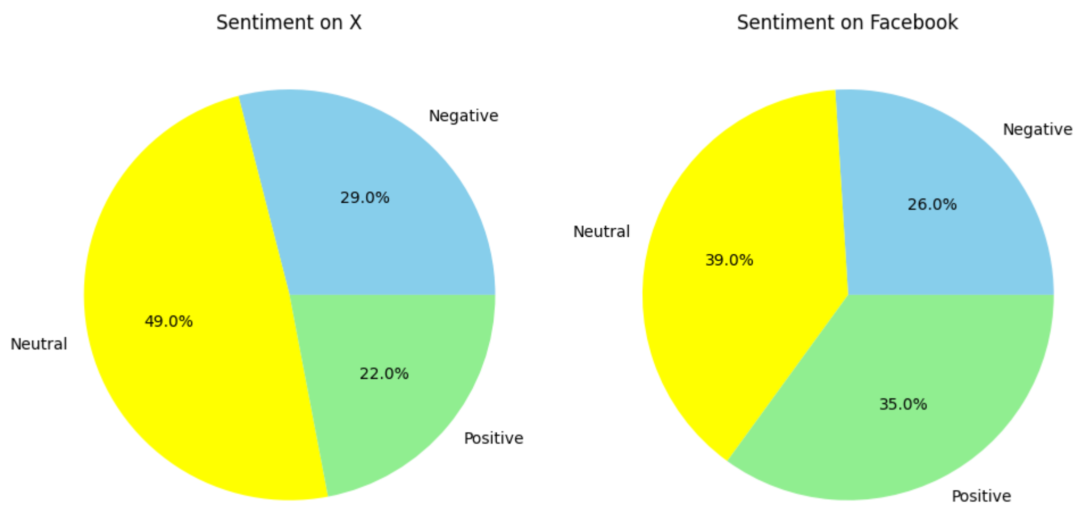

Figure 7: Sentiment polarities compared between Facebook and X platforms for 100 posts.

According to the analytics for this paper, the total percentage of positive sentiments on the X platform is lower than the total percentage of positive sentiment on Facebook platform. This indicates that users on Facebook are absolutely different from the users on Twitter. Second, users' feelings on Facebook also are different from X. For negative sentiment, the negative sentiment on X platform was higher than Facebook platform. On the other hand, neutral sentiment on X platform is roughly 50% of the total sentiment count. However, on Facebook, neutral polarity is roughly 40%. It means comments published on X (Twitter) platform, according to the analytics, tend to be more neutral. However, on Facebook, comments tend to be more positive and negative compared to neutral.

CONCLUSION

This paper proposed a novel system for comparing data published by media organisations on two popular social media platforms, Facebook and X platform, Twitter before, to address this proposition. It handled several necessary tasks and representatives of investigation, for investigating their comparison between data on social media published by the same media organisation. The paper defined three metrics. The first metric is the overlapping score. It is used to detect the percentage of shared and similar users that interact with content published by the page itself in two different places. This has gone by using the Jaccard coefficient metric. The second one is the hate speech percentage between Facebook and X. To detect hate speech on data that exists online in terms of comments and published on social media in particular, there are a lot of existing models, open source, that are based on deep learning solutions. These models have proven their proficiency in online data. To better test these models, the paper selected two well-known models, which are Bert and Roberta, and utilised the one that has higher performance on selected hate speech types. In this paper, the hate speech types handled threatening, dehumanisation, and animosity. The third metric studies the descriptive analytics part. In this part, the paper presented the analytics between features of both social media platforms. These analytics have been applied to data collected from content published by predefined pages on both platforms. The analytics have been done to understand the interactivity between the two pages. Consequently, being able to know which platform users are more active, will aid in decision-making, future posts, and user targeting strategies.

BIODATA

Hussein Y. Hazimeh is an Associate Professor of journalism and digital media at Al Maaref University. He finished his PhD in 2019 in Computer and Data Science from the University of Fribourg, Switzerland. His research interests include Artificial Intelligence and Social Media Data Analytics. Email: hussein.hazimeh@mu.edu.lb

Fatima Mazloum is a Master 2 Research student in Data Science at the Lebanese University, Faculty of Sciences, Beirut, Lebanon. She finished her Bachelor's in Data Science at the same university in 2022. Her research interests include Machine Learning, Information Retrieval, and RAG. Email: fatima.malzoum.4@st.ul.edu.lb

REFERENCES

Aichner, T., Grunfelder, M., Maurer, O., & Jegeni, D. (2021). Twenty-five years of social media: A review of social media applications and definitions from 1994 to 2019. *Cyberpsychology, Behavior, and Social Networking, 24*(4), 215-222. https://doi.org/ghknq2

Akobiarek, G. C. (2024). Patterns of social media choice and use and their impacts on online political participation in Sarawak. *Jurnal Komunikasi: Malaysian Journal of Communication, 40*(4), 1-18. https://doi.org/10.17576/JKMJC-2024-4004-01

Al-Mashhadani, M. I., Hussein, K. M., & Khudir, E. T. (2022). Sentiment analysis using optimized feature sets in different Facebook/Twitter dataset domains using big data. *Iraqi Journal for Computer Science and Mathematics, 3*(1), 64-70.

Amalia, F. R., Harani, N. H., & Prianto, C. (2024). Sentiment analysis of hate speech against presidential candidates of the Republic of Indonesia in the 2024 election using BERT. *Jurnal Sistem Cerdas, 7*(3), 339-355.

Anisa, N., Ariana, P. M., & Ramadhani, A. (2024). Digital marketing strategy through social media: Analysis of the effectiveness of health seminar promotion on Instagram account @Indonesianmedicalcenter. *Jurnal Komunikasi: Malaysian Journal of Communication, 40*(4), 466-484. https://doi.org/10.17576/JKMJC-2024-4004-26

Bao, H., Cao, B., Xiong, Y., & Tang, W. (2020). Digital media's role in the COVID-19 pandemic. *JMIR Mhealth Uhealth, 8*(9), e20156. http://doi.org/10.2196/20156

Lomborg, S., & Bechmann, A. (2014). Using APIs for data collection on social media. *The Information Society, 30*(4), 256-265.

Neely, S., Eldredge, C., & Sanders, R. (2021). Health information seeking behaviors on social media during the COVID-19 pandemic among American social networking site users: Survey study. *J Med Internet Res, 23*(6), e29802. http://doi.org/10.2196/29802

Parry, D. (2024, May 09). Restrictions on data access impede crucial societal research. *University World News*. https://www.universityworldnews.com/post.php?story=20240509000643443

Sachdeva, P., Barreto, R., Bacon, G., Sahn, A., Von Vacano, C., & Kennedy, C. (2022, June). The measuring hate speech corpus: Leveraging rasch measurement theory for data perspectivism. *Proceedings of the 1st Workshop on Perspectivist Approaches to NLP @LREC2022*, 83–94. Marseille, France: European Language Resources Association. https://aclanthology.org/2022.nlperspectives-1.11/

Saima, A., Iqbal, N., & Ishaq, R. (2022). Social media as a news source: An analysis of Facebook. *Global Multimedia Review, V*(I), 24-46. http://dx.doi.org/10.31703/gmmr.2022(V-I).03

Saroj, A., & Pal, S. (2023). E-governance through social media: An analysis on the use of Facebook and Twitter by Indian Government. *Electronic Government, An International Journal, 19*(3), 304-331. https://doi.org/10.1504/EG.2023.130585

Sörensen, I., Fürst, S., Vogler, D., & Schäfer, M. S. (2023). Higher education institutions on Facebook, Instagram, and Twitter: Comparing Swiss universities' social media communication. *Media and Communication, 11*(1), 264-277. https://doi.org/qfsn

Subekti, D. (2024). The movement of opposition political parties in social media: Case study of relocation of the Indonesia nation's capital policy. *Jurnal Komunikasi: Malaysian Journal of Communication, 40*(4), 181-200. http://doi.org/10.17576/jkmjc-2024-4004-10

Vidgen, B., Thrush, T., Waseem, Z., & Kiela, D. (2020). *Learning from the Worst: Dynamically Generated Datasets to Improve Online Hate Detection*. Cornell University. https://doi.org/10.48550/arXiv.2012.15761

Yu, H., Murat, B., Li, L., & Xiao, T. (2021). How accurate are Twitter and Facebook altmetrics data? A comparative content analysis. *Scientometrics, 126*, 4437-4463. https://doi.org/gjxrq4