

Hateful and Other Negative Communication in Online Commenting Environments: Content, Structure and Targets

Vasja Vehovar , Dejan Jontes 

Faculty of Social Sciences, University of Ljubljana, Kardeljeva pl. 5, 1000 Ljubljana, Slovenia

Corresponding author: Vasja Vehovar (vasja.vehovar@fdv.uni-lj.si)

Abstract

Information and communication technologies are increasingly interacting with modern societies. One specific manifestation of this interaction concerns hateful and other negative comments in online environments. Various terms appear to denote this communication, from flaming, indecency and intolerance to hate speech. However, there is still a lack of an umbrella term that broadly captures this communication. Therefore, this paper introduces the concept of socially unacceptable discourse, which serves as the basis for an empirical study that evaluated online comments scraped from the Facebook pages of the three most-visited Slovenian news outlets. Machine-learning algorithms were used to narrow the focus to topics related to refugees and LGBT rights. Ten thousand comments were manually coded to identify and structure socially undesirable discourse. The results show that about half of all comments belonged to this type of discourse, with a surprisingly stable level and structure across media (i.e., right-wing versus mainstream) and topics. Most of these comments could also be considered a potential violation of hate speech legislation. In the context of these findings, the political and ideological consequences and implications of mediated emotions are discussed.

Keywords

Facebook; Online comments; Socially unacceptable discourse; LGBT; Refugees; Social media; Social informatics; Emotional audience response.

Citation: Vehovar, V., & Jontes, D. (2021). Hateful and Other Negative Communication in Online Commenting Environments: Content, Structure and Targets. *Acta Informatica Pragensia*, 10(3), 257–274. <https://doi.org/10.18267/j.aip.165>

Special Issue Editors: Vasja Vehovar, University of Ljubljana, Slovenia
Zdenek Smutny, Prague University of Economics and Business, Czech Republic
Alice R. Robbin, Indiana University, USA

Academic Editor: Stanislava Mildeova, University of Finance and Administration, Czech Republic

Copyright: © 2021 by the author(s). Licensee Prague University of Economics and Business, Czech Republic.

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution License (CC BY 4.0).

1 Introduction

The interaction between information and communication technologies (ICTs) and society is one of the major challenges of modern social science research. The literature addresses this interaction from various angles, such as information society, social computing, science and technology studies, internet studies, computer-mediated communication, human–computer interaction and many others. In this context, the term *social informatics* can serve as the broadest umbrella term that covers all types of this interaction, whether at the macro (society, nation), micro (organisation, community, group) or individual (person) levels (Smutny and Vehovar, 2020). The changes that ICTs bring to modern societies¹ at these three levels are particularly noticeable in the modification of human communication patterns. Computer-mediated communication has led to profound differences in the ways people interact.

One specific and particularly critical aspect of these processes is the deterioration in the quality of online communication, which Quandt (2018) termed *dark participation*. Proposing an alternative approach to understanding this discourse, Rossini (2020, p. 2) warned that ‘little is known about the extent to which online discourse is inherently toxic, or whether it is simply characterised by rudeness and profanities’. More importantly, Wahl-Jorgensen (2020) claims that the expanded opportunities for participation have contributed to the questioning of traditional distinctions between news audiences and producers and have led to new forms of emotional expression that have spilled over into news production practices.

Lünenborg and Maier (2018, p. 1) similarly argued that phenomena such as online hate speech² and ‘shitstorms’ via social media should be understood as explicit public articulations of emotions. Orgeret (2020) emphasised that emotions have received significant new attention in the online environment. As Döveling et al. (2018, p. 1) argued, insight into this complex spectrum of mediatised emotions is important because we witness a stream of online effects that resonate with political campaigns, terrorist attacks, natural disasters, celebrity deaths, etc. In this context, the normative ideal of public communication, in which citizens participate in commenting on news, exchanging arguments and expanding their knowledge, seems too optimistic. This is also true of the vision of participatory journalism, in which citizens share their expertise and invest their time in supporting news media at different stages of news production (see Frischlich et al., 2019).

Patterns of news consumption have changed even more dramatically. Su et al. (2018) observed that Facebook (FB) pages and associated user comments have become an inseparable part of online news consumption in the United States. They also noted that online commenting focuses on FB, as more news organisations are removing the comment section of their websites and instead focusing on building sustainable online communities around their FB pages. Some authors have argued that ‘social media platforms have largely taken over as the sites where active participation with the news takes place’ (Westlund and Ekström, 2018, p. 2).

Studying user comments on social media can therefore provide insights into and help understand user engagement in the online environment. This current paper contributes to the understanding of hateful and other negative online communication in two ways. First, it proposes an alternative conceptualisation of such communication via the notion of *socially unacceptable discourse* as the broadest umbrella term that can provide a basis for standardised empirical investigation. The other contribution is an examination of

¹ The Eurostat figures for 27 member states of the European Union (EU) show that active participation on social media (i.e., creating a user profile, posting messages or other contributions to Facebook or Twitter, etc.) involved 57% of the population aged 15–75 years in 2020, while 89% of this population have used the internet. See Eurostat: Science, technology, digital society. <https://ec.europa.eu/eurostat/data/database>

² An official Facebook (FB) report states that due to hate speech violations, more than 25 million comments were removed in the first quarter of 2021 (Facebook, 2021).

the extent and structure of hateful and negative communication in a case study of FB comments about two major, recent media issues in recent years: the refugee crisis in Slovenia and LGBT rights. Both are interesting from a broader European perspective.³

The empirical study focused on the Slovenian context. A large dataset of online comments from the three most-visited Slovenian news outlets' FB pages was used; they present a typical local manifestation of the global FB presence. It is important to note that, from a social research perspective, Slovenia consistently holds the position of a median European country. This is true for attitudes³ and official socioeconomic indicators among the 27 EU countries.⁴ It should be added that the general trends of the rise of right-wing populist media⁵ can also be observed in Slovenia (see Pajnik and Sauer, 2018); one of the three media outlets examined in this empirical study belongs to this type. Based on its profile, the Slovenian case study thus offers a revealing insight into the European situation.

2 Literature Review

The literature examining the prevalence and characteristics of hateful and negative commenting is rapidly expanding (cf. Meza et al., 2019; Pohjonen, 2019; Haapoja et al., 2020; Paasch-Colberg et al., 2021; Westlund, 2021). Initially, the question of online interaction patterns was typically addressed in the context of web forums (e.g., Petrovčič et al., 2012), but later on Twitter (e.g., Kapidzic et al., 2019; Ozalp et al., 2020), and particularly on FB (Su et al., 2018). In contrast, commenting sections of newspapers' websites (e.g., Coe et al., 2014; Kenski et al., 2020) have traditionally been well researched, as they were the most widespread online platforms for public discussion before the rise of social media. However, despite growing empirical research, the related concepts, terminology and definitions are still relatively diverse.

Valenzuela et al. (2019) contended that a decade of research addressing news consumption on social media platforms has produced two consistent findings. First, social media has become a major source of news about public affairs. Second, using social media for news increases individuals' political engagement. However, besides positive contributions, these media involvements also have a role in anti-democratic processes: 'Hate groups, online harassment, hyperpolarisation and state-sponsored propaganda are all examples of illiberal forces exploiting the informational value of social platforms' (Tucker et al., cited in Valenzuela et al., 2019, p. 1). Thus, Lewis and Molyneux (2018) challenged one of the most widely held assumptions in social media and journalism research that social media would be a net positive; they argued that 'journalist–audience interactions may be overwhelmingly negative for journalists (let alone for users), and in ways not fully captured in the literature thus far' (p. 15).

Another potential factor that might contribute to the expansion of hateful and other negative communication is the intentional production of negative comments. Gagliardone (2019) discussed two types of commenters, labelling them swarms or armies, and warned that they are not necessarily exclusive because swarming behaviour can gradually turn into coordinated efforts against specific targets. Similarly, Quandt's (2018) notion of dark participation encompasses both unintentional and strategic aspects, whereas those engaging in dark participation can range from individuals to small or large groups. In this context, an example of organised hate was described by Mihaylov et al. (2018), who observed that commenters were paid for politically biased comments on social networks and news-comment sections.

³ European Social Survey (ESS) shows that attitudes on these topics is a very important and controversial issue for European citizens. See ESS: <https://www.europeansocialsurvey.org/>.

⁴ For example, the share of internet users in Slovenia was 88% (2020), which is very close to the EU average (89%).

⁵ Schroeder (2019) used the example of the United States and Sweden to demonstrate how right-wing populists successfully and legitimately use digital media to circumvent traditional media.

They called them ‘opinion manipulation trolls’ and identified their characteristics with respect to nicknames and patterns of communication.

Research has demonstrated the importance of the media type. A longitudinal study by Su et al. (2018) on incivility in the FB-news environment emphasised the differences between liberal, conservative and local news outlets. Examining the comments of 42 US news outlets’ FB pages over 18 months in 2015–2016, they showed that liberal news elicited the highest proportion of civil discourse, which is also true for national news pages compared to conservative news sites and local news sites. In contrast, all three types of pages studied had a greater number of impersonal, uncivil comments than personal ones.

Some authors have also highlighted a platform’s affordances as an important factor in the rise of hateful and other negative communication. Ben-David and Matamoros-Fernandez (2016) used the example of FB to argue that hate speech and discriminatory practices are also formed by a network of ties between a platform’s policies, its technological affordances and the communicative acts of its users. Massanari (2015) employed the term *toxic technocultures* to describe the cultures enabled by and propagated via sociotechnical networks such as Reddit, 4chan, Twitter and online gaming. Massanari argued that these toxic cultures accelerate retrograde ideas of gender, sexual identity, sexuality and race, and push against issues of diversity, multiculturalism and progressivism.

This raises important questions about the social, political and ideological consequences of online discourse for public communication. In this context, Wahl-Jorgensen (2019) specifically contended that social media is characterised particularly by a distinctive emotional regime or normative emotions. She argues that mediated emotional expression can be carefully staged for a particular purpose and can become a fundamental driver of social and political actions: ‘Precisely because emotions are in part socially constituted and profoundly shaped by power relations, their public articulation—particularly in mediated contexts—tells us about more than merely how individuals feel: it tells us about how we collectively and socially narrate emotions for larger purposes’ (Wahl-Jorgensen, 2019, p. 10).

Wahl-Jorgensen claimed that public expression of negative and divisive emotions may undermine broader forms of public debate, even as it strengthens bonds within particular communities premised on exclusionary identities (Wahl-Jorgensen, 2019). It is thus not surprising that the rise of dark participation on the internet can relate to the most recent wave of populism in Western democracies (Quandt, 2018). Similarly, Wahl-Jorgensen (2019) stated that the populist turn is also marked by a shift in the emotional climate of public discourse.

3 Conceptualising Socially Unacceptable Discourse

The above-mentioned notions addressing hateful and other negative communication are relatively limited in scope (e.g., incivility) or too broad (e.g., dark participation), so they cannot serve as a general umbrella for empirical research on the entire spectrum of hateful and negative communication. Also, they are often faced with diversity and lack coherence (Paasch-Colberg et al., 2021).

For example, for the notion of incivility, Kenski et al. (2020) demonstrated that one’s perception of incivility is influenced by gender, ideology and personality traits. Su et al. (2018, p. 3,680) summarised the popular definition of incivility as the use of inappropriate vocabulary and an absence of courtesy, which is usually operationalised in terms of insulting language, dramatic language and emotional display (ibid.). In contrast, Papacharissi (2004, p. 265) used a normative approach and defined uncivil comments as ‘a set of behaviours that threaten democracy, deny people their personal freedoms, and stereotype social groups’. Alternatively, Coe et al. (2014, p. 660) defined it as a ‘feature of discussion that conveys an unnecessarily disrespectful tone towards the discussion forum, its participants or its topics’. However, incivility does not fully cover intolerant communication; in some situations, incivility is not necessarily negative (Rossini, 2020).

Another typical notion encountered in this context is *flaming*. However, Moor et al. (2010) highlighted inconsistencies in its definition and operationalisation. Still, they identified some of its essential characteristics, such as expressing hate by using insults, cursing or some other type of aggressive communication.

When discussing hateful and negative communication, various other terms appear, particularly cyberbullying, trolling, cyber-hate, micro-aggression and improper, abusive, offensive, rude, foul, impolite or indecent language, and speech that is labelled as problematic, dangerous, fearsome, violent, offensive, extremist or extreme (see Gagliardone, 2019). Like incivility or flaming, all these notions typically suffer from limited scope, have vague definitions or lack structure.

We should emphasise that the hateful and negative communication that we address in this paper refers to communication that hits the target *directly* (Quandt, 2018), so it produces direct harm to citizens or the general level of public communication. Therefore, its negative status can be established immediately from the corresponding statements and their contexts (i.e., in related posts or comments), without the need for further checking, contextualising, investigation and verification. The latter is usually required in the case of various indirect harms and if misinformation-based types of dark participation occur, which should be separated from hateful-based types of dark participation (Westlund, 2021). Besides dark participation, online communication can have numerous other negative aspects, such as violence and child abuse, and damaging communication related to suicides, self-injuries, firearms, drugs and terrorism (Facebook, 2021). Almost all crimes from penalty codes also have their specific extension, modification, counterparts or adjacency in the online context.

In this context, we focused specifically on the discourse aspects of corresponding online communication. This means that we observed how language is used and how it expresses the related thinking, meaning, relations, knowledge, etc., in the context of hateful and other negative communication.

We started from the legal perspective because at the core of any restrictions of freedom of communication lies so-called *hate speech*. We relied on the Council of Europe's (1997, p. 7) definition of hate speech as an extreme violation of free speech rights, referring to aggressive, violent and insulting communication aimed at specific groups that share protected characteristics, such as religion, gender, race, nationality, disability, refugee status or sexual orientation (i.e., *protected groups*).

In this context, the Council's 47 participating European countries⁶ implemented hate-speech legislation for the protection of these groups. The corresponding decisions of the European Court for Human Rights⁷ well illustrate the nuances and complexity of this matter. The notion of hate speech is otherwise very delicate, even within the legal context. For example, contrary to Europe, it is not used much in the US, where the formal focus is rather against 'defamation' or 'inciting lawless actions'. Outside the legal context (i.e., the fields of sociology, communication science, psychology, philosophy and religion), the understanding of hate speech is even more diverse and controversial. The latter is particularly true for the jargon used in politics and popular media, where this term is often misused (e.g., as a label for impoliteness) or abused (e.g., as a label for arguments we dislike). Still, at least from European legal viewpoints, hate speech is relatively well defined (i.e., according to Council of Europe documentation) and can be prosecuted by corresponding national and international hate-speech legislation.

⁶ Except Belarus, Kazakhstan and Vatican City.

⁷ An international court, established in 1959 on the bases of the European Convention of Human Rights, launched by the Council of Europe in 1950, see <https://echr.coe.int/>.

In this specific context, the term *speech* is not restricted to words spoken orally but is a synonym for any type of communication. Similarly, it is clear from the context that we narrow this discussion to public and computer-mediated communication on social media.

In addition to hate speech, other abuses of freedom of speech (e.g., threats, insults, obscenity) can be prosecuted according to legislation and related measures (e.g., fines, civil lawsuits). In this context, we denote any speech that can be the subject of legal actions undertaken by authorities (e.g., law enforcement, courts, inspections) as *prosecutable speech*.

Prosecutable speech encompasses only a small part of hateful and negative communication that can potentially be regulated. The bulk of this communication, which can be perceived as unacceptable to the owners of the online venues, will be denoted here as *improper communication*. This communication can be regulated by corresponding moderation rules or standards, which depend on each online venue (e.g., social media). These rules vary greatly and can be ideological, political or religious and inconsistent, unclear, unpredictable, strange or even unfair. If they also become unlawful (e.g., discrimination against protected groups), improper communication turns to prosecutable speech. In this context, the Facebook Community Standards (Facebook, 2021) are of particular importance, as they are widely applied to communication that is unacceptable according to FB's perceptions.

Figure 1 upgrades the structure from Vehovar et al. (2012) and summarises the above discussion on communication that can be subject to certain interventions (e.g., moderation, removal, warning); this classification also serves as the starting point for structuring the comments in the empirical study.

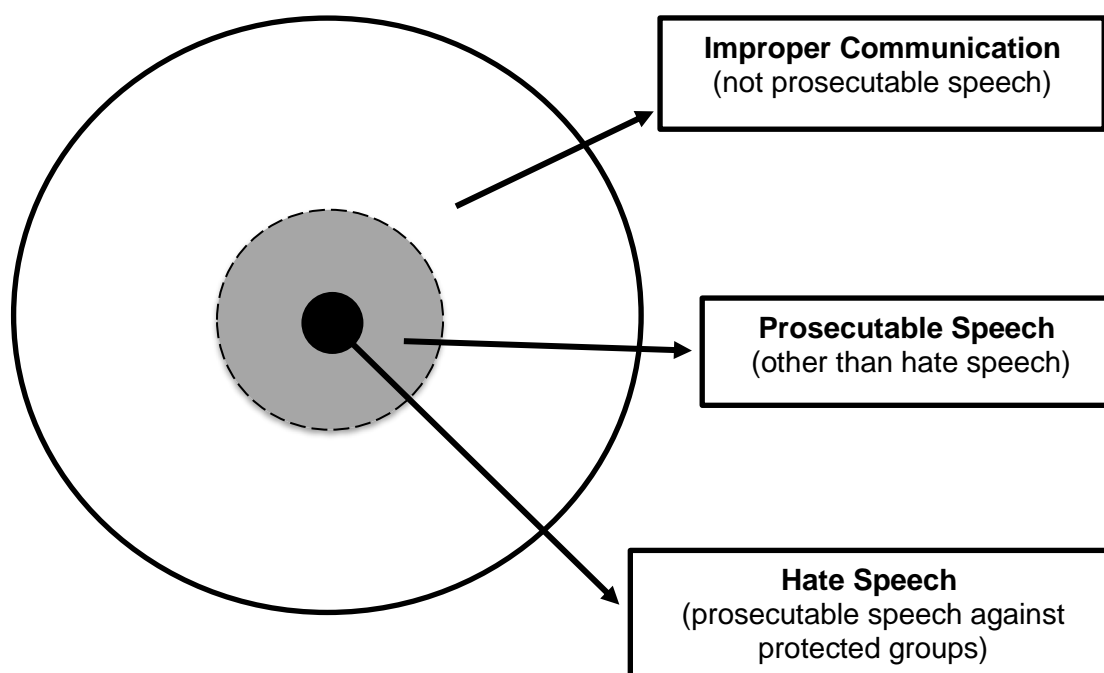


Figure 1. Hate speech, prosecutable speech and improper communication.

To meet the needs of our empirical study, we replaced the term *hateful and negative communication* with the more precise notion of *socially unacceptable discourse* (SUD). SUD covers the entire spectrum of hateful and negative communication, where the corresponding *discourse itself directly harms targets* (e.g., internet users). Consequently, this discourse is potentially *unacceptable* for society or its subgroups because its *societal consequences* are *net negative* to the extent that society or its subgroups have developed mechanisms for its potential regulation (e.g., moderation, intervention, removal and censoring).

Net negativity, which is the main criterion for identifying SUD, is manifested by the following characteristics:

- a) Cumulative vulgarisation, toxication and degradation of communication, which then prevents objective public discussion
- b) Unacceptable and unnecessary harm (resulting from the communication discourse itself) to citizens is not outweighed by an alternative added value, such as benefits to scientific, artistic, polemical or satirical discourse, and therefore cannot be justified.

SUD thus includes a full range of hateful and negative communication, which can be potentially regulated, from hate speech, threats, defamation and insults to negative stereotyping, obscenity, flaming, intolerance, incivility and vulgarity. Alternatively, as noted by De Maiti et al. (2020, p. 13), SUD can thus be an umbrella term for the majority of prosecutable, hateful, abusive and discriminatory speech and indecencies and rude language.

The perception of communication, which is socially unacceptable, varies not only across societies, their subgroups, and online venues but also across individuals who typically have their own, unique perceptions and understanding of the above-mentioned criteria. Thus, SUD has countless versions. Nevertheless, building on the above-mentioned conceptualisation, we further elaborate on specific SUD criteria that are suitable for our empirical research. We rely on a vision that citizens participate in a public discussion with the deliberate aim of contributing to the exchange of ideas, arguments and opinions to increase their knowledge. This ideal (or idealistic) vision was the basis for evaluating communication (i.e., each FB comment) against the above-mentioned criteria of net negativity.

4 Research Design

Within the above-described framework, we addressed the following research questions in the empirical study:

- RQ1: What is the extent of SUD, and which types and targets predominate?
- RQ2: How are SUD comments connected to a media and a topic?

The empirical study analysed comments from FB, which currently dominates as a global online venue for online commenting and discussion.

4.1 The Data

The empirical work was conducted via the FRENK project,⁸ which comprehensively collected user comments from the FB pages of the three most-visited Slovenian news portals.⁹ At that time, two of them, 24ur.com (M1) and Siol.net (M2), belonged to mainstream (more or less liberal) media websites, while Nova24TV.si is a typical populist right-wing (RW) media outlet strongly associated with one of the largest political parties. The monthly reach of each portal is around half of the Slovenian internet population (MOSS, 2019), while the number of FB followers differs.¹⁰

In the first research step, all the available posts (N = 38,000) on the three FB pages, with their comments (N = 268,000), were harvested with data scraping procedures in October 2017. Therefore, all posts that

⁸ See <https://nl.ijs.si/frenk/english/>.

⁹ The national TV website (rtvslo.si) was omitted because at that time, it handled comments predominantly with its own system and without a sizeable FB page.

¹⁰ The number of FB followers in 2020 were as follows: M1: 217,232; M2: 45,361; RW: 22,096.

existed on these FB pages in October 2017 were included; they were created somewhere in the period 2010–2017.

We further summarise the technical details from Ljubešić et al. (2019). By ‘post,’ we mean the FB *entry* (performed by media FB editors) with a link to the corresponding news article on the original media portal. The FB *comments* (made by users) on these posts are structured in *threads* (i.e., groups of related comments under the same initial comment). A new thread appears whenever someone comments on a particular post with a new comment (and not within an existing thread).

Most of the FB posts were from 2013–2017, with nearly half from 2017. Relatively few posts were from 2010, 2011 and 2012, while no posts were from 2009 or earlier; this is either because the transition to FB was not yet complete or a particular portal (e.g., RW) did not exist. This is a unique and comprehensive dataset of posts and comments because we disposed of all comments, not just a sample.

In the second research step, the machine-learning algorithm classified all posts according to their relevance to two selected topics: migrants and LGBT. The goal was to obtain around 5,000 comments for each topic, assuring a critical mass for machine-learning procedures and a feasible amount for manual coding. The algorithm identified 957 posts (with 43,000 comments) for the migrant topic, of which the top 30 posts (with 6,545 comments) showing the highest relevance for the migrant topic were included in the analysis. As there were only 93 posts (with 4,571 comments) scraped for the LGBT topic, all were included. Out of 123 posts, 64 were from RW (Nova24TV.si), 14 from M1 (24ur.com) and 11 from M2 (Siol.net). In the third research step, all 11,118 comments were manually coded by 32 annotators who produced 93,251 annotations.

The initial data collection was followed by lengthy data cleaning, complex machine-learning algorithms, extensive coding and very complex analysis. Despite a certain delay in the data collection period, the data are still extremely valuable because, as of 2018, this type of scrapping is no longer possible due to GDPR restrictions. The data are also unique because they are rich in variables and include the entire population of comments (not merely a sample) over the eight years.

We believe that the genuine patterns of emergence, evolution and structure of SUD are analogous today because they are based on how people process emotions in an online environment. An additional value of the data is that they provide insight into how hate speech and negative comments emerge in an online environment in which there is little or no moderation.

We also added context about the two topics. From the autumn of 2015 to the first quarter of 2016, over 400,000 refugees entered Slovenia¹¹, a country with two million inhabitants. Regarding LGBT, two referendums regarding same-sex marriage were carried out in 2012 and 2015, polarizing the voting population. The topic of refugees has been much more turbulent and negatively exposed in the media than LGBT issues (toleration is growing steadily). Therefore, the results from a nationwide general social survey (SJM, 2019) showed that 64% of respondents were concerned about refugees, while most respondents (58%) supported LGBT marriage.

4.2 Annotation Process

The selected comments (6,545 and 4,571) were evaluated by the poll of 32 specially trained evaluators—annotators (Ljubešić et al., 2019, p. 6). Not all annotators evaluated every comment; however, each evaluated roughly eight to nine comments, so reliable coding was achieved. The corresponding coding criteria were in accordance with the definition and structure of SUD (see Figure 1). The criteria were further developed specifically for this research study according to an optimistic vision of public

¹¹ For an analysis of reporting about the refugee crisis in Slovenia, see Bajt (2016) and Vezovnik (2018).

communication in which citizens make online comments to share ideas or knowledge and discuss this in a decent manner (Frischlich et al., 2019). Thus, in the following section, we call SUD a specific SUD-FRENK implementation.

The criteria for identifying comments that can be classified as net negative (according to SUD-FRENK implementation) were further elaborated in the corresponding coding instructions developed exclusively for this research exercise (FRENK, 2018). Decisions were often difficult, delicate and arbitrary, particularly when the added value of polemic or satiric components might outweigh negativity. Nevertheless, all of these situations were treated with a common approach and in a standardised way.

This specific implementation was expected to provide much more comments classified as SUD compared to the usual moderation practice on social media, particularly the practice on FB in the years 2010–2017, where FB had yet to establish strict community standards. Some illustrations are provided below.

During the coding process, each comment was first checked to determine whether it belonged to SUD; otherwise, it was denoted as '**not SUD**'. The SUD comments were then structured for the dimension of type (see Figure 2), which is based on Figure 1.

- The initial separation depended on whether a SUD comment was directed towards groups with protected characteristics. These were further classified according to the prevailing **background**, either as containing elements of potential **violence** or threats (e.g., 'All refugees should be put in concentration camps and executed.') or **insults**, discrimination, stereotyping and intolerance (e.g., 'Muslims are like monkeys and rapists.'). In extreme cases, both categories could be, in principle, a potential subject of legal hate-speech prosecution, particularly cases of violence.
- When SUD comments were directed at specific persons or groups outside protected characteristics (e.g., medical doctors, firefighters, political parties, journalists, FB commenters), they were classified as **other** and then again according to the prevailing presence of **violence** and threats (e.g., 'I will find you and beat you up.'). or **insults** and other severely negative attitudes, such as defamation, stereotyping, mocking and so on (e.g., 'You are such an idiot.'). In extreme cases, these two types of communication could also be legally prosecuted (e.g., via civil lawsuits).
- The remaining SUD comments, which were not directed at any people or groups, were coded as **indecent language** (e.g., cursing and rude language). Formally, this slightly departs from the notion of improper language (see Figure 1) because indecent language can still be unlawful (e.g., obscenity and bestiality) and thus prosecutable. This category refers only to residual SUD comments, which were without a target and thus outside the main SUD categories denoted above as 'background' and 'other'. Otherwise, indecent language accompanies many SUD comments with a target. Nevertheless, the comments in the main categories of SUD (i.e., background and miscellaneous) do not necessarily have to contain indecencies. Intolerance can also be expressed politely without being indecent (Rossini, 2020).

Most of the comments in the above-described SUD-type categories do not reach effective standards for legal prosecution. At the same time, all SUD comments are potential subjects of intervention from the owners of online venues, depending on their specific moderation rules. It is still possible (but unlikely) that a certain comment was classified as 'not SUD' (based on SUD-FRENK implementation), while it might be found socially unacceptable (and removed) on another online venue with extremely high standards or with a very specific understanding of net negativity.

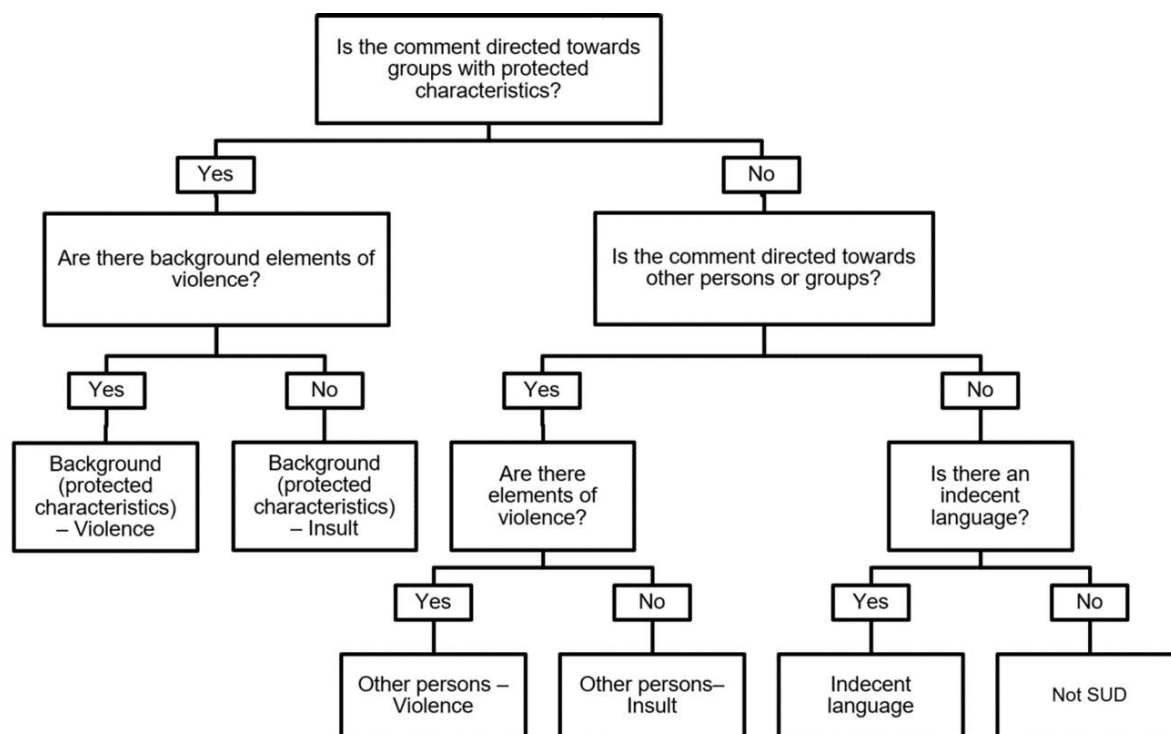


Figure 2. Coding scheme for the type of SUD.

In the second coding dimension, the annotators allocated each comment to a specific target. There were five potential targets: (i) direct comments to protected groups (i.e., migrants or LGBT), (ii) indirect comments to supporting public entities (e.g., lawyers, NGOs, public bodies), (iii) comments to journalists or media, (iv) comments to FB commenters or (v) to some other target.

The annotation process resulted in 93,251 annotations, 56,171 for 6,545 migrant-related comments and 37,080 for 4,571 LGBT-related comments. On average, there were roughly eight (LGBT) and nine (refugee) annotations per comment. The reliability of the annotations was measured with Krippendorff's alpha, and it was above the threshold, $\alpha > 0.66$. For details on Krippendorff's alpha and further references, see Ljubešić et al. (2019, p. 9). The annotation process was also rechecked with expert evaluation from a research team member who controlled a random half of the annotations.

5 Results

5.1 Posts and Annotations

The 30 posts on the refugee topic had an average of 218 comments (maximum: 1,030; minimum: 5) and 135 commenters (maximum: 486; minimum: 3), while the 93 posts on the LGBT topic had an average of 50 comments (maximum: 409; minimum: 3) and 33 commenters (maximum: 347; minimum: 2). The posts about refugees had more comments because they had more threads per post, while the number of comments per thread was similar, around 3, which was true for each media (i.e., RW, M1 and M2).

The raw shares of SUD among 93,251 annotations (see Table 1) show that around half of the annotations belong to some type of SUD. This seems surprisingly high, especially since before data harvesting certain SUD comments had been removed by the authors or FB. Only 47% of all annotations were 'not SUD', slightly more for LGBT (54%) than migrants (42%).

Table 1. Shares (%) of annotations across the SUD types.

Topic	Background – violence	Background – insult	Other – violence	Other – insult	Indecent language	Not SUD	Total ^a
Migrants	8	28	1	19	2	42	100
LGBT	2	20	1	20	2	54	100
Total	5	25	1	20	2	47	100

^a Due to rounding, the totals may not add up to 100%.

We can observe from Table 1 that insults (totalling 25% and 20%) were the most common annotations. Elements of violence (5% and 1%) were observed much less frequently, which was also true for ‘indecent language’ (2%). We can recall here that the latter includes only SUD comments without a target; otherwise, all SUD types are more or less indecent. The prevailing SUD sentiment is thus linked to insulting, which also covers mocking, discriminating, stereotyping, defamation and so on.

5.2 Types and Targets of Comments Across Topics and Media

For further analysis, each comment was assigned a modal (or mode) SUD type, which was the category that most annotators had assigned. On average, these modal categories contained 80% of all annotations assigned to a certain comment, so roughly seven out of nine annotations (and annotators) agreed with the prevailing SUD type assigned to a comment. Specific discrepancies in the distribution of the annotations across comments exist; however, they do not interfere with a content-oriented discussion. There are also some differences in the SUD structure based on comments (i.e., modal annotations) in Tables 2 and 3 compared to the raw SUD structure of annotations (see Table 1); however, these discrepancies are also negligible for a substantive discussion.

We analysed the SUD comment types across media and SUD targets, where each comment was preliminarily assigned a modal SUD target category (a similar process to assigning a modal SUD type). Regarding the media, the difference was surprisingly small (see Table 2). The rare specific was RW on the refugee topic (see Table 2a), where violence against the groups with protected characteristics (‘background-violence’) was encountered more often (10%) compared to M2 (2%) or M1 (5%), while corresponding insults (‘background-insult’) were only slightly higher (23% versus 21% or 18%). Somewhat surprisingly, on the refugee topic, the most-visited mainstream media, M1, had a similarly low share of comments that did not belong to SUD as RW (43%). This is mainly due to the remarkably high ‘other-insult’ category (33%) in M1, which was likely based on heavily polarised discussions.

Table 2. Shares (%) of comments for media across the SUD types for refugee and LGBT topics.

Refugee (Table 2a)

Media	Background – violence	Background – insult	Other – violence	Other – insult	Indecent language	Not SUD	Total ^a
M1	5	18	0	33	1	43	100
M2	2	21	0	24	1	53	100
RW	10	23	0	23	1	43	100

^a Due to rounding, the totals may not add up to 100%.

LGBT (Table 2b)

Media	Background – violence	Background – insult	Other – violence	Other – insult	Indecent language	Not SUD	Total ^a
M 1	1	21	0	17	1	59	100
M 2	0	22	0	16	0	61	100
RW	3	22	1	30	1	42	100

^aDue to rounding, the totals may not add up to 100%.

Regarding LGBT (see Table 2b), the specific of RW was related to a high share (30%) of insults ('other-insults'), which also resulted in a much lower overall number of comments that did not belong to SUD (42%) compared to M2 (61%) or M1 (59%). Slightly more exposed 'background-violence' for RW (3%) versus M2 (0%) or M1 (1%) was also observed, although to a much lesser extent than the refugee topic.

Next, we compared the SUD types of comments across the media and targets (see Table 3), where 'not SUD' and 'indecent language' were excluded, as they have no targets (by definition). Only three SUD targets are thus presented in Table 3: protected characteristic group (i.e., refugee/LGBT), journalists (including media) and other FB commenters. The category of related supporting public entities (e.g., institutions and lawyers) and the remaining category (i.e., other targets) were also omitted because they had almost no variability and thus brought no added value to our discussion. We can observe at media M2 that among all refugee-related comments (see Table 3a) belonging to the SUD type 'background-insult', 78% of these comments were directed towards refugees, 2% towards journalists or media and 1% towards other FB commenters, while the remaining 19% of comments were omitted as they belonged to indirect targets or to other targets.

Table 3. Shares (%) of comments for SUD types across media and targets for refugee and LGBT topics.

Refugee (Table 3a)

	M1	M2	RW	M1	M2	RW	M1	M2	RW
SUD Type	Aimed at Refugees			Aimed at Journalists			Aimed at FB Commenters		
Background-violence	97	97	98	0	0	0	0	1	0
Background-insult	83	78	83	0	2	0	2	1	2
Other-violence	4	13	7	2	13	3	27	52	37
Other-insults	1	1	1	8	14	28	42	49	38

LGBT (Table 3b)

	M1	M2	RW	M1	M2	RW	M1	M2	RW
SUD Type	Aimed at LGBT			Aimed at Journalists			Aimed at FB Commenters		
Background-violence	83	88	86	0	0	0	0	1	0
Background-insult	88	85	82	0	0	0	1	1	2
Other-violence	6	0	4	0	0	3	33	52	21
Other-insults	0	1	0	2	6	13	35	53	33

We can also observe that with the refugee topic (see Table 3a), in the case of 'background-violence', more than 95% of all SUD comments were directed towards refugees, with little difference across the media. Little differences between the media can be observed in these aspects (although at a somewhat lower level) for the LGBT topic (see Table 3b). A similar lack of differentiation existed for the 'background-insult' type of SUD. Somewhat more diversity can be noted with 'others-violence' and 'others-insult'. Here, the main target across all three media was other FB commenters, which reflected high polarisation and impolite discussions. This is particularly notable in M2, which shows the extreme polarisation of the commenter community.

In contrast, for RW, polarisation seems least exposed, likely because of the users' more homogeneous (commenting) profile. The only point where RW media strongly differed were the insults towards journalists and media: in refugee topics (see Table 3a) in the RW media in 'other-insults', 28% of the comments targeted journalists (or media), much more than in M1 (8%) or M2 (14%). Similar ratios were observed for the LBGT topic (13% versus 2% and 6%; see Table 3b). This can be interpreted either as ('domestic') commenters attacking journalists from other media or ('nondomestic') commenters disagreeing with the hostile reporting of ('domestic') journalists.

6 Discussion

6.1 Answers to the Research Questions

Regarding the first research question (RQ1: What is the extent of SUD and which types and targets predominate?), approximately half of the comments were identified as SUD. Among the SUD types, insults against protected groups (25%) or other persons (20%) dominated, followed by violence against protected groups (5%) or other persons (1%) and (non-targeted) indecent language (2%). Regarding targets, most SUD comments were directed towards protected groups. When other persons (not belonging to protected groups) were targeted, these were predominantly insults towards other FB commenters.

Although not directly comparable, we can parallel this SUD structure with the proportions of incivility in other studies, where it ranged from 20% to 40% (e.g., Papacharissi, 2004; Coe et al., 2014; Santana, 2015; Rossini, 2020). Given that the focus of our current study was on two very exposed, controversial topics and that a very broad concept of SUD was used, it seems that our results roughly fit into this range. It is worth noting that intolerance, which includes all SUD comments characterised as 'violence' or 'insult' (but can also encompass some comments from 'indecent language'), clearly prevails, as it amounted to at least 85% of all SUD comments. This differs somewhat from Rossini (2020), where the share of posts with incivility, where there was no intolerance, dominated over the share of posts with intolerance; however, in different contexts, the topics analysed in Rossini's study were much more general (i.e., the minority rights topic was only one among six topics) and were not focused exclusively on protected groups (e.g., refugees, LGBT).

Regarding differences (RQ2: How are SUD comments connected to the media and the topic?), somewhat surprisingly, there were relatively few differences across the media regarding the internal structure of comments and the type of SUD. The share of SUD comments on the RW populist media outlet Nova24TV.si was 57%, very similar to the leading mainstream media M1 (24ur.com) with 56%, and with only moderate lagging (47%) in the other mainstream media M2 (Siol.net). However, very clear differences between the media appeared regarding the number of initial SUD-generating news, which was then republished as posts on FB, where they generated a stream of (SUD) comments. Specifically, the RW media hosted most of these posts (64 out of 123 included in the analysis), although it produced by far the smallest amount of daily news.

Thus, the RW media differed in having a radically higher share of initial SUD-generated posts, which compensated for the smaller absolute number of news and threads (and, in sum, the smaller overall

number of comments following these posts). However, the commenting itself (i.e., internal size and structure of the commenting threads, levels and types of SUD) contained relatively little specifics in RW media, except for a somewhat higher share of violent comments towards protected groups on refugee topics and a higher share of insults towards other persons when discussing the LGBT topic.

Regarding topics, commenting was generally more intensive in the case of the refugee topic than LGBT. Ten times more comments (refugees: 43,000 versus LGBT: 4,571) were identified by machine-learning algorithms among all 268,000 scraped comments. This can be explained by the much more invasive nature of the refugee problem and the higher media exposure. The share of SUD among the analysed comments on the refugee topic was also higher (58%) compared to LGBT (46%).

6.2 Broader Context

The relatively high share of SUD can be largely explained by very specific and antagonistic topics (e.g., refugees and LGBT), which flame polarisation. The unrestricted FB environment additionally contributed to this, as until 2018, interventions from FB were very limited. At that time, the pressure from the European Commission for hate-speech regulation had not yet been effectively transformed into corresponding measures of FB-moderation practice (European Commission, 2021). Nevertheless, it is worth adding that similarly high levels of SUD were also found in the preliminary results from the replication of this same FRENK study in the UK and Croatia (Ljubešić et al., 2019). Thus, it seems that in a relatively unrestricted online environment, when exposed to polarising topics, social media users tend to produce sizeable amounts of SUD despite a considerable lack of anonymity.

We should add that Vehovar et al. (2020) demonstrated for this same Slovenian FRENK dataset that half of all commenters could be identified as commenters who predominantly produce SUD comments. Thus, it seems that we do not have a situation in which SUD comments are produced only by a small prolific group of users. Instead, we could talk about the latent potential for producing SUD, which is evenly distributed across the entire population of online commenters. This can be justified by the above statement (RQ2) regarding minor differences in SUD levels and structures between the media. Despite the three different FB media target groups, SUD exhibits relatively stable characteristics in terms of level, type and target within the commenting threads. This confirms that the propensity to create SUD does not essentially depend on the media (at least when talking about general news media) or the corresponding audience but is evenly distributed across all online commenters (or possibly across all internet users).

High shares (i.e., half) of SUD comments and SUD-producing commenters raise issues about the nature of contemporary computer-mediated communication, the role of mediated emotions and the related consequences for public communication and corresponding political and ideological effects. For Wahl-Jorgensen (2019, p. 13), a discursive climate dominated by negative emotions, articulated with deliberative exclusionary intent, can have serious consequences for political participation, particularly in the context of a political culture characterised by a cynical attitude towards political engagements. Those who engage in a political discussion are often dismissed as so extreme or even insane in their positions that their contributions can be ignored (*ibid.*).

Therefore, it could be argued that the dominance of negativity in the commenting sections of popular news outlets' FB pages contributes to rising cynicism about political matters. The danger thus exists that negative emotions within a community are gradually becoming normative as the discourse demarcations of communal boundaries are constructed (Döveling et al., 2018, p. 4). Orgeret (2020, p. 294) argued that a combination of strong emotions and a lack of media and digital literacy may lead to populist turbulence, particularly in weakly regulated environments.

In this context, it is interesting to compare the results with consistent findings from representative surveys in Slovenia (SJM, 2019; Vehovar et al., 2012), according to which public respondents—when confronted

with certain SUD statements (such as the examples from Section 4.2) – almost uniformly declared SUD statements unacceptable. This contrasts starkly with the high level of SUD comments in the current study. It appears that social media triggers and incites SUD comments from otherwise decent citizens, and/or this specific (negative) online communication disproportionately attracts SUD-oriented commenters.

In the language of social informatics, perhaps the interaction between ICTs and modern societies has led to specific negative consequences (i.e., a deterioration in the quality of public discourse). Interestingly, it seems that another interaction of ICTs with society was needed to remedy these effects, namely, the development of advanced ICT-based moderation policies. For example, FB reports that in the first quarter of 2021, 97% of all user-content interventions were automatically executed by computer algorithms. This contributed significantly to a decrease in the prevalence of hate speech, which was less than 0.1% in 2020 and less than 0.05% in 2021, meaning that of all the views users made on FB, only 0.05% were exposed to hateful content (Facebook, 2021). This suggests that rigorous moderation and control are necessary to enhance public discussions in computer-mediated environments and steer them towards the promise of participatory journalism and the normative ideal of public communication (Frischlich et al., 2019).

6.3 Limitations and Future Research

This study has several limitations. We focused on only two topics in which antagonistic public debate was to be expected, so the level of SUD comments is likely to differ from the general situation. Also, this study was limited to three specific news outlets' FB pages in one country. Due to the operationalisation of SUD (i.e., SUD-FRENK), a direct comparison with other SUD-related terms (incivility, intolerance, indecency, flaming, etc.) was impossible.

Nonetheless, this case study is very informative in understanding the actual patterns of SUD comment creation in an unrestricted online environment with weak monitoring, as was the case with FB commenting in 2010–2017. On the one hand, this is a limitation (lack of contemporary insight), but on the other hand, it is also a particular advantage, as it provides insight into an unrestricted environment. However, since SUD comments are closely related to emotions, we can expect relatively stable principles and patterns. Therefore, these results are very valuable for understanding how people create SUD comments in unrestricted online environments.

Future research could overcome these limitations and extend this study. One of these is the replication of this study by including the last few years (i.e., 2018–2021), when social media introduced much stricter moderation measures, especially under the pressure of the European Commission and its monitoring (European Commission, 2021). Another interesting extension could be the comparative replication of our analysis for already collected data (using the same FRENK methodology and identical data scraping procedures) in the UK and Croatia (Ljubešić et al., 2019) to observe country differences. Similar is true for an analysis that would compare refugee and LGBT topics with other issues, and also for the analysis of overlaps between respective media audiences.

It would also be interesting to compare comments posted on FB with those posted directly on the corresponding original media website (i.e., the three news outlets) and with comments posted on other online platforms (e.g., online forums and other social media platforms). It would also be useful to apply surveys or cognitive techniques to test the definition of the SUD-FRENK approach and to further compare it to existing moderation practices. Qualitative research approaches could be used to illuminate the apparent discrepancy between the high prevalence of SUD comments and the extreme public aversion to SUD. It would also be very informative to additionally recode the SUD comments in this study according to alternative terms used in dealing with hateful and negative online communication, specifically indecency, cyberbullying, trolling, inappropriate language and intolerance.

7 Conclusion

In this paper, we addressed hateful and other negative aspects of public online communication, focusing specifically on social media. First, we summarised related concepts ranging from incivility and flaming to dark participation and hate speech. Then, we introduced the concept of socially unacceptable discourse (SUD), which is the main contribution of this study.

The corresponding SUD criteria were further elaborated, starting from the normative ideal of public communication in which citizens participate in commenting on news and exchanging arguments. The criteria were implemented in a special coding manual that served the SUD-FRENK empirical research. This is another important contribution of this study and can be used in further research efforts to develop standardised measures of hate speech and other negative communications.

In an empirical study, we analysed all the users' comments on refugees and LGBT issues on the FB pages of the three most-visited news websites in Slovenia during the period 2010–2017. The main conclusion is that the quality of the analysed comments was surprisingly low, with about half of the comments belonging to SUD. An equally remarkable finding is that SUD does not depend strongly on the media or the target audience but seems evenly distributed. Thus, it appears that in an online environment, the mediatisation of emotions stimulates the creation of SUD comments evenly among users. The high prevalence of SUD comments in (unrestricted) social media seems to contradict the general attitude of the population, which is (at least in principle) against the use of SUD; this paradox is one of the main challenges for future research.

The results imply that the specifics of an online environment strongly encourage impulsive behaviour and negative emotions, which steers online communication towards SUD. Therefore, continuous monitoring is generally needed to avoid compromising public discourse in the online environment.

Additional Information and Declarations

Funding: This paper was prepared thanks to grants from Slovenian Research Agency P5-0399, J7-8280 and V5-1736.

Conflict of Interests: The authors declare no conflict of interest.

Author Contributions: V.V.: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing. D.J.: Conceptualization, Writing – original draft.


Data Availability: Ljubešić, Nikola; Fišer, Darja; Erjavec, Tomaž and Šulc, Ajda, 2021, Offensive language dataset of Croatian, English and Slovenian comments FRENK 1.1, Slovenian language resource repository CLARIN.SI. Available at <http://hdl.handle.net/11356/1462>.

References

- Bajt, V. (2016). Online hate speech and the 'refugee crisis' in Slovenia. In Atienza, B. B., Alonso, J. A. (Eds.), *Migration and Asylum: New Challenges and Opportunities for Europe*. Thomson Reuters Aranzadi.
- Ben-David, A., & Matamoros-Fernandez, A. (2016). Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-Right Political Parties in Spain. *International Journal of Communication*, 10, 1167–1193.
- Coe, K., Kenski, K., & Rains, S. A. (2014). Online and uncivil? Patterns and determinants of incivility in newspaper websites comments. *Journal of Communication*, 64, 658–679. <https://doi.org/10.1111/jcom.12104>
- Council of Europe. (1997). *Recommendation No. R (97) 20 of the Committee of Ministers to Member States on Hate speech*. Retrieved October 1, 2021, from <https://rm.coe.int/1680505d5b>
- De Maiti, K. P., Fišer, D., & Ljubešić, N. (2020). Nonstandard linguistic features of Slovene socially unacceptable discourse on Facebook. In Fišer, D., & Smith, P. *The Dark Side of Digital Platforms: Linguistic Investigations of Socially Unacceptable Online Discourse Practices*, (pp. 12–35). Ljubljana University Press.

- Döveling, K., Harju, A. A., & Sommer, D. (2018). From mediatized emotion to digital affect cultures: New technologies and global flows of emotion. *Social Media + Society*, 4(1), 1–11. <https://doi.org/10.1177/2056305117743141>
- European Commission. (2021). *The EU Code of conduct on countering illegal hate speech online*. Retrieved July 31, 2021, from https://ec.europa.eu/info/policies/justice-and-fundamental-rights/combating-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en
- Facebook. (2021). *Transparency Centre: Community Standards Enforcement Report*. Retrieved July 31, 2021, from <https://transparency.fb.com/data/community-standards-enforcement>
- FRENK. (2018). Guidelines for unacceptable speech annotation in online comments. Unpublished coding manual (Available on request from authors).
- Frischlich, L., Boberg, S., & Quandt, T. (2019). Comment Sections as Targets of Dark Participation? Journalists' Evaluation and Moderation of Deviant User Comments. *Journalism Studies*, 20(14), 2014–2033. <https://doi.org/10.1080/1461670X.2018.1556320>
- Gagliardone, I. (2019). Defining Online Hate and Its "Public Lives": What Is the Place for "Extreme Speech"? *International Journal of Communication*, 13, 3068–3087.
- Haapoja, J., Laaksonen, S. M., & Lampinen, A. (2020). Gaming Algorithmic Hate-Speech Detection: Stakes, Parties, and Moves. *Social Media + Society*, 6(2), 1–10. <https://doi.org/10.1177/2056305120924778>
- Kapidzic, S., Neuberger, C., Stiglitz, S., & Mirbabaie, M. (2019). Interaction and influence on Twitter. *Digital Journalism*, 7(2), 251–272. <https://doi.org/10.1080/21670811.2018.1522962>
- Kenski, K., Coe, K., & Rains, S. A. (2020). Perceptions of Uncivil Discourse Online: An Examination of Types and Predictors. *Communication Research*, 47(6), 1–20. <https://doi.org/10.1177/0093650217699933>
- Lewis, S. C. & Molyneux, L. (2018). A decade of research on social media and journalism: Assumptions, blind spots and a way forward. *Media and Communication*, 6(4), 11–23. <https://doi.org/10.17645/mac.v6i4.1562>
- Ljubešić, N., Fišer, D., & Erjavec T. (2019). *The FRENK datasets of socially unacceptable discourse in Slovene and English*. Available at: <https://arxiv.org/pdf/1906.02045.pdf>
- Lünenborg, M., & Maier, T. (2018). The Turn to Affect and Emotion in Media Studies. *Media and Communication*, 6(3), 1–4. <https://doi.org/10.17645/mac.v6i3.1732>
- Massanari, A. (2015). #Gamergate and The Fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media & Society*, 19(3), 329–346. <https://doi.org/10.1177/1461444815608807>
- Meza, R., Vincze, H. O., & Mogos, A. (2019). Targets of Online Hate Speech in Context. A Comparative Digital Social Science Analysis of Comments on Public Facebook Pages from Romania and Hungary. *Intersections. East European Journal of Society and Politics*, 4(4), 26–50. <https://doi.org/10.17356/ieejsp.v4i4.503>
- Mihaylov, T., Mihaylova, T., Nakov, P., Márquez, L., Georgiev, G. D., & Koychev, I. K. (2018). The dark side of news community forums: opinion manipulation trolls. *Internet Research*, 28(5), 1292–1312. <https://doi.org/10.1108/IntR-03-2017-0118>
- Moor, P. J., Heuvelman, A., & Verleur, R. (2010). Flaming on YouTube. *Computers in Human Behavior*, 26, 1536–1546. <https://doi.org/10.1016/j.chb.2010.05.023>
- Orgeret, K. S. (2020). Discussing Emotions in Digital Journalism. *Digital Journalism*, 8(2), 292–297. <https://doi.org/10.1080/21670811.2020.1727347>
- Ozalp, S., Williams, M. L., Burnap, P., Liu, H. & Mostafa, M. (2020). Antisemitism on Twitter: Collective Efficacy and the Role of Community Organisations in Challenging Online Hate Speech. *Social Media + Society*, 6(2), 1–20. <https://doi.org/10.1177/2056305120916850>
- Paasch-Colberg, S., Strippel, C., Trebbe, J. & Emmer, M. (2021). From Insult to Hate Speech: Mapping Offensive Language in German User Comments on Immigration. *Media and Communication*, 9(1), 171–180. <https://doi.org/10.17645/mac.v9i1.3399>
- Pajnik, M. & Sauer, B. (2018). *Populism and the Web: Communicative Practices and Movements in Europe*. Routledge.
- Papacharissi, Z. (2004). Democracy online: civility, politeness, and the democratic potential of online political discussion groups. *New Media & Society*, 6(2), 259–283. <https://doi.org/10.1177/1461444804041444>
- Petrovčič, A., Vehovar, V. & Žiberna, A. (2012). Posting, quoting, and replying: a comparison of methodological approaches to measure communication ties in web forums. *Quality & Quantity*, 46(3), 829–854. <https://doi.org/10.1007/s11135-011-9427-z>
- Pohjonen, M. (2019). A Comparative Approach to Social Media Extreme Speech: Online Hate Speech as Media Commentary. *International Journal of Communication*, 13, 3088–3103.
- Quandt, T. (2018). Dark participation. *Media and Communication*, 6(4), 36–48. <https://doi.org/10.17645/mac.v6i4.1519>
- Rossini, P. (2020). Beyond Incivility: Understanding Patterns of Uncivil and Intolerant Discourse in Online Political Talk. *Communication Research*, (in press). <https://doi.org/10.1177/0093650220921314>
- Santana, A. D. (2015). Incivility dominates online comments on immigration. *Newspaper Research Journal*, 36(1), 92–107. <https://doi.org/10.1177/073953291503600107>
- Schroeder, R. (2019). Digital Media and the Entrenchment of Right-Wing Populist Agendas. *Social Media + Society*, 5(4), 1–11. <https://doi.org/10.1177/2056305119885328>

- SJM.** (2019). *Slovensko javno mnenje/Slovenian Public Opinion (2019): Module Hate Speech*. Retrieved October 1, 2021, from <https://www.adp.fdv.uni-lj.si/opisi/sjm191/>
- Smutny, Z., & Vehovar, V.** (2020). Social Informatics Research: Schools of Thought, Methodological Basis, and Thematic Conceptualization. *Journal of the Association for Information Science and Technology*, 71(5), 529–539. <https://doi.org/10.1002/asi.24280>
- Su, L. Y. F., Xenos, M. A., Rose, K. M., Wirz, C., Scheufele, D. A., & Brossard, D.** (2018). Uncivil and personal? Comparing patterns of incivility in comments on the Facebook pages of news outlets. *New Media & Society*, 20(10), 3678–3699. <https://doi.org/10.1177/1461444818757205>
- Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P.** (2019). The Paradox of Participation Versus Misinformation: Social Media, Political Engagement, and the Spread of Misinformation. *Digital Journalism*, 7(6), 802–823. <https://doi.org/10.1080/21670811.2019.1623701>
- Vehovar, V., Motl, A., Mihelič, L., Berčič B., & Petrovčič, A.** (2012). Zaznava sovražnega govora na slovenskem spletu. *Teorija in praksa*, 49(1), 171–189.
- Vehovar, V., Povž, B., Fišer, D., Ljubešić, N., Šulc, A., Jontes, D.** (2020). Družbeno nesprejemljivi diskurz na Facebookovih strani novinarskih portalov. *Teorija in praksa*, 57(2), 622–645.
- Vezovnik, A.** (2018). Securitizing migration in Slovenia: a discourse analysis of the Slovenian refugee situation. *Journal of Immigrant & Refugee Studies: International, National, and Regional Theory, Research, and Practice*, 16(1/2), 39–56. <https://doi.org/10.1080/15562948.2017.1282576>
- Wahl-Jorgensen, K.** (2019). *Emotions, Media and Politics*. Polity.
- Wahl-Jorgensen, K.** (2020). An Emotional Turn in Journalism Studies? *Digital Journalism*, 8(2), 175–194. <https://doi.org/10.1080/21670811.2019.1697626>
- Westlund, O., & Ekström, M.** (2018). News and Participation through and beyond Proprietary Platforms in an Age of Social Media. *Media and Communication*, 6(4), 1–10. <https://doi.org/10.17645/mac.v6i4.1775>
- Westlund, O.** (2021). Advancing Research into Dark Participation. *Media and Communication*, 9(1), 1–10. <https://doi.org/10.17645/mac.v9i1.1770>

Editorial record: The article has been peer-reviewed. First submission received on 10 August 2021. Revision received on 9 October 2021 and 8 November 2021. Accepted for publication on 20 November 2021. The editor in charge of coordinating the peer-review of this manuscript and approving it for publication was Stanislava Mildeova .

Special Issue: Perspectives of Social Informatics.

Acta Informatica Pragensia is published by Prague University of Economics and Business, Czech Republic.

ISSN: 1805-4951
