

**Citation:** Khan, S., Khan, M. R., & Ghauri, M. J. (2023). Hate Speech through Social Media: A Systematic Review. *Global Digital & Print Media Review*, VI(II), 175-182.

[https://doi.org/10.31703/gdpmr.2023\(VI-II\).11](https://doi.org/10.31703/gdpmr.2023(VI-II).11)

- Vol. VI, No. II (Spring 2023)
- Pages: 175 – 182
- p- ISSN: 2788-4988
- e-ISSN: 2788-4945

▪ URL: [http://dx.doi.org/10.31703/gdpmr.2023\(VI-II\).11](http://dx.doi.org/10.31703/gdpmr.2023(VI-II).11)

▪ DOI: 10.31703/gdpmr.2023(VI-II).11



Cite Us



Shanza Khan \*

Muhammad Riaz Khan †

Muhammad Junaid Ghauri‡

## Hate Speech through Social Media: A Systematic Review

**Abstract:** *Hate speech is a complicated concept and there is no locally recognized definition for it in Pakistan. But there are some academic publications and court precedents that tell us about what comes in the ambit of 'hate speech'. According to them any danger and damage caused by certain forms of expressions have been globally acknowledged in defining hate speech. Our study explores existing research findings by using systematic review analysis on how social media may or may not create an opportunity for online hate speech and which kind of hate speech is mostly being disseminated on social media. A sample of 20 studies out of a total of 50 research papers found in the searches was analyzed that discussed online hate speech from 2015 to 2020. The reviewed studies provide exploratory data about reasons for hate speech happening on social media and how social media make space for hate speech and cyber hate. The findings of this study provide recommendations to counter hate speech on social media.*

**Key Words:** Hate Speech, Cyber Hate, Social Media

**Corresponding Author:** Muhammad Junaid Ghauri (Lecturer, Department of Media and Communication Studies, International Islamic University, Islamabad, Pakistan.  
Email: [Muhammad.junaid@iiu.edu.pk](mailto:Muhammad.junaid@iiu.edu.pk))

### Introduction

Social media is abundant with hate speech. An overlook of any social media platforms' comments section especially a YouTube video demonstrates how prevalent the problem is. Even though main social media corporations such as Facebook, Instagram and Twitter have their own guidelines about to what extent hate speech is acceptable on their sites but most

guidelines are difficult to understand for users (Elizabeth, 2013). So the most basic human right (freedom of expression) is being misused on social media platforms. Freedom of expression is the right of every individual to express his/ her feelings, share opinions and should have access to knowledge and information. It gives the right to citizens to participate in the democratic process but there

\* MS Scholar, Department of Media and Communication Studies, International Islamic University, Islamabad, Pakistan.

† English Language Instructor, Jazan University, Saudi Arabia.

‡ Lecturer, Department of Media and Communication Studies, International Islamic University, Islamabad, Pakistan.

is more hate found on social media than any healthy discussion and participation.

But we cannot ignore the role and importance of social media in changing the mode of communication in Pakistan. Social media with all its new features has revolutionized our lifestyle, it has modernised our ways of getting news and it has helped us get involved in the news production and dissemination process. It has provided us with new and effective tools to talk to other people. It has a scattered as well as democratic nature, which let people produce and share ideas and information. Because it has democratic nature, so many websites are being activated to spread hate against particular groups and communities (Tariq, [2016](#)). As a result of the mushroom growth of these websites, the chances of the creation of hate speech have increased. There are very visible conflicts in Pakistani online space which include general abuse, discrimination and creating propaganda and hate against specific groups, institutions and ideas. The most common hate which is found in Pakistani online spaces is religious, political and ethnic hate. Private hate is also generated and spread in online media (Haque & Ilyas, [2014](#)).

The inclinations of online networking in Pakistan are the same as that of any developed country of the world (Riaz, [2010](#)). By the end of 2016, Facebook was a mainstream social network with 25 million users in Pakistan. And in 2017 Facebook users escalate to 30 million and Twitter 3.1 million according to Pakistan Annual Social Media Report (Kemp, [2021](#)). Not only in Pakistan but other countries also facing such issues in their online space. According to a civil rights organization that tracks and sues such activities (Southern poverty law centre), from 2000 to 2012 number of active hate groups rise by 69% (SPL, [2021](#)).

Simon Wiesenthal Center shared a report detailing hate and terror on online platforms during 2012 which says that there are approximately 15,000 hate speech-spreading websites that also include games and different applications (Simon Wiesenthal Center, [2012](#)). Most of these websites create and spread hateful rhetoric, videos and music while others

ask for physical violence as well. For example, there is a movement known as 'creativity movement kids' which encourages people to arm themselves to fight against the mud races by labelling the fight as a 'divine fight' (Creativity Movement Kids, 2012).

Pakistanis are frequently becoming a target of hate speech coming from bordering countries, mostly India. Because of such online arguments sometimes consequences in attacks on Islam, Muslims and the ministry in Pakistan (Haque & Ilyas, [2014](#)). Ethical, racial and religious kind of hate content is very common on social media platforms including Facebook, Twitter, YouTube, and Yahoo groups (British Institute, [2012](#)).

Although every social media platform has its own rule of what is being accepted and what is not, however, Twitter's approach is very different from all of them as it tends to be more liberal (Institute, [2012](#)). However, there is difficult to understand what these policies contain and, more significantly, how they are being practised. In 2012 Google was criticised for removing the film from its video-sending sites in some countries but not in others, which was based on hate-propaganda named, "The Innocence of Muslims" (Stephen, [2017](#)). The one-hour film in which Prophet Muhammad S.A.W was taunted got the reactions of the Muslim community because it was so offensive for the Muslim community. Throughout the Middle East, there were so many protests. So what Google did that, it blocked the hate content in several countries that includes Libya and Egypt also but it was left it up in so many other countries by saying that this content doesn't meet the criteria of hate content (Sengupta, [2012](#)). In Pakistan, many people were murdered as a result of this hate content and what critics said is that Google is accountable for permitting such content, which was responsible for global turmoil, to remain posted for a long period of time. Irrespective of who is responsible, this incident tells us that power corporations like Google, Facebook etc. have to make policies because there are no such regulations from the US to prohibit hate speech. All these online companies must rely on their own created

guidelines to determine on a case-by-case basis which content is considered hate speech and therefore not permit them (Elizabeth, 2013).

Along with the physical harm, psychological harm is another major reason why hate speech needs to be banned online. Hate speech affects the marginalized groups attacked by various social, political, cultural, ethnic and religious propaganda (Klein, 2012). If people are regularly exposed to hate and abusive content on social media and personally attacked, so then ultimately they will begin to distrust their own self-worth (Clark, 1965). Critical theorist Mari Matsuda believes that the worst effect of hate speech is the tendency of defaming a person, a group and/or an idea which can result in the form of hate creation that can ultimately affect the self-esteem of the victims. According to various social research findings hate speech against a particular social or racial group can cause a sense of inferiority among the victims. Such dichotomous feelings can further cause polarization in society. Such polarization can cause discrimination, racial attacks, and violation of basic human rights in society (Matsuda & Matsuda, 2012).

Despite the fact that there exist clear theoretical gaps to relate such behaviour with internet use and we cannot exclusively claim that there is any definite link between the use of the internet and the creation of hate but all this literature intrigues the researchers to look into such relationships. In modern-day times, studying the use and impact of the internet and particularly social media in the dissemination of hate speech is an important area of research. So based on this need, this paper has analyzed various research findings explaining how social media/social networking sites along with internet use are related to online hate speech. This would be the main contribution of this systematic review that it would prove to be a baseline to understand the relationship between the spread of hate speech and social media. The researcher would benefit from this systematic review in way that they can find various research findings related to the use of social media in spreading hate speech in one place.

This systematic review would be helpful in devising recommendations to cope with the problem of hate speech being spread through social media.

## Research Design and Methodology

### Data Collection and Analysis

---

The studied literature was collected in August, September, and October 2021 from Google Scholar and the Research Gate. The Research Gate was selected because researchers mainly prefer to publish their work on this website and there are so many good research articles on Research Gate. Google Scholar was selected due to its significance in the area of article citations as a metric of propagation.

Our study is limited to the following five points sampling criteria for the research articles; (1) articles published in the form of academic research articles, (2) publications focusing on online hate speech and/or cyber hate, (3) publications written in English, (4) publications focusing only on social media platforms, and (5) publications that were published during 2015-2020. Articles that focused on hate speech on offline platforms or explored and analyzed similar terms such as cybercrime, cyber terrorism, populism on cyberspace, web nationalism, cyberbullying or sexting were excluded.

In this study, we have employed grounded theory to analyze the literature and information in it with an inductive analysis orientation. To get a summary of the material, there were no pre-settled themes before the analysis started (Strauss, 1997). We have created theoretical models and used emergent categories, where theory is inductively derived from data and also it has given the pieces of evidence for the conclusion. Systematic methods of data collection have been employed to collect the relevant selected literature. The final output of this review is a qualitative analysis of the findings (Rodgers & Chabrol, 2009).

### Findings and Analysis

---

Following the selection criteria we collected a total of 50 research articles from both the

websites; Google Scholar and Research Gate. Out of a total of 50, we excluded 10 articles that fell in the exclusion criteria and 20 were excluded because of the duplicate and irrelevant data and materials. The remaining 20 articles came under the umbrella of selection criteria and were analyzed. Information has been categorized into the three main types that came from data by resemblance of themes which are; (1) definitions, (2) types of online hate speech, and (3) methods of evaluation used in the articles. This type of method to analyze diverse data has been used in different research by different researchers (Roger, [2014](#)).

## Definitions

---

The definition of hate speech is very complex. There is no globally accepted definition for it. But ranges of definitions that vary in each country exist. Although there are different definitions of hate speech, every country has its own values to measure and prosecute those suspected of hate speech. In Pakistan, Article 19 gives freedom to speak and exchange information along with limitations, "Every citizen shall have the right to freedom of speech and expression, and there shall be freedom of the press, subject to any reasonable restrictions imposed by law in the interest of the glory of Islam or the integrity, security or defence of Pakistan or any part thereof, friendly relations with foreign States, public order, decency or morality, or in relation to contempt of court, or incitement to an offence<sup>12</sup>" (Pakistan constitution, [2021](#)). The constitution of Pakistan clearly prohibits hate speech and restricts its spread. Along with this, constitution article 31 titled, 'Islamic way of life' says that the state of Pakistan shall attempt with respect to the Muslims of the country, 'to encourage observance and unity of benchmarks', which may be taken to apply limitations on religious hate speech specifically about Islam (Pakistan constitution, [2021](#)). So, the hate speech is also prohibited by the constitution of Pakistan.

Pakistan penal code also addresses hate speech in article 153- A with the title, 'promoting enmity between different groups'

and stated that, 'whoever commits or organizes or incite any other person to organize violence or any such activity will be punished for up to five years with fine' (Pakistan Penal, [2013](#)). So, the Pakistan Penal Code also prohibits hate speech and restricts it with punishments.

## Types of Online Hate Speech

### Religious Hate Speech

---

By religious hate speech, we meant; any statement or expression that spread hatred against any specific group or community on the basis of their perceived faith which calls for violence against that group.

On the basis of religion, there are so many religious groups which are being targeted in online space. Muslim religious-political leaders and clerics are also targeted by hate speech and it is mostly a Muslim community that spread hate against them for their extremist perceptions and along with this, their involvement in some corruption cases. Under religious targets, there is sectarian hate in Pakistani online space and in Pakistan, the sectarian division is being spread by social media it has been observed that sectarian division is being escalated with the innovation of social media tools (PIPS, [2019](#)). The target of such hate is Shia Muslims. The Shia community has been the target of hate in an online space along with this they are being targeted physically in the form of terror attacks (Ebrahim, [2014](#)).

Another community that is the target of online hate is the Ahmadi community the basis of religion. There is a constant war between India and Pakistan and as a result, the Hindu religion is the main target in online space in Pakistan (Haque & Ilyas, [2014](#)).

### Terrorism as an Online Hate Speech Trigger

---

Unregulated cyber-space has increased the influence of terrorism. Now terrorist organizations use the internet to spread hate speech and activate violence in society (Malik, [2020](#)). Though, the threat of terrorism is now

more powerful because of its covert nature (Khan, [2019](#)).

### **Gender-based Hate Speech**

---

Hate speech on the basis of gender is even more common nowadays, especially when the month of March comes along with the 'Aurat March'. Females have been the target of hate speech and the basis of gender discrimination and differences in all circles of life (Ejaz & Ara, [2011](#)). Along with this LGBT also faces hatred in Pakistani cyberspace (Haque, [2013](#)). Pakistan penal code has criminalized 'unnatural offences'. In such an environment LGBT community members are the targets of online hate speech (Pakistan Penal, [2013](#)).

### **Race-based Hate Speech**

---

This type of hate speech is common, especially targeting local ethnicities of Pakistan including Pakhtoons that belong to the province of KPK and Balochs from Baluchistan. Words terrorist and extremist have been used for them in cyberspace. Along with these words, most other used words for local ethnicities are stupid and illiterate. On Facebook and Twitter, hate speech for Pakhtoons was 1% and for Balochs, it was less than 1% (Haque & Ilyas, [2014](#)).

### **Political Hate Speech**

---

Political hate speech is also very common in Pakistan. It is defined as a speech with political agenda. The violation of digital laws and ethics is now escalating by political parties. Their campaigns through social media encourage hate content and loosely define ethical parameters (Haque & Ilyas, [2014](#)). There is freedom of expression in online space which is being misused by social media users by expressing hate in online space and spreading negativity instead of some healthy material. Specifically, cyber propaganda is the use of cyberspace to spread hate content by different agencies (Khan, [2019](#)).

According to the findings of a research study the tweets containing hate content and targeting Pakistan Tehreek-i-Insaf (PTI) workers were more frequently re-tweeted and

in the same case for other parties. Various researchers and scholars have also argued that different political parties mostly used Facebook and Twitter for their campaigns and spread hate through these social media platforms. Researchers have also stated that Pakistani political parties don't miss any chance to turn down opponents without having any idea that it would definitely encourage and spread hate speech among people (Rao, [2020](#)).

### **Most Used Methods to Evaluate Online Hate Speech**

---

This section contains a summary of the methods employed by the researchers in the research articles under study. One of the most commonly used methods was a content analysis of different issues related to cyber hate. The thematic and critical discourse analysis was also employed for the qualitative analyses of cyber hate. The thematic analysis was used to detect the themes (Mondal & Ryan, [2018](#)), racism recognition, racial micro violence, inter-ethnic hate incidents, female and male victims, sexual harassment, hate against women and others. Some of the researchers have used gratification theory to analyze when and how people seek actively specific media to gratify specific needs (Khalil, [2016](#)).

There is significant evidence of the use of mixed analysis methods where most of the researchers have combined quantitative analysis techniques in order to deliver a complete model of topics (Merrill & Åkerlund, [2018](#)) along with qualitative methods i.e. in-depth interviews, discourse analysis etc. Overall, the most commonly used methods to evaluate online hate speech are based on qualitative approaches that are grounded in theory and discourse analysis etc. It is essential to improve empirical methods to evaluate cyber hate, especially among youngsters as the most regular consumers of the Internet.

### **Recommendations**

---

Pakistan already has a legal framework that empowers the state to tackle hate speech

online and hate mongers. What is actually needed is that there should be enough resources along with good training programs for law enforcement agencies to track and handle hate base content. The capacity of law enforcement agencies should be enhanced in order to deal with new innovations in cybercrimes. That data should be openly available for a continued public inquiry. There is very little literature on cyber security and cyberspace in Pakistan. Government should encourage setting up more opportunities and scholarships for students to research in cyberspace area.

"Cyberhate" has become a serious concern that needs to be addressed by educating social media users (Khalil, 2016). There should be public awareness campaigns by the government to literate and aware public about media and cyberspace rights and ethics. Many social media users in Pakistan are not media literate to understand the consequences of internet usage. Daily thousands of people fall prey to online scams through the internet. So for this purpose, advertisements and awareness campaigns are required to help the public how to keep themselves aware and secure in cyberspace.

On the national level, the establishment of cyber command is really needed to handle the issues of cyber warfare. This can be done under National Security Council to take all concerned persons on board and prepare defensive and offensive cyber war capabilities. Along with this, regulation of Pakistani cyberspace is critical for the national security of Pakistan. The illegal use of untraced pirated software, IPS and VPNs should be banned (Khan, 2019).

### Addressing Online Hate Speech in Pakistan

---

Hate speech in Pakistan's online space has to be addressed through different methods like a multi-pronged strategy approach that includes educating the public by creating awareness

about discouraging hate and intolerance and by criminalizing hate speech through law. But more specifically this goal to regulate cyberspace should not be sponsored and manipulated by political agendas. There is a need to collaborate with other international organizations that are working on this issue to define what exactly hate speech is and which type of actions comes under hate speech.

The role of media cannot be ignored in this case as media is the major player because there is a wide range of areas of local cyberspace which are being operated by media industries. Particularly mostly spread hate speech area in Pakistan is on Facebook and Twitter, so there is a dire need to regulate these platforms along with engaging the public to make them follow the code of conduct and best practices with respect to the online community. However, this is a very tricky path. While controlling hate speech it has to be kept in mind that the peoples' basic human right of freedom of expression must not be violated. People's right to freedom of speech must not be violated in the name of controlling hate speech. This is where the responsibility of the government and state institutions has increased.

The regulation of the most extreme level of online hate speech is needed and it would be one course of action. As Gill and Shareef (2020) stated that the nature of the internet gives itself another solution to the problem i.e. counter-speech. A collective and organized effort by the state along with society to make positive and progressive narratives to counter hate material in online space.

Overall, Pakistan needs a broader level campaign, outside the online context, starting from the education sector to the state level. But one of the biggest hurdles between these efforts is the lack of research in the context of cyberspace and hate speech. So, the government of Pakistan should encourage research in online hate speech and cyber security to make counter-narratives against hate speech.

## References

- Best, P., Manktelow, R., & Taylor, B. (2014). Online communication, social media and adolescent wellbeing: A systematic narrative review. *Children and Youth Services Review*, 41(41), 27–36. <https://doi.org/10.1016/j.childyouth.2014.03.001>
- Chen, Y., & Yang, D. Y. (2019). The Impact of Media Censorship: 1984 or Brave New World? *American Economic Review*, 109(6), 2294–2332. <https://doi.org/10.1257/aer.20171765>
- Clark, K. B. (1965). *Dark Ghetto: Dilemmas of Social Power*, 31.
- Ebrahim, Z. T. (2012). 'If this isn't a Shia genocide, what is?'. Dawn <http://www.dawn.com/news/744925/i-f-this-isnt-shia-genocide-what-is>,
- Ejaz, S. S., & Ara, A. A. (2011). Gender discrimination and the role of women in Pakistan. *Journal of Social Sciences and Humanities*, 50(1), 95-105.
- Gill, & Shareef. (2020). The cost of hate speech. *Minorityright.org*
- Haque, J. & Ilyas. (2014). *Hate Speech-A study of Pakistan's cyberspace*. Islamabad, Pakistan: bytes4all.
- Haque, J. (2013). *Pakistan's internet landscape*. Bytes for All, Pakistan.
- Institute, B. (2012). *Mapping Study on Projects Against Hate Speech Online Institute of Human Rights*. British Institute of human rights.
- Kemp, S. (2021). *Datareportal*. <https://datareportal.com/reports/digital-2021-pakistan>
- Khali, L. (2019). Identifying cyber hate: an overview of online hate speech policies & finding possible measures to counter hate speech on the internet. *Journal of Media Studies*, 31(1).
- Khan, M. I. (2019). Cyber-warfare: Implications for the national security of Pakistan. *NDU Journal*, 117-132.
- Klein, A. (2012). Slipping Racism into the Mainstream: A Theory of Information Laundering. *Communication Theory*, 22(4), 427–448. <https://doi.org/10.1111/j.1468-2885.2012.01415.x>
- Malik, D. A. (2020). Governing Abuse of Free Speech during Elections in Pakistan. *South Asian Studies*, 33(1), 239-251. [http://pu.edu.pk/images/journal/csas/PDF/15\\_v33\\_1\\_18.pdf](http://pu.edu.pk/images/journal/csas/PDF/15_v33_1_18.pdf)
- Matsuda, M. K., & Matsuda, M. K. (2012). *Pacific Worlds: a history of seas, peoples, and cultures*. Cambridge University Press.
- Merrill, S., & Åkerlund, M. (2018). Standing Up for Sweden? The Racist Discourses, Architectures and Affordances of an Anti-Immigration Facebook Group. *Journal of Computer-Mediated Communication*, 23(6), 332–353. <https://doi.org/10.1093/jcmc/zmy018>
- Pakistan Constitution, L. (2021, April). Freedom of expression. *Pakistan constitution law.com*.
- Pakistan Penal, c. (2013). Pakistan Penal Code (Act XLV of 1860). (n.d.). *Pakistan.org*.
- PIPS. (2019). Pakistan's Sectarian Mire and The Way Forward. *Pak Institute for Peace Studies (PIPS)*.
- Rao, M. F. (2020). Hate speech and media information literacy in the digital age: A case study of 2018 elections in Pakistan. *Global Media Journal*, 18(34), 1-10.
- Riaz, S. (2010). Effects of New Media Technologies on Political Communication. *Journal of Political Studies*, 17(2). <https://link.gale.com/apps/doc/A253609193/AONE?u=googlescholar&sid=bookmark-AONE&xid=f23b77d0>
- Ring, C. E. (2013). *Hate speech in social media: An exploration of the problem and its proposed solutions* (Doctoral dissertation, University of Colorado at Boulder).
- Rodgers, R., & Chabrol, H. (2009). Parental attitudes, body image disturbance and disordered eating amongst adolescents and young adults: A review. *European Eating Disorders Review*, 17(2), 137–151. <https://doi.org/10.1002/erv.907>
- Sengupta, S. (2012). *On the Web, a fine line on free speech across the globe*. New York Times.

- Silva, L., Mondal, M., Correa, D., Benevenuto, F., & Weber, I. (2021). Analyzing the Targets of Hate in Online Social Media. *Proceedings of the International AAAI Conference on Web and Social Media*, 10(1), 687–690. <https://doi.org/10.1609/icwsm.v10i1.14811>
- Simon Wiesenthal center. (2012). *Simon Wiesenthal Center hate report 2012*. California: Simon Wiesenthal Center.
- SPL. (2021). *Hate and Extremism*. South Asia: SPLCenter.org.
- Stephen, C. (2017). *Controversial Film Sparks Protests and Violence Across the Muslim World*. Guardian.
- Strauss, A., & Corbin, J. M. (1997). *Grounded theory in practice*. Sage.
- Tariq, R. (2019). Hate speech or healthy speech, user friendly or user anxiety: An analysis of cyber world speech discourse. *Journal of Media Studies*, 31(1).