'Made in Italy': Online hate speech targeting gender-normative defiance at the Sanremo music festival

Citation:
Borgna, C. and Martella, A. (2025), "'Made in Italy': Online Hate Speech Targeting Gender-Normative Defiance at the Sanremo Music Festival", Goglio, V. and Biancalana, C. (Ed.) *Mapping the Evolution of Platform Society,* Emerald Publishing Limited, Leeds, pp. 65-86. https://doi.org/10.1108/978-1-83608-028-220251005

[“Made in Italy”: online hate speech targeting gender-normative defiance at the Sanremo music festival]
[Camilla Borgna, Antonio Martella]

PLEASE PUT THE CORRESPONDING AUTHOR IN BOLD
[Camilla Borgna, Department of Cultures, Politics and Society, University of Turin & Collegio Carlo Alberto]
[camilla.borgna@unito.it]
[**Antonio Martella**, Department of Cultures, Politics and Society, University of Turin]
[**antonio.martella@unito.it**]

Authors Biographies:

Camilla Borgna, University of Turin, Italy, is Assistant Professor of Sociology at the Department of Cultures, Politics, and Society at the University of Turin, Italy. She is also Affiliate of the Collegio Carlo Alberto and Fellow of the College for Interdisciplinary Educational Research. Her research lies at the intersection of social stratification, sociology of education, and migration.

Antonio Martella, University of Turin, Italy, is a researcher at the Department of Cultures, Politics, and Society at the University of Turin. He is a member of the Observatory on Political and Public Communication at the same department and an associate researcher at MediaLaB | Big Data in Social and Political Research Lab at the Department of Political Science, University of Pisa. Additionally, he serves on the board of the Italian Association of Political Communication, is the secretary of the scientific journal The Lab’s Quarterly, and member of the Editorial board of the scientific Journal ComPol. His research interests primarily focus on digital journalism and innovation in the production, dissemination, and consumption of information through platform analysis, data science, and artificial intelligence.

ABSTRACT:
This study examines online hate speech targeting gender-normative defiance during the 2023 Sanremo Music Festival in Italy, focusing on the performance of the rapper Rosa Chemical, who faced backlash after kissing a male singer on stage. Using machine learning analysis techniques, we compare user comments across three social media platforms: Facebook, Instagram, and TikTok. Results reveal significant differences in the prevalence and nature of hate speech across these platforms, with Facebook exhibiting the highest levels of hate speech (35.9%), predominantly driven by anger and disgust, while TikTok had the lowest (1.9%). Hate speech was strongly correlated with negative emotions like anger and disgust, particularly on Facebook. Moreover, while on Facebook comments characterized by negative emotions produced more reactions, on TikTok comment negativity was not correlated with the number of responses. These findings are consistent with the interpretation that older audiences on platforms like Facebook feel more threatened by gender-normative challenges and resort to online hate speech as a form of cultural backlash. Moreover, platform-specific moderation policies, content distribution mechanisms, and social norms about the perceived appropriateness of negative content may influence the amount of hate speech and the degree to which users decide to engage with it. This research contributes to the understanding of the 'supply side' of online hate speech by highlighting how platform architecture and user demographics influence the production and reaction to hate speech.

KEYWORDS:
(Please supply 6-10 keywords for your Chapter to help with depository and online searches)

1. Hate speech
2. Gender norms
3. Cultural backlash
4. Facebook
5. TikTok
6. Instagram

1. Introduction

Over the past decade, internet exposure in developed countries has grown tremendously: while in 2002 only about half of the population reported using the internet, at least occasionally, ten years later virtually all adults had become internet users (PEW Research Center, 2022). In particular, social media have become widespread not only among young people, but also among older individuals. In Italy, which is the empirical context of the analyses presented in this chapter, more than half of the population aged 50 or over report using social media in their daily lives (PEW Research Center, 2022). While social media have many obvious benefits in terms of social connectivity and information sharing, they have also created more opportunities for aggressive discourse to circulate (Blazak, 2009; Hawdon et al., 2017).

Online hate speech – i.e., the expression of offensive, harassing, or threatening attitudes based on some social identity factor – negatively affects its victims (Wachs et al., 2022), its bystanders (Reichelmann et al., 2021) and society at large (Bilewicz & Soral 2020). While previous studies have prevalently focused on online hate speech detection (Fortuna & Nunes, 2018), research on the processes whereby this discourse is produced is less common. Within this emerging field, two methodological approaches prevail. The first strategy relies on self-reported behavior measured by means of targeted surveys (e.g., Inguanzo et al., 2021): this approach makes it possible to collect rich background information on potential perpetrators, but it is prone to obvious social desirability biases. The second strategy uses social media data and applies automated content analysis techniques to identify hateful discourse. In order to circumvent the lack of background information on potential perpetrators, the extant research has compared hate speech incidence across regional contexts (e.g. Vargo & Hopp, 2017).

In this chapter, we contribute to the emerging literature on the 'supply side' of hate speech by adopting another analytical strategy, i.e., by comparing online behavior across social media platforms. Although this strategy has been seldom applied in the literature, partly due to methodological constraints (Siegel, 2020; Schemer & Reiners, 2023), it is promising because platforms differ in terms of at least three factors that may affect the amount of online hate speech, as well as counter-reactions to it: (1) their typical content and usage; (2) users' demographics; and (3) moderation and censorship policies.

In this regard, we conduct a case study on an event where a public figure challenged traditional gender norms during a mainstream televised event in Italy and thus became the target of violent attacks, both online and offline. We analyzed online reactions to the event across the three social media platforms where the victim was active.

The chapter is organized as follows. Section 2 provides the background for the case study and discusses recent research on online hate speech production, while Section 3 presents our research questions and contribution to the literature. In Section 4, we describe the case that we analyzed. Section 5 details our empirical strategy and Section 6 presents and discusses the results. Section 7 concludes and suggests avenues for further research.

2. Background and previous studies

Online hate speech can be defined as a form of communication (including text, videos, and photos) that expresses hatred or degrading attitudes towards individuals or groups of people because of their gender identity, sexual orientation, body shape, race, ethnicity, religion, national origin, or some other social identity factor (Blazak 2009; Hawdon et al., 2017). Clearly, violent and discriminatory discourse also exists and has always existed in the offline world. Yet, the online context facilitates its expression and spread by making it easier to maintain anonymity and to access broader audiences. This is particularly the case of social media platforms, where individuals can connect with each other while avoiding face-to-face interactions. However, it is important to note that the same characteristics also create opportunities for resisting and countering hate speech (Barlińska et al., 2018; Eschmann, 2021; Inara Rodis, 2021).

Most previous research on this topic has focused on artificial intelligence tools with which to detect and thus censor hate speech (for recent reviews, see Schmidt & Wiegand, 2017; Fortuna & Nunes, 2018; Siegel, 2020;

European Union Agency for Fundamental Rights, 2023; Gandhi et al., 2024), as well as on identifying the individual- and contextual-level predictors of this discourse, in order to design effective prevention and intervention measures (Munger 2017; Àlvarez-Benjumea & Winter, 2018; Hangartner et al., 2021; Stahel & Weingartner, 2024).

Less investigated are the processes whereby online hate speech is produced. Content analyses of hateful discourse reveal that typical targets are individuals who challenge the status quo by not conforming to the traditional norms associated with their group, such as minority women who work in male-dominated domains or are active in the public sphere (Sobieraj, 2020; Martella & Pavan, 2023). In fact, online hate speech can be an expression of cultural backlash against ongoing social change (Inguanzo et al., 2021). As societies become more inclusive and diverse, some individuals who feel threatened by these changes may react by attacking those that they perceive as challengers to the status quo power structure, in order to reassert their perceived superiority or dominance. Hence, online hate speech can take the form of racism, sexism, homophobia, transphobia, or xenophobia, among others, and often overlaps with radically conservative views (Council of Europe, 2022).

The extant research furnishes some insights into the 'supply side' of online hate speech (Hawdon et al., 2019), i.e., the characteristics of its perpetrators and the contexts that facilitate its production. By analyzing data from a US online survey and by focusing on young people (aged 15 to 36), Costello & Hawdon (2018) found that men are more likely than women to have produced, at some point in their life, online material that other people would likely interpret as hateful or degrading. The authors did not detect any significant differences in hate speech production by age, race, educational attainment, religion, political ideology, or employment status, although it is important to note that their sample was not fully representative of the US population. Interestingly, while the amount of time spent online is not a direct predictor of hate speech production, the use of specific social media platforms, such as Reddit, is. Inguanzo et al. (2021) adopted a similar research strategy to study the characteristics of hate speech perpetrators, i.e., they designed and implemented a survey where respondents were asked to report on their online behavior and attitudes. However, differently from Costello & Hawdon (2018), Inguanzo and colleagues) did not restrict themselves to one country, but collected data on the USA, Spain, Germany, and the UK, and employed sampling procedures that ensured statistical representativeness. Their results indicated that individuals holding hostile sexist views are more likely to generate political content online. As a consequence, although some form of political polarization of online discourse is in place, sexist discourse prevails over countering (feminist) discourse. In the same vein are the results of similar survey-based studies, such as Kaakinen et al. (2018), Frischlich et al. (2021), and Bernatzky et al. (2022), which indicate that gender, socio-economic position and political orientation are important predictors of online hate speech production at the individual level.

Despite the informative character of these studies, self-reported measures of aggressive online behavior are likely affected by social desirability biases, with a consequent underestimation of hate speech. In contrast, research based on social media data uses automated content analysis techniques to identify online hate speech directly. The main advantages of this strategy – i.e. the direct observation of users' online behavior in a natural setting and the large sample sizes that can be obtained – come at the cost of very limited information about users' background characteristics. At the individual level, perpetrators of online hate speech have been found to have more short-lived profiles with fewer followers, because the type of content they produce often leads to banning (ElSherief et al., 2018; Ribeiro et al., 2018). However, they tend to be very active and embedded in highly dense networks, factors that enhance their visibility (Ibid.). Unfortunately, social media platforms generally do not allow researchers to systematically retrieve users' socio-demographic information except (and only partially) for gender identity.

Hence, in order to study hate speech production using social media data, researchers often compare its spread across regional contexts. For instance, Vargo & Hopp (2017) studied online political discourse during the 2012 US presidential campaign by analyzing 'tweets' and found that aggressive discourse was more frequent in districts characterized by poorer socio-economic conditions, low levels of partisan polarity and high levels of racial diversity. Similarly, Rosenbusch et al. (2020) used Twitter data to test whether the level

and dispersion of outgroup biases based on race and sexual orientation predict differences in social media hostility, and they found evidence of a positive association at the regional level. Denti & Faggiani (2021) applied the same approach to a different context (Italy) and found that hateful discourse on Twitter is more common in localities where income inequality is greater, even net of regional differences in education. This, they argued, suggests that online hate speech is fueled by the perceived threat of instability in social positions.

In a rare example of methodological triangulation, Stahel & Weingartner (2024) combined social media data and survey data to study the processes whereby aggressive online speech is produced. First, they applied supervised machine-learning techniques to classify comments posted in response to online petitions put forward on a German platform. Second, they conducted a voluntary survey among a subsample of users that had posted aggressive or non-aggressive comments. They found that online aggression is more common among historically advantaged groups (men, older people, and national citizens) who have experienced a loss of social status (in terms of education or socio-economic conditions). Moreover, many of these structural effects are mediated by political inclination, meaning that these groups lean more than others towards right-wing ideologies, and use online aggression as a form of cultural backlash to defend their dwindling privilege.

3. Contribution and research questions

In this chapter, we contribute to the emerging literature on the 'supply side' of hate speech, which has so far mainly relied on self-reported behavior (survey data) or comparisons of actual behavior across geographical units (social media data). We do so by adopting another analytical strategy, i.e. comparing online behavior across social media platforms. Although this strategy has been seldom applied in the literature, partly due to methodological constraints (Siegel, 2020; Schemer & Reiners, 2023), it is promising because platforms differ in terms of at least three factors that may affect the amount of online hate speech as well as counter-reactions to it.

Typical content and usage. Social media platforms may be primarily designed and/or used to create and share content, to discuss specific topics or trends, to connect with others both socially and professionally, to build online communities centered around shared interests, to play interacting games, etc. Depending on these features, some platforms may be more conducive to hate speech. Particularly problematic are discussion forums and chat communities where users tend to remain anonymous and where political discourse is prevalent (e.g. Reddit or 4chan). Even within mainstream social media, platforms that are mainly used to share content (be it textual, as on X/Twitter, or visual, as on Instagram or TikTok or YouTube) may facilitate the expression of hate speech (as well as counter-reactions) to a greater extent than networking platforms (e.g., Facebook or LinkedIn), where users are typically not anonymous. Even among platforms that are typically used to share content, differences may exist due to the prevailing social norms. By comparing emotions disclosure, Masciantonio & Bourguignon (2024) showed that Instagram tends to foster a more positive environment compared to Twitter, although it is also associated with 'low' negative emotions such as boredom and lassitude. In similar vein, Waterloo et al. (2018) evidenced that expressing positive emotions is deemed more appropriate on Instagram than on Facebook, especially among women, whereas the expression of negative emotions is considered more acceptable on Facebook than on Instagram. They also linked these findings to platforms' characteristics, such as the emphasis on self-promotion and the visual nature of Instagram. Finally, platforms differ in how content becomes visible to users. For example, on Facebook, users typically see the content posted within their social networks, while on X/Twitter and Instagram users 'follow' specific content producers. TikTok is fundamentally different in this respect, because content is algorithmically proposed in the 'For you' page based on individual preferences regardless of the following/follower relationship. In other words, videos are distributed by the platform in clusters of probably like-minded individuals identified algorithmically by their behavior on the platform (Martella & Cepernich, 2024). Nevertheless, users adopt 'corrective' practices to make sure that sensitive content (i.e., LGBTQ+

community videos) is suggested to desired/imagined audiences in a way that avoids hate speech (Karizat et al., 2021).

User socio-demographics. Partly because of their content, and partly because of the historical phase in which they were created, social media platforms also differ in terms of the population they attract (Meltwater & We Are Social, 2023; Schemer & Reiners, 2023; PEW Research Center, 2024). Among mainstream platforms, Facebook is used by a wide range of social groups, but it is particularly popular among adults older than 30. X/Twitter is more common among younger adults, especially those with higher levels of education and income. By contrast, TikTok is very popular among teenagers and young adults aged under 30 who are more varied in terms of socio-economic backgrounds. Finally, Instagram is extremely widespread among adolescents, young adults, and, increasingly, adults in their 30s and 40s. Since previous research has indicated that socio-demographic characteristics are significant predictors of attitudes and behavior related to online hate speech, social platforms catering to different user populations may result in a larger or smaller amount of hateful content and counter-reactions. In particular, given that hate speech can be an expression of cultural backlash (Inguanzo et al., 2021), we may expect it to be more common in platforms with older audiences, who are more likely to feel threatened by social change.

Moderation and censorship policies. Arguably the most common institutional strategy to combat online hate speech is content moderation, which may result in the removal of content and the banning of accounts that violate platforms' terms of service or stated rules. Content moderation may reduce the diffusion of hate speech in two ways: first, by discouraging users from producing it in the first place (and possibly moving to other online communities), and second, by diminishing its visibility because, even if hateful content is produced, it is quickly removed. Since the specific rules of conduct adopted by the different platforms vary,[1] the amount of hate speech may also vary across platforms. Unfortunately, the extant research reports mixed results in regard to how these bans are actually implemented by the different platforms and how effective they are in reducing hate speech (Fortuna & Nunes, 2018; Siegel, 2020).

Against this background, this chapter asks whether the production of online hate speech (supply side) and the reactions to this kind of discourse (demand side) vary across social media platforms. Given the emerging status of research in this field and the mixed results of previous studies, we adopt an exploratory approach and refrain from formulating precise research hypotheses.

4. Case study

To address our research questions, we investigate a specific case that occurred during the 2023 edition of the Sanremo Music Festival, in which the rapper Rosa Chemical[2] competed with a song, titled 'Made in Italy', which celebrated sexual freedom and gender fluidity as part of the Italian identity. During his performance on February 11th, Rosa Chemical kissed a male singer and influencer, Fedez, after twerking on him. The episode generated polarized reactions in the public debate, reaching up to the national parliament. On the one hand, conservative politicians and commentators criticized the performance as offensive, dangerous for children, and 'a form of gender propaganda'.[3] Critics also questioned the genuine nature of the performance and whether the kiss was consensual. Yet another controversial issue was the fact that Fedez was married to

[1] Since 2016, the European Commission has promoted a voluntary Code of Conduct on Countering Illegal Hate Speech Online jointly with the most popular social media platforms. This Code of Conduct is meant to harmonize moderation and censorship strategies for online content visible in the European Union, but major differences remain among the specific platform guidelines. See for example: Facebook & Instagram: https://transparency.meta.com/policies/community-standards/hate-speech/ ; X/Twitter: https://help.twitter.com/en/rules-and-policies/violent-speech

[2] https://en.wikipedia.org/wiki/Rosa_Chemical

[3] See, for example, the speech by Maddalena Morgante, a Member of Parliament belonging to the government party Fratelli d'Italia, calling for Rosa Chemical to be banned from the Sanremo Music Festival: https://www.fanpage.it/politica/fuori-rosa-chemical-da-sanremo-fa-propaganda-gender-la-richiesta-di-fratelli-ditalia-alla-rai/

the well-known influencer Chiara Ferragni, who was in fact one of the hosts of the 2023 Sanremo Music Festival, and may have been embarrassed, both personally and professionally, by the 'adulterous' kiss. On the other hand, Rosa Chemical and his supporters argued that the performance was inherently political, as an 'innocuous kiss among friends' successfully exposed the bigotry present in mainstream television.[4]

In the immediate aftermath of the event,[5] Rosa Chemical posted a number of pictures and videos on the three social media platforms where he was active, namely Instagram, TikTok, and Facebook. The content, partially overlapping between platforms, included: (1) a picture of the controversial kiss, without any accompanying text; (2) a video of his contest song, 'Made in Italy', dubbed in sign language; (3) a photo gallery of himself together with Alex Mucci, an OnlyFans content creator, where the two pose with fetish dresses in public spaces,[6] with the caption 'This is just the first step. Change has just begun'; (4) a video of himself and Alex Mucci, wearing bathrobes in a dressing room and singing the contest song 'Made in Italy'; (5) a photo gallery of himself together with Rose Villain, a female singer also competing at Sanremo, where the two pose wearing bathrobes in a hotel room, with the caption 'Sanremo is not only the Italian Music Festival. It is also the Italian Sex Festival.' Table 1 summarizes the content of the posts and their comparability across social media platforms.

TABLE 1 HERE

This case study is exemplary for at least three reasons. First, it involves highly visible public figures, who are likely targets of hate (ElSherief et al., 2018; Isbister et al., 2018), especially when they challenge the norms associated with the social category they embody (Sobieraj, 2020). Second, the kiss occurred during the most important musical event and TV show in Italy, which has a crucial symbolic role for many social issues, including controversies over populistic discourse, and is widely followed across generational lines (Magudda, 2020), so that it is possible to track the reactions of younger and older individuals to the same, contentious event. Importantly, many online reactions emerged immediately, because Sanremo is one of the most commented television programs through 'dual-screen'. This practice, which consists in engaging with online content while watching television, often reflect the identity narratives of online audiences (Airoldi, 2016). Third, taken together, Rosa Chemical's actions (kiss performance, contest song, and social media posts) constituted a kind of gender-normative defying behavior which is often targeted by online hate speech (Ging & Siapera, 2019; Butler, 2024). This is even more relevant in the Italian context, where the so-called 'anti-gender' movements, proclaiming ultraconservative views on women's and LGBTQ+ rights, are particularly strong (Paternotte & Kuhar, 2017; Garbagnoli & Prearo, 2018; Prearo, 2024).

We exploit this case to address our research questions. To do so, we analyze the number and the type of comments that social media users posted below the visual content shared by Rosa Chemical in the immediate aftermath of the kiss, as well as the number of reactions generated by such comments, in order to investigate whether similar content posted on different social platforms provoked different degrees of hostility.

5. Empirical strategy

We manually collected users' comments to the posts produced by Rosa Chemical during the time frame 2023-02-11 to 2023-02-12. In total, we collected 7 posts (2 on Facebook, 3 on Instagram and 2 on TikTok) reporting

[4] See https://www.fanpage.it/politica/perche-il-bacio-tra-rosa-chemical-e-fedez-sul-palco-di-sanremo-non-e-solo-spettacolo/

[5] Specifically, we restrict our analytical window to the content posted on February 11th to 12th, 2023.

[6] More specifically, Rosa Chemical is dressed in latex and is held on a leash and Alex Mucci is wearing a veil shirt that allows her breasts to be seen. The pictures were taken at daytime in everyday public places in Milan (shopping galleries, subway, etc.).

about 10k users' comments. We merged the two Facebook posts because they were two identical photo galleries shared within a few minutes. We removed from the analysis 4 Facebook comments without text.
To study the production of online hate speech, we analyzed the number and type of these comments. For this purpose, we automatically analyzed the text through several deep learning techniques, trained to the Italian language, to identify: (a) sentiment polarity (negative or positive); (b) basic emotions (fear, joy, sadness, anger), following Bianchi et al. (2021); and (c) hate speech (dichotomous classification), following Nozza et al. (2022). Indeed, as the literature has shown, hate speech is difficult to identify automatically and is often accompanied by negative feelings and emotions (Siegel, 2020). We then manually checked classified comments to correct misclassification. After the manual check, we concluded that the classified emotions were joy-enthusiasm, anger-disgust, and sadness-disappointment. To explore the content of comments more in depth, we preprocessed the text removing URL, non-ASCII characters, etc., and we used the R library SpacyR to perform POS tagging and identify relevant lemmas (nouns, proper nouns, adjectives, and verbs).
Next, in order to study the reactions to online hate speech, we tested the association between the type of comment (as identified in the previous steps) and the number of responses to it. We did so in a multivariate regression framework, estimating a generalized-linear-model with negative binomial distribution to account for the highly skewed distribution of the outcome (>90% of comments produced zero responses) (Hilbe, 2011).

6. Results

Table 2 reports the posts collected, along with some metadata. As apparent from the last column, Rosa Chemical is more popular on Instagram than on other platforms, especially Facebook, based on number of followers. This is obviously reflected in the number of interactions (N Likes, N comments) produced by the posts. It would therefore be misleading to compare the absolute number of interactions across platforms.
Within platforms, however, some interesting patterns emerge. The Kiss post produced relatively few Likes (almost 50 percent fewer than those produced by the post featuring Alex Mucci), but it was by far the one most commented on. This apparent contradiction can be explained by the different meanings associated with the two interactions: Likes express acknowledgment or approval, while comments often indicate debates or arguments on topics (Martella & Bracciale, 2021).
If we calculate the ratio between comments and Likes, Facebook emerges as the platform with the highest number of comments per Like to the same posts (Facebook posts ratio: 0.36 and 0.82; Kiss on Instagram: 0.03), potentially indicating a larger degree of contention over the posts. Concerning the ability to engage the community in commenting (N comments / N followers), the second gallery on Facebook and the kiss on Instagram reached the highest ratio (0.02) among all posts. Finally, on analyzing the ability to engage the community (N Likes / N followers), it emerges that Rosa Chemical was more able to gain likes on Instagram and TikTok compared to Facebook, especially with the post featuring Alex Mucci (N Likes / N followers: 1.34). Therefore, this overview of the posts seems to highlight that Facebook audiences were overall more willing to comment on Rosa Chemical's posts, while Instagram and TikTok users liked his posts the most. Furthermore, the Kiss post – which unfortunately only exists on Instagram – fueled a large number of comments, suggesting that this content was more controversial than those of the other posts.

TABLE 2 HERE

6.1 Hate and emotions across platforms: quantitative analyses

Table 3 reports the shares of comments classified in terms of their hate speech content, sentiment, and prevalent emotion. Comments on the Facebook posts show the highest percentages of hate (36%), negative sentiment (65%), and anger-disgust (54%). Instagram comments are the ones that expressed most joy or

enthusiasm (72%). TikTok stands out as the least negative platform (12%) and with the lowest percentage of hate (2%). In terms of emotions, joy and enthusiasm prevailed (68%) and comments were more sad or disappointed (5%) than angry or disgusted (3%). These patterns reflect the relationship between hate and emotions. Indeed, on Facebook hate is rather related to anger and disgust (ρ=0.55) and slightly negatively correlated with sadness and disappointment (ρ= - 0.09). Comments on Instagram show a similar correlation between anger and hate but they are not associated with any specific emotion. TikTok comments, on the other hand, exhibit a less intense relationship between anger and hate (ρ=0.49), which, compared to other social media, is somewhat more related to sadness (ρ=0.07). Finally, moving from Facebook to TikTok shows that negative sentiment is less and less connected to hate speech, with Pearson's coefficient varying from ρ=0.51 (Facebook) to ρ=0.45 (Instagram) and ρ=0.3 (TikTok).
Thus, generally speaking, hate is much more present on Facebook than on other platforms, and it is mainly connected to anger. Moreover, Facebook audiences express far less joy – only one third of their comments do so. On the other hand, TikTok emerges as the least 'hateful' environment, with most of the negative sentiment being related to sadness or disappointment.

TABLE 3 HERE

6.2 Qualitative exploration of the content of comments

To understand the underlying emotions expressed by individuals, we extracted the lemmas most frequently used in the comments on each post. We present the results by prevalent emotion.

Anger - Disgust

On Facebook, the most common words associated with anger was the verb 'do' (fare) accompanied by the noun 'suck' (schifo), which together convey the sentiment 'you suck'. Additionally, adjectives such as 'ridiculous' and 'asshole' were prevalent. Anger was also linked to the verb 'to see', referring to the Sanremo broadcast, as well as to the general discussion on the (stolen or performed) kiss and twerking by Rosa Chemical on Fedez. Indeed, the word 'Fedez' was also quite frequent, even though the post was not about the kiss, indicating that the users visited Rosa Chemical's webpage with the intention of commenting on the controversial kiss episode and gathered below the latest post.
Moreover, users' criticisms of the entire performance mainly revolved around scandal rather than music, disqualifying Rosa Chemical as a musician.
On Instagram, anger related to different issues depending on the post being referred to. Comments on the post featuring the kiss were quite similar to the ones on Facebook. Here the words 'kiss' (both noun and verb) and 'Italy' are among the most frequent lexical items, arguing that with his behavior, Rosa Chemical does not represent the country, contrary to what his contest song claimed ('Made in Italy'). Among recurrent words, we also find 'shit' and 'wife', referring to Fedez being married to Chiara Ferragni, who was hosting Sanremo. In contrast, anger in comments on the post with OnlyFans content creator Alex Mucci mostly related to the word 'change', challenging the claim, present in the post caption, that the photo gallery of the couple, wearing fetish dresses and exposing the female body in public places, actually constituted social change. In other instances, users wrote that the alleged 'change' only generated disgust. Among the most frequent words were 'woman', 'naked', 'tits', 'exhibit', 'freedom' and 'sucks' ('fare schifo'). This result highlights that almost all the anger was directed at the creator, Alex Mucci, who appeared undressed in the photos in public places showing her breasts. Many angry comments point out that this was not true freedom and that Mucci is a 'plastic surgery woman'. Angry comments related to the post featuring Rose Villain were very few, as were the frequency of words. However, some criticism of both Rosa Chemical and Rose Villain emerges here because they 'suck' and represent a 'cliché' and a 'carnival' instead of 'sex'.

Finally, on TikTok only the comments to the post featuring Alex Mucci expressed some anger, which was mainly directed at both her and Rosa Chemical. Interestingly, both criticisms seem to be motivated by an alleged violation of the social norms connected with relational monogamy. On the one hand, comments attacked Rosa Chemical for ruining the Ferragnez 'family' by kissing Fedez on stage. On the other hand, users argued that the photo gallery where Mucci had Rosa Chemical on a leash suggested that she was cheating on the Italian celebrity 'Marra', to whom she is engaged.

Sadness - Disappointment

On Facebook, disappointment and sadness mostly related to the 'song', the singer and 'Fedez' because they do not represent 'Italian sex', he is not an 'artist' and he sucks (fare pena). This is coherent with the caption openly referring to the contest song ('Sanremo is not only the Italian Music Festival. It is also the Italian Sex Festival').

On Instagram, starting with the Kiss post, sadness concerned the kiss not being consensual and Fedez being heterosexual and married. The most frequent words related to sadness and disappointment on the post featuring Alex Mucci were quite similar to the ones that emerged in posts expressing anger. However, here the emotional tone was more disappointed and saddened by the 'naked' 'woman' showing her 'breasts' despite the 'cold' and who did not represent any 'change'. Very few comments were sad about the Instagram post featuring Rose Villain, and they mostly claimed that she is 'boring' rather than provocative.

On TikTok, the sad comments were not necessarily negative; on the contrary, many users expressed sadness or disappointment because they were 'jealous' of Mucci or Rosa Chemical, while others pointed out that the 'Made in Italy' song should have won Sanremo or would have been successful in the Eurovision Song Contest. Most disappointed comments related to what had happened with 'Fedez'.

Joy – Enthusiasm

Enthusiastic comments on Facebook mainly related to the 'song' 'Made in Italy', expressing appreciation ('bravo', 'bravissimo', etc.). Therefore, once again, it seems that on Facebook emotions and users' expressions dealt with the post content to a lesser extent.

On Instagram, joyful comments on the Kiss post clearly expressed 'love' of 'Rosa' Chemical and 'Fedez' also because of the 'kiss' ('kissing', 'mouth'), to the extent that there were many users tagging others to share the gallery. Many users expressed their appreciation by using online slang expressions (e.g. 'you are my father', 'mother', 'brother'). Joy concerning the gallery featuring Alex Mucci was mainly expressed by people tagging others and commenting 'wonderful', 'good looking', 'sexy', 'hot', etc. Hence, users did not seem to follow up on the call for social change mentioned in the caption text. Comments on the gallery featuring Rose Villain expressed joy through puns on the names 'Rosa' and 'Rose' and by saying 'you are my mother', 'father', 'parent', you represent 'Italian sex' and we 'love' you.

Also on TikTok, enthusiasm was mainly expressed by tagging. Appreciation of the post where the contest song was dubbed in sign language was directed towards the 'interpreter' and the 'song'. The video with Alex Mucci is mainly commented on with the above-mentioned expressions: you are my 'mother', 'father', 'brother', 'family tree', and 'chiropodist', with this last referring to Rosa Chemical's obsession with feet.

In sum, how do these analyses respond to our first research question, the one concerning hate speech creation and its differences across social media platforms? The number of hateful comments varied dramatically among platforms, ranging from more than one third on Facebook to less than 2% on TikTok. Hate was most related to anger and disgust in every platform, but this relationship was weaker on TikTok, where hate was also related to sadness and disappointment. Television content was a driver of hate comments on Facebook, as revealed by the analysis of the most frequent lemmas, which did not relate to post content, but rather to the kiss incident. However, it is worth bearing in mind that Rosa Chemical did not post anything else on Facebook in those days.

Interestingly, the most hated post on Instagram and TikTok was the one featuring Alex Mucci, and this hate was mainly driven by sadness and disappointment. Mucci's physical appearance and the way she dresses

provoked most of the hate and disappointment. However, comments on the post featuring the kiss between Fedez and Rosa Chemical (only available on Instagram) emerged as the angriest and most disgusted ones, thus highlighting that homophobic discourse and sexist discourse generate the same percentage of hate but differ in the kinds of emotions that lead to hate.

6.3 The effects of hate and emotions on online debates

In order to understand whether and to what extent hate or other emotions trigger online reactions, we built four negative binomial regression models: one for each social media and one general (Table 4).

TABLE 4 HERE

All models show that positive emotions, namely joy and enthusiasm, do not affect the number of responses; indeed, the IRR is not significant across platforms. For Facebook and Instagram, a quite clear pattern emerges from our results, one which is largely in line with the findings of previous research. Only negative emotions increase responses to comments. Hate speech is positively correlated to the number of responses, but, somewhat surprisingly, this correlation disappears once emotions are controlled for, suggesting that negative emotions, and in particular anger, mediate the effect of hate speech on engagement.[7] Among negative emotions, sadness and disappointment in particular appear to be strongly related with responses to comments. These results confirm those of previous studies on negativity on Facebook (Bene, 2017; Martella & Bracciale, 2021) and that positive emotions do not foster commenting by Facebook users (Martella & Bracciale, 2021). However, our results extend this finding to Instagram. Moreover, our analyses disentangle the influence of various negative emotions, something which – to the best of our knowledge – has not yet been done in the literature. We show that sadness and disappointment powerfully induce users to respond to comments. Indeed, except for TikTok, IRRs of sadness and disappointment are higher than those of anger. By contrast, TikTok audiences exhibit a behavior that is very different from that of the users of the other platforms, and from what has been found by previous studies relying on traditional platforms. First, anger and disgust – which triggered more users' responses on both Facebook and Instagram – were not even significant on TikTok, despite the presence of a similar post featuring Alex Mucci. More importantly, sadness and disappointment *negatively* affected the number of responses, in the sense that comments imbued with these negative emotions received about 20% fewer responses ($e^{\beta} = 0.2$).

7. Discussion and conclusion

In this chapter, we have explored the processes that lead to the creation of online hate speech and its consequences by analyzing a case where, during the most important Italian music festival, a male singer kissed another male singer. This act challenged traditional gender norms and produced polarized reactions in the public debate, making the singer a likely target of the hate speech which proliferates around public figures who push the boundaries of their social identity (ElSherief et al., 2018; Sobieraj, 2020). Moreover, gender norms are a particularly polarizing topic in Italy (e.g. Paternotte & Kuhar, 2017). We analyzed online reactions to the event on the three social media platforms where the singer is active (Facebook, Instagram, and TikTok). Specifically, we collected users' reactions below the five posts that the singer made right after the event, and we analyzed the number and the type of comments and the number of responses.

It should be borne in mind that, among the various posts made by the singer Rosa Chemical within our selected time frame, only one, made on Instagram, was about the contentious 'kiss' episode. Yet, as evident

[7] Due to possible multicollinearity among anger, hate, and joy, we run several regression models excluding anger. Our results confirm that without anger, both hate and joy are significant ($p<0.01$) but hate IRR is lower than anger ($e^{\beta} = 1.8$) and joy IRR is negative, thus showing that: 1) anger is a more powerful driver than hate for commenting and 2) positive emotions decrease the number of responses ($e^{\beta} = 0.3$).

from our exploration of the most frequent lemmas, many hateful comments on Facebook directly referred to the episode. This result has two implications. First, it highlights the importance of traditional media content in generating online hate speech through 'dual screening' (television/smartphone or computer). Second, it suggests that homophobic attitudes are a significant driver of hateful discourse. This is further backed up by the fact that the 'kiss' post, compared to the other two posted on Instagram, produced the largest share of comments characterized by Anger and Disgust (20.3%). Yet, the 'kiss' post and the post with the (female) OnlyFans creator were met by a similar share of hateful comments on Instagram, suggesting that, at least in Italy, sexism is just as important as homophobia in generating online hate speech (see also Martella & Pavan, 2023).

Moving to the cross-platform comparison, our findings indicate that hateful comments were definitely not common on TikTok (1.9%), rare on Instagram (7.6%), and very frequent on Facebook (35.9%). This ranking holds true also when we narrow the comparison down to identical or very similar posts that Rosa Chemical simultaneously made on various social platforms. Similarly, the general sentiment was more negative than positive on Facebook, while the reverse is true for the other social platforms. More specifically, on Facebook the prevailing emotions were Anger and Disgust (53.7%), while on Instagram and TikTok they were Joy and Enthusiasm (71.9% and 68%, respectively).

The stronger presence of hateful and, more generally, negative reactions on Facebook cannot be traced back to the nature of the platform. In fact, Facebook is strongly based on users' personal networks, while Instagram and TikTok are mostly designed and used as 'content communities' (Bhandari & Bimo, 2022) and, as such, offer more opportunities for aggressive behavior to anonymous users. However, TikTok differs from other platforms in its content distribution mechanism, which is highly personalized and disconnected from the relationships between users. Therefore, one factor that may have affected the lower negativity present on the platform is that the content was seen mostly by like-minded users. It is still true that the content was also searched for on the platforms, as evidenced by the comments on Facebook that did not address the content of the post. Therefore, it seems more likely that Facebook audiences, who tend to be older than those of Instagram and TikTok, feel more threatened by gender-normative defying behaviors and resort to hate speech as a form of cultural backlash (Inguanzo et al., 2021).

The socio-demographic composition of the different social platforms is probably strictly intertwined with the prevailing social norms about which type of content 'belongs' to the platform, and with which one can engage, and which type of content, instead 'does not belong' and therefore should be ignored (or possibly reported as inappropriate). In fact, our analyses on second-order reactions produced a different pattern of results for Facebook and TikTok (with Instagram somewhere in between). On Facebook, comments characterized by negative emotions produced more reactions, thus replicating the results of previous studies (Bene, 2017; Martella & Bracciale, 2021). In contrast, on TikTok, comment negativity was not correlated with the number of responses. This result is consistent with qualitative insights on adolescent online behavior, indicating unwillingness to participate in hatred disputes in the comments section (Astuti & Partini, 2019; Borgna et al., 2024), but also, more generally, with the perceived appropriateness of positive / negative content in different platforms (Waterloo et al., 2018).

In conclusion, although recent evidence indicates that TikTok is also beginning to be slightly affected by the spread of hatred, particularly against minorities and women by far-right wingers and incel movements (Weimann & Masri, 2023), our case study centered on a gender-norm defying episode points to Facebook and, to a lesser extent, Instagram, as the social media where more hatred is expressed and generates further engagement.

From a methodological perspective, our study has sought to contribute to investigation of the 'supply side' of hate speech, which has so far mainly relied on self-reported behavior based on survey data. As we have shown, comparing the spread of online hate speech across social media platforms is promising because they differ in their architecture (typical content, social norms, moderation policies), as well as in the prevailing socio-demographics of their users. However, an obvious limitation of our approach, as opposed to survey-based research, is that we lack direct information on users' socio-demographics. Therefore, the

interpretation of cross-platform differences as indicative of different behaviors prevailing among younger vs. older online audiences remains, at this stage, a speculation. Moreover, as mentioned, we cannot exclude that the very low presence of hate and negativity on TikTok is due to its content distribution mechanism. Finally, our study suggests multiple avenues for further research. In particular, future studies could investigate the possible reasons why adolescents refrain from engaging in online hate speech, by exploring the role of values associated with social change (e.g. gender fluidity and sexual freedom) and the role of online social norms.

References:

Airoldi, M. (2016). Studiare i «social media» con i «topic models»: Sanremo 2016 su Twitter. *Studi Culturali*, 3(2016), 431–448. https://doi.org/10.1405/85342

Álvarez-Benjumea, A., & Winter, F. (2018). Normative change and culture of hate: An experiment in online environments. *European Sociological Review*, *34*(3), 223-237.

Astuti, F., & Partini, P. (2019). The Hate Speech Behavior of Teenagers on Social Media Instagram. *International Summit on Science Technology and Humanity (ISETH2019) Advancing Scientific Thought for Future Sustainable Development*, 252–259.

Barlińska J, Szuster A, Winiewski M (2018) Cyberbullying among adolescent bystanders: role of affective versus cognitive empathy in increasing prosocial cyberbystander behavior. *Frontiers in Psychology,* 9. https://doi.org/10.3389/fpsyg.2018.00799

Bene, M. (2017). Go viral on the Facebook! Interactions between candidates and followers on Facebook during the Hungarian general election campaign of 2014. *Information, Communication & Society*, 20(4), 513–529. https://doi.org/10.1080/1369118X.2016.1198411

Bernatzky, C., Costello, M., & Hawdon, J. (2022). Who produces online hate?: An examination of the effects of self-control, social structure, & social learning. *American journal of criminal justice*, *47*(3), 421-440.

Bhandari, A., & Bimo, S. (2022). Why's Everyone on TikTok Now? The Algorithmized Self and the Future of Self-Making on Social Media. *Social Media + Society*, 8(1).

Bianchi, F., Nozza, D., & Hovy, D. (2021). FEEL-IT: Emotion and Sentiment Classification for the Italian Language. In O. De Clercq, A. Balahur, J. Sedoc, V. Barriere, S. Tafreshi, S. Buechel, & V. Hoste (Eds.), *Proceedings of the 11th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis* (pp. 76–83). Association for Computational Linguistics.

Bilewicz, M., & Soral, W. (2020). Hate speech epidemic. The dynamic effects of derogatory language on intergroup relations and political radicalization. *Political Psychology*, *41*(S1), 3-33.

Blazak, R (2009). Toward a Working Definition of Hate Groups. *Hate Crimes, 3*(1): 133–162.

Borgna, C., Charitopoulou, E., Miglio, M. (2024) Online hate speech and prejudice among adolescents: a field experiment on Italian schools [Unpublished manuscript].

Butler, J. (2024). *Who's Afraid of Gender?*. Farrar, Straus and Giroux.

Costello, M., & Hawdon, J. (2018). Who are the online extremists among us? Sociodemographic characteristics, social networking, and online experiences of those who produce online hate materials. *Violence and gender*, *5*(1), 55-60.

Denti, D., & Faggian, A. (2021). Where do angry birds tweet? Income inequality and online hate in Italy. *Cambridge Journal of Regions, Economy and Society*, *14*(3), 483-506.

ElSherief, M., Kulkarni, V., Nguyen, D., Wang, W. Y., & Belding, E. (2018, June). Hate lingo: A target-based linguistic analysis of hate speech in social media. In *Proceedings of the international AAAI conference on web and social media* (Vol. 12, No. 1).

Eschmann, R (2021). Digital Resistance: How Online Communication Facilitates Responses to Racial Microaggressions. *Sociology of Race and Ethnicity*, *7*(2), 264-277.

European Union Agency for Fundamental Rights (2023). *Online Content Moderation – Current challenges in detecting hate speech*. Luxembourg: Publications Office of the European Union.

Fortuna, P., & Nunes, S. (2018). A survey on automatic detection of hate speech in text. *ACM Computing Surveys (CSUR)*, *51*(4), 1-30.

Frischlich, L., Schatto-Eckrodt, T., Boberg, S., & Wintterlin, F. (2021). Roots of incivility: How personality, media use, and online experiences shape uncivil participation. *Media and Communication*, *9*(1), 195-208.

Gandhi, A., Ahir, P., Adhvaryu, K., Shah, P., Lohiya, R., Cambria, E., Poria, S., & Hussain, A. (2024). Hate speech detection: A comprehensive review of recent works. *Expert Systems*, *41*(8). https://doi.org/10.1111/exsy.13562

Garbagnoli, S., & Prearo, M. (2018). *La crociata" anti-gender". Dal Vaticano alle manif pour tous* (pp. 1-84). Kaplan.

Ging, D., & Siapera, E. (Eds.). (2019). *Gender hate online: Understanding the new anti-feminism*. Springer.

Hawdon, J, Bernatzky, C, & Costello, M. (2019). Cyber-routines, political attitudes, and exposure to violence-advocating online extremism. *Social Forces*, *98*(1), 329-354.

Hawdon, J., Oksanen, A. & Räsänen, P. (2017). Exposure to Online Hate in Four Nations: A Cross-National Consideration, *Deviant Behavior*, *38*(3), 254-266.

Karizat, N., Delmonaco, D., Eslami, M., & Andalibi, N. (2021). Algorithmic Folk Theories and Identity: How TikTok Users Co-Produce Knowledge of Identity and Engage in Algorithmic Resistance. *Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2)*, 1–44.

Inara Rodis Paulina d. C. (2021). Let's (Re)Tweet about Racism and Sexism: Responses to Cyber Aggression toward Black and Asian Women. *Information, Communication & Society 24*(14), 2153–73.

Inguanzo, I., Zhang, B., & Gil de Zúñiga, H. (2021). Online cultural backlash? sexism and political user-generated content. *Information, communication & society*, *24*(14), 2133-2152.

Isbister, T., Sahlgren, M., Kaati, L., Obaidi, M., & Akrami, N. (2018). Monitoring targeted hate in online environments. arXiv.org. https://arxiv.org/abs/1803.04757

Kaakinen, M., Räsänen, P., Näsi, M., Minkkinen, J., Keipi, T., & Oksanen, A. (2018). Social capital and online hate production: A four country survey. *Crime, Law and Social Change*, *69*, 25-39.

Magudda, P. (2020). Populism, music and the media. The Sanremo Festival and the Circulation of Populist Discourse. *PArtecipazione e COnflitto*, *13*(1), 132–153.

Martella, A., & Bracciale, R. (2022). Populism and emotions: Italian political leaders' communicative strategies to engage Facebook users. *Innovation: The European Journal of Social Science Research, 35*(1), 65–85.

Martella, A., & Cepernich, C. (2024). "Dacci oggi il nostro TikTok quotidiano". Strategie di pubblicazione dei quotidiani italiani su Tik Tok. *Problemi Dell'informazione, Rivista Quadrimestrale*, *2024*(1), 65–92.

Martella, A, & Pavan, E (2023). "We hate her . . . and you too": Polarized intersectionality in Italy throughout changing political scenarios. *New Media & Society*, 14614448231160706.

Masciantonio, A., & Bourguignon, D. (2024). Too Positive to Be Tweeted? An Experimental Investigation of Emotional Expression on Twitter and Instagram. *Media Psychology*, *27*(2), 243–270. https://doi.org/10.1080/15213269.2023.2236935

Meltwater, & We Are Social. (2023). Digital 2023 Global Overview Report. https://datareportal.com/reports/digital-2023-global-overview-report

Munger, K. (2021). Don't@ Me: Experimentally reducing partisan incivility on Twitter. *Journal of Experimental Political Science*, *8*(2), 102-116.

Nozza, D.,Bianchi, F., Attanasio, G. (2022). HATE-ITA: New Baselines for Hate Speech Detection in Italian. In Proceedings of the Sixth Workshop on Online Abuse and Harms (WOAH), pages 252–260, Seattle, Washington (Hybrid): Association for Computational Linguistic.

PEW Research Center (2022). “Internet, smartphone and social media use”, Report December 6 2022. Retrieved from https://www.pewresearch.org/global/2022/12/06/internet-smartphone-and-social-media-use-in-advanced-economies-2022/ on May 5 2024.

PEW Research Center (2024): “Americans’ Social Media Use”, Report January 31 2024. Retrieved from https://www.pewresearch.org/internet/2024/01/31/americans-social-media-use/ on May 5 2024.

Paternotte, D., & Kuhar, R. (2017). The anti-gender movement in comparative perspective. In R. Kuhar & D. Paternotte (Eds.), *Anti-gender campaigns in Europe: Mobilizing against equality* (pp. 253–276). Rowman & Littlefield International.

Prearo, M. (2024). *Anti-gender Mobilizations, Religion and Politics: An Italian Case Study*. Taylor & Francis.

Reichelmann, A., Hawdon, J., Costello, M., Ryan, J., Blaya, C., Llorent, V., Oksanen, A., Räsänen, P. & Zych, I. (2021). Hate knows no boundaries: Online hate in six nations. *Deviant Behavior*, *42*(9), 1100-1111.

Ribeiro, M., Calais, P., Santos, Y., Almeida, V., & Meira Jr, W. (2018, June). Characterizing and detecting hateful users on twitter. In *Proceedings of the International AAAI Conference on Web and Social Media* (Vol. 12, No. 1).

Rosenbusch, H., Evans, A. M., & Zeelenberg, M. (2020). Interregional and intraregional variability of intergroup attitudes predict online hostility. *European Journal of Personality*, *34*(5), 859-872.

Schemer, C, & Reiners, L (2023). Challenges of comparative research on hate speech in media user comments: Comparing countries, platforms, and target groups. In C. Strippel, S. Paasch-Colberg, M. Emmer, & J. Trebbe (Eds.), Challenges and perspectives of hate speech research (pp. 127-139). Berlin.

Schmidt, A., & Wiegand, M. (2017). A survey on hate speech detection using natural language processing. In *Proceedings of the fifth international workshop on natural language processing for social media* (April, pp. 1-10).

Siegel, A. (2020). Online Hate Speech. In N. Persily & J. A. Tucker (Eds.), *Social Media and Democracy. The State of the Field, Prospects for Reform* (pp. 56–88). Cambridge University Press.

Sobieraj, S(2020) *Credible Threat. Attacks against women online and the future of democracy*. Oxford University Press.

Stahel, L., & Weingartner, S. (2024). Can Political Orientation Explain the Social Structure of Online Aggression? Integrating Social Media and Survey Data. *Sociological Inquiry*, *94*(1), 149-169.

Vargo, C. J., & Hopp, T. (2017). Socioeconomic status, social capital, and partisan polarity as predictors of political incivility on Twitter: A congressional district-level analysis. *Social Science Computer Review*, *35*(1), 10-32.

Wachs, S., Gámez-Guadix, M., & Wright, M. F. (2022). Online hate speech victimization and depressive symptoms among adolescents: The protective role of resilience. *Cyberpsychology, Behavior, and Social Networking*, *25*(7), 416-423.

Waterloo, S. F., Baumgartner, S. E., Peter, J., & Valkenburg, P. M. (2018). Norms of online expressions of emotion: Comparing Facebook, Twitter, Instagram, and WhatsApp. New Media & Society, *20*(5), 1813–1831. https://doi.org/10.1177/1461444817707349

We Are Social & Meltwater (2023). Digital 2023 Global Overview Report. Retrieved from https://datareportal.com/reports/digital-2023-global-overview-report on May 5 2024.

Weimann, G., & Masri, N. (2023). Research Note: Spreading Hate on TikTok. *Studies in Conflict & Terrorism*, *46*(5), 752–765. https://doi.org/10.1080/1057610X.2020.1780027

INSERT ANY TABLES / FIGS / PICTURES / ILLUSTRATIONS / TABLES HERE (or attach them in separate documents)

Notes for Chapter Authors:
Please ensure that each Table/Fig/Picture/Illustration you use is fully sourced. Where you have used a Table/image that has been published before (this can be online or in another publication) please ensure that you obtain a written permission for use in this publication prior to submitting your chapter. If in doubt as to whether you require a permission, please ask your Volume Editors.

Table 1. Comparability of content posted by Rosa Chemical across platforms

	Facebook	Instagram	TikTok
Kiss (picture)		X	
Song dubbed in sign language (video)			X
Feat. Alex Mucci, fetish dresses (pictures)		X	
Feat. Alex Mucci, bathrobes (video)			X
Feat. Rose Villain, bathrobes (pictures)	X	X	

Table 2. Posts and reactions across platforms

Platform	URL[1]	Post	date	N. Likes	N. comments[2]	Retrieved comments	Followers
Facebook	https://www.facebook.com/RosaChemical/posts/pfbid02vHZUT5tuvmvsTTsMq58DrRL4HmHM2KyfjhLvzMiWnDWX1i22RmaBaFzsFB3KRpywl	Feat. Rose Villain, bathrobes (pictures)	Feb 11	745	270	341	22,226
	https://www.facebook.com/RosaChemical/posts/pfbid0ufQvDDyg4x97763VCvdZDgTVLEoKeyh7cvtGSAVgpTb8DTFVcjkaPKNyHuhkYhZ6l		Feb 11	609	498		
Instagram	https://www.instagram.com/p/CokjauWKTKY/?img_index=1	Feat. Alex Mucci, fetish dresses (pictures)	Feb 12	552,000	4,768	2753	411,000
	https://www.instagram.com/p/CoifZqQLWAl/?img_index=1	Kiss (picture)	Feb 11	288,000	8,893	3822	
	https://www.instagram.com/p/CogB2CGq4vk/?img_index=1	Feat. Rose Villain, bathrobes (pictures)	Feb 11	269,000	1,793	1507	
TikTok	https://www.tiktok.com/@gipsyboirosa/video/7199002417428499718	Song dubbed in sign language (video)	Feb 11	211,600	896	581	309,800
	https://www.tiktok.com/@gipsyboirosa/video/7199339655605587206	Feat. Alex Mucci, bathrobes (video)	Feb 12	204,200	959	629	

Notes: [1]Unfortunately, as we are writing this chapter, the Instagram posts are not visible anymore because Rosa Chemical removed almost all the content prior to 2024. [2]The number of comments provided by the platform include nested comments that have been not retrieved for the analysis

Table 3. Hate, sentiment and emotion distribution by social media

Social	**N**	**Hate speech**	**Sentiment**	**Emotion[1]**		
		% of hateful comments	% of negative comments	% of comments Anger/Disgust	% of comments Sadness/Disappointment	% of comments Joy/Enthusiasm
Facebook	337	**35.9**	**64.7**	**53.7**	9.5	30.3
Feat. Rose Villain, bathrobes (pictures)	337	35,9	64,7	53,7	9,5	30,3
Instagram	8082	7.6	27.9	16.8	7.7	**71.9**
Feat. Rose Villain, bathrobes (pictures)	1507	3.8	12.5	7.7	3.6	85.1
Feat. Alex Mucci, fetish dresses (pictures)	2753	8.6	30.8	16.8	11.2	67.5
Kiss (picture)	3822	8.4	31.9	20.3	6.9	69.9
TikTok	1210	1.9	12.1	2.8	5.5	68.0
Feat. Alex Mucci, bathrobes (video)	629	2.9	13.5	4.3	7.9	66.8
Song dubbed in sign language (video)	581	0.9	10.7	1.2	2.8	69.4

Note: [1]Not all comments expressed the emotions under analysis; therefore, the total sum of emotions does not equal 100.

Table 4. Negative binomial regression models on the number of responses to comments (odds ratios).

	General	**Facebook**	**Instagram**	**TikTok**
Predictors	*Incidence Rate Ratios*	*Incidence Rate Ratios*	*Incidence Rate Ratios*	*Incidence Rate Ratios*
(Intercept)	0.45 *	0.24 *	0.47 *	0.37 ***
Hate	1.47	0.89	1.54	0.54
Anger – Disgust	2.58 ***	5.68 **	2.24 *	0.56
Sadness – Disappointment	3.61 ***	15.43 ***	3.25 **	0.21 *
Joy- Enthusiasm	0.66	1.36	0.55	0.89
Social [Instagram]	0.89			
Social [TikTok]	0.93			
Observations	9633	341	8082	1210
R^2 Nagelkerke	0.091	0.166	0.092	0.021
AIC	9515.227	753.921	7131.163	1448.858
log-Likelihood	-4749.614	-370.960	-3559.581	-718.429

** p<0.05 ** p<0.01 *** p<0.001*