

## Free speech or hate speech? A legal analysis of the discourse about Roma on Twitter

Therese Enarsson & Simon Lindgren

**To cite this article:** Therese Enarsson & Simon Lindgren (2019) Free speech or hate speech? A legal analysis of the discourse about Roma on Twitter, Information & Communications Technology Law, 28:1, 1-18, DOI: [10.1080/13600834.2018.1494415](https://doi.org/10.1080/13600834.2018.1494415)

**To link to this article:** <https://doi.org/10.1080/13600834.2018.1494415>



© 2018 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 02 Jul 2018.



Submit your article to this journal [↗](#)



Article views: 16520



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 9 View citing articles [↗](#)

# Free speech or hate speech? A legal analysis of the discourse about Roma on Twitter

Therese Enarsson<sup>a</sup> and Simon Lindgren<sup>b</sup>

<sup>a</sup>Department of Law, Umeå University, Umeå, Sweden; <sup>b</sup>Department of Sociology, Umeå University, Umeå, Sweden

## ABSTRACT

This article draws on material gathered from Swedish tweets about the Roma population, in order to map different discourses. Based on this material, a legal analysis was made focusing on how the legal protection under the European Convention on Human Rights (ECHR) and Swedish law for such types of expression may vary depending on the wider discursive context. This article concludes that the legal protection for hateful expressions against for instance Roma, will vary depending on the discursive context. When the expression is a direct part of a political discussion, the protection of freedom of expression will be higher. However, emphasis must also be placed on the aim, value, and accuracy of the statement, even in a political context. This will increase the possibilities to legally intervene against speech that may perhaps be triggered by an ongoing political debate, but is hateful and without value to that debate.

## KEYWORDS

Hate speech; freedom of expression; Roma; social media; discourse; text analysis

## Introduction

The problem of online *hate speech* has become a new factor to handle legally, in Sweden as well as in other countries. An overarching question in doing so is what should be protected as *free speech* and what should not, due to the victimization of groups or individuals that some expressions could cause. In social media, many ethnic and religious groups have been subjected to discussion and hatred, such as for example Muslims<sup>1</sup> and people of Jewish descent.<sup>2</sup> In Sweden, particularly in recent years, one such targeted group is the Roma population.

Sweden, which is a signatory state to the *European Convention on Human Rights*<sup>3</sup> (ECHR, the convention), has a legal responsibility not to enact any law or regulation in breach of the convention. The convention has been incorporated as law in Sweden since 1994 and has a strong constitutional status. It has been implemented in fundamental law through

**CONTACT** Therese Enarsson  [Therese.enarsson@umu.se](mailto:Therese.enarsson@umu.se)

<sup>1</sup>See for example Anton Törnberg and Petter Törnberg, 'Muslims In Social Media Discourse: Combining Topic Modeling And Critical Discourse Analysis' (2016) 13 Discourse, Context & Media 132–142.

<sup>2</sup>Amos Guiora & Elizabeth A. Park, 'Hate Speech on Social Media' (2017) 45 Philosophia, September 2017, Volume 45, Issue 3, 957–971.

<sup>3</sup>The Convention for the Protection of Human Rights and Fundamental Freedoms, drafted in 1950 and came into force in 1953.

the *Instrument of Government*, which means that legislating in breach with the convention is an unconstitutional act.<sup>4</sup> Sweden, like other signatory states, must take the convention and the rulings made by the *European Court of Human Rights* (ECtHR, the Court) into consideration when enacting or applying legislation in regard to expressions online, for instance. The Court's rulings and reasonings do, as will be shown, take into consideration many contextual factors – such as the wider discourse in which a statement is made – when deciding if a limitation of a persons' freedom of expression is legitimate, and maybe even strongly called for.

This study analyses discussions about Roma in social media. More specifically, it draws on material gathered from Swedish tweets that mention the Roma population in Sweden. We map what different discourses are expressed, with a particular focus on seeing whether, or to what extent, the identified discourses contain hateful speech or expressions against Roma. The aim is to provide an analysis of how the legal protection under the ECHR and Swedish law for such types of expression may vary depending on the wider discursive context. For instance, if aggressive tweets identified in on-going political discussions enjoy stronger protection, and if aggressive ones without such a context could be intervened against more easily. The present article is part of two larger research projects, where one focuses on understandings of victimization online<sup>5</sup> using a combination of sociology and victimology, and the other highlights legal aspects of racially motivated victimization<sup>6</sup> – especially against Roma and Sami in Sweden – by combining legal studies with victimology.

## Background: Roma in Sweden

The Roma are a national minority in Sweden, and their history in the country, as well as in other European countries, is marked by oppression, threats, hate speech and discrimination against them, from both individuals and the state. Today, Roma are still a group that is subject to discrimination on, for example, the labour market. The situation with a larger number of Roma people travelling to Sweden from poorer countries in the EU has also added a dimension of vulnerability for Roma people, and this has made the debate on Roma interconnect with the broader issue of street begging.<sup>7</sup> A political debate about the Roma has been ongoing in Sweden over the last few years, and in the autumn of 2017, the Moderate Party suggested that begging on the streets should be made illegal, causing that issue to yet again be discussed by the media, politicians and on social media.<sup>8</sup>

Victimization of the Roma people in Sweden can therefore be seen both historically and today. Studies have also shown that victimization can raise feelings of insecurity, difficulty in understanding information, and poor self-esteem, although the extent of these reac-

<sup>4</sup>See The Instrument of Government, Chapter 2, Section 19, and Lag (1994:1219) om den europeiska konventionen angående skydd för de mänskliga rättigheterna och de grundläggande friheterna.

<sup>5</sup>The project 'Understandings of online victimization', funded by The Swedish Crime Victim Compensation and Support Authority, was carried out between 2011 and 2016 at the Department of Sociology, Umeå University, Sweden.

<sup>6</sup>The 'Victims of Racism' (swe. Offer för rasism) project is a research project combining legal analysis with victimology, and focuses especially on the victimization of Roma and Sami. The project is conducted by five researchers at the Department of Law, Umeå University, and is funded by the Swedish Research Council.

<sup>7</sup>Jacqueline Bhabha, 'Realizing Roma Rights: An Introduction', *Realizing Roma Rights* (University of Pennsylvania Press 2017) 17–56; Karin Åström, 'Skyddet För Utnyttjade Tiggare Är Statens Ansvar' (2015) 92 *Socialmedicinsk Tidskrift* 286–295.

<sup>8</sup>See for instance SVT Nyheter, 'Moderaterna Vill Ha Ett Nationellt Tiggeriförbud' (5 September 2017 updated 23 October 2017) <<https://www.svt.se/nyheter/inrikes/moderaterna-vill-ha-ett-nationellt-tiggeriforbud>>; Svenska Dagbladet, 'M vill förbjuda tiggeri – får kritik' (5 September 2017) <<https://www.svd.se/just-nu-m-presenterar-ny-kriminalpolitik>>.

tions at the individual level may be due to both the offense and the personal characteristics of the victim.<sup>9</sup> Being exposed to a *racially motivated* crime, which targets something deeply personal such as skin colour, ethnicity, cultural origin or religion, can cause an even greater pain and more harm to the victim. This may be because the crime is motivated by hatred directed towards a deeply personal part of an individual, something that the victim cannot influence or change. Studies have also shown that people within targeted groups experience fear and anxiety, and that such victimization can linger in a community at large, even with people that have not themselves been victimized, and can also lead to people from such groups or communities keeping parts of themselves hidden, and not show religious or cultural symbols or avoiding certain contexts and spaces where an increased risk of victimization may be suspected.<sup>10</sup> One such space could then be the internet, since social media platforms have become a new arena for not only political discussions but also victimization.<sup>11</sup>

## Hate speech on the internet

During the last couple of decades, the internet has given rise to a new arena where it is easy to share information, opinions and thoughts. This availability can have a positive effect on people's lives since it offers accessible ways and means of self-expression, hence giving people a chance to access and participate in society and politics. However, it has also led to the spread of hatred and also a whole new arena for networks spreading messages based in hate, such as racist propaganda, to get a foothold and share their views quickly and widely.<sup>12</sup>

The occurrence of *hate speech* has raised the question of the responsibility of social media platforms for facilitation of the distribution of such speech, and since 2016 there is an agreed Code of Conduct between some of the largest social media platforms (Twitter, Facebook, YouTube and Microsoft) and the European Commission, where the social media platforms agree to strive to remove hateful content defined as hate speech within 24 hours. In the fall of 2017 these efforts intensified in an attempt to decrease the amount of speech inciting hatred and violence,<sup>13</sup> and in the spring of 2018 the issue got even more attention when both Facebook and YouTube received severe criticisms for not removing hateful content.<sup>14</sup>

<sup>9</sup>Jo Goodey, *Victims and Victimology: Research, Policy and Practice* (Pearson Education Limited 2005) 121–122; Christian Diesen, *Tereapeutisk juridik* (Liber 2011) 109; Annemarie ten Boom and Karlijn F Kuijpers, 'Victims' Needs as Basic Human Needs' (2012) 18 *International Review of Victimology* 155–179.

<sup>10</sup>Paul Iganski, 'Hate Crimes Hurt More' (2001) 45 *American Behavioral Scientist* 626. 626–638; Paul Iganski and Spiridoula Lagou, 'Hate Crimes Hurt Some More Than Others: Implications for the Just Sentencing of Offenders' (2015) 30 *Journal of Interpersonal Violence* 1696–1718; Mika Andersson and Caroline Mellgren, 'Anmälningsbenägenhet vid utsatthet för hatbrott' (2015) 3–4 *Socialvetenskaplig tidskrift* 283–301.

<sup>11</sup>Teo Keipi *et al*, *Online Hate and Harmful Content: Cross-national perspectives* (Routledge 2017).

<sup>12</sup>Daniel J Solove, *The Future of Reputation: Gossip, Rumor, and the Privacy on the Internet*, (Yale University Press 2007) 17; The Swedish Government, *A Comprehensive Approach to Combat Racism and Hate Crime: National Plan Against Racism, Similar Forms of Hostility and Hate Crimes* (2016) 60.

<sup>13</sup>For more on this, see European Commission, Press release, Security Union: Commission steps up efforts to tackle illegal content online, Bryssel September 28, 2017, European Commission, Press release, Countering online hate speech – Commission initiative with social media platforms and civil society shows progress, Bryssel June 1, 2017.

<sup>14</sup>See for instance The Verge (Adi Robertson), 'YouTube bans neo-Nazi channel after criticism over hate speech rules' (28 February 2018) <<https://www.theverge.com/2018/2/28/17062002/youtube-ban-atomwaffen-neo-nazi-channel-hate-speech-rules>>; The Guardian, 'Myanmar: UN blames Facebook for spreading hatred of Rohingya' (13 March 2018) <<https://www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya>>.

At a more profound level, this development has to do with how the internet and social media have crucially transformed the conditions for social interaction. First of all, the ways in which people communicate things to each other have become increasingly *asynchronous*. Even if people may use texting, chats, direct messaging, emails, and even discussion forums, for communicating at a steady pace, or to respond to each other the very moment that they receive a message, the digital tools and platforms allow for delays. These possibilities for asynchronicity lead to a 'conversational relaxation'.<sup>15</sup> Asynchronicity makes it possible for large groups of people to have sustained interaction. This is what happens on social media such as Twitter, and in various online forums and communities.

Asynchronous communication also means that one does not have to deal with the immediate reaction of those that one is interacting with. Psychologist John Suler argues that this makes users disinhibited. The possibility of moving in and out of a conversation – returning when and if we want to – and the absence of 'a continuous feedback loop that reinforces some behaviours and extinguishes others' – enables us to feel safer, and to formulate our thoughts more freely.<sup>16</sup>

Online interaction can therefore be more democratic than face-to-face interaction. At least in some contexts, it can contribute to minimizing the role played by status and authority. Suler's overarching argument is that people will say and do things online that they would not say and do in offline and face-to-face settings. He writes of *the online disinhibition effect* – the effect that people tend to be less restrained and to express themselves more openly online. Obviously, this has its advantages and disadvantages: It might enable hate speech, as well as it may promote participation, intimacy and self-disclosure.

Another characteristic of digital sociality, especially when communication is text-based, is that it makes embodied people largely invisible. When one uses services, such as Twitter, one cannot be seen directly by others, nor can we see who else is there at the same time. Even though some sites, platforms, and services have indicators that display whether people are online or not, one is still physically invisible (though sometimes represented by a profile photo or other avatar). As discussed earlier, the lack of non-verbal cues and other information about the communication partner, and the social context of interaction, has been thought by some to increase uninhibited communication, such as being aggressive or using harsh language.<sup>17</sup>

Anonymity can be both good and bad for social interaction. As psychologist Philip Zimbardo showed in the infamous Stanford prison experiment, anonymity in groups can lead to de-individuation – the process by which individuals, unknown to each other, and with concealed identities become immersed in a group dynamic – a state where people can be impulsive, blatantly aggressive towards one another and even sadistic.<sup>18</sup> On the internet, the 'illusion of large numbers' might make us overestimate how many people share

<sup>15</sup>Joseph B. Walther, Joseph B. 'Computer-Mediated Communication: Impersonal, Interpersonal, and Hyperpersonal Interaction' (1996) *Communication Research* 23(1) 3–43, 26.

<sup>16</sup>John Suler 'The Online Disinhibition Effect' (2004) *Cyberpsychology & Behavior* 7(3) 321–325, 323.

<sup>17</sup>M.J. Culnan & M.-L. Markus, 'Information Technologies' in L. Putnam & D. Mumby (Eds.), *Handbook of Organizational Communication* (Beverly Hills, CA: Sage 1987) 420–443, 429.

<sup>18</sup>Philip Zimbardo *The Lucifer Effect: Understanding How Good People Turn Evil* (New York Random House 2013).

our views.<sup>19</sup> We might see, for example, the number of views of a YouTube video or the number of times a tweet has been retweeted and make a rough translation between such numbers and the general legitimacy of the content.

When anonymity removes personal responsibility and generates a perceived loss of individuality, it can also lead to people becoming more altruistic and more willing to help others. In digital society then, anonymity can be a force of unity and solidarity as well as of fragmentation and nihilism. This is a long-standing debate. Some claim that anonymous interaction in digital media is a major cause of hate speech, racism, sexism etc., while others like to focus on how anonymity online can facilitate things like grassroots political action in places where censorship and surveillance make such mobilization difficult.

The question of anonymity online has many dimensions, as seen above, and another one is that of anonymity in relation to *victimization*. Experienced victimization online or a generally hateful environment can cause someone to feel that they either have to hide their identity or withdraw from the internet. This can be the case when people feel threatened online, causing them to limit or miss out on taking part in participatory social and political aspects of the internet.<sup>20</sup> This, and the issue of victimization online, its effects, and how to tackle it have been the focus of intense legal debate in Sweden and other European and western countries over the last decade, and the demand for legislative actions has been prominent. The legal aspects that have been focused on in this debate have taken their starting point in the need to protect people from victimization. However, one constantly relevant and overarching question has also been that of reaching a careful balancing by international and national courts between the right to freedom of expression and the right for people to be protected from such hateful expressions that should be prohibited in a democratic society.<sup>21</sup>

## Data and method

In terms of methodology, this study positions itself within the wider field of critical discourse studies (CDS), aiming to analyse the process by which social reality is ‘conceptually mediated’. It is about analysing how language use relates to the symbolic construction of institutions, ideologies, power relations, and identities.<sup>22</sup> CDS does not dictate any one specific research method, but this study combines computational topic modelling with legally oriented close readings of the data. Such combinations of methods within CDS are not novel, as there has long been talk of research in the vein of corpus-assisted discourse studies, or CADS.<sup>23</sup> In line with the argument of Baker and colleagues, that none

<sup>19</sup>Katelyn YA McKenna and John A Bargh, ‘Plan 9 From Cyberspace: The Implications of the Internet for Personality and Social Psychology’ (2000) 4 *Personality and Social Psychology Review* 57–75, 64.

<sup>20</sup>Danielle Keats Citron, ‘Civil rights in our information age’ in Saul Levmore and Martha C Nussbaum (Eds.) *The Offensive Internet* (Cambridge: Harvard University Press 2010) 31–49.

<sup>21</sup>See for instance Michael Salter and Chris Bryden, ‘I Can See You: Harassment and Stalking on the Internet’ (2009) 18 *Information & Communications Technology Law* 99–122. In Sweden several legislative changes has come into force over the last few years, and the balancing of freedom of expression and the protection of victims’ privacy has had a prominent place in the legal working leading up to it, see for example: The Swedish Government, Prop. 2016/17:222 Ett starkt straffrättsligt skydd för den personliga integriteten.

<sup>22</sup>Norman Fairclough ‘Critical Discourse Analysis’ In *The Routledge Handbook of Discourse Analysis*, James Paul Gee and Michael Handford (Eds.) (Routledge Handbooks in Applied Linguistics. London: Routledge 2012) 9–20, 9.

<sup>23</sup>Nicholas Close Subtirelu and Paul Baker ‘Corpus-Based Approaches’ in John Flowerdew and John E. Richardson (Eds.) *The Routledge Handbook of Critical Discourse Studies* (Milton Park, Abingdon, Oxon ; New York, NY: Routledge 2017) 106–119.

of the methods ‘need be subservient to the other (as the word ‘assisted’ in CADS implies),’<sup>24</sup> this study sees the computational analysis and the close readings as contributing ‘equally and distinctly to a methodological synergy’.

The empirical procedure for this study can be described in terms of three different phases: (1) data collection and text pre-processing; (2) topic modelling; and, (3) interpretation and critical legal analysis.

### **Data collection and text pre-processing**

Data for this study was accessed through Twitter’s public Search API, which provides a near-complete sample of real-time tweets for most queries that do not return huge data volumes. Data was streamed for a set of filter words related to Roma issues, during a week in September (8th – 15th) of 2017. The particular week was randomly chosen, but coincidentally happened to include the heated debate about the Moderate Party’s suggestion about imposing a ban on street begging. The collection resulted in a dataset consisting of 19,474 tweets. The data was then processed through a set of text cleaning procedures which were applied in order to transform the tweets into the appropriate format for further processing. Such operations were implemented in Python using relevant libraries, most prominently NLTK (Natural Language Toolkit).<sup>25</sup> The cleaning pipeline included lowercase conversion, whitespace stripping, removal of punctuation, filtering of stopwords (a generic Swedish list) and sparse terms.

### **Topic modelling**

The next phase employed topic modelling through Latent Dirichlet Allocation (LDA), to identify common thematic categories in the tweets, and to map how they were distributed across them.<sup>26</sup> This provided an overview and insight into the corpus that helped spot areas of interest for the following closer critical analysis.

LDA assumes, first, that there are a fixed number of topics by which a corpus can be described. Topics are defined as patterns or groups of terms that have a tendency to appear together in the documents. Second, it assumes that each document in the analysed collection exhibits the topics in different degrees. The result of an LDA is information about the topical structure: Which set of topics is likely to best describe the corpus? How are the topics distributed across documents?

The LDA analysis was performed using the gensim library for Python.<sup>27</sup> There are some statistical metrics that can be calculated and used to validate the relevance and coherence of topic models. In the case of this analysis, a set of models with different numbers of topics (5–150) were created and evaluated using the  $C_V$  coherence metric.<sup>28</sup> This indicated the best model coherence at 20 topics, and this was the number of topics chosen for the

<sup>24</sup>Paul Baker and others, ‘A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press’ (2008) 19 *Discourse & Society* 273–306, 274.

<sup>25</sup>Steven Bird, Ewan Klein, and Edward Loper *Natural Language Processing with Python* (Cambridge MA: O’Reilly 2009).

<sup>26</sup>David M Blei, Andrew Y. Ng, and Michael I. Jordan ‘Latent Dirichlet Allocation’ (2003) *Journal of Machine Learning Research* 3, 993–1022.

<sup>27</sup>Radim Rehurek and Petr Sojka. ‘Software Framework for Topic Modelling with Large Corpora’ (2010) in *Proceedings of the Lrec 2010 Workshop on New Challenges for Nlp Frameworks* 45–50.

<sup>28</sup>Michael Röder, Andreas Both, and Alexander Hinneburg. ‘Exploring the Space of Topic Coherence Measures’ (2015) in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, ACM 399–408.



model analysed in this present paper. Furthermore, experimental research has shown that such measures tend to be vastly outperformed by evaluation through actual human judgment.<sup>29</sup> When it comes to the different configurations that can be made in order to arrive at a reasonable set of topics, it has been commonly suggested that ‘human-in-the-loop methods’ are recommendable.<sup>30</sup>

Critical close reading and contextualized analysis of the texts makes it possible to go past the topic structure to discuss what the algorithmically extracted categories express in the social context that the analysed corpus was constructed to represent. Such analysis entails paying attention to intertextual, historical, and political contexts – and in our case, the legal context – that go way beyond the mere analysis of linguistic units.

### **Interpretation and critical legal analysis**

Traditionally, legal methodology can be described as the qualitative analysis of legal material for which there are established interpretation methods based on an internal judicial assessment – a theoretical hierarchy of the status of legal sources.<sup>31</sup> For this study, this will be done by analyzing the mapping of correlations between hateful or aggressive speech and certain topics where such expressions were more frequently occurring, in relation to the ECHR and court rulings from the European Court of Human Rights (ECtHR, the Court). Since this study positions itself in a Swedish context and Swedish tweets, this will be exemplified in relation to Swedish legislation, its underlying purpose and relevant court rulings. The main focus in doing so will, however, be the scope of *freedom of expression* in article 10 ECHR and what can and can’t be included in this as part of a political discussion online.

Legally, the field of political discourse demands a contextual analysis, taking into account such factors as the value and relevance of the expression for a political debate, and also *who* or *what* the discourse is targeting.<sup>32</sup> For an understanding of this contextuality, court rulings of the ECtHR, and exemplifying court rulings from Swedish courts, will be of great significance in identifying and analyzing discourses. This will enable an analysis of the level of aggressive tweets – based on the frequency of typically aggressive words – concerning Roma, and if they especially occur in a discourse which revolve around current political discussions and thus within a context in which they enjoy stronger protection.

### **Identifying discursive themes through topic modelling**

The created topic model was explored using the Python version of the LDAvis tool.<sup>33</sup> Using the visualization technique suggested by LDAvis enables interpreting the meaning,

<sup>29</sup>Jonathan Chang, Sean Gerrish, Chong Wang, Jordan L. Boyd-graber, and David M. Blei. ‘Reading Tea Leaves: How Humans Interpret Topic Models’ (2009) in *Advances in Neural Information Processing Systems* 22, Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta (Eds.), 288–296.

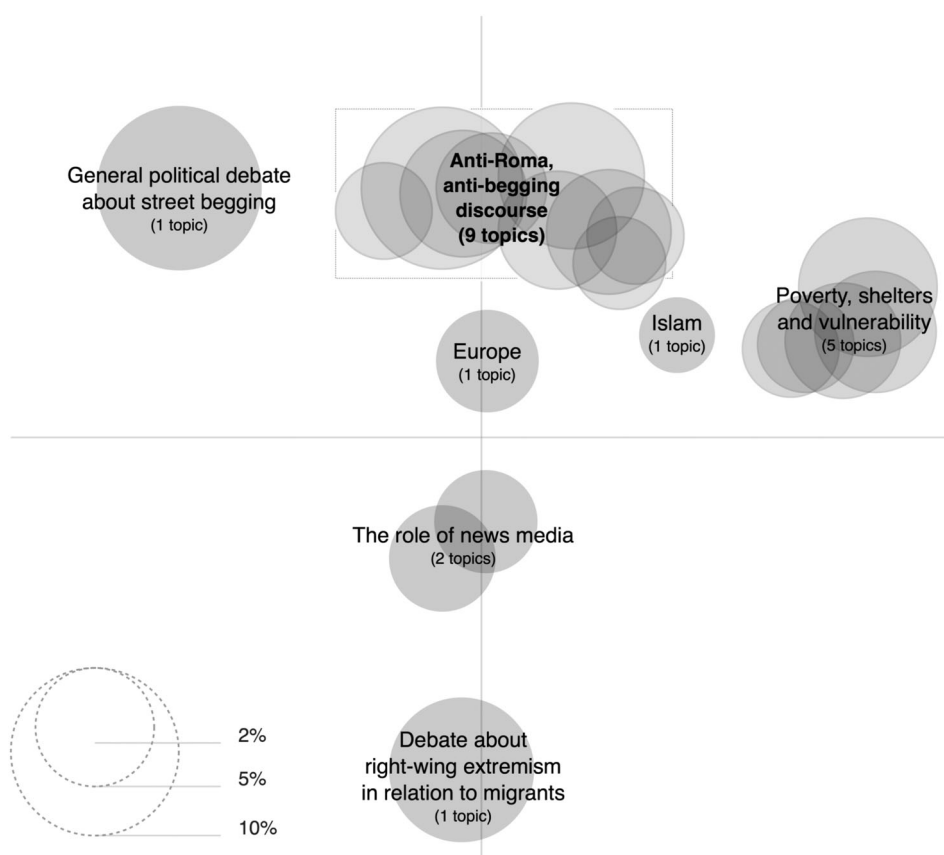
<sup>30</sup>Tak Yeon Lee and others, ‘The Human Touch: How Non-Expert Users Perceive, Interpret, and Fix Topic Models’ (2017) 105 *International Journal of Human-Computer Studies* 28–42, pp. 29–30.

<sup>31</sup>Aleksander Peczenik ‘Vad är rätt?: om demokrati, rättssäkerhet, etik och juridisk argumentation’ (Stockholm, Norstedts juridik AB 1995).

<sup>32</sup>Therese Enarsson and Markus Naarttijärvi, ‘Is It All Part of the Game? Victim Differentiation and the Normative Protection of Victims of Online Antagonism under the European Convention on Human Rights’ (2016) 22 *International Review of Victimology* 123–138.

<sup>33</sup>See <https://github.com/cpsievert/LDAvis>, and <https://github.com/bmabey/pyLDAvis>.





**Figure 1.** LDA topic model of Roma tweets.

prevalence, and relationships of topics. Figure 1 shows the topics in our dataset visualized as circles in a two-dimensional plane, where the sizes of circles indicate the overall prevalence of each respective topic in the model. The distances between topics are determined, in LDAvis, by computing the so-called Jensen-Shannon divergence between topics, and then projecting the inter-topic distances onto two dimensions using multidimensional scaling.<sup>34</sup> The visualization reveals how the discourse about the Roma population in Sweden on Twitter at the time of this study can be represented through a set of topics and topical clusters.

The most prevalent single topic is the one representing a *general political debate about street begging*. Indeed, the issue of street begging among the Roma population – which also runs through most other topics – is *the* key topic in the analysed discourse at the time of this study. Tweets rating high within this topic expressed various opinions and criticisms relating to how the street begging problem should be addressed by political means. Many of these tweets were part of a discussion about the issue of whether begging should be prohibited by law or not, and to what extent a ban would be an adequate way of addressing the issue. By extension, this debate also concerned if street

<sup>34</sup>Sievert Carson, and Kenneth Shirley 'LDAvis: A Method for Visualizing and Interpreting Topics' (2014) in *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 63–70, 63.

begging among Roma in Sweden is a case of organized crime, or if that view is a myth. In these polemic discussions, left- and right-wing parties in the Swedish parliament were often mentioned, compared and pitted against one another.

In relation to the debate about how Swedish politicians have, or should, deal with street begging, there is also the topic of the situation in Sweden as compared to in Europe in general. This part of the discourse is generally connected to the argument that Swedish policy is not doing enough to deal with the issue. Tweets related to this topic express things such as that street begging is more commonly seen in Sweden than in other European countries, and once again to the discussion about whether a ban would be the right way to go or not.

Furthermore, the two-topic cluster connecting discourse about *the role of news media* includes tweets of a similar character, but which are to a larger degree responding to statements made in opinion pieces in mainstream media outlets. This general debate was also sometimes expressed by comparing the Roma minority to other minority groups, such as Muslims (cf. the topic *Islam*). The debate became the most heated and polemic in tweets related to the topic which was connected to the *debate about right-wing extremism in relation to migrants*, where some of those in opposition of passing a law against street begging questioned why 'Nazis' were allowed to hold political manifestations, while 'poor people' should not be allowed in public. Conversely, those in favour of a ban questioned whether anyone not defending street begging was automatically to be categorized as a 'Nazi'.

Some of the issues covered in the topics outlined above were dealt with in more detail in the five-topic cluster collecting discourse on the issues of *poverty*, *shelters*, and *vulnerability* in relation to the social and political issue of street begging among the Roma population in Sweden. Basically, all of the tweets connected to these topics are variants of comments to either the issue of whether a ban would solve the problem of poverty, whether it is reasonable to use government funding to build shelters for street beggars, or to what extent street begging is an expression of organized crime or of true vulnerability.

The most prominent – and for this study also the most interesting – pattern illustrated in [Figure 1](#) is however the 9-topic cluster collecting expressions of *anti-Roma and anti-begging discourse*. This group of topics include rather aggressive expressions of opinions related to the character of the issue of street begging, as well as to the character of Roma as a group. For instance, in our dataset, one person who commented on the issue of begging in regard to Roma focused the tweet on comparing beggars with vermin, and that the problem therefore should be 'treated accordingly'. On a similar note, another tweet described begging as a disease that could not be cured. Yet another wanted to kill all beggars, however, in that tweet it was not clearly stated that the person connected begging particularly to Roma or any other particular group, leaving the expression wider and less likely to be seen as targeting a national group or ethnic group. By using the methods used in this study it was also not possible to connect certain tweets to a possible string of tweets from that same individual, and then, by context clarifying if that was the intended meaning. Yet another tweet stated that 'gypsies' do nothing but deceive, lie and beg. It is on these types of tweets that we focus the critical legal analysis to follow.

## Contextualization and legal analysis of identified discourses

To avoid that any tweets be identifiable in relation to specific individuals we are not citing complete specific tweets in this text. Instead, we discuss the importance of the *context* in which hateful or aggressive expressions and discourses take place, when determining whether expressions in tweets might be legally protected or not. In doing so, we will provide indirect examples from specific tweets.

### *Freedom of expression in relation to different discourses*

To understand the question of context and victimization in regard to the identified discourses, one must understand the legal concept of freedom of expression. As mentioned, this study has a Swedish and European outlook and will therefore be based in article 10 of the ECHR. The article states that everyone has the right to hold and express their opinions, to receive and convey information, and form and give expression to ideas, without interference of the state. The ECtHR has stated that this entails airing such ideas and expressions that can be seen as offensive by the State or a group in society, as well as expressions that shock or disturb people – all in order to protect an open democratic society, that enables pluralism and diversity of thought.<sup>35</sup> This means that political discussions – like the different political discourses identified about Roma on Twitter, regarding such overarching issues as immigration, whether street begging should be allowed or not, or if politicians are doing a good job – fall under the protection of article 10 on a more general level.

However, *if individual expressions* in such discussions are protected is another matter. This is due to the fact that the protection for expressions in article 10 is highly contextual and in no way absolute.<sup>36</sup> Having the right to freedom of expression carries with it duties and responsibilities, and signatory states are allowed to make a number of exceptions with regards to the freedom of speech, to protect other interests, such as for example that of national security, the prevention of disorder or crime and for protection of health or morals as well, as the protection of reputations or rights of others. States are allowed to impose restrictions, penalties etcetera that are *necessary in a democratic society* and *prescribed by law* (ECHR article 10:2). However, article 10 also has the purpose of strengthening other rights under the convention, making sure that people can access, and exercise other rights as well, mainly article 8, with the right to respect for private and family life, article 9 and its protection of freedom of thought, conscience and religion, and article 11 that specifically protects the freedom of assembly and association. This shows the scope of the purpose of freedom of expression: Individuals are supposed to be protected within their private sphere, but also to freely practice their religion with others, to speak up in public, and enjoy the freedom of assembly.<sup>37</sup>

<sup>35</sup>*Handyside v. United Kingdom* (App no 5493/72) ECHR 7 december 1976 § 49.

<sup>36</sup>Therese Enarsson and Markus Naarttijärvi, 'Is It All Part of the Game? Victim Differentiation and the Normative Protection of Victims of Online Antagonism under the European Convention on Human Rights' (2016) 22 *International Review of Victimology* 123–138.

<sup>37</sup>Alastair Mowbray, *Cases, Materials, and Commentary on the European Convention on Human Rights* (Oxford University Press 2012) 694–730; Macovei, Monica, 'Freedom of expression: a guide to the implementation of Article 10 of the European Convention on Human Rights' Human Rights Handbooks No. 2 (Council of Europe Publishing 2004).

Generally speaking, expressions that in some way contribute to, or are part of, political discussions, social debate or an exchange of information will be protected to a larger extent, whereas for expressions of a commercial nature, for instance, the opposite is the case.<sup>38</sup> This especially strong protection for expressions of value in a public or political debate is due to the fact that the Court has defined this as essential in a democratic, pluralistic society, and a foundation in protecting human rights in a broader sense.<sup>39</sup> For that reason, this stronger protection includes expressions that criticize, for instance, a government.<sup>40</sup> This can be exemplified by the case of *Lindon, Otchakovski-Laurens and July v. France*, where the Court states that,

There is little scope under Article 10 § 2 of the Convention for restrictions on freedom of expression in the area of political speech or debate – where freedom of expression is of the utmost importance [...] – or in matters of public interest [...].

Furthermore, the limits of acceptable criticism are wider as regards a politician as such than as regards a private individual. Unlike the latter, the former inevitably and knowingly lays himself open to close scrutiny of his every word and deed by both journalists and the public at large, and he must consequently display a greater degree of tolerance.<sup>41</sup>

Such identified debates on Twitter as mentioned above, of polemic discussions where left- and right-wing parties in the Swedish parliament were discussed, compared and pitted against one another, should therefore – generally speaking – have a strong protection, in order to promote an open discussion. That also means that political figures may have to endure more criticism and harsh statements about them, because of the protection of political debates and since they have willingly put themselves in a public role.<sup>42</sup>

However, the opposite can be the case if an expression is directed at, and expresses intolerance against, certain protected individuals or groups in society. Such intolerant expressions can, according to the ECtHR, be limited by signatory States to a larger extent, as discussed in *Gündüs v. Turkey*:

... the Court would emphasise, in particular, that tolerance and respect for the equal dignity of all human beings constitute the foundations of a democratic, pluralistic society. That being so, as a matter of principle it may be considered necessary in certain democratic societies to sanction or even prevent all forms of expression which spread, incite, promote or justify hatred based on intolerance (including religious intolerance), provided that any ‘formalities’, ‘conditions’, ‘restrictions’ or ‘penalties’ imposed are proportionate to the legitimate aim.<sup>43</sup>

The case revolves around Mr Gündüs, who appeared on television, speaking out as a member of an Islamist sect, and his statements were critical against democracy and secular institutions. Mr Gündüs was prosecuted and convicted in Turkey on the grounds that the statements made on the television show had incited to hatred and hostility.

<sup>38</sup>Bernadette Rainey and others, *Jacobs, White & Ovey, The European Convention on Human Rights*, (Oxford 2014) 438.

<sup>39</sup>See for instance *Lingens v. Austria* (App no 9815/82) 8 July 1986; *Brasilier v. France* (App no 71343/01) 11 April 2006; Alastair Mowbray, *Cases, Materials, and Commentary on the European Convention on Human Rights* (Oxford University Press 2012) 627.

<sup>40</sup>See for instance *Castells v. Spain* (App no 11798/85) 23 April 1992.

<sup>41</sup>*Lindon, Otchakovski-Laurens and July v. France* (App no 21279/02) 22 October 2007 § 46.

<sup>42</sup>*Lingens v. Austria* (App no 9815/82) 8 July 1986; *Lindon, Otchakovski-Laurens and July v. France* (App nos. 21279/02 and 36448/02). This is developed further in Therese Enarsson and Markus Naarttijärvi, ‘Is It All Part of the Game? Victim Differentiation and the Normative Protection of Victims of Online Antagonism under the European Convention on Human Rights’ (2016) 22 *International Review of Victimology* 123–138.

<sup>43</sup>*Gündüs v. Turkey* (App no 35071/97) 14 June 2004 § 40.

However, the ECtHR argued that Turkey had violated Mr Gündüs' freedom of expression under article 10 and that Gündüs expressions did not constitute *hate speech*.<sup>44</sup> This, even though the Court did state that Turkish people could feel attacked in their way of living in an offensive manner. The Court placed weight on the fact that it was a live broadcast where he did not have time to rethink and rephrase his answers, and that he took part in a 'lively public discussion' at the time.<sup>45</sup>

The *Gündüs* case demonstrates the contextual nature when determining if States have legitimate reasons for limiting citizens freedom of expression. This means that other expressions in other contexts, that target ethnicity or religion, can be limited or restricted, as seen in *Norwood v. the United Kingdom*. In *Norwood* a man had placed a picture in his window, in the months after 9/11, that portrayed the Twin Towers in flames, a symbol of a prohibition sign with a crescent and a star together with the words 'Islam out of Britain – Protect the British People'.<sup>46</sup> This was seen as a public expression and an attack of Muslims in general, linking this group to a terrorist attack. The Court therefore agreed with domestic courts in that this should not fall under the protection of article 10.

Similarly, in *Garaudy v. France*, a historian had authored a book in which he denied the Holocaust. The Court stated that the denial of well-documented crimes against humanity must be seen as the most serious racial defamation of Jewish people, and that such statements also incited hatred towards this group.<sup>47</sup> In both *Norwood* and *Garaudy*, the Court made this assessment in accordance with article 17 of the ECHR, that states that no person, group or State can rely on the protection of the convention to perform 'any act aimed at the destruction of any of the rights and freedoms' stated in the convention.<sup>48</sup> This means that the protection of expressions targeting, for instance ethnicity or religion, it not limitless, and some will be removed from the protection of article 10 altogether.<sup>49</sup> The Court has nevertheless made it clear that such remarks, even under article 10, must not go beyond certain limits, particularly as regards respect for the reputation and rights of others. In the case of *Jersild v. Denmark*, the Courts stated that 'it is of the utmost importance to combat racial discrimination in all its forms and manifestations'.<sup>50</sup>

Statements that generally constitute a breach of article 10 are such that incite or endorse violence against the state or citizens. This is seen in, *inter alia*, *Zana v. Turkey*, where a man proclaims support for the PKK, a terrorist organization, and the Court does not consider imprisoning him for such an endorsement in conflict with article 10. They found that it was legitimate in protecting national security and public safety, but also emphasized that the statement must be seen in its context of extreme political tension at the time of the statement.<sup>51</sup>

<sup>44</sup>*Gündüs v. Turkey* (App no 35071/97) 14 June 2004 § 51–52.

<sup>45</sup>*Gündüs v. Turkey* (App no 35071/97) 14 June 2004 § 49.

<sup>46</sup>*Norwood v. the United Kingdom* (App no 23131/03) 16 November 2004.

<sup>47</sup>*Garaudy v. France* (App no 65831/01) 7 July 2003 § 23. See also, for example, *Witzsch v. Germany* (App no 7485/03) 13 December 2005 (decision).

<sup>48</sup>For more on this balancing and the relation between article 10 and 17, see Inessa Shahnazarova, 'Criminalisation of Genocide Denial and Freedom of Expression' (2013) 1 International Journal of Human Rights and Constitutional Studies 322–340 (note 328–330) and Clotilde Pégiorier, 'Speech and Harm: Genocide Denial, Hate Speech and Freedom of Expression' (2018) 18 International Criminal Law Review 97–126.

<sup>49</sup>Also note the case of *Glimmerveen and Haagenbeek v. the Netherlands* (App nos 8348/78 & 8406/78) 11 October 1979 in regard to incitement to discrimination and the exclusion of protection under article 10 by virtue of article 17.

<sup>50</sup>*Jersild v. Denmark* (App no 15890/89 23) 23 September 1994 §§ 30–31.

<sup>51</sup>*Zana v. Turkey* (App no 69/1996/688/88) 25 November 1997 §§ 58–61.

Hateful expressions against Roma in the context of article 10 have also been actualized in case law from the ECtHR. In the case *Le Pen v. France* the applicant claimed that there had been an interference with his freedom of expression when he was convicted for statements about the Roma population. In an official speech the former leader of the National Front, Jean Marie Le-Pen, made remarks about Roma, indicated that they did not want to integrate into European societies. The Court noted that even though these remarks were made in a political context, and even though he referenced facts on alleged delinquency amongst Roma to support his case, the remarks could not be seen as having a value strong enough for the public debate, and the factual basis could not be seen as substantial enough to question the judgment made by national courts. The ECtHR expressed that the remarks by Le Pen must be seen as potentially invoking feelings of hostility towards Roma. Therefore, the Court rejected his complaint as manifestly ill-founded, and declared the application inadmissible.<sup>52</sup>

The ECtHR has also identified, through its case law, several groups that require special protection, such as Roma. In this context, the court has stated 'as a result of their turbulent history and constant uprooting the Roma have become a specific type of disadvantaged and vulnerable minority'<sup>53</sup> and that they, because of this, also should have stronger protection.<sup>54</sup> Several other groups have also been identified as vulnerable, due to other reasons, such as the mentally disabled or children, who in different ways can have difficulties protecting their rights.<sup>55</sup> So whom an expression targets can make a difference in deciding its legitimacy, and also, as seen above, cultural, historical, and religious factors can be relevant in deciding if an expression is protected under article 10.

### Swedish legislation in relation to hateful tweets in identified discourses

As seen above, the protection for expressions is strong but not unlimited, and, as seen above, highly contextual. If States enforce restrictions, limitations or penalties these must therefore be in order to uphold democratic values and be *prescribed by law*. In some cases, States are not only *allowed* to impose laws that protect people from victimization, they have an obligation to do so. Sweden, like other signatory states to the convention, has *positive obligations* as well as *negative obligations*, where the latter entails that States must, themselves, refrain from human rights violations. The positive obligations, however, also entails that States actively must take measures that protect and enable access to these human rights, which can include criminalizing behaviour that restricts people's rights under the convention or otherwise tries to undermine the protected values under the convention, and also that States provide law enforcement agencies

<sup>52</sup>*Le Pen v. France* (App no 45416/16) 23 March 2017.

<sup>53</sup>*D.H and others v Czech Republic* (App no 57325/00) 13 November 2007 § 182.

<sup>54</sup>*D.H and others v Czech Republic* (App no 57325/00) 13 November 2007 § 182.

<sup>55</sup>See for instance *K.U v. Finland* (App no 2872/02) 2 December 2008 about children, and *Alajos Kiss v. Hungary* (App no 30696/09) 20 May 2010 § 44 about the mentally disabled. See also a more extensive discussion about identified vulnerability under the convention in Lourdes Peroni and Alexandra Timmer, 'Vulnerable Groups: The Promise of an Emerging Concept in European Human Rights Convention Law' (2013) 11 International Journal of Constitutional Law 1056–1085; Therese Enarsson and Markus Naarttijärvi, 'Is It All Part of the Game? Victim Differentiation and the Normative Protection of Victims of Online Antagonism under the European Convention on Human Rights' (2016) 22 International Review of Victimology 123–138.

with legal tools to satisfactorily investigate crime.<sup>56</sup> This means that Sweden, for instance, should have an effective legal framework to handle hateful, racially motivated incidents, that also allows for a balancing with the freedom of expression.

As mentioned, the most prevalent single topic of tweets in this study is the one representing *general political debate about street begging*. Within this topic we found representations of this spectrum, with expressions that should fall under the protected area of article 10, such as those generally debating solutions to the issue of street begging or how politicians have worked with or discussed the matter, but also hateful and aggressive ones that might *not* fall under the protected area. It has not been possible to identify threats or statements about, or directed at, specific individuals in the discourses studied.<sup>57</sup> That means that the hateful and aggressive speech found will be analysed in relation to *agitation against a national or ethnic group*. This is criminalized and regulated in the Swedish penal code (SFS 1962:700), Chapter 16 Section 8, and states that, someone who, in public, spreads statements or other forms of communication that threatens or expresses contempt for a group of people based of their nationality, ethnic origin, religious beliefs or sexual orientation will be sentenced for *agitation against a national or ethnic group* (Swedish: 'hets mot folkgrupp').<sup>58</sup> The focus of this provision is, and historically was, to prevent defamation and hate propaganda against groups of people based on their national or ethnic heritage or religion. This provision had its origin in the 1940s when more political attention was given to the spreading of anti-Semitic propaganda.<sup>59</sup> In 1948 a bill was passed, criminalizing spreading such propaganda, and it has been illegal in Sweden since, and Roma as a group are one such group that is protected under this law.<sup>60</sup> Several cases of *agitation against a national or ethnic group* in Sweden has been against Roma,<sup>61</sup> but none of them has later been tried by the ECtHR.

However, in 2003 sexual orientation was included in the provision,<sup>62</sup> and one case of anti-gay propaganda, that was deemed *agitation against a national or ethnic group* in Sweden, has made it all the way to the ECtHR. In this case, *Vejdeland v. Sweden*, several people left anti-gay propaganda leaflets in school lockers, with statements such as AIDS and HIV having a foothold in society due to a 'promiscuous lifestyle' carried out by gay people, and claims that 'homosexual lobby organizations' try to down play paedophilia. The ECtHR stated that Sweden were within its right to sanction this speech, and that in doing so they did not restrict the rights under article 10 in an unlawful manner. This case is interesting both in a broader way in that it examines whether or not homophobic

<sup>56</sup>This development of positive obligations have mainly emerged through case law, and continues to do so: See for instance *Abdulaziz, Cabales and Balkandali v. United Kingdom* (App nos 9214/80; 9473/81 & 9474/81) 28 May 1985; *X and Y v. the Netherlands* (App no 8978/80) 26 March 1985; *Christine Goodwin v. United Kingdom* (App no 28957/95) 11 July 2002; *K.U v. Finland* (App no 2872/02) 2 December 2008.

<sup>57</sup>If such statements do exist, they could also be of a nature that would be criminalized and a racial motive could be seen as an aggravating circumstance when assessing penal value under Swedish law (Swedish penal code, Chapter 29 Section 2).

<sup>58</sup>The sentence for agitation against a national or ethnic group is imprisonment for, at most, two years or, if the crime is petty, a fine.

<sup>59</sup>Karin Åström & Görel Granström, 'Den svenska regleringen av hatmotiverade brott: i linje med internationella normer?' in *Jubileumsskrift till Juridiska institutionen 40 år* (2017) Örjan Edström, Johan Lindholm & Ruth Mannelqvist (Eds.) Umeå: Juridiska institutionen, Umeå universitet 285–301.

<sup>60</sup>See judgement by the Swedish Supreme Court: NJA 1982 s. 128.

<sup>61</sup>Over the years 2012 and 2016 there were between 12 and 44 reported incidents every year in Sweden of *agitation against a national or ethnic group* with anti-Roma motives. The Nation Council for Crime Prevention (BRÅ), 'Rapport 2017:11, Hatbrott 2016, 'Statistik över polisanmälningar med identifierade hatbrottsmotiv och självrapporterad utsatthet för hatbrott' 77.

<sup>62</sup>See SFS 2002:800.



statements can amount to hate speech (the Court ruled not in this case), but also since the Court declares that:

[...] inciting to hatred does not necessarily entail a call for an act of violence, or other criminal acts. Attacks on persons committed by insulting, holding up to ridicule or slandering specific groups of the population can be sufficient for the authorities to favour combating racist speech in the face of freedom of expression exercised in an irresponsible manner<sup>63</sup>

Therefore, this case demonstrates that signatory States have a real possibility in limiting hateful expressions against individuals or groups, even if the expressions do not, for instance, incite to violence.

Just like the case law from the ECtHR, the Swedish case law reflects the contextuality that needs to be taken into account when handling cases concerning freedom of expression. What constitutes an expression according to article 10 is very wide, and basically anything can be an expression under the ECHR<sup>64</sup> and the Swedish penal code, in regard to the crime of *agitation against a national or ethnic group*. This means that tweets clearly stipulate *expressions* in that sense and would as such fall under such regulation. But as shown, whether or not individual statements expressed in the tweets could be seen as *protected* by freedom of expression must be assessed on a case to case basis. Tweets, even aggressive and offensive ones, connected to the identified anti-begging discourse for example, containing criticism about a political situation, or about how political parties handle the question of begging Roma, for instance, could then be seen as having a stronger protection. This can however be contrasted with the previously mentioned *Vejdeland v. Sweden*, where the claimed purpose of leaving such leaflets in schools was to stir up a political debate of the 'objectivity' of what the school was teaching, but looking at the entire context of the situation – that the recipients of the messages were children, that the school was an arena relatively free from politics and that the leaflets were offensive in an unmotivated way – caused for a sanction against it.<sup>65</sup>

But, some expressions would – according to the analysis of ECtHR case law above – be seen as having significantly weaker protection under the convention, and Swedish law. Those are generally hateful ones, that do not contribute to a democratic discourse, and has no value in a political debate, and that also attack a minority such as Roma. Such expressions like the previously mentioned tweets comparing beggars with vermin, or that 'gypsies' do nothing but deceive, lie and beg. To decide if such expressions could be seen as *agitation against a national or ethnic group* a first step is to consider whether or not a tweet could be seen as spreading a statement in public, and according to both case law and preparatory work to the existing legislation, this form of publicly posting a statement on Twitter could – in general – be seen as *in public*. This is due to the fact that it should be enough with a quite limited group of people for this to be actualized, as long as it is more than a few, and not strictly in a private conversation between a group of people.<sup>66</sup> The potential group of people that could access public tweets must certainly be regarded as large enough.<sup>67</sup> It is also not required that the people that in some

<sup>63</sup>*Vejdeland v. Sweden* (App no 1813/07) 9 February 2012 § 55.

<sup>64</sup>*Hashman and Harrup v. United Kingdom* (App no 25594/94) 25 November 1999.

<sup>65</sup>See the judgement from the Swedish Supreme Court: NJA 2006 s. 467.

<sup>66</sup>See judgement by the Swedish Supreme Court: NJA II 1988 s 541.

<sup>67</sup>See a similar reasoning in the judgement by the Court of Appeals (Svea Hovrätt), Case nr. 11651-17, date 2018-01-22, where posts on Facebook were made within a closed group that had thousands of members.

way were targeted with this message – like all of the tweeters’ own followers or the followers of a certain hashtag – necessarily accessed the content.<sup>68</sup> However, to establish if tweets like the above mentioned could fall under the scope of *agitation against a national or ethnic group* one must also consider if such statements *threaten* or express *contempt* for a group of people, the Roma, in this case. And the Swedish case law must relate to the case law of the ECtHR, meaning that Swedish courts in these cases must make an overall balancing between the freedom of expression and the protection of vulnerable groups, and in doing so seeing to the wider discursive context in individual cases, with factors such as the content of the expression, who it is targeting, the context in which it was stated and who is spreading it.<sup>69</sup>

The courts in Sweden have handled several similar cases over the last decade, where people have been charged for hateful speech online. One such case is that of a politician posting negative comments about the Koran and Islam. The overall message of the post was that Muslims coming to Sweden would consider it their right to rape women, and that this was in accordance with the Koran. This post was made in a clear political setting, since it was posted on a Facebook page of a political party where he himself was active as a politician, but the Court of Appeals stated that this post was not part of an ongoing debate in the party or related to a clear political idea, but instead was offensive to the targeted group. The court also stated that there, of course, must be room for a critical discussion about religion, but that the statements made by the politician not could be seen as such.<sup>70</sup>

In another case, a man posted several posts on Facebook in which Muslims, Africans and refugees were targeted with allegations such as raping woman and children, that they were traitors in their home countries and in different ways implying that they trick the Swedish state or Swedes for money. This was regarded as *agitation against a national or ethnic group* due to the facts that he clearly was portraying Muslims, Africans and refugees as living criminal lifestyles, it was not within the scope of objective criticism and he must have understood the great reach these messages were going to get.<sup>71</sup>

In a similar case, a man expressed disrepute towards Muslims and Africans on Facebook, claiming, *inter alia*, that there was a connection between immigrating Muslims and the number of rape-cases in Sweden. The court devoted some reasoning to the question on the validity of the statements. The accused claimed to believe that the statements that he had made were factually true, and that he did not intend to agitate against these groups. However, the Swedish court ruled that he had not provided any evidence for these claimed factually correct statements, and he was therefore convicted.<sup>72</sup> Such assessments by the Swedish courts could be seen as in accordance with how the ECtHR have reasoned over the years, where the Court has differentiated between expressions that are purely a value statements, which are subjective and cannot be proven, and such expressions that claim to be factually correct. When making such statements a

<sup>68</sup>See for instance judgement by the Swedish Supreme Court: NJA 1999 s.702 and The Swedish Government, ‘Prop. 2001/02:59 Hets mot folkgrupp m.m.’ 15.

<sup>69</sup>See judgement by the Court of Appeals (Svea Hovrätt), Case No. 11651-17, date 2018-01-22, and judgements by the Swedish Supreme Court: NJA 2006 s. 467 och NJA 2007 s. 805.

<sup>70</sup>Court of Appeals (Svea Hovrätt), Case No. B 4509-16, date 2016-12-20.

<sup>71</sup>Court of Appeals (Svea Hovrätt), Case No. 11651-17, date 2018-01-22

<sup>72</sup>Court of Appeals (Hovrätten över Skåne och Blekinge), Case No. B 1632-17, date 2017-11-06.

person could be asked to prove or show that there are valid reasons to make such a statement.<sup>73</sup>

This exemplification shows that tweets, similar to the studied ones, in some cases could be seen as illegal under Swedish law, especially when they do not contribute to a political discussion, are beyond objective criticism, not factually true (or substantiated) and degrading to, for instance, Roma. Worth noting in regard to both Swedish legislation and the handling of this issues in Swedish courts, as well as the case law of the ECtHR, is that cases like those mentioned above rarely lead to more severe penalties than fines. The proportionality of the criminal sanction and the offense will be examined extra closely by the ECtHR, and imprisoning someone for such offenses will be harder to justify for the signatory States. The seriousness of the offense and any previous offenses by the applicants will be taken into account in doing so.<sup>74</sup>

## Conclusions

It is important to keep a strict balancing in regard to freedom of expression, where individual victims or targeted groups also have strong protection. Since, in not doing so, individuals as well as larger communities may suffer great emotional damage, and loss of freedom themselves, such as withdrawing from social media or hiding parts of themselves as not to be subjected to hateful behaviour. From this study, it is clear that the legal protection under the ECHR and Swedish law for such types of hateful expression against a group of people such as Roma, on basis of their cultural or ethnical origin for example, varies a great deal depending on the wider discursive context. When an expression can be seen as a direct part of a political discussion, is presented in a way that is not unnecessarily offensive, or when it presents objective criticism that is in some way factually true and substantiated – then the protection of freedom of expression, such as a tweet about Roma, must be higher. Even though that same tweet may be perceived as offensive to a reader or viewer.

However, it is not in any way impossible to protect people and groups of people from hateful speech against them, on the contrary the signatory States must do so. Many of the studied tweets about Roma, if not all, could be said to have some kind of basis in a political debate, and could therefore have a stronger protection, but as shown by, for instance, both the case of *Le Pen v. France* and that of the Swedish politician, the courts will place emphasis on the aim, value, and accuracy of the statement as well, making it possible to legally intervene against speech that is perhaps triggered by an ongoing political debate, but hateful in nature and without value to that debate as such. That a statement was made within the scope of a political discourse may not make the tweet itself a political statement. That would also mean that tweets containing, for instance, political hashtags will not automatically become ‘political’ in nature. The complexity and the fact that any tweet, or other expressions online, must be viewed and assessed individually to be able to take into account the surrounding context and discourse, also makes it hard to give clear guidelines for what expressions are allowed and not. That is a challenge for courts

<sup>73</sup>David Harris and others, *Harris, O’Boyle & Warbrick: Law of the European Convention on Human Rights* (Oxford University Press 2014) 697–707.

<sup>74</sup>Bernadette Rainey and others, *Jacobs, White & Ovey, The European Convention on Human Rights*, (Oxford 2014) 451.

and law enforcement agencies faced with countering hate speech. In a time where large social media platforms are criticized it is also worth noting that it can be a challenge for them as well, since it in some cases can be very difficult to draw a line between offensive, but protected speech, and plain hateful speech.

In regard to the arena of the crime, i.e. hateful speech online, one could also raise the question of what type of communication tweets really are? Are they always as static as we tend to perceive written messages as, where people should be expected to have time to take a step back and reflect on the wording and expressions used? Or could tweets, when they are part of a continuous debate on Twitter, be seen as a kind of 'live-debate' in line with the case of *Gündüs v. Turkey*? No answers can be given to that questions here, but it is clear that the internet truly has changed and challenged the landscape of communication and therefore also have challenged and will continue to challenge the legal handling of communication between people.

Legally we are, and have been for several years, facing a new arena – in Sweden and in the rest of Europe. To access this arena and, if necessary, make legal adjustments we might need to expand our methodological approach. The tools used for this study could help legislators identify problem areas that need addressing (legally) and could also be used as a political tool in terms of showing the discourses surrounding certain people or groups of people, in order to address those discourses.<sup>75</sup>

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by Brottsofferfonden [grant number 2605/2013] and Vetenskapsrådet [grant number 2016-04195\_VR].

---

<sup>75</sup>Note of course, that there are other ways to find out if and what kind of illegal, or potentially illegal, content exists on social media platforms, such as statistics of reported crimes, interviews or questionnaires focusing on victimization etc. However, the tweets that were identified as potentially unlawful were mainly to be considered as hate speech, which in Sweden is believed to be massively underreported. It is important to reveal some content behind these hidden statistics. See Sveriges Radio, (Gustav Wirtrén) 'Stort mörkertal för hatbrott på nätet' (18 October 2018) <<http://sverigesradio.se/sida/artikel.aspx?programid=96&artikel=6801370>>.