

The Psychological Boomerang: Investigating the Mental Health Impacts of Spreading Anti-LGBTQ+ Misinformation and Hate Speech Online

Man Shu Katie Lai^{1,a,*}

¹*Harrow International School Hong Kong, Hong Kong SAR, 1 Castle Peak Road, Block 31 2A, Hong Kong, 000000, China*

a. katielaims@ldy.edu.rs

**corresponding author*

Abstract: The increasing prevalence of hate speech in digital spaces has become a growing concern, particularly with the advancement of technological tools, with sexual minorities being a primary target of this growing phenomenon. While prior research has primarily focused on the psychological impact of hate speech on victims, the effects on those who spread such hate remain underexplored. This study aims to examine the relationship between actively spreading LGBTQ+-related misinformation and the mental health of message spreaders. To investigate this, Twitter user accounts that used the #superstraight hashtag, a meme popularized in February 2021 to express transphobic sentiments, were analyzed over a three-month period. Sentiment analysis of posts from 421 users, conducted before and after the period of spreading the messages, revealed a negative correlation between the dissemination of misinformation and users' mental health. These findings suggest that the dissemination of hate speech is associated with a decline in the mental well-being of individuals who propagate such messages. This research highlights the harm of online hate on all and provides insights into combating hate speech.

Keywords: Misinformation, Online Activism, LGBTQ+, Hate Speech, Mental Health.

1. Introduction

In the modern digital age, online activism has become a powerful tool for marginalized communities, especially the LGBTQ+ community, to advocate for acceptance and social change. On the other hand, hate speech and false information have surged online as a result of this increased visibility. Although a great deal of study has been done on the detrimental effects that hate speech and disinformation have on the mental health of sexual minorities, there is still a clear knowledge vacuum about the effects that spreading hate speech has on the mental health of those who disseminate it [1][2]. This study seeks to fill this gap by examining the following research question: Does promoting misinformation or hate speech about LGBTQ+ advocacy online correspond with harmful mental health outcomes? Drawing on social identity theory, the online disinhibition effect, and the frustration-aggression hypothesis, this study suggests that disseminating misinformation may be the result of complicated psychological processes including group identification and lower social inhibitions in online contexts. Noting the lack of empirical evidence in the Freudian theory, recent

research suggests that expressing aggression online may reinforce hostile tendencies and increase emotional distress for the perpetrators [3]. By integrating these frameworks with the cumulative stress model, this study posits that engagement in online hate speech against LGBTQ+ communities leads to deteriorating mental well-being for those spreading such content [4]. The purpose of this study is to shed light on the psychological effects that go unresearched on those who spread false information about LGBTQ+ populations online. Through an analysis of this overlooked aspect, this research seeks to contribute to a multidisciplinary strategy to counteract digital prejudice and to the creation of more comprehensive intervention solutions. Furthermore, this research offers perspective on how we may rethink how we approach digital well-being in order to promote an inclusive online community and meet the psychological needs of those who participate in the hateful cycle of the internet.

2. Literature Review

There is paucity of research on the relationship between the mental health of those who participate in sexual minority advocacy and the dissemination of hate speech. Relevant research mostly focuses on examining the psychological effects of prejudice against sexual minorities. For instance, it has been shown that a variety of discriminatory practices, including homophobia, microaggressions, and institutional bias, can lead to emotional discomfort in lesbian, gay, and bisexual individuals [1][2]. As shown by 69% of the participants displaying depressive symptomology assessed through PSS-4 and CES-D-10 scales, harassment and exclusion based on sexual orientation and gender identity significantly harm LGBTQ+ mental health [1]. Despite the research, there is limited knowledge about how people spreading hate speech toward sexual minorities might experience mental health issues. This gap is important because understanding the psychological factors driving hate speech could inform more effective intervention strategies.

Understanding the influence of spreading hate speech toward sexual minorities is intrinsically linked to understanding the influence of online discrimination on sexual minority activism. Freedom of communicative exchange and accessibility of modern digital media platforms surged the emergence of online activism, which to some extent addresses the discrimination toward marginalized communities. Online activism, manifesting in forms such as posts, content sharing, and hashtags, has been shown to have a positive impact on marginalized groups, whose feelings of support are enhanced through activism and increased online visibility [5].

While collective online advocacy for the acceptance of socially marginalized groups brings together supportive communities and drives change, it also unfortunately attracts online hatred. For example, the sharing of distorted information and false narratives about selected groups proposes consequential issues; it reinforces existing biases and limits exposure to diverse perspectives [6]. Although active efforts exist in moderating and banning hate posts on social media platforms, vast amounts of hateful content remain visible. Furthermore, cyber hatred has the potential to rapidly escalate and contribute to an increase in real-life violence, posing a significant threat to the physical safety of marginalized populations.

A significant unfavorable consequence of online hatred lies in its impact on mental health, especially towards LGBTQ+ activism. As highlighted in Keighley's study, online hate elicits various detrimental emotional and behavioral reactions among LGBTQ+ young individuals, notably including feelings of sadness, shame, and inferiority [7]. This leads to the internalization of blame for victimization within LGBTQ+ communities, as they feel pressured to adhere to the identity norms of dominant society. This further impedes their ability to express their unique differences [7].

Scholarly works suggest that the impact and the level of hate perceived by people differ due to various factors. Individuals who encounter hate material more frequently tend to find it more distressing, and those who exhibit greater acceptance of behaviors that defy social norms tend to be less affected by online hate [8]. Also, certain personality traits may be attributed to susceptibility to

cyber hate and sensitivity towards information perceived online. For example, the capability of discernment between true and false information is associated with higher levels of open-mindedness and agreeableness [9]. However, there is limited knowledge about how messages specifically targeting sexual minority communities influence those who spread such messages.

In existing scholarly works on mental health and targeted, cyber hate towards sexual minority communities, the focus tends predominantly to be on addressing mental health consequences of receiving cyber hate, superseding the negative mental health associations for the perpetrators of hate speech. However, it is also important to investigate the instigators of hate speech, as they may play a crucial role in contextualizing and understanding the importance of discrimination. Future efforts in advocacy for the prevention of online discrimination also require an examination of the potentially alarming psychological impacts of actively spreading hate speech.

3. Theory

Social identity theory could be applied to suggest why people tend to engage in spreading hateful commentaries. The innate human nature of social categorization and identification in which we perceive and then adopt the set of values and attitudes of ingroups is exemplified in digital media. In the context of supporting hate speech, individuals who recognize themselves as having similar values of a certain hate group are motivated to display commitment through aggressive digital behaviors, such as aggravating hate speech comments, as it provides a boost of self-esteem and belonging [10]. To elucidate further, the online disinhibition effect, where people feel less restrained and more likely to act inappropriately online compared to in-person interactions, might explain why hate comments online are often more intense and malicious than hate encountered in real life.

Another trigger for spreading hate comments can be explained through the Frustration and Aggression hypothesis [11]. The increase in strong emotions of frustration can often lead to aggressive behavior. In the context of online hatred, the inclination to spread hateful comments online may be provoked by cultivated annoyance or resentment towards online LGBTQ+ activism content. However, purposefully expressing and venting one's anger through outbursts of hateful information online may be disadvantageous in relieving intense emotions. Extended research works on Freud's Catharsis theory challenge the explanation in that aggressive outbursts may reduce psychological tension [12]; rather, expressing aggression tends to reinforce than alleviate hostile feelings. Specifically, the expression of discontentment often reinforces aggressive tendencies and increases feelings of anger, leading to greater emotional arousal [3]. As individuals repeatedly express anger, they cognitively reinforce their hostility [3]. Online interactions involving aggressive behavior can further validate and justify these hostile beliefs through feedback loops, leading to desensitization toward the harm caused by their expressions of hatred. Consequently, not only does the targeted community suffer emotional harm from receiving hate speech, but the perpetrators themselves often feel more stressed and dissatisfied after posting hateful comments, rather than experiencing relief.

Furthermore, by integrating the cumulative stress model by Lazarus and Folkman, where detrimental effects on mental health may be accumulated from continuous exposure to stressors, it could be inferred that the aggressive nature of perpetrators in spreading online hate speech may also possibly be experiencing deteriorating mental well-being [4]. Consequential effects of anxiety, depression, and other stress-related disorders may result from the prolonged activation of the stress response and damaged interpersonal relationships through online hatred behaviors. In accordance with the compiled theoretical frameworks above, this study hypothesizes that actively spreading hate speech against LGBTQ+ activism online is correlated with negative mental health conditions.

4. Research Design

Hypothesis: Individuals spreading LGBTQ+-related misinformation are correlated with a decline in their mental health.

To examine the hypothesis about the relationship between spreading LGBTQ-related misinformation and the mental health of the people spreading the messages, this study uses text data from X (previously Twitter) by focusing on the event “Super Straight”. The “Super Straight” meme, popularized in February 2021 by TikTok user Kyle Royce, quickly became a transphobic trope embraced by anti-trans social media users. Disguised as a legitimate sexual orientation, it falsely equates itself with LGBTQ identities, claiming to be neutral and non-hateful. However, it functions as a form of whataboutism and false equivalency, allowing anti-LGBTQ sentiment to spread under the guise of personal identity [13]. Relevant texts are targeted using the hashtag #superstraight during the period between February 2021 to April 2021, which covers the major event period.

Data is collected using APIFY, a platform that provides automated web scraping services. Specifically, Tweet Scraper V2, a Twitter scraper, is used to extract text data from X [14][15]. This scraper allows for “lightning-fast search, URL, and profile scraping options,” and query outputs include information such as the post URL, text, time, and retweet count, enabling a comprehensive analysis of the discourse around relevant events [16]. The procedures of data collection are as follows. Firstly, tweets containing the hashtag #superstraight between February 2021 to April 2021 were scraped. Second, the user IDs of those who tweeted or posted these messages at that time were determined by utilising the post URL. For example, the user ID "helloworld" is contained in the URL "https://twitter.com/helloworld/status/123456789123456789." It is the users of these posts that disseminate the false information against LGBTQ people. Thirdly, more searches were carried out with the time period shifted to include one month prior to and one month following the incident, utilising the user data that was gathered. To make sure user data is thoroughly sentiment analysed, time periods of one month prior to and one month following the event are employed. Fourth, TextBlob was used to do sentiment analysis on the postings these users made both before and after the occurrence, which is a Python library that provides a simple API for natural language processing (NLP) tasks such as part-of-speech tagging and sentiment analysis [17]. It works by processing text and assigning a polarity score ranging from -1 to 1, where -1 indicates a highly negative sentiment, and 1 indicates a highly positive sentiment. With this tool, the positive vs. negative sentiment of these posts was examined, allowing for examining whether users spreading misinformation exhibited more negative sentiments after the message spread. This approach provides a way to analyze the psychological impact on individuals who spread misinformation, with a focus on how their mental health might be reflected in their other online activities and posts.

Between February 2021 and April 2021, 645 people were discovered using the hashtag #superstraight; 421 of them posted both before and after the hashtag movement's peak phase. Of the 421 users, 129,382 posts were made one month before the time and 162,122 posts one month subsequent to it.

The independent variable, which is the action of spreading misinformation related to LGBTQ+ people, is operationalised by counting the number of posts made by users about LGBTQ+ people between February 2021 and April 2021. This is done in order to test the hypothesis that there is a relationship between mental health and the act of spreading misinformation. To be more precise, this entails monitoring the quantity of related posts made by each user to ascertain the degree to which each user participated in the dissemination of material pertinent to this subject. The dependent variable, representing potential changes in mental health, is measured through sentiment analysis. This is done by calculating the difference between the sentiment scores of posts made in the month following April 2021 (i.e., May 2021) and those made in the month preceding February 2021 (i.e.,

January 2021). The sentiment score comparison allows for evaluating shifts in emotional tone over time, helping to identify whether increased activity in spreading LGBTQ+ related misinformation correlates with more negative, potentially indicating deteriorating mental well-being among those users.

5. Findings

The analysis revealed a negative correlation of -0.12 between the number of LGBTQ+-related misinformation posts shared by users and the change in their sentiment scores. This correlation suggests that as users increased their activity in spreading such misinformation, their overall sentiment tended to become slightly more negative. While a correlation of -0.12 is relatively weak, it still provides some evidence of a link between engaging in misinformation and shifts in emotional well-being. It also aligns with previous findings from studies on COVID-19 misinformation, which indicated that individuals who actively spread false information experienced a slight deterioration in sentiment over time [18].

Overall, the negative trend indicates the potential emotional burden associated with engaging in this behavior, possibly reflecting anxiety, hate, or social consequences related to spreading harmful or misleading content. While the correlation is not robust, it reinforces the idea that participating in the spread of false information — especially on sensitive topics like LGBTQ+ issues — may be associated with a negative impact on users' mental health. Further research could use a wider demographic and collect a larger user group than 421 users in a longitudinal study to explore additional variables or mechanisms driving this relationship. Additionally, a more elaborate emotional analysis could be used to explore the specific mental health issues and underlying emotions spreading hate speech may correlate to, thus providing a more comprehensive understanding.

The findings of this study suggest several potential future directions. Firstly, public education initiatives could focus on the mental health risks faced by perpetrators of online hate speech, increasing self-awareness and potentially reducing harmful behaviors. The study also highlights the need for expanding research on examining motivational factors on hate-speech perpetrators to develop more effective interventions for mitigating online hate. Activism strategies, peer support networks, and safe online spaces may need to be adjusted accordingly.

6. Conclusion

This study successfully investigated the correlational effect of spreading online hate speech on LGBTQ+ activism and worsening mental health. The findings show a negative correlation between the mental health symptoms of users and the usage of misinformed LGBTQ+ hashtags on tweets. Thus, the results confirm the research hypothesis that spreading LGBTQ+-related hate speech misinformation is correlated with a decline in mental health. However, this study also has some limitations. Firstly, it is not a causal analysis and does not account for demographic factors like age, gender, or other user attributes. Future research could benefit from analyzing a broader dataset, including a wider range of user demographics and hashtags over longer periods. Another limitation is that the study only looked at hashtag usage related to a single event. Further studies should explore more recent events and consider analyzing user activity over extended periods before and after such events. Finally, emotional analysis could build on these findings to provide more insights into the mental health impacts associated with online hate.

References

- [1] Kneale, D., & Becares, L. (2021). *The mental health and experiences of discrimination of LGBTQ+ people during the COVID-19 pandemic: Initial findings from the Queerantime Study*. *MedRxiv* (2020): 2020-08.

- [2] Marciano, L., & Antebi-Gruszka, N. (2022). Offline and online discrimination and mental distress among lesbian, gay, and bisexual individuals: The moderating effect of LGBTQ facebook use. *Media Psychology*, 25 (1), 27-50.
- [3] Bushman, B. J. (2002). Does venting anger feed or extinguish the flame? Catharsis, rumination, distraction, anger, and aggressive responding. *Personality and Social Psychology Bulletin*, 28 (6), 724-731.
- [4] Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer Publishing Company.
- [5] Matsick, J. L., Kim, L. M., & Kruk, M. (2020). Facebook LGBTQ pictivism: The effects of women's rainbow profile filters on sexual prejudice and online belonging. *Psychology of Women Quarterly*, 44(3), 342-361.
- [6] Frank, L., & Mohamed, S. (2024). *Systemic Approaches to Strategic Planning: Exploring other systemic approaches to strategic planning, such as Systems Thinking, Complexity Theory, and Cybernetics*.
- [7] Keighley, R. (2022). Hate hurts: Exploring the impact of online hate on LGBTQ+ young people. *Women & Criminal Justice*, 32(1-2), 29-48.
- [8] Costello, M., Hawdon, J., Bernatzky, C., & Mendes, K. (2019). Social group identity and perceptions of online hate. *Sociological inquiry*, 89(3), 427-452.
- [9] Calvillo, D. P., León, A., & Rutchick, A. M. (2023). Personality and misinformation. *Current Opinion in Psychology*, 101752.
- [10] Tajfel, H., & Turner, J. C. (2004). *The Social Identity Theory of Intergroup Behavior*. In J. T. Jost & J. Sidanius (Eds.), *Political psychology: Key readings* (pp. 276–293). Psychology Press. <https://doi.org/10.4324/9780203505984-16>
- [11] Berkowitz, L. (1989). Frustration-aggression hypothesis: examination and reformulation. *Psychological bulletin*, 106(1), 59.
- [12] Konecni, V. J., & Doob, A. N. (1972). Catharsis through displacement of aggression. *Journal of Personality and Social Psychology*, 23(3), 379.
- [13] GLAAD. "Online Anti-LGBTQ Hate Terms Defined: 'Super Straight.' " <https://glaad.org/super-straight-definition-meaning-anti-lgbt-online-hate/>
- [14] Twitter Scraper. <https://console.apify.com/actors/61RPP7dywgiy0JPD0/information/latest/readme>
- [15] X. <https://x.com>
- [16] API Dojo. <https://apify.com/apidojo>
- [17] Loria, S. (2018). *textblob Documentation*. Release 0.15, 2.
- [18] Verma, G., Bhardwaj, A., Aledavood, T., De Choudhury, M., & Kumar, S. (2022). Examining the impact of sharing COVID-19 misinformation online on mental health. *Scientific Reports*, 12(1), 8045.