HATE SPEECH IN SOCIAL MEDIA

Shomaila Sadaf

Master's Thesis

Intercultural Management and Communication

Department of Language and Communication Studies

University of Jyväskylä

Spring 2020

UNIVERSITY OF JYVÄSKYLÄ

Tiedekuta-Faculty Faculty of Humanities and Social Sciences	**Laitos-Department** Department of Language and Communication Studies
Tekijä–Author Shomaila Sadaf	
Työn nimi-Title Hate speech in social media	
Oppiaine-Subject Intercultural Management and Communication	**Työn laji-Level** Master’s thesis
Aika-Month and year March 2020	**Sivumäärä-Number of pages** 82 + 1 appendix

Tiivistelmä – Abstract

Social networking sites (SNS) play a substantial role in facilitating online communication and social interaction a global scale. Simultaneously, these social networks become platforms that not only spread the rhetoric of hate but also normalize it. This study systematically uncovers the existing literature related to hate speech in social media. The focus of this study is to explore the approaches used in data collection and data analysis along with theoretical frameworks used in the existing studies.

The method used in this study is a systematic literature review, and a total of 30 articles met the inclusion criteria and are used for the final analysis. The findings related to data collection and data analysis methods show that the most common data collection method is that, the data is taken directly from online social networks in the form of user comments, content shared on these networks, tweets, blog posts, etc., while the quantitative data is collected through surveys and questionnaires. Analysis further show that the qualitative studies provide in-depth descriptive analysis of the discourses online, whereas few studies used quantitative data analysis methods. The two most common areas of research on hate speech online according to the analysis are political and ethnic/racial issues. A total of 17 studies out of 30 used theories or models to support their central ideas. The present study further aims at exploring the use of positioning theory with regard to hate speech in social media. The analysis reveals that only one study has used positioning theory as the theoretical base, and only one article utilized the idea of positioning analysis as a discursive process at various levels.

Considering the basic aim of this study, important data is gathered within the area of hate speech in social media and information is extracted that would help in exploring various other areas for future research.

Asiasanat-Keywords
Hate speech, social media, positioning, positioning theory

Säilytyspaikka-Depository
University of Jyväskylä

Additional information

LIST OF TABLES

TABLE OF CONTENTS

1 INTRODUCTION

“Social networks are the frenzy of the twenty-first century” (Alkiviadou, 2019, p. 19). Egalitarian in nature, the Internet is a communication medium that has the potential to communicate beyond borders. It also has the defining traits of being interactive, globalized and decentralized (Banks, 2011). The Internet is considered a global giant, also known as ‘a network of networks’ (Vajić & Voyatzis, 2012), which provides unique opportunities for communication through online social networks (Silva et al., 2016). These social networks allow individuals to interact at various levels.

Social networking sites (SNS) play a substantial role in facilitating online communication and social interaction on a global scale. Simultaneously, these social networks become platforms that not only spread the rhetoric of hate but also normalize it. While the sentiment of hate and hate speech existed long before the ascent of SNS, their rise has arguably introduced another dimension to the already existing complex phenomenon of hate speech (Timofeeva, 2002). Coliver (1992) refers to hate speech as any expression and manifestation that is directed to abuse, insult, intimidate or harass, led by an open or underlying message of violence, discrimination and hatred towards an individual’s belonging to a group of different race, nationality, ethnicity or religion, etc.

The Internet being multi-mediated in nature, including photos, videos, online games, words, etc. allows for various forms of communication. This also helps in conveying hatred and derogatory feelings aimed at a specific group of people (Foxman & Wolf, 2013). In this age of global media, where we have the right to choose our own personal media landscape, this sometimes makes us inclined to gravitate towards like-minded people. We tend to surround ourselves with copies of ourselves, meaning that we share similar thinking for matters under consideration. It has been argued that hateful and negative communication presents the biggest threat to the development of tailored communities and groups online (Altonen, 2017).

Hate speech online becomes detrimental and pernicious as it not only constructs but also politicizes ingroups and outgroups. In this process of hate, the outgroup is made the “other”, who is automatically detached from the dominant opinions of the ingroup (Gagliardone, 2014).

This leads to the concept of positions and positioning; where, an individual takes a particular position, and that person inexorably views the world from his standpoint (Davies & Harre, 1990). They further support the notion that individuals take positions in accordance to their own narrative experiences that include beliefs, emotional state, histories and schools of thought, along with the knowledge of their rights, duties, expectations, obligations and roles in the social structures they belong to. Positioning takes place in two phases; the first phase is prepositioning. One can preposition himself or the other by “listing and sometimes justifying attributions of skills, character traits [and/or] biographical ‘facts’, deemed relevant to whatever positioning is going forward” (Harré et al. 2009, p. 10), while positioning in the true sense takes place at the second phase when the real interaction starts.

This scenario of social media, positioning and hate speech made me realize that this topic is vast enough. And in order to understand these complex phenomenons I needed to start looking at the available literature systematically. Hence, I decided to do a systematic literature review.

1.1 Statement of purpose

The purpose of this study is to systematically find the existing literature related to hate speech in social media. To achieve this aim, multiple criterions are set, such as identifying the methods used in collecting and analyzing the data, the focus of the previous research studies, and theories and frameworks that are used by the research already completed on hate speech in social media.

1.2 Research questions

Drawing on the concept of hate speech and the theorizing done on positioning, this study aims to answer the following research questions:

RQ1: What approaches, viewpoints and methodologies are used to study hate speech in social media?

RQ2: How is positioning theory used in the context of hate speech in social media?

To gain insight and answers on the research questions, a systematic literature review is conducted. The results are presented in various sections, starting from the methodologies used in data collection and data analysis. Further, the focus of the studies is discussed, and finally, an in-depth review of the theories and frameworks used by the studies is presented.

Continuing from the introduction as chapter one, this thesis is further divided into the following chapters: chapter two comprises the theoretical framework, while the chapter three entails the methodology and elaborates on how the data is collected and finalized for analysis. In the chapter four, the results from the review are presented and discussed. The chapter answers the main research questions. Finally, chapter five provides a discussion of the results obtained by answering the research questions. The last chapter, chapter six explains the limitations of the study along with conclusions and future recommendations.

2 THEORETICAL FRAMEWORK

2.1 Hate speech

Hate speech is conceptualized as any expression that spreads, incites, promotes or justifies hatred towards a race, xenophobia or any other form of hate primarily based on intolerance expressed through aggression, discrimination and antagonism against minority groups and immigrants (Timofeeva, 2002). The basic motivation behind hate speech is prejudice towards an individual or a group of people who share similar characteristics of race, gender, sexual orientation, religious beliefs and so forth (Gagliardone et al., 2015). The concept of hate speech does not hold a single and common definition. According to Stakić (2011), there has been an extensive debate on hate speech in academic and political circles, but a universally established and agreed upon definition of hate speech does not exist. Typically, the concept of hate speech revolves around two main features, the tone and style in which the message is composed and the grounds towards which the message is directed. According to the Council of Europe (2013), hate speech:

> “Covers all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on in tolerance, including: intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility against minorities, migrants and people of immigrant origin”.

As per Nockleby (2000), the form of communication that belittles an individual or a group’s characterization on the basis of complexion, ethnicity, cultural background, nationality, creed or any other distinguishing feature, is defined as hate speech. Hate speech is further described as any degrading and abhorrent speech targeting a person or a group sharing similar attributes or ideology (Boeckmann & Turpin-Petrosino, 2002). The two elements that are common in most of the views are that hate speech is directed towards anyone who is distinguished as inferior on the basis of some innate characteristic including sex, gender, ethnicity, race etc., and that hate speech intends to aggravate violence, produce prejudiced treatment, and incite offence to the dignity of the targeted group(s) or individual(s) (Stakić, 2011).

The characteristics that shape the concept of hate speech are prejudice, negative stereotypes and stigma, and perceived hierarchies and boundaries between groups laid the foundation of hate speech. It is built on the rhetoric of elimination, fear and disrespect for individuals and groups that are different from their personal perspective (Perry, 2001). Perry further explains that the purpose of this behavior is to safeguard and highlight the perceived boundaries among the groups and to remind individuals and groups about them being "the other" in the social structure. Hence, in order to understand hate speech, the tone of the message, the rhetoric built around the speech and the target of the speech needs to be examined.

Various scholars agree that hate speech strongly expresses, promotes, advocates and encourages hatred towards individuals who are distinguished based on some particular features (Hernández, 2011; Townsend, 2014; Traum, 2014). This term refers to the verbal conduct and other communicative and symbolic actions, which express intense hostility towards an individual or a group on the mere innate connection to that group (Simpson, 2013). As a matter of fact, hate speech is not always a verbal act, rather it is also expressed via nonverbal communication. Taking Waldron's (2012) work into account, it can be said that any expression that is considered hateful, for example, by the use of text, sound or images, its function is to dehumanize and weaken the members that belong to the target group.

Before World War II, discrimination and hate speech were often accepted in one form or another. Hate speech is strictly regulated in the world except the USA after the Second World War (Bleich, 2011; Parekh, 2006). The definition of hate speech is modified and used in different countries, and all these countries have passed national and international regulations regarding the use of hate speech (Gagliardone, et al. 2015). For example, Norway has a strict stance against the use of hate speech. The Norwegian Penal Code section 185 defines and protects individual and groups from hate speech and discrimination based on skin color, ethnic background, nationality, religion, sexual orientation or disability. But this characterization does not mean that any other expression that is hateful towards individuals and groups is allowed; rather, they are taken into consideration under some other rules and laws that include the laws of defamation and threat of law on discrimination (see Wessel-Aas et al. 2016). Hence, hate speech intends to hold a strong message for the receiver.

Hate speech is always disseminated face to face or through some medium. The Internet is one of those platforms that allow communication among individuals, most evidently through social networking sites. Hate speech online has been escalating and activists have been expressing their apprehensions towards social networking sites due to their usage for spreading various forms of discrimination (Simon Wiesenthal Center, 2012). For a long time, social media operating companies have not done much to keep their platforms free of hate speech; as a result, these platforms end up being major hubs of hate speech (Knowledge-Wharton, 2018).

2.2 Internet and social media

“The Internet is the decisive technology of the Information Age” (Castells, 2014, p. 127). In today’s globalized world, people’s lives are significantly affected by the Internet. On October 24, 1995, the Federal Networking Council (FNC) defined the Internet as a “global information system”. According to the FNC, the Internet is linked together through internet protocols. It supports the transfer of messages online by using Transmission Control Protocol (TCP) or Internet Protocol (IP). These protocols are the rules that govern the movement of data from the source to the receiver or the internet (“The TCP/IP Reference Model”, n.d.). It also offers accessibility to an improved level of services that depends on communication and the infrastructure related to it (Leiner et al. 2009).

The development and extension of the Internet has created numerous openings for individuals to communicate and participate in the social networking platforms. This development picked up speed in the early 2000s, and could be seen, for example, in the creation of Friendster in 2002. Later on, Facebook, Instagram and many other social media platforms solidified the idea of social media on the Internet. Today, all kinds of human activities are taking place on these social networking sites, ranging from personal and social interaction, politics, work, business, etc. (Castells, 2014).

Since the networking sites facilitate communication, they have become an integral part of our daily routine. The social media platforms have also transformed the users from being passive to an active audience, who hold authority to comment publicly on the events they are interested in. According to Allen (2012):

> "Today social media is beginning to change the form and nature of 'the media' in turn presenting many new and different challenges. In the social media sphere, we have recently seen existing boundaries being pushed, not just in what can and cannot be said, but so too by whom and to which audiences." (p. 3).

2.3 Hate speech online

Hate speech is a commonly occurring phenomenon on the Internet (Kettrey & Laster, 2014). Along with social media's significant role in negotiating communication and social interaction on a global scale, it has also facilitated negative behavior (Oksanen et al., 2014). Individuals use the social media space for addressing a wider audience by using hate disguised by anonymity, letting them surpass and circumvent editorial control and regulations (Citron, 2014). Consequently, the Internet becomes a platform that provides opportunities for cyber hate (Jaishankar, 2008) and cyber bulling (Kowalski, et al., 2012). As the sentiment of hate and hate speech continue to grow online (Sood et al., 2012a), the social media platforms continue to encounter the problem of recognizing and censoring offensive posts (Moulson, 2016). People are still not well aware of the content that falls under hate speech (Ma, 2015). Groups are targeted systematically, which affects the world around us at individual, group and societal levels (Brennan, 2009). Various social networking sites may become a space for spreading hate online, and their visibility enhances, as they are used by a significant number of users (Oksanen et al., 2014). For example, in the US between 2000 and 2010, the active hate groups online increased by 66% and there were more than 1000 active hate groups online in 2010 (Potok, 2011, p. 41).

Victims of hate on the internet have varied experiences (Awan & Zempi, 2015; Chakraborti & Garland, 2009). Through hate online, victims are harassed and intimidated, along with experiencing devious crimes (Christopherson, 2007). Hence, the Internet has proven to be an important tool, holding the power to influence the users to behave in a specific manner. According to Iganski (2012), online hate crime can become a means of creating space for communicating messages whose effects can be witnessed in the physical world, well beyond the virtual world. Coliandris (2012) suggests that hate crime perpetrators are capable of targeting a particular community. The early adopters of the Internet have used this medium as a tool for building communities, reaching newer audiences and making new members (Gerstenfeld et al., 2003). Likewise, some of them have also used social networking sites for propagating racist propaganda and inciting violence offline (Chan et al., 2014).

Social media tends to operate as a corporate platform that helps in defining hate speech, establishing a code of conduct and its implementation. Foxman & Wolf (2013) argue that since the popularity of social media platforms like Facebook, Twitter and YouTube is on the rise, the challenges related to hate speech on these platforms are also significant. Hate speech online use electronic communication technology to spread hate messages and information that is related to ethnicity, religion, etc. Websites, blogs, social networking sites, email, instant messages, WhatsApp, etc. all constitute electronic communication technologies. In order to address these challenges, various legislations and regulatory policies are designed to protect freedom of expression and distinguish hate speech from free speech (Banks, 2010). He further reports that there has been a gradual increase in the number of ethnic hate groups online along with activities related to hate speech online. By October 2019, there were almost 4.48 billion internet users, meaning that, "58 percent of the global population was active internet users" (Clement, 2019). According to statistics, "around 2 billion internet users are using social networking sites" and these figures are expected to rise as there is a significant increase in the usage of reformulated mobile devices and mobile social networks (Clement, 2019). This can also be linked to the remarkable surge of Internet usage.

According to Banks (2011), the Internet has the potential and ability to virtually cross borders and break the barriers of "real life". Along with the benefits, there are some perils linked to this ideology. What makes the Internet an important tool for promoting hate speech is the underlying characteristics of anonymity and immediacy, along with its global nature. The interaction between individuals is characterized by polarization. They connect with each other by putting them into certain blocks that may or may not differ from themselves. This clear distinction between us and them, insider and outsider, normal and deviant defined "the other" (Staszak, 2008).

2.4 Positioning "the other" online

The term positioning has its roots in Foucault's (1969) idea of "subject positions" that can be occupied in certain discourses (Depperman, 2015). While the idea of positioning in social psychology was first used by Wendy Hollway in 1984. She is regarded as one of the first scholars to use the notion of position and positioning, people take up when negotiating gender related places in discourses. She considers positioning as an analytical tool that can help in understanding how individuals see themselves in interactions focused of gender and sexuality (Depperman, 2015).

Davies and Harré (1990) are the first ones to bring positioning to bear on interactive exchanges and to relate it to narratives. According to them, positioning is the basic mechanism by which a self and identity is acquired in social interaction in terms of practical, emotional, and epistemic commitment to identity-categories and associated discursive practices. They argue that position is "the appropriate expression with which to talk about the discursive production of a diversity of selves" (p. 47). Davies and Harré (1990) further explain:

> "Once having taken up a particular position as one's own, a person inevitably sees the world from the vantage point of that position and in terms of the particular images, metaphors, story lines and concepts which are made relevant within the particular discursive practice in which they are positioned." (p. 46)

Positioning theory as defined by Harré and Langenhove (1999) is a "study of local moral orders as ever-shifting patterns of mutual and contestable rights and obligations of speaking and acting". It revolves around intergroup relations, identity construction of the individual and individual narratives, and analyzes the fact that individuals participating in interaction easily change positions (Harré & Moghaddam, 2003; Harré & Langenhove, 1999). When there is a change in the situation, interactants are considered as active agents who tend to construct and change interactions. The process of positioning is like a thread that weaves social interaction and wraps the entire interactive situation.

Positioning signifies the activity in which competent individuals are positioned within a system of rights and obligations through interaction. Hence, positioning takes place during socialization and unfolds during interaction. In this respect, positioning and socialization tend to be synonyms. Positioning in interaction corresponds or amounts to a form of socialization

(Tirado & Gálvez, 2007). Harré and Langenhove (1999) further elaborate the concept of position as a:

> “cluster of generic personal attributes, structured in various ways, which impinges on the possibilities of interpersonal, intergroup and even intrapersonal action through some assignment of such rights, duties and obligations to an individual as are sustained by the cluster". (p. 1)

Positioning is a relational process that is formed in interaction with other people (Hollway, 1984). For positioning, the links and continuity between different episodes of interaction are very important. There is strong continuity between positionings if they interactional episodes in which they occur immediately follow upon one another (Harré & Moghaddam, 2003). Episodes hold an important place in positioning theory, as they helped in shaping social reality. A complete picture, making sense and meaning, was the compilation of episodes based on a series of interactions. Episodes were derivatives of social interactions and helped define social reality (Harré and Langenhove, 1999). In every episode there were two main elements: position and positioning. Position was the relationship between the self and the other, while positioning was the result of positions and their negotiations. Position is never static; it is negotiated, and changes according to the opinions of others.

Positioning theory uses triangulation of three units of analysis in order to look at discourse. First unit is positions, where rights and duties are determined as acts in a storyline. While the second unit is speech-acts described as expressions with illocutionary force. They help in shaping the storyline. And the third unit is the storyline which is unfolded in episodes (Warren & Maghaddam, 2018). Potter and Wetherall (1987), in order to perform discourse analysis use the idea of illocutionary force in speech act. In discourse analysis interpretive repertories or patterns are searched in the transcribed scripts. While these techniques are approached by the social scientists from the critical movement as critical discourse analysis, where positions of power are assigned. Later it analyses the discourse to develop the understanding on the use of language for promoting the power of one group over the other.

Discursive practice is the rudimentary idea behind positioning theory; the background in this regard is provided by Bakhtin, Benveniste and Wittgenstein (Harré & Secord, 1973). Speech acts and social actions are the core issues when analyzing social reality. Not having any specific structure, they are connected and associated to each other through pace and rhythm involved in the specific interaction. Conversations, institutional practices and the use of rhetoric are the three things in discursive practices where social reality is raised. And these conversations are essential to social reality, where the reality of everyday is made, reproduced and transformed.

Positioning theory conceptualizes and studies discourse as the institutional use of language. This institutionalization of language occurs at various levels that include disciplinary, cultural, political and small group levels (Krogh, 2016). Discourse as a process tends to be dynamic in nature, which is neither intended nor confined to a particular space. It actively constructs, acquires and transformes meanings. Discourse is characterized by its ability to provide its subject a position (Tirado & Gálvez, 2008). On account of this idea, the theory claims that positioning is the product of conversation. Positioning is a dynamic process that adapts to changes easily. Changes in positions depend on narratives, images and metaphors by which they are made and constructed.

Another important element that we need to closely look at is the sociolinguistic symbols that people use to position themselves and their audience. According to Davies and Harré (1990), when a person takes on a position and owns it, he views the world around him from the specific viewpoint of his position. Being in the position of his role, certain concepts, images, metaphors and storylines become relevant to him. Therefore, the act of positioning, is the discursive construction of personalized stories by allocating roles and duties to one's own self and the audience (Harre´ & Langenhove, 1999). Positioning theory is also used to analyze interactions that take place online. It is further used in studying how stereotypes are produced and how social identity is created (Sabat & Harré, 1999) and how intergroup relations are developed (Tan & Moghaddam, 1999).

Positioning can be an important conceptual and methodological tool to study interaction in social media. As already discussed, positioning in interactions is considered as a discursive and narrative phenomenon that keeps on changing according to the context. Positioning in interactions on social media work the same way. It can be a helpful tool to study conflicts in social media, and hate speech is a type of conflict that also takes place online. Tirado and Gálvez (2008) suggest that positioning is a model with the help of which we can analyze conflicts. They further explain it as a situationally developed interactive process, whose analysis is based on agent's active role in the process.

This literature about the key concepts that we intend to explore in this study provide a detailed background of how diverse the concepts of hate speech, social media, hate speech online and dynamics of positioning theory are. We were able to identify some gaps, and in order to make the understanding of the topic under discussion more substantial, we will now move on to conduct a systematic literature review based on clear research questions.

3 METHODOLOGY

This section provides an overview of the systematic literature review in detail, followed by an explanation of the methodology used in this study. Later, the process of systematic literature review is applied to the finalized data.

As an important research methodology, the systematic literature review gained popularity in the 1990s. A systematic literature review is important in research as it provides objective outlines of previously researched topics. Systematic literature reviews are valuable in those research areas where literature already exists and publications focus on certain aspects of the field (Budgen & Brereton, 2006).

According to Wright et al. (2007), the systematic literature review is an analysis of the corroboration of a distinctly devised question. It uses explicit and systematic methods to determine, choose and critically assess the most relevant primary search. It then helps in extracting and evaluating the data that is included in the review. Kitchenham and Charters (2007) defines a systematic literature review as:

> “a means of identifying, evaluating and interpreting all available research relevant to a particular research question, or topic area, or phenomenon of interest" (p. 3).

He further elaborated on the reasons for conducting a systematic literature review, which are:

- For encapsulating and summarizing the existing literature about the topic under study.
- To find gaps in the research.
- To help define a framework that can aptly position the updated research activities.

The basic motivation behind the current research is to understand hate speech online thoroughly. To obtain a profound understanding of the said concept, its definitions and the various elements that constitute hate speech online are explored. Its working and ways of circulation in the social media are further investigated. This systematic literature review in particular is aimed at:

- Understanding the concept of hate speech in social media.
- Collecting and summarizing the available studies concerning hate speech in social media.
- Identifying the gaps that exist in the published research about hate speech in order to suggest areas for future research.

Kitchenham and Charters (2007), also define the important features of a systematic literature review, which according to them differs from a conventional literature review; the factors that contribute to making a systematic literature review valuable are:

- Review protocol. This is the main component of a systematic literature review; it postulates the research questions under consideration and the methods used in undertaking the review.
- Conducive research strategy. This is for extracting the maximum amount of relevant literature.
- Documentation of research strategy and results for future research.
- Assessing the criterion defined for the inclusion and exclusion of research to be finally included in the study.
- Extract information through data extraction forms and tools to provide consistent and desired information.

To find the answer related to the research questions, a systematic literature review method is employed. Specifically, this study presents a synthesis of research about hate speech in social media with a special focus on positioning. As a method, a systematic literature review has the potential to let the researcher extract general information and details about a research topic.

Tranfield, Benyer and Smart's (2003) three phase process is used to conduct a systematic literature review on hate speech in social media. This process is divided into "planning, conducting and reporting". Over the next paragraphs, each phase will be presented in detail.

3.1 Phase 1; Planning

In the first phase, a review panel is formed, the purpose of which is to find experts from the field who define review protocols along with the inclusion and exclusion criteria for the literature to be used (Tranfield et al. 2003). Since they do not define any particular size of the review panel, a team of at least two reviewers is proposed by Carter and Ellram (2003). The activities in the review process should not be planned meticulously - rather they should have enough flexibility to adjust to the needs of the review (Tranfield et al. 2003). The flexibility in protocols for a systematic literature review are inherent to the process (Moher et al. 2009).

For the present study, the review panel consists of myself and the supervisor of the study. In addition, input from colleagues in seminars and other discussions may be counted in this stage, as they informed the final decision-making. The inclusion/exclusion criteria for this study are:

1. The search concentrated on research articles published between January 1, 2010 and December 31, 2018. The reason for choosing this time frame is that around 2015, there was a significant influx of immigrants in the European countries, and many hate movements were also mushrooming in these countries, where social media is used as an active tool.

2. Research articles are published in peer-reviewed scholarly journals. Peer-reviewed journal articles provide a “critical and ethical assessment of the quality of a manuscript” (Swartz, 2008).
3. Only research reported in English is included. Thus, restricting the studies to the English language carried the risk of not taking potential data into consideration.

3.2 Phase 2; Conducting

In the next stage, the search terms need to be formulated. According to the guidelines of the University of York (2009), the search terms should be articulated considering the scope of the study and they should support the research with respect to obtaining the answers to the research questions. Defining a search strategy is also crucial, which is done after deciding the search terms. A search strategy typically includes the measures that are taken to detect the relevant literature that answers the research questions and thus holds critical value for validity of the findings and the success of the review (Bettany-Saltikov, 2010).

In this study, we received help from a professional librarian at the University of Jyväskylä library for matters such as choosing databases as well as search terms. The key search terms are: positioning theory, hate speech and social media (including combinations of those, e.g., “positioning theory” AND “hate speech” AND “social media”; “positioning theory” AND “hate speech” etc.).

Three publication databases are searched. These databases are suggested by the librarian as well. These are up to date and established research databases that include a wide range of peer reviewed journals. A variety of keywords are used to find the articles in the three selected databases.

i. EBSCOhost’s Academic Search Elite and Communication & Mass Media Complete databases.
ii. Directory of Open Access Journals database.

iii. ProQuest

The details of the search are mentioned below:

Databases	**Results**
EBSCOhost's Academic Search Elite and Communication & Mass Media Complete databases:	i. "Positioning Theory" AND "Hate Speech" AND "Social Media" No results found. ii. "Positioning Theory" AND "Hate Speech" No results found. iii. "Positioning Theory" AND "Social Media" 2 results. iv. "Hate Speech" AND" Social Media" 17 results. v. Positioning AND "Social Media" 42 results.
Directory of Open Access Journals database (DOAJ):	i. "Positioning Theory" AND "Hate Speech" AND "Social Media" No results found. ii. "Positioning Theory" AND "Hate Speech" No results found. iii. "Positioning Theory" AND "Social Media" 22 results. iv. "Hate Speech" AND" Social Media" 32 results. v. Positioning AND "Social Media" 128 results.
ProQuest	i. "Positioning Theory" AND "Hate Speech" AND "Social Media" No results found.

	ii. “Positioning Theory” AND “Hate Speech” No results found. iii. “Positioning Theory” AND “Social Media” No results found. iv. “Hate Speech” AND” Social Media” 7 results. v. Positioning AND “Social Media” No results found.

Table 1 The initial search results and the frequency of hits per search term

A 12-step framework has been developed by Kable et al. (2012) for conducting a literature review. It is a structured approach for the formulation and documentation of a search strategy. The primary focus of the framework is on the elements that need to be documented in the manuscript so that the specific strategy can be replicated by other researchers. Detailed documentation of the search strategy helps the readers in understanding and comprehending the rationale of the study. Another benefit of the framework is that it directs the reviewers through its development phase and warrants that all the important aspects are incorporated in the review. Thus, this framework is considered to be a valuable tool for new researchers.

The 12-step framework has the following suggested steps, of which all the steps have been followed except for the 10th step i.e., quality assessment of retrieved literature in this review. Since peer reviewed articles are used in this systematic review, the need for assessing the quality of the retrieved literature is nonobligatory.

1) Purpose statement

2) Databases, search engines used

3) Search limits

4) Inclusion and exclusion criteria

5) Search terms

6) Exact searches per database, search engine and the results

7) Relevance assessment of retrieved literature

8) Table reporting literature included in the review, accompanied with key data such as title, author, but also research subject and findings

9) Document final number of search results

10) Quality assessment of retrieved literature

11) Review

12) Accurate, complete reference list

Kable et al., (2012)

At the very end, when the literature is retrieved and the search process is finished, the data needs to be assessed particularly for relevance. This can be done by going through the title and abstracts and comparing them with the inclusion and exclusion criteria already set (Bettany-Saltikov, 2010). Then comes the second stage of assessment, where all the studies that have passed the preliminary phase and qualify the first round are scanned thoroughly. This process helps in saving time and energy as we do not thoroughly go through the literature retrieved, rather shortlist the articles by going through their titles and abstracts only. By doing this a large bulk of literature can be evaluated rather quickly. We followed the same procedure and examined the titles and abstracts in the search database. By doing so, we were able to shortlist the articles that are included in the second stage of assessment. A total of 30 articles met the inclusion criteria, as presented in Appendix 1.

Finally, the researcher scans the literature for specific information and records the information that he gains from reading into a form. The form is used to list the answers related to the research questions. Wynstra (2010), in his review paper, has provided examples of the categories that he uses in data extraction form. The main categories he employed are topic, data collection, analysis of data, product type, purchase type etc., and the main categories are further divided into subcategories.

In the present study, after the initial screening, the full text of the 30 articles is reviewed in detail. The following data is extracted from each included article: aim/focus of the research, theory/framework used in the research, method used for analysis (qualitative/quantitative), major findings of the research and future recommendation (if any).

3.3 Phase 3; Reviewing

The final stage of a literature review is the synthesis phase that summarizes all the findings extracted in the previous stage. According to Tranfield et al. (2003), there are two methods by which data synthesis can be done: narrative and meta-analysis. A narrative synthesis simply helps in identifying what has been written and researched on a topic or area earlier (Greenhalgh, 1997). while meta analysis helps in obtaining reliability by synthesizing the findings from various studies (Tranfield et al., 2003). The present study uses narrative synthesis, as it suits the research aim. Over the following chapter, the findings and their subsequent analysis are presented.

4 ANALYSIS AND FINDINGS

4.1 Data collection methods used in studying hate speech in social media

The complete list of approaches and methodologies used in data collection and data analysis along with the focus of the studies selected for analysis is offered in Appendix 1. The data collection methods used in the selected publications are summarized in Table 2.

Methods used	**Publications**
Various forms of qualitative data collection from online social networks that included reader comments on news websites, comments on social networking sites, content on Facebook pages, tweets, blog posts and messages on social media accounts.	Badarneh & Migdadi (2018); Sayımer & Derman (2017); Ben-David & Matamoros-Fernández (2016); Özarslan (2014); Ott (2017); Meza (2016); Aguilera-Carnerero & Azeez (2016); Uysal, Schroeder & Taylor (2012); Al-Tahmazi (2015); Schaffar (2016); Horbyk (2018); Maweu (2013); Abraham (2014); Burnap & Williams (2015)
Survey method: data collection through questionnaires, physically and online.	White II & Crandall (2017); Harell (2010); Piechota (2014); Pitsilis, Ramampiaro & Langseth (2018); Näsi, et al. (2015); Alam, Raina & Siddiqui (2016)
Various forms of qualitative data that did not involve content from social networking sites and news websites; rather, they were essays, document analysis, round table discussions and reviews.	Chetty & Alathur (2018); Mantilla (2013); Langford & Speight (2015); Shepherd et al. (2015)

Mixed methodology: data collection through questionnaires and blog posts, questionnaires and focus group discussions.	Kimotho & Nyaga (2016); Alakali, Faga & Mbursa (2017)
Ethnographic study of online video of a movement.	Yamaguchi (2015)

Table 2 Methods of data collection used in the selection of publications

Five main approaches related to data collection methods are recognized after summarizing the data from the shortlisted literature. The most common approach to qualitative data collection is that the data taken directly from online social networks. This includes, for example, reader comments on news websites and social networking sites, content on Facebook, Twitter and blog posts. Out of the 30 selected publications, 14 belong to the category where data is collected directly from social networking sites. Of all the social media platforms, Facebook and Twitter are used the most for data collection. Some studies focus on the content on Facebook pages and Twitter accounts, while the others focused on user comments only. Few studies used both Facebook and Twitter as the primary source of their data. Maweu (2013), for example, used a total of 30 hateful messages exchanged during the months of January to May 2013, on Facebook and Twitter, in order to examine the use of these platforms by the citizens of Kenya involved in political discussions online. Meza (2016) in order to examine the instances of hate speech in Romanian language comments to online media used user comments published on 25 Facebook pages along with 10 blogs. He also used comments on the news section of five online news websites between 1st January 2015 and 30th June 2015. Suntai and Targema (2017) also used information available on online media that included news websites, social media and web blogs about the general elections of 2015.

In addition, Badarneh and Migdadi (2018) used 500 reader comments as their sample data from eight different news stories during the years 2014 and 2015. They took two Jordanian news websites into consideration: Ammon News and Khaberni. Similarly, Horbyk (2018) collected data from Ukrains'ka Pravda (Ukranian Truth), which is a leading news website in Ukraine. He used 3000 reader comments posted on the evening of Euromaiden protests as his data. With these comments, he tried to investigate how ethnolinguistic identities were constructed online.

Ben-David and Matamoros-Fernández (2016) used the official Facebook pages of seven extreme right political parties and the majoritarian party PP in Spain between 2009 and 2013 as their sample data. Al-Tahmazi (2015), used comments on Facebook pages of Iraqi political commentators as the data for his research. While some studies used user comments and content available on Facebook pages, Schaffar (2016) used the screenshots of posts along with the user comments on the Facebook page of Rubbish Collector Organization in 2015.

Özarslan (2014) used Twitter as the source of data. He used the case study of hate speech against the Kurds situated in Turkey and used the tweets posted on 23rd October 2011 as his data. Ott (2017) used the twitter feed of Donald Trump on 10th November 2012, to explore the public discourse. Similarly, Arguilera-Carnerero & Azeez (2016), studied how an average netizen articulated Cyber Islamophobia through 10,025 tweets around the hashtag #jihad during the month of April, 2013. To study the use of Twitter as a public relations strategy by government officials, Uysal, Schroeder and Taylor (2012) analyzed the personal and official accounts of the top three Turkish government officials. Pitsilis, Ramampairo and Langset (2018) and Burnap and Williams (2015) also used publically available tweets as their dataset.

The second most commonly used method of data collection was through surveys and questionnaires. These studies were quantitative in nature. For example, Piechota (2014) conducted a survey of 200 students selected through random sampling. Students from Germany and Poland, having various levels of multiculturalism in their local community were selected, to investigate the role of new media in overcoming the prejudice of students. Likewise, Alam, Raina and Siddiqui (2016), studied the perspective of people on free speech in social media, through questionnaires filled out by 200 social media users selected randomly. Näsi et al. (2015) conducted an online survey. The data was collected from Facebook users who were Finnish nationals, of ages between 15 and 18 years. The study was aimed at finding how material on hate available online affects respondent's trust towards people around them.

With regards to mixed methods of data collection, Kimotho and Nyaga (2016) and Alakali, Faga and Mbursa (2017), used mixed methods in their studies. The former study investigated how ethnic hate speech is propagated among Kenyans through citizen journalism. It used data from questionnaires filled out by students at universities in Kenya, along with the content available on eight social networking sites between the months of January and April 2013. The latter study used questionnaires and focus group discussions as their data collection technique. The study was aimed at seeking answers as to why hate speech plagues social media in Nigeria, along with the consequences of such practices.

Overall, the content available on social networking sites in the form of user comments, blog posts and tweets have been of high importance to the researchers. Researchers have been interested in studying user generated content and responses on the content. The context of all these qualitative studies were different and the results could not be generalized. Focus groups have rarely been used, although Rubin and Babbie (2010) suggests that they provide in-depth understanding of issues under research, as they help discover unanticipated factors. Ethnography as a method has not been used much. Only one study by Yamaguchi

(2015) used ethnographic data. This study is based on the fieldwork of the author with Action Conservative Movement (ACM) groups in Japan. This study seeks answers related to the use of communication modes online and social media in connection with those groups. Brewer (2000) considers ethnography a method that captures meaning to naturally occurring activities in the field.

4.2 Data analysis techniques

Various types of data analysis techniques were used in the short-listed studies. Since not all the included articles were using empirical data, not all of them have a specific data analysis technique. All the articles with quantitative data used statistical methods while the studies that used qualitative methodology used various analysis methods. The data analysis methods used in the selected publications are summarized in Table 3.

Data analysis methods	**Publications**
Qualitative data analysis	
Qualitative analysis of positioning online	Badarneh & Migdadi (2018)
Content analysis	Sayimer & Derman (2017); Meza (2016)
Network analysis	Ben-David & Matamoros-Fernández (2016)
Multimodal content analysis	Ben-David & Matamoros-Fernández (2016)
Critical discourse analysis	Özarslan (2014); Aguilera-Carnerero & Azeez (2016); Horbyk (2018)
Co-occurrence analysis	Meza (2016)
Qualitative content analysis	Uysal, Schroeder & Taylor (2012); Maweu (2013)
Positioning analysis	Al-Tahmazi (2015)

Corpus Linguistics	Aguilera-Carnerero & Azeez (2016)
Quantitative data analysis	
Multinomial logistic regressions	Harell (2010)
10-fold cross validation approach	Burnap & Williams (2015)
A meta study of eight studies	White II & Crandall (2017)
Graphical and descriptive analysis	Piechota (2014), Alakali, Faga & Mbursa (2017)
Descriptive interpretive design (Data analyzed through analyzed using IBM SPSS version 21)	Kimotho & Nyaga (2016)
Kolmogorov–Smirnov (K-S) Z test	Alam, Raina & Siddiqui (2016)
Developed an algorithm-based approach (RNN) for detecting hate speech online	Pitsilis, Ramampiaro & Langseth (2018)
Univariate analysis of variance (ANOVA)	Näsi, et al. (2015)

Table 3 Methods of Data Analysis used in the selected publications.

The qualitative data analysis is descriptive in nature. In the articles in which qualitative data is used, the authors sought answers to the research questions by providing an in-depth descriptive analysis. In most of the studies, the authors made attempts to explore the prevailing phenomenon of hate speech online. In addition, some have tried to create a conceptual framework by identifying themes and patterns in the content available online. Discourses online have been of particular interest to some of the authors, while some are focused on other types of content available online besides conversations.

For example, Meza (2016), uses content analysis in order to analyze the dataset, comprising of comments on Facebook pages, blogs and online news media. The aim of this study is to identify hate speech directed to the public figures who belong to minority ethnic groups and the representatives of the nationalist political groups. The application of content analysis is considered valuable in order to explore the sensitive topics like prejudice and discrimination in communication content (Das & Bhaskaran, 2008). Content analysis is defined as a method that uses a set of procedures to help make inferences from the text about the sender of the message, the messages and the audience of the message (Weber, 1985).

Content analysis is also used by Sayimer and Derman (2017) in order to show how hate speech about Syrian refugees is dispersed in Poland and Turkey in online debates. Their data comprise of comments in Polish and Turkish language on YouTube videos. These videos are about refugees and have more than 10,000 views. Although the studies Meza (2016) and Sayimer and Derman (2107) invariably differ in their purpose and focus, they reflect how the application of content analysis is possible on sensitive topics catering to hate speech online. Prasad (2008) suggests that content analysis is a context sensitive method that helps in processing symbolic meanings from data.

As per the findings Critical Discourse Analysis (CDA) is used by Horbyk (2018), Özarslan (2014) and Aguilera-Carnerero and Azeez (2016). Horbyk (2018), is applying critical discourse analysis using discourse-historical approach, committed to CDA. The prime focus of his study is that how ethnolinguistic identities are formed in social media. He has taken a particular focus of interactions of social media users on the eve of the Euromaiden protests in Ukraine. The application of the discourse-historical approach is done with the use of thematic analysis along with an in-depth analysis of the strategies used in arguments called topo. He also used the material surrounding the text in the form of foreground and background. In discourse-historical approach one can integrate texts of various genres about the subject being investigated, along with the historical dimension (Wodak, 1999). The main distinguishing feature of discourse-historical approach is that it has the ability to work with several approaches and methods along with diverse background information and a wide variety of empirical data (Wodak, 2001). While Horbyk (2018) has used CDA by applying discourse-historical approach, Özarslan (2014), is using CDA method to analyze the case of hate speech in Twitter. The hate speech is targeted towards the Kurds, who are located in Van (a city in the east of Turkey). The hate speech is spread on Twitter after an earthquake in Van on Oct. 23rd, 2011. Critical discourse analysis according to Fairclough (2001):

> "Analyses texts and interactions, but it does not start from texts and interactions. It starts rather from social issues and problems, problems which face people in their social lives." (p.26).

CDA is concerned with investigating how structural relationships of power and dominance are created and manifested in language use Wodak (2001). Hence, CDA critically investigates how social inequalities are expressed and legitimized in discourse and language use. In the study by Özarslan (2014), the tweets are sent by common people and not the racist groups. Those people link the natural disaster like earthquake with the battle in the east of Turkey with Kurds. Since CDA has effectively helped in analyzing hate speech spread through mainstream media, the author is convinced that CDA can equally be useful in analyzing hate speech in social media. In his analysis he is suggesting that hate speech as a term needs revision and considers Web 2.0 as the new era of hate speech. Also, that "hate speech acts" and "hate discourse" could be added to the concept of hate speech, as according to Özarslan (2014), hate speech is not only speech, but an act with huge repercussions.

Another study that used Critical discourse analysis and Corpus linguistics methodology is by Aguilera-Carnerero and Azeez (2016). The study investigates how an average netizen articulates Cyber Islamophobia discursively. The dataset in this study is comprised of

10,025 tweets compiled around hashtag #jihad, posted from 1st till 30th of April 2013 in English language. The aim of the study is to identify the virtual communities that are built surrounding some religious and socio-political values. It also uncovers the correlation among them and see how Muslims and Islam is evaluated by social media users. Considering the way data from the tweets has been analyzed, it is more closely resembles to content analysis.

Ben-David and Matamoros-Fernández (2016) in their study use the methods of network analysis and longitudinal multimodal content analysis of text, images and links (Kress &Van Leeuwen, 2001). This study combines the rise in the popularity of social media and popularity of political extremism in order to investigate how explicit hate speech and hidden discriminatory practices circulate on social media especially Facebook, where Facebook has a strict policy on hate speech. By using these two analysis methods, the authors try to evaluate the ways in which discriminatory and hate speech is circulated on the Facebook pages of political parties in Spain. In order to identify the patterns and compare the co-occuring terms and most frequently used words related to overt hate speech in Facebook pages, Ben-David and Matamoros-Fernández (2016) performed textual analysis. Parallel to this, they analyzed 272 images and 306 links manually. These were the links with highest engagements for the political parties. They also use network analysis to study the relationship between political parties and the Facebook pages they liked.

Badarneh and Migdadi (2018) in their study focus on providing an in-depth and theoretically analysis of the comments and the responses to those comments by the Jordanian readers on the news related to politics and the economy. In the study the readers perform the act of positioning the other by commenting and responding to the comments. To do so Jordanian readers employee three discursive strategies: face attack and impoliteness, invoking of national identity and invoking of religious identity. Uysal, Schroeder and Taylor (2012) and

Maweu (2013) used qualitative content analysis method to analyze the findings of their studies. Both the studies focus on analyzing content on social networking sites, while the study by Uysal et al. (2012) focuses on Turkey's use of Twitter to spread the image of the country as being a soft power. The study by Maweu (2013), evaluates the use of Facebook and Twitter by the audience, which involves inciting and vulgar content. Another analysis method, positioning analysis, is applied by Al-Tahmazi (2015). The purpose of the study was to analyze how political discussion polarizes subsequently constructing socio-political communities.

The quantitative data analysis involves statistical analysis. In the shortlisted data, almost all the researchers employed surveys (online/offline) and questionnaires as strategies of inquiry. Statistical analysis of the data enables the researchers to accept or reject the hypotheses about the topic in question. In only one article by Harell (2010), multinomial regression analysis is used. The aim of the study was to evaluate how influence in diversity in ethnic and racial networks can have an impact on the attitude of Canadian youth regarding their speech rights. Though it is also a statistical analysis, this classification method helps in generalizing logistic regression to multiple problems, with a possibility of more than two outcomes (Greene, 2012).

Burnap and Williams (2015), for example, used a 10-fold cross validation approach in their study. The study aimed to develop a machine learning classifier for hateful content in Twitter. By using this approach for classification, the researchers were able to achieve high levels of performance. Alam, Raina & Siddiqui (2016), applied Kolmogorov–Smirnov (K-S) Z test to their data. They examined the stance of individuals on conveying free speech through Facebook and found that the posts and messages with hate are on a rise: They also found that the number of users is also increasing. Näsi et al., 2015, used Univariate analysis of variance (ANOVA) in order to make a comparison between the trust level among social groups, where they have exposure to hate material online. Piechota (2014) and Alakali, Faga and Mbursa (2017) applied graphical and descriptive analysis methods in their studies.

4.3 Focal points of research

The analysis revealed that the studies covered seven different areas. They are presented in Table 4.

Areas of research	**Publications**
Political hate speech	Badarneh & Migdadi (2018); Sayımer & Derman (2017); Ben-David & Matamoros-Fernández (2016); Ott (2017); Uysal, Schroeder & Taylor (2012); Al-Tahmazi (2015); Schaffar (2016); Maweu (2013); Suntai & Targema (2017)
Ethnic/Racial hate speech	Özarslan (2014); Meza (2016); White II & Crandall (2017); Yamaguchi (2015); Harell (2010); Langford & Speight (2015); Piechota (2014); Kimotho & Nyaga (2016); Alakali, Faga & Mbursa (2017); Jakubowicz (2107); Alam, Raina & Siddiqui (2016); Burnap & Williams (2015); Näsi, et al. (2015)
Religious hate speech	Aguilera-Carnerero & Azeez (2016)
Gendered hate speech	Mantilla (2013); Shepherd et al. (2015); Chetty & Alathur (2018)
Hate speech in online social interactions	Antoci et al. (2016); Abraham (2014); Pitsilis, Ramampiaro & Langseth (2018)
Legal frame for international bodies	Chetty & Alathur (2018)
Language and Linguistics	Horbyk (2018) Reader's comments concerning language issues in Ukraine's news website.

Table 4 Areas of research of the selected publications

The two most common areas of research on hate speech online, according to the analysis, are political and ethnic/racial issues. The studies use different platforms, varying from online news sites, Facebook groups and Twitter. The studies are done while taking into consideration diverse elements of online communication, i.e., by analyzing comments, tweets, and content shared on these sites.

Almost half of the research publications focus on hate speech related to ethnicity and race. In these studies, either hate speech is targeted towards a specific group of people belonging to a certain race, or the points of view of people from a specific ethnic background are considered regarding hate speech. For example, Özarslan's (2014) article is a case study targeting Kurds as an ethnic group, while Meza (2016) examines the occurrence of hate speech in online posts and comments, targeted towards the Roma group in Romanian language. Studies by Harell (2010) and Näsi, et al. (2015) focus on the influence of ethnic networks online on social media users.

4.4 Theories and frameworks used in the studies

Theories and frameworks used	**Studies**
Positioning theory	Badarneh & Migdadi (2018)
The works of Jeremy Waldron (2012), Susan Benesch (2012a, 2012b) and Antoine Buyse (2014), used as a theoretical base.	Sayımer & Derman (2017)
Actor-network theory	Ben-David & Matamoros-Fernández (2016)
Speech act theory	Özarslan (2014); Kimotho & Nyaga (2016)
Essayistic approach	Ott (2017)
Theoretical framework of computer mediated communication, which has two types: Synchronous and asynchronous.	Meza (2016)
Systemic Functional Linguistics	Aguilera-Carnerero & Azeez (2016)

Social Actor Theory	Aguilera-Carnerero & Azeez (2016)
Justification-suppression model of the experience and expression of prejudice.	White II & Crandall (2017)
Model of social network effects	Harell (2010)
Mean field evolutionary framework.	Antoci et al. (2016)
Critical theory	Langford & Speight (2015)
Political discourse	Al-Tahmazi (2015)
Ethnolinguistic identity theory	Horbyk (2018)
Descriptive interpretive design	Kimotho & Nyaga (2016)
Mediamorphosis theory and public sphere theory	Alakali, Faga & Mbursa (2017)
Social Responsibility Theory	Suntai & Targema (2017)

Table 5 Theoretical backgrounds and Frameworks used

Not all of the 30 studies use a specific theoretical background as their foundation. A total of 17 studies use theories or models to support their central ideas. Because these studies examined various elements of hate speech in social media, the theories and frameworks also vary.

Speech act theory is the only theory used in two different shortlisted studies, by Özarslan (2014) and Kimotho and Nyaga (2016). Both of these studies base their arguments on the notion that language is referential, informative and performative. According to Austin (1975), language is not just a medium of verbal expression and is not just used to say things, rather it is an action and things are done with words. Özarslan (2014) in his study considers illocutionary force of hate speech in social networking sites as a means of transformation. The article explores hate speech that is communicated via Twitter after the earthquake in Van (Turkey) mostly populated by Kurdish people, on Oct 23, 2011. The article presents the idea that people sending tweets have dramatically wounded the victims of the earthquake. The research also argues that hate messages are destructive and racist against Kurds; it also justifies the use of speech act theory. Similarly, a study by Kimotho and Nyaga (2016) also focuses on the illocutionary force of hate speech in social media. More specifically, the research explores the different types of illocutionary acts, and the illocutionary force held by these acts that are present in the ethnic hate speech in social media in Kenya. Since the speech act theory emphasizes that speech generally has some specific meaning to the listener, the study also presented the notion that the disseminators of digitized hate speech in Kenya that intends to spur hatred and violence.

Positioning theory is used by Badarneh and Migdadi (2018) in their research that explores how the self and the other are positioned in comments and their responses on Jordanian news sites. Other theories taken into consideration in the selected data include ethnolinguistic identity theory, which is used by Horbyk (2018), and social responsibility theory, used by Suntai and Targema (2017), etc. Table 5 is listed above for reference.

Besides theories, certain frameworks and models are also used by the authors. Since the focus of the studies is diverse, so is the use of frameworks and models. For example, Harell (2010) uses the model of social network effects to examine the attitudes of young people in Canada influenced by diverse ethnic and racial networks. In order to analyze the civil and uncivil ways of interaction online and explore the effects on collective behavior, Antoci et al. (2016) defines an evolutionary framework. However, Ott (2017) uses an essayistic approach to his study, in which he examines Twitter from a particular focus of media ecology.

In this chapter, the data is analyzed, and all the findings are reported. The data collection and analysis methods, focal points of the research and theoretical frameworks used by the shortlisted studies are reviewed in detail. Further, in the next section, the results analyzed will be discussed along with future recommendations in the particular area of study.

5 DISCUSSION

The aim of this study is to uncover the theoretical frameworks and methods used for data collection and data analysis by previous studies regarding hate speech in social media. One of the important questions that is kept into consideration is about the use of positioning theory in social media research. By restricting the research to the use of positioning theory has affected the results. We have used hate speech, social media and positioning theory as keywords, thus, limited amount of data is collected. After gathering the background information related to these keywords, the following research question are formed:

RQ1: What approaches, viewpoints and methodologies are used to study hate speech in social media?

RQ2: How is positioning theory used in the context of hate speech in social media?

While examining the data, an important finding is that out of 30 studies, only 7 studies are from the years from 2010 till 2014, while the remaining 24 studies are from the year 2015 till 2018. This drew an interesting comparison between the pre and post-2015 literature.

5.1 Data collection and data analysis methods

As per the findings, most of the studies analyzed in this systematic literature review use qualitative data, which includes user comments in social networking sites, tweets, blog posts, etc. Few studies also use reader's opinions on online news websites. What is interesting here is that the studies under scrutiny have their focus on user comments and opinions, which makes for a very interesting narrative overall. Ethnography and reviews are also used in collecting the data for analysis. In the initial search there are a few articles that did not have a clear-cut methodology for data collection and analysis. Those studies are think pieces and essays. For that reason, they are not included in this systematic literature review. As the purpose here is to look for the answers related to methods of data collection and data analysis. Most of the studies are qualitative in nature, and this reinforces the fact that the data analyzed cannot be generalized directly. Unlike qualitative methods, the use of quantitative data is very limited in the corpus.

One of the key findings from the data is that within the qualitative data analysis methods, the most commonly used method is critical discourse analysis. Three studies by Özarslan (2016), Aguilera-Carnerero and Azeez (2016) and Horbyk (2018) have used critical discourse analysis as their analysis method. None of the studies have a common focus; as, one study centered around ethnic hate speech, the other one was related to religion, while the third explored the linguistic situation in Ukraine. Of these three articles, only two take into consideration the tweets on Twitter, while the last one used user comments on news websites as their sample. There are two studies that use content analysis: Sayimer and Derman (2017) and Meza (2016). Both these studies used user comments during a certain time period as their sample to identify hate speech using content analysis. However, one focuses on comments about the political issue of Syrian refugees in Turkey and Poland and the other focuses on hate speech in the Romanian language in social media. Other qualitative data analysis methods are discourse analysis, network analysis, multimodal content analysis, co-occurrence analysis, qualitative content analysis, positioning analysis etc.

Considering the research studies that use quantitative data analysis methods, one study uses multimodal logic regression: Harell (2010). The aim of this study is to examine how ethnic and racial network diversity influence the attitude of young individuals in Canada. Another quantitative study, White and Crandall (2017), uses an experimental setup in which a total of seven experiments are conducted to examine anti-black prejudice. The results indicate that people use free speech as a justification for prejudice. Almost all other use statistical analysis methods using SPSS. The studies are mostly dealing with either identifying hate speech or people's attitudes towards hate/free speech.

In qualitative and quantitative data collection and analysis methods, where qualitative data gives us a very focused picture of an issue, handling the data quantitatively helps us understand a broader view of the problem under research. Since the use of quantitative data methodologies has been limited, the attention needs to be drawn towards this. As much as we need to understand qualitative methods, the importance of quantitative methods cannot be denied. The quantitative research in the corpus dealing with hate speech in social media is mostly focusing on detecting hate speech. There is a need to define the type and ways of hate speech is disseminated on various social media platforms. The research studies in the corpus also highlight the fact that social media platforms encounter the problem of identifying hate speech (Moulson, 2016). The awareness of the sentiment of hate speech is limited and needs attention (Ma, 2015).

5.2 Important areas of research

One of the important findings in this research is related to the focal points of earlier research. Initially, when planning this systematic literature review, the intended focus was only on religious hate speech and at that time, a very limited amount of data was retrieved from the shortlisted databases. Since there needs to be a considerable amount of data to be studied, the search criteria was enhanced to include more sensitive subjects other than just religious hate speech. At the initial screening and review, the studies covering hate speech in politics, race, religion, gender, etc., all are included for the final study. During the in-depth analysis, it is found that research on hate speech is done in seven different major areas. Ethnic and racial hate speech is the most concentrated unit. Thirteen out of thirty-one studies focuse on this area particularly. Similarly, the area of political hate speech also has a significant number of studies. Eight studies examine political hate speech from various angles. Three studies focus on hate in online social interactions. Furthermore, three studies have their focus on gendered hate speech and only one study has its focus on religious hate speech.

In this systematic literature review, the focal points of the studies could be categorized into certain areas such as race and ethnicity, gender, politics, religion, etc., but the scope of individual studies still varies. For example, if thirteen studies are studying hate speech from an ethnic and racial perspective, few particularly studied hate speech against some ethnic group, e.g., hate speech about Kurds in Turkey (Özarslan, 2014), hate speech in the Romanian language in online media regarding the Roma group (Meza, 2016), etc. One study is an essay about the social media campaign #BlackLivesMatter (Langford & Speight, 2015), whereas a couple of studies investigate the nature of digitized hate speech and foul language in Kenya and Nigeria (Kimotho & Nyaga, 2016; Alaklai, Faga & Mbursa, 2017).

All the other groups have the same dynamics; the studies that are included in this systematic literature review do fall in one focal area, but their execution is in several different dimensions. Another example of this is political hate speech, where out of eight studies, two focus on how hate speech and dangerous speech is disseminated on social media about Syrian refugees and on Facebook pages of right-wing political parties in Spain (Sayimer & Derman, 2017; Ben-David & Matamoros-Fernández, 2016). Similarly, one case study is an essay that reflects on the Twitter practices of President Donald Trump (Ott, 2017), whereas another article explores Turkey's use of Twitter for public relations strategy (Uysal, Schroeder & Taylor, 2012). In the same category of political hate speech, one study explores how new media plays its role in the entrenchment of democracy in Nigeria (Suntai & Targema, 2017).

Out of the three studies on gendered hate speech, one is an essay that covers various events in the past to identify the features of gender trolling (Mantilla, 2013). The other article offers a dialogue among digital culture scholars considering #Gamersgame campaign aims at women in video games (Shepherd et al. 2015), while the third is a review on hate speech on social networking sites, where a part of the review also looks at gender-based hate speech (Chetty & Alathur, 2018).

Hence, the study of hate speech in social media is broad in scope. Even if we divide it into certain groups and focus areas, the studies within each area would stand alone. And the importance of each study cannot be denied, as they make the research substantial and worthwhile due to its diversity and immensity. Furthermore, these studies contribute to a clearer understanding of hate speech in social media related to politics, ethnicity, race, religion and gender.

5.3 Theoretical frameworks used

Since the aim of the study is also to contemplate the theoretical frameworks used in the shortlisted studies, the findings in that section hold prodigious importance. The analysis further reveals that almost half of the studies did not use any particular approach and theoretical framework. Digging into the details of these studies, it is revealed that six of them are case studies that took into consideration a particular case or event. For example, the study by Ott (2017) was an essay, where the author examines the platform of Twitter from a media ecology perspective. He based his arguments on one of Donald Trump's tweets on 10th Nov 2012. Similarly, Yamaguchi (2015) also based his research on a particular case of online video sharing in June 2010 by ACM, an activist group in Japan. Besides case studies, a review article by Chetty and Alathur (2018) also did not use any theoretical framework. Few qualitative and quantitative studies do not fall under the category of essays and review articles or use any particular theoretical framework. These articles are based on the gaps identified by the authors, which become the research questions whose answers are sought through the best suitable method. There are many studies that are done without any theory and the content is analyzed considering the research questions based on a certain observation. This does not make the study less valuable. It simply is a different way of conducting research.

Interestingly, there are hardly any studies that have used the same theory to back their arguments. Only the studies by Özarslan (2014) and Kimotho and Nyaga (2016) used speech act theory, where the former deliberates that the illocutionary force of hate speech in social media platforms works as a means of transformation, whereas the latter investigates the various types of illocutionary acts and the force held by those acts. Other theories include: ethnolinguistic theory, social responsibility theory, critical theory, social actor theory, etc.

5.4 Positioning theory and hate speech in social media

The analysis revealed that among the reviewed studies, only one, Badarneh and Migdadi (2018), has used positioning theory specifically as its theoretical base. Apparently, positioning is a characteristic of most of the conversations taking place in general and in social media. This does not mean that the other studies would not have considered positioning at all or related to it in some way. Indeed, positioning may be viewed as a characteristic of most of the conversations taking place in general and in social media. Davies and Harré (1990) originally presented the metaphors of positions and positioning in interactions. People involved in interactions understand positions as per their own experiences that include beliefs, histories, norms and emotions, etc. This creates the environment of echo chambers, ingroup and outgroup, or us versus them. According to Badarneh & Migdadi (2018):

> "If we are to come close to understanding how it is that people actually interact in everyday life we need the metaphor of an unfolding narrative, in which we are constituted in one position or another within the course of one story, or even come to stand in multiple or contradictory positions, or to negotiate a new position by "refusing" the position that the opening rounds of a conversation have made available to us." (p. 53)

Badarneh and Migdadi (2018) further argue that the accomplishment of self and the other is attained through the application of three strategies online that include impoliteness, invoking of national identities and invoking of religious identities. They illustrate how social media gives room for anti-social behavior that includes online harassment and trolling, and expressing hate, symbolically and verbally. The comments and rebuttal analyzed in the study show the existence of a certain stance where the ‘other’ holds a different stance. Overall, the article works as an example of how positioning theory may be used in context of social media.

While not referring to positioning theory as such, Al-Tahmazi’s (2015) study utilized what they label as positioning analysis. The study aimes at finding out how political discussions are polarized by pursuit of power on Facebook, which ends in creating sociopolitical communities online. The author in this article argues that the gap between the macro analytical discourse approaches and micro analytical approaches can be filled by a multi-tiered positioning theory. Here they refer to Michael Bamberg (1997), who built his idea on the concept of positioning by Davies and Harré’s (1990), where they define positioning as a “discursive process whereby selves are located in conversations as observably and subjectively coherent participants in jointly produced story lines” (p. 48). Bamberg, in turn views positioning as a discursive process that takes place at three levels. Al-Tahmazi (2015) applied the three-tiered process of positioning described in Bamberg’s (2004) study to analyse his data at three different levels. His analysis reveals that at the first level of positioning, de/legitimization takes place, while at the second level of positioning, alignments are established and political fronts are shaped, while at the third level of positioning, socio-political communities are formed. The analysis gives a concrete example of how commentators on Facebook categorize themselves and others into opposite communities, consciously or unconsciously. This is in line with the original viewpoint of the positioning theory.

Positioning theory is a relatively new theory and it still needs to be used within the educational research. Rather it has a concrete purpose, where the individuals are positioned, position others, define audiences and the attitude they have before them (Tirado & Gálvez, 2008). This opens endeavors of research on social media to a great extent. The literature shows that positioning theory has mostly been used in the area of linguistics and linguistic signs. Positioning theory has various branches and they cover certain areas of research, but it still needs to be explored in the area of multiple modalities like videos, images, films, etc. In addition, language researchers should use this theory in their work that is transdisciplinary, which will help in strengthening the actual use of the theory of positioning, and not just position or positioning as a metaphor.

6 CONCLUSION

Considering the core aim of this study, important data are gathered in the area of hate speech in social media. And this information is extracted to help in exploring various other areas for future research. Hate speech has always been an important area of research and studying it purely from the context of online media and social networking sites opens more opportunities in this field.

6.1 Limitations and validity threats

As a method, a systematic literature review has several limitations that need to be considered while reporting the findings. This particular study has the following limitations:

- The study is limited to research articles only.
- Only peer reviewed articles are included.
- Book chapters are not made a part of the study.
- Only articles that are accessible in the databases are included.
- Only full text articles are included in the review
- Only articles with a clear methodology are included.

Validity threats are the factors that influence the accuracy of the research in a negative manner. It is crucial to recognize these threats to make the results of the review reliable. This research study has various validity threats that are classified into three main categories, which are: researcher bias, biasness related to primary research studies, and the threats related to the data extraction process and the results.

As this research has been conducted by an individual, there is certainly an increased threat to validity when compared to reviews conducted by a group of researchers. To minimize the risk of validity biasness, certain tasks are carried out twice. The abstracts are read twice to ensure that none of the relevant research studies are left out.

In order to minimize the threat of biasness related to primary research studies, the researcher uses all the possible studies that could have been included in this research. The titles and abstracts are read numerous times to make sure that the right studies are included for this systematic literature review. Validity in the data extraction phase is crucial to a systematic literature review. For this particular research study, the data extraction process is well defined before conducting the research, which ensures that all the necessary information is recorded. This curtails the data extraction process bias.

6.2 Recommendations and future directions

Although this systematic review has its limitations, this research still offers a useful synthesis of previously used theoretical frameworks and adopted approaches for data collection and analysis. There is a considerable amount of research on the topic of hate speech in social media; the topic is complicated and needs to be studied and understood deeply. It is recommended that research in some more areas is performed in order to help researchers appreciate the vastness of the topic.

In the very beginning, when the keywords for the research are decided, the aim is to examine hate speech, commonly understood as hate in spoken and written expressions. It is observed during the initial search that many other terminologies are also used to address the concept including hate, free speech, incivility, inappropriate language, etc. Future research in this area can employ more keywords to obtain an even richer and comprehensive outcome. Since the systematic research is limited to three databases, some articles have not become a part of this particular research. Using more databases like Elsevier, etc., could uncover some relevant data, and made part of the systematic review. This research is restricted to the English language, so conducting a literature review in the same area but with other languages included could add some important research to the review.

While examining the studies shortlisted for this systematic literature review, a few fascinating areas are also found that could be researched further. Kimotho and Nyaga (2016), in their study about ethnic hate speech suggested examining certain other areas of hate speech such as “racist hate speech, religious hate speech and gender”. Further, they also suggested studying the effects of online hate speech on the targeted groups. Some more possible topics for future research in online discourse related to language, the working of language and events shaping the language landscapes in social media are suggested by Badarneh and Migdadi (2018) and Horbyk (2018). Schaffar (2016) presented a case study related to the emergence of Fascist Vigilante Groups on Facebook in social media in Thailand. For future research, he suggests digging into the idea of linkage between online media, political polarization and Fascist vigilantism.

Similarly, in the research study by Piechota (2014), the aim is to investigate social media’s role in overcoming prejudice through intercultural dialogue. Being quantitative in nature, the study showed the difference in the attitudes of students who are surveyed. Considering the nature of the study, it would be interesting if the idea of positioning is applied in the same area. Comparing how students position themselves and other students while engaging in intercultural dialogue on social media involving hate speech could generate interesting findings.

Näsi et al. (2015), in their study inspect the correlation of Finnish youth’s exposure to hate material online with their trust towards people in their close family circle, friends circle, colleagues etc. This study is limited to a very specific age group and sample size; however, it dealt with online hate material. If this particular study uses positioning theory, it could help bring a clear understanding of the positioning of family, friends, colleagues, neighbors, etc. Finally, expanding on the sample could bring forward some interesting findings.

REFERENCES

Abraham, B. 2014. Challenging hate speech with Facebook Flarf: The role of user practices in regulating hate speech on Facebook. *The Fibreculture Journal,* 23, 47–72.

Aguilera-Carnerero, C., & Azeez, A. H. (2016). 'Islamonausea, not Islamophobia': The many faces of cyber hate speech. *Journal of Arab & Muslim media research*, *9*(1), 21-40.

Alam, I., Raina, R. L., & Siddiqui, F. (2016). Free vs hate speech on social media: the Indian perspective. *Journal of Information, Communication and Ethics in Society, 14*(4), 350-363. doi: 10.1108/JICES-06-2015-0016

Alakali, T. T., Faga, H. P., & Mbursa, J. (2017). Audience perception of hate speech and foul language in the social media in Nigeria: Implications for morality and law. *Academicus - International Scientific Journal, 15*, 166-183. doi: 10.7336/academicus.2017.15.11

Allen, C. (2012). A review of the evidence relating to the representation of Muslims and Islam in the British media. *Birmingham, AL: Institute of Applied Social Studies.* Retrieved from https://www.birmingham.ac.uk/Documents/college-social-sciences/social-policy/IASS/news-events/MEDIA-ChrisAllen-APPGEvidence-Oct2012.pdf

Alkiviadou, N. (2019). Hate speech on social media networks: towards a regulatory framework? *Information & Communications Technology Law, 28*(1), 19-35. doi: 10.1080/13600834.2018.1494417

Al-Tahmazi, T. H. (2015). The pursuit of power in Iraqi political discourse: Unpacking the construction of sociopolitical communities on Facebook. *Journal of Multicultural Discourses*, *10*(2), 163-179. doi: 10.1080/17447143.2015.1042383

Altonen, J. P. (2017, May). Positions of opposition. Using the subject of immigration as a tool for identity construction for self and other (Unpublished master’s thesis). Retrieved from https://helda.helsinki.fi/bitstream/handle/10138/193687/Altonen_Jussi_Pro-Gradu_2017.pdf?sequence=2&isAllowed=y

Antoci, A., Delfino, A., Paglieri, F., Panebianco, F., & Sabatini, F. (2016). Civility vs. incivility in online social interactions: An evolutionary approach. *PLoS One*, *11*(11). doi: 10.1371/journal.pone.0164286

Austin, J. L. (1975). *How to do things with words*. Harvard University Press.

Awan, I., & Zempi, I. (2015). *We fear for our lives: Offline and online experiences of anti-Muslim hostility.* Retrieved from https://www.tellmamauk.org/wp-content/uploads/resources/We%20Fear%20For%20Our%20Lives.pdf

Badarneh, M. A., & Migdadi, F. (2018). Acts of positioning in online reader comments on Jordanian news websites. *Language & Communication*, *58*, 93-106. https://doi.org/10.1016/j.langcom.2017.08.003

Bamberg, M. 1997. Positioning between structure and performance. *Journal of Narrative and Life History, 7*(1-4)*,* 335-342.

Bamberg, M. (2004). Positioning with Davie Hogan. In C. Daiute & C. Lightfoot (Eds.), *Narrative analysis: Studying the development of individuals in society*. (pp. 135-158). Thousand Oaks, CA: Sage Publications.

Banks, J. (2010). Regulating Hate Speech Online. *International Review of Law, Computers & Technology*, *24*(3), 233–239.

Banks, J. (2011). European Regulation of Cross-Border Hate Speech in Cyberspace: The Limits of Legislation. *European Journal of Crime, Criminal Law and Criminal Justice, 19*(1), 1–13.

Ben-David, A., & Matamoros-Fernández, 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.

Bettany-Saltikov, J. (2010). Learning how to undertake a systematic review: part 2. *Nursing Standard, 24*(51), 47-56. doi:10.7748/ns2010.08.24.51.47.c7943.

Bleich, E. (2011). What is Islamophobia and how much is there? Theorizing and measuring an emerging comparative concept. *American behavioral scientist*, *55*(12), 1581-1600.

Boeckmann, R. J., & Turpin-Petrosino, C. (2002). Understanding the Harm of

Hate Crime. *Journal of Social Issues*, *58*(2), 207-225. doi: 10.1111/1540-4560.00257

Brennan, F. (2009). Legislating against Internet race hate. *Information & Communication Technology Law, 18*(2), 123-153.

Brewer, J. D. (2000). *Ethnography.* Philadephia: Open University Press.

Budgen, D., & Brereton, P. (2006, May). Performing systematic literature reviews in software engineering. In *Proceedings of the 28th international conference on Software engineering* (pp. 1051-1052).

Burnap, P., & Williams, M. L. (2015). Cyber hate speech on twitter: An application of machine classification and statistical modeling for policy and decision making. *Policy & Internet*, *7*(2), 223-242.

Carter, C. R., & Ellram, L. M. (2003). Thirty-rve years of the Journal of Supply Chain Management: Where have we been and where are we going?. *Journal of Supply Chain Management*, *39*(1), 27-39.

Castells, M. (2014). The impact of the internet on society: a global perspective. *Change*, *19*, 127-148.

Chakraborti, N., & Garland, J. (2009). *Hate Crime: Impact, Causes and Responses*. Sage Publications.

Chan, J., Ghose, A., & Seamans, R. (2014). *The Internet and hate crime: Offline spillovers from online access.* Rochester, NY: Social Science Research Network. Retrieved from http://papers.ssrn.com/abstract=2335637

Chetty, N. & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and violent behavior*, *40*, 108-118. doi: https://doi.org/10.1016/j.avb.2018.05.003

Citron, D. K. (2014). *Hate crimes in cyberspace*. Harvard University Press.

Clement, J. (2019). *Worldwide Digital Population as of November 2019*. Retrieved from https://www.statista.com/statistics/617136/digital-population-worldwide/

Clement, J. (2019). *Global Social Networks Ranked by Number of Users 2019*. Retrieved from https://www.statista.com/statistics/272014/global-social-networks-ranked-by-number-of-users/

Coliandris, G. (2012). Hate in a cyber age. *Policing Cyber Hate, Cyber Threats and Cyber Terrorism*, 75-94.

Coliver, S. (1992). *Striking a Balance: Hate Speech, Freedom of Expression and Non-discrimination.* London: University of Essex Press

Council of Europe. (2013, August). No Hate Speech Movement: Campaign for Human Rights Online. Retrieved from https://rm.coe.int/09000016806cae52

Christopherson, K. (2007). "The Positive and Negative Implications of Anonymity in Internet, Nobody Knows You're a Dog." *Computers in Human Behavior*, *23*(6), 3038-56.

Das, D. K., & Bhaskaran, V. (2008). *Research methods for social work*. New Delhi: Rawat, 173-193.

Davies, B., & Harré, R. (1990). Positioning: The discursive production of selves. *Journal for the Theory of Social Behaviour, 20*, 46-63.

Depperman, A. (2015). Positioning. In A. De Fina & A. Georgakopolou (Eds.), *The handbook of narrative analysis.* Chichester, West Sussex: Wiley Blackwell, pp. 369-387.

Fairclough, N. (2001). Critical discourse analysis. *How to analyse talk in institutional settings: A casebook of methods*, 25-38.

Foxman, A. H., & Wolf, C. (2013). *Viral hate: Containing its spread on the Internet*. New York, NY: St. Martin's Press.

Gerstenfeld, P. B., Grant, D. R., & Chiang, C. P. (2003). Hate online: A content analysis of extremist Internet sites. *Analyses of social issues and public policy*, *3*(1), 29-44.

Gagliardone, I. (2014). *Mapping and analysing hate speech online.* Retrieved from https://www.researchgate.net/publication/314552833_Mapping_and_Analysing_Hate_Speech_Online

Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. Unesco Publishing. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000233231

Greenhalgh, T. (1997). 'Papers that Summarise Other Papers (systematic reviews and meta analyses), *British Medical Journal, 315* (7109), 672–675.

Harré, R., & Langenhove, L. V. (1991). Varieties of positioning. *Journal for the theory of social behaviour*, *21*(4), 393-407.

Harré, R., & Langenhove, L.V. (Eds.). (1999). *Positioning theory: Moral contexts of intentional action*. Oxford, UK: Blackwell.

Harré, R., & Moghaddam, F. (2003). *The self and others: Positioning individuals and groups in personal, political and cultural contexts.* Westport, CT: Praeger.

Harré, R., Moghaddam, F. M., Cairnie, T. P., Rothbart, D., & Sabat, S. R. (2009). Recent advances in positioning theory. *Theory & psychology*, *19*(1), 5-31.

Harre, R. and Secord, P. (1973), *The Explanation of Special Behaviour*, Totowa, NJ.: Adams Littlefield

Harell, A. (2010). Political tolerance, racist speech, and the influence of social networks. *Social Science Quarterly*, *91*(3), 724-740.

Hernández, T. K. (2011). Hate speech and the language of racism in Latin America: a lens for reconsidering global hate speech restrictions and legislation models. *University of Pennsylvania. Journal of International Law, 32*(3), 805-841.

Hollway, W. (1984). Gender difference and the production of subjectivity'in Henriques, J, Hollway, W, Urwin, C, Venn, C and Walkerdine, V (eds) Changing the Subject: Psychology, Social Regulation and Subjectivity. *Social Regulation and Subjectivity, London: Methuen*.

Hinduja, S., & Patchin, J. W. (2008). Cyberbullying: An Exploratory Analysis of Factors Related to Offending and Victimization. *Deviant behavior*, *29*(2), 129-156.

Hirvonen, P. (2016). Positioning theory and small-group interaction: Social and task positioning in the context of joint decision-making. *Sage Open*, *6*(3), doi: 10.1177/2158244016655584

Horbyk, R. (2018). Discourses on Languages and Identities in Readers' Comments in Ukrainian Online News Media: An Ethnolinguistic Identity Theory Perspective. *East/West: Journal of Ukranian Studies*, *5*(2). doi: https://doi.org/10.21226/ewjus417

Iganski, P. (2012). *Hate Crime: Taking Stock: Programmes for Offenders of Hate*. Belfast: Northern Ireland Association for the Care and Resettlement of Offenders.

Jaishankar, K. (2008). Cyber hate: Antisocial networking in the Internet. *International Journal of Cyber Criminology*, *2*(2), 16.

Jakubowicz, A. (2017). Alt_Right White Lite: trolling, hate speech and cyber racism on social media. *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, *9*(3), 41-60. doi: http://dx.doi.org/10.5130/ccs.v9i3.5655

Kable, K, A., Pich, J., and Sian, E., Maslin-Prothero. A. (2012). Structured approach to documenting a search strategy for publication: *A 12 step guideline for authors. Nurse Education Today, 32*(8), 878-886. doi:10.1016/j.nedt.2012.02.022.

Kettrey, H. H., & Laster, W.N. (2014). Staking territory in the “World White Web” an exploration of the roles of overt and. Color-blind racism in maintaining racial boundaries on a popular web site. *Social Currents, 1*(3), 257-274.

Kimotho, S. G., & Nyaga, R. N. (2016). Digitized ethnic hate speech: Understanding effects of digital media hate speech on citizen journalism in Kenya. *Advances in Language and Literary Studies*, *7*(3), 189-200. doi:10.7575/aiac.alls.v.7n.3p.189

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering.

KnowledgeWharton. (2018). *How Can Social Media Firms Tackle Hate Speech?* Retrieved from https://www.fairobserver.com/region/north_america/social-media-free-speech-mark-zuckerberg-latest-facebook-news-this-week-23439/

Kowalski, R. M., Limber, S. P., & Agatston, P. W. (2012). *Cyberbullying: Bullying in the digital age*. John Wiley & Sons.

Krogh, S. (2016). Anticipating organizational change – A positioning theory perspective (Doctoral dissertation). Retrieved from https://openarchive.cbs.dk/bitstream/handle/10398/9365/Simon_Krogh_ISEOR2016.pdf?sequence=1

Langford, C. L., & Speight, M. (2015). # BlackLivesMatter: Epistemic Positioning, Challenges, and Possibilities. *Journal of Contemporary Rhetoric*, *5*(3/4), 78-89.

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., ... & Wolff, S. (2009). A brief history of the Internet. *ACM SIGCOMM Computer Communication Review*, *39*(5), 22-31.

Ma, A. (2015). *Global Survey Finds Nordic Countries Have the Most Feminists.* Retrieved from https://www.huffpost.com/entry/global-gender-equality-studyyougov_n_564604cce4b045bf3deeb96d?guccounter=1&guce_referrer=aHR0cHM6Ly9kdWNrZHVja2dvLmNvbS8&guce_referrer_sig=AQAAAAUbp_7alP1-Ws_p4KPuSRGCP8Ed2dagiSXiW2qenBLY52Drl-bhhMg3ae5B65KHKU1BTTIM5ttFDSHHGEFyUO-qgWn51q7pMEo2tUjks42f9l_AUKGrS98w6ZGHnLh7nK5vjCqqirvk-2wfRU_BCY6PCMlGNIRbGh_B1LDTHHasE0

Mantilla, K. (2013). Gendertrolling: Misogyny adapts to new media. *Feminist Studies*, *39*(2), 563-570.

Maweu, J. M. (2013). The Ethnic Hate Speech was Networked: What Social Media Political Discussions Reveal about the 2013 General Elections in Kenya'. *Index. Comunicación Journal*, *3*(2), 37-52

McVee, M. B., Silvestri, K. N., Barrett, N., & Haq, K. S. (2018). Positioning theory. *In Theoretical models and processes of literacy*, 397-416. Routledge.

Meza, R. (2016). Hate-Speech in the Romanian online media. *Journal of Media Research-Revista de Studii Media*, *9*(26), 55-77.

Moghaddam, F., & Harré, R. (2010). *Words, conflicts and political processes.* Santa

Barbara, CA: Praeger.

Moher, D, Liberati, A., Jennifer, Tetzla. and Altman, G.D. (2009). Preferred reporting items for systematic reviews and meta-analyses: The prisma statement. *Annals of Internal Medicine, 151*(4), 264-269, 2009. doi:10.7326/0003-4819-151-4-200908180-00135.

Moulson, G. (2016). Zuckerberg in Germany: No place for hate speech on Facebook. Retrieved Nov 16, 2019, from https://www.seattletimes.com/business/zuckerberg-no-place-for-hate-speech-on-facebook/

Näsi, M., Räsänen, P., Hawdon, J., Holkeri, E., & Oksanen, A. (2015). Exposure to online hate material and social trust among Finnish youth. *Information Technology & People*, *28*(3), 607-622. doi: 10.1108/ITP-09-2014-0198

Nockleby, J. T. (2000). Hate speech. *Encyclopedia of the American constitution*, *3*(2), 1277-1279.

Oksanen, A., Hawdon, J., Holkeri, E., Näsi, M., & Räsänen, P. (2014). Exposure to online hate among young social media users. *Sociological studies of children & youth*, *18*(1), 253-273.

Ott, B. L. (2017). The age of Twitter: Donald J. Trump and the politics of debasement. *Critical studies in media communication*, *34*(1), 59-68.
doi: 10.1080/15295036.2016.1266686

Özarslan, Z. (2014). Introducing two new terms into the literature of hate speech:" Hate Discourse" and" Hate Speech Act" Application of" speech act theory" into hate speech studies in the era of Web 2.0. *Ileti-s-im*, (20), 53-75.

Parekh, B. (2006). Hate speech. *Public policy research*, *12*(4), 213-223.

Perry, B. (2001). *In the name of hate: Understanding hate crimes*. London, UK: Routledge.

Piechota, G. (2014). The Role of Social Media in Creating Intercultural Dialogue and Overcoming Prejudice–a Comparative Analysis of Pilot Survey Results. *KOME − An International Journal of Pure Communication Inquiry, 2*(2), 37-63.

Pitsilis, G. K., Ramampiaro, H., & Langseth, H. (2018). Effective hate-speech detection in Twitter data using recurrent neural networks. *Applied Intelligence*, *48*(12), 4730-4742. https://doi.org/10.1007/s10489-018-1242-y

Potok, M. (2011). Hate groups top 1,000: The year in hate and extremism. Intelligence Report, 14, 40-67.

Potter, D., and M. Wetherell. (1987). Discourse and Social Psychology. London: Sage.

Prasad, B. D. (2008). Content analysis. *Research methods for social work*, *5*, 1-20.

Rubin A. and Babbie, E. 2010. Essential Research Methods for Social Work. 2nd ed. Belmont (Calif.): Brooks/Cole: Cengage Learning.

Riffe, D., Lacy, S., Fico, F., & Watson, B. (2019). *Analyzing media messages: Using quantitative content analysis in research*. Routledge.

Sabat, S., & Harré, R. (1999). Positioning and the recovery of social identity. *Positioning theory: Moral contexts of intentional action*, 87-101.

Schaffar, W. (2016). New social media and politics in Thailand: The emergence of fascist vigilante groups on Facebook. *Austrian Journal of South-East Asian Studies*, *9*(2), 215-234. doi 10.14764/10.ASEAS-2016.2-3

Sayimer, İ., & Derman, M. R. (2017). Syrian refugees as victims of fear and danger discourse in social media: A Youtube analysis. *Global Media Journal: Turkish Edition*, *8*(15).

Shepherd, T., Harvey, A., Jordan, T., Srauy, S., & Miltner, K. (2015). Histories of hating. *Social Media+ Society*, *1*(2), 1-10. doi: 10.1177/2056305115603997

Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2016, March). *Analyzing the targets of hate in online social media.* In Tenth International AAAI Conference on Web and Social Media, North America. Retrieved from https://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13147/12829.

Simon Wiesenthal Center. (2012, March 27). *Social Media Must Do More to Thwart Subculture Of Hate Fueling Lone Wolf Terrorism - Simon Wiesenthal Center Debuts 2012*

Digital Hate Report. Retrieved from http://www.wiesenthal.com/about/news/social-media-must-do-more-to.html

Simpson, R. M. (2013). Dignity, harm, and hate speech. *Law and Philosophy, 32*(6), 701-728.

Sood, S., Antin, J., & Churchill, E. (2012, May). *Profanity Use in Online Communities.* In Proceedings of th4 SIGCHI Conference on Human Factors in Computing Systems. doi: 10.1145/2207676.2208610

Stakić, I. (2011). Homophobia and hate speech in Serbian public discourse: How nationalist myths and stereotypes influence prejudices against LGBT minority. *The Equal Rights Review*, *7*, 44–65.

Staszak, J., F. (2008). Other/Otherness. International Encyclopedia of Human Geography.

Suntai, D. I., & Targema, T. S. (2017). New media and democracy in Nigeria: an appraisal of the opportunities and threats in the terrain. *Brazilian Journal of African Studies*, *2*(4), 198-209.

Swartz, M. K. (2008). The Importance of Peer Review. *Journal of Pediatric Health Care, 22*(6), 333-334. https://doi.org/10.1016/j.pedhc.2008.08.004

Tan, S. L., & Moghaddam, F. M. (1999). Positioning in intergroup relations. *Positioning theory*, 178-194.

The TCP/IT reference model. (n.d.). Retrieved from: https://www.studytonight.com/computer-networks/tcp-ip-reference-model

Timofeeva, Y. A. (2002). Hate Speech Online: Restricted or Protected-Comparison of Regulations in the United States and Germany. *Journal of Transnational Law & Policy, 12*(2), 253-268.

Tirado, F., & Gálvez, A. (2007). Positioning Theory and Discourse Analysis: Some Tools for Social Interaction Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research, 8*(2). doi:http://dx.doi.org/10.17169/fqs-8.2.248

Tirado, F., & Gálvez, A. (2008). Positioning Theory and Discourse Analysis: Some Tools for Social Interaction Analysis. *Historical Social Research / Historische Sozialforschung, 33*(1 (123), 224-251.

Townsend, E. (2014). Hate Speech or Genocidal Discourse? An Examination of Anti-Roma Sentiment in Contemporary Europe. *PORTAL Journal of Multidisciplinary International Studies, 11*(1).

Tranfield, D., Denyer, D. and Smart, P. (2003) Towards a methodology for developing Evidence: Informed Management Knowledge by Means of Systematic Review. *British Journal of Management, 14,* 207-222. doi: http://dx.doi.org/10.1111/1467-8551.00375

Traum, A. (2014). Contextualising the hate speech debate: The United States and South Africa. *Comparative and International Law Journal of Southern Africa, 47*(1), 64-88.

University of York. (2009). Systematic Reviews. *CRD's guidance for undertaking reviews in health care.* Retrieved from https://www.york.ac.uk/media/crd/Systematic_Reviews.pdf

Uysal, N., Schroeder, J., & Taylor, M. (2012). Social media and soft power: Positioning Turkey's image on Twitter. *Middle East Journal of Culture and Communication*, *5*(3), 338-359. doi: 10.1163/18739865-00503013.

Vajić, N., & Voyatzis, P. (2012). The Internet and freedom of expres sion: a "brave new world" and the ECtHR's evolving case law'. *Freedom of Expression: Essays in Honor of Nicolas Bratza.* Oisterwijk: Wolf Legal Publishers.

Waldron, J. (2012). *The Harm in the Hate Speech.* Cambridge, Massachusetts & London, England: Harvard University Press.

Warren, Z., & Moghaddam, F. M. (2018). Positioning theory and social justice. *The Oxford handbook of social psychology and social justice*, 319-331.

Weber, R. P.(1985). *Basic content analysis*, Newbury Park, CA: Sage.

Wessel-Aas, J., Fladmoe, A., & Nadim, M. (2016). Hate speech, report 3. The boundary between freedom of speech and criminal law protection against hate speech. *Rapport Institutt for samfunnsforskning.*

White, M. H. II, & Crandall, C. S. (2017). Freedom of racist speech: Ego and expressive threats. *Journal of Personality and Social Psychology, 113*(3), 413-429. https://doi.org/10.1037/pspi0000095

Wodak, R. (1999). Critical Discourse Analysis at the End of the 20th Century. *Research on Language & Social Interaction, 32*(1-2), 185-193, doi: 10.1080/08351813.1999.9683622

Wodak, R. (2001). The discourse-historical approach. In R. Wodak & M. Meyer (Eds.), *Methods of critical discourse analysis* (pp. 63-94). London: SAGE Publications Ltd doi: 10.4135/9780857028020.n4

Wright, R. W., Brand, R. A., Dunn, W., & Spindler, K. P. (2007). How to write a systematic review. *Clinical Orthopaedics and Related Research®*, *455*, 23-29.

Wynstra, F. (2010). What did we do, who did it and did it matter? a review of fifteen volumes of the (European) journal of purchasing and supply management. *Journal of Purchasing & Supply Management, 16*(4), 279-292. doi: 10.1016/j.pursup.2010.09.003.

Yamaguchi, T. (2013). Xenophobia in Action: Ultranationalism, Hate Speech, and the Internet in Japan. Radical history review, (117), 98-118. doi: 10.1215/01636545-2210617

APPENDICES

APPENDIX 1

S. #	Authors	Aim/Focus	Theory/ Framework	Sample	Method	Findings	Future Recommendations
1	Badarneh & Migdadi (2018)	The focus of this study was to explore reader comments and responses on local online news sites in Jordan and how readers respond to, comment on, or challenge the news source, specifically regarding much debated issues pertaining to the political, economic and social landscape of the country.	Positioning theory	500 reader comments	Qualitative analysis of positioning	The readers seek to accomplish self and other positioning through three main strategies: impoliteness and face attack, invoking of national identity, and invoking of religious identity.	To investigate other aspects of Jordanian, and Arab, online reader comments on news so as to examine more workings of language and interaction in Arabic-language online discourse.
2	Sayımer & Derman (2017)	The aim of this paper is to reveal how the dangerous speech and fear speech towards Syrian refugees is disseminated from online debates in two different countries: Poland and Turkey.	The works of Jeremy Waldron (2012), Susan Benesch (2012a, 2012b) and Antoine Buyse (2014) were used as a theoretical base.	The sample covered the comments published between December the 25th, 2015 to December the 25th, 2016 (in total 18,563 comments – 6190 comments from the Polish and 12,373 comments from the Turkish videos).	Content Analysis	Hate speech was identified in 855 Polish and 1705 Turkish comments, which, in both data sets, established exactly the same proportion of hate speech – 13.8 per cent.	
3	Ben-David & Matamoros-Fernández (2016)	This study considers the ways that overt hate speech and covert discriminatory practices circulate on Facebook despite its official policy that prohibits hate speech.	Actor-network theory	Official Facebook pages of seven extreme-right political parties in Spain between 2009 and 2013.	Network analysis and multimodal content analysis	The Spanish extreme-right political parties primarily implicate discrimination, which is then taken up by their followers who use overt hate speech in the comment space.	

4	Özarslan (2014)	The aim of this paper is to explain the need for a revision of the term “hate speech” in the era of Web 2.0 and to introduce two new terms into the literature of hate speech, that is “hate discourse” and “hate speech act.”	Speech-act theory	Case study-hate speech communicated through Twitter after the earthquake in Van, a city situated in the east of Turkey and populated mostly by Kurds, on 23 October 2011	Critical discourse analysis	Revision of the term ‘hate speech’ from the perspective of ‘speech act theory’ could provide effective ways to combat against hate speech in the era of Web 2.0. Hate speech is not only ‘speech’ anymore, it is an ‘act’.	New media literacy with special emphasis on critical thinking could contribute to the development of more democratic acts, common sense, in Web 2.0 and so more works should be done to develop critical new media literacy not only by academics but also by the institutions such as media, schools, municipalities, etc.
5	Ott (2017)	This essay explores the changing character of public discourse in the Age of Twitter. The essay highlights how Twitter priviliges discourse that is simple, implusive and uncivil. Based on this claim, the author examines the platform of Twitter from the perspective of media ecology. The author further reflects upon the Twitter practices of President-Elect Donald J. Trump.	Case Study	Donald Trump's twitter feed on Nov 10, 2012; “Thanks- many are saying I’m the best 140 character writer in the world.”	Essay	The author concludes that Twitter is producing most self-involved people in history by treating everything one does or thinks as newsworthy. Television may have assaulted journalism, but Twitter killed it. If Twitter is treated as a legit source of news, it will have its consequences. Firstly, Twitter's underlying logic will continue to supplant television. Secondly, we will continue to witness the rise and mainstreaming of divisive and incendiary public discosurse. Thirdly, we are likely to witness a growing intolerance for cultural and political platforms. And fourthly, we will see more dangerous demagogues rise to prominence.	
6	Meza (2016)	This research explores new methodologes for automatically identodying and classifying online hate speech, both on popular social network sites like Facebook, and on web content management system driven dynamic webistes like blogs or online news sites. The goal of the research is to identify and classify instances of hate-speech in Romanian language comments to online media (posts and articles).	The author refers to the theoretical framework of computer mediated communication, which has two types; Synchronous and asynchronous.	All the comments published on 25 Facebook pages, 10 blogs and the news sections of 5 major online news outlets between January 1 2015 and June 30 2015.	Content analysis and Co-occurrence analysis	The most frequently referenced target group in online comments is the Roma group, mostly through the term(s) ”țigan/i”. There were also significant numbers of references to Hungarians, Jews and members of the LGBT community, some of them through use of the derogative terms ”bozgori”, ”jidani” or ”poponari”, cases in which these can be considered hate-speech by themselves. Violent or offensive language was encountered in varying degrees in comments posted on Facebook (2%), on blogs (6.3%) or on news websites (8.3%). The most	This research opens up new methodological pathways in researching online hate-speech in Romania. The analysis methods may be replicated and extended to cover more time and more contexts for online computer mediated communication. The author considers discussion groups, Web forums or Facebook groups popular among teenages such as Toti Pentru Unu (tpu.ro) or Junimea to be of particular interest. Also, Facebook pages

						frequent negative, violent or offensive terms detected were those in the semantic areas of ”stupidity” and ”debility”. A higher frequency of obscene explicit language was detected in comments posted on blogs or online news outlets. The frequency of co-occurrences of terms referencing targets of hate-speech with violent and offensive language is below 1% in the 2.6 million comments which were analyzed – 0,1% in Facebook comments, 0.14% on blogs and 0.28% on online news websites. Still, it is worth noting, that more in-depth analysis may allow precise pin-pointing of contexts in which these co-occurrences surge.	belonging to other public figures, political parties or civil society groups might be of interest to future researchers.
7	Aguilera-Carnerero & Azeez (2016)	The aim of this article is to study how Cyber Islamophobia is articulated discursively by the average netizen (as opposed to the mainstream media).	Systemic Functional Linguistics, Social Actor Theory	A corpus of more than 10,025 tweets compiled around the hashtag #jihad between April 1 and 30, 2013. Also, only the tweets in English were retrieved.	Critical Discourse Analysis, Corpus Linguistics methodology	The conception of ‘jihad’ and the stereotypes of Muslims and Muslim culture associated with it in our corpus reflect the ways Muslims and ‘jihad’ has been represented in the mainstream media in the recent past. Muslims are portrayed as being inherently violent, backward and oriented to the destruction of the West. The ‘otherness’ of Muslims is what Ameli et al. (2007: 14) call ‘new ways of racism’, defined by Van Dijk (2000) as being more subtle and of a symbolic nature; discursive and expressed in text and in everyday talk.	A re-analysis of the corpus at more recent date may shed insight into how the discourse around #jihad has been impacted by the emergence of ISIS.
8	White II & Crandall (2017)	The study investiagtes whether the claim of "free speech" provides cover and justification for prejudice? The aim of the research is to find whether prejudiced people strategically use freedom of speech as a justification for - or defense against - these punishments for racism? Two main hypothesis were considered, (a) Learning someone else was punished for a prejudice that one shares threatens one's self-image and (b) Seeing someone	Justification-suppression model of the experience and expression of prejudice.	Seven studies were conducted, with 1078 participants in total. These experimental studies were held considering some racist events that went viral on the internet.	Survey	The main finding of the research is that prejudiced people justify another person's prejudiced speech. It was found that explicit racial prejudice is a reliable predictor of the “free speech defense” of racist expression. Participants endorsed free speech values for singing racists songs or posting racist comments on social media; people high in prejudice endorsed free speech more than people low in prejudice. This endorsement was not principled— high levels of prejudice did not predict endorsement of free speech values when identical speech was directed at coworkers or the police. Participants low in explicit	

		else punished threatens one's sense of freedom, triggering reactance.				racial prejudice actively avoided endorsing free speech values in racialized conditions compared to nonracial conditions, but participants high in racial prejudice increased their endorsement of free speech values in racialized conditions. Three experiments failed to find evidence that defense of racist speech by the highly prejudiced was based in self-relevant or self-protective motives. Two experiments found evidence that the free speech argument protected participants' own freedom to express their attitudes; the defense of other's racist speech seems motivated more by threats to autonomy than threats to self-regard. These studies serve as an elaboration of the Justification-Suppression Model (Crandall & Eshleman, 2003) of prejudice expression. The justification of racist speech by endorsing fundamental political values can serve to buffer racial and hate speech from normative disapproval.	
9	Yamaguchi (2015)	This article investigates the use of online communication an social media in connection with the ACM in Japan. The primary focus of the study is the significance of the Internet and online video streaming and sharing in particular for the ACM. It also examines the function of those media in the making of the movement's action styles, by fostering real-time, synchoronous communicatio between activists and spectators. This research also explored the problems resulting from the mmovement'S excessive dependency on online videos.		Demonstration as performance by ACM with online video sharing on in June 2010	Ethnographic descriptionn of the influence of online video on the movement and its actions.	The ACM successfully used the internet to spread its racist agenda, but such tactics also had negative effects. To appeal to a wider audience, ACM activits sought to present themselves as "ordinary" citizens, yet, at the same time, the extensive recording and dissemination of aggressive hate speech to attract viewers created a form og celebrity that undermined the very movement that spawned it. The style also has caused serious problems to the movement itself and to people influenced by such actions and speeches.	
10	Chetty & Alathur (2018)	This article is a review on hate speech in the context of online social netwroks. Initially the definitions of hate speech by different researchers are reviewed. In this		Definition of hate speech by researchers. Interna-	Review	The study concludes that the existence of online social networks led to increase in features such as contact establishment, message exchange, infromation sharing and news posting with the penalties such as	In the future the researchers can work towards any of the approaches to counter hate speech.

		article legal framework on hate speech from international bodies is also observed. Gender based hate speech is also reviewed. And finally cyber-terrorist networks are also discussed.		tional legal frameworks for hate speech from India, Canada, UK, Poland, UAE and USA. Comparison of works on religious hate speech, comparison of works on hybrid hate speech targetting multiple identities.		hate speech, hate crime, cyberterrorism and extremism. It has been identified that by framing proper policies from the government in association with the Internet Service Providers (ISPs) and online social networks, countering both hate speech and terrorism is efficient and effective. Therefore, there is a necessity to develop policies and methods to prevent and control these online activities. As women are one of the targets of online hate speech, it is necessary to have mandatory gender information while creating online social network accounts. In case of any suspect, this gender identity information can be used to watch internet traffic to and from female accounts while maintaining the freedom of expression. With this knowledge, the possibility of joining a female to any terrorist organizations can be reduced. Other possible approaches to counter hate speech are speech vs. speech, education and training, public awareness meeting on hate speech, making public more tolerant, usage of hate speech monitoring systems, and television broadcast programmes.	
11	Harell (2010)	This study examines the influence of ethnic and racial network diversity on young people's attitudes about speech rights in Canada by examining the impact of diversity on racist groups' speech compared to other objectionable speech.	Model of social network effects	The data are drawn from the Canadian Youth Study, a sample of 10th- and 11th-grade students in Quebec and Ontario (N53,334).	The study presents multinomial logistic regressions to assess the impact of network diversity on three types of political tolerance dispositions.	The analysis suggests that exposure to racial and ethnic diversity in one's social networks decreases political tolerance of racist speech while simultaneously having a positive effect on political tolerance of other types of objectionable speech. The dual effects arguably represent an evolving norm of multicultural political tolerance, in which citizens endorse legal limits on racist speech.	Future work should assess the extent to which target group distinctions in political tolerance judgments have evolved over time and across age cohorts.

12	Antoci et al. (2016)	This research study defines an evolutionary game framework to analyse the dynamics of civil and uncivil ways of interaction in online social networks and their consequences for collective behaviour. The purpose of the study is to define incivility as a manner of offensive interaction that can range from aggressive commenting in threads, incensed discussion and rude critiques, to outrageous claims, hate speech and harassment.	Mean field evolutionary framework.	Homogenous population, were individuals have the same access to technologies, but can pursue three different strategies of social interaction.		The findings of the study state that, when the initial share of the population of polite users reaches a critical level, civility becomes generalized if its payoff increases more then that of incivility with the spreading of politeness in online interactions. Otherwise, the spreading of self-protective behaviours to cope with online incivility can lead to economy to non-socially optimal stationary states.	Future research should consider relaxing the mean-field assumption that the researchers adopted in their framework. Furthermore, the future research should address te role of homophily by analysing how P and the H strategies interact with other users' personal features such as, their opinions.
13	Mantilla (2013)	This essay attempts to identify the distinct features of gendertrolling and bring attention to recent examples from a range of internet communities.		Case of Kathy Sierra, 2007. Melissa McEwan 2007, Anita Sarkeesian, Daniel Tosh 2012, Zerlina Maxwell 2012	Essay	The characteristics of these online campaign against outspoken women echo the misogynistic responses to the "Who Needs Feminism?" campaign. Gendertrolling has much in common with other offline targeting of women such as sexual harassment in the workplace and street harassment. In those arenas, as is the case with gendertrolling, the harassment is about patrolling gender boundaries and using insults, hate, and threats of violence and /or rape to ensure that women and girls are either kept out of, or play subservient roles in, male-dominated arenas. Sexual harassment of women is a behaviour that functions to inhibit women from fully occupying professional environments and fully competing with men.	
14	Langford & Speight (2015)	In this essay the researcher argues that the #BlackLivesMatter hashtag provides a rhetorical space to rescript Black bodies. It begins by discussing how the hashtag can be considered a grassroots movement. Next it discusses the counter movements that seek to invalidate the #BlackLivesMatter movement.	Critical theory	Social media campaign #BlackLivesMatter	Essay	The research reveals an epistemological logic of #BlackLivesMatter that moves from granting Black individuals presence to creating a rhetorical space to re-script the Black body as valuable. First, Black individuals have a positive presence—they are not invisible or portrayed as a negative stereotype. Second, violence against the Black body is news—the violence against this marginalized community cannot be ignored. Third, white privilege is unmasked by calling attention to the violence and marginalization	

						perpetuated against Black individuals. Fourth, color-blind rhetoric, which argues that we live in a post-racial society, advances the civil rights and civil liberties of African Americans.	
15	Uysal, Schroeder & Taylor (2012)	This article explores how Turkey is using social media via Twitter, a public relations strategy, to spread its messages and to establish itself within the international community.		Three top Turkish governmental officials' personal and official Twitter accounts	Qualitative content analysis	Turkey is wielding its soft-power in both the West and the Middle East/North Africa regions. Yet the quantitative analysis reveals that the western emphasis is more prominent in the messages. In its Twitter messages, the Turkish government follows an image cultivation and information subsidy approach in public diplomacy. Contrary to the highly interactive features of this social media tool, Turkish bureaucracy is not engaged in building relationships with its publics through Twitter.	Future studies could examine the use of other social media tools and social media activism in the context of public diplomacy. Alternative methods, such as surveys, experiments and interviews with the tweeters and followers would no doubt provide additional insight into the reach of soft power and the role of public relations in public diplomacy.
16	Al-Tahmazi (2015)	The research aims to show how the pursuit of power polarizes political discussions on Facebook and consequently constructs online sociopolitical communities. The article investigates how the pursuit of power, by means of de/legitimization, is produced and perceived in the Iraqi political discourses produced in social media as discourses of ethno-sectarian and cultural contestations.	Political discourse	The corpus analyzed in this paper represents three comment-threads consists of 396 individual comments (comprising 8322 words in total) selected from three publically available Facebook pages of leading Iraqi political commentators.	Positioning analysis	The results show that recontextualizing political actions and actors to de/legitimize particular interpretations of political reality based on differentiation and exclusion polarizes the discussions on Facebook. The delegitimization process that is based on differentiation and exclusion emphasizes the distinction between in-groups and out-groups and motivates the commentators to categorize themselves in oppositional sociopolitical communities that are discursively constructed. These sociopolitical communities range from completely imagined communities to the online recreation of actual ethno-sectarian groups.	
17	Piechota (2014)	This article studies the role of new media in overcoming schemata and prejudice of students in two different cities Berlin(Germany) and Krakow (Poland) with different levels of multiculturalism in the local community was carried out.		200 randomly selected students	Survey	The carried out pilot survey revealed differences in attitudes of students from Berlin and Krakow. Students in Krakow more often communicate with the use of social media than students in Berlin. The latter at the same time declare that they more often use social media to search for information connected	An interesting area for qualitative research that may be continuation of the carried out pilot study, is the observation of communication in social media in different groups and communities whose aim is to promote tol-

						with multiculturalism, promoting tolerance and helping immigrants to assimilate with the society. At the same time both groups declared a rather low level of interest and activity in groups, whose aim is opposing to building multicultural societies, what may be treated as a positive effect. The analysis how students in both cities use social media has shown that they do not use them to start new relations but only move relations existing in real life to the Internet. We cannot therefore definitely say that students' activity in social media influences overcoming stereotypes and eliminating prejudice, although in the long run it may be important, particularly in the light of increasing educational mobility of students.	erance, equality, intercultural dialogue, and supporting immigrants in their assimilation with the environment.
18	Schaffar (2016)	The paper analyses the genesis of two vigilante Facebook groups, Social Sanction group and Rubbish Collector Organization in Thailand. The aim of these groups is to expose political opponents by accusing them of lèse-majesté, which can result in a prison sentence of 15 years or more.	Case study	Screenshots of positings found in the RCO Facebook group in summer 2015.	Analysis of the screenshots, interviews and informal talks with Thai friends and colleagues who were threatened or targeted by violence attacks conected with Facebook.	The analysis of online communication in July 2015 shows that despite the large number of several hundreds comments connected to one post, each comment was responded from Rienthong's personal account - a clear sign that there is professional staff behind this account. Also in stark contrast to the image of the 'common man of the streets' is the militancy and violence that was apparent in the language of the RCO's official proclamations and Facebook posts. The ritual performance of indignation, followed by hate speech and the documentation of actions, under the guidance of a fatherly but uncompromising and rigorous leader, was increasingly combined with calls for and documentation of mass mobilization of members 'performing' their loyalty to the monarchy. In this respect too, the RCO page constitutes a new development compared to the SS page. Whereas older Facebook pages served as fora for the documentation of private initiatives, the RCO's, with its prominent individual members and its mass membership, triggered a new effect.	Further studies on similar groups will be needed to get a mroe complete picture of the recent rise of vigilante groups on the internet. A crucial question to ask will be in how far the specific features of Facebook, the general trend toward political polarization, and more or less dormant legacies of Fascist vigilantism are interlinked.

19	Horbyk (2018)	The main objective of this research is to investigate how different ethnolinguistic identities were constructed in informal public online communication on the eve of the Eurimaidan protests. The research compared the self perception of Ianukovzch's controvertial language policy. It also examined the linguistic situation in Ukraine.	Ethnolinguistic identity theory	Nearly 3,000 readers' comments concerning language issues posted to Ukraine's leading news website *Ukrains'ka pravda* (*Ukrainian Truth*) in 2010-12	Critical discourse analysis	At first sight, news readers' comments on Ukrains'ka pravda during the sampling timeframes embodied a classical East European srach, or, to use its apt English equivalent, a "shitstorm." The Ukrains'ka pravda commenters had both optimist and pessimist perspectives of Ukrainian language vitality. However, this ambiguity should be interpreted in relation to the status of the competing language, Russian. An evaluation of the comments posted revealed that there was not much concern about the vitality of the Russian language in Ukraine and there was considerable concern about the vitality of the Ukrainian language in Ukraine. This analysis shows that Ukrainophones' assimilation into the Russophone group in 2010-12 was likely obstructed by factors such as language proximity and ease of code-switching, but also by the unique official status of the Ukrainian language that increased its perceived vitality (in line with Bilaniuk and Melnyk's findings).	This study presents an avenue for future research: going beyond the virtual space into the real world, with individual biographies, case studies, in-depth interviews and focus groups aimed at locating personal motivations and strategies, could help understand how Ukrainian society accumulated energy for its outpouring of anger during Euromaidan and how its subsequent events are shaping the current media and language landscapes.
20	Maweu (2013)	This article examines if the increased political discussions on social media especially Twitter and Facebook before and after the March 4th, 2013 general elections in Kenya translated to a more robust alternative public sphere that broke the hegemony of the traditional media as agenda setters or an alternative space for the audience to vent out their frustrations and grievances about the election. The main aim was to examine how citiyens used new media (Twitter and Facebook) to fight out their ethnic wars online.		A purposive sample of 30 hate messages exchanged between January 2013 and May 2013 was chosen based on two categoroies used by Umati to monitor hate speech: Offensive speech and Extremely dangerous speech. The research sampled 15 messages from each category.	Qualitative content analsysis	From the analysis it was evident that immediately after the elections on March 4th, there was an increase in extremely inciting messages targeted at three main tribes (Kikuyus, from which the current president, Uhuru Kenyatta hails; Kalenjin, from which the Deputy President William Ruto hails; and Luo from which Raila Odinga, the main loser comes from) as well as supporters of the two main political parties (Jubilee supporters and CORD supporters). It was also evident that most of the inciting speech online was as a response to events happening on the ground as reported by the mainstream media. The highly inciting speech ranged from extremely vulgar language directed to members of a particular tribe, to calling members of one tribe to kill the other to advocating for eviction of a particular tribe from their	

						land. The pattern of hate speech circulated in 2013 was very similar in tone to that circulated in 2007/2008 though this didn’t culminate to physical violence, but very fierce soft ethnic war online. There were several Social Media pages such as ‘Not another Kikuyu President’ and ‘STOP Raila NOW’ where supporters of either presidential candidate traded insults and offensive remarks.	
21	Kimotho & Nyaga (2016)	This paper investigated the nature of digitized hate speech by: describing the forms of ethnic hate speech on social media in Kenya; the effects of ethnic hate speech on Kenyan’s perception of ethnic entities; ethnic conflict and ethics of citizen journalism.	Descriptive interpretive design, and Speech Act Theory	Purposive sampling was used to pick two public and two private universities in Kenya. Questionnaires were administered to students in the four universities. Content published between January and April 2013 from six purposefully identified blogs was analysed. The datasets from the eight sites yielded 35,915 speech acts. Data appearing on these sites between 04.11.12 and 16.05.2013 were analysed.	Descriptive interpretive design by using qualitative and quantitative appraoches. Qualitative data were analysed using NVIVO 10 software, while responses from the questionnaire were analysed using IBM SPSS version 21.	The findings indicated that Facebook and Twitter were the main platforms used to express ethnic hatred. Hate speech incited hatred and conflict for political gain. Ethical issues raised included moral subordination and incivility. The digital platforms mostly used for hate speech in were Facebook, twitter and personal blogs and instagram and they accounted for 96.6 of total posts. This compares with the Umati report which indicated that over 90% of all online inflammatory speech captured by Umati was on Facebook, making it the highest source of such content. This study demonstrated that digital media hate speech disseminators had varied intentions raging from inciting hatred, violence, or moral subordination among others. Nevertheless, the magnitude of incivility that accompanied hate messages on digital platform in Kenya was appalling. Hate speech, and the accompanying ethical issues it raises, are detrimental to the welfare of a nation and its people. Digitized hate speech adds speed and volume to such messages and can only be doubly destructive.	Further research need to be done on other types of hate speech including racist hate speech, religious hate speech and gender. Another area that deserves further investigation is on the effects of digitized hate speech on the target individuals or groups.
22	Alakali, Faga & Mbursa (2017)	The problem this paper intends to study therefore includes why hate speech and foul language plague the social media in	Mediamorphosis theory and public sphere theory	384 respondents	Used questionnaire and focus group discussion as	This study indicate that promoting hate speech and foul language on social media have moral consequences in the society and to journalism practice.	

		Nigeria and what ramifications this nuisance has in the society and for the journalism profession. Most importantly, the paper investigates the consequences of these practices in the social media, to morality, ethics and law in the society.			instruments for data collection. Also, the paper adopted the qualitative, doctrinal and analytical methodology to discuss the legal consequences and obligations created against perpetrators of hate speech and foul language in Nigeria.	These consequences include loss of credibility, diverting media from fulfilling their primary role of serving the public interest and increasing moral decadence in the society. Further findings indicate that freedom of speech on social media and political interest are the major factors that motivate the posting of hate speech and foul language on social media platforms in Nigeria and that majority of hate speech prevalent on social media platforms in Nigeria is politically motivated hate speech. Findings also reveal that hate speech and foul language has negative implications on social media as it leads to unwanted censorship of social media platforms among others. The study also found that although, most people in Nigeria are aware that there need to enact law to regulate the increasing spate of hate speech and foul language on the social media, however, they are unaware if there are already any existing legal measures against the practice in Nigeria. Finally, findings of the study established that hate speech and foul language on social media platforms cannot be constricted to conform to the ethical standards of journalism practice in Nigeria because most perpetrators of this practice are not journalist.	
23	Shepherd et al. (2015)	This article presents a dialogue between digital culture scholars on the seemingly increased presence of hating and hate speech online.		Primarily revolves around #GamerGame campaign of intensly misogynistic discourse aimed at women in video games.	Roundtable discussion	The discussion suggests that the current moment for hate online needs to be situated historically. From the perspective of intersecting cultural histories of hate speech, discrimination, and networked communication, we interrogate the ontological specificity of online hating before going on to explore potential responses to the harmful consequences of hateful speech. Finally, a research agenda for furthering the historical understandings of contemporary online hating is suggested in order to address the urgent need for scholarly interventions into the exclusionary cultures of networked media.	

24	Jakubowicz (2107)	This article charts the most recent rise and confusion of the Internet under the impact of the Alt_Right and other racist groups, focusing on an Australian example that demonstrates the way in which a group could manipulate the contradictions of the Internet with some success.	An analytical model.	An Australian Study, 'Cyber Racism and Community Resilience' (Jakuwicz et al. 2017)	Draws and reflects on one aspect of the Australian study of online racism, namely antisemitism, and the rise of online neo-nazism	There are three areas of law that could be addressed. At the global level, Australia could withdraw its reservation to article 4 of the International Convention to Eliminate All Forms of Racial Discrimination. Such a move has been flagged in the past, but stymied by relentless opposition from an alliance of free speech and social conservative activists and politicians. Australian law could move to recognise European legislation on Cyber Crime, and adopt the Additional Protocol as it has for the overall legislation. Finally, Australia could adopt a version of New Zealand's approach to cyber hate, where platforms are held ultimately accountable for the publication of online content that seriously offends, and users can challenge the failure of platforms to take down offensive material in the realm of race hate. There are many initiatives in civil society that would empower those who are currently the targets, and disempower those who are the current perpetrators of race hate. Firstly, people who are targeted by racists need support and affirmation; this approach underpins the approach that the E Safety commissioner has undertaken in the development of a Young and Safe portal. There could be a CyberLine for tipping and reporting race hate speech online, for follow up and possible legal action. Anti-racism workshops (some have already been run by the E Safety commissioner) have aimed to pushback against hate, while building structures where people can come together online.	
25	Suntai & Targema (2017)	The study explores the contribution of the new media in the entrenchment of democracy in the country, and critically assesses issues and matters arising with the adaptation of the platform by both the government and the masses.	Social Responsibility Theory	Arsenal of information dissemination on social media during the general elections of	Case study	While the new media appears to provide vibrant discursive channels that will facilitate democracy in the country, a careful observation of the trend reveal quite a number of threats that are not only worrisome, but have the capacity to diminish the opportu-	

				2015 in Nigeria and its aftermath.		nities which they offer to countries with budding democracies like Nigeria. The issue here is that, although the divide between North and South has existed in the country for long, new media platforms accentuated the division, and created an atmosphere full of enmity for one another during the 2015 elections. Sentiments that lie latent in the minds of people were given a voice, and widely expressed. This development poses a great threat to the fragile democracy which the country is striving to consolidate. Conclusively, new media platforms are formidable forces in the consolidation of democracy. The information gap, which they help to bridge, benefits democracy in no small measure, and serves to strengthen the cherished principles of transparency and accountability in the process of governance. Nigerian democracy is speedily heading towards this enviable destination courtesy of the new media. Similarly, the new media platforms have extended the frontiers of political participation and interaction between the rulers and the ruled. This is a positive development that needs to be acknowledge, as it makes democracy in the country to live up to the expectations of its generic definition, as government of the people, for the people, and for the people.	
26	Abraham (2014)	This article makes a case study of ‘flarfing’ in order to contribute to an understanding of the potentials and limitations facing users of online social networking sites who wish to address the issue of online hate speech. The research explores one case of users acting creatively within Facebook’s technical and regulatory environment to take small-scale actions against hate speech.		Facebook user activities online over a period of years. The majority of examples are drawn from the year 2012 which had the most flarf activity.		Facebook flarf presents a useful case study for theories of regulating and responding to hate speech online. Facebook flarf has some ability to drown out hate speech practically and aesthetically, but perhaps more importantly it can serve to communicate social opprobrium and community limits on acceptable discourse online. Facebook flarf represents an encouraging attempt by users to ‘take responsibility’ for online hate speech and online culture in the spaces they frequent, through personalisation and the per-	

						formance of an expertise within the platforms affordances. It also communicates a meta-textual and reflexive awareness of the medium of communication itself. The research situated the practice of Facebook flarfing for activist ends within a contemporary context of ubiquitous memes and the uncertainty around the sincerity of online comments and discourse, viewing flarf as an example of discursive activism that repurposes the tropes and practices of troll culture.	
27	Alam, Raina & Siddiqui (2016)	This paper aims to examine the take of people on the “Free Speech via SocialMedia” issue and their attitude towards the way sensitive messages/information are posted, shared and forwarded on social media, especially, Facebook.		200 social media users (100 males and 100 females), randomly picked from five Indian states/Union Territories.	Quantitative analysis (Kolmogorov-Smirnov Z test).	The findings indicate that hate posts/messages are on the rise, and more and more users Are joining in. Besides, prosecution happens only when the aggrieved party is influential or powerful. The findings of this research give a strong insight into the social media behaviour of users in relation to hate contents/posts. The study establishes the fact that Indian people are in favour of free speech, but with a sense of restraint and responsibility.	The work could form the basis for future research on various aspects of hate speech on social media. Researchers could study the trials and prosecutions that have happened over the past few years and whether punishment has acted as a deterrent.
28	Pitsilis, Ramampiaro & Langseth (2018)	This research addresses the important problem of discerning hateful content in social media. The research question addressed in this work is, how to effectively identify the class of a new posting, given the identity of the posting user and the history of postingd related to that user?	A detection scheme was proposed that is an ensemble of Recurrent Neural Network (RNN) classifiers, and it incorporates various features associated with user related information, such as the users’ tendency towards racism or sexism.	16 thousand tweets publically available	This data is fed as input to the RNN classifiers along with the word frequency vectors derived from the textual content.	The experimental results have shown that this approach outperforms the current state-of-the-art approaches, and no other model has achieved better performance in classifying short messages. Also, the results have confirmed the original hypothesis of improving the classifier’s performance by employing additional user based features into the prediction mechanism.	Future research can investigate other sources of information that can be utilized to detect hateful messages.

29	Burnap & Williams (2015)	In this article a supervised machine learning classifier for hateful and antogonistic content in Twitter is developed. The purpose of the classifier is to assist policy and decision makers in monitoring the public reaction to large-scale emotiive events.	Case stduy (Murder of Drummer Lee Rigby in Woolwich, London, UK.	The study data set was collected from Twitter during a two-week time window following the "trigger" event -the murder of Drummer Lee Rigby in Woolwich, London, UK on May 22, 2013. Total 450,000 tweets were collected and a sample of 2000 were coded.	10-fold cross validation approach	The classification results showed very high levels of performance at reducing false positives and produced promising results with respect to false negatives. The implementation of individual probabilistic, rule-based, and spatial classifiers performed similarly across most feature sets, but the combination of the classification output of these base classifiers using a voted meta-classifier based on maximum probability matched or improved on the recall of the base classifiers in every experiment, suggesting that an ensemble classification approach is most suitable for classifying cyber hate, given the current feature sets. This could be due to the noise and variety of types of response within the data, with some features proving more effective with different classifiers. Also, an illustrative example using cyber hate as classified by a machine as a predictive feature in a statistical regression model is developed. The model produced IRRs for retweet activity given a set of features for each tweet. The model showed a reduction in retweet rate ratio when a tweet contained a hateful or antagonistic response, suggesting a stemming of the flow of content on Twitter when a tweet contained cyber hate.	This article could act as a clarion call for further research into cyber hate and its manifestation in social media around events, and the development of technical solutions that are informed by such research.
30	Räsänen, et al. (2015)	How exposure to hate material in the internet correlates with Finnish youths' particularized and generalized trust toward people who have varying significance in different contexts of life. Hence, the purpose of this paper is to provide new information about current online culture and its potentially negative characteristics. We investigate the relationship between exposure to online hate material and respondents' trust in their		15 to 18 year old Finnish Facebook users, in the spring of 2013. Sample size 723.	Online survey using three Facebook advertisement campaigns in April-May 2013.	The results indicate that online hatred can have social impacts and influence young people's trust toward other people. In particular, exposure to online hate material clearly influences levels of both particularized and generalized trust. It is noticeable that young Finns have relatively It also appears that witnessing hate material online has a greater effect on the levels of particularized trust than generalized trust. The results indicate that while exposure to online hate materials does reduce generalized trust, its influence is greatest on particularized trust. It further indicates that exposure to online hate material is	In terms of suggestions for future research, as earlier researchers have found, levels of trust and levels of happiness appear to be positively related. It is therefore likely that exposure to online hate material would have a similar correlation with the levels of happiness. In addition, future research should examine how online hate material influences different age groups in terms of their perceived trust. Similarly, researchers

		family, close friends, other acquaintances, work or school colleagues, neighbors, people in general, and people they met only online.				not only relatively common, but it also has consequences for the young people who witness such material in their daily lives.	should compare levels of trust across different age groups to see if older individuals are more trusting toward online acquaintances than younger individuals are.