

I 🍑 your hate: Emojis as infrastructural platform violence on Telegram

Esteban Morales
University of Groningen
e.morales.velasquez@rug.nl

Anatoliy Gruzd
Toronto Metropolitan University
gruzd@torontomu.ca

Jaigris Hodson
Royal Roads University
jaigris.Hodson@royalroads.ca

Philip Mai
Toronto Metropolitan University
philip.mai@torontomu.ca

Abstract

Emojis are a ubiquitous form of online expression. In this paper, we explore emojis as affordances that enact and sustain discursive violence via toxic content. We take a case study approach by focusing on Chismes Frescos Medellin (Fresh Gossip Medellin), a Colombian Telegram group with over 125,676 members. Relying on Communalistic, we collected 98,729 publicly accessible posts. Next, we subdivided the posts into 3,155 toxic and 95,574 non-toxic posts using Detoxify, a popular machine-learning classifier. We explored and compared the two subsets through statistical analysis and thematic analysis. Our findings show that emojis—and specifically, emojis suggesting positive emotions such as 🍑 and 😊—are often used to accompany toxic speech in ways that indicate the approval and normalization of toxic speech. Overall, our study points to the need to pay closer attention to how affordances can enable symbolic forms of violence on digital platforms in unexpected ways.

Keywords: Telegram, Emojis, Toxic speech, Platforms, Violence

1. Introduction

Emojis are now ubiquitous on the internet, with over 3,782 digital symbols available for users across platforms, covering everything from emotions and actions to national flags, animals, and food (Unicode, n.d.). Indeed, emojis have become a key affordance on digital platforms to express affect, developed under a standard that enables its readability across platforms and devices (Abidin & Gn, 2018; Berard, 2018). As a technological affordance, emojis are a core component of how we define and understand the internet—and, thus, our contemporary cultural milieu. Such importance was best demonstrated by the Oxford

Dictionary, which made the face with tears of joy 😊 the 2015 word of the year, arguing that it properly reflected the “ethos, mood, and preoccupations” of that year (PBS, 2015, para 2).

Nevertheless, scholars and activists increasingly highlight the role various platform affordances—such as emojis—have in enabling and sustaining antisocial behaviour (Munn, 2020; Recuero, 2024; Riedl et al., 2024). For instance, Weigel and Gitomer (2024) discuss how users of the platform Gab use the share function to amplify and spread hateful speech. Likewise, Are (2024) showcased how the flagging feature (a mechanism for reporting offensive content) in platforms like TikTok and Instagram is used to silence others. This line of research has emphasized the consequences of platform design, whereby certain features that may seem mundane or positive at first can be used to facilitate and even encourage online harm (Munn, 2020).

In this context, we aim to build on previous research that has explored the role of emojis in the enactment and propagation of antisocial behaviour (Ghaffari, 2022; Lazareva, 2023; Matamoros, 2018). Here, we scrutinize emoji use as a tool of *infrastructural platform violence*—a concept put forward by Riedl et al. (2024) that emphasizes how violence is both propagated *on* platforms (that is, when people use platforms to bully and harass others violence) and propagated *by* platforms (that is, the platform affordances that are actively or passively enabling harm). As such, infrastructural platform violence marks that, while the infrastructure of our digital landscape might not be created with the intent to harm, it can nonetheless be harmful in its implementation—thus suggesting that platforms bear some responsibility for the negative uses of such affordances.

In what follows, we explore emojis as affordances that enact and sustain discursive violence as they relate to other harmful content. While previous research has

explored how emojis act as mechanisms to enact harm (e.g., Al-Rawi, 2022; Matamoros-Fernandez, 2018), our study aims to emphasize how emojis accompanying toxic speech can sustain, expand, and normalize harmful behaviours. Through this analysis, we seek to better grasp the social gratifications (Walther, 2024) that affordances such as emojis have in relation to online toxic speech. To achieve this, we focus on two research questions:

RQ1 How are emojis used differently on toxic compared to non-toxic posts?

RQ2: How do group members use emoji-based reactions on toxic comments?

As a case study, we examine interactions on the Telegram group called Chismes Frescos Medellin (Fresh Gossip Medellin). By focusing on Telegram, we focus our inquiry on a platform often recognized for its anti-sociality (e.g., Guhl & Davey, 2020; Urman & Katz, 2022). This platform has increasingly become a site where bad actors migrate due to its lax content moderation policies (Rogers, 2020; Vergani et al., 2022). As such, studying platform affordances within Telegram holds a particular opportunity to understand how affordances can relate to expressions of antisocial behaviour.

Moreover, the Telegram group analyzed in this study is a popular venue in Medellin (Colombia) for discussing and making sense of events occurring in the city. The group was selected due to its scale, popularity, and contextual importance. At the time of data collection (October 2023), it had 125,676 enrolled members discussing what is happening in Medellin—the second biggest city in Colombia and one of the epicentres of violence in the country during the past century (Centro Nacional de Memoria Histórica, 2017). Emphasizing a context so closely connected to cultures of violence presents an opportunity to unpack the mediation processes of harm on digital platforms (Morales, 2024).

2. Literature review

2.1. Emojis in the digital landscape

Ubiquitous in our digital landscape, emojis are often seen as one of the vital building blocks for online communication (Abidin & Gbn, 2018; Berard, 2018). Their history traces to emoticons—a linguistic feature that represents faces through the creative use of punctuation, such as :-), and which has existed since at least the 19th century (Bai et al., 2019). From this context came emojis, which cannot be typed with keyboards but require predefined names and code (Yoo & Rays, 2021). Emojis—images measuring 12 by 12 pixels, which appear as a single character within inline

text on digital devices and platforms—serve as graphical symbols capable of depicting facial expressions, abstract concepts, emotions/feelings, as well as animals, plants, activities, gestures/body parts, symbols and objects (Berard, 2018). Emojis can be used as independent message content or to react to others' messages—for example, by adding a heart to another user's message on platforms like WhatsApp or Facebook.

Amid their growing popularity, emojis are an important affordance with which people can express and communicate affect to others (Abidin & Gbn, 2018). For instance, Kaye and colleagues (2017) found that research into emoticons has underscored their utility as valuable nonverbal communication tools, shedding light on the cognitive and neural processes underlying digital communication. However, to portray emojis as *only* an affective mechanism is somewhat incomplete. Emojis also represent an illocutionary force—that is, they convey actions within digital messages, whether an intention of playfulness, sarcasm or harm (Matamoros-Fernandez, 2018).

Accordingly, as cultural objects and technical constructs (Stark & Crawford, 2018), emojis have presented opportunities for researchers to study online communication practices. Indeed, researchers increasingly rely on ever more complex analytical systems to analyze emojis, aiming to better grasp the emotional intricacies of digital communication, including sentiment analysis (Grover, 2022). Emojis are also a valuable tool for researchers interested in the qualitative shifts in how people negotiate meaning and culture online—for example, by seeing how emojis act to promote or inhibit conversations on politics (Riddick & Shivener, 2022). Amid these methodological possibilities, emojis remain a critical study artifact for understanding the intricacies of digital communication.

2.2. Emojis and antisocial behaviour

In the context of such cultural relevance, emojis are often used to accompany and even enact antisocial behaviours. Examples of these antisocial uses of emojis abound. For instance, Ghaffari (2022) discusses how fandoms use emojis like 🤒 (sick), 😷 (masked) or 🐷 (pig) as forms of hate speech against female celebrities on social media. Similarly, Matulewska and Gwiazdowicz (2020) discuss how emojis are employed to cyberbully those seen as different. Overall, these examples illustrate how emojis can be employed as part of hate speech, cyberbullying, and other forms of antisocial behaviour.

The impact of emojis as tools for antisociality draws from their role as affective affordances of platforms—affordances that are often read as innocuous but may be used to promote antagonism and even harm

others (Sturm Wilkerson et al., 2021). For instance, Matamoros-Fernandez (2018) illustrated how far-right activists use emojis to incite anger and hate, thus increasing the range of their racist discourses on Facebook. Emojis then become essential for amplifying or provoking emotions against others. Al-Rawi (2022), for example, shows how Twitter and Instagram users relied on emojis such as 🧠 (skull) and 🗡️ (crossed swords) to threaten Muslims. In this context, emojis act as a “subsidiary mechanism used for various purposes: establishing a power hierarchy, bonding between like-minded readers, conveying unfavorable emotions, and marking dehumanizing humor” (Lazareva, 2023; p. 42).

Researchers have aimed to use the preeminent role of emojis in antisocial behaviour to better grasp their pervasive nature on digital dialogues (e.g., Wiegand & Ruppenhofer, 2021; Kirk et al., 2022). These methodological approaches often rely on emojis that are explicitly suggestive of harm and violence—such as 🧠 (skull) and 🗡️ (water gun)—as ways of understanding possibly toxic emojis. However, this kind of analysis is not enough to capture the role of emojis in antisocial behaviour, as they are often part of complex cultural processes of adaptation and cooption of various visual (and seemingly non-threatening) representations. For instance, Bartolo and Matamoros-Fernandez (2023) discuss how, in 2022, racists appropriated monkey and banana emojis to attack black footballers in the 2022 World Cup. The harmful use of emojis such as these ones would not necessarily be flagged by automated systems without a full understanding of cultural and contextual cues. While current methodological approaches aim to address more complex readings of emojis concerning hate speech—such as when emojis are used as positive cofounders of toxic speech (Kirk et al., 2022)—there is little research on such contextual processes with real-life data.

Indeed, identifying and addressing emojis in antisocial contexts is not an easy task—as they are overly context-dependent. For instance, McMahon and Kirley (2019) discuss the difficulties of introducing emojis in criminal trials, as their interpretation often depends on both the discourses surrounding the emoji and the cultural background of both the perpetrator and the target. Similarly, Highfield (2018) discusses how emojis often contain semiotic ambiguity that constrains a shared meaning even when used with community hashtags. He illustrates that, for example, users often respond to efforts to moderate specific language by substituting the specific harmful word with a corresponding emoji, which is, in turn, more challenging to police. In this sense, Patton and colleagues (2019) argue that “the perceived threats in violent emojis or images are a ‘puzzle’ that requires careful translation of meaning and understanding of

context, symbols, and placement of content.” (p. 70; see also Wagner et al., 2020).

3. Conceptual framework: Emojis as infrastructural platform violence

Violence that occurs due to the passive affordances of social media platforms has been conceptualized as infrastructural platform violence (Reidl et al., 2024). This type of violence occurs when a “platform’s infrastructure neglects—or does not sufficiently protect—vulnerable user populations” (p. 3), for example, when bad actors use the specific affordances of a platform to perpetuate online violence. For instance, Reidl and colleagues discussed how trolls abuse WhatsApp’s reporting features to harass targets. Similarly, emojis are an affordance that can both spread and reinforce online violence via a mechanism termed *value capture* (Nguyen, 2024). Value capture refers to a phenomenon whereby metrics, such as the step counts via FitBits, likes and shares on Facebook, or the accumulation of positive emojis on a Telegram post tend to rise in value without a person consciously being aware they are adopting the value specified by the metric being tracked. Put simply, value capture occurs when what is being measured becomes what matters. Nguyen (2024) gives the example of a couple who started to cancel dinner dates with friends because their FitBit data reported they had not achieved their step count for the day.

As part of the many affordances that people can use to judge the way others receive their posts, otherwise positive emojis can encourage a kind of value capture that not only reinforces the strength of antisocial online sentiment in the eyes of readers; but also, and perhaps more importantly, could encourage the poster to ramp up their antisocial or violent posting habits. By providing feedback, often positive, to the original poster, emojis can unintentionally incite the production of more violent content, in addition to being a means of communicating negative sentiment towards specific groups directly. When emojis serve to reinforce a message, they can create a type of violent value capture: a feedback loop of online violence that is more difficult to deal with than the simple use of negative emojis as a means to demean marginalized groups. In this manner, they can become part of a platform infrastructure that rewards negative, toxic and antisocial posting behaviours.

4. Method

To respond to our two research questions—exploring the difference between emojis used across

toxic and non-toxic posts and the use of emoji-based reactions to toxic comments—we relied on Communalytic (Gruzd & Mai, 2024). Using this platform, we collected 98,729 publicly accessible posts (with their accompanying emoji-based reaction counts) from the Chismes Frescos Telegram group. The data collection processes focused on messages posted between April and September 2023. To further explore the most antisocial content in the group, we relied on one of Communalytic's integrated toxicity analyzer modules (Detoxify). This module automatically assigns a toxicity score ranging from 0 to 1 to each post in the dataset. Following previous studies that have relied on similar analysis procedures (Gruzd et al., 2020; Pascual-Ferrá et al., 2021), posts with a toxicity score of 0.7 or higher were deemed likely to be toxic. Of the total corpus, 3,221 posts (constituting 3.2% of the dataset) surpassed this threshold. A further qualitative review of all the toxic posts removed 66 other posts because their contents were falsely tagged as toxic. To achieve this, we read the comments in context to better understand and assess the use of toxic language in our data. For instance, a post saying, "Here it's crazy," was coded as toxic by Detoxify. Under further scrutiny, it was clear that the message was referring to a traffic jam and not as an insult. As a result, this was removed from the sample. Through this process, we had a total of 98,729 posts in the study sample—3,155 toxic posts and 95,574 non-toxic posts.

We relied on two analytic processes to respond to the research questions of this study. For RQ1, we exported the totals of emoji reactions for each post in the sample—a total of 235,718 emoji-based reactions. We relied on descriptive statistics to explore these emoji-based reactions, particularly examining their distribution across toxic and non-toxic posts. To find if the differences between the emoji use across toxic and non-toxic posts were significant—and following previous studies that have relied on similar processes (Gruzd et al.; 2018)—we conducted an independent samples *t*-test to compare the means of each of the seven emojis across the two categories.

To make sense of the emotion portrayed by emojis, we follow previous work on sentiment analysis of emojis (e.g., Brito et al., 2019; Novak et al., 2017; Yoo & Rayz, 2021), examining the prevailing attitudes towards different posts in the sample. Nevertheless, we are aware that emotions portrayed across emojis are highly contextual. For instance, Brito et al. (2019) discuss how the shy face emoji (😳) can be read as sickness, indecision, silence, or wealth by readers across various contexts and demographics. Similarly, Częstochowska and colleagues (2022) found that while some emojis are often unambiguous (such as the heart emoji ❤️), others pack more complex semantics that

inhibit their understanding across people (such as religious and astrological symbols). We consider such contextual richness of emojis in our study, aiming to nuance the findings of our study to better account for this limitation.

To respond to RQ2, we qualitatively analyzed the messages displaying toxic speech. More specifically, we relied on thematic analysis (Braun et al., 2019), creating coherent themes and narratives across the toxic posts that generated the most emoji-based reactions. To achieve this, we categorized participants' posts by following an open coding scheme, where the first author identified, grouped, and refined a series of themes across the sample. Following the work of Braun and Clark (2019), we focus on the narratives that are prevalent across our analysis of the data.

Finally, it is worth noting that, as the data from the Telegram group was in Spanish, the first author translated the quotations showcased in this paper. In the translation process, the objective was to preserve participants' intended meaning, achieve functional equivalency between the texts in both languages and capture the overall meaning of users' posts (Munday, 2016; Roth, 2013).

5. Findings

In this section, we use toxic quotes verbatim in order to illustrate how emojis occur on toxic posts. For that reason, we caution the reader that some content in this section may be offensive, but the context is necessary for both the data richness and to remark on the ethical implications of antisocial behaviour on digital platforms (Jane, 2015).

5.1. Comparison of the use of emojis between toxic and non-toxic posts

Among the 98,729 posts that comprise our sample, there were a total of 235,718 emoji-based reactions. For our study, we only include those presented across the sample >1%, comprising seven different emojis. The most used emoji in the sample is 👍 (thumbs up), which was used 101,417 times (43% of all emoji-based reactions). Other popular emojis included the face with symbols on mouth emoji 😬 (37,090, 15.7%), the grinning face with smiling eyes emoji 😄 (31,236, 13.3%), the heart emoji ❤️ (30,666, 13%), the face screaming in fear emoji 😱 (17,471, 7.4%), and the clapping hands emoji 🙌 (11,592, 4.9%). Overall, the emojis often associated with positive sentiments (👍, ❤️, 🙌, and 😄) represent the majority of the reactions in the Telegram group (74.6%), while emojis often associated with negative sentiments (😱, 😬, and 😳) are less prevalent (24.7%). These results suggest that the

group often exhibits positive sentiments towards other people's posts.

A look at the distribution of emojis across toxic and non-toxic posts further complicates the image of positivity among members. Indeed, our results show that positive emoji-based reactions to toxic posts are slightly more likely to occur than those to non-toxic posts. As shown in Figure 1 (below), the mean number of emojis used per post is higher for toxic posts than for non-toxic posts. For instance, while the mean of finding the face with symbols on mouth emoji (👤) is 0.37 across non-toxic posts, it was 0.43 across toxic posts. Moreover, the increased likeliness of finding emojis in toxic messages is particularly evident across the emojis usually assumed to be of positive sentiment. For instance, the mean use of the thumbs-up emoji (👍) was 1.01 across non-toxic posts but 1.44 across toxic posts.

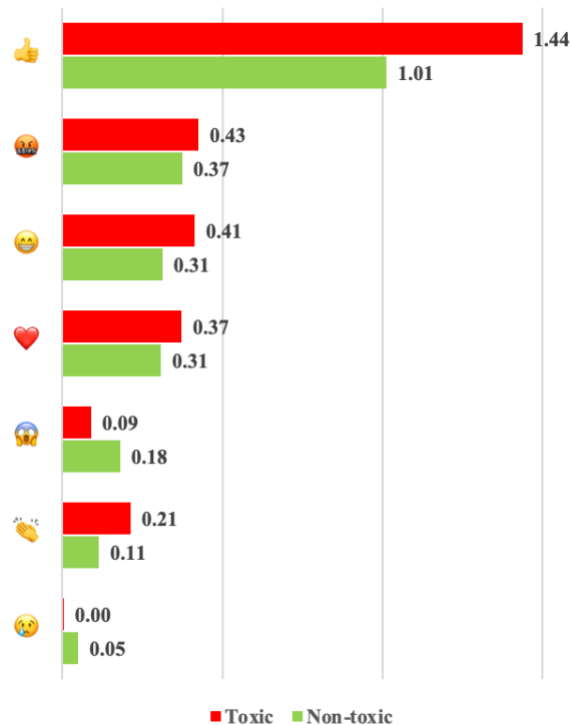


Figure 1. Mean emoji use per post across toxic and non-toxic posts

To determine if the differences between emoji use across toxic and non-toxic posts were statistically significant, we conducted an independent samples t-test. To do this, we first ran Levene's Test for Equality of Variances to determine the homogeneity of the variance across the sample of each emoji (see Table 1).

	F	Sig.
--	---	------

👍	98.471	<.001
❤️	6.017	0.014
👏	38.496	<.001
😊	9.758	0.002
😱	26.756	<.001
👤	2.122	0.145
😄	27.853	<.001

Table 1. Levene's Test for Equality of Variances

Based on the results from Levene's test of equality of variances, the analysis of the 👤 emoji assumes equal variances, while for the rest of the emojis, equal variances are not assumed. With these results, we conducted an independent samples t-test (Table 2). The results of this test show that, except for ❤️ and 👤, most emojis (👍, 👏, 😊, 😱, and 😄) were found to have a statistically significant difference of means across toxic and non-toxic posts. This suggests that toxicity can play a role in determining whether or not people react to the messages on social media platforms with emojis—and more so when it concerns the use of emojis with positive connotations.

	t	df	Sign. (2-tailed)	Mean Diff.	Std. Error Diff.
👍 Equal variances not assumed	7.5	3378.8	<.001	0.4	0.1
❤️ Equal variances not assumed	1.3	3376.0	0.20	0.1	0.0
👏 Equal variances not assumed	3.1	3399.5	0.00	0.1	0.0













 Equal variances not assumed	3.0	3567.5	0.00	0.1	0.0
 Equal variances not assumed	-3.7	3662.3	<.001	-0.1	0.0
 Equal variances assumed	.754	1019.48	.451	.050	.066
 Equal variances not assumed	-11.2	1705.9.6	<.001	0.0	0.0




Table 1. Independent samples t-test for equality of means






5.2. Exploration of the use of emojis in toxic posts

A close examination of the toxic posts that generate emoji-based reactions further shows how emojis, as a technological affordance, can be used to enact, reinforce, and occasionally challenge harm. Here, we first found that negative emojis as a response to content denouncing people who enact violence. For instance, a user shared a photo of a person who was allegedly hurting animals, saying “Get to know the face of the most despicable and psychopathic being in Villavicencio, the animal killer.” Members of the group reacted to this message with 97  emojis, thus illustrating their disapproval of the violence against animals. These responses serve to construe the refusal of such enactment of violence.

In contrast, and more often in our dataset, when users posted negative content directed at other members, the most common emoji responses were positive. For example, when a user posted a threat to another user (“I want to take the silly face off this fool”), people reacted with  38 times and  two times, illustrating what might be construed as an approval of the threat. Likewise, when a user defended a politician, another user replied, “You are a scumbag just like him,” which generated a positive response among members of the community, with them reacting with  14 times.



People’s positive reactions were not limited to toxic speech towards other users—they also occurred in responses to calls addressing insecurity in the city with further violence. For instance, when a user posted a call for killing and torturing those who commit gender-based violence (“This should be the end for abusers, rapists and femicides”), community members reacted 89 times with the  emoji, 47 times with  emoji, 24 times with the  emoji, and 14 times with the  emoji.

Similarly, when a user posted that it would be great to hurt thieves, saying, “It would be great to knock out those teeth with three kicks in that head,” people reacted with the  emoji 25 times, the  emoji five times, and the  emoji two times. These reactions showcase an approval of the call for taking justice into citizens’ hands by members of the group.

Finally, another common evidence of positive reactions was seen in emojis tagged on xenophobic posts—for instance, when discussing what to do with the surge of migrants from Venezuela in their city. When a user posted, “Venezuelans are the worst that the earth could sprout. Polluting all countries where they arrive, GARBAGE!!!!,” other members of the community reacted with the  emoji 16 times and the  emoji three times. Likewise, when a user argued that people should be “tough on those bastards [Venezuelans]. What a pest they are, why don’t they just go back to their own country and stop fucking around the world?” members of the group reacted with 16  emojis, one  emoji, and one  emoji. These reactions suggest a tacit acceptance and even approval of the xenophobic comments toward migrants in the country, implying that violent responses toward them would be positively received, and increasing the chances this type of behaviour will be reinforced.

6. Discussion

As mentioned in our introduction, the analysis of emojis is like completing a puzzle, particularly when researchers are trying to understand how harm is perpetuated against marginalized groups on online platforms (Patton et. al, 2019). This is because, even if one can use toxicity and similar analysis tools to understand the surface-level communication of an emoji, it must actually be understood in context to fully understand the implications of its use. Accordingly, in this paper we have combined automated toxicity analysis and statistical analysis with qualitative analysis of emojis in context to understand the complex ways that affordances such as emojis contribute to online harm against others.

Emojis often serve as a symbol of what it means to communicate with others in the postdigital world—a way to depict and share affect and meaning with others through simple images. However, they can and are often used to enact harm to others. In this context, our findings illustrate how emojis are an understudied way that hate and harm are reinforced and escalated. Certainly, in our data, emoji-based reactions (particularly those suggesting positive emotions such as  and ) are more often used to accompany toxic speech than non-toxic speech in ways that suggest the approval of toxic speech. Through the lens of infrastructural platform

violence, emojis act as affordances that enable the normalization of harm across everyday uses of digital platforms. To grasp the implications of our findings, it is necessary to explore two related issues: first, the role of emojis and toxic speech in the political economy of digital platforms, and second, the social gratification people expect and obtain when using them.

As we have already discussed at length, emojis are one of the most commonly used features to express affect on social media platforms (Abidin & Gn, 2018; Berard, 2018). However, affect through the use of emojis does not flow without barriers. Emojis are constrained by platforms' economic and political interests, which decide which features better drive user engagement and thus translate into profits for them—without much attention to how they might be used for harm (Munn, 2020). As noted by Stark and Crawford (2015), “emoji offer us more than just a cute way of ‘humanizing’ the platforms we inhabit: they also remind us of how informational capital continually seeks to instrumentalize, analyze, monetize, and standardize affect” (p. 8). Emojis are then not only a way of cultural expression and communication across digital platforms—they are a key component in these platforms' economic and political functioning. As such, the use of emojis and toxic speech displayed in this study are not just accidental manifestations of how people communicate online. Rather, they are inherent parts of our digital ecosystem.

Indeed, the affective forces displayed throughout emojis in our study—joy, anger, hate—have long been considered key drivers of engagement within digital platforms (Boler & Davis, 2018). The implications of platforms' affective force on harm and hate—as the one displayed here—cannot be underestimated. Examples of the role of affect in anti-social behaviours in digital spaces abound. For instance, Harel and colleagues (2020) illustrated how affective polarization on social media is used to dehumanize others in order to justify harming them. In this context, positive emojis on toxic messages are an intentional design feature to drive user engagement, whether with the intention of sustaining harm or not, making them a key site of unintentional violence propagated by platforms (Riedl et al. 2024). Thinking of our findings in this context means that we need to consider how emojis can escalate already negative and antisocial behaviour further, providing fuel for the fires of online violence via a demonstration of community support for it. Emojis are part of the infrastructure that links online engagement to the propagation of symbolic online violence—a technological decision with profound and contradictory social implications.

Thus, it is necessary to consider the social gratifications people expect and obtain when engaging

in such antisocial behaviour. Value capture appears to be part of the perverse incentives for toxic posts on Chismes Frescos Medellin. Our findings show that toxic posts most commonly receive the most positive emojis, suggesting that those who post violent content might be receiving positive reinforcement through the use of emojis for posting the most negative content. While people may come to this Telegram group to learn more about their neighbourhood or to help advocate for safety, if once in the group they are rewarded with positive feedback for posting racist or otherwise toxic content, they might be encouraged to keep sharing toxicity. With value capture, one of the key dangers is that the values being inculcated through metrics are externally motivated and cannot take into account the complexity of human interaction (Nguyen, 2024). This is likely exacerbated by the fact that Telegram is famous for a lack of moderation.

Seen through the lens of value capture, our findings highlight the social forces at play in enacting toxic speech. Indeed, our findings align with previous research that has discussed how people engage in antisocial behaviours to create bonds with like-minded individuals (Marwick, 2021; Walther, 2022). In this line of research, Walther (2024) discussed how people who engage in antisocial behaviour can do so for any number of social gratifications, including having fun, acquiring social approval, or sustaining social support. Through the use of positive emojis on toxic messages, our study points out how such social gratifications might be actually materialized—for example, by reacting with a thumbs-up emoji (👍) to a message that diminished the existence of others.

Our analysis has implications for information design researchers. With respect to the design of online communities, we have shown how the affective affordance of emojis may be used to escalate and normalize digital manifestations of violence against others. This use goes beyond the substitution of emojis for banned words on platforms since even an emoji that is used to symbolize a like (👍 and 🍷, for example) can be used unironically to reinforce support for toxic content. This suggests that information designers may want to carefully consider the unintentional consequences of affective tools like emojis. While the ability to add emojis to a post is not a negative or toxic affordance in and of itself, it can be used in unanticipated ways to cause or reinforce online harms. It may be necessary to accompany emojis with strategies to diminish the possibility of using them to normalize and enact violence, such as developing community standards for their use in groups where people might be at risk for harassment or other toxic online behaviour.

For information researchers, it is important to note the complexities of toxic speech analysis on digital

platforms and recognize the importance of engaging in mixed-methods research, complementing big data analysis of content with small sub-sample qualitative coding. Otherwise, researchers risk missing important contextual clues that can change ❤️ to 🤡. Though often assumed to be trivial or innocuous, emojis are important tools for self-expression and affect. As such, emojis offer clues to better grasp how hate and polarization grow online, thus warranting further attention from the scholarly community as a whole.

7. Conclusion

Whereas other platform features can be used in ways that more directly enact harm to others—such as surveillance (Dragiewicz et al., 2018) and harassment (Marwick, 2021)—the impacts of emojis as infrastructural platform violence cannot be overstated. Here, the findings of our study point to emojis' vital role in enacting symbolic violence—that is, where abuses and imbalances of power are legitimized and normalized (Kramsch, 2020). In brief, we have showcased how emojis are used to legitimize harmful discourses across this Telegram group, such as xenophobia, insecurity, and personal insults. More significantly, our analysis illustrated how emojis concerning positive emotions are used along with toxic messages in ways that further emphasize their symbolic harm. Through these processes, platforms contribute to weakening the social fabric by spreading, legitimizing, and enacting harm (Recuero, 2024).

Scholars could take the insights of this study into future work. First, given that we presented a single case study, future studies could explore the role of emojis (particularly positive ones) in enacting and naturalizing toxic speech across various sociocultural contexts and platforms. Second, as the non-toxic post in our sample comprised 95,574 messages, we could not systematically screen which ones were falsely coded as non-toxic. Thus, future work could analyze the non-toxic posts to better contextualize their toxicity. Additionally, future work could take a more detailed approach to emoji use concerning toxic speech. For instance, future work could include variations in emoji use concerning toxic speech across communities, topics, and identities, to name a few.

These processes of symbolic violence displayed in this study have severe implications in post-conflict scenarios like Colombia. Amid a political struggle to define the future of the armed conflict in the country and its fight to make sense of its histories of violence (Gutierrez-Sanin, 2020), mechanisms that reinforce the legitimization of harmful discourses hinder the possibility of fomenting a culture of peace in the country. Moreover, as the processes of meaning-making

of citizens concerning peace and violence are increasingly mediated through social media platforms (Morales, 2024), the findings of this study then point to emojis' vital role in sustaining and expanding existing cultures of violence and polarization in the country.

8. Acknowledgements

This research is funded in part by the Social Sciences and Humanities Research Council of Canada (PIs: Gruzd, Hodson, Jacobson)

9. References

- Abidin, C., & Gn, J. (2018). Between art and application: Special issue on emoji epistemology. *First Monday*. <https://doi.org/10.5210/fm.v23i9.9410>
- Are, C. (2024). Flagging as a silencing tool: Exploring the relationship between de-platforming of sex and online abuse on Instagram and TikTok. *New Media & Society*. <https://doi.org/10.1177/14614448241228544>
- Bai, Q., Dan, Q., Mu, Z., & Yang, M. (2019). A Systematic Review of Emoji: Current Research and Future Perspectives. *Frontiers in Psychology*, 10. <https://doi.org/10.3389/fpsyg.2019.02221>
- Berard, B. (2018). I second that emoji: The standards, structures, and social production of emoji. *First Monday*. <https://doi.org/10.5210/fm.v23i9.9381>
- Boler, Megan, and Elizabeth Davis. "The Affective Politics of the 'Post-Truth' Era: Feeling Rules and Networked Subjectivity." *Emotion, Space and Society* 27 (May 2018): 75–85. <https://doi.org/10.1016/j.emospa.2018.03.002>.
- Brito, P. Q., Torres, S., & Fernandes, J. (2019). What kind of emotions do emoticons communicate? *Asia Pacific Journal of Marketing and Logistics*, 32(7), 1495–1517. <https://doi.org/10.1108/APJML-03-2019-0136>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Braun, V., Clarke, V., Hayfield, N., & Terry, G. (2019). Thematic analysis. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 843–860). Springer Singapore. https://doi.org/10.1007/978-981-10-5251-4_103
- Centro Nacional de Memoria Histórica. (2017). *Medellín, Memorias de una Guerra Urbana*. <https://centrodememoriahistorica.gov.co/medellin-memorias-de-una-guerra-urbana/>
- Częstochowska, J., Gligorić, K., Peyrard, M., Mentha, Y., Bień, M., Grütter, A., Auer, A., Xanthos, A., & West, R. (2022). On the Context-Free Ambiguity of Emoji. *Proceedings of the International AAAI Conference on Web and Social Media*, 16, 1388–1392. <https://doi.org/10.1609/icwsm.v16i1.19393>
- Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N. P., Woodlock, D., & Harris, B.

- (2018). Technology facilitated coercive control: Domestic violence and the competing roles of digital media platforms. *Feminist Media Studies*, 18(4), 609–625. <https://doi.org/10.1080/14680777.2018.1447341>
- Ghaffari, S. (2022). Discourses of celebrities on Instagram: Digital femininity, self-representation and hate speech. *Critical Discourse Studies*, 19(2), 161–178. <https://doi.org/10.1080/17405904.2020.1839923>
- Grover, V. (2022). Exploiting Emojis in Sentiment Analysis: A Survey. *Journal of The Institution of Engineers (India): Series B*, 103(1), 259–272. <https://doi.org/10.1007/s40031-021-00620-7>
- Government Information Quarterly*, 35, 579–587.
- Gruzd, A., Lannigan, J., & Quigley, K. (2018). Examining government cross-platform engagement in social media: Instagram vs Twitter and the big lift project. *Government Information Quarterly*, 35, 579–587.
- Gruzd, A., & Mai, P. (2024). *Communalytic – A Research Tool For Studying Online Communities and Online Discourse*. [Computer software]. <https://communalytic.org/>
- Gruzd, A., Mai, P., & Vahedi, Z. (2020). Studying anti-Social behaviour on Reddit with Communalytic. In *The SAGE Handbook of Social Media Research Method* (pp. 503–520).
- Guhl, J., & Davey, J. (2020). *A safe space to hate: White supremacy mobilisation on Telegram*. Institute for Strategic Dialogue. <https://www.isdglobal.org/wp-content/uploads/2020/06/A-Safe-Space-to-Hate.pdf>
- Gutiérrez-Sanín, F. (2020). ¿Un nuevo ciclo de la guerra en Colombia? Debate.
- Harel, T. O., Jameson, J. K., & Maoz, I. (2020). The normalization of hatred: Identity, affective polarization, and dehumanization on Facebook in the context of intractable political conflict. *Social Media + Society*, 6(2). <https://doi.org/10.1177/2056305120913983>
- Highfield, T. (2018). Emoji hashtags // hashtag emoji: Of platforms, visual affect, and discursive flexibility. *First Monday*. <https://doi.org/10.5210/fm.v23i9.9398>
- Jane, E. A. (2015). Flaming? What flaming? The pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, 17(1), 65–87. <https://doi.org/10.1007/s10676-015-9362-0>
- Kaye, L. K., Malone, S. A., & Wall, H. J. (2017). Emojis: Insights, Affordances, and Possibilities for Psychological Science. *Trends in Cognitive Sciences*, 21(2), 66–68. <https://doi.org/10.1016/j.tics.2016.10.007>
- Kirk, H. R., Vidgen, B., Röttger, P., Thrush, T., & Hale, S. A. (2022, May 6). *Hatemoji: A Test Suite and Adversarially-Generated Dataset for Benchmarking and Detecting Emoji-based Hate*. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Seattle, United States. <http://arxiv.org/abs/2108.05921>
- Lazareva, S. (2023). Emojis as graphical hate speech markers. *Journal of Language Works*, 8(1), 29–44.
- Marwick, A. E. (2021). Morally motivated networked harassment as normative reinforcement. *Social Media + Society*, 7(2), 205630512110213. <https://doi.org/10.1177/20563051211021378>
- Matamoros-Fernández, A. (2018). Inciting anger through Facebook reactions in Belgium: The use of emoji and related vernacular expressions in racist discourse. *First Monday*. <https://doi.org/10.5210/fm.v23i9.9405>
- Matulewska, A., & Gwiazdowicz, D. J. (2020). Cyberbullying in Poland: A case study of aggressive messages with emojis targeted at the community of hunters in urbanized society. *Social Semiotics*, 30(3), 379–395. <https://doi.org/10.1080/10350330.2020.1731194>
- McMahon, M., & Kirley, E. (2019). When Cute Becomes Criminal: Emoji, Threats and Online Grooming. *Minnesota Journal of Law, Science and Technology*, Fall. <https://ssrn.com/abstract=3451553>
- Morales, E. (2024). Social media and the mediation of everyday violence: A study of Colombian young adults' experiences. *New Media & Society*, 14614448241236756. <https://doi.org/10.1177/14614448241236756>
- Munday, J. (2016). *Introducing translation studies: Theories and applications* (4th ed. edition). Routledge
- Munn, L. (2020). Angry by design: Toxic communication and technical architectures. *Humanities and Social Sciences Communications*, 7(1), 1–11. <https://doi.org/10.1057/s41599-020-00550-7>
- Nguyen, C. (2024). Value capture. *Journal of Ethics and Social Philosophy*, 27(3).
- Novak, P. K., Smailović, J., Sluban, B., & Mozetič, I. (2015). Sentiment of emojis. *PLOS ONE*, 10(12). <https://doi.org/10.1371/journal.pone.0144296>
- Pascual-Ferrá, P., Alperstein, N., Barnett, D. J., & Rimal, R. N. (2021). Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211023533>
- Patton, D. U., Leonard, P., Elaesser, C., Eschmann, R. D., Patel, S., & Crosby, S. (2019). What's a Threat on Social Media? How Black and Latino Chicago Young Men Define and Navigate Threats Online. *Youth & Society*, 51(6), 756–772. <https://doi.org/10.1177/0044118X17720325>
- PBS. (2015, November 17). Oxford Dictionaries 2015 word of the year is an emoji. *PBS NewsHour*. <https://www.pbs.org/newshour/nation/oxford-dictionary-says-the-2015-word-of-the-year-is-an-emoji>
- Recuero, R. (2024). The platformization of violence: Toward a concept of discursive toxicity on social media. *Social Media + Society*, 10(1), 20563051231224264. <https://doi.org/10.1177/20563051231224264>
- Riddick, S., & Shivener, R. (2022). Affective Spamming on Twitch: Rhetorics of an Emote-Only Audience in a Presidential Inauguration Livestream. *Computers and Composition*, 64, 102711. <https://doi.org/10.1016/j.compcom.2022.102711>
- Riedl, M. J., El-Masri, A., Trauthig, I. K., & Woolley, S. C. (2024). Infrastructural platform violence: How women and queer journalists and activists in Lebanon experience abuse on WhatsApp. *New Media & Society*, 14614448241248372. <https://doi.org/10.1177/14614448241248372>

- Rogers, R. (2020). Deplatforming: Following extreme Internet celebrities to Telegram and alternative social media. *European Journal of Communication*, 35(3), 213–229. <https://doi.org/10.1177/0267323120922066>
- Roth, W.-M. (2013). Translation in qualitative social research: The possible impossible. *Forum: Qualitative Social Research*, 14(2). <https://doi.org/10.17169/fqs-14.2.1986>
- Stark, L., & Crawford, K. (2015). The Conservatism of Emoji: Work, Affect, and Communication. *Social Media + Society*, 1(2), 2056305115604853. <https://doi.org/10.1177/2056305115604853>
- Sturm Wilkerson, H., Riedl, M. J., & Whipple, K. N. (2021). Affective affordances: Exploring Facebook reactions as emotional responses to hyperpartisan political news. *Digital Journalism*, 9(8), 1040–1061. <https://doi.org/10.1080/21670811.2021.1899011>
- Unicode. (n.d.). *Full Emoji List*, v15.1. Retrieved April 23, 2024, from <https://unicode.org/emoji/charts/full-emoji-list.html>
- Urman, A., & Katz, S. (2022). What they do in the shadows: Examining the far-right networks on Telegram. *Information, Communication & Society*, 25(7), 904–923. <https://doi.org/10.1080/1369118X.2020.1803946>
- Vergani, M., Martinez Arranz, A., Scrivens, R., & Orellana, L. (2022). Hate Speech in a Telegram Conspiracy Channel During the First Year of the COVID-19 Pandemic. *Social Media + Society*, 8(4), 205630512211387. <https://doi.org/10.1177/20563051221138758>
- Wagner, A., Marusek, S., & Yu, W. (2020). Emojis and Law: Contextualized flexibility of meaning in cyber communication. *Social Semiotics*, 30(3), 396–414. <https://doi.org/10.1080/10350330.2020.1731198>
- Walther, J. B. (2022). Social media and online hate. *Current Opinion in Psychology*, 45, 101298. <https://doi.org/10.1016/j.copsyc.2021.12.010>
- Walther, J. B. (2024). Making a case for a social processes approach to online hate. In J. B. Walther & R. E. Rice (Eds.), *Social Processes of Online Hate* (1st ed.). Routledge.
- Weigel, M., & Gitomer, A. (2024). Hate-sharing: A case study of its prevalence and impact on Gab. *New Media & Society*, 14614448241245349. <https://doi.org/10.1177/14614448241245349>
- Wiegand, M., & Ruppenhofer, J. (2021). Exploiting Emojis for Abusive Language Detection. *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 369–380. <https://doi.org/10.18653/v1/2021.eacl-main.28>
- Yoo, B., & Rayz, J. T. (2021). Understanding emojis for sentiment analysis. *The International FLAIRS Conference Proceedings*, 34(1). <https://doi.org/10.32473/flairs.v34i1.128562>