

Digital Content Regulations: A Select Study in Indian Context

Thesis

submitted in partial fulfillment of the requirements for the degree
of

DOCTOR OF PHILOSOPHY

by

NAGANNA CHETTY



**SCHOOL OF MANAGEMENT
NATIONAL INSTITUTE OF TECHNOLOGY
KARNATAKA,
SURATHKAL, MANGALORE - 575025**

MARCH, 2022

Digital Content Regulations: A Select Study in Indian Context

Thesis

submitted in partial fulfillment of the requirements for the degree
of

DOCTOR OF PHILOSOPHY

by

NAGANNA CHETTY

(165094SM16F03)

Under the guidance of

Dr. Sreejith A



**SCHOOL OF MANAGEMENT
NATIONAL INSTITUTE OF TECHNOLOGY
KARNATAKA,**

SURATHKAL, MANGALORE – 575025

MARCH, 2022

DECLARATION

I hereby *declare* that the Research Thesis entitled “**DIGITAL CONTENT REGULATIONS: A SELECT STUDY IN INDIAN CONTEXT**” which is being submitted to the **National Institute of Technology Karnataka, Surathkal** in partial fulfilment of the requirements for the award of the Degree of **Doctor of Philosophy in Management** is a *bonafide report of the research work carried out by me*. The material contained in this thesis has not been submitted to any University or Institution for the award of any degree.



Mr. Naganna Chetty
Reg. No.: 165094SM16F03
School of Management

Place: NITK, Surathkal

Date: 08-03-2022

CERTIFICATE

This is to *certify* that the Research Thesis entitled “**DIGITAL CONTENT REGULATIONS: A SELECT STUDY IN INDIAN CONTEXT**” submitted by Mr. Naganna Chetty (Register Number: 165094SM16F03) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of Doctor of Philosophy.



Dr. Sreejith A
Research Guide
Assistant Professor
School of Management
NITK, Surathkal - 575025



Chairman, DRPC
(Signature with Date and Seal)

Rajesh Acharya H.
Head, School of Management
National Institute of Technology Karnataka
Post Srinivasnagar, Surathkal D.K.-575025

DEDICATION

This thesis is dedicated to the memory of my father

Somaraya Chetty

ACKNOWLEDGEMENT

The thesis is completed with the support and cooperation from multiple people along the way.

I express my heartfelt gratitude to my research advisor Dr. Sreejith A, Assistant Professor, School of Management, National Institute of Technology Karnataka (NITK), Surathkal for his supportive guidance during this research work. I will be highly obliged to him for his situational motivation and mentoring.

I thank my research progress assessment committee members, Dr. Beneesh P. B., Assistant Professor, Department of Chemistry, NITK, Surathkal, Dr. Suprabha K. R., Assistant Professor, School of Management, NITK, Surathkal, Dr. Sheena, Associate Professor, School of Management, NITK, Surathkal and our former DRPC secretaries Dr. Savita Bhat, Assistant Professor, School of Management, NITK, Surathkal and Dr. Gopalakrishna B. V., Assistant Professor, School of Management, NITK, Surathkal for their guidance and support during research work. I also thank all the Faculty Members of the School of Management, NITK Surathkal, for their cooperation.

I am immensely grateful to Dr. Rajesh Acharya H., Head & Associate Professor, School of Management, NITK Surathkal, for his support in the successful completion of this research work. I am also grateful to our previous heads of the department Prof. Aloysius Henry Sequeira, Prof. K. B. Kiran and Dr. S. Pavan Kumar, School of Management, NITK, Surathkal for their support during this research work.

I would like to thank Dr. Rajesh R. Pai, Vanitha P. S., and other research colleagues for their cooperation during the research work. I extend my gratitude to my Teachers, Colleagues, and Friends for their support, and motivation. I also thank the Students and social media users for their participation in research-related surveys. I thank the Reviewers, Editors, and Publishers for their comments and timely publication of the research articles. My special thanks to the non-teaching Staff of NITK Surathkal for their academic, administrative, and official assistance throughout the research work.

I am indebted to my wife Dr. Nagamma Patil, son Neil Chetty, and daughter Saanvi Chetty for their patience and cooperation throughout the research work. I express my

heartfelt thanks to my Mother, Brother, Sisters, Father-in-law and all Family Members for their good wishes towards the completion of this research.

I am very much grateful to my Father, Grandmothers and Mother-in-law, who could not see this time of success but showered their heavenly blessings on me for the successful completion of this research work.

Above all, I would like to thank God Almighty for the blessings bestowed and for giving me the strength and wisdom to reach this milestone in my life. I will always be grateful to you. Thank you.

Naganna Chetty

EXECUTIVE SUMMARY

Hate contents are the digital data that is harmful due to its inciting nature towards the protected characteristics. In the Indian context, the need for research on social media and citizens' e-participation with a focus on hate content reduction is often reported. This is because society and nations get affected by the hate content generation and propagation. Hate content at the workplace affects both the mental and physical health of the victims (employees), productivity, and financial status of the organization. Therefore, this research tries to address issues around online hate content with the objectives to (a) understand the nature and forms of digital hate content in the Indian context, (b) understand the influence of digital hate content in online safety, (c) understand policies and regulations addressing digital hate content, and (d) suggest recommendations for improving digital hate content regulation in India.

Two different quantitative studies are conducted to understand the ecosystem of online hate content and the role of organizations in hate content mitigation. Survey data is collected through both online and offline modes from appropriate respondents. The data is analyzed for respondents' views about the nature of online hate content and its mitigation separately. The results indicate that alienated, inclination, and personal experience contribute to gender hate content, which in turn contribute to online hate content. The religious hate content contributes to online hate content with the support from the stance of superiority, ignorance, religionizing the nation, and the clothing style contributing variables. The inherited nature, peer pressure, stereotype, unfamiliarity of ethnicity, personal experience, skin color, and poverty contributes to racist hate content, which in turn lead to online hate content. Similarly, disability hate content contributes to online hate content with the support of ignorance, envious, the fright of health, and able-bodied contributing variables. The analysis results towards the mitigation of hate content reveal the role of government and non-government authorities for digital hate content reduction. The non-governmental communities- civil societies, private sectors, and intermediaries are more important to reduce digital hate content.

Multiple case studies are conducted using social media data. The case study on related incidents of race, gender, religion, and disability indicated the existence of racism, gender hate, religious hate, and disability hate content. Another case study on an

ideological incident reported that the terrorist act is one of the causes of online hate content.

This thesis contributes to the development of a mixed-method design involving the quantitative data (questionnaire survey) and qualitative data (Twitter data) to answer a research question “*How to improve digital hate content regulations in the Indian context?*”. The identification of race, gender, religion, and disability as the drivers of online hate content contributes to theory.

TABLE OF CONTENTS

EXECUTIVE SUMMARY	i
LIST OF FIGURES	vii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	xi
CHAPTER 1 INTRODUCTION	
1.1 Introduction	1
1.2 Digital Content and Importance of its Regulations	7
1.3 Governance Framework for Digital Content	10
1.3.1 International Frameworks	11
1.3.2 National Frameworks	13
1.4 India's Digital Environment and Hate Content Control Strategies	16
1.5 Select indices of India's Hate Content	17
1.6 Justification for Study in Indian Context	22
1.7 Motivation for Research	23
1.8 Outline of the Study	23
1.9 Organization of the Thesis	24
1.10 Concluding Remarks	24
CHAPTER 2 LITERATURE REVIEW	
2.1 Introduction	27
2.2 Related Concepts of Hate Content and Governance	27
2.3 Models of Hate Content	29
2.3.1 Gender-based Hate Content	29
2.3.2 Religious Hate Content	31
2.3.3 Racist Hate Content	34
2.3.4 Disability Hate Content	37
2.4 Hate Content Detection Techniques	39

2.5	An Act of Violence and Governance Policies for Hate Content	45
2.6	Research Gaps	48
2.7	Concluding Remarks	49
CHAPTER 3 RESEARCH DESIGN		
3.1	Introduction	51
3.2	Research Question	51
3.3	Research Objectives	51
3.4	Conceptual Framework and Hypotheses	52
3.5	Context and Scope of the Study	55
3.6	Research Methodology	55
3.6.1	Justification for Mixed methods Approach	57
3.6.2	Questionnaire Design and Administration	58
3.6.3	Justification for Studies using Social Media Data	61
3.6.4	Case Studies	62
3.6.4.1	Case Study on Gender-based Hate Content	62
3.6.4.2	Case Study on Religious Hate Content	63
3.6.4.3	Case Study on Racism	63
3.6.4.4	Case Study on Influence of Physical Violence on Hate Content	64
3.7	Concluding Remarks	64
CHAPTER 4 ONLINE HATE CONTENT: A SURVEY AND ANALYSIS		
4.1	Introduction	65
4.2	Online Hate Content: Nature, Forms and Hypotheses	65
4.2.1	Research Model	66
4.2.2	Reliability and Validity of the Survey Instrument	70
4.2.3	Sample Characteristics	70
4.2.4	Survey Sample and Justification	73

4.2.5	Results and Discussions	74
4.3	Analysis for Policy Perspectives	84
4.3.1	Model and Methodology	84
4.3.2	Results and Discussion	85
4.4	Concluding Remarks	88
CHAPTER 5 CASE STUDIES		
5.1	Introduction	89
5.2	Case of Gender based Online Hate Content	89
5.2.1	Case Context	90
5.2.2	Methodology	91
5.2.3	Results and Discussion	93
5.3	Case of Religious Online Hate Content	102
5.3.1	Case Context	103
5.3.2	Methodology	104
5.3.3	Results and Discussion	104
5.4	Case of Racist Online Hate Content	109
5.4.1	Case Context	109
5.4.2	Methodology	111
5.4.3	Results and Discussion	112
5.5	Case of Physical Violence on Online Hate Content	118
5.5.1	Case Context	119
5.5.2	Methodology	119
5.5.3	Results and Discussion	120
5.6	Concluding Remarks	122
CHAPTER 6 SYNTHESIS, IMPLICATIONS AND RECOMMENDATIONS		
6.1	Introduction	125
6.2	Synthesis	125
6.2.1	Synthesis of Survey Findings	132
6.2.2	Synthesis of Gender and Religious cases	133

6.2.3	Synthesis of Racist and Physical Violence Cases	134
6.2.4	Synthesis of Survey and All Cases Findings	134
6.3	Implications	135
6.3.1	Implications for the Government and Online Service Providers	135
6.3.2	Implications for Citizens	138
6.3.3	Implications for Management	139
6.3.4	Theoretical Implications	140
6.4	Recommendations for Improving the Digital Content Governance	140
6.5	Concluding Remarks	144
 CHAPTER 7 CONCLUSION		
7.1	Introduction	145
7.2	Summary of Research	145
7.3	Revisiting Research Question and Objectives	153
7.3.1	Research Question Revisited	153
7.3.2	Research Objectives Revisited	156
7.4	Key Findings from the Synthesis of the Studies	159
7.5	Significant Research Contributions	160
7.6	Limitations and Scope of Future Works	162
7.7	Concluding Remarks	164
 REFERENCES		165
 APPENDIX		195
 BIO-DATA		203

LIST OF FIGURES

Figure No.	Name of Figure	Page No.
1.1	Violent activities in a tree structure	2
1.2	Online social networks for destructive activities	6
1.3	Stages of hate speech following a trigger event	7
3.1	A conceptual framework for hate content analysis	53
3.2	Research Methodology	56
4.1	A Conceptual Framework to identify determinants of hate content	67
4.2	Steps to identify determinants of hate content	68
4.3	Structural model	68
4.4	Measurement models	77
4.5	A structural model with weights	80
4.6	Path analysis model to identify determinants	82
4.7	A developed conceptual model for policy perspectives	86
5.1	A methodology to analyze social reform movements	92
5.2	Emotions on a government-initiated social reform movement	94
5.3	Wordcloud of emotions for a government-initiated social reform	95
5.4	Emotions on a citizen-initiated social reform movement	96
5.5	Wordcloud of emotions for a citizen-initiated social reform movement	98
5.6	Percentage of emotions for a government-initiated social reform movement	99
5.7	Percentage of emotions for a citizen-initiated social reform movement	99
5.8	Causal loop diagram of MeToo movement	101
5.9	Honor based and hate-based violence with different violent acts	107
5.10	Emotions for honor-based hate violence	108
5.11	A wordcloud with emotions for honor-based hate violence	108
5.12	Racism influencing factors	110
5.13	Datasets for racism analysis	111

5.14	An approach for racism analysis	112
5.15	Graph of racism analysis (Keywords: African or racism)	113
5.16	The score of racism (Keywords: African or racism)	113
5.17	Graph of racism analysis (Keywords: African and Racism or black)	114
5.18	The score of racism (Keywords: African and racism or black)	114
5.19	Graph of racism analysis (Keywords: Bollywood or racism or hate)	115
5.20	The score of racism (Keywords: Bollywood or racism or hate)	116
5.21	Graph of racism analysis (Keywords: Bollywood and racism or hate)	116
5.22	The score of racism (Keywords: Bollywood and racism or hate)	117
5.23	Emotions on the problematic incident	120
5.24	Wordcloud for emotions on the problematic incident	121
5.25	Percentage of emotions on the problematic incident	122
6.1	Proposed system architecture for digital hate content reduction	141

LIST OF TABLES

Table No.	Name of Table	Page No.
1.1	Fake news vs. hate content	9
1.2	Evolution of hate content in India	9
1.3	International legal frameworks for hate content	12
1.4	Select a few countries' legal frameworks for hate content	14
1.5	The laws around hate content	15
1.6	Hate crimes in India	18
1.7	Select societal and information and communication technology indices for India	19
2.1	Concepts of hate and governance	27
2.2	Comparison of works on religious hate speech	32
2.3	Literature for identified variables	38
2.4	Comparison of works on hate content detection	42
2.5	Major findings from previous works (Category wise)	43
2.6	Listing of research gaps	48
3.1	Basic constructs of a questionnaire	59
4.1	List of measurement variables used in the study	69
4.2	Structural model in square matrix format	69
4.3	Reliability testing of the survey instrument	70
4.4	Gender and age of respondents	71
4.5	State-wise respondents	71
4.6	Tools and Internet used by the respondents	72
4.7	Respondents response to hate content	73
4.8	Factors and unidimensionality	74
4.9	Measurement model evaluation factors	75
4.10	Coefficient of determination	78
4.11	Correlations between the latent variables	79
4.12	Results of path analysis	80
4.13	Total effects of the relationships	81

4.14	Derived relationships from measurement and structural models	83
4.15	Factors and unidimensionality for policy perspectives	85
4.16	Coefficient of determination for policy perspectives	87
4.17	Path analysis results for policy perspectives	87
5.1	Government-initiated movement vs. citizen-initiated movement	100
5.2	Keyword wise analysis of tweets with the percentage	118
6.1	Variables and practices identified from social media data analysis	128
6.2	Synthesis of survey findings	132
6.3	Synthesis of gender-based and religion-based case findings	133
6.4	Synthesis of racist and physical violence incidents case findings	134
6.5	Integration of findings	135

LIST OF ABBREVIATIONS

4GLTE:	Fourth Generation Long Term Evolution
API:	Application Programming Interface
AASMA:	Advanced Application for Social Media Analytics
BHIM	Bharat Interface for Money
DHC:	Disability Hate Content
ECHR:	European Convention on Human Rights
FIR:	First Investigation Report
GCI:	Global Cybersecurity Index
GHC:	Gender Hate Content
GoF:	Goodness of Fit
GPS:	Global Positioning System
H1:	Hypothesis1
H2:	Hypothesis2
H3:	Hypothesis3
H4:	Hypothesis4
H5:	Hypothesis5
H6:	Hypothesis6
HBV:	Honor-Based Violence
HCR:	Hate Content Reduction
ICCPR:	International Covenant on Civil and Political Rights
ICERD:	International Convention on the Elimination of all forms of Racial Discrimination
ICERT:	Indian Computer Emergency Response Team
ICTs:	Internet Communication Technologies
IPC:	Indian Penal Code
ISPs:	Internet Service Providers
IT:	Information Technology
ITU:	International Telecommunications Union
MANDOLA:	Monitoring and Detecting OnLine Hate Speech
NGOs	Non-Governmental Organisations

NRC:	National Research Council
OHC:	Online Hate content
PLS-PM:	Partial Least Square-Path Modeling
PLS-SEM:	Partial Least Square SEM
RaHC:	Racist Hate Content
ReHC:	Religious Hate Content
SEM:	Structural Equation Modeling
SHI:	Social Hostilities Index
SMART:	Simple, Moral, Accountable, Responsive and Transparent
SPSS:	Statistical Package for the Social Sciences
UAE:	United Arab Emirates
UDHR:	Universal Declaration of Human Rights
UK:	United Kingdom
URL:	Uniform Resource Locator
USA:	United States of America
VIF:	Variance Inflation Factor

CHAPTER 1

INTRODUCTION

1.1 Introduction

Digital content is any information available for download or distribution on electronic media such as an ebook, audio, video, graphics, animations, images, online conversations, and cyberhate (Parajuli, 2007). The existence of Internet Technologies, online social media, and the easy availability of computing devices increased the growth of digital content. Digital content is getting important as it can be recorded and used as a complement to other forms of information to support decision making on related incidents (Allegra et al., 2011; Wang, 2007).

Advances in Internet Technologies (ITs) and online social networks have made more benefits to humanity. At the same time, the dark side of this growth/benefit has increased hate content and terrorism as the most common and powerful threats globally. The current study aims to research online hate content. Therefore, hate and other violent acts are briefed. Hate is a dislike feeling about someone or something. Hate intends to devalue others through diminishing or destroying their well-being (Rempel and Burris, 2005). The triangular theory of hate is based on the “negation of intimacy, passion, and commitment” (Rempel and Burris, 2005; Sternberg, 2003, p. 307). According to the dictionary, the meaning of hate is “to have a strong dislike or ill will for; loathe; despise”.

Hate is one of the antecedents of violent activities such as massacres and genocides (Sternberg, 2003). Figure 1.1 shows violent activities with tree structures. For simplicity and conciseness of the report, only a few activities are discussed here. The relationship such as *is a / kind* nature moves from bottom to top of the tree.

Terrorism is a global phenomenon that results in the loss of innocent lives and public properties on a larger scale. Two main objectives of terrorism are creating terror in the minds of targeted victims and attracting media and world power. Terrorism presents a threat to humanity in common, without differentiating between race, gender, religion, and nationality. It is an international problem by challenging communities of the entire world. Emanuel Gross said that “The majority of the definitions have a common basis - terrorism is the use of violence and the imposition of fear to achieve a particular purpose” (Gross, 2001, p. 97).

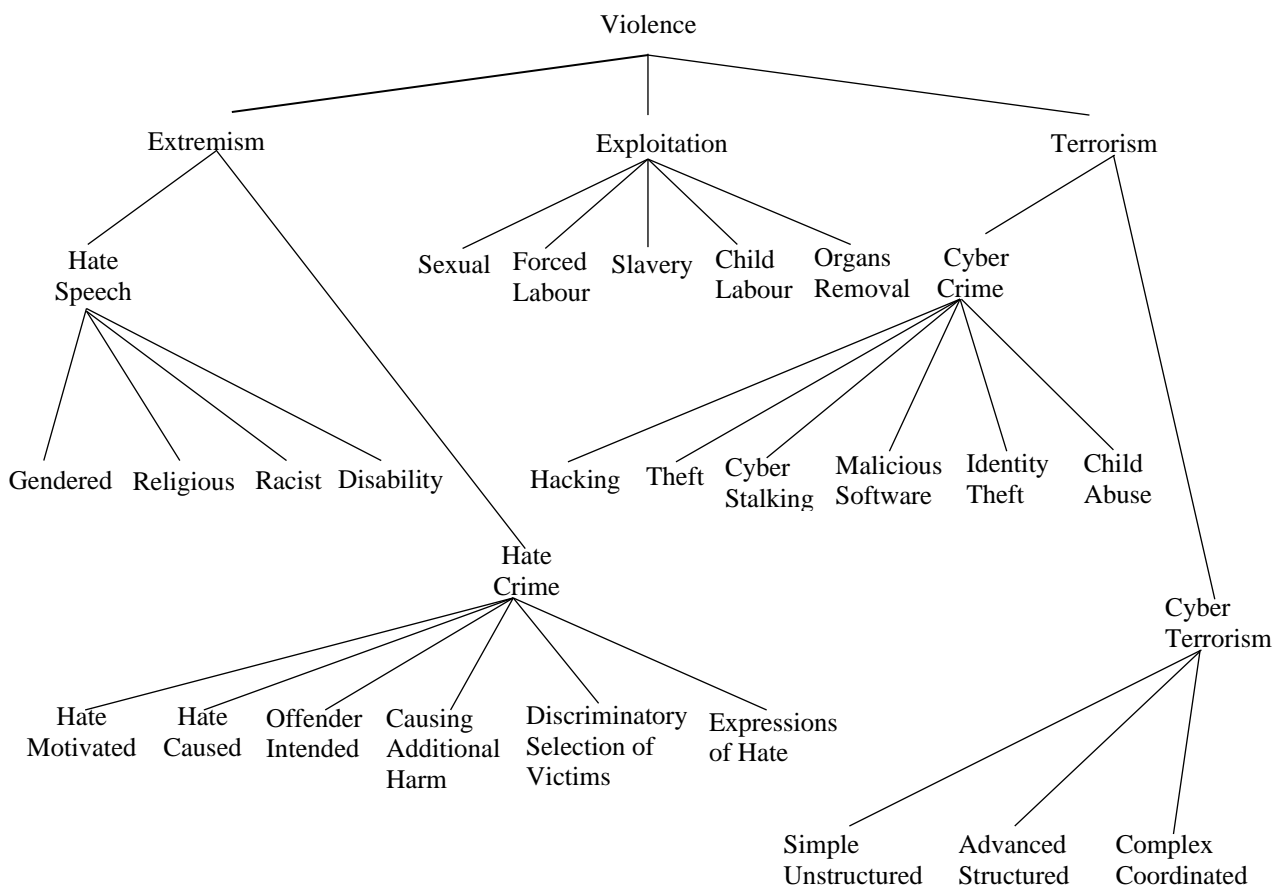


Figure 1.1: Violent activities in a tree structure

Cyberterrorism is a novel approach to damage the victims of the attack. It makes use of computers and related technologies to attack the target. Cyberterrorism activities are very common due to a lack of international resolve. Technological tools like social networks and related websites help terrorist groups to increase and improve their terrorist activities by exchanging harmful information. Cyberterrorism is defined by Denning as: “cyberterrorism is the convergence of terrorism and cyberspace. It is generally understood to mean unlawful attacks and threats of attacks against computers, networks, and the information stored therein when done to intimidate or coerce a government or its people in furtherance of political and social objectives” (Denning, 2000, p. 1). From the views of Denning, the impact of cyberterrorism is treated as similar to traditional terrorism.

Extremism is a political and religious ideology with opposition towards societal norms and its nature is almost the same as terrorism. According to Liebman, extremism is “a desire to expand the scope, detail and strictness of religious law, social isolation and

the rejection of the surrounding culture” (Liebman, 1983, p. 75). In some countries, the violence of creating terror is called extremism. Its extreme political and religious views lead towards hate speech and hate crimes.

Hate crime is a hate-motivated physical attack on a person, property, or group concerning the identities like gender, race, religion, nation, and disability. Walters et al. (2016, p. 11) argued that “the criminal offense, which is perceived by any person, motivated by hostility and prejudice is called hate crime”. Hate crimes take away the victim’s civil rights. It is a kind of extremist crime and punishable by the constitutional law of each country, whereas hate speech is a verbal attack and not punishable easily with the legal framework.

Exploitation is an act of treating others by an unfair means to get benefit from them (Sample, 2003). This is one of the common harm to society and is practiced by unethical people. The harassments such as sexual, forced, and child labor, slavery, and organ removal are the different forms of exploitation (Cockbain and Bowers, 2019; Joffres et al., 2008).

The concept of hate is contradictory. There is no universal definition of hate. Some researchers argue that hate is based on motive (Rempel & Burris, 2005), emotion (Opotow & McClelland, 2007), and attitude (Brudholm, 2010). The conflicts between the different groups are the impact of a variety of emotions. Hate has multiple dimensions such as disgust, anger/fear, and devaluation/diminution and that may take different forms based on the situation (Sternberg, 2003). For intergroup conflicts, anger is a common and powerful emotion (Halperin & Gross, 2011; Bar-Tal 2007). As anger possesses aggressive behavior, sometimes it is assumed as a destructive force for the occurrence of conflicts between the groups (Halperin et al., 2011).

Hate is bias-motivated and may act against the member of a protected group such as gender, religion, race, and disability (Jay, 2009; Masucci & Langton, 2017). Hate speech/content is protected or allowed until its impact turns into a threat. Victims of hate content feel threatened by their victimizers (Jay, 2009). Intimidating content may not be accepted by the law and may form a biased crime. The bias-motivated crimes may be restricted by the criminal law with more penalties than non-bias motivated crimes (Dharmapala & Garoupa, 2004; Lawrence, 1994).

Some content may be offensive but not harmful (Hargrave et al., 2009). There is no universal statement to indicate a speech/content expressed as offensive in nature (Jay & Janschewitz, 2008). The meaning and impact of expressed content are situational, depend on how persons are related to each other and the content of the topic. Exchange of contents like nigger, cunt, or fuck between the couples is non-offensive at an appropriate place and time (Jay, 2009).

The impact of hate content on victims may be direct or indirect. Direct hate content leads to an immediate effect on victims whereas indirect hate content leads to delayed or immediate harm. Agents as perpetrators induce delayed harm instead of an original actor. As an example, the expression of hate content against any identity during public events may make others act accordingly in the future (Seglow, 2016).

Bigotry is a basic common factor associated with all hate offending groups (McDevitt et al., 2002) and acts as a principal motivating factor for the occurrence of hate. There are four motives such as thrill, defensive, mission, and retaliatory behind the hate (McDevitt et al., 2002). Hatred is exhibited with a thrill motive to indicate the desire for excitement and power, with a defensive motive to indicate protection against threats, with a mission motive to indicate destruction of evilness, and with a retaliatory motive to indicate avenge on perceived assault of their group.

Hate speech affects a portion of society, some people suffer from it and some enjoy it without sympathy. It targets mostly the minority groups to exhibit an opposing behavior on them (Benesch, 2014). The supremacist organizations believe that their superiority is natural and while making hatred statements will forget that the minority groups also have equal rights (McVeigh, 2004). When hate speech targets traditional identities results in severe effects (Tseis, 2009). Criticism of an individual concerning the protected characteristics has a greater impact than criticism on the personal information of an individual (Tseis, 2009; Bhandari and Bhatt, 2012).

Cohen-Almagor (2011, p. 1) has defined hate content “as bias-motivated, hostile, malicious speech aimed at a person or a group of people because of some of their actual or perceived innate characteristics”. The European Court of Human Rights adopted a definition on hate content as “all forms of expression which spread, incite, promote or justify racial hatred, xenophobia, anti-Semitism or other forms of hatred based on

intolerance, including intolerance expressed by aggressive nationalism and ethnocentrism, discrimination and hostility towards minorities, migrants and people of immigrant origin” (Council of Europe, 1997, p. 107).

With this concept, hate content is defined as “content, which attacks an individual or a group intending to hurt or disrespect based on the identity of a person”. Once the hate content is expressed, hurting or disrespecting depends on the perception of the victim. For some people, it may or may not affect them. Generally, the effect of hate content depends on the originator, its content, and the targeted one.

If hate content does not incite discrimination (do not hurt the targeted one), then, there arises a question that whether this kind of content is hatred or not? Here it is accepted as hate content because of the intention and its content. For clarity consider a legal framework, in which an attempt to murder is treated as a crime, the accused will be penalized and the victim will be provided more protection. Here purpose and action performed by the murderer are counted. Similar ideology can be applied in the context of hate content.

Generation of hate content has become a trend and people are using this as a shortcut way to get instant popularity without putting in more effort. Hate content creates a situation to test the limits of free speech. Hate content is handled by different regulations from different countries (Kasakowskij et al., 2020). Hate content usually opposes freedom of speech and violates the fundamental rights of a human being. The broader goal of the freedom of expression is to help an individual to attain self-fulfillment, assist in the discovery of truth and strengthen the capacity of an individual by participating in decision making, provide a mechanism by which it would be possible to establish a reasonable balance between stability and social change. It also allows all members of the society to form their own beliefs and communicate them freely to others (Bhandari and Bhatt, 2012). Hate content will act as an obstacle to these goals.

Online social networks are a special form of social network and help to establish the relationship among users of the networks globally (Cheung et al., 2011). These networks are one of the most important points of growth for the Internet. Traditionally online social networks are meant for maintaining the existing relationship, enhancing

the existing relationship, and creating a new relationship based on common interests. Nowadays these are used as a rich set of the database for decision making and as a media for communication (Haythornthwaite, 2005). As a communication media, these can be used for generating and spreading healthy and unhealthy information among connected users. A small percentage of users use a portion of the networks for unhealthy activities such as hate content and terrorism but the impact of this small percentage of users is more and harmful.

Figure 1.2 shows the role of online social networks for destructive activities such as hate speech, hate crime, extremism, and terrorism. Hate speech is made spreadable by posting a message, reposting a message, and responding to a message on social networks. Hate crime is a hate-motivated physical attack and social networks are used for planning and executing the attack-related activities. Extremists and terrorists use social networks for contacting and recruiting like-minded persons, spreading propaganda, planning and executing the attacks.

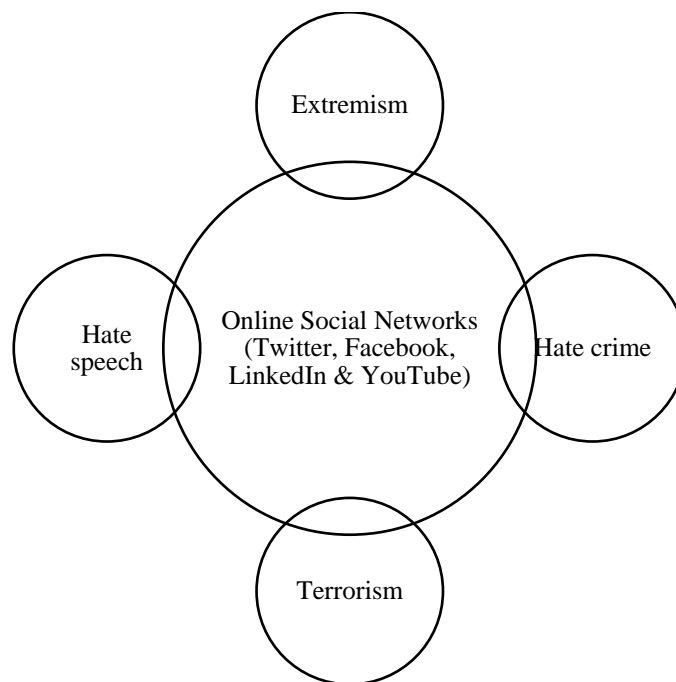


Figure 1.2: Online social networks for destructive activities

Different stages of hate speech following a trigger event are shown in Figure 1.3. Hate speech, immediately after the event (influence stage) will flow heavily on social networks, after few days (intervention stage) will get reduced, after some more days (response stage) reduces to zero levels and after a long time once again it may appear.

This indicates that after a particular event, people will be more excited and gradually will get a normal state or behavior. The rebirth stage is shown with a dashed line to indicate an optional stage. Based on the type and impact of an event, the hate speech may or may not appear once again after a long time.

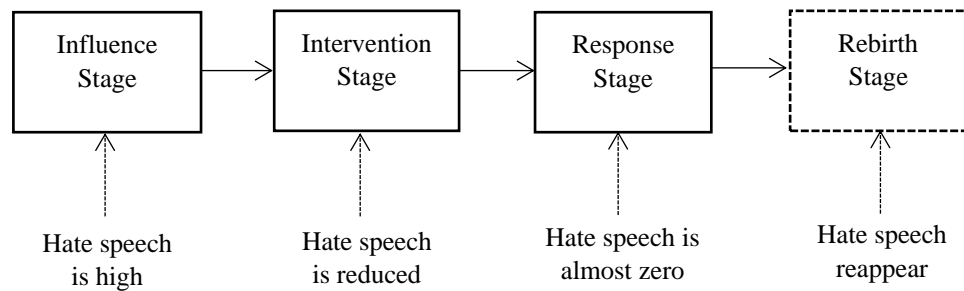


Figure 1.3: Stages of hate speech following a trigger event

1.2 Digital Content and Importance of its Regulations

Digital content is a kind of data available from electronic devices. Digital content is available for broadcasting, streaming, and storing in electronic storage. Rowley is defined digital content as “bit-based objects distributed through electronic channels” (Rowley, 2008, P. 517). Anyone who is connected with the internet or electronic devices may create digital content. The delivery of digital content to the audience is important than its creation. The selection of an appropriate content delivery approach depends on the requirement and control of the digital content. Based on the purpose of the content, the availability of digital content may be free or paid.

Regulations are the legal landscape (Cohen, 2014). The digital content regulations are measures to ensure the appropriate use of digital content. An increase in the access of online content, particularly the use of social networking sites by individuals and corporate bodies leads to draw a legal framework and some of the key issues that affect users of social media. The regulations can be framed jointly by experts in technology and policymaking. Regulation is a major subset of governance and deals with the control of the flow of events and behavior, as opposed to providing and distributing. Governance is about providing, distributing, and regulating (Braithwaite et al., 2007).

The people think that digital convergence ends the ownership and the regulation of digital content by blurring the boundaries between the different sectors of communication (Iosifidis, 2002). Often, the digital content which is generated and

distributed through social media platforms incites violence in society (Patton et al., 2013). The governance of online hate content by social media platforms is more oriented towards technological solutions than political and judicial (Siapera and Viejo-Otero, 2021). Online hate content could be controlled by converged efforts from different sectors such as politics, the judiciary, and technology.

Fake news has got more attraction from recent incidents such as 2016 the United States presidential election (Haciyakupoglu, 2018). The assessment from the National Intelligence Council of the United States in January 2017 revealed that the outside Government and non-governmental agencies played a major role in influencing the United States presidential election through campaigns (ICA, 2017). The news is a major source of exact information, which informs about ongoing global affairs and plays a constructive role in society. When it is fake, misinforms the citizen and acts as a destructive mechanism rather than constructive. Fake news is a kind of information, which is intentionally created and propagated to make others believe falsehoods or doubt actuals.

Fake news is a very much exciting term in the era of post-truth, the era in which we live now. Fake news fuels destructive activities such as propaganda hate speech and violence (McGonagle, 2017). Fake news has greater impacts on third persons than the creator themselves and the groups they belong (Jang and Kim, 2018). Fake and hate are closely related. If you hate something, hatred can be represented by creating fake information (Farkas and Neumayer, 2017). The Pew Research Center, a popular public opinion assessor, observed that uncivil and manipulative behaviors on online platforms would persist and maybe worst in the future. Terrorists and politicians are benefitting from online activities such as bot-based misinformation/fake news or any other persuasion tactics (Rainie et al., 2017). A comparison of fake news and hate speech is made in Table 1.1.

In India, the administrators of social media groups are made accountable for the dissemination of fake news. Recently, one of the district magistrates in India issued an order highlighting the first investigation report (FIR) can be filed against the administrator of social media group if fake news exists/circulate in that group (Haciyakupoglu, 2018).

Table 1.1: Fake news vs. hate content

Factor	Fake news	Hate content
Objective	Conveys wrong information to others	Exhibits hatred against others
Popularity	Since 2016	Since long back
Destructive acts	Support for violence, terrorism, hate speech, etc.	Support for violence, terrorism, etc.
Impact is on	Third-person	Second person
Negative impact	Indirect	Direct/indirect
Relationship	Cause	Effect

The existence of hate content in India becomes apparent during the British rule before 1947. As a result of the division made by the British rule, after the independence of India, the expression of hate content start flourishing. Table 1.2 shows the evolution of hate content in India. The hate, particularly religious hate originated during British rule evolved by organizations and political parties against the religious minority communities. Even the mode of expressing hate is evolved from offline to online. Often, the content is available in both online and offline modes.

Table 1.2: Evolution of hate content in India

Period	Authors	Statement on hate	Remarks
Before 1947	(Hartnack, 2012; Alam et al., 2016)	The concept of religious hate has been originated during the British rule in India as a result of their divide and rule strategy to control citizen	Division of people initiated hatred among them
1947-2000	(Kapur, 1996)	Communities of minority religions are discriminated during 1992-1994	Religious hatred flourished
	(Cossman & Kapur, 1997)	Religious hatred by organizations and political parties to gain votes	
2001-2010	(Sundar, 2004)	Often, some private schools emphasize and teach more on nationhood	Extreme nationhood and inter-religious marriages lead to hatred
	(Lankala, 2006).	Hate speech and systematic killing of minorities	
	(Dhavan, 2007)	Artists are targeted by assuming that they had hurt religious sentiments	
	(Gupta, 2009),	Campaigns to foster hate against inter-religious marriages are organized by some groups	
2011-2018	(Goolsby et al., 2013)	Shifting of hatred activities become apparent on social media	The transition from offline to online hatred
	(Mirchandani, 2018)	Digital hatred: Hate flourished on social media	

In India, the majority of youths are involved with the virtual environment. Social media may be used by the youths to share their emotions such as love, hate, aggression, and violence in digital form (Tripathi, 2017). Therefore, regulating digital content is important to control the dissipation of toxic content over social media platforms.

1.3 Governance Framework for Digital Content

The concept of governance denotes the use of political authority and exercise of control in society about the management of its resources for social and economic development. This broad definition encompasses the “role of public authorities in establishing the environment in which economic operators function for determining the distribution of benefits as well as the nature of the relationship between the ruler and the ruled” (Weiss 2000). With the help of electronics or digital mechanisms, the performance of the governance can be improved.

E-governance is the performance of the government to ensure a process of service delivery. It is often defined as the application of Internet Communication Technologies (ICTs) to the Government processes for optimizing the delivery of services and bringing Simple, Moral, Accountable, Responsive, and Transparent (SMART) governance and the government (Reji, 2021). In general, governance mainly deals with decision-making. Good governance depends on knowledge and its recognition by the decision-makers. This knowledge is digitized over a network connecting every individual and the decision-makers of the organization to form digital governance. Governance is fundamental and digital governance is instrumental.

The goal of digital governance is to provide equal rights to common citizens in decision-making processes related to their lives with quality. This indicates that citizens are not just passive consumers of services, they have the right to participate actively in deciding the kinds of services they desire.

Freedom of speech is considered one of the important freedoms globally. As a part of the legal frameworks, some of the commonly acceptable activities related to expressions like free and hate content by national and international bodies are discussed. The legal frameworks contain a set of rules to permit or prohibit activities or ideas based on their nature.

1.3.1 International Frameworks

The legal information on hate speech can be found by accessing international human rights law with internationally accepted declarations and conventions supporting fundamental rights to every human being. Article 19 from the Universal Declaration of Human Rights (UDHR) states that “Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive and impart information and ideas through any networks and regardless of frontiers”.

The whole universe is agreed upon the freedom of expression. To make effective and appropriate use of freedom of speech, article 29(2) of the UDHR states that, “In the exercise of his rights and freedoms, everyone shall be subject only to such limitations as are determined by law solely to secure due recognition and respect for the rights and freedoms of others and of meeting the just requirements of morality, public order and the general welfare in a democratic society.” It opposes the use of text, content, theory, and practice of free speech as liberty of an individual in modern societies. Similarly, other international bodies stated their views on free speech rights and/or hate speech restrictions in the form of articles. A summary of articles from international bodies for free speech rights and/or hate speech restrictions are outlined in Table 1.3 ([UDHR, 1948](#); [ECHR, 1950](#); [ICCPR, 1976](#); [ICERD, 1969](#)).

The statements made by 3 international communities UDHR, European Convention on Human Rights (ECHR), and International Covenant on Civil and Political Rights (ICCPR) on free speech rights are almost the same. International Convention on the Elimination of all Forms of Racial Discrimination (ICERD) does not state any article on free speech rights.

UDHR stated the minimum and general restrictions on hate speech. ECHR refines the conditions and expands the restrictions to cover more phrases on anti-hate speech as a threat to national security, territorial integrity, the information disclosing against the confidentiality, preserving impartiality and judicial authority. Initially, ICCPR is phrased with minimal terms on hate speech and later a paragraph is added to cover more on hate speech. Added paragraph prohibits propaganda for war and hatred advocacy on nationality basis, racism or religion.

Table 1.3: International legal frameworks for hate content

S. No.	International Body	Adaptation Year	Free speech Article No.	Rights on Free Speech	Hate speech Article No.	Restrictions on Hate Speech
1	UDHR	1948	19	<ul style="list-style-type: none"> • Hold opinions • Seek information and ideas • Receive information and ideas • Impart information and ideas 	29(2)	<ul style="list-style-type: none"> • Law determines the restrictions • Restrictions are meant to respect the rights and freedom of others • Restrictions are to attain morality, public order, and the welfare of society.
2	ECHR	1950	10(1)	<ul style="list-style-type: none"> • Hold opinions • Receive information and ideas • Impart information and ideas 	10(2)	<ul style="list-style-type: none"> • Retaining national security, territorial integrity, or public safety • Prevention of disorder or Crime • Protection of health or morals • Protection of the reputation or rights of others • Preventing the disclosure of information received in confidence • Maintaining the authority and impartiality of the judiciary
3	ICCPR	1976	19(2)	<ul style="list-style-type: none"> • Seek information and ideas • Receive information and ideas • Impart information and ideas 	19(3)	<ul style="list-style-type: none"> • To maintain respect for the rights or reputations of others • Protection of national security or public order or public health or morals
					20(1)	<ul style="list-style-type: none"> • Prohibit propaganda for war
					20(2)	<ul style="list-style-type: none"> • Prohibit advocacy of national, racial, or religious hatred
4	ICERD	1969	--	--	4	<ul style="list-style-type: none"> • Condemn propagandas based on ideas of superiority of one race or group of persons of one color or ethnic origin • Condemn the attempt to justify or promote racial hatred and discrimination in any form • Undertake to adopt immediate and positive measures designed to eradicate all incitement to, or acts of such discrimination.

ICERD stated more on an anti-hate speech by prohibiting the ideas disseminated with racial superiority, whether this dissemination was likely to lead towards violence or hostility or not. The discussion on legal frameworks of international bodies shows that the views of all the treaties are almost the same with some added restrictions on hate speech by ICERD.

1.3.2 National Frameworks

Apart from the international standards to control hate speech, it is also essential to have national laws to combat hate speech. The constitutional and penal code laws of few countries to combat hate speech are discussed in this section.

Hate speech laws in India aim to avoid conflicts among the various religions in the country. These laws lead towards a punishment when a citizen unrespect others based on race, religion gender, disability, language, occupation, or any other identity. The laws of hate speech also obstruct the expressing mechanisms, which harms the citizen. Article 19 of the Indian constitution provides the right to every citizen on freedom of speech and expression with the constraints to preserve morality, public interest, or decorum ([Indian Penal Code, 1860](#); [Law Commission of India, 1971](#); [The Constitution of India, 2007](#)). Similarly hate speech laws of Canada ([Walker, 2013](#)), United Kingdom ([Public Order Act 1986](#); [Criminal Justice and Public Order Act 1994](#)), Poland ([The Constitution of the Republic of Poland, 199](#); [Penal Code of Poland, 1997](#)), United Arab Emirates ([UAE Anti-discriminatory Law, 2015](#)) and United States of America ([Ruane, 2014](#); [Office of General Counsel, 2009](#)) are referred and summarization is made as shown in Table 1.4.

In Canada, if a person justified his hatred expression as true with faith or in the public interest, then he will not be punished. Similarly, in the UK, a person will be punished if the hatred expression is threatening not just abusive or insulting. In Poland, if a person commits guilty with hatred expression, is liable for punishments through fine, imprisonment, or restriction of liberty. In the UAE hate speech law is implemented recently in 2015 as an obstacle to hate speech and promotion of violence. According to this law, an accused will be punished with a fine and jail term. The constitution of the USA emphasizes more free speech than hate speech and still, the hate speech laws are not clear.

Table 1.4: Select a few countries' legal frameworks for hate content

S. No.	Country	Constitutional role to control hate speech	Penal code role (Punishment) to control hate speech
1	India	<ul style="list-style-type: none"> • Article 25(1) states, all are having equal freedom and the right to freely profess, practice, and propagate religion. • Article 19 provides the right to freedom of speech and expression to all citizens with restrictions for preserving public order, decency, and morality. • According to article 28 conveying any religious message in educational institutions is prohibited. • Article 51A (h) imposes on every citizen the duty to develop the scientific temper, humanism, and the spirit of inquiry and reform. 	<ul style="list-style-type: none"> • According to section 153(A), from the Indian Penal Code (IPC), hatred expression against any identifiable group is accountable to punish by imprisonment of a maximum of 3 years or fine or both. • Section 295(A) from IPC says hatred expression exclusively on religion is accountable to punish with imprisonment up to 3 years or fine or both.
2	Canada	<ul style="list-style-type: none"> • Section 2 of the Charter, grants freedom of conscience and religion, thought, belief, opinion, and expression. • Section 1 restricts the granted freedoms by making them subject "only to such reasonable limits prescribed by law as can be demonstrably justified in a free and democratic society". 	<ul style="list-style-type: none"> • Section 318 states imprisonment up to 5 years for anyone who advocates genocide. • Section 319 mentions penalties as fine or imprisonment with a maximum of two years to the exhibitor of hatred towards any identifiable group. This section also nullifies the guiltiness of the accused, if the accused proves that the statements made were true in the public interest. • According to section 320, a judge can seize publications containing propaganda for hate.
3	United Kingdom (UK)	<ul style="list-style-type: none"> • Act 1986 prohibits racial hatred expression based on a group's color, nation, race, or ethnicity. • Section 4A inserted into the Public Order Act 1986 by the Criminal Justice and Public Order Act 1994 prohibits anyone from causing alarm or distress. 	<ul style="list-style-type: none"> • Section 18 of the Act says a person is accountable to punish with imprisonment of a maximum of 7 years or fine or both if he threatens, abuses, or insults others. • Section 4A says a person to harass, distress, or alarm others is treated as guilty and will be punished with imprisonment up to 6 months or fine, or both.
4	Poland	<ul style="list-style-type: none"> • Article 54 of the Constitution protects freedom of speech. • Article 13 prohibits political parties and other organizations which have programs based upon totalitarian methods. It also prohibits any programs or activities which promote racial or national hatred. • Article 35 gives national and ethnic minorities the right to establish educational and cultural institutions and 	<ul style="list-style-type: none"> • Article 196 states that anyone offending religious feelings intentionally is accountable to fine, liberty restriction, or imprisonment up to 2 years. • Article 256 states that, if anyone inciting hatred concerning nationality, race, ethnicity, or religion is liable to fine or liberty restriction or imprisonment up to 2 years. • Article 257 states that, if anyone found guilty of insulting a group or an

		institutions designed to protect religious identity.	individual publicly is liable to fine, liberty restriction, or imprisonment up to 3 years.
5	United Arab Emirates (UAE)	<ul style="list-style-type: none"> The law, No. 02 of 2015, criminalizes any act that stokes religious hatred and/or insults religion through any form of expression, be it speech or written word, books, pamphlets, or via online media. 	<ul style="list-style-type: none"> The law states that the penalties for violating various provisions of it are imprisonment between 6 months and 10 years with a fine of Dirham 50,000 to 2 million. The law bans promotional activities on hate speech and punishes receiving financial support for such activities.
6	United States of America (USA)	<ul style="list-style-type: none"> Congress shall make no law respecting an establishment of religion, or prohibiting the free exercise thereof, or abridging the freedom of speech, or of the press, or the right of the people peaceably to assemble and to petition the Government for a redress of grievances. 	<ul style="list-style-type: none"> Public universities adopted speech codes as societal implementation for regulating discriminatory speech by faculty and students.

To maintain discipline at universities, authorities adopted some speech codes as societal implementation. As these speech codes oppose the First Amendment of the constitution of the USA, they do not stand in the courts. Table 1.5 lists different sections and acts around hate content in India (Alam et al., 2016; Yadav, 2018).

Table 1.5: The legal framework for hate-related content in India

Name of the Act	Section Name	Purpose
Indian Penal Code, 1860	124A IPC	Penalizing seditious
	153A IPC	Penalizing promotion of enmity between different groups based on different protected characteristics.
	295A IPC	Penalizes deliberate acts to insult religious beliefs
	505(1) and (2) IPC	Penalizes circulation of any content which causes enmity, hatred, etc. between classes.
Representation of The People Act, 1951	8	A person who misuses freedom of expression will be disqualified to contest the election.
	123(3A) and 125	Enmity based on religion, race, caste, community, or language in connection with the election is prohibited.
Protection of Civil Rights Act, 1955	7	Penalizing incitement to untouchability through words or signs or visible representations.
Religious Institutions (Prevention of Misuse) Act, 1988	3(g)	Prohibits the use of the institutional property to promote disharmony or hatred among different groups or communities in society.

Cable Television Network Regulation Act, 1995	5 and 6	The transmission of the program against the prescribed program code is prohibited.
Cinematograph Act, 1952	4, 5B and 7	A film is regulated by a certification board before its screening.
Code of Criminal Procedure, 1973	95	State Government is allowed to penalize published content that is restricted under different sections of constitutional acts.
	107	Executive Magistrate is empowered to prevent a person from disturbing public peace.
	144	District Magistrate and other state Government authorities are empowered to issue an order in urgent cases to maintain public order.
Information Technology Rules 2021	Rule 4	The ministry of electronics and information technology instructed the social media intermediaries to strictly follow the guidelines under rule 4 such as appointing persons at a different level, identifying the originator of the content, keeping a record of received complaints, etc.

As hate content lacks an appropriate definition, in India, there is no specific law that punishes the expression of hate content. There are some laws under constitutional acts to control incidents that incite hatred among the communities.

1.4 India's Digital Environment and Hate Content Control Strategies

Different schemes, models, and missions are initiated by the government of India to support digital activities in the country. The initiatives such as aadhar based payment, smart cities, Bharat Interface for Money (BHIM) app, and promotion of higher education are few to name from implemented digital initiatives in India. Sharing of e-resources among higher education students, academicians and researchers is a prevalent digital initiative (Cholin et al., 2003; Sasireka et al., 2011). As there is a cheaper data rate and easier access to communication networks, for every three seconds a new user associates with the Internet in India (Kant, 2021). Recently, the coronavirus pandemic has increased the use of digital technology in India (Sharma and Sengupta, 2020).

Indian Computer Emergency Response Team (ICERT), a unit of the Ministry of Communication and Information Technology is responsible for reviewing and blocking access to specific websites or information. After reviewing the received requests for blocking, ICERT will instruct internet service providers (ISPs) for the purpose.

The hate speech laws in India aim to avoid conflicts among the various religions in the country. These laws lead towards a punishment when a citizen unrespect others by race, religion gender, disability, language, occupation, or any other identity. The laws of hate speech also obstruct the expressing mechanisms, which harms the citizen. Article 19 of the Indian constitution provides the right to every citizen on freedom of speech and expression with the constraints to preserve morality, public interest, or decorum ([Indian Penal Code, 1860](#); [Law Commission of India, 1971](#); [The Constitution of India, 2007](#)).

According to section 153(A), from the Indian Penal Code (IPC), hatred expression against any identifiable group is accountable to punish by imprisonment of a maximum of 3 years or fine or both. Section 295(A) from IPC says hatred expression exclusively on religion is accountable to punish by imprisonment up to 3 years or fine or both.

1.5 Select indices of India's Hate Content

Every year, crimes, particularly hate crimes are increasing. Table 1.6 shows the status of different crimes in India. Religious hate crimes are taking place based on the religion of the victim. Mostly, the minority religious people are the victim of this crime. Now a day, crimes against women are prevalent and increasing. Women are the victims of crimes such as rape, kidnapping, assault, insult, and dowry deaths. The crimes against senior citizens are based on the age of the victim and increase every year. The other category of crimes is a crime based on the race of a person, and it is also increasing yearly. As technology advances, the usage of mobile phones in India is increasing drastically, and on the other side, cybercrimes are also increasing. The international ranking based on a particular index depicts the developmental status of the nation. International rankings of India on different indices related to society and information and communication technology are shown in Table 1.7.

The gender gap index is used to measure equality among different genders. The ranking on the gender gap index for India in 2018 is 108 and indicates that still more differences are there between the genders. The social hostilities index measures societal hostilities such as violence, harassment, and terrorism involving religion. The social hostilities index score of India in 2016 is greater than 7.2 out of 10 points.

Table 1.6: Hate crimes in India

Crimes head and source	Yearwise crimes						Based on	Remarks
	2013	2014	2015	2016	2017	2018		
Hate crimes								
Religious hate crimes (IndiaSpends factchecker, https://p.factchecker.in/)	---	18	30	42	74	92	The target of the crime is an individual or a group based on the religious identity	The count indicates the religious hate crimes in India are increasing yearly. According to the hate crime checker, mostly, the minority communities are the victims, and majority communities are the perpetrators of the crimes.
Crimes against women (National Crime Records Bureau, India)	309546	337922	327394	---	---	---	Rape, kidnapping, assault, insult, dowry, etc.	Includes all kinds of crimes that can be perpetrated against women
Dowry deaths (National Crime Records Bureau, India)	8083	8455	7646	7628	---	---		The victims are women
Crimes based on religion, race, and place (National Crime Records Bureau, India)	---	373	378	447	---	---	Targets religion, race, and birthplace	Hate crimes based on these characteristics are increasing
Crimes against a senior citizen (National Crime Records Bureau, India)	---	18714	20532	21410	---	---	Based on the age	The victim's age considered is 60 years and above
Information and communication technology								
Cybercrimes (Open government data platform India, https://data.gov.in)	---	9622	11592	12317	---	---	Use of electronics and communication devices and technology for the purpose	Cybercrimes increased as the advances in information and communication technology take place year by year.

Table 1.7: Select societal and information and communication technology indices for India

Index and Source	Yearwise Global Rank					Based on	Remarks
	2014	2015	2016	2017	2018		
Societal							
Global gender gap (World Economic Forum's The Global Gender Gap Reports)	---	108	87	108	108	Economic opportunity, political empowerment, educational attainment, and health and survival	Though, an improvement in wage equality (72 ranks) in 2018, India, still has the worst ranks of 142 out of 149 countries in economic opportunity and third lowest on health and survival.
Social hostilities index (Pew Research Center)	7.9	8.7	≥7.2	----	----	Involves religion as the major measure	The SHI score is calculated for 198 countries and is out of 10 points.
The human freedom index (Cato and Fraser Institutes)	---	---	87	102	110	Movement, expression, security and safety, association, assembly, etc.	Measured for 162 countries on a scale of 1 to 10. Value 10 showing the highest degree of freedom
Global terrorism index (Institute for Economics and Peace)	6	6	8	8	7	Measured by considering terrorist activities in the country	The GTI is measured referring to a global database on terrorism maintained by the START. India ranks among the top 10 countries in terrorist activities.
World happiness index (UN Sustainable Development Solutions Network)	111	117	118	122	133	Based on inequality, life expectancy, GDP per capita, social freedom, public trust, and social support.	Measures the happiness of the citizen and their immigrants. A continuous fall in the happiness index indicates disturbed social life.
Trust index (Edelman Trust Barometer Global Report)	61	68	65	72	68	Based on the performance of the government, media, business, and NGOs.	The outcome of the measurement is trust (60-100), neutral (50-59), and distrust (1-49). India is one of the most trusted countries.

Worried about fake news as a weapon (Edelman Trust Barometer Global Report)	---	---	---	----	71-75%	In search of truth	Approximately, 7 out of 10 countries worry about fake news being used as a weapon.
Human Capital Index (World Economic Forum)	72	100	105	103	158	Based on four thematic dimensions and five distinct age groups	India ranks lower than BRICS countries for the knowledge and skills people possess that enable them to create value in the global economic system.
Information and communication technology							
Global Cybersecurity index (International Telecommunications Union-ITU)	---	05	---	23	---	Legal, technical, organizational measures, capacity building, and cooperation	Though the ITU report on the global cybersecurity index (GCI) of 2015 motivates to increase the ranking, the GCI rank of India in 2017 falls from 5 to 23.
ICT Development Index (International Telecommunication Union)	131	135	138	134	---	Three clusters of ICT indicators-access, use, and skills (11 indicators)	The indexing is based on the agreed information technology indicators for 176 countries calculated to know the ICT strength of a country.
Global Innovation Index (World Intellectual Property Organization and Cornell University)	76	81	66	60	57	Based on 80 indicators including intellectual property, education, research, and publications	The ranking is calculated for 126 countries to know the innovative strength of a country.
E- participation Index (United Nations)	40	---	27	---	15		It considers 193 countries and ranks based on ICT-supported participation processes for government and governance.
4G LTE penetration (OpenSignal.com)	---	62	24	14	---	Coverage of 4G LTE	The ranking indicates the coverage of 4G LTE is increasing in India.

The human freedom index is used to measure the freedom of a person about personal, economic, and civil factors in a country. In 2018, India had the 110 rankings for the human freedom index, which depicts less freedom for a human being. The global terrorism index is assigned to the countries by measuring the associated terrorist incidents. India ranks among the top 10 countries with a ranking of 7 in 2018 for terror-involved activities. The world happiness index is used to measure the happiness of the citizen and its immigrants. The ranking of the world happiness index of India is 133 in 2018. The trust of the citizens in the government, media, business and Non-Governmental Organisations (NGOs) is measured using the trust index. India's trust index ranking is acceptable.

Worried about fake news as a weapon index indicates the percentage of worriedness of a country about the impact of fake news on social life. India is more worried about fake news as it spreads quickly and escalates the violence. The human capital index is used to measure the knowledge and skill of the citizen which contributes value to the global economic system. India possesses a low ranking for the human capital index.

The global cybersecurity index is used to measure the capability of a country for providing security in cyberspace to its citizen. The global cybersecurity index of India in 2017 is 23. The information communication and technology capability of a country is measured through the ICT development index. ICT development index of India in 2017 is 134. The quality of the intellectual property, education, research, and publications of a country is measured as the global innovation index. In 2018, India had a ranking of 57 for the global innovation index. The level of e-participation for governing purposes is measured as the E-participation index. India acquired the 15th rank for the E-participation index globally in 2018. 4G LTE penetration index is used to measure the coverage of a 4G LTE network in a country. India held the good rank of 14 for 4G LTE penetration in 2017.

Most of the indices' values are below average and not good for the social life of the citizen. The indices tabulated and discussed exhibit some kind of hatred among the citizen. For example, the global gender gap and social hostilities indices indicate the possibilities of hatred existence. Hence, the reported values of these indices motivate the study of online hate content in the Indian context.

1.6 Justification for Study in Indian Context

India is the largest democratic and second-largest populated country in the world. The people of various religions and castes live in India. Although, India is a secular republican country with equal rights to all citizens, is contending with religious nationalism towards a national identity (George, 2016). Colonial policies of the British during their rule in India created a competitive environment among the Hindus and Muslims. This competitive environment is represented through the publications and communal acts hurting religious sentiments. To avoid its effect and maintain public order, laws were introduced to restrict speech by authorities (Nair, 2013).

The concept of religious hate has been originated during the British rule in India as a result of their divide and rule strategy to control citizens (Hartnack, 2012; Alam et al., 2016). Based on this strategy, both India and Pakistan have been separated and still fighting with religious hatred. Even today also the attempts of dividing the citizen in the name of religion are taking place in India.

In India, hate is using to divide the citizen based on protected groups or characteristics, particularly based on religion. Several times the attempts have been made to divide the citizen in terms of Hindus and Muslims for political benefits. The political parties and other groups with the motives of acquiring power and egotism divide the citizen by posting hate content on online social media such as Facebook, Twitter, YouTube, and Instagram.

Facebook can be used for spreading hate content through posting messages, sharing the messages, and liking the messages. Similarly, Twitter can be used for tweeting, re-tweeting, and liking the content. The prevalent use of YouTube for hate content propagation may be through the distribution of stereotypes in the form of videos over the Internet. Instagram may be used to post and sharing disputable images.

Being the largest democratic country, India allows free speech within the laws and constitutional codes. On the other hand freedom of expression is restricted online for the betterment of the citizen in terms of defamation, national security, and communal harmony. The laws and techniques used to restrict free speech are undermining the freedom of expression in India. This restriction is a result of social and political

pressure. In Information Technology (IT) Act with related amendments, there is no clear definition of offensive and no clear information on how to behave online and offline (Patry, 2013). This unclear information is causing, the generation, reporting, and removal of offensive content.

According to the Pew Research Center, India is one of the countries with a very high social hostilities index (SHI). The SHI of India is 9.0 with a 10 point scale as of 2013 (Pew Research Center, 2015).

The discussions from the preceding paragraphs indicate that possibilities for the expression of hate content over the social networks within the proximity of India are more. Therefore it is necessary to identify the causes, and impacts of hate content for crafting the policy framework to control hate content.

1.7 Motivation for Research

All the governing authorities public as well as private intended to maximize revenue, improve customer/citizen experience, minimizing costs, protecting human rights, and maintaining equality. These can be achieved easily through the use of digital governance. Another motivating factor for digital governance is advances in computing and storage technologies such as cloud computing, mobile devices, analytics, etc. The political, economic, social, cultural, geographical, managerial, citizen's expectations, regional comparisons, and demographical reasons are also the motivating factors to digital governance. Digital governance is essential to fulfilling the requirements of modern citizens like customized services, on-demand services, services at any time, services everywhere, and services as needed. Digital contents are violated due to the availability of ITs at reachable cost and the destructive behavior of the users. These violations can be monitored and controlled with the help of digital content governance by framing some regulations. In the Indian context, even though digital governance is initiated a decade back, still requires more improvement.

1.8 Outline of the Study

The research attempts to address the research question “How to improve digital hate content regulations in the Indian context?”

To address this question, the following research objectives are set.

- To study the nature and forms of digital hate content in the Indian context.
- To understand the influence of digital hate content on online safety.
- To study policies and regulations addressing digital hate content.
- To bring out recommendations for improving digital hate content regulations in India.

The portion of data for the present study is collected from internet users using a questionnaire designed for the purpose. Another portion of data is collected from a social media giant Twitter as case studies. To fulfill these objectives, the data collected is analyzed by a mixed-method approach.

1.9 Organization of the Thesis

After the introductions of the research in previous sections, the remaining portion of the report is organized as follows. Chapter 2 provides a literature review on hate content based on gender, religion, race and disability identities, hate content regulations, and other associated topics. The conceptual model and research methodology are outlined in chapter 3. In chapter 4, the quantitative survey and analysis results are discussed. Chapter 5 presents qualitative analysis and results of different case studies. The synthesis of results and a framework for hate content control is drawn in Chapter 6. Finally, chapter 7 concludes the work carried out.

1.10 Concluding Remarks

Different terms like digital content, hate content, e-governance, hate content regulations, and related demeaning expressions are briefed. After reviewing the hate content definitions, hate content is defined as “The content, which attacks an individual or a group intending to hurt or disrespect based on the identity of a person or group”. Different violent acts such as terrorism, extremism, hate crimes, and exploitation are discussed. The role of social media for demeaning content expressions and different stages of hate content after the trigger event are outlined. The transition of hate content from offline to online mode is represented through its evolution. As hate content is a global issue, different international organizations' efforts to control it are studied and

compared. Select nations' hate content regulations were also studied and compared for the purpose.

Hate has multiple dimensions such as disgust, anger/fear, and devaluation/diminution. India is not free from crimes based on protected characteristics. The international ranking of India for several indices is not good and needs to be improved. As digital content is vast and can reach anyone, it is essential to regulate for the betterment of society. From the discussion on fake news vs. hate speech, it has been observed that fake news causes hate speech. Finally, it has been observed that India is equipped with ICERT and hate speech laws such as sections 153(A) and 295(A) of IPC for countering hate speech.

The observed index values of India are inadequate for the comfortable social life of citizens and online hate content control laws are less effective. Therefore, the study is presented to make some suggestions to improve the social life of citizens by framing the above research question and attaining the research objectives.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

The background of the research is introduced in the previous chapter. A detailed literature survey is carried in this chapter. Hate content can have actual consequences in real life. The target of hate content is not a single identity, it can be expressed concerning gender, religion, race, and disability (Seglow, 2016). The content which is expressed over the internet and its associated social media based on any of the protected characteristics such as gender, religion, race, and disability is referred to as online hate content.

Developments with mobile and web technologies and ease of posting opinions are increasing hate content on social media platforms (Mossie and Wang, 2020). The growth of social media users is also increasing the hate content on social media platforms (Corazza et al., 2020; Modha et al., 2020). Expression of hatred against any of the protected characteristics over the Internet and social media is less expensive, faster, easier, and efficient (Alkiviadou, 2019; Guiora and Park, 2017). Therefore, in this chapter, a literature review on hate content involving these identities is made. The rest of the chapter reviews different theories of hate and governance, hate content models, hate content detection techniques, and governance policies. After the literature review, the research gaps are presented at the end.

2.2 Related Concepts of Hate Content and Governance

As the theoretical frameworks are important for researching (Ngulube et al., 2015), it is necessary to review related theoretical concepts of hate and governance. Table 2.1 briefs some of the related concepts of hate and violence.

Table 2.1: Concepts of hate and governance

Theory	Description	Authors	Remarks
Slacktivism	Slacktivism consists of slacker and activism words and denotes the disconnection between awareness and action using social media	Glenn (2015)	This requires less time and effort to solve a social issue. With this approach, the maximum geographic area can be reached for the purpose.
Hate and love	Love is intended to value others and is concerned with the preservation/promotion of	Sternberg (2003), Rempel	These approaches indicate that love tries to bring closeness among the parties

	other's well-being. Similarly, hate is intended to devalue others and is concerned with the goal of diminishing/destroying the other's well-being. The triangular hate theory involved intimacy, passion, and commitment	and Burris (2005)	and hate tries to maintain the distance between the parties. The hate exhibits disgust, anger, fear, devalue and diminution.
Racism	Racism is based on an ideology of dividing people into racial groups and serving the interests of a group.	Bonilla-Silva (1997)	Ideologies are generated by the superior/dominating groups and exhibit their interests
Gender-based violence (GBV)	GBV results from normative role expectations of each gender and the existence of unequal power between the genders in society.	Bloom (2008)	The perpetrators in GBV may be an intimate partner or anybody other than an intimate partner. Generally, this violence is on women and girls. This violence can take place at different stages like infancy, childhood, adolescence, adulthood, and old age
Perception of Disability	Societal attitudes and perceptions of disability are driven by education and knowledge; if there is no exposure to a disability, it indicates that society would have a narrow understanding of it	Esmail et al. (2010)	Education plays an important role in controlling the negative attitude towards disability. Women with disabilities face more violence than normal people.
Religious communication	Every religion tries to represent itself as different, superior, complete, and genuine than any other religions	Pace (2011)	The people, who belong to the superior religion, try to exhibit hatred expression towards the other religious people.
Conflict resolution and neutralization	Conflicts resolution and neutralization can be made by justifying and excusing the guilt of individuals so that they feel good about themselves	Hamlin (1988)	Mistakes are common to a human being, can be corrected by conflict resolution and neutralization techniques
Organizational Theory	A system in which higher authorities supervise and control subordinates,	Heady (2001)	The hierarchical structure will be followed in governance
Political elasticity theory of governance	An integrated system of soft and hard forms of political power and decentralization of power.	Werlin (2003)	Reduces the confusion and tension of administration

In general, governance is a political activity to attain the purpose of a national government (Asaduzzaman and Virtanen, 2016). Expression of hate can be controlled by governance activities.

2.3 Models of Hate Content

The expression of hate is not concerned with a single identity rather targets multiple identities. The literature of hate content based on some of the protected characteristics such as gender, religion, race, and disability are reviewed in the following sections.

2.3.1 Gender-based Hate Content

This is an expression, which is made on the grounds of gender or sex. The victims of this kind of hate are generally women and girls. There is intended violence on women and girls due to their gender identity. This is also known as sexist hate content and is a kind of social shaming that intends to disrespect women, introduce fear and insecurity among women in society.

The easy availability of the Internet, the rapid growth of information and communications technologies, and the common use of social networks made addressing violence against women and girls much simple. These advancements are being used as tools to harm women and girls. Gender-based hate violence results from the existence of expectations from one gender over the other and unequal power of two genders in a societal environment (Bloom, 2008).

Social networks are the primary medium for online harassment based on gender. This kind of harassment of women affects the personal lives and professional careers of women (McLaughlin et al., 2017). Both women and minority groups are targeted by online hate than any other gender and community. Internet may be an unsafe space for some academics, particularly those who face societal inequalities, like being a woman or a person of a minority group (Barlow and Awan, 2016). The active women on social media may receive many comments directly attacking their gender identity, safety, and right to an opinion. For a woman, being online is like walking on the streets of an Indian city, town, or village to transgress an unwritten social patriarchal law by crossing over the space which is dedicated to men only (Kovacs et al., 2013).

The act of bullying, whether conventional or through digital devices is influenced by individual personality and contextual factors (Casas et al., 2013). An act of bullying can be in any form such as traditional bullying and cyberbullying. Both the forms of bullying, traditional as well as cyber involve changing patterns of gender. Sometimes, during bullying bystanders are intended to help victims in case of a more severe incident and sometimes intended to promote bullying when other bystanders were good friends rather than acquaintances (Bastiaensens et al., 2014). Both the behaviors of helping and reinforcing during bullying are gender-dependent.

In a home with boys and girls, children are made to socialize into different domains based on gender. Females are motivated to socialize for taking care of others and better communication, whereas males are motivated towards non-communal, leadership, and achievement-oriented (Ridgeway, 2011). With this differentiated treatment at home, men and women will choose their convenient profession. Based on the profession, most likely women will have contacts with the people on female-dominated jobs like teacher, nurse, cashier, and hairdresser, whereas men most likely will have contacts with the people on male-dominated jobs like computer programmer, banker, security guard and factory operator occupationally (Chua et al., 2016).

The high-speed evolution of online social networks has weakened the laws that control them and result in a difficult situation for victims of online attacks. Feminist campaigners are also facing abuse and harassment through the usage of online social networks (Hardaker and McGlashan, 2015). One of the solutions to online harassment such as rape threats against feminist campaigners is to adopt the do-it-yourself approach (Jane, 2016). Hate crimes are increased by legal inequalities because they lead to bias, discrimination, and violence. Legal equality will result in reduced violence (Levy and Levy, 2016).

Simons (2015) highlighted that there is a need to have analytical research for providing insights to empower victims, discourage perpetrators, and increase awareness among the public. Barlow and Awan (2016) suggested that the social networks companies, like Twitter, should take corrective measures to counter online abuse against women and minorities. Edwards (2017) identified that women are recruited by terrorist organizations to establish physical relationships with men.

Based on the identified relationship among the predictors of traditional bullying and cyberbullying, Casas et al. (2013) suggested that educational programs can be used as a tool to counter abuses of both bullying and cyberbullying. Factors involved such as personality, contextual and roles are closely related to both the acts. Beckman et al. (2013) determined the role of youngsters with gender differences engaged in traditional bullying and cyberbullying using data samples of size 2989 from school students of Sweden to control cyberbullying. Bastiaensens et al. (2014) examined the effect of contextual factors on bystander's behavioral intentions towards helping the victim or reinforcing the bullying during the harassment using Facebook with the data collected from 453 secondary school students of Flemish.

After analyzing the attitude towards gender, statements such as *women are dedicated caretakers and mothers* and *men are facility providers* are made by Ridgeway (2011). Similarly, Chua et al. (2016) identified the nature of women and men towards contact establishment with others in the society.

Levy and Levy (2016) after analyzing the effects of 3 policies on a partnership of same-sex, non-discriminated employment and laws of hate crime with annual data from 2000-2012, show that hate crimes are affected by public policies related to sexual orientation. Hardaker and McGlashan (2015) investigated the sustained period of abuse and harassment towards a feminist campaigner and journalist, Caroline Criado-Perez via her Twitter account using an interdisciplinary approach with quantitative and qualitative analysis. Jane (2016) examined the responses of feminists to increasing problems of online hate with a focus on female gamers and the responses of Australian gamer Alanah Pearce with alert messages to their mothers against sexual violence threats from young male Internet users.

2.3.2 Religious Hate Content

This is a type of hatred expression against religions such as Islam, Hindu, and Christian. As religion contains a group of people, the hate content against this is more harmful than against an individual. Every religion tries to represent itself as different, superior, complete, and genuine than any other religion (Pace, 2011). People belonging to the superior religion, try to exhibit hatred expression towards the other religious people.

Minority people are demonized and vilified online through negative attitudes, stereotypes, discrimination, physical threats, and harassment to create violence. An analysis of online communities is possible by observing their activities such as information they post, share, and like (Awan, 2016). Minorities are being used as a model to depict homogeneous out-group, involved in conflict, violence, and extremism characteristics. The internet forum serves as an online amplifier by reflecting and reinforcing available discourses in traditional networks which leads to stronger polarizing effects on public discourses (Tornberg and Tornberg, 2016).

Microblogs being posted during disasters include situational information and the emotions/opinions of the public. It is necessary to emphasize non-situational tweets/communal tweets rather than only situational tweets, i.e., abusive posts towards a specific religion or racial group. Generally, communal tweets are made by common and popular users with more followers. Communal tweets get more exposure (retweets) than situational tweets (Rudra et al., 2016). To summarize the review on religious hate speech a comparison is made in Table 2.2.

Table 2.2: Comparison of works on religious hate speech

S. No.	Authors	Data source	Purpose of Work	Methodology	The outcome of the work
1	Awan (2016)	Facebook	To address the issue of Islamophobia on social networks.	Mixed methodology (qualitative data collecting techniques with grounded theory).	<ul style="list-style-type: none"> • 494 incidents of online hate targeted towards Muslim communities. • Typology of five characteristics (opportunistic, deceptive, fantasists, producers, and distributors) of anti-Muslim hate adopted on Facebook.
2	Tornberg and Tornberg (2016)	Swedish Internet forum	Examining different representations of Muslim and Islam words in a large corpus of 105 million words.	<ul style="list-style-type: none"> • Topic modeling and discourse analysis 	<ul style="list-style-type: none"> • Minorities are being used as a model to depict homogeneous out-group, involved in conflict, violence, and extremism characteristics

					<p>originating from Islam.</p> <ul style="list-style-type: none"> • Internet forum serves as an online amplifier by reflecting discourses in traditional networks.
3	Rudra et al. (2016)	Twitter	To emphasize non-situational/communal tweets rather than only situational tweets, (abusive posts towards a specific religion or racial group)	<ul style="list-style-type: none"> • Twitter Search API to collect tweets • 10-fold cross-validation approach to build classifiers 	<ul style="list-style-type: none"> • Communal tweets are made by common users and as well as popular users with more followers. • Communal tweets got more exposure through re-tweets than situational tweets. • A technique to counter communal tweets during disasters such as anti-communal tweets by the non-communal user.
4	Awan and Zempi (2016)	Twitter	To examine how minorities are affected by online/ offline hate crimes and the affinity between online and offline crimes.	<ul style="list-style-type: none"> • Mixed methodology (qualitative data collecting techniques with grounded theory). 	<ul style="list-style-type: none"> • Some minority people withdrew their presence from online social networks. • Hostility towards minorities in both the cyber and the real world is a continuous process.
5	Rahman et al. (2016)	Canadian online news media	To know the relationships between the Islamic culture (attire) and the opinions of the online readers.	<ul style="list-style-type: none"> • Sentiment and content analysis methods 	<ul style="list-style-type: none"> • hijab is an indicator of individual integration level • Educating policy on Islamic culture to avoid misinterpretation and confusion.

The clothing style of an individual can provide some information on a person. The people will think that the hijab is an indicator of individual integration level. This misinterpretation and confusion can be avoided by educating the general public on

Islamic culture. The initiatives like art and design activities of religion and collaborated fashion projects to exhibit multiple cultures could be proposed and implemented to serve the purpose. (Rahman et al., 2016). There were increased attacks on minorities after terrorist attacks on Paris, Tunisia, and Woolwich. In these attacks mosques have been vandalized, hijab or niqab of women were pulled off, physical torture is given to minority people and their properties have been destroyed (Awan and Zempi, 2016).

For victims of an attack, it is difficult to separate online threats from offline threats. Victims will live in a fear of the materialization of online threats into real-world incidents. With this affinity between online and offline hate crimes, some minority people have withdrawn their presence from online social networks to be safe. Hostility towards minorities in both the cyber and the real world is a continuous process (Awan and Zempi, 2016).

2.3.3 Racist Hate Content

An expression towards the appearance of a person or group is known as racist hate speech. Usually, this kind of speech takes place at the international level. The frequency of occurrence and impact of this speech depends on the intention and perception of the government of a particular nation and varies from one leadership to another leadership. Tatum has argued that “racism is a system involving cultural messages and institutional policies and practices as well as the beliefs and actions of individuals” (Tatum, 2001, p. 103). Wodak and Reisigl (1999, p. 181) assumed that “racism is both an ideology of a syncretic kind and a discriminatory social practice that could be institutionalized and backed by the hegemonic social groups”. This indicates that in an environment or a system, people of one group exhibit their power against another group/individual based on physical appearance such as skin color.

Apart from providing freedom of expression, social media amplifies hate content with the re-tweeting facility (Kwok and Wang, 2013). When subjects were approved by a white male having high followers, there is a significant decrease in a racist slur on Twitter (Munger, 2016). After the death event of a young Black man Mike Brown, who was killed by a shooting from Darren Wilson, more black people tweeted than white people (Chaudhry, 2016). Social networks have a significant role in racism and they act as sources to understand it. Social networks provide a context for learning, challenging,

and addressing issues related to racism. Social networks act as a platform for students to discuss identity and the ways to produce and consume networks that reflect and shapes societal attitudes and constructions of the race (Nakagawa and Arzubiaga, 2014).

From a study on different minorities, only blacks of ethnic minorities have shown negative views about the US than white people (Uemura, 2017). Discrimination towards ethnicity by adults during school age and by peers was associated with more negative perceptions of one's ethnic group. There is a strong association between ethnic-racial parent socialization and ethnic identity for persons with more adult discrimination (Rivas-Drake et al, 2009). For the adolescents who consider their mother's role is important in socialized decision making, there exists a positive association between racial barrier socialization and engagement among adolescents. In contrast, there exists a negative association between barrier socialization and engagement among adolescents who consider mothers' role as less democratic-involved parenting (Smalls, 2009).

The stress resulting from existing racism with discrimination and hatred expression in America is affecting the physical and mental health conditions of African Americans. A study on African American women indicated that racism on a person is more prone to alcohol consumption by the victim whereas group racism is less prone to alcohol consumption (Kwate et al., 2010). Depressive symptoms are possible with perceived racial discrimination. A higher level of coping reduces the depressive symptoms whereas a lower level of coping increases depressive symptoms. A better coping effort normalizes the association between perceived racial discrimination and depressive symptoms (West et al., 2010).

Racism is associated with drug usage among African American women and the socio-economic status of the individual or a group. Lower socioeconomic status increases the risks of drug usage and racism in that particular group (Stevens-Watkins et al., 2012). Racism has caused negative moods and consumption of alcohol daily in the evening by African American college students. Men are more negative and addicted to alcohol drinking than women (O'hara et al., 2015). Racial socialization is also associated with the academic performance of the students (Banerjee et al., 2011).

It is difficult to define American with the context of immigration but the students in America dared to define it. Even though the school follows a multicultural ideology, the students defined *American* as a person who is US-born, white-colored, and able to speak in the English language. This definition itself conveys a kind of racism (Lash, 2017). During an electoral process, the division of voters was on the grounds of racism and sexism, particularly by whites without higher education. Even under proper supervision and control, racism and sexism were proved as powerful forces for the 2016 presidential election (Schaffner et al., 2016).

In December 2011 the number of hate crimes recorded by the England police department and represented by Roberts with his team is 43,748. Out of these hate crimes, 35,816 (82%) were racial, 1,621 (4%) were religious, 4,252 (10%) were sexual oriented, 1,744 (4%) were disability-based and 315 (1%) transgender based. A lot of research has focused/emphasized victims and their perspectives but not on persons responsible for hate crimes (Roberts et al., 2013).

Tolerance is the only force that allows different classes and races to settle down disputes and work towards reconstruction. Intolerance in India is initiated by the British community and leading towards the reaction, harshness, retaliation, revenge, misunderstanding, and hatred. The hostile attitude of one community towards another made living difficult (Yousafzai & Khan, 2011). Indian population has been categorized based on the occupation and social structure by the British officials forming different castes in India. For the sake of control, certain tribes in India have been criminalized by the law and considered people belonging to those tribes are habitually criminals (Bates, 1995). The result of this hatred criminalization of certain tribes during British rule might lead to the possible racism currently in India. In the nineteenth century during the British rule in India, racism is directed against Indians by discriminating with education, employment, salary, and respect (Kumar, 1982).

The people from the Northeast part of India migrate towards mainland India for better employment opportunities. The histories of Northeasters' presence in Delhi indicate different forms of racial discrimination against them. Physical appearance, cultural difference, and belongingness to the outer part of the country may be the reasons for the discrimination against the Northeasters in India (Sitlhou & Punathil, 2014). Racial

discrimination in India is leading to increased illegal migration and Human Trafficking (Pandya & Pandya, 2011).

Munger (2016) identified and collected some Twitter users who harassed others and used “bots” created accounts with control for sanctioning the harassers. After analyzing the tweets following the death event of a black man, Chaudhry (2016) identified that the tweets from the concerned racist community (victim group) will be more than the perpetrating group. Nakagawa and Arzubaga (2014) discussed how to promote racial literacy with more emphasis on intersectionality.

2.3.4 Disability Hate Content

The hostility or incitement made towards the physical and mental conditions of a person is referred to as hate content on disability. Disability is not an isolated entity associated with a medical field, is a social category like race and gender. Disability means a physical or mental problem associated with an individual limiting some of the life activities. The synonyms for disability are mutilated, deformed, monstrous and deviant.

Disability is increasing due to medical advances which made people survive longer with the help of impairments that get worst with age. As disability can happen to anyone at any time during their lifetime, it touches all races, genders, nationalities, and generations. With this view, some activists refer to non-disabled people as temporarily able-bodied (Kudlick, 2003). Proper perception of disability by society is made through education. If the disability is not exposed, then there is a narrow understanding of disability by society (Esmail et al., 2010). Education plays an important role in controlling negative attitudes towards disability. The way of perceiving the disability by the perpetrator leads to hatred against a person with a disability (Hollomotz, 2013). Denying parental rights and avoidance of sexual freedom are the structural barriers against persons with a disability (Stevens, 2011). Persons with disabilities are more vulnerable to hatred than able-bodied people.

As persons with a disability can be targeted easily and response for the crimes is low, they are the common victims of hate crime (Roulstone et al., 2011). Even though persons with a disability are more vulnerable to hate violence, reporting the incidents is less as compared to others. For the sake of dignity and safety of persons with a

disability, concerned governments should provide an appropriate system for reporting and control the crimes (Macdonald, 2017; Thomas, 2011).

Violence against women with intellectual disabilities at home is more and required to treat like violence in public. Usually, an able-bodied man will establish a relationship with a woman with an intellectual disability, initially start being pleasant, and gradually moves towards controlling her (McCarthy, 2017). Domestic violence is an issue of gender involving mostly men as perpetrators and women as victims. Domestic violence on women with disabilities is even more than the able-bodied women. The women with disabilities are tortured by their partners in multiple ways, such as denying access to transportable aids like a wheelchair, avoiding access to medicines, personal care, and home facilities (Thiara, 2011). Table 2.3 summarizes the literature of identified variables.

Table 2.3: Literature for identified variables

Authors	Category of hate content	Contributing variables
Bloom, 2008; Women, 2002; Simons, 2015; Barlow and Awan, 2016; Hardaker and McGlashan, 2015; Kovacs et al., 2013; Campbell, 1981	Gender hate content	Alienated, inclination and personal experience
Pace, 2011; Awan and Zempi, 2016; Rahman et al., 2016; Rudra et al., 2016; Rahman et al., 2016; Mulligan, 2018	Religious hate content	The stance of superiority, ignorance, religionizing nation, and clothing style of a person
Bonilla-Silva, 1997; Bianchi, 2014; Munger, 2016; Chaudhry, 2016; Yousafzai & Khan, 2011; Kumar, 1982; Sitlhou & Punathil, 2014; Jones, 2013; Catholic Charities, 2008; Miller and Garran, 2017; Banton, 1992; Crawford, 1998	Racist hate content	Inherited nature, peer pressure, personal experience, stereotypes, unfamiliarity, skin color, and poverty
Hollomotz, 2013; Stevens, 2011; Roulstone, 2011; Thiara, 2011; Walter-Brice, 2012; Wendell, 1989; Hannon, 2007	Disability hate content	Ignorance, envious, the fright of health and able-bodied
Hanes & Machin, 2014; Heyman, 2008; Olteanu et al., 2018; Qin et al., 2011	Online hate content	Communal nature, freedom of expression, terrorism, and extremism

Mate crime is an act of a crime on persons with disabilities which is carried out by the friends or relatives or familiar persons of the victims and is more similar to domestic violence. The women with disabilities have been dually victimized by their partners and failed systems. These kinds of victimizations made women feel powerless and self-blamed (Walter-Brice et al., 2012). Cyber harassment of persons with disabilities is based on their disability, family presence, perceived complexity, ignorance, and misrepresentations by perpetrators. Disablism and its vulnerability could be addressed by emphasizing these factors (Alhaboby et al., 2016).

2.4 Hate Content Detection Techniques

Flames are abusive messages and lead to both frustration and time wastage of Internet users. Exchange of flames is an enjoyment to some users and for most users, it is insulting and abusive (Spertus, 1997). Causes of flames are variations in demographic, psychological, and behavioral characteristics of the user. Flaming will get an increase when both parties in communication are unknown to each other. The flaming intensity is dependent on the age of a person and is more with the adolescents than other age groups. It is difficult to get a rigid system for flame detection with all the purposes but possible to have a tolerant and interactive system (Razavi et al., 2010).

Freedom of speech is fully utilized by Internet commentators through online social media like Twitter, Facebook, Yahoo, Linked In, and newsgroups. There are terms of service to prohibit hate content with all social media. If user posts or messages are filtered for offensive words to block by the social media, there does not exist any public and separate hate content detection systems (Warner and Hirschberg, 2012).

Sometime a user may treat others comment inappropriate irrespective of whether it is meant to him or not. Insults and other forms of hatred expressions posted against the tradition of people or religion hurt more than against an individual (Ismail and Bchir, 2015). Social networks and dedicated websites associated with a particular group are the major media to communicate and spread hateful messages. Dangerous hate content usually motivates the public to involve in violent acts towards the targeted group which leads to destructive activities like beating, looting, and killing (Gitari et al., 2015).

Whenever a person associates with online social media, there are more possibilities of risks such as online harassment and ridicule irrespective of the media type. An abusive language consists of hate content, derogatory and profanity expressions. Most commercial methods use black lists and regular expressions for detection purposes, but these measures are not suitable for more subtle kinds of hate content. As the growth of online content is high, it is required to have accurate and automated methods to process it (Nobata et al., 2016). Because of natural language processing, detecting hate content is considered a task of classification requiring the training of a classifier model. Annotating the tweets by multiple users makes the classification model as regression to predict the degree of hatefulness as a fuzzy value instead of hard values such as hate and non-hate (Ross et al., 2017).

Online social media like Twitter provides useful and almost real-time data sources to analyze hatred expressions as responses to trigger events. These data sources are very much important to determine the online social behavior and emotion of individuals following a destructive and violent terrorist attack. Big data plays an important role in making policies and decisions. A supervised machine learning text classifier is developed to identify online hate content through Twitter data following Lee Rigby's murder incident. Generally, a combination of words as n-gram produces better results (Burnap and Williams, 2015). The Naïve Bayes classifier used by Kwok and Wang (2013) on Twitter data produced only 76% of average classification accuracy on individual tweets because the classifier is built using only unigrams instead of n-grams.

Negative contents with angry, hostile, or abusive messages are barriers to users from participating in the social web. Hence encouraging users to participate in online communities is a challenging task for technological companies and developers of the applications (Sood et al., 2012). Social media are beneficial to adolescents by facilitating the activities such as interaction with others and learning from others. As a penalty to these benefits, adolescents are at risk of exposure to offensive content online. With the view of adolescents' safety, offensive contents detection is essential because negative impacts of harmful and biased content on adolescents are more than the adults (Chen et al., 2012).

Twitter provides a micro-blogging facility with a maximum size of 140 characters to share the tweets. Bag-of-words, part-of-speech (POS) and pattern-based methods are not suitable for processing tweets with significant noise. Bag of words-based methods is also inappropriate when abusers use simple but tricky forms to convey the messages. These messages are difficult to detect by the systems but convey a clear intent to the recipients (Djuric et al., 2015).

Sexists, religious and political extremists usually target to hate particular individuals/organizations through online social media. The content posted by the users on social networks reflects their interests. Conventional methods for detecting community, use only explicit information on users. Unlike traditional machine learning methods, ontology-based methods learn without training (Fu et al., 2012). The learnability of the classifier depends on the set of features used to train. There is a necessity to improve overall performance by increasing classification accuracy, changing parameters, and optimal kernel functions (Warner and Hirschberg, 2012).

Cyberbullying is directed at an individual and resembles a threat. Methods for automatic detection of cyberbullying are limited, hence the negative consequences of cyberbullying are increasing daily. Cyberbullying targets teenagers as victims and results in depression and suicide of the teen users. Using machine learning techniques, it is possible to develop a system along with a set of rules to detect cyberbullying automatically (Reynolds et al., 2011).

A terrorist attack is one of the antecedent/parental trigger events for the production and dissemination of hate on online social media like Twitter. Following an attack, the hate speech will be more at the time of impact stage, will start to reduce at the inventory stage, and will vanish during the reaction stage (Williams and Burnap, 2015). Online social networks being the richer sources of data are important to scientists for researching human behavior. Following an attack on Charlie Hebdo weekly in Paris 2015, tweets were collected and classified as tweets with violent or hate and general tweets (Miro-llinares and Rodriguez-sala, 2016). Nowadays identification of abusive statements in online content has become a common trend. As the growth of online content is high, it is required to have accurate and automated methods to process it (Nobata et al., 2016). Table 2.4 shows a comparison of works on hate content detection.

Table 2.4: Comparison of works on hate content detection

S. No.	Authors	Data Source	Purpose of Work	Methodology	The outcome of the work
1	Williams and Burnap (2015)	Twitter	To analyze cyberhate with online social media after Woolwich terrorist attack.	<ul style="list-style-type: none"> • Twitter streaming API to collect data • Supervised Machine Classifier 	<ul style="list-style-type: none"> • The terrorist attack leads to hate speech • After a terrorist attack hate speech occurs, reduces, and vanishes gradually.
2	Burnap and Williams (2015)	Twitter	To help decision-makers by monitoring the reaction of the public during emotive events.	<ul style="list-style-type: none"> • Twitter API to collect data • 10-fold cross-validation method for training and testing machine learning model Statistical regression models 	<ul style="list-style-type: none"> • n-gram typed dependencies produced better results
3	Warner and Hirschberg (2012)	Yahoo and American Jewish Congress (AJC)	To detect hate speech in online text.	Support Vector Machine (SVM) classifier	<ul style="list-style-type: none"> • Bigram and trigram templates reduced the performance of the classifier.
4	Miro-linares and Rodriguez-sala (2016)	Twitter	To categorize different representations of violence and hate speech.	Classification followed by qualitative and quantitative analysis	<ul style="list-style-type: none"> • The hashtag is identified as a variable for predicting violence and hate messages from the tweet.
5	Nobata et al. (2016)	From Yahoo	Detect abusive language in online content.	Supervised classification with Natural Language Processing (NLP) features	<ul style="list-style-type: none"> • Hate speech detection method for online user comments. • Corpus of user comments.
6	Cohen-Almagor (2011)	--	Analysis of the ways for countering hate speech on the internet	--	<ul style="list-style-type: none"> • A framework to identify problems and resolve them by considering ethical and social responsibilities. • Possible solutions to counter hate speech like speech vs. speech, education, hate-watch, etc.

Summarization of previous works in the form of major findings on different identities has been made in Table 2.5.

Table 2.5: Major findings from previous works (Category wise)

Category	Works	Major Findings	Remarks
Gender Hate content	Simons (2015), Barlow and Awan (2016), Casas et al. (2013), Bastiaensens et al. (2014), Ridgeway (2011), Chua et al. (2016), Levy and Levy (2016), Hardaker and McGlashan (2015), Jane (2016)	<ul style="list-style-type: none"> • Social media are required to take measures to counter online abuse • Victims to be empowered and perpetrators to be discouraged • Women are used as sexual satisfying things by the men in IS • Education to control bullying and cyberbullying • Women are dedicated caretakers and men are facility providers • A do-it-yourself approach to counter hate content 	There are fewer analyses and measures to empower victims (Women) and discourage perpetrators
Religious Hate content	(Awan, 2016), (Tornberg and Tornberg, 2016), (Rudra et al., 2016), (Rahman et al., 2016) and (Awan and Zempi, 2016).	<ul style="list-style-type: none"> • Recent researches recorded that Muslims are the common targets for hatred expression • Communal tweets get more exposure (retweets) than situational tweets • Internet forums serve as an online amplifier by reflecting discourses in traditional networks • Educating policy on Islamic culture to avoid misinterpretation and confusion. 	Religious hate content analyses were made without considering its causes like clothing style and religionizing the nation
Racist Hate content	(Kwok and Wang, 2013), (Chaudhry, 2016), (Nakagawa and Arzubiaga, 2014), (Uemura, 2017), (Rivas-Drake et al, 2009), (Smalls, 2009), (Kwate et al., 2010), (West et al., 2010), (Stevens-Watkins et al., 2012), (O'hara et al., 2015), (Banerjee et al., 2011), (Lash, 2017), (Schaffner et al., 2017), (Roberts et al., 2013).	<ul style="list-style-type: none"> • After a racial attack tweets from the victim group will be more than the perpetrators' group • Racism leads to consumption of alcohol, drugs and poor academic performance by the victim • Coping with racism is essential to avoid depression • 82% of hate crimes are racist as recorded by the England police department in 2011 	Most of the researches have focused only on the hate crime victims and their views but not on the perpetrators or responsible persons for hate crimes

Disability Hate content	(Kudlick, 2003), (Hollomotz, 2013), (Stevens, 2011), (McCarthy, 2017), (Macdonald, 2017), (Alhaboby, 2016), (Roulstone, 2011), (Thiara, 2011), (Thomas, 2011), (Walter-Brice, 2012)	<ul style="list-style-type: none"> • Hate events with disabled people are more than the able-bodied. • Domestic violence is more on women than men • Disable hate events are underrepresented • Actual harassment is more than cyber harassment. 	Disable hate events are underrepresented and need to be represented appropriately
Hate content Detection Techniques	(Spertus, 1997), (Razavi et al., 2010), (Hardaker and McGlashan, 2015), (Warner and Hirschberg, 2012), (Ismail and Bchir, 2015), (Gitari et al., 2015), (Nobata et al., 2016), (Ross et al., 2017), (Burnap and Williams, 2015), (Kwok and Wang, 2013), (Sood et al., 2012), (Chen et al., 2012), (Djuric et al., 2015), (Fu et al., 2012), (Reynolds et al., 2011)	<ul style="list-style-type: none"> • Classified content as strongly hate, weakly hate or non-hate • Detailed content annotation is required instead of only yes or no • Decision Tree, Support Vector Machines, K- nearest neighbor, and Naïve Bayes are commonly used for the classification of hate content • Almost all works used online data itself. 	Most of the hate content detection techniques are performed based on the concept of hard classification but not based on the soft classification

As online hate is increased, a framework is needed to identify problems on hate speech and can be resolved by societal and moral responsibilities. Some of the possible policy approaches like speech vs. speech, education, hate-watch, content filters, blocking programs at work and school and responsible ISPs can be used to counter hate-related problems (Cohen-Almagor, 2011).

The General Theory of Crime (GTC) depicts that the lack of self-control causes the production of online hate content (Gottfredson and Hirschi, 1990). GTC extends its application to other behaviors such as the use of substances (Ford and Blumenstein, 2013), and bullying (Moon & Alarid, 2015). Pauwels and Svensson used GTC to explain offline extremism (Pauwels and Svensson, 2017). According to social structure-social learning (SSSL) theory, exposure to online hate content is a structural activity that brings an individual closer to the content and makes a favorite (Akers, 2009). SSSL theory supports different status and associations in society with four different components such as differential location, social location, differential association, and differential reinforcement.

2.5 An Act of Violence and Governance Policies for Hate Content

The awareness announcement of the developed new technologies by the Government and communication media alerts cybercriminals and cyber terrorists for developing approaches to attack new technologies. Nowadays, cyberterrorism is an everyday activity of terrorists and will continue with political, religious, or ideological support (Awan and Blakemore, 2012). Terrorism is referred to as an *upward crime*, with the perpetrator belonging to a lower social group compared to a targeted group, whereas hate crimes referred to *downward crimes*, in which perpetrators belong to a higher social or powerful group than the victim's group in the society (Deloughery et al., 2012).

Cyberterrorism is a special type of terrorism to harms physically or creates terror in the minds of the victim using electronic and internet tools. Terrorism or cyberterrorism is one of the causes for the occurrence of hate speech (Deloughery et al., 2012). Like cousins, hate crime and terrorism are very close (Mills et al., 2017). Generally, the generation and spreading of hate speech start by following a terror event. Cyberterrorism is a global issue that disturbs world peace and security. Terrorist organizations make use of existing cyberspace tools, like social networks and websites for posting illegal/harmful content, sharing information, and directing new members. As the nature of cyberterrorism is transactional, only local regulations cannot control its attacks, this requires a collective international solution.

Cyberterrorist networks are the kinds of social networks along with the features like high secrecy and hidden relationships of their members. Social Network Analysis (SNA) is an appropriate mechanism to analyze the nature of cyber terrorists using the concept of graph theory (Roberts and Everton, 2011). An appropriate method to counter transactional crime like cyberterrorism is to develop an international legal framework through multiple cooperating nations (Tehrani et. al, 2013). It is also essential to develop effective approaches for identifying cyber terrorist organizations, to know their hierarchical structures and operating strategies (Saidi et. al, 2017).

An act of violence, the narrative used as justification to attack, and messages to describe the way an organization achieves its goals are the languages used to communicate by terrorists (Wilkinson and Barclay, 2011). Terrorist organizations would-be members are persuaded using any or all of these languages. Terrorist group's positioning policy

is to create a death-centered militant with struggles in life for afterlife benefits (Baines and Shaughnessy, 2014). These organizations organize contests by providing rewards and prizes to a terrorist group with the best performance. Members of terror organizations will feel like they are in a contest and put their full effort to attain the target of more victims. With this style of attack, the number of victims in the current year will be more than the previous year (Caruso and Schneider, 2013).

The radicalization process is a very common terrorist activity and is difficult to understand its mechanisms. Hence, there is a necessity to identify and analyze the approach used by the radical Islamic groups to attract and polarize more disaffected individuals to block or reduce the spreading of extremism and terrorism (Torok, 2013). With the terrorist groups, women may be considered as active agents but voiceless victims. As their roles are situational and depend on multiple factors, cannot be judged based on the current situation. Generally, women are undervalued by terrorist organizations (Chatterjee, 2016).

Nowadays, most terrorist organizations are connected to the internet through the platforms such as chat rooms, electronic mail, forums, and tools like Twitter, Facebook, and Google Earth. Youth are the target of terrorist organizations for their criminal activities such as propaganda, incitement, and recruitment purposes (Weimann, 2010). Social networks have provided a digital platform to terrorist organizations, which enables cyber-based attacks through message dissemination (Bertram, 2016). Recently Twitter is considered a terrorist's favorite online tool to propagate criminal activities to online users and enable communication among members of terrorist organizations. The Twitter social network is promoting radical activities such as propaganda and recruitment of members around the world (Chatfield et. al, 2015).

Currently, there is less idea of the factors related to information propagation through social networks succeeding the terrorist events. Social features of a tweeter and content features of the tweets are the most influencing factors for heavy and long-lasting information flow respectively through Twitter social networks following a terrorist attack (Burnap et. al, 2014).

A social media is maliciously used to spread fake images of the Hurricane Sandy 2012 disaster as rumors, which created more panic in people associated (Gupta et. al, 2013).

During the terrorist event, people will use online social networks for exchanging information, uniting against the event, pressurizing the government to punish the actors, collecting funds for recovery of victims, etc. After a terrorist attack in Kenya at Westgate shopping mall, an analysis is made for the number of tweets, geographic location of tweets, response (tweet, retweet, or reply) of users in developing countries, reach, and impression of the tweet (Ishengoma, 2013).

The majority of users use online social networks for good services like interactions and information exchange but a fraction of users may misuse them for terrorism. This small set of terrorists will have the worst impact on several people around the world, such as a 9/11 terrorist attack on the USA's international trade center. Recruiting members and retaining sympathizer's loyalty, planning attacks and sharing information, gathering intelligence, training for specific attacks, raising funds for maintenance, propagating fear to enemies, and engaging in counterintelligence are the online social networks activities in favor of terrorists. Some of the online social networks activities against terrorist groups are flagging based on keywords, analyzing the sentiment, honeypots, and facial recognition (Mahmood, 2013).

Young people's exploitation approaches such as child-abusing, solicitation, and radicalization over the internet are common in nature (Quayle and Taylor, 2011). Cyberspace is a convenient and suitable venue for terrorist groups for their criminal activities. Social networks companies should know their social responsibilities like, how to filter and handle the content of terrorist groups. To manage this social responsibility it is required to have internet censorship (through ISPs and social networks companies) and the government's proactive measures to prevent terrorist groups from the dissemination of information (Bieda, 2015).

Terrorism and the Internet are important international phenomena, which reflect and shape different views of world politics (Conway, 2007). Evan Kohlmann cyber-terrorism expert argued that, in present days, 90 percentages of terrorist activities are carried out using social media over the Internet (Noguchi, 2006). These tools safeguard the identities of participants and facilitate contacting terrorist representatives, asking questions, and contributing towards helping the cyber jihad (Weimann, 2014). All terrorist organizations are using YouTube, Twitter, Facebook, and Instagram as online

platforms for their activities. Terrorists use social media to perform the activities such as spreading propaganda, psychological motivation, and tutoring weapons usage (Weimann, 2014).

As the policies to counter online hate are not adequate, online hate is enjoying its presence on social media platforms. Though the intermediaries are having the freedom from select governments to eliminate online hate content, they are tolerating its presence for human rights. The governance of hate content can be made by eliminating the content, using counter speech, educating and empowering the communities (Citron and Norton, 2011). Internet policymaking is constrained by social and cultural factors and structural factors of the internet itself. In India, internet policymaking is complex as it involves multiple stakeholders and discussions at different levels (Agur et al., 2015). Internet policymaking requires stringent participation by the concerned government (Baird, 2002). Overall, there are no adequate policies to govern the internet and online hate content (Agur et al., 2015; Baird, 2002; Best, 2004; Citron and Norton, 2011).

2.6 Research Gaps

Based on the literature review, the research gaps identified have been shown in Table 2.6.

Table 2.6: Listing of research gaps

Authors	Research Gaps	Remarks
Agur et al. (2015), Citron & Norton (2011), Baird (2002), Best (2004)	Inadequate governance policies for hate content	Suggestions are made to improve governance policies
Matsuda (2018)	Fewer studies on hate content in the Indian context	Studies conducted in the Indian context
Ridgeway (2011), Kwok and Wang (2013), Macdonald (2017)	Not many studies detected hate content influencing factors	Hate influencing factors are identified in this research
Simons (2015), Barlow and Awan (2016), Casas et al. (2013), Bastiaensens et al. (2014), Ridgeway (2011), Chua et al. (2016), Levy and Levy (2016), Hardaker and McGlashan (2015), Jane (2016)	There are fewer analyses and measures to empower victims and discourage perpetrators	Based on the results, empowerment approaches are presented
(Awan, 2016), (Tornberg and Tornberg, 2016), (Rudra et al., 2016), (Rahman et al., 2016) and (Awan and Zempi, 2016).	There are few studies with the emphasis on the causes of religious hate	The current research considered the causes of religious hate

(Kwok and Wang, 2013), (Chaudhry, 2016), (Nakagawa and Arzubiaga, 2014), (Uemura, 2017), (Rivas-Drake et al, 2009), (Smalls, 2009), (Kwate et al., 2010), (West et al., 2010), (Stevens-Watkins et al., 2012), (O'hara et al., 2015), (Banerjee et al., 2011), (Lash, 2017), (Schaffner et al., 2017), (Roberts et al., 2013).	Few studies have focused on the hate crime perpetrators or persons responsible for hate crimes	Emphasis on hate crime perpetrators is considered
(Kudlick, 2003), (Hollomotz, 2013), (Stevens, 2011), (McCarthy, 2017), (Macdonald, 2017), (Alhaboby, 2016), (Roulstone, 2011), (Thiara, 2011), (Thomas, 2011), (Walter-Brice, 2012)	Disable hate events are underrepresented	Needs to be encouraged to report
(Spertus, 1997), (Razavi et al., 2010), (Hardaker and McGlashan, 2015), (Warner and Hirschberg, 2012), (Ismail and Bchir, 2015), (Gitari et al., 2015), (Nobata et al., 2016), (Ross et al., 2017), (Burnap and Williams, 2015), (Kwok and Wang, 2013), (Sood et al., 2012), (Chen et al., 2012), (Chen et al., 2012), (Djuric et al., 2015), (Fu et al., 2012), (Reynolds et al., 2011)	Most of the hate content detection techniques are performed based on the concept of hard classification but not based on the soft classification	With soft classification, the coverage of the analysis will be more and needs to be applied

After identifying these research gaps, the current study tries to fill the gaps to a certain extent by addressing the issues around online hate content in the Indian context.

2.7 Concluding Remarks

The review of gender-based hate speech indicates that abuse and harassment against females disappoint them and as a consequence, they move towards terrorist organizations to join as a member. Based on the clothing style of a person, people will think that he belongs to a particular religion and start hating him verbally after a man-made or natural disaster event. Racist hate speech takes place concerning the natural appearance of a person and the consequence is minority groups will feel very bad about their natural status. Hate speech on persons with disabilities leads to more disability of the victim in terms of mental and physical conditions. Women with disabilities are more vulnerable to hatred attacks than able-bodied women. A review showed that terrorist attacks lead to the generation and propagation of hate speech over the internet. Human behavior can be predicted by analyzing social networks contents following terrorist events.

Cyber-terrorist networks contain features like high secrecy and hidden relationships of their members. Different languages like an act of violence, the narratives, and messages are used by the terrorists to describe an incident or to persuade would-be members of their organization. A terrorist organization intends to create death-centered militants

with benefits in the afterlife. With terrorist groups, women are undervalued and considered differently based on the situation. Youth are used for criminal activities by terrorist organizations. Only a fraction of users misuse the benefits of social networks, which results in tremendous loss to society with life and property-related threats.

The overall conclusion is, the existence of online social networks led to increases in features such as contact establishment, message exchange, information sharing, and news posting with the penalties such as 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.

CHAPTER 3

RESEARCH DESIGN

3.1 Introduction

In the previous chapter, a literature review on different categories of hate content and related terms is made. The literature review is a base on which the constructs of this chapter are designed. The chapter comprises a research question, research objectives, conceptual framework, context and scope of the study, and research methodology. The chapter emphasizes the necessity of the current study in India and the usage of mixed-method for analysis. The primary data for the present study is collected from internet users using a questionnaire designed for the purpose. The secondary data is collected from the social media giant Twitter for case studies.

The rest of the chapter is structured as follows. Section 3.2 outlines a research question and section 3.3 briefs the research objectives. A conceptual framework of the research work is provided in section 3.4. In section 3.5, the context and scope of the study are presented. Section 3.6 discusses the research methodology involving questionnaire design, justification for mixed methods approach and social media data, and different case studies. In the end, the research design is concluded in section 3.7.

3.2 Research Question

From the literature, though the possible mechanisms to control hate content are present, the dissipation of hate content over social media is prevalent. Therefore, the following research question is set to study the online hate content and its control.

Research question: How to improve digital hate content regulations in the Indian context?

3.3 Research objectives

To answer the above research question, the following research objectives are set.

Research objective 1: To study the nature and forms of digital hate content in the Indian context.

Understanding the nature and different forms of digital hate content is essential to crafting a strategy to control it. The different forms such as gendered hate, religious hate, racist hate, and disability hate can be studied under this objective.

Research objective 2: To understand the influence of digital hate content on online safety.

The hate content often affects the safety of online users. Therefore, the study on how hate content impacts the safety of online users is essential.

Research objective 3: To study policies and regulations addressing digital hate content.

The previous two objectives establish the ecosystem of online hate content. The policies and regulations play an important role in mitigating hate content. Therefore, the understanding of existing policies to control online hate content is necessary.

Research objective 4: To bring out recommendations for improving digital hate content regulations in India.

After knowing the ecosystem of hate content and existing regulations to control it, possible recommendations can be made to improve the regulations in India to control online hate content.

3.4 Conceptual Framework

With the help of a literature review, different dependent and independent variables related to hate content research have been identified. The conceptual framework containing different forms of hate content, their characterizing attributes, and the relationship among them is shown in Figure 3.1.

Dependent variable: This variable is dependent on the other independent variables which are measured through several indicators. *Online Hate content (OHC)* is a form of communication that disparages people based on their affiliation with a particular identifiable group.

Independent variables: This category of variables is independent and is measured with the help of instruments like questionnaires, surveys, etc. Based on the literature survey, four hypotheses have been formed involving different independent variables.

Gender hate content (GHC): This hate is exhibited based on gender. The measurement variables for gender hate content are alienated, inclination, and personal experience (Campbell, 1981).

H1: There is a positive association between gender hate content and online hate content.

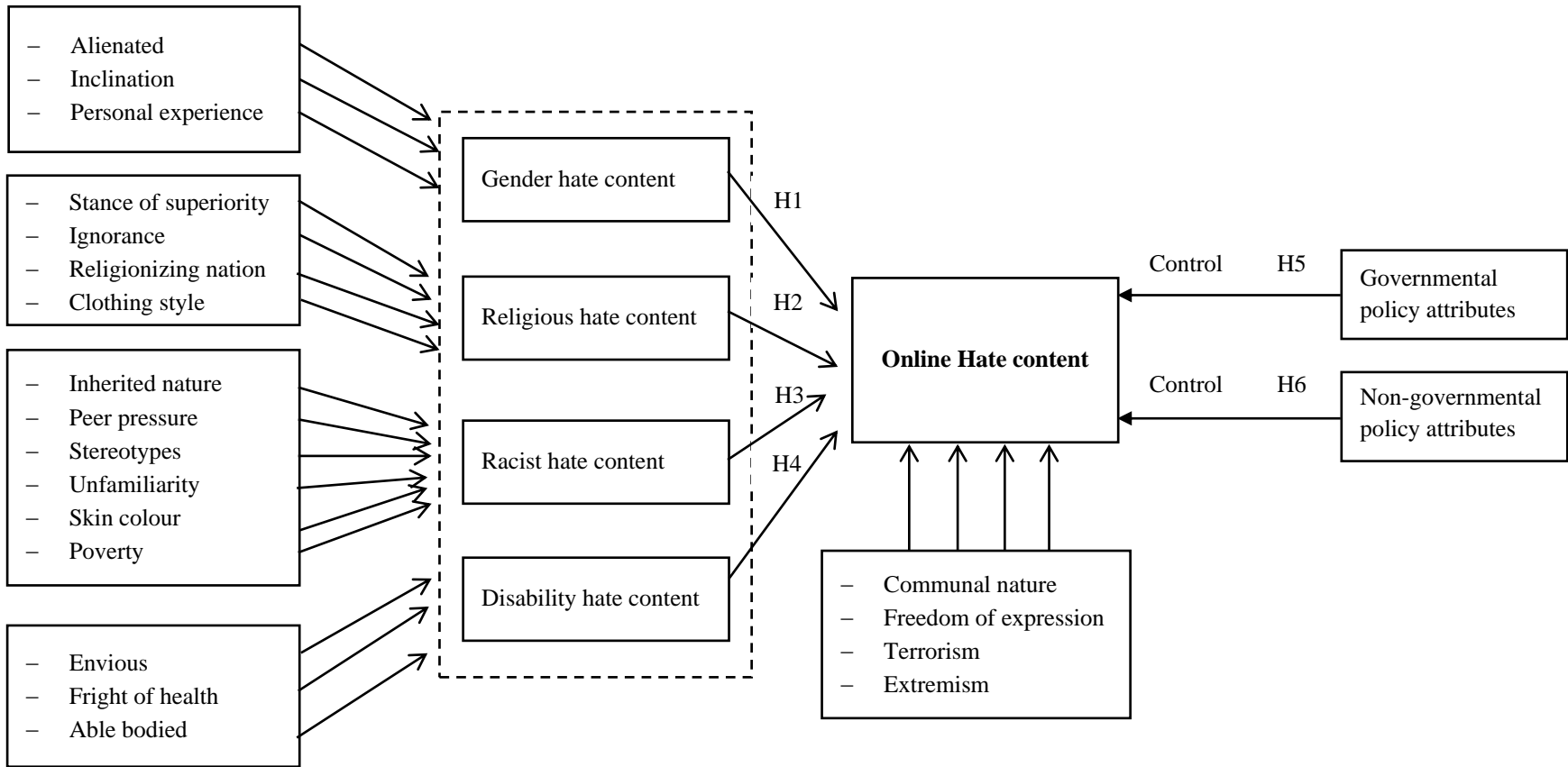


Figure 3.1: A conceptual framework for hate content analysis

Religious hate content (ReHC): This kind of hate is expressed against other religious persons. Because of the group of people associated with religion, an expression of hate content against the religious sentiments is more impactful than an individual. *The stance of superiority, ignorance, religionizing nation and clothing style of a person* may be the measurement indicators of this hate content (Rahman et al., 2016; Mulligan, 2018).

H2: There is a positive association between religious hate content and online hate content.

Racism or Racist hate content (RaHC): Racism is a discriminatory practice with the support of social groups and can be institutionalized (Wodak and Reisigl, 1999, p.181). The measurements such as *inherited nature, peer pressure, personal experience, stereotypes, unfamiliarity, skin color, and poverty* may form racism leading towards online hate (Jones, 2013; Catholic Charities, 2008; Miller and Garran, 2017; Banton, 1992; Crawford, 1998).

H3: There is a positive association between racism and online hate content.

Disability hate content (DHC): This kind of hate is expressed against persons with a mental or physical disability. Disability is not only a medical entity; it is also a social category like other protected characteristics such as race and gender. The indicators such as *ignorance, envious, the fright of health, and being able-bodied* may form disability hate content leading towards online hate (Wendell, 1989; Hannon, 2007).

H4: There is a positive association between disability hate content and online hate content.

H5: There is a positive association between governmental policy attributes and the reduction of online hate content (Breen and Nel, 2011; Dharmapala and Garoupa, 2004).

H6: There is a positive association between non-governmental policy attributes and the reduction of online hate content (Gagliardone et al., 2015; Henry, 2009).

Apart from these indirect measurement variables, the online hate content can be measured with the manifest variables such as *communal nature, freedom of expression, terrorism, and extremism* (Hanes & Machin, 2014; Heyman, 2008; Olteanu et al., 2018; Qin et al., 2011).

3.5 Context and Scope of the Study

The study concerns governance, and regulations of digital content in general, hate content in particular. Hate content spreads rapidly and affects the social life of the citizen by establishing a hatred nature in the surrounding environment. Hate content originates and dissipates over the Internet. The scope of the present study includes but is not limited to the following.

- Online users, who use different online applications for communication, entertainment, education, news, and shopping purposes.
- The state and central government authorities, whose purpose is to provide a hassle-free environment to the citizen.
- The policymakers, who frame policies and regulations to act against the perpetrators' post-hate content incidents.
- The educational institutes, which want to maintain an unbiased environment on the campus by controlling hate content.
- Public and private organizations, who want to provide a healthy environment to their employees.
- The parents, who care about the future of their children and want to keep them away from unwanted contacts which spread hate content.

3.6 Research Methodology

As a part of the research methodology, the set of procedures used in the study are briefed. In this research work, a mixed-methods approach is used and shown in Figure 3.2. Mixed methods research collect, analyze, and interpret data with the help of both quantitative and qualitative methods as a single study to answer a research question. The quantitative analysis is made on the data collected using a questionnaire through both online and offline modes. Social media data on hate content is analyzed using a qualitative method.

In both the analysis, the data on online hate content based on four identities such as gender, religion, race, and disability is analyzed. The questionnaire survey data is processed and analyzed using a statistical and predictive approach. A content analysis approach is used for social media data analysis. In the end, the outcome of both

approaches are integrated, interpreted, discussed and conclusions are drawn from the discussion.

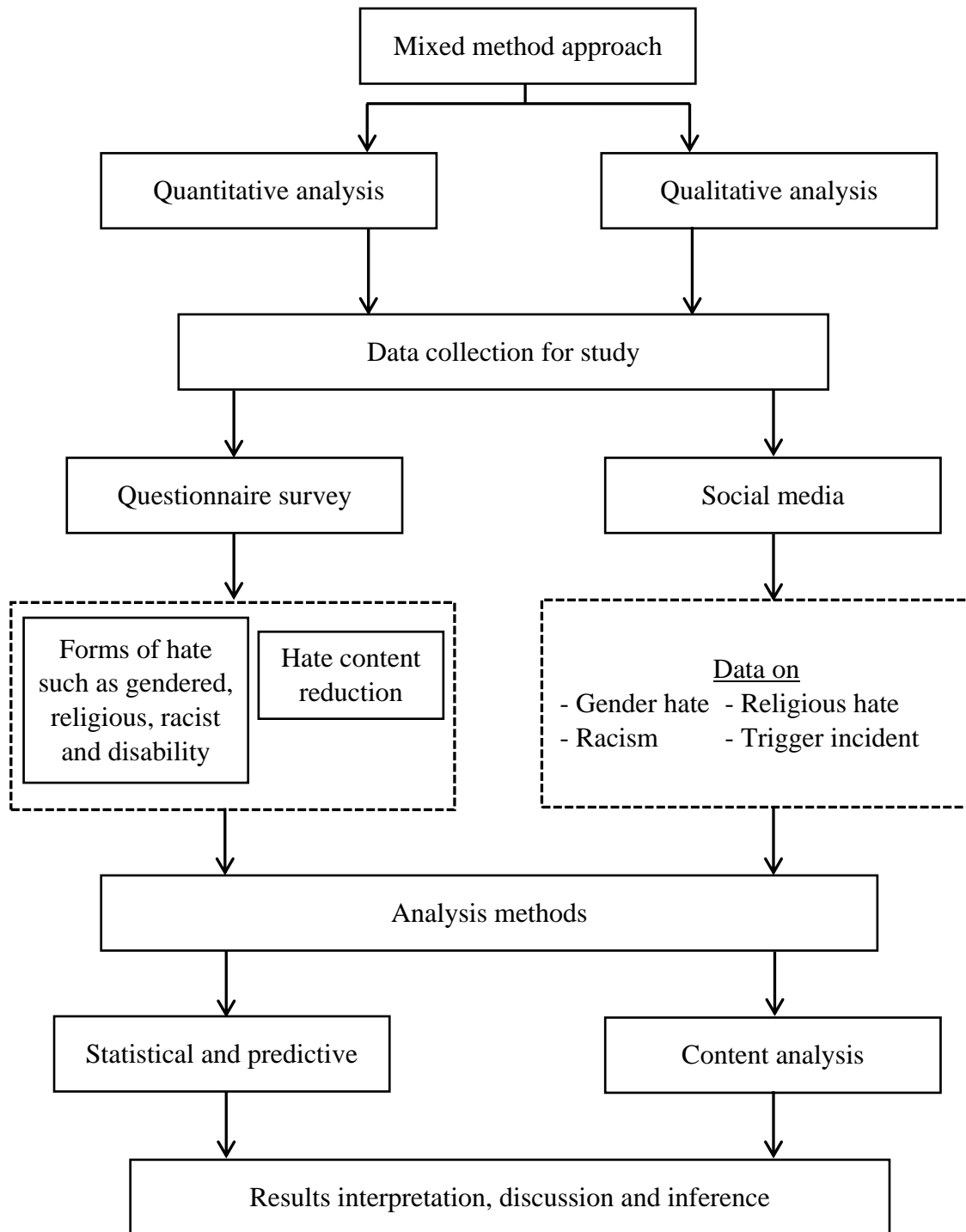


Figure 3.2: Research Methodology

3.6.1 Justification for Mixed Methods Research

The qualitative methods suffer from a lack of objectivity and generalizability whereas quantitative methods suffer from a lack of participant's voice and meaningful interpretation. As the weaknesses of a method can be overcome by the strengths of another method, many researchers moved towards mixed methods research (McKim, 2017). The mixed-method approach uses both "qualitative and quantitative analysis methods to collect data for the study, analyze, integrate results and draw inferences from the integrated result" (Tashakkori and Creswell, 2007, p 4).

Mixed methods add value to findings by increasing validity, supporting secondary data sources, and knowledge creation. The mixed-methods approach provides a better understanding of the situation than studies that use only quantitative or qualitative approaches (Hurmerinta-Peltomaki and Nummela, 2006). Articles that used mixed methods are cited more than the studies without mixed methods (Molina-Azorin, 2011). This increased citation adds more value to the mixed methods approach. This research poses a broader research question demanding the analysis of data containing numbers, words, symbols, and pictures. The use of a mixed-methods approach is based on the research problem rather than the research topic.

The research question which is required to answer in the present study is "How to improve digital hate content regulations in the Indian context?" The question can be answered with detailed knowledge on influencing factors of hate content, how it dissipates, and through which platforms it passes. Therefore, in the current study, the simultaneously mixed methods (quantitative + qualitative) approach in which quantitative is for foundation and qualitative is for complementary information is used. The quantitative analysis is performed on the data which is collected using the questionnaire survey. Social media is used as a secondary source of data. The case studies are conducted to supplement the results of quantitative research. Further, more justification for the mixed methods research is supported by the following requirements.

The inefficiency of a single data source: The use of data source either quantitative or qualitative for research results in either general or detailed analysis of a problem in hand respectively. In quantitative analysis, many individuals will be analyzed but an in-depth

understanding of individuals' views will be missed. The qualitative analysis supports a detailed understanding of an individual against the problem at hand but with a restriction to examine limited individuals which results in poor generalization of the result. The research should provide detailed and generalizable results. The issue can be addressed better with multiple shreds of evidence than single evidence. Therefore, to overcome the limitations of one approach with the strengths of another approach and to provide a comprehensive outcome, a mixed-methods approach is adopted in the current study.

To support and extend initial results: In some cases, the results of a study may be incomplete and requires detailed explanations. The outcome of the first phase of a study may be determining the influencers of some events or acts. In the second phase of a study, these influencers can be elaborated. In the present study, the quantitative analysis is performed to determine the influencers of online hate content and policies to control it whereas the qualitative analysis is applied to explore their presence and forms.

3.6.2 Questionnaire Design and Administration

As the target respondents of the study are online users, the questionnaire is developed in the English language assuming that all respondents understand the language. The questionnaire has been designed to make the respondents convenient and enthusiastic during the process of answering the questions. Though the study under consideration is attractive, to the maximum extent nonthreatening words are used and organized to form pleasant questions. The sensitivity of the questions is in increasing order from the beginning. To make respondents feel that the survey is emphasizing the study under consideration, the questionnaire is started with questions governing the study, and respondents' demographic-related questions are located in the end. The phrases such as social media and the internet are used in the questionnaire to depict the dependent variables.

While designing the questionnaire, possible efforts are made to keep the length minimum by incorporating necessary items of the study. The questionnaire contains questions that do not require more effort to recall and answer. An outline of the questionnaire is provided in Table 3.1. The questionnaire involves mostly the ranking type questions and some open-ended questions. The questionnaire consists of a set of

questions on causes of different hate content, the aftermath of hate, hate content reduction, the purpose of Internet use, victimization, blocking while using the Internet, and demography of respondents.

Table 3.1: Basic constructs of a questionnaire

Category of Questions	Sample Question	The scale used to measure
Causes to online hate content	Extremism is one of the causes for hate content occurrence on the Internet	Five-point scale from strongly disagree to strongly agree
Causes to gender hate content	Expression of hate is common on/by the divorced person	
Causes to religious hate content	Hate is because of a lack of religious knowledge	
Causes to racist content	Racial hate is because of exposure to stereotypes on racism	
Causes of disability hate content	Disability hate is because of jealousy on disabled person's benefits	
Causes of online hate content	A terrorist attack on any public event/gathering leads to hate expression on the Internet	
Aftermath of hate	Online gendered hate leads to change of name by the victim	
Hate content reduction	Strengthen legislation relating to hate content expression will reduce hate content	
Purpose of Internet use	I use social networking sites (such as Facebook and Twitter) frequently	
Victimization	I/my friends/my relatives harassed with gendered hate content	
Blocking while using the Internet	I was blocked by organizational level security	
Demography of respondents	How do you describe your gender?	<input type="checkbox"/> Male <input type="checkbox"/> Female <input type="checkbox"/> Transgender (Other)

The use of social media by college students is increasing. The occurrence of hate speech with college students is prevalent (Saha et al., 2019). Cyberbullying, a form of hate content exists among college students (Kowalski et al., 2012). Therefore, the college students and social media users are identified as target respondents to answer the questionnaire. With this view for collecting the data through offline mode, the educational institutes are approached. The questionnaires are distributed to both

students and faculty and collected back after filling it. For the convenience of answering, the key constructs of the questionnaire/related terms are briefed to the respondents inside the classroom. For wider coverage of the nation, offline data has been collected from different educational institutes of the country.

For collecting the data through online mode, a Google document is created containing all the questions of the questionnaire. Three different online mechanisms such as E-mail, Twitter, and Facebook are used to collect data in online mode. An E-mail request to fill out the questionnaire with the Uniform Resource Locator (URL) link has been sent to select persons. To ensure their response, a request message for confirmation is appended at the end of the E-mail message. The respondents, who filled out the questionnaire, have been confirmed back.

Twitter is one of the top ten most visited sites in the world with the use of short and fast communication (Wolny, 2016). By logging into the Twitter account, the URL link to the questionnaire is tweeted along with a request to fill the questionnaire.

Facebook is more advantageous than the other constituents of social media brands (Stoycheff et al., 2017) and supports the social presence of the user by providing real-time communication among the users (Cheung et al., 2011). It acts as a platform for diverse viewpoints with an authentic conversation over it (Weedon et al., 2017) and enables the well-being of its users (Burke and Kraut, 2016). Facebook is one of the most popular networking sites with 1,871 Million users as of January 2017. Facebook users are almost double that of its immediate competitor YouTube. Due to its instant communicative nature, support for diverse viewpoints, and more number of users, the Facebook social networking site is selected and used as a primary platform for online data collection.

By logging into the Facebook account, a personal request for filling the questionnaire with the Uniform Resource Link (URL) has been sent to active Facebook users. Request messages are sent to users during their active/on state on Facebook instead of off state. To identify the active users, the Facebook account is accessed throughout the clock during the period of data collection. During data collection, it has been observed that in the late evening, users are more than the other parts of the day. As the users will be active on Facebook at different parts of the day based on the nature of their job, the

request messages are sent to the users at different times to collect the data from different kinds of users. Some of the users responded immediately and some other took time and few more requests to respond. As most of the online users are active in the late evening, this has been observed as an effective time for online data collection.

The appropriate numeric values are entered into the excel sheet for collected data. Value labeling has been done in an excel sheet or code sheet according to the codebook procedure. The labels or questions without any responses or missing values are filled with the value 999.

3.6.3 Justification for Studies using Social Media Data

Social media platforms are a prevalent mode of communication and share information quickly. Social media are important in enabling an individual to express freely and interact with the entire world without bothering about the physical location and social status (Sawyer & Chen, 2012; Amedie, 2015). Every business or organization is dependent on social media for its growth. Social media is a set of online platforms which supports communication among the users by generating and exchanging digital content. The use of social media makes an individual innovative and creative (Correa et al., 2010). Social media is based on the social presence and media richness theories which indicate the interaction time and reduction of uncertainty or conflicts respectively (Kaplan and Haenlein, 2010). Twitter is a commonly used social media application that supports the collection and analysis of business data (Culnan et al., 2010), disaster data, health data (Berry et al., 2017), etc.

The darker side of social media is which may affect negatively the behavior of some individuals in making them antisocial (Amedie, 2015). Increased reporting of cybercrimes indicates that the existing laws are less effective to control them (Tripathi, 2017). Online social media provides support to express free and hate speech (Sangsuvan, 2013). Like free speech, other social rights such as equality, dignity, liberty, harass free life, mutual respect, and individual's name/honor are required to protect under the constitutional law for maintaining the justified balance of all human rights (Baer, 2009).

The preceding discussion indicates that social media have a lot of benefits and some issues for humanity. Issues such as online abuse, demeaning expressions, and hate content are often possible through social media applications. Few studies have analyzed online hate content using social media applications such as Facebook, Twitter, and YouTube.

Twitter is commonly used social media to express the feelings of an individual. As there is a limitation on text size, the feelings are required to be expressed wisely. Hence the tweets contain acceptable information and are convenient to process and analyze. Online users consider that Twitter social media is a safer space to express their feelings (Berry et al., 2017). This feeling of safety may increase the expression of problematic content such as online abuse, demeaning expressions, and hate content. Therefore, the present study considered Twitter social media content as secondary data to address the research question.

3.6.4 Case Studies

After analyzing online hate content quantitatively, to provide in-depth information on the identified individual entities, the case studies on gender hate, religious hate, racism, and terrorist acts are conducted.

3.6.4.1 Case Study on Gender-based Hate

Both gender and religious hate are the highest contributors to online hate content as resulted from the quantitative analysis. To know further how this hate content influences online safety, two studies have been conducted to analyze gender and religious hate using Twitter social media. These studies are based on social reforming movements and honor-based violence. Social reforming movements attempt to empower the disadvantaged people in a society. The purpose of this study is to identify the impact of social-reform initiatives on gender and religion-based hate content generation.

Twitter sentiments for government and citizen-initiated movements are analyzed for emotions using an algorithm developed for the purpose. The score and contributing words for each emotion are examined, and a comparison of both the reforming movements is made. The citizen-initiated reforming movement shares more hate

content than the government-initiated reforming movement. The emotions like anger, disgust, and sadness towards gendered social-reform initiatives are more in citizen initiatives than the government.

The nature of emotions observed can be used in policymaking to maintain gender equality and control hate content generation. Social-reform initiatives should also focus on strategies to rebuild citizen initiatives in a hate-free environment. The citizen-initiated reforming movements support the sharing of online hatred information with high magnitude. This sharing of hate with high magnitude may result in reduced hate content in the future. There are fewer studies reported in the Indian context to assess the role of social reforming movements on gender hate content generation. This study analyzed online information on Twitter through social reforming movements and attempted to address the problem.

3.6.4.2 Case Study on Religious Hate

Honour is a kind of respect associated with an individual or a family or a community in a society. Honor-based hate content is predominantly generated from family hate content and may affect people. The purpose of this study is to identify and understand the influencing factors and emotions of honor-based hate content. A review of literature, news articles on honor killing, and the analysis of Twitter content are made to attain the purpose. In India, factor like marrying a person against family members' ideologies is observed as dominating among other factors of honor-based hate content. It has been also observed that emotions such as anger, fear, disgust, and sadness are used to express hate. Possible impacts of honor-based hate content on family and society are discussed.

3.6.4.3 Case Study on Racism

Racism is a form of hatred and expressed against the ethnicity and skin color of the target. From the quantitative analysis using a questionnaire survey, it is identified that the presence of racism influences online hate content. Therefore, further, to examine the presence of racism in India a case study is conducted using Twitter content. The data is extracted from Twitter social media through an application programming interface (API). To accommodate possible tweets and make the results general, the Twitter API is queried with different keywords to extract tweets.

For convenient and proper classification of tweets as racist, neutral and non-racist, first, the tweets are pre-processed to remove some unwanted information. After pre-processing, a classification algorithm that is written using the sentiment analysis concept is applied to the Twitter content. The classification results are represented graphically and some of the tweets are analyzed for their polarity. The results showed that the Twitter content contains all three types of tweets such as racist, neutral, and non-racist. The keen observation of some tweets revealed that even though some tweets contain problematic words, still, they are non-racist.

3.6.4.4 Case Study on Influence of Physical Violence on Hate Content

The problematic act or terrorism damages the target and seeds fear in the neighborhood. Social media sites are used for planning and coordinating problematic acts. The problematic act is a trigger event that influences hatred feeling. The objective of the case study is to analyze the aftermath of a recent problematic incident in the southern part of the Asian continent from Twitter content. After the problematic incident, citizens used to share their views over social media sites. The opinions shared through Twitter social media are collected and analyzed with the help of software. The results show hatred against the problematic act through different related emotions. Results also contain more negative tweets which are almost thrice the positive tweets. Fear and anger emotions exhibit a high degree of emotions than the other.

3.7 Concluding Remarks

Digital content over the internet is affected by the reduced cost of IT products and the destructive behaviors of online users. Based on the literature review, a research question on the improvement of prevention measures for online hate content is framed. To answer the research question, the objectives involving nature, forms, influence, policy challenges, and governance architecture for hate content are set. A conceptual framework is drawn with influencing factors for hate content based on different protected identities. Increased population, democratic presence, British colonial policy, the occurrence of hate crimes, low ranking for international indices, and a high score of SHI made India the context of the present study. A mixed-methods approach is proposed for a deeper understanding and analysis of the problem with justification. Different case studies are conducted to support and explore primary analysis results.

CHAPTER 4

ONLINE HATE CONTENT: A SURVEY AND ANALYSIS

4.1 Introduction

After outlining the research design with methodological discussion in the previous chapter, this chapter deals with the analysis and hypothesis testing. Social media sites provide convenient interaction among the users irrespective of their background, and location (Amedie, 2015; Kapoor et al., 2018; Sawyer & Chen, 2012). The negative side of social media use may affect human behavior resulting in antisocial activities (Amedie, 2015), depression (Shensa et al., 2017), and generation of problematic content. Some of the problematic contents are harassment, pornography, hate speech (Gillespie, 2018), and fake news (Caplan et al., 2018). The expression of hatred through social media is a prevalent and serious issue because of the difficulty to provide barriers (Mondal et al., 2018) and rapid amplification (Jubany & Roiha, 2016) to hate content.

To reduce the complexity and increase the understandability, the conceptual model is divided into two parts. The first part of the model is concerned with the ecosystem of online hate content and the second part deals with the reduction of online hate content. The data for both the model testing is collected at the same time using a single questionnaire that demands the responses of online hate content drivers and controllers.

The rest of the chapter is organized as follows. Section 4.2 presents a survey analysis for identifying different determinants of online hate content. The policy perspective analysis is presented in section 4.3 and section 4.4 concludes the chapter.

4.2 Online Hate Content: Nature, Forms and Hypotheses

The objective of this quantitative data analysis is to identify the determinants of online hate content by understanding the nature, forms, and dispersion of online hate content. To attain this objective, the data has been collected from 716 internet users using the research instrument designed for the purpose both online and offline. A comprehensive analysis is made using partial least squares path modeling (PLS-PM) package from an open-source software R. By analyzing a structural model for different protected characteristics, the gendered hate content, religious hate content, racist hate content, and disability hate content are identified as determinants of online hate content. Further, the analysis of measurement models revealed the role of measuring indicators in terms of contribution to the determinants.

The hate speech/content is bias motivated with hostile and malicious nature and exhibited against a group or an individual. This content is expressed by considering the innate characteristics of the victim such as gender, religion, race, or disability (Cohen-Almagor, 2011; Seglow, 2016). Summarization of literature towards hypotheses design is made in Table 2.3. The remark on the hate content category column indicates that the group of researchers emphasized their work on a particular identity. Based on the detailed literature review on gender, religious, racist and disability hate content in Chapter 2, the following hypotheses have been set.

H1: There is a positive association between gender hate content and online hate content.

H2: There is a positive association between religious hate content and online hate content.

H3: There is a positive association between religious hate content and online hate content.

H4: There is a positive association between disability hate content and online hate content.

4.2.1 Research model

Dependent variable: This category of the variable is dependent on the independent variables which are measured through several indicators. *Hate content (HC)* is a form of communication that disparages people based on their affiliation to a particular identity group. The research model with hypotheses is shown in Figure 4.1. The online hate content is characterized by the terrorist attack, freedom of expression, communal nature and extremism of an individual.

Independent variables: These categories of variables are independent and are measured with the help of instruments like questionnaires, surveys, etc. Different types of independent variables are gender hate content, religious hate content, racist hate content and disability hate content.

Hypotheses H1-H4 shows the association between the hate content against different protected characteristics and online hate content.

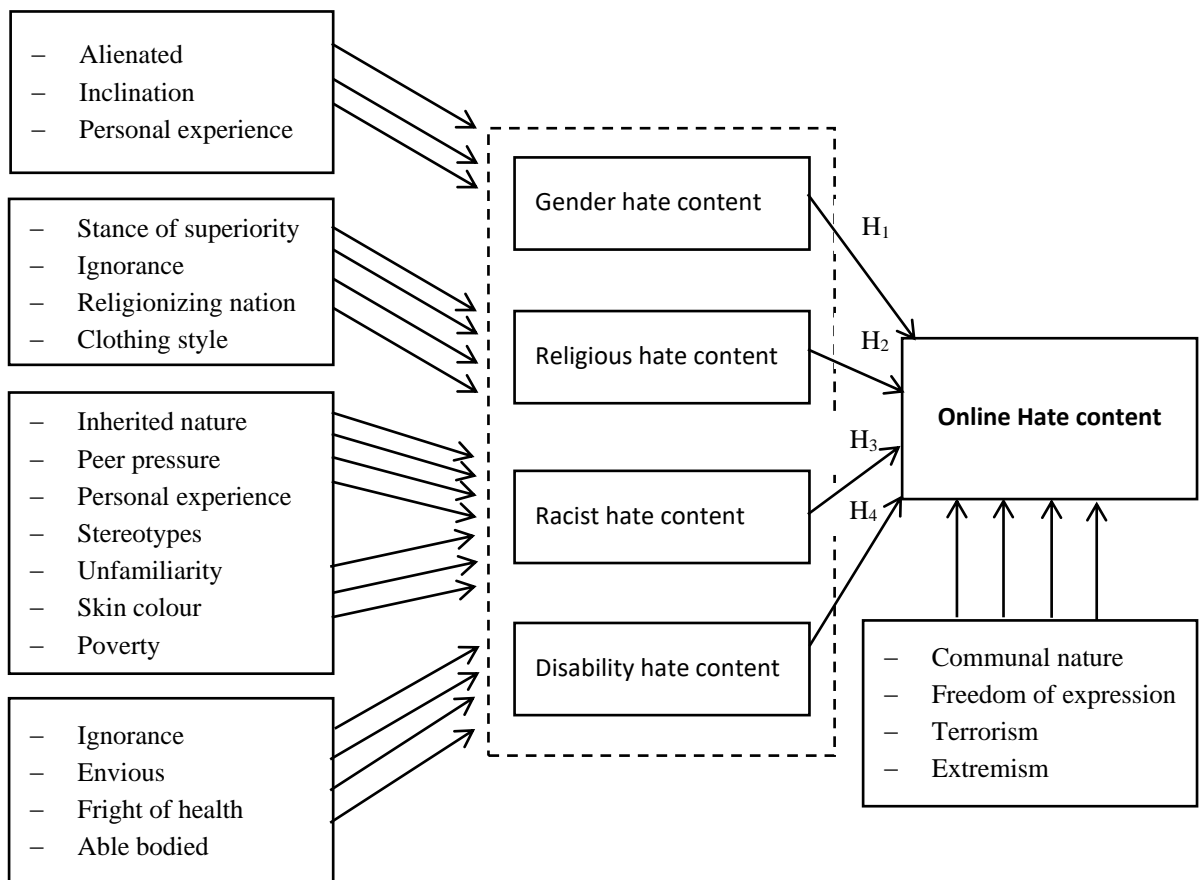


Figure 4.1: A Conceptual Framework to identify determinants of hate content

Methodology

The different stages carried out to identify the determinants of online hate content are shown in Figure 4.2. The questionnaire used to collect views from different respondents about online hate content is given in Appendix. The detailed data collection approach is discussed in Chapter 3.

As this work involves multilevel causal/formative model fitting, structural equation modeling (SEM) is adopted for the study. Unlike the regression technique, SEM takes into account the relationship between the number of variables simultaneously and the measurement errors (Schumacker and Lomax, 2004). The variance-based SEM, i.e. partial least square SEM (PLS-SEM) is used to develop the model. The structural model drawn using the PLS-PM package is shown in Figure 4.3.

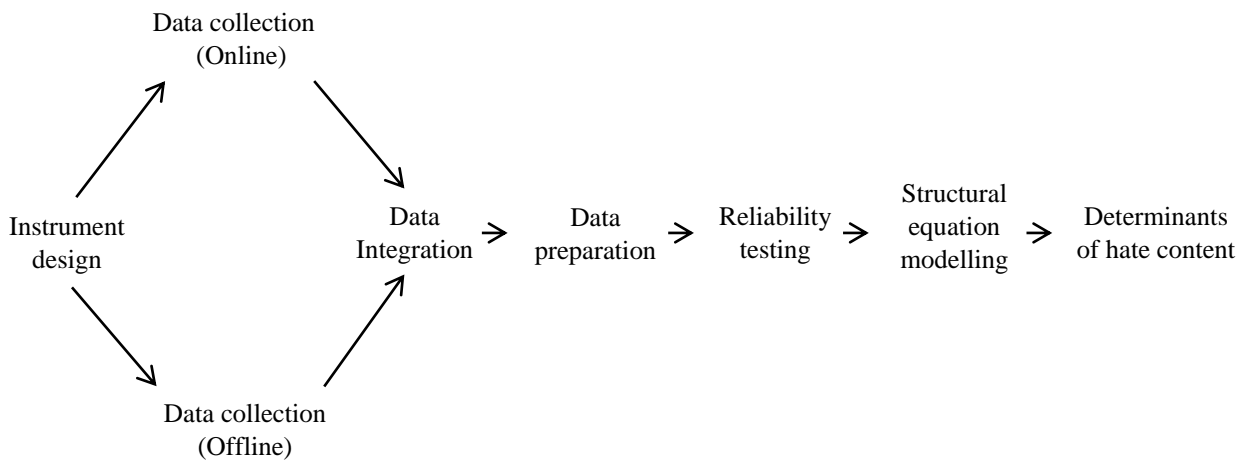


Figure 4.2: Steps to identify determinants of hate content

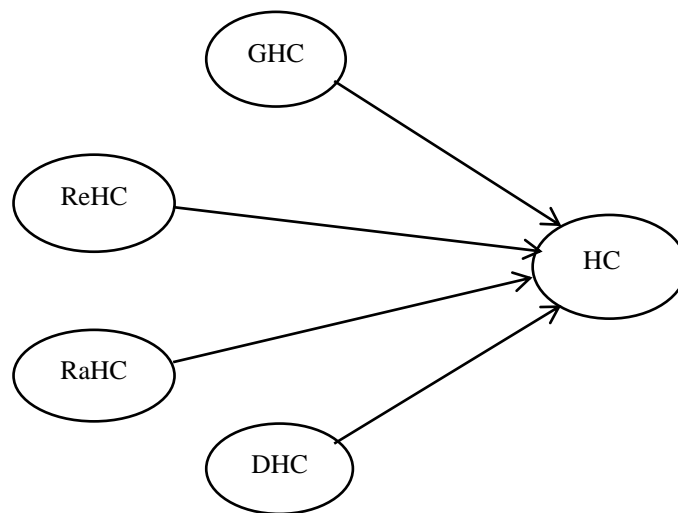


Figure 4.3: Structural model

The model is developed in R software using the partial least square-path modeling (PLS-PM) package. The model involves two parts such as structural/inner model and measurement/outer model. The different variables used for measuring the formative constructs for the study are described in Table 4.1. As these are the formative measurements, they can contribute to more than one latent variable. The *personal experience* of being the victim of hatred is considered for both gendered and racist hate content. Similarly, *ignorance* is considered for religious and disability hate content. To develop a structured model the causal effect has to be represented in a square matrix.

Table 4.1: List of measurement variables used in the study

Name of indicator	Description/purpose
Communal nature	The communal nature of a person may lead to hate content
Terrorist attack	The occurrence of a terrorist attack may lead to hate content
Freedom of expression	Freedom to express anything may lead to hate content
Extremism	The extremist behavior of a person may lead to hate content
Alienated	Gendered hate content may be more on/from a divorced person
Inclination	The inclination is an attitude of a person and may lead to gendered hate
Personal experience-1	Hate experienced person may exhibit more gendered hate
Stance of superiority	Being prouder about the own religion may lead to hatred against other religions
Ignorance-1	Lack of knowledge about the different religious nature may lead to religious hate
Religionizing the nation	An intention of religionizing the nation may lead to religious hate
Clothing style	The clothing style of a person may lead to religious hate
Inherited nature	The inherited nature of a person may lead to racism
Peer pressure	Pressurization from friends/colleagues may lead to racism
Stereotype	Exposure to stereotypes may lead to racism
Unfamiliarity of ethnicity	The lack of familiarity about ethnicity and traditions may lead to racism
Personal experience-2	Racism experienced person may be racist
Skin color	The skin color of a person may lead to racism
Poverty	The poverty of a person may lead to racism
Ignorance-2	Lack of knowledge about the disability and related diseases may lead to disability hate
Envious	Jealousness of a person may lead to disability hate
Fright of health	Fear of health may lead to disability hate
Able-bodied	Being proud of the non-disability may lead to disability hate

Table 4.2 shows the square matrix with the values 0 and 1. The causes of this matrix will be read from columns to rows.

Table 4.2: Structural model in square matrix format

	GHC	ReHC	RaHC	DHC	HC
GHC	0	0	0	0	0
ReHC	0	0	0	0	0

RaHC	0	0	0	0	0
DHC	0	0	0	0	0
HC	1	1	1	1	0

For a matrix of “i” rows and “j” columns, it is read that column “j” causes row “i” wherever the entries of the cell contain the value 1.

4.2.2 Reliability and Validity of the Survey Instrument

First, a pilot study is conducted to ensure the reliability/validity of the research/survey instrument. The Cronbach’s alpha is tested against initially collected 40 samples using a statistical package for the social sciences (SPSS) software. The Cronbach’s alpha values for all the different categories of hate content (latent variables) are shown in Table 4.3.

Table 4.3: Reliability testing of the survey instrument

Variable name	No. of items	Alpha value
Gender hate content	3	0.72
Religious hate content	4	0.79
Racist hate content	7	0.89
Disability hate content	4	0.69
Hate content	4	0.76

As all Cronbach’s alpha values are equal to or above 0.7, the survey instrument is considered reliable.

4.2.3 Sample Characteristics

A sample is a smaller collection of units from a population used to determine truths about that population. In the present study, a probability-based simple random sampling method is used to select respondents for quantitative analysis. Simple random sampling provides a chance of consideration for every item or event in the population.

Demography of Respondents

The purpose of analyzing the demographics of respondents is to know the characteristics of the selected population. The characteristics of any item or human

being play an important role in policymaking for the purpose. The gender and age-related information of the respondents are shown in Table 4.4. Gender and age information is gathered separately for both online and offline respondents. The genderwise respondents in the study are male, female, and transgender with their participation percentages of 61.03, 38.12, and 0.28 respectively. Similarly, the participation percentage of respondents with the age span 15-19 years, 20-29 years, 30-39 years, 40-49 years, and above 40 years is 5.17, 78.91, 12.43, 2.23, and 0.70 respectively. The unknown term indicates that gender and age information is unavailable for some of the respondents.

Table 4.4: Gender and age of respondents

Gender/Age	Online Response		Offline Response		Total	
	Respondent	Percentage	Respondent	Percentage	Respondent	Percentage
Male	276	68.14	161	51.77	437	61.03
Female	127	31.36	146	46.95	273	38.12
Transgender	2	0.50	0	0	2	0.28
Unknown	0	0	4	1.29	4	0.56
15-19 Yrs.	5	1.23	32	10.29	37	5.17
20-29 Yrs.	300	74.07	265	85.21	565	78.91
30-39 Yrs.	84	20.74	5	1.61	89	12.43
40-49 Yrs.	14	3.46	2	0.64	16	2.23
>49 Yrs.	2	0.50	3	0.96	5	0.70
Unknown	0	0	4	1.29	4	0.56

The native state or union territory of the respondents is shown in Table 4.5. The respondents are from almost all the major states and union territories of India. Most of the respondents are from Karnataka, Uttarakhand, and Uttar Pradesh states covering the nation from south to north.

Table 4.5: State-wise respondents

State/Union Territory	Online Respondent	Offline Respondent	Total Respondent
Andhra Pradesh	6	14	20
Assam	0	1	1
Bihar	7	2	9
Delhi	17	4	21
Goa	3	1	4
Gujarat	5	2	7
Haryana	2	0	2

Himachal Pradesh	3	0	3
Jammu and Kashmir	4	1	5
Jharkhand	0	1	1
Karnataka	90	105	195
Kerala	23	16	39
Madhya Pradesh	7	7	14
Maharashtra	9	14	23
Odisha	0	5	5
Punjab	1	0	1
Rajasthan	2	6	8
Sikkim	1	0	1
Tamil Nadu	4	13	17
Tenangana	0	8	8
Tripura	0	1	1
Uttarakhand	75	49	124
Uttar Pradesh	49	7	56
West Bengal	0	5	5
Total	308	262	570

Tools and Response of Respondents

Understanding the usage of technology by the respondents is also essential to obtain more insights for policymaking. Therefore, the Internet and device usage by the respondents are gathered and shown in Table 4.6. The percentage of respondents with 0-3 hours, 4-6 hours, 7-9 hours, 10-20 hours, 21-40 hours and more than 40 hours of Internet usage per week is 7.74, 10.86, 15.33, 25.45, 22.17, and 18.45 respectively. Similarly, the percentage usage of electronic devices such as simple phone, smartphone, tablet, and the laptop is 1.70, 81.84, 1.13 and 15.32 respectively for Internet access by the respondents.

Table 4.6: Tools and Internet used by the respondents

Internet Used/week			Tool Used		
Used Hours	Respondent	Percentage	Tool Name	Respondent	Percentage
0-3 Hrs.	52	7.74	Simple phone	12	1.70
4-6 Hrs.	73	10.86	Smartphone	577	81.84
7-9 Hrs.	103	15.33	Tablet	8	1.13
10-20 Hrs.	171	25.45	Laptop	108	15.32
21-40 Hrs.	149	22.17	---	---	---
>40 Hrs.	124	18.45	---	---	---

The more part of the Internet is accessed by the respondents who will be online for 10-20 and 21-40 hours per week. Similarly, most of the respondents are online through their smartphones.

The genderwise response of respondents to different categories of hate content is shown in Table 4.7. The genderwise percentage of responses to different categories of hate is calculated from the data collected through questionnaire survey. The response of female respondents on gender and disability hate content is around two percentages more than the response of male respondents.

Table 4.7: Respondents response to hate content

Hate category	Percentage of Male respondents	Percentage of Female respondents
Gender hate	38.67	41.02
Religious hate	48.28	61.17
Racist hate	42.56	46.89
Disability hate	19.68	21.61

About religious hate content, female respondents responded 13 percent more than the response of male respondents. Similarly, the responses on racist hate content are also more by female respondents than the male respondents.

4.2.4 Survey Sample and Justification

As a part of the data collection, first, the sample size required for the study is identified. According to the Internet and mobile association of India, the internet users in India are 481 million as of December 2017 (IAMAI, 2018). The sample size of the study has been determined using the following Slovin's formula (Tejada & Punzalan, 2012).

$$n = N / (1 + N \times e^2)$$

Where n is a sample size, N is total population and e is the error margin. With a 95% of confidence level, the resulted error margin is 5% to determine the sample size for the study.

$$n = 481 \text{ million} / (1 + 481 \text{ million} \times 0.05^2) = 400$$

According to this, a minimum of 400 samples is required to generalize the work. The samples collected for the present study are 716, which are almost double the required minimum number of samples. Alam et al. (2016) are used a sample size of 200 to determine the attitude of Indian people towards hate speech on social media. With this information, it is determined that collected samples are sufficient to carry out the present study.

4.2.5 Results and Discussions

The entire model is evaluated with measurement and structural models.

Assessment of measurement models

As the measurement model constructs are formative, the model is evaluated only for unidimensionality and contribution of the measuring indicators. Unidimensionality is evaluated by reporting the values of Cronbach's alpha and principal component analysis. Cronbach's alpha reports the average correlation among the measuring indicators of a construct. To illustrate homogeneity among the indicators, the Cronbach's alpha value to be 0.70 or more. The principal component analysis mechanism is another approach to access unidimensionality. To indicate the unidimensionality of a set of variables, the first eigenvalues should be larger than 1 while the subsequent ones should be smaller than 1. Cronbach's alpha and principal component analysis values for evaluating the unidimensionality are shown in Table 4.8. As required the Cronbach's alpha for all the constructs is greater than the specified value 0.7 and the first Eigenvalues are also greater than 1. Hence the measurement model is evaluated as unidimensional.

Table 4.8: Factors and unidimensionality

Factor	Mode	Type	Size	C. alpha	Eigen value 1 st	Eigen value 2 nd
GHC	Formative	Exogenous	3	0.72	1.38	0.96
ReHC	Formative	Exogenous	4	0.79	1.58	1.00
RaHC	Formative	Exogenous	7	0.89	1.93	1.07
DHC	Formative	Exogenous	4	0.69	1.57	0.90
HC	Formative	Endogenous	4	0.76	1.86	0.80

The measurement model can also be evaluated by accessing the contribution of individual indicators to a concerned latent variable. Table 4.9 reports the contribution of each measurement variable to its related construct in the form of weight. The GHC latent construct is measured with the variables such as alienated, inclination, and personal experience-1. The contributions from alienated, inclination, and personal experience-1 to GHC are 14%, 56%, and 63% respectively. These results indicate that the divorced status of a person, an attitude of a person, and the personal experience of being the victim of hatred all play a role in generating GHC.

Table 4.9: Measurement model evaluation factors

Latent variable	Indicator	Weight	VIF
GHC	Alienated	0.14	1.06
	Inclination	0.56	1.36
	Personal experience-1	0.63	1.35
ReHC	Stance of superiority	0.44	1.30
	Ignorance-1	0.51	1.40
	Religionizing the nation	0.43	1.28
	Clothing style	0.05	1.09
RaHC	Inherited nature	0.11	1.13
	Peer pressure	0.19	1.18
	Stereotype	0.36	1.29
	Unfamiliarity of ethnicity	0.58	1.33
	Personal experience-2	0.03	1.07
	Skin color	0.26	1.20
	Poverty	0.10	1.08
DHC	Ignorance-2	0.78	1.31
	Envious	0.26	1.19
	Fright of health	0.32	1.18
	Able-bodied	0.09	1.17
HC	Communal nature	0.33	1.27
	Terrorist attack	0.48	1.43
	Freedom of expression	0.22	1.20
	Extremism	0.40	1.38

The variables stance of superiority, ignorance-1, religionizing the nation, and clothing style of a person are indicators and contribute to ReHC 44%, 51%, 43%, and 5% respectively. This indicates that being proud about the own religion is a strong contributor, lack of knowledge about the religious traditions is a very strong

contributor, an intention of religionizing the nation is also a strong contributor, and clothing style of a person is the least contributor to the generation of ReHC

The RaHC measuring indicators such as inherited nature, peer pressure, personal experience, stereotypes, the unfamiliarity of ethnicity, skin color, and poverty are having the prediction weights 11%, 19%, 36%, 58%, 2% 26%, and 10% respectively. These weights indicate that personal experience, stereotypes, and skin color contributes to the generation of racist hate content more than the other indicators. The contribution of inherited nature of a person, peer pressure, and poverty associated with a person towards the RaHC is moderate, whereas the contribution of the unfamiliarity of ethnicity is least.

The variables ignorance, envious, fright of health, and being able-bodied are measuring indicators and contribute to DHC 78%, 26%, 32%, and 9% respectively. This indicates that the lack of knowledge about the disability and associated diseases contribute strongly to the generation of DHC. The other indicators jealousy and fear of health contribute moderately whereas the feeling of being able-bodied contributes least to DHC.

The indicators such as communal nature, terrorist attack, freedom of expression, and extremism resulted in weights 33%, 48%, 22%, and 40% respectively. Here the terrorist attack and extremism indicators contribute more to the generation of hate content than others, whereas freedom of expression contributes less. The contribution of the communal nature of a person is closer to the contribution of extremism. The graphical representation of the measurement models is shown in Figure 4.4.

The personal experience indicator is associated with more than one latent variable such as GHC and RaHC. As discussed above it contributes 63% to GHC and 36% to RaHC. This indicates that the victim of gendered and racist hate content may be the perpetrator of the respective hate content in the future. The reason behind the gendered hate may be to take revenge by one sex over another if the former is already a victim of the same. Similarly, the same reason may hold for racist hate also to take the revenge by one ethnic/tradition group on another.

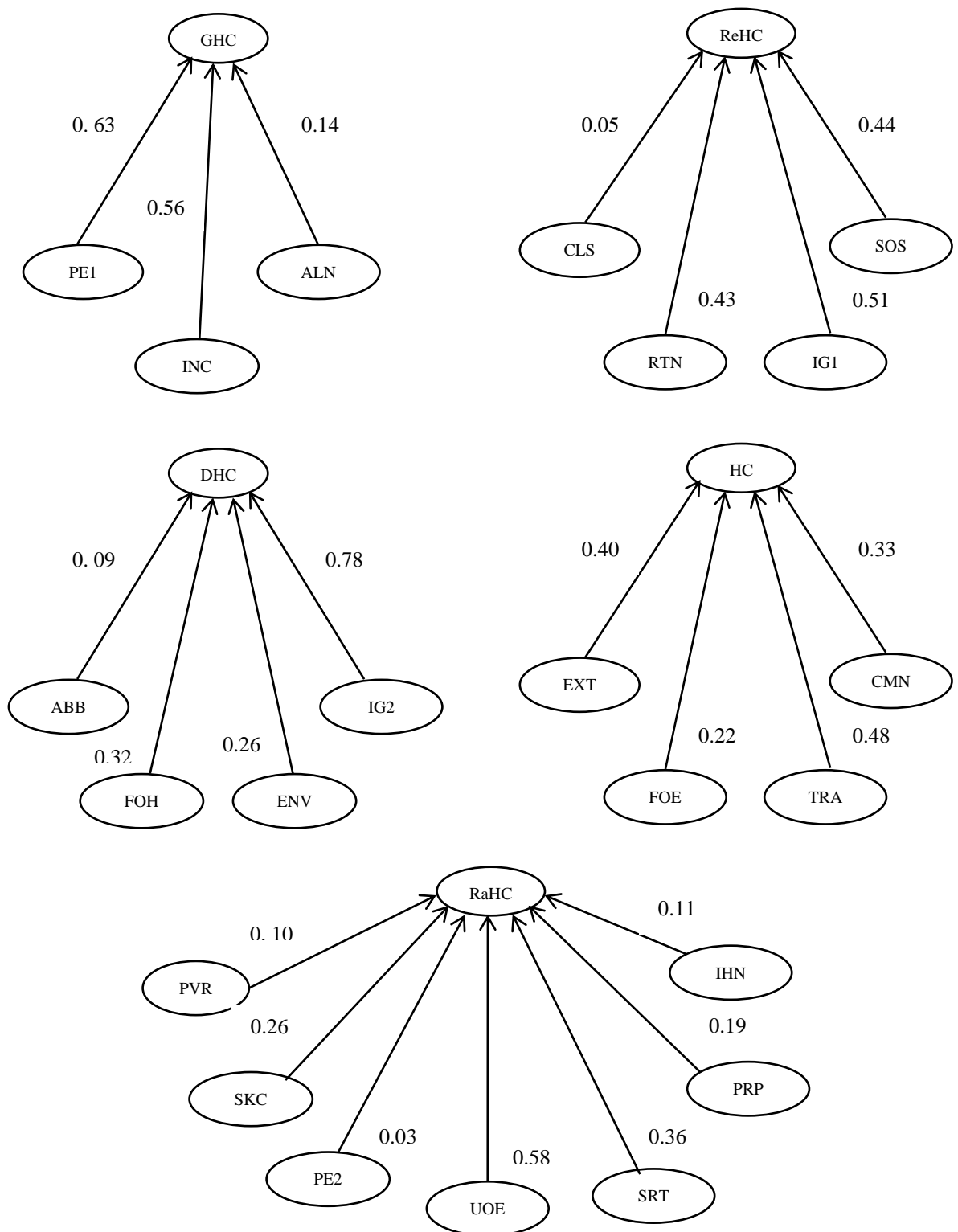


Figure 4.4 Measurement models

Legends: ALN: Alienated, INC: Inclination, PE1: Personal experience-1, SOS: Stance of superiority, IG1: Ignorance-1, RTN: Religionizing the nation, CLS: clothing style, IG2: Ignorance-2, ENV: Envious, FOH: Fright of health, ABB: Able-bodied, CMN: Communal nature, TRA: Terrorist attack, FOE: Freedom of expression, EXT: Extremism, IHN: Inherited nature, PRP: Peer pressure, SRT: Stereotype, UOE: Unfamiliarity of ethnicity, PE2: Personal experience-2, SKC: Skin colour, PVR: Poverty.

Similarly, ignorance indicators are also supporting more than one latent construct such as ReHC and DHC. It contributes 51% and 78% to ReHC and DHC respectively. In both cases, the contribution is more and indicates that the lack of knowledge on religious traditions and disability with associated diseases plays a major role in generating religious and disability hate content.

Colinearity is another mechanism to assess measurement models. The variance inflation factor (VIF) is used for checking the collinearity among the measuring variables. VIF value for all the indicator variables is less than 2 and is far away from the value 10 which is required to indicate the presence of colinearity among the variables.

The direction and magnitude of the edges represent cause and weight respectively. For example, an arrow on an edge (Alienated, GHC) indicates that alienated /divorce causes GHC, and the weight on this indicates the contribution of alienated to GHC is 14%.

Assessment of structural model and hypothesis testing

An evaluation of structural models can be performed by analyzing the results of coefficients of determination/ R^2 , the fitness of good, and the regression. R^2 reports the quantity of variance in the dependent latent variable due to its independent latent variables. The coefficient of determination for the model is 0.38 and is shown in Table 4.10. Concerning the determination coefficients, the model is moderate as R^2 lies between 0.30 and 0.60 (Sanchez, 2013) and is substantial/high as it exceeds 0.26 (Cohen, 1998).

Table 4.10: Coefficient of determination

Block	Type	R^2
Gender hate content	Exogenous	0.00
Religious hate content	Exogenous	0.00
Racist hate content	Exogenous	0.00
Disability hate content	Exogenous	0.00
Online hate content	Endogenous	0.38

The other index to assess the model is the goodness of fit (GoF). GoF is a global criterion to measure the overall (both measurement and structural) quality of the model. In another way, it provides the prediction rate of the entire model. Usually, there is no threshold of GoF for its statistical significance but some authors suggested the possible range of values. The GoF achieved for our model is 0.37 which is just above the cut-off 0.36 specified for substantial/large models (Wetzels et al. 2009).

The correlations between the latent constructs such as GHC, ReHC, RaHC, DHC, and HC are shown in Table 4.11.

Table 4.11: Correlations between the latent variables

	Gender hate content	Religious hate content	Racist hate content	Disability hate content	Online hate content
Gender hate content	1.00				
Religious hate content	0.48	1.00			
Racist hate content	0.30	0.42	1.00		
Disability hate content	0.34	0.44	0.40	1.00	
Online hate content	0.47	0.50	0.43	0.39	1.00

Results of path analysis for the structural model are shown in Table 4.12. These analysis results are obtained by executing the `path_coefs()` function with the object created over the PLS-PM package in R software. In path analysis, a predictor with a low P-value indicates that the variations in the predictor variable affect the response variable. A path $GHC \rightarrow HC$ with coefficient 26.07 at $t\text{-value}=7.57$ and $p\text{-value}<0.001$ represents that the predictor GHC is significant and hence by rejecting the null hypothesis the alternate hypothesis H1 is accepted. To decide about hypothesis H2 a path $ReHC \rightarrow HC$ is observed. The observation reveals that the path $ReHC \rightarrow HC$ is significant with coefficient 23.12 at $t\text{-value}=6.25$ and $p\text{-value}<0.001$. Therefore by rejecting the null hypothesis the alternate hypothesis H2 is accepted.

Hypothesis H3 is tested by analyzing the path $RaHC \rightarrow HC$. The analysis with coefficient value 21.10 at $t\text{-value}=6.20$ and $p\text{-value}<0.001$ indicates that there is a

positive association between RaHC and HC. Hence the hypothesis H3 is accepted. Similarly, a path DHC→HC is analyzed and found that it is significant with coefficient value 11.40 at t-value=3.30 and p-value<0.001. Therefore hypothesis H4 is also accepted by rejecting the null hypothesis.

Table 4.12: Results of path analysis

Hypothesis	Path	Path coefficient	t-value	Result
H1	Gendered hate content →Hate content (GHC→HC)	26.07***	7.57	Supported
H2	Religious hate content →Hate content (ReHC→HC)	23.12***	6.25	Supported
H3	Racist hate content →Hate content (RaHC→HC)	21.10***	6.20	Supported
H4	Disability hate content →Hate content (DHC→HC)	11.40***	3.30	Supported

*** Significance at the P < 0.001 level

A structural model obtained after running the code with R software is shown in Figure 4.5. Each edge/path in the model is associated with magnitude and direction. Magnitude indicates the weight/coefficient of the predictor whereas direction indicates effect from predictor to response variable.

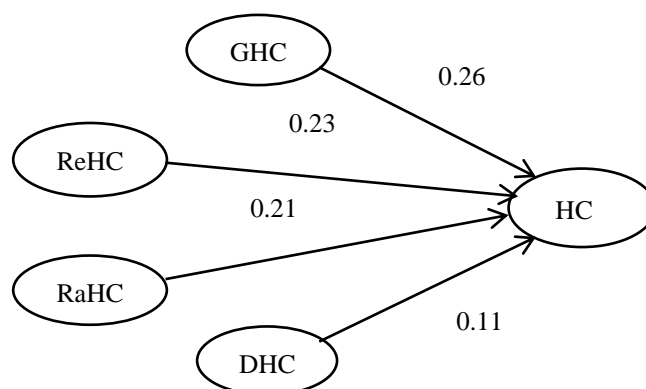


Figure 4.5: A structural model with weights

The effects of predictors such as GHC, ReHC, RaHC, and DHC on the response variable HC are shown in Table 4.13. The predictor GHC is having a 26% direct effect on HC without any indirect effect. Similarly, the direct effects of ReHC, RaHC, and

DHC are 23%, 21%, and 11% on HC without indirect effects respectively. As there is no indirect effect from any of the predictors on response variable HC, the total effect for each predictor is the same as its direct effect.

Table 4.13: Total effects of the relationships

Relationships	Direct effect	Indirect effect	Total effect
GHC → ReHC	0.00	0	0.00
GHC → RaHC	0.00	0	0.00
GHC →DHC	0.00	0	0.00
GHC → HC	0.26	0	0.26
ReHC → RaHC	0.00	0	0.00
ReHC →DHC	0.00	0	0.00
ReHC → HC	0.23	0	0.23
RaHC →DHC	0.00	0	0.00
RaHC → HC	0.21	0	0.21
DHC →HC	0.11	0	0.11

A discussion on effects indicates that whenever online hate content is present, there is a major role of gendered hate content. Religious and racist hate content play the next major roles in the generation of online hate content. Lastly, the disability hate content plays the least role in the generation of online hate content.

A complete path analysis from characterizing attributes of different hate content predictors to the online hate content is shown in Figure 4.6. The derived relationships/paths from measurement and structural models by applying transitive property are shown in Table 4.14. The transitive property indicates that if $x=y$ and $y=z$ then $x=z$. The derived paths indicate that the measuring variables of different categories of hate content indirectly affect the generation of online hate content.

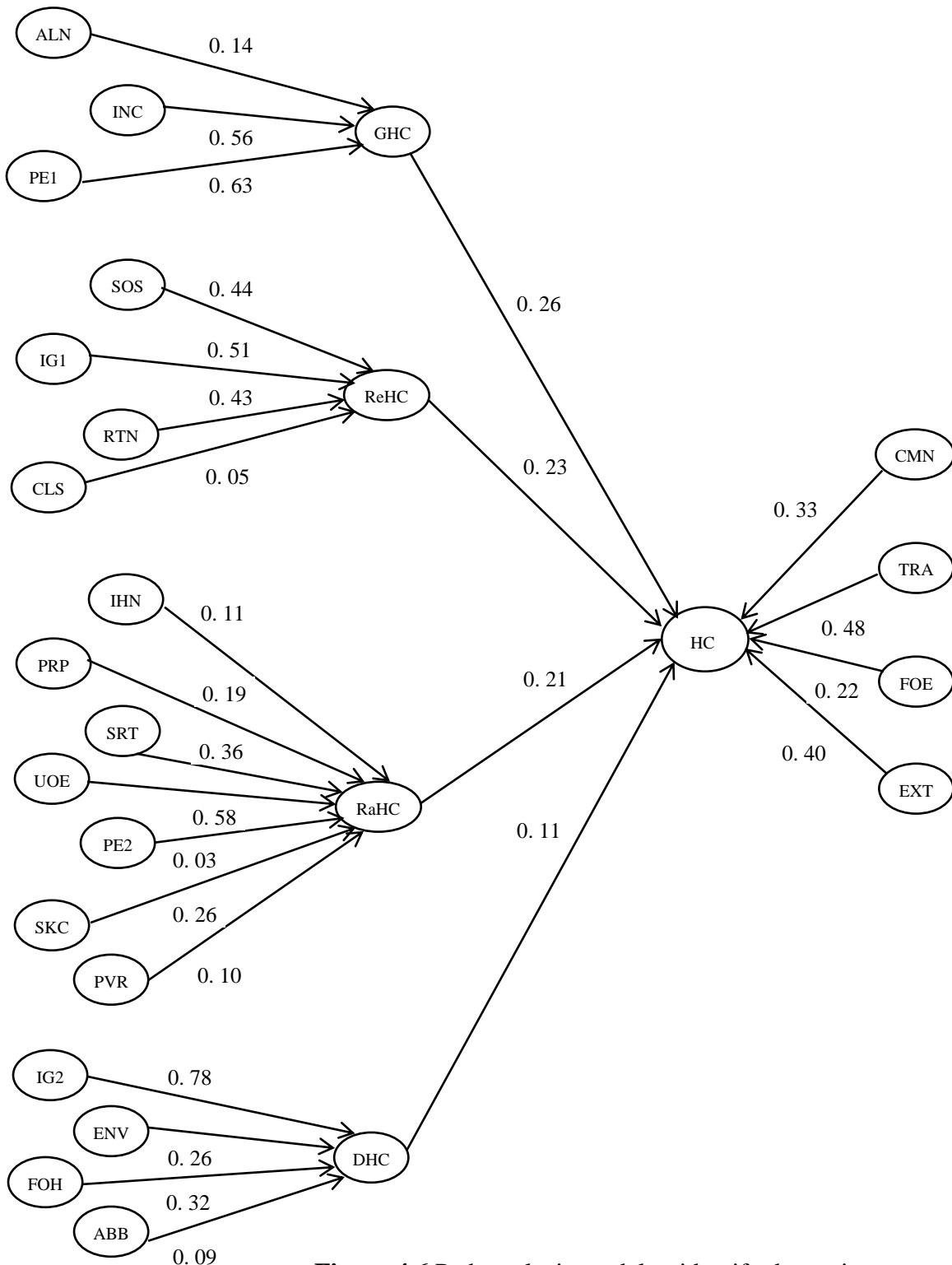


Figure 4.6 Path analysis model to identify determinants

Legends:

ALN: Alienated, INC: Inclination, PE1: Personal experience-1, SOS: Stance of superiority, IG1: Ignorance-1, RTN: Religionizing the nation, CLS: clothing style, IG2: Ignorance-2, ENV: Envious, FOH: Fright of health, ABB: Able-bodied, CMN: Communal nature, TRA: Terrorist attack, FOE: Freedom of expression, EXT: Extremism, IHN: Inherited nature, PRP: Peer pressure, SRT: Stereotype, UOE: Unfamiliarity of ethnicity, PE2: Personal experience-2, SKC: Skin colour, PVR: Poverty, GHC: Gender hate content, ReHC: Religious hate content, RaHC: Racist hate content, DHC: Disability hate content, HC: Hate content.

Table 4.14: Derived relationships from measurement and structural models

Paths from measurement models	Paths from a structured model	Paths derived by applying the transitive property	Remarks
Alienated → Gendered hate content Inclination → Gendered hate content Personal experience → Gendered hate content	Gendered hate content → Hate content	Alienated → Hate content Inclination → Hate content Personal experience → Hate content	In the generation of online hate content, alienated, inclination and personal experiences plays the role through gendered hate content
The stance of superiority → Religious hate content Ignorance → Religious hate content Religionizing the nation → Religious hate content Clothing style → Religious hate content	Religious hate content → Hate content	The stance of superiority → Hate content Ignorance → Hate content Religionizing the nation → Hate content Clothing style → Hate content	In the generation of online hate content, a stance of superiority, ignorance, religionizing the nation, and clothing style plays the role through religious hate content
Inherited nature → Racist hate content Peer pressure → Racist hate content Stereotype → Racist hate content The unfamiliarity of ethnicity → Racist hate content Personal experience → Racist hate content Skin color → Racist hate content Poverty → Racist hate content	Racist hate content → Hate content	Inherited nature → Hate content Peer pressure → Hate content Stereotype → Hate content The unfamiliarity of ethnicity → Hate content Personal experience → Hate content Skin color → Hate content Poverty → Hate content	In the generation of online hate content, inherited nature, peer pressure, stereotype, the unfamiliarity of ethnicity, personal experience, skin color, and poverty play the role through racist hate content
Ignorance → Disability hate content Envious → Disability hate content Fright of health → Disability hate content Able-bodied → Disability hate content	Disability hate content → Hate content	Ignorance → Hate content Envious → Hate content Fright of health → Hate content Able-bodied → Hate content	In the generation of online hate content, ignorance, envious, the fright of health, and being able-bodied plays the role through disability hate content

4.3 Analysis for Policy Perspectives

As the increased hate content causes several health issues, it is necessary to mitigate it. Hate content mitigation may attain sustainable development goals of the United Nations. Therefore, the objective of this analysis is to identify the possible policies to mitigate online hate content. In this regard, online and offline opinions from the Indian respondents are gathered through the questionnaire designed for the purpose. The analysis results reveal the role of government and non-government authorities for digital hate content reduction.

Regulation is a subset of governance and tries to control events flow and behavior. Governance is meant to “provide, distribute and regulate” (Braithwaite et al., 2007). As the traditional technologies of digital content communication are already regulated separately, it is essential to regulate digital content, in particular, digital hate content. Anonymity and different legal jurisdiction in the online world made it difficult to track the originator of the digital hate content. In India, the possible strategies for controlling hate content are helplines, internet disconnection, and counter-speech (Arun and Nayak, 2016).

4.3.1 Model and Methodology

With the help of a literature review outlined in Chapter 2, different dependent and independent variables related to hate content regulations have been identified. These variables are used in model design.

Dependent variable: Hate content reduction (HCR) is identified as the dependent variable and can be measurable by governmental and no-governmental policy attributes.

Independent variables: governmental and non-governmental policy attributes are identified as independent variables and can be measured through the questionnaire surveys. Strengthening legislation, increasing penalties, and campaigns by the government are the contributing factors to governmental policy attributes (Breen and Nel, 2011; Dharmapala and Garoupa, 2004). Campaigns by private sectors, campaigns by civil societies, and efforts by social media are the contributing factors to non-

governmental attributes (Gagliardone et al., 2015; Henry, 2009). With this information following hypotheses have been set.

Hypothesis H5: Governmental policy attributes reduce online hate content.

Hypothesis H6: Non-governmental policy attributes reduce online hate content.

As discussed in Chapter 3, the data is collected from both online and offline modes. A conceptual model is designed using identified attributes. To ensure the strength and contribution of policy attributes to reduce hate content, the model is tested with the help of software developed using the R programming language. Structural equation modeling (SEM) takes into account the relationship between more than one variable and the measurement errors simultaneously (Schumacker and Lomax, 2004). The variance-based SEM, i.e., partial least square SEM (PLS-SEM) is used to develop the model. The model is developed in R software using the partial least square-path modeling (PLS-PM) package. The model involves two parts such as structural/inner model and measurement/outer model.

4.3.2 Results and Discussion

As the SEM involves two different sub-models such as measurement and structural models, the model evaluation also can be performed separately. The evaluation of the model for outer and inner models is presented below.

Measurement models assessment

These models are evaluated for unidimensionality and contribution of the measuring indicators. Cronbach's alpha and principal component analysis approaches are used to assess unidimensionality. Unidimensionality evaluation values are shown in Table 4.15. As Cronbach's alpha values are greater than the specified value 0.7, the first and second Eigenvalues are larger and smaller than 1 respectively, the measurement model is unidimensional.

Table 4.15: Factors and unidimensionality for policy perspectives

Factor	Mode	Type	Size	C. alpha	Eigen value 1 st	Eigen value 2 nd
GPA	Formative	Exogenous	3	1.0	3	0.000224
NPA	Formative	Exogenous	3	1.0	3	0.000260
HCR	Formative	Endogenous	2	1.0	2	0.000347

Further, the measurement models are assessed by measuring the contribution of individual indicators of a latent construct. The measurement models are shown in Figure 4.7. Each path of a model possesses magnitude and direction. Magnitude indicates the weight of the predictor whereas direction indicates the effect of the predictor on the response variable.

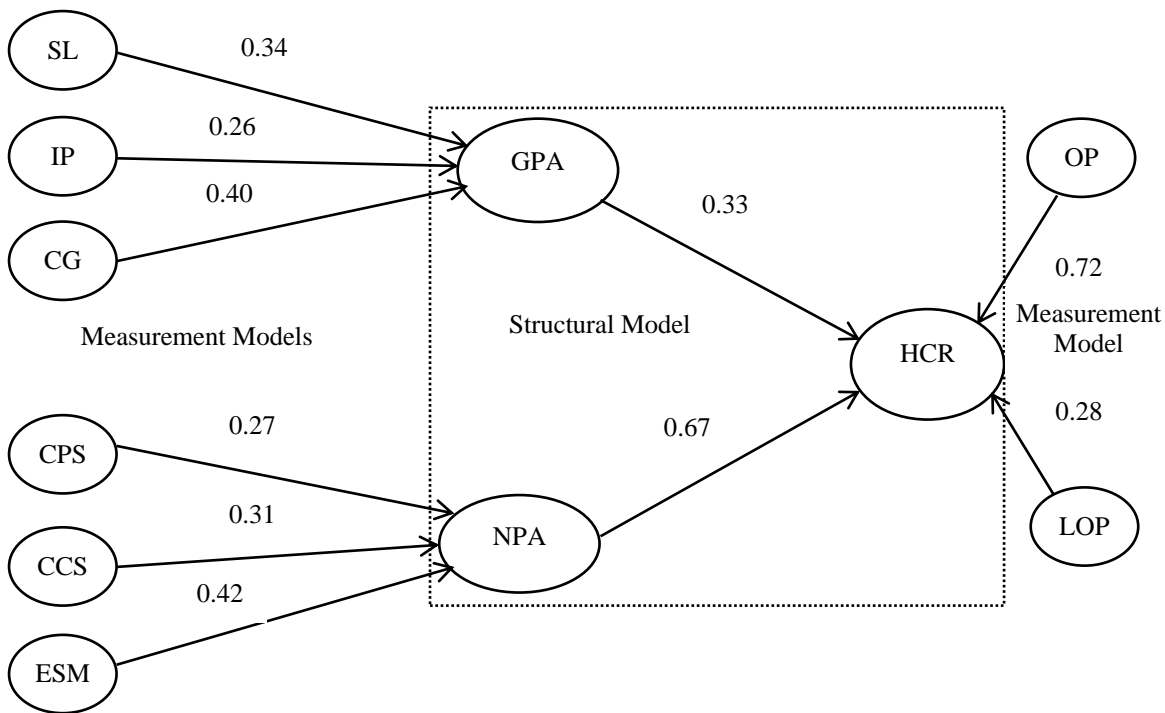


Figure 4.7: A developed conceptual model for policy perspectives

Legends:

GPA: governmental policy attributes; NPA: non-governmental policy attributes; SL: strengthening legislation; IP: increasing penalties; CG: Campaigns by government; CPS: Campaigns by private sectors; CCS: Campaigns by civil societies; ESM: Efforts by social media; HCR: Hate content reduction; OP: Online precaution; LOP: Limited online presence.

The construct governmental policy attribute is measured by strengthening legislation, increasing penalties, and campaigns by the government. The contribution from strengthening legislation, increasing penalties, and campaigns by the government is 34%, 26%, and 40% respectively. The results indicate that strengthening legislation, increasing penalties, and campaigns by the government all play an important role in forming governmental policies to reduce hate content.

The indicators campaigns by private sectors, campaigns by civil societies, and efforts by social media are the contributing factors to non-governmental policy attributes with

individual contributions of 27%, 31%, and 42% respectively. This indicates that all the indicators play an important role in constructing non-governmental policies. The contribution of online precaution to hate content reduction is 72%, and the limited online presence of an individual is 28%. The contribution of online precaution reveals that by being careful while accessing online media the hate content and its impact can be reduced largely.

Structural model assessment

The structural model is shown in Figure 4.7 within a dotted rectangle. Coefficients of determination/ R^2 , the fitness of good, and regression analysis are used to evaluate structural models. The exhibited R^2 value of the developed model is 1.0 and is shown in Table 4.16. Referring to the R^2 value, the model is good as it is above 0.60 (Sanchez, 2013) and is substantial/high as it exceeds 0.26 (Cohen, 2013). The R^2 value 1 also indicates that all necessary covariates are considered for model development (Miaou et al., 1996; Shtatland et al., 2002).

Table 4.16. Coefficient of determination for policy perspectives

Block	Type	R^2
Governmental policy attribute	Exogenous	0.00
Non-governmental policy attribute	Exogenous	0.00
Hate content reduction	Endogenous	1.0

The goodness of fit (GoF) is another criterion to measure the overall performance of the model. The GoF value of the developed model is 0.99 which is much more than the cut-off 0.36 specified for substantial/large models (Wetzels et al., 2009). The path analysis results of the structural model are shown in Table 4.17.

Table 4.17: Path analysis results for policy perspectives

Hypothesis	Path	Path coefficient	t-value	Result
H5	Governmental policy attribute → Hate content reduction (GPA→HCR)	33.5***	7.56	Supported
H6	Non-governmental policy attribute → Hate content reduction (NPA→HCR)	66.5***	15.0	Supported

*** Significance at the $P < 0.001$ level

A path GPA→HCR with coefficient 33.5 at t-value=7.56 and p-value<0.001 represents that the predictor GPA is significant and hence by rejecting the null hypothesis the alternate hypothesis H5 is accepted. Hypothesis H6 was tested by a path NPA→HCR. The path values reveal that it is significant with coefficient 66.5 at t-value=15.0 and p-value<0.001. Therefore, the null hypothesis is rejected and the alternate hypothesis H6 is accepted.

4.4 Concluding Remarks

From the discussions on results, we know the predictors of online hate content along with their contributing variables. Online hate content will be generated more by the persons who experienced hatred, having an attitude, lack of religious knowledge, the stance of superiority, ethnic unfamiliarity, the fright of health, ignorance, extremist behavior, communal nature, watching stereotypes, differentiating skin color and observing terrorist incidents. The hatred against different protected characteristics such as gender, religion, race and, disability leads to online hate content. Therefore, the attitude towards gender, religion, race and, disability are the determinants of online hate content.

To mitigate online hate content, both governmental and non-governmental policy attributes play an important role. The variables strengthening legislation, increasing penalties, and campaigns by the government are the influencers to governmental policy attributes. Similarly, the non-governmental policy attribute is influenced by the variables campaigns by private sectors, campaigns by civil societies, and efforts by social media. Apart from these governmental and non-governmental policy attributes, two standalone variables- online precaution and the limited online presence of an individual also influence hate content reduction. Non-governmental policy attributes such as campaigns by private sectors, civil societies, and efforts by social media are vital to reducing hate content than the governmental policy attributes. As the policies are suggested based on the respondents' views, implementing them may reduce online hate content.

CHAPTER 5

CASE STUDIES

5.1 Introduction

The previous chapter analyzes the responses of respondents for understanding and identifying the determinants and policy requirements of online hate content. To supplement and strengthen the results of quantitative analysis, different case studies are conducted and outlined in this chapter. The rest of the chapter is organized as follows. Section 5.2 presents the case of gender online hate content. The case of religious hate content is presented in section 5.3. In section 5.4, the case of racist hate content is presented. The case of physical violence concerning online hate content is discussed in section 5.5. Section 5.6 concludes the chapter.

5.2 Case of Gender-based Online Hate Content

The purpose of this case study is to identify the influence of technology-assisted social reform initiatives on gender-based hate content generation. In India, most of the social reforms took place between the late 19th and early 20th centuries. The majority of reforms were formed against the devaluing of Indian societal culture. The social reforms exist to maintain liberty, freedom, and equality of human beings by overcoming the differences between gender, religion, race, or caste. The modernity of social reforms can be achieved with a cost of degraded tradition. Social reforms prefixes with increased grievances and ideology development (McCarthy and Zald, 1977). In India, the social reform movements created a positive image, and the reformers are accepted as great contributors to the betterment of women (Srivastava, 2000).

In 1987, after the immolation of a woman, several reform movements occurred in India, which got support as well as opposition from the people (Kumar, 1997). Based on the initiators, social reforms can be classified as citizen-initiated and government-initiated reforms. Both the reforms take place for the betterment of society, to protect individual's rights, and to maintain societal equality among the citizen.

Social movements make concerned authorities work for the welfare of disadvantaged people (d'Anjou and Van Male, 1998). Neo social movements focus on life-related acts rather than class-related. Neo social movements struggle for human rights, such as women's rights and unlike traditional movements, neo social movements are less involved in politics (Carroll and Ratner, 1995). Often, contemporary social movements

are initiated and controlled by social media. The current study identifies and analyses the movements for human rights through social media.

Online hate can be reduced by considering the approaches which belong to "technological, legal and Internet literacy" (Blaya, 2019). Upon receiving the complaint from the users on hate content, the intermediaries review the content and try to block it if identified as hate content. This violates the freedom of expression of perpetrators. As an ethical approach to control hate content, the quarantining method can be used to temporarily block the content and alert the receiver like malicious software reporting (Ullmann and Tomalin, 2019).

Passing hate content law from the legislation to authorize intermediaries to moderate content by maintaining the freedom of expression is difficult. Instead, a regulation on consumer protection can be formed and passed by the legislations (MacCarthy, 2019). Moreover, policing for hate content by a single social media worsens the situation (Johnson et al., 2019). When policing at one platform takes place, online hate groups may move to other platforms. Therefore, simultaneous global policing is essential by all social media platforms. The hate content can be controlled by identifying and blocking the hate groups, creating anti-hate groups, and allowing counter-speech to neutralize it (Johnson et al., 2019).

With this background, the purpose of the study is set to identify the impact of social-reform initiatives on gender-based hate content generation. Social media content for government and citizen-initiated movements is extracted through Twitter API. The collected Twitter contents are analyzed for hatred nature in terms of emotions using software in the R programming language.

5.2.1 Case Context

With the ability to post, share, re-post, and like the messages, social media platforms are popularizing communication among citizens. Social media platforms play a major role in political reforms by uniting against the government (Bhuiyan, 2011). In 2009, social media helped in joining against the accused of murdering a person (Harlow, 2012).

The MeToo movement is a social mass movement and an important phenomenon to end “sexual harassment and assault” (Wolfe, 2018, p. 2) as a platform to share stories by the victims. Even though the movement is initiated in 2006 on MySpace by a woman, it went viral and popularized in October 2017 when a Hollywood actress in the USA wrote a tweet on sexual harassment expecting the response as MeToo (Wolfe, 2018). MeToo, as online activism or reform, is more effective and largely accepted by the victims for fighting against injustice. The MeToo movement raises a voice against sexual harassment irrespective of gender, race, religion, and disability, but the responses to it are biased based on these identities (Onwuachi-Willig, 2018).

In India, during 1920 and 1930, the temple entry movements succeeded in providing entry to lower caste people into the temples (Manmathan, 2013). The citizen-initiated movement, such as the anti-corruption (Lokpal) bill, is initiated in anticipation of overall social good (Alathur et al., 2016). Recently, the MeToo movement has been launched by victims in Bollywood and supported by millions of change seekers in society. The coverage of the MeToo movement is not limited to the film industry; it has been covered by media, government, and workplaces also (The Hindu, 2018). According to search giant Google, India is shining to the brightest for the trending of #MeToo searches (Radhakrishnan, 2018).

Social reforms play a significant role in framing hate crime laws (Platts et al., 2019). The justice system is improved to control hatred against immigrants through reforms (Johnson and Ingram, 2012). Reforms can be formed by governmental or non-governmental organizations. Many times, non-governmental organizations supported the framing of the legal reforms through communities to address hate incidents. The existing research indicates that there exist online gender hate content and social reforms assisting in bringing changes in the society by favoring marginalized people. The citizen support towards social reform movements can be analyzed through social media.

5.2.2 Methodology

As the study is intended to analyze social reform movements and gender hate content, the data is collected from Twitter social media on different social reform movements favoring equality among the genders. A methodology used for the analysis of social reform movements is shown in Figure 5.1.

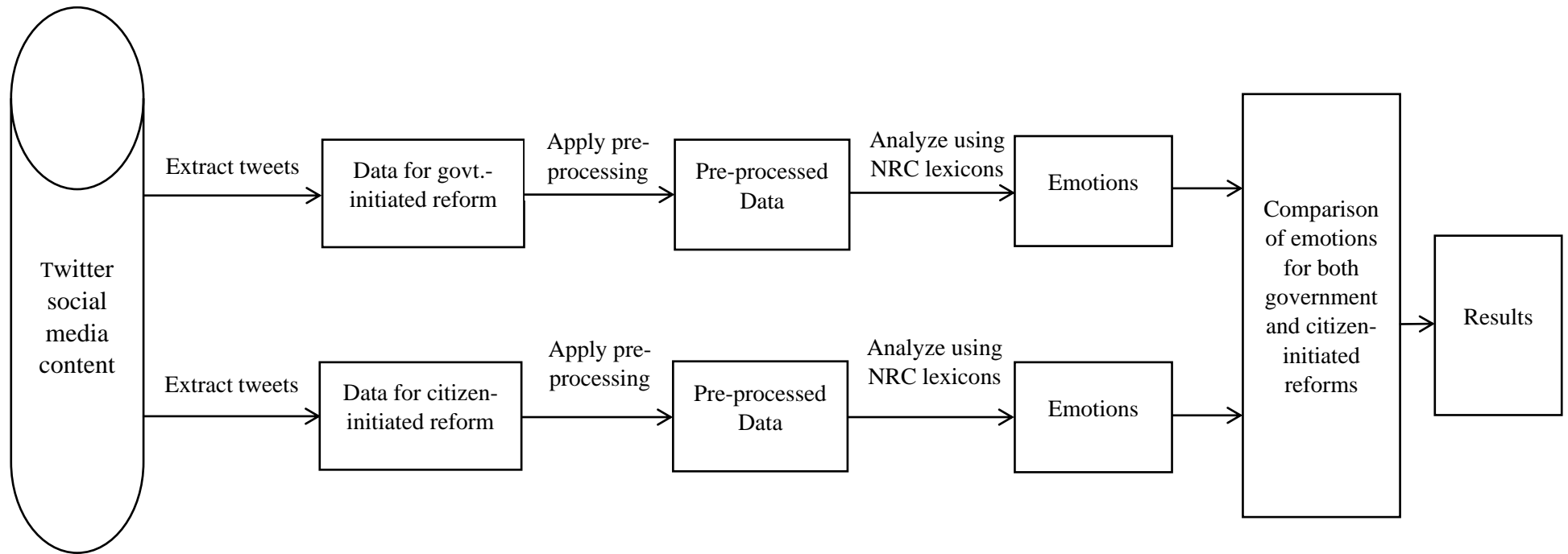


Figure 5.1: A methodology to analyze social reform movements

The government and citizen-initiated reforms are considered for the study. As a case for the government-initiated reform movement, the emphasis is made on a pilgrimage site (the divine place) in the Southern part of India. The Sabarimala is a divine tourist place that attracts global tourists. Though the Sabarimala issue is specific to South India, the views (tweets) collected for analysis covered the entire India. The analysis is based on the general respondents. On the other hand, the MeToo movement in India is considered a citizen-initiated reform movement.

To query Twitter social media, the trending words were identified through the search engines when the reforming movements were initiated. The keywords/hashtags that are pertinent to women and equality in temple case context the “Lord” is used to extract tweets on government-initiated reforming movement. Similarly, the “#Metoo” hashtag is used to collect data for the citizen-initiated reforming movement. The extraction of tweets on Metoo movement is limited to Indian geographical area by providing latitude and longitude global positioning system (GPS) coordinate values (latitude: 21.0000 longitude: 78.0000) with the search query. The data collection is initiated immediately after the birth of reforming movements and continued for 45 days spanning over October and November months in 2018. The Twitter data collected for government and citizen-initiated reform movements are 112577 and 58370 tweets, respectively.

The extracted tweets are analyzed using the software. Each tweet contains values for 16 different attributes like text, created, id, etc. As the opinions are conveyed through the textual part of the tweet, the study considered only the values of text attributes for analysis. The collected tweets are pre-processed to eliminate non-contributing information such as punctuations, and digits. The national research council (NRC) emotion lexicon ([Khoo and Johnkhan, 2018](#); [Mohammad and Turney, 2013](#)) is used to obtain the sentiment and emotions of the tweets.

5.2.3 Results and Discussion

The results of both the reforms are discussed separately, and in the end, a comparison between the reforms is made.

Government Initiated Social Reform

In government-initiated social reform, an attempt will be made to provide justice to the disadvantaged section of the society by incorporating appropriate changes in the existing system. The verdict on the divine place in Southern India is considered as a case for this reform movement. The NRC lexicons are used to analyze collected tweets and the emotional scores for respondents' opinions are shown in Figure 5.2.

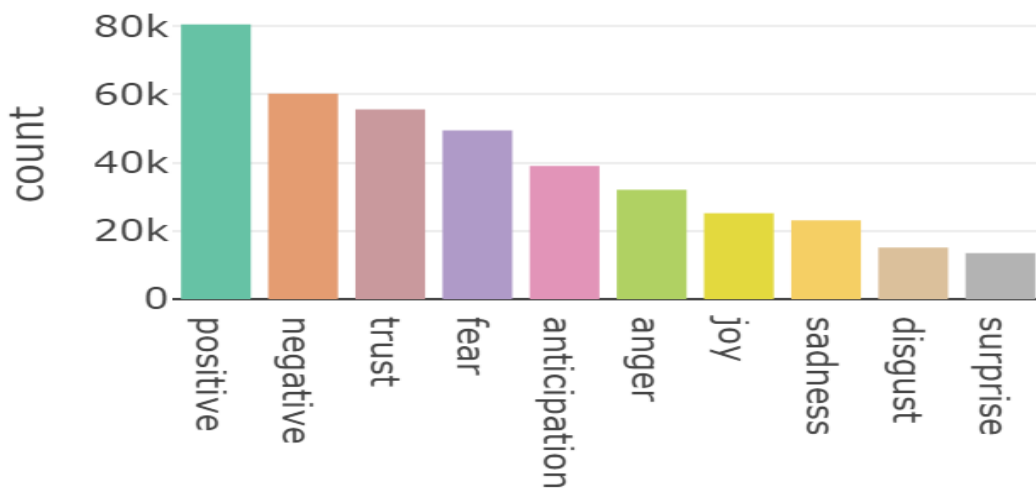


Figure 5.2: Emotions on a government-initiated social reform movement

The tweets shared within the social media on the shrine verdict, contain more positive sentiments. This shows that the majority of citizens are accepted the judgment on the shrine entry case. Among the other eight emotions, trust is exhibited with a greater number of tweets and indicates that the people believe or trust in this movement towards gender equality.

Fear results when some danger occurs with the living bodies and may lead to behavioral changes. The next highest impacting emotion is fear and depicts that the people are fearful of the implementation of a system for change. The presence of anticipation indicates that even with a lot of trust and fear still, people have some expectations from the movement. Anger is associated with a negative feeling and less exhibited compared to trust, fear, and anticipation.

Joy and sadness emotions are opposite as the former belongs to a positive feeling and later belongs to a negative feeling. The scores for joy and sadness from the analysis are almost the same. This reveals that one portion of the citizen is happier, and another portion is not happy with the movement. Overall, the movement results in neutrality regarding joy and sadness. The disgust emotion's score value shows that part of the citizens is unpleasant with the movement and trying to reject it. The score of surprise indicates that few citizens considered this movement as unexpected.

Every emotion is expressed using words. The wordcloud for different emotions is shown in Figure 5.3. As anticipation is concerned with the future events, expressed with the words like time, temple, wait, long, etc. The anger emotion being aggressive, expressed with the words like a court, supreme, fight, attack, etc.



Figure 5.3: Wordcloud of emotions for a government-initiated social reform

Trust is an emotion of positive feeling and expressed using words like the leader, support, sir, religion, etc. The words good, hope, pray, judgment, etc. can be used to represent surprise in case of unexpected incidents. The sadness emotion dealing with negative feelings can be expressed using words like a case, violence, problem, hurt, etc.

Joy being opposite to the sadness emotion can be exhibited through the words save, faith, god, respect, etc. The frightened response is expressed using the words police, verdict, hear, petition, etc. The disgust emotion being unpleasant is expressed through the set of words such as shame, discrimination, hell, insult, etc.

Citizen Initiated Social Reform

Citizen-initiated social reform will attempt to raise the voice against the injustice that occurred to the marginalized section of the society. The initiators of this movement expect that justice will be provided to the weaker section from governing authorities by introducing appropriate laws and punishing the perpetrators. The MeToo movement, which is initiated in the United States of America in 2017 and currently in India, is considered as a case for this type of movement. The tweets are analyzed like government-initiated movement and the scores for different emotions are shown in Figure 5.4.

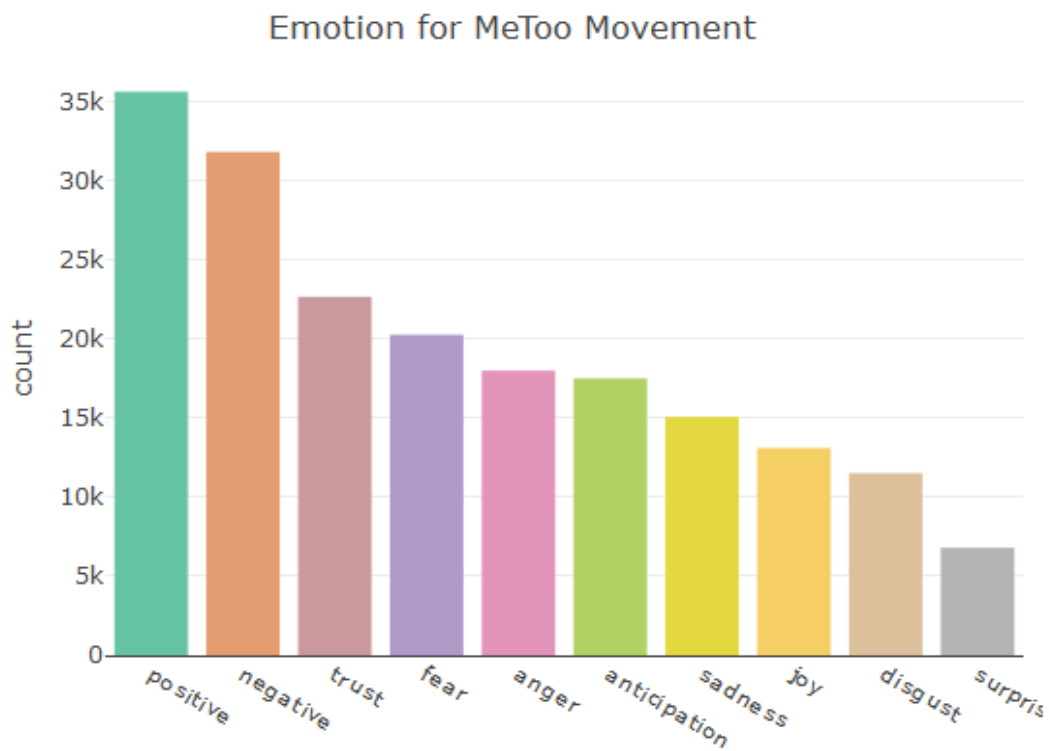


Figure 5.4: Emotions on a citizen-initiated social reform movement

The tweets collected through social media on the MeToo movement in India contain more positive sentiments than negative. This indicates that most of the citizens are accepted the MeToo movement initiation. The higher score value of the trust emotion reveals that the citizen believes or trusts in the movement towards gender equality. After the trust, fear emotion is having more count value than the other emotions and indicates that some people are afraid of the initiation of movement for changes in society. The next highest scored emotion anger is associated with a negative feeling and exhibited as a response to their perceived threat or hurt or harassment. The score value of anticipation indicates that people have some expectations from the movement. Sadness emotion score value indicates that some people are not happier with the incidents they faced and expressing the worst time they had in the past. As the movement is intended to bring some changes in society, some people feel joy about it. The score of disgust emotion shows that some citizens are unpleasant with the incidents they faced and trying to reject them. The surprise emotion's score shows that a part of the citizens is considered the incident/movement as unexpected.

The wordcloud for different emotions is shown in Figure 5.5. The anticipation emotion is expressed through the set of words like time, come, start, thought, etc. The words such as accuse, assault, fight, harass, and attack are used to represent anger emotion by the citizen in the movement. The trust emotion is expressed with the words like support, truth, show, real, law, and justice. The words good, hope, inspire, young, surprise, resignation, and India are used to representing surprise emotion for unexpected incidents.

The sadness represents the negative feeling and is expressed with words like the case, violence, problem, the victim, prison, black, and lost. Joy is exhibited through the words sex, love, white, share, respect, and music. The words used for representing the fear emotion are sexual, change government and fire. The disgust emotion is expressed through the set of words such as rape, abuse, misconduct, power, shame, and disgust.

One of the tweets involving the MeToo hashtag is "In the wake of the #MeToo movement, more sexual harassment scandals on college campuses have come to light". The nature of the tweet reveals a positive opinion about the MeToo movement. The victim of the incident may feel happy as it could be possible to expose the incident in

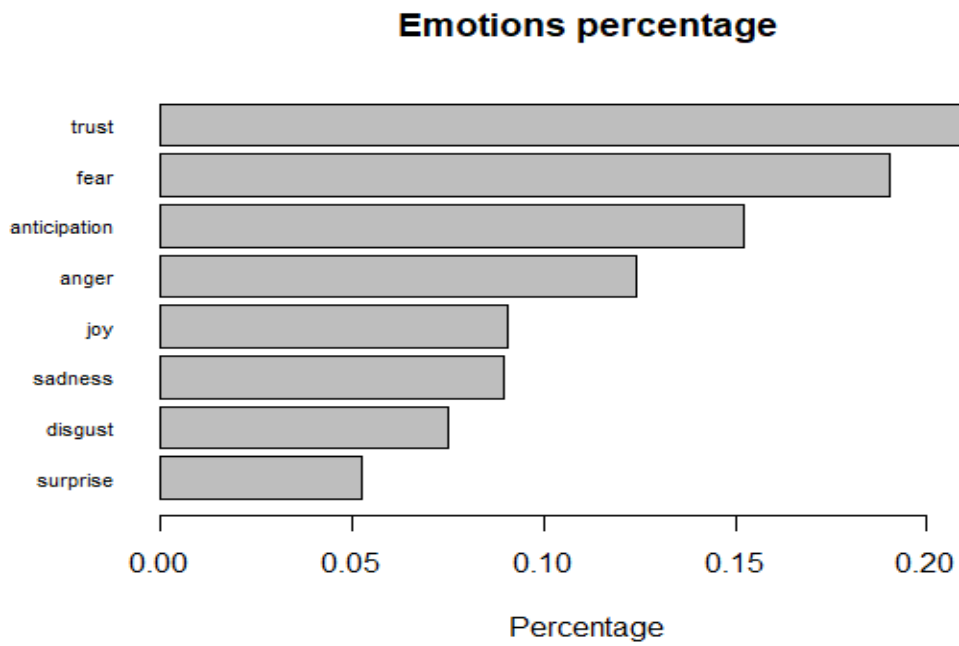


Figure 5.6: Percentage of emotions for a government-initiated social reform movement

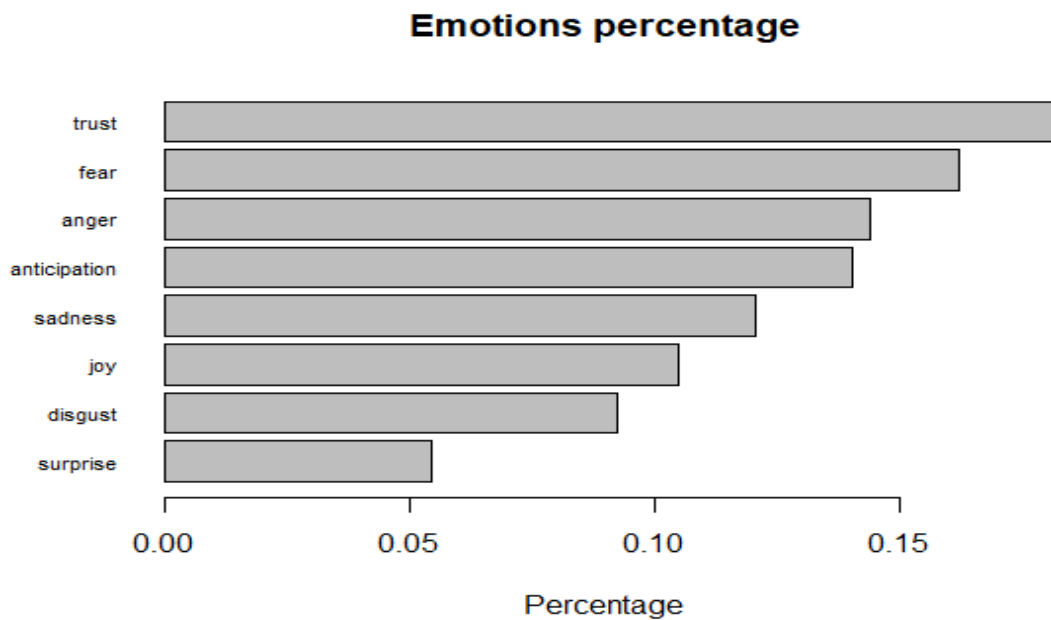


Figure 5.7: Percentage of emotions for a citizen-initiated social reform movement

Table 5.1 reveals that there are many words for these emotions in both movements. In general, these recorded emotion values indicate that there exists some hatred in the content exchanged over both movements. The exchange of hatred emotions is more in the citizen-initiated social reform movement.

Table 5.1: Government-initiated movement vs. citizen-initiated movement

Feature/Emotion	Govt.-initiated movement (Shrine verdict)	Citizen-initiated movement (MeToo)
Initiators	Initiators are lawmakers	Initiators are victims
Impact	The positive impact is more than the negative	The positive impact is more than the negative
Trust	As the movement is initiated by the government and lawmakers, trust is more (25 percentage)	Trust is less (18 percentage) as the citizen think more about trusting each other
Fear	As the movement is supported legally, fear is more (19 percentage)	Fear is less (17 percentage) to tweet as the movement is not yet covered under the legal framework
Anticipation	As the people expect more from the government and lawmakers, the anticipation is more (15 percentage)	Comparatively, the anticipation is less (14 percentage) as the lawmakers, and the government is not involved in the movement
Anger	Anger is less (12 percentage) as no particular citizen is victimized	As the people (victims) are harassed, the anger is dominating (15 percentage)
Joy/Sadness	As the movement is initiated by the government, without the effort of citizens and making changes in the society, the citizens have more joy than the sadness	Being the victims of harassment or well-wishers of the victims, the people expressed more sadness than the joy
Disgust	Disgust is less (7 percentage) as no hostile or unpleasant behavior towards a particular person	Act as a platform to share their worst experiences in life, hence, content possess more (12 percentage) disgust than in government-initiated movement
Surprise	As the movement is planned and implemented, there is less surprise	As the movement is not planned and suddenly occurred, the surprise emotion is more

The cohesiveness of Hate and Harassment

The relationship between hate and harassment is identified by examining the citizen-initiated movement MeToo. The causal loop diagram for the results of the MeToo

movement is shown in Figure 5.8. The emotions which contribute more to the hate content such as anger, fear, disgust, and sadness (Mohammad, 2012) are considered to examine this relation. Representing the magnitude of harassment and urging for the action against the harasser is the central theme of the MeToo movement. Harassment is originated from different dyads and the dyads from different emotions. The theme of the Metoo movement “harassment” is identified by observing the text part of the tweets. The existing studies revealed that there is a relationship between the different emotions and dyads such as contempt, remorse, despair, and shame (Hutcherson and Gross, 2011; Sommer et al., 2019; Vartanian et al., 2013).

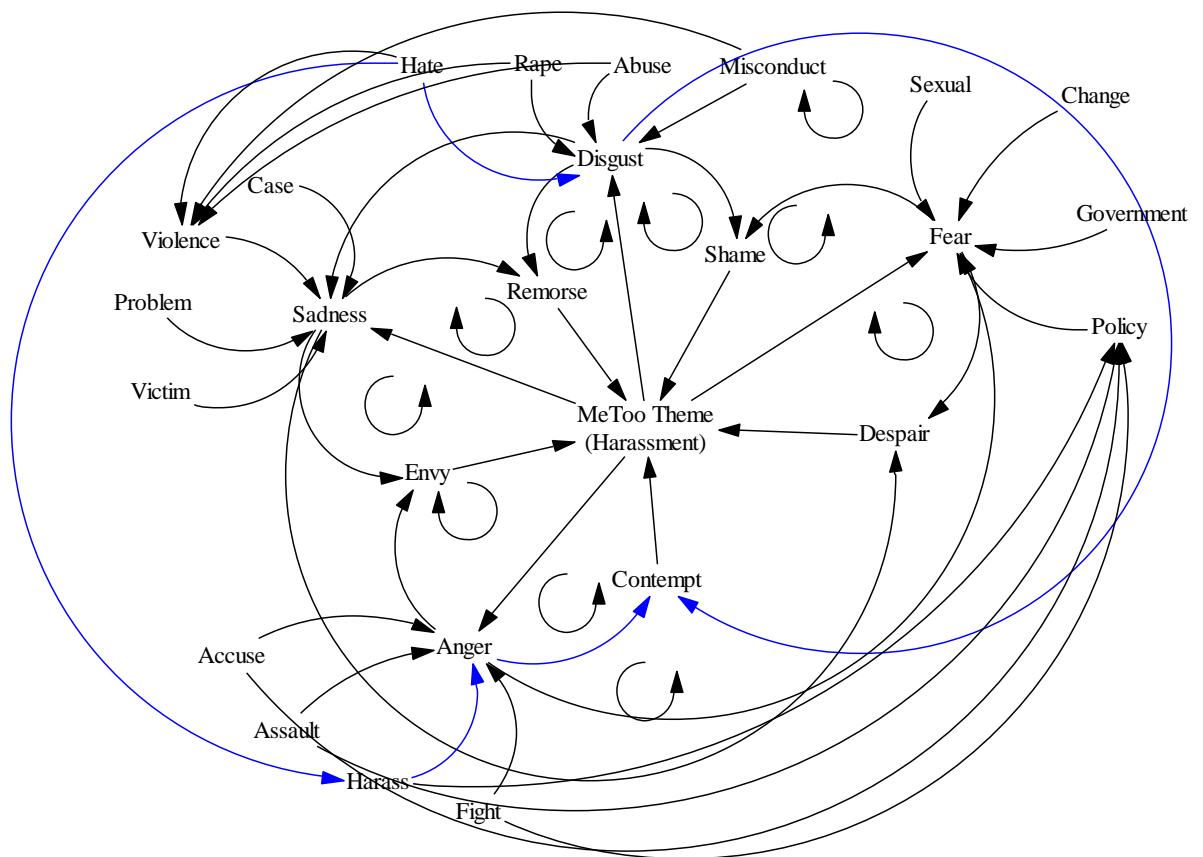


Figure 5.8: Causal loop diagram of MeToo movement

Based on the frequency of occurrence, the most occurring words related to different emotions such as anger, disgust, sadness, and fear are identified from the wordcloud of

citizen-initiated reform. Frequently occurring words are connected to their respective emotions, emotions are connected to different dyads, and dyads are connected to the theme of MeToo reform. The connections between the related words of different emotions and from the MeToo theme to different dyads are drawn to a complete causal loop diagram using the Vensim software tool.

The contempt, a strong dislike feeling, is a dyad and originated from the combination of anger and disgust emotions (Athar et al., 2011; Heerdink et al., 2019). The anger emotion is contributed by words such as accuse, assault, fight, and harassing. Similarly, words such as abuse, hate, misconduct, and rape are contributors to disgust emotion.

As the citizen-initiated movement MeToo is initiated in extreme situations (after being the victims of severe hatred), the movement carries more hatred expressions against their harassment in the fast. This reform movement acts as the platform to express hatred against the haters. The government-initiated movement, the divine place verdict, is initiated to maintain equality in society (to avoid severe hatred). Therefore, it carries fewer hatred expressions. Hence, the research objective of understanding the influence of digital hate content in online safety through social reform movements is addressed.

5.3 Case of Religious Online Hate Content

Traditionally words are considered more powerful than weapons. Hate content is a form of weapon for violence and in the digital world as violence itself. The violence can be defined as "a communication of hate with a motive of retaining honor of the family/communities and for exhibiting hatred behavior attitudes of propagator/s, and may result in meaningful content with expressions of emotions such as anger, fear, disgust, and sadness". Thus Hate content can be considered as extreme aggression and is a global threat that can occur anywhere anytime. Hate violence is categorized as self-directed, interpersonal, and collective hate content (Krug et al., 2002). Self-directed hate content results and is exhibited on oneself through suicide and self-abuse (Boyd et al., 2011). Interpersonal hate content exists between different persons, families, or communities (Ichheiser, 1943). Collective hate content is expressed in larger groups with multiple motives.

Multiple sources like self-report, convictions, and information from a relative may be used to achieve the comprehensive history of hate violence (Harris et al., 2013).

Sometimes in the literature, the term aggression is used instead of hate-related violence (Serper, 2011). Aggression is “any behavior directed toward another individual to cause harm” (Anderson & Bushman, 2002, p. 298). This behavior can be treated as aggressive even in the absence of behavioral expression or any physical harm (Serper, 2011). Accordingly, hate-related violence is generally considered an extreme form of aggression (Harris et al., 2013). The heterogeneous nature of violent behavior leads to a major challenge for the measurement of hate content in aggression research (Khan and Cooke, 2013).

Inter religion and inter-caste marriages are one of the causes of honor-based violence in India (Deol, 2014; Goli et al., 2013). The honor-based violence after some time persists as hatred among the communities involved. Therefore, the honor-based violence case is considered as an example of religious hate content.

5.3.1 Case Context

Honour is the respect involving norms and traditions associated with an individual, family, or community. The assumption is that honor can be achieved by following societal expectations. Otherwise, it leads to the shame of the concerned (Gill et al., 2012). Shame is a “set of emotions like embarrassment, guilt, and shyness and is a threat to the social bond” (Scheff, 2003, p. 255). The definition relates individuals with emotional reactions to society. Honor-based violence (HBV); the form of violence against women for the social construction of honor and is associated with a male-dominant family/community (Gill et al., 2012). Social self-feeling is the root of hostility and makes us live in others' minds (Retzinger, 1991). HBV may be exhibited against women to maintain the honor of the family.

HBV is a collection of activities, which are performed to control behavior and exhibit power in the family to safeguard assumed cultural and religious concepts of honor. This kind of violence may occur when perpetrators assume that the behavior of a member shamed the entire family by breaking their code of conduct. HBV is an abuse of human rights. There is no honor with the practices of violence in the name of honor (Asquith, 2015). HBV is a cultural phenomenon and impacts communities. Honor crimes are considered as part of domestic violence, intimate partner violence, and family violence (Hayes et al., 2016).

Honor-based violence is gendered violence with cultural perspectives. Youths are well educated and wish to live a stylish life with freedom. The traditional and cultural views of the family or community act as barriers to this way of living. With the changed lifestyle, the youths may neglect the cultural framework of family or community. In India, the ban on intimate partner harming is confusing because it is backed by a legal mechanism but with limited social support (Tichy et al., 2009). To control HBV, it is essential to understand the supporting factors to HBV and its persistence.

5.3.2 Methodology

A review of literature, news articles on honor killing, and the analysis of Twitter content are made to understand the religious hatred online. The honor-based violence incidents which are reported through different sources from 2010 to 2018 in India are studied and analyzed for their nature. The incidents which are analyzed cover different parts of India.

The honor of a family and violence exhibited against the victim are associated (Brown et al., 2018; Thrasher and Handfield, 2018). Hatred behavior against a person based on protected characteristics and violence exhibited are interrelated (Morrow, 2001). Similarly, honor, hate and, violence are interrelated (Hayes et al., 2018). Therefore, Twitter content is extracted using a combination of keywords for honor, hate, and violence. The data collection is carried out for two weeks to extract 21000 tweets through the Twitter API. A pre-processing mechanism is applied on Twitter content to remove punctuations, digits, and to stem the words. For each tweet, the sentiment and emotions are obtained using the national research council (NRC) lexicon. The sentiment and emotions for the entire content are based on the sentiments of these individual tweets.

5.3.3 Results and Discussion

The occurrence of HBV is prevalent in different cultures and communities (Faqir, 2001). It has a long history of persistence. Honor killing, the most extreme form of honor-based violence is more attracted by Media. Apart from killing, other forms of violence also persist. “Forced marriage, selling of females into slavery, mutilation, and

denying rights such as freedom, education, or friendship are the other forms of honor-based violence” (Gill, 2006).

HBV is assumed as a tradition to maintain the honor of the family or community (Almosaed, 2004). Often the victims of this violence are female, and the offenders are male and his relatives (Wikan, 1984). The offenders justify their crimes by highlighting that the victim has dishonored the entire family in society (Kirti et al., 2011). It is assumed that violation of honor codes by individuals dishonors the entire family and continues as a threat to a family’s survival in the future (Gill and Brah, 2014). The major consequence of this violation is the exclusion of the family from the social group (Kirti et al., 2011). HBV is an artifact of an uncivilized code of morality (Asquith, 2015). More broadly, HBV covers every form of violence against women in a view of the patriarchal family and social structures.

The stems of honor killings are initiated long back and too old. An honor killing may involve all members of a family (Meetoo and Mirza, 2007). A woman can defile the honor of her family on the grounds of choosing a partner to marry, seeking a divorce, having a boyfriend, refusing an arranged marriage, homosexuality, and trying to live independently (Meetoo and Mirza, 2007). The root cause of this violence is considered as immoral behavior such as independent partner selection by women, having premarital/extramarital affairs, being a rape victim, and talking to a non-relative man (Khan, 2007).

The different causes of honor-based violence such as marriage against family members, inter-caste relationships, love affairs against family members, and relationship with a person against family members are identified (Krishnan, 2018; DNA, 2017, TOI, 2018). According to the reported incidents, the different forms of honor-based violence are burning, stabbing, strangulation, hitting by axe, hitting by sickle, and mass attack by family members. The influencing factors on honor-based violence range from love affairs to marriage against family or community members.

Honor-based violence is often supported by the khap or caste panchayats which are prevalent in northern India. Khap panchayats are formed with senior members of the community; mostly the members are from the upper caste (Vishwanath and Palakonda, 2011; Yadav, 2009). Inter-caste marriages are hated because the people belong to

different caste often think that these marriages will raise the status of other caste people socially as well as economically (Kumar, 2012). Therefore, if a relationship establishes between the male and female of different caste or religions, the dominating community may act violently with strong dislike against the other community. With the introduction of the Hindu marriage act, the status of khap panchayats is diminishing. “As the khap panchayats are managed by senior men, they use young and less-educated men for the purpose” (Kaur, 2010, p.16).

The increased honor killings with the support of khap panchayats, lead to a proposal for a separate law against honor killings by the national commission for women. As a response to this proposal, the Supreme Court of India announced that “there are no restrictions on adults to marry persons of their interest and no social group should hurt them” (Rajagopal, 2018). In this regard, finally, the Supreme Court in its judgment dated 27 March 2018 strictly instructed that “the khap panchayats are not permitted to take the law into their hands to punish the crime committer or persons who married against family members/community” (SCJ, 2018, p. 41). The Supreme Court also instructed that the concerned governments should take the necessary actions to prevent killings in the name of honor.

Figure 5.9 shows the different influencing factors for honor-based violence (Khan, 2007; Meetoo and Mirza, 2007), forms of honor-based violence (Gill, A., 2006), and forms of hate-based violence. The honor killing is committed as a final resort to the problem, or the shame or dis-honor occurred to the family. However, there exist other forms of violence such as forced marriage, forcing to slavery, mutilation, and removal of rights on freedom or education or friendship as a substitute for honor killing.

In the case of honor killings, the harm is intended because of gender identity only. Even though the honor killings have common roots with spousal abuse, the actual killing of girls and women originates from a very harsh form of male dominance that takes the shape of a hate crime. Murder in the name of honor and the dowry motivated crimes are still reported in some parts of the world including India. Domestic violence may start in the name of honor and continue as hate violence in the form of acid throwing (Eisler, 2015). Domestic violence leads to poor quality of health of the victim (Brewer

et al., 2010). Often, honor violence ends with honor killing; otherwise, it persists as hate violence in different forms.

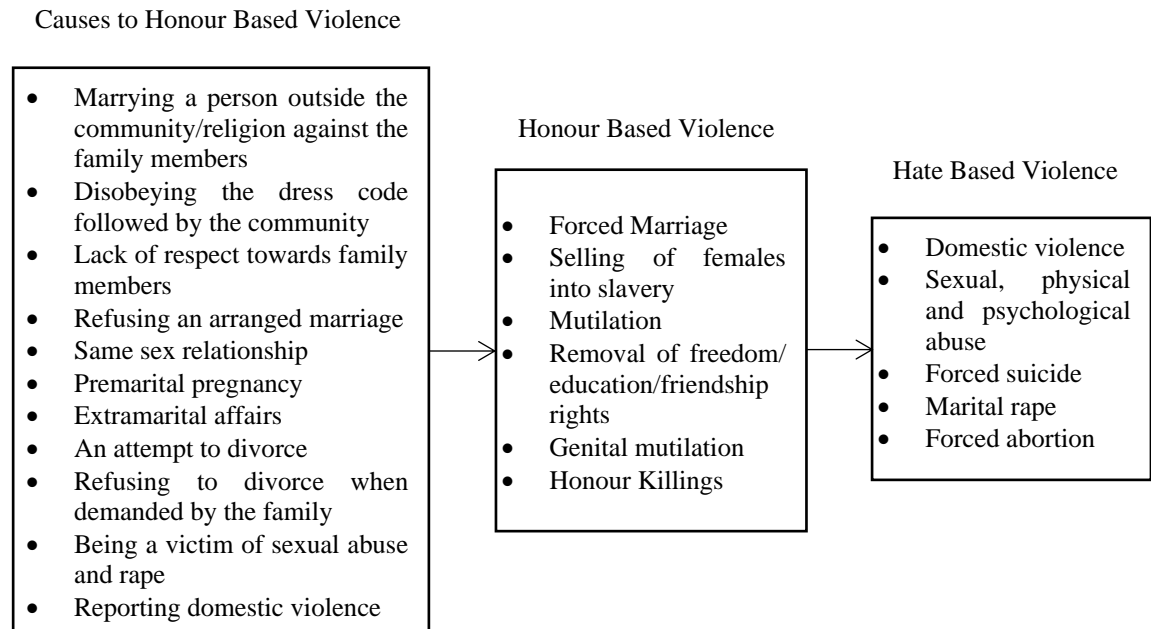


Figure 5.9: Honor based and hate-based violence with different violent acts

Social Media Content Analysis

Emotion is “both physical feeling and a conscious making sense of that feeling” (Askins, 2009, p. 9). Emotion supports religious life by inducing rituals (Corcoran, 2015). Positive societal interactions create group supportive emotions with sacrificial behavior (Lawler et al., 2000; Lawler, 2001). This group's supportive emotion may lead to hatred and violence against other groups. Therefore, the analysis of Twitter content concerning different emotions such as anger, disgust, sadness, trust, fear, joy, anticipation, and surprise is made. A graphical representation for different emotions which are obtained with the help of software developed for the purpose is shown in Figure 5.10. There are more negative words than positive words.

The conflicts between the different groups are the impact of a variety of emotions. Hate has multiple dimensions such as disgust, anger/fear, and devaluation/diminution and that may take different forms based on the situation (Sternberg, 2003). In Figure 5.10, the word count of anger, fear, sadness, and disgust of Twitter content is more than the other emotions. These values show that the collected social media data contain more

hate content than non-hate content. These emotions are in line with the result of Sternberg (2003), the dimensions such as disgust, anger/fear, and devaluation /diminution are associated with hate content. A wordcloud containing the frequent words in collected Twitter data is shown in Figure 5.11.

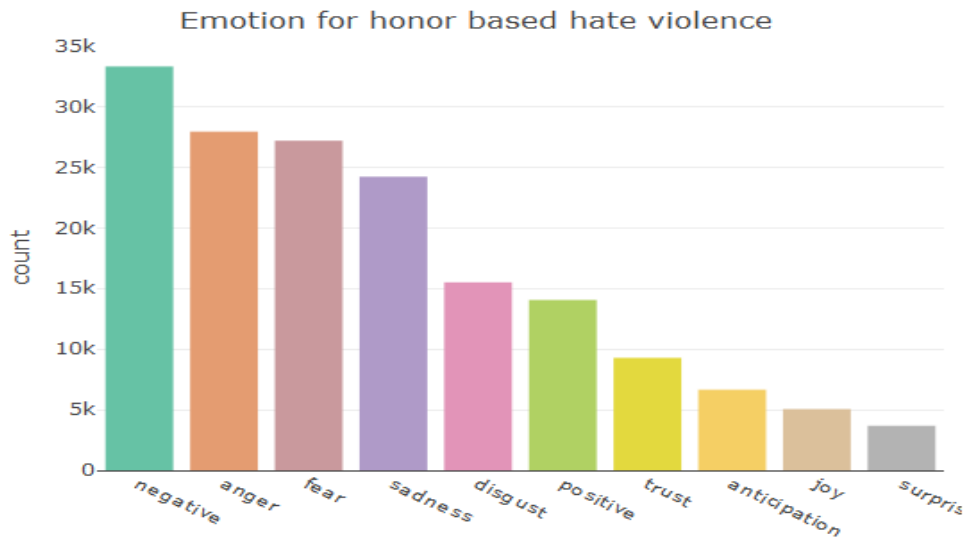


Figure 5.10: Emotions for honor-based hate violence

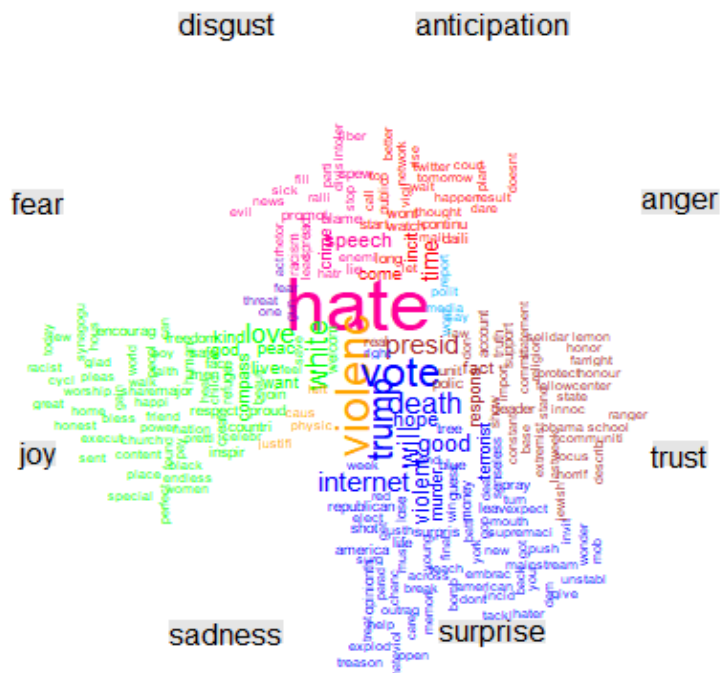


Figure 5.11: A wordcloud with emotions for honor-based hate violence

In wordcloud, the words honor, hate, and violence lie within the emotion dimensions such as trust, disgust, and sadness respectively. According to wordcloud the honor resulted in trust, hate resulted in disgust, and violence resulted in sadness. This indicates that the honor-based hate violence is initiated with trust/belief, continued/persist with disgust, and ended with sadness.

5.4 Case of Racist Online Hate Content

Racism is a kind of hatred behavior, exhibited in written, verbal or physical form against the ethnicity or physical appearance of a group or an individual. The case context, methodology, and results are discussed in the following sections.

5.4.1 Case Context

There is much more affinity between racism and ethnicity. Ethnicity is a part of racism (Hosseini & Zohdi, 2016). Ethnicity refers to cultural differences among people. Racism and its consequences originate early from school days (Tatum, 2001). Even with the efforts to nullify it, “racism exhibiting a sufficient capacity to mutate and evolve like bacteria’s response to antibiotics” (Acuna, 2005, p. 2). Racism is invisible to perpetrators but visible and harmful to the target community with physical, economic, social, and psychological losses. Racism divides people into different parts by creating misunderstandings among them. An ideology behind racism is the concept of white superiority and black inferiority. It reduces opportunities, peace of mind, and safety of an individual of black ethnicity (Miller and Garraan, 2017).

Different influencing factors on racism are shown in Figure 5.12 (Banton, 1992; Catholic Charities, 2008; Jones, 2013). With the *upbringing/inheritance* factor, “children start representing their parent’s behavior in society by inheriting features of the parents” (Dunn, 1975, p. 86). *Peer pressure* is a force from friends or another individual at the same level towards racism. *Personal experiences* of the racial victim make them behave as racist towards another ethnicity.

Stereotypes are created using verbal and visual information to represent a group. *Stereotypes* make superior groups feel proud of their ethnicity/race and exhibit racism against another ethnicity/race (Crawford, 1998). *Skin color* is associated with the physical appearance of a person and common cause of racism. *Poverty* is the state of a

person with financial deficiency and causes inferiority among poor people. If the children are not taught to respect others and their race, there is more possibility of those children becoming racists in the future (Derman-Sparks, 1980).

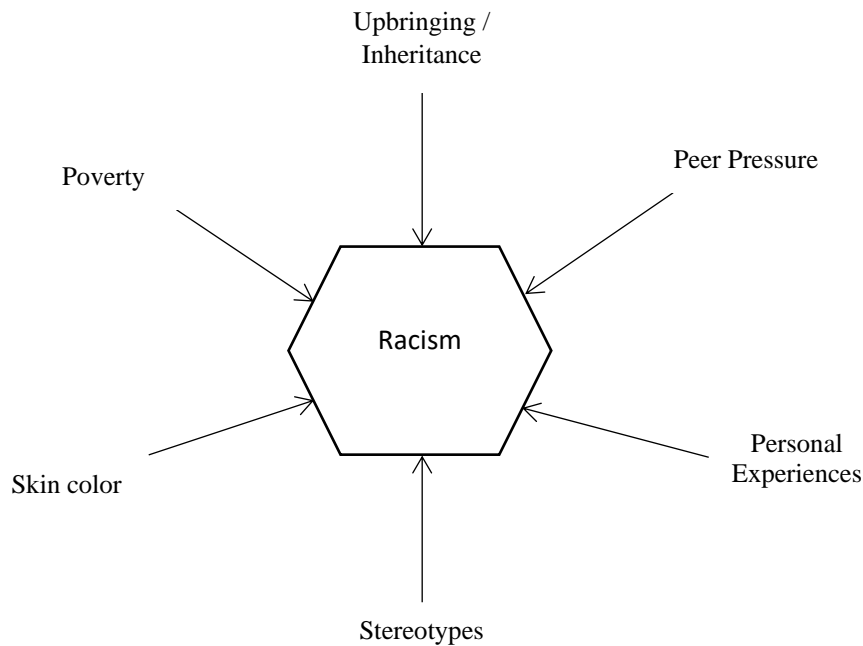


Figure 5.12: Racism influencing factors

Indian news media often had drawn articles on racism. A few articles point to the root cause of racism in India that may include the Bollywood industry (Jyoti, 2017). This made us consider Bollywood also as a keyword while extracting tweets from Twitter. Bollywood is an integrated name, derived from the words Bombay and Hollywood for the Hindi film industry. The history of Bollywood started in 1913 with a simple silent movie to sound and technically advanced movies of the present day. Bollywood movies are often “spicy” with a combination of different attributes such as music, dance, songs, and violence (Dwivedi, 2017; Srivastava, 2009).

Apart from entertainment, Bollywood affects the lifestyle, culture, and behavior of the people in India (Goswami & Kumari, 2013; Jain et al., 2015). As an impact on lifestyle, the patterns of dresses and ornaments worn by the star cast in the movie are appeared in the markets and are used by the common public. The different changes in the societal culture such as widow remarriage, pre and post-marriage programs, and the voice of youth against corrupt politics are also the impact of Bollywood.

Often, racism is expressed against people of colour. At the time of the case study on racism, the news media were reporting the incidents of racism concerning the Africans in India. Therefore, the study considers the use of keywords Africans during the search for tweets on Twitter social media.

5.4.2 Methodology

The dataset for the analysis is extracted from the Twitter social media using the *twitterSearch* function. As this work is based on different contexts, we extracted the tweets by following different ways to meet the requirements. A tree structure of the dataset used in this work is outlined in Figure 5.13.

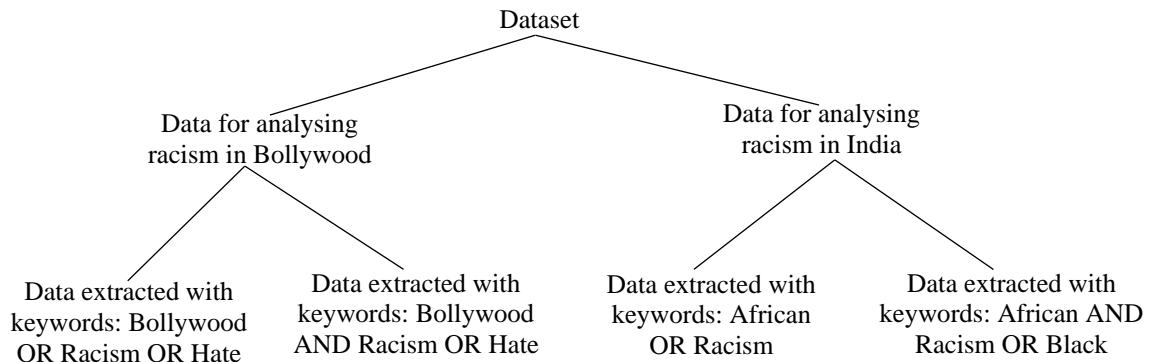


Figure 5.13: Datasets for racism analysis

While extracting the tweets longitude and latitude information is used to ensure that tweets are obtained from the geographic location of India. Each dataset contains 16 attributes. As the present work is to analyze racism with different contexts in India we considered only the text part of the dataset. An outcome of work is highly dependent on the approach used for the purpose. A list of activities followed for analysis of tweets on racism in India is shown in Figure 5.14. The activities of Figure 5.14 are implemented in the R programming language.

Tweets are obtained by querying Twitter (requesting for tweets) with a different combination of keywords using the *twitterSearch* function. Initially, retweets are also part of the tweets obtained by *twitterSearch* function. Retweets are removed from the original set of tweets by applying the filter function. After extracting only the text part

of the tweets, some preliminary operations are applied to remove punctuations, control characters, and digits from the text.

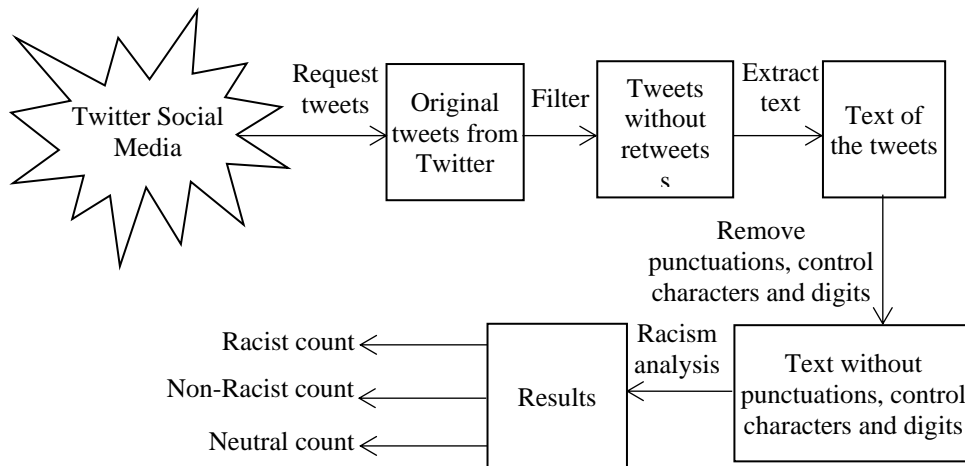


Figure 5.14 An approach for racism analysis

A simple racism detection mechanism is used to detect and analyze tweets as racist, non-racist, and neutral polarities. The algorithm classifies the tweets as racist and non-racist based on the negative and positive polarities of the tweets respectively.

5.4.3 Results and Discussion

In the following subsections, a discussion on racism with different contexts in India is made. For each context, the results are based on the searching style of tweets with different keyword combinations and the analysis.

Racism analysis with Keywords African OR Racism

The racism detection method is applied to the dataset, which is built by searching Twitter social media with the keywords *Africans or Racism*. This dataset consists of 1000 relevant tweets for the topic of study. Graphical representation of the predicted result as racist, non-racist, and neutral is shown in Figure 5.15. The columns of Figure 5.15 show the polarity of the tweets such as racist, non-racist, and neutral with the values 416, 219, and 365 respectively. These polarity values represent that out of 1000 Twitter users 416 users are tweeted with racist messages, 219 users tweeted with non-racist messages and 365 users are tweeted with neutral statements.

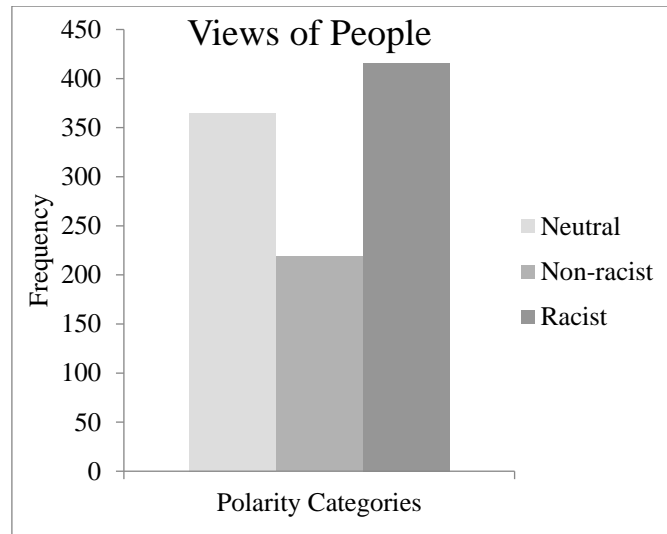


Figure 5.15: Graph of racism analysis (Keywords: African or racism)

Racism score of the tweets with the keywords: *African or racism* is shown in Figure 5.16. Racism is represented by the “-” symbol, which indicates the amount of racism, the “+” symbol indicates the amount of non-racism and *zero* indicates that a specified number of tweeters are neutral. The strength of racism increases from 4 to -4.

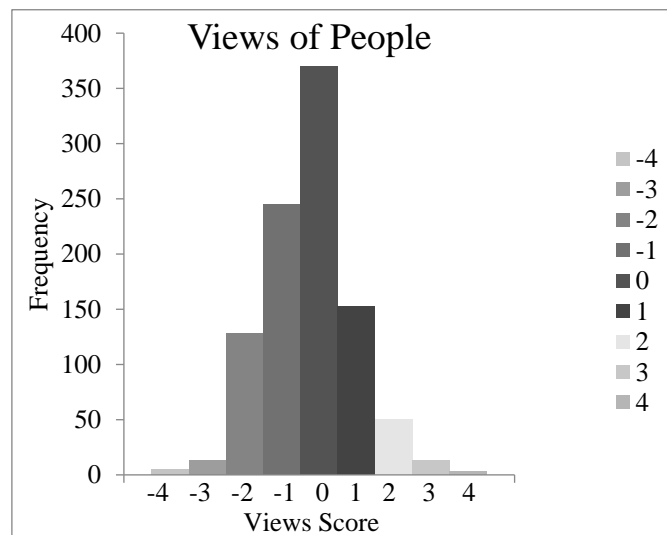


Figure 5.16: The score of racism (Keywords: African or racism)

Racism analysis with Keywords African AND Racism OR Black

For the prediction of racism in this category a detection method is applied to the tweets dataset collected with keywords *Africans and racism or black*. This dataset consists of

only 37 tweets that are more relevant to the topic of study and spanned over 10 days. Figure 5.17 shows a graphical representation of predicted results as racist, non-racist, and neutral.

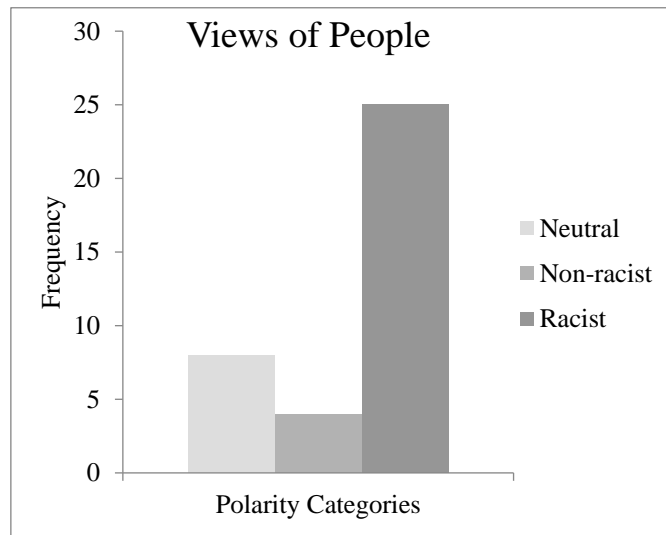


Figure 5.17: Graph of racism analysis (Keywords: African and Racism or black)

The columns of Figure 5.17 show the polarity of the tweets such as racist, non-racist, and neutral with the values 25, 4, and 8 respectively. These polarity values represent that out of 37 Twitter users 25 users are tweeted with racist messages, 4 users tweeted with non-racist messages and 8 users are tweeted with neutral statements. In Figure 5.18, a score of racism for the tweets with keywords: *African and racism or black* is shown as a bar graph.

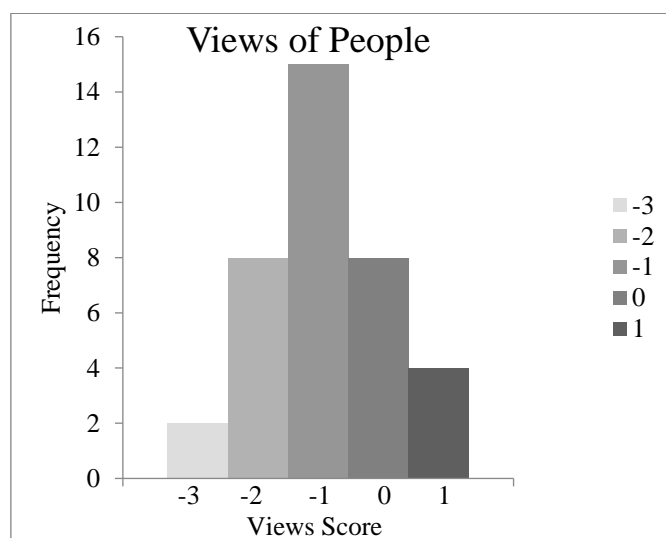


Figure 5.18: The score of racism (Keywords: African and racism or black)

Racism is represented by the “-“ symbol, which indicates the amount of racism, the “+” symbol indicates the amount of non-racism and *zero* indicates that a specified number of tweeters are neutral. The strength of racism increases from 1 to -3.

Racism analysis with Keywords Bollywood OR Racism OR Hate

After querying Twitter social media with the keywords *Bollywood or racism or hate*, the racism prediction method is applied to the collected tweets dataset. This dataset also consists of 1000 relevant tweets for the topic of study. Figure 5.19 shows a graphical representation of predicted results as racist, non-racist, and neutral. The polarity of the tweets is shown in Figure 5.19. The number of user tweets for racist polarity is 543, the non-racist polarity is 160, and the neutral is 297.

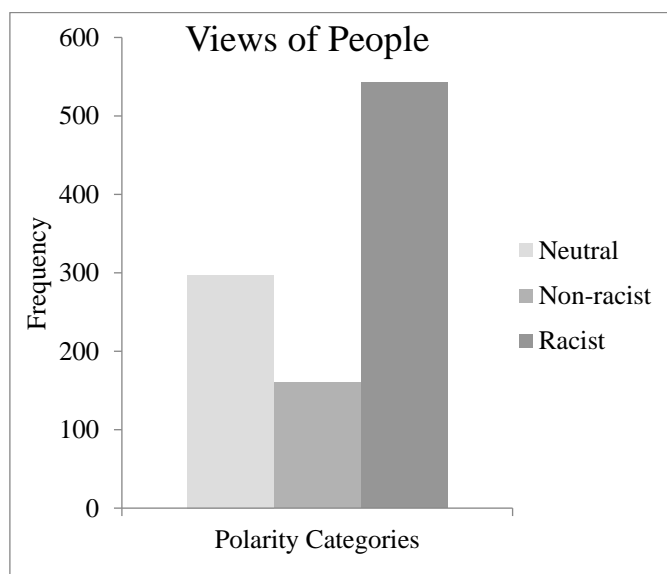


Figure 5.19: Graph of racism analysis (Keywords: Bollywood or racism or hate)

Racism score of the tweets with the keywords: *Bollywood or racism or hate* is shown as a bar graph in Figure 5.20. Racism is represented by the “-“ symbol, which indicates the amount of racism, the “+” symbol indicates the amount of non-racism and *zero* indicates that the specified numbers of tweeters are neutral. The strength of racism increases from 5 to -8.

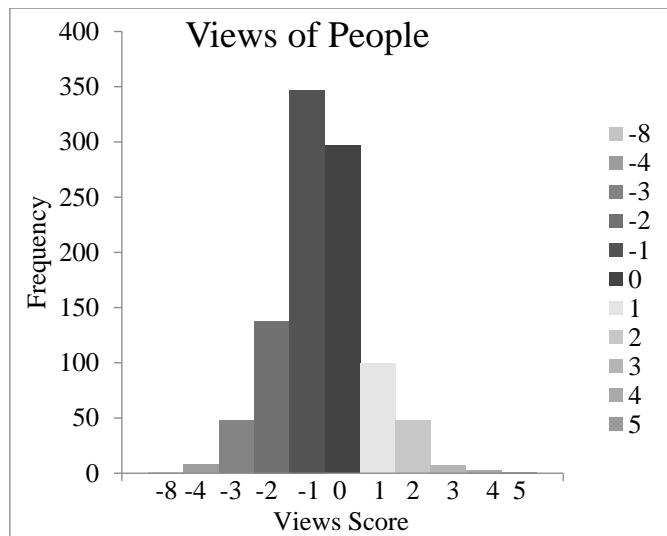


Figure 5.20: The score of racism (Keywords: Bollywood or racism or hate)

Racism analysis with Keywords Bollywood AND Racism OR Hate

To predict racism from the tweets with keywords *Bollywood and racism or hate*, a racism detection method is applied. This dataset consists of only 65 tweets that are more relevant to the topic of study and spanned for 10 days. Figure 5.21 shows a graphical representation of predicted results as racist, non-racist, and neutral. The columns of Figure 5.21 show the polarity of the tweets such as racist, non-racist, and neutral with the values 37, 6, and 22 respectively. These polarity values represent that out of 65 Twitter users 37 users are tweeted with racism, 6 users tweeted with non-racist messages and 22 users are tweeted with neutral statements.

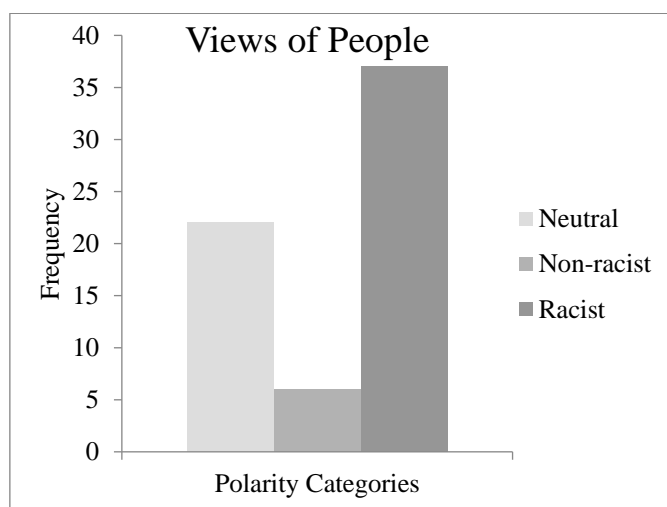


Figure 5.21: Graph of racism analysis (Keywords: Bollywood and racism or hate)

In Figure 5.22, a score of racism for the tweets with keywords: Bollywood and racism or hate is shown as a bar graph. Racism is represented by the “-“ symbol, which indicates the amount of racism, the “+” symbol indicates the amount of non-racism and *zero* indicates that a specified number of tweeters is neutral. The strength of racism increases from 2 to -4.

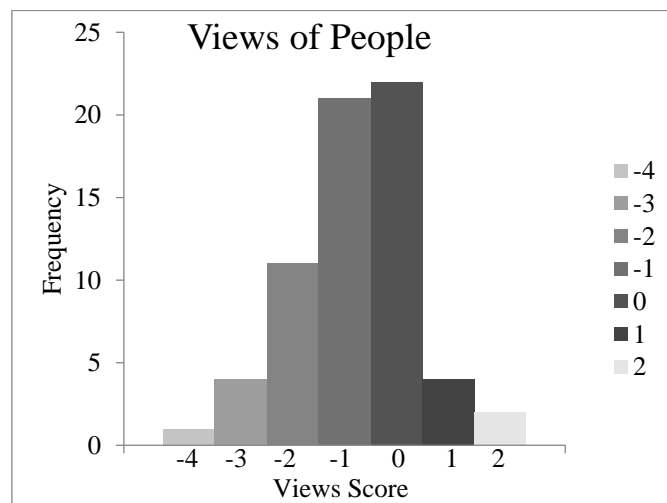


Figure 5.22: The score of racism (Keywords: Bollywood and racism or hate)

Figures for a score of racism reveal that, in all the four cases of analysis, the frequency/count of score values 0 and -1 are more than the other score values. This indicates that the majority of user tweets are neutral (‘0’ score) and very less negative (‘-1’). These score values of tweets depict the behavior of the citizen.

The summarization of an entire analysis is made in Table 5.2. For a specific set of keywords polarity, frequency/count, and the percentage of tweets are shown. The percentage of tweets for non-racist polarity varies from 9 to 22 based on the set of keywords used for extracting tweets. Racist polarity resulted in between 42 to 68 percentages concerning the keywords used. Similarly, neutral polarity exhibits a range of percentages from 22 to 36. According to Table 5.2, the racist tweets for all four different datasets are more than the non-racist and neutral tweets. This information leads towards the perception that there exists some amount of racism on social media in the Indian geographical area.

Table 5.2: Keyword wise analysis of tweets with the percentage

Keywords	Polarity	Frequency	Percentage
African OR Racism	Non-Racist	219	22
	Racist	416	42
	Neutral	365	36
African AND Racism OR Black	Non-Racist	4	11
	Racist	25	68
	Neutral	8	22
Bollywood OR Racism OR Hate	Non-Racist	160	16
	Racist	543	54
	Neutral	297	30
Bollywood AND Racism OR Hate	Non-Racist	6	9
	Racist	37	57
	Neutral	22	34

From the analysis, the racist tweet “*Why are Black #AFRICAN farmers protesting in our India by drinking piss?*” indicates that a direct attack on African based on color. As the tweet contains ethnic disrespect, it hurts the entire black community. Another example of a racist tweet is “*having black skin does not make you African*”. This expression attacks African citizen indirectly based on the color of skin.

Even though a tweet “Allah blesses everyone whether he is black, white, Muslim, Hindu, Christian, Asian, African, etc.” contains black words it is non-racist because it does not carry any derogatory message.

Similarly, an example from the analysis for a neutral tweet is “*Even when you hate him, you still love him*”. As the tweet contains two different aspects of expression such as hate and love, it is difficult to categorize as racist or non-racist.

5.5 Case of Physical Violence on Online Hate Content

In this section, a physical violence incident is considered for the case study. Problematic acts damage the target and create an environment of fear around the target. Hatred can be triggered by a problematic act. Problematic groups, often plan and coordinate their activities through social media. Problematic groups target youth to perform their problematic activities such as propaganda and incitement (Weimann, 2010). Social media is commonly used by problematic groups to enable communication among their members. Social media is promoting radical ideological activities like propaganda and the recruitment of members for problematic groups (Awan, 2017).

5.5.1 Case Context

Cyberterrorism is a kind of problematic act that uses electronic devices and Internet technology to create fear in society. The problematic acts influence (Deloughery et al., 2012) and are very “close like a cousin” (Mills et al., 2017) to hate content. Sharing of hate content through social media follows the problematic event. The transactional nature of cyberterrorism throws a challenge to local regulations and invites collective international effort to control it.

In the aftermath of any problematic incident, people exchange information, unite against the trigger event, pressure the government to act against problematic groups, collect funds for victims through social media services. Social networks will support counter problematic acts through flagging, sentiment analysis, honeypots, and facial recognition (Mahmood, 2013).

Social media users are the first to represent their reaction against the problematic groups after the trigger events such as problematic incidents (Williams et al., 2017). The online environment with millions of users provides insight into opinions on a large scale within a short time of the trigger event. Concerning problematic acts, several users of social media are arrested for their anti-national comments (Holbrook and Taylor, 2019; Karyotis, 2007). The literature review reveals that the problematic incidents are trigger events to generate hate content.

5.5.2 Methodology

Data collection is initiated immediately after the recent problematic incident in southern Asia. To collect data, Twitter social media is queried with appropriately identified keywords through its application programming interface (API). The data collection continued for a week after the problematic incident and resulted in 48819 tweets.

The tweets are extracted using different keywords through the Twitter application programming interface. After collecting tweets, a pre-processing mechanism is applied to Twitter content to remove punctuations, digits, and to stem the words. For each tweet, the sentiment and emotions are obtained using the national research council (NRC) lexicon. The sentiment and emotions for the entire content are based on the sentiments

of these individual tweets. The hatred feeling from the tweets against a violent act is interpreted through related emotions.

5.5.3 Results and Discussion

The tweets collected are analyzed for emotions using NRC lexicons. A graphical representation of different emotions for the entire tweets dataset is shown in Figure 5.23. The tweets which are posted over Twitter social media on a problematic incident contain more negative sentiments than positive. This reveals that most of the respondents are not happy and opposing the incident. Among the emotions, the fear is expressed through more tweets and indicates that people are frightened after the incident. The anger being associated with negative feelings is the next highest exhibited emotion through tweets.

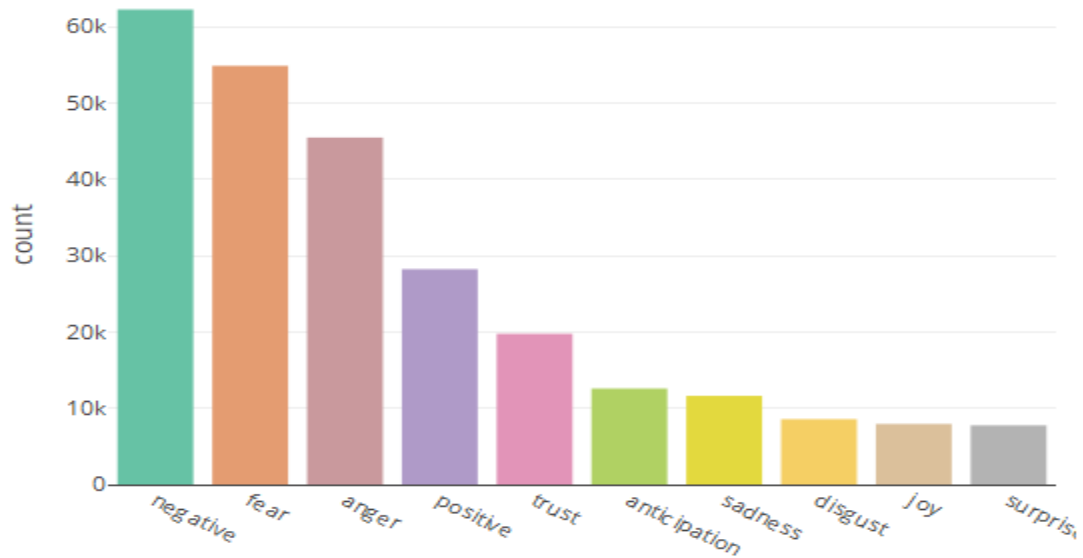


Figure 5.23: Emotions on the problematic incident

The presence of trust and anticipation emotions indicate that even with a lot of fear and anger still, people believe and have some expectation in the system. The score for sadness is more than joy and reveals that respondents are not happy with the problematic incident. The score for disgust emotion from the graph indicates that some respondents are unpleasant and condemning the act. Lastly, surprise is a reactive

emotion and results as a response to an unexpected event. The count value of the surprise emotion in the graph indicates that a small portion of the respondents is considered this incident as unexpected.

Each emotion is exhibited with a set of words. The contributing words for each emotion are shown in Figure 5.24 as wordcloud. As anticipation is associated with the prediction and represented through words like time, watch, happen, expect, etc. The anger possesses an uncomfortable reaction against a threat and is expressed using words such as an attack, shoot, and strike.

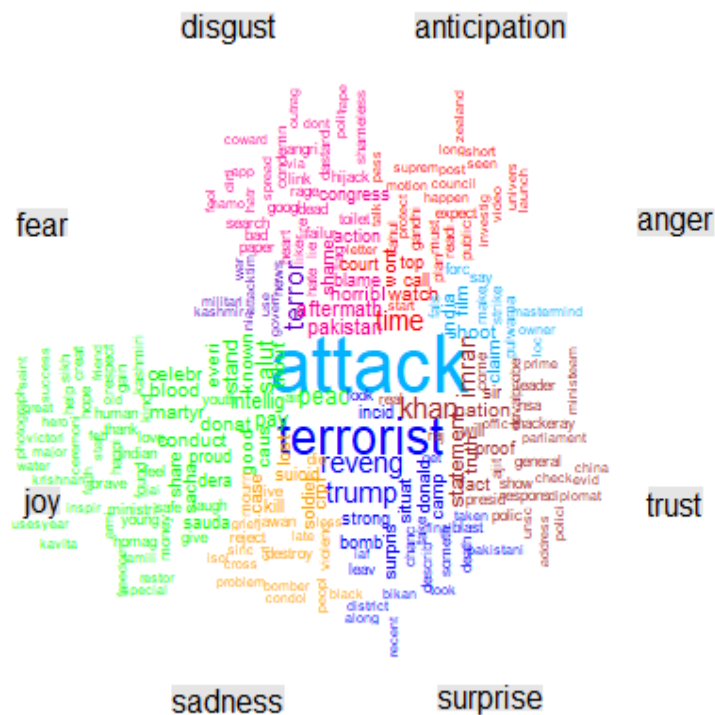


Figure 5.24: Wordcloud for emotions on the problematic incident

Trust is related to positive feelings and expressed with words like probe, truth, policy, and parliament. The words bomb, death, blast, etc. are used while reacting to unexpected incidents as surprise emotion. The sadness emotion being associated with negative feelings expressed using words like lost, destroys, dies, and mourn. Joy as opposite to sadness is exhibited using words peace, salute, good, pay, and intelligence. Fear is expressed using the words terror, military, war, government, etc. Unpleasant

emotion disgust is expressed through the set of words aftermath, horrible, shame, hate, and condemn.

The percentage of different emotions on the violent act is shown in Figure 5.25. Hate content shows more percentage of words with the emotions of anger, fear, disgust, and sadness (Mohammad, 2012). Hatred is expressed with the emotions of disgust, anger/fear, and devalue (Sternberg, 2003). In general, the score of emotions indicates that social media was contained some hatred against problematic groups after their problematic act.

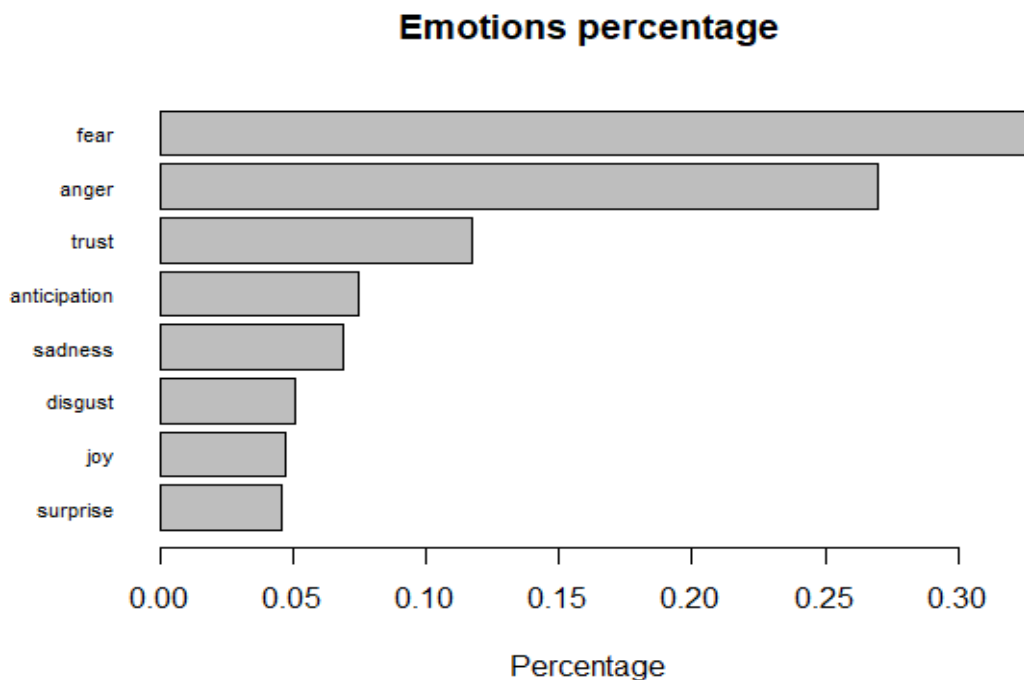


Figure 5.25: Percentage of emotions on the problematic incident

Out of the total tweets extracted from the Twitter media, 34177 are negative, 4658 are neutral, and 9984 are positive tweets. The quantity of negative tweets shows the possible presence of hatred content in social media after the problematic incident.

6. Concluding Remarks

The case of gender-based hate presents gender-based hatred analysis through social reform movements. Hatred-associated emotions such as anger, disgust, and sadness are expressed more through the MeToo movement. This indicates that the citizen-initiated

movements encourage more to express feelings of victims against the harassment or the worst situation they faced than the government-initiated movements. As the social reforms are assisted by information and communication technology, it can be concluded that technology acts as a vehicle for disseminating sentiments or emotions deeper into society at a faster rate. The sharing of hatred grievances is present in both the movements, but more with the citizen-initiated movement. The sharing of information on harassment or injustice by individuals with high magnitude acts as a mechanism to counter hate content. In the future, this may result in reduced hate content. Both hate and harassment are cohesive and combined as contempt feelings.

From the case of religious hate content, HBV is assumed as a cultural behavior and tries to preserve the honor of the family or community at the cost of the life of a family member. The basic cause for the occurrence of honor killing is the mechanism the victims used for their partner selection. Perpetrators assume that the partner selection mechanism of a family member brought shame to the entire family/community and that shame often results in HBV. With this ideology, the family members (individual or in the group) acted on the victim with cruelty. These incidents highlight that marriage against family members can be a major cause of the occurrence of HBV. HBV is prevalent in high context communities, in which the family/community is more social than the individual. It has been also identified that, after the initial occurrence, HBV continues as hate content in different forms. As the emotions such as fear, anger, disgust, and sadness are different dimensions to express hate content, they are supported by the social media analysis results. Increased abuse against communities can be controlled by addressing their issues on social media and with the identification of online communities by observing their behaviors such as messages they post, share, and like.

The case of racist hate content analysis reveals that the tweets extracted with specific keywords from Twitter are not only racist, but some are non-racist and neutral. This indicates that the part of Indians has thoughts towards the neutralization of harmful events. Results of score analysis on different datasets show that score values either 0 or -1 are having the highest frequency, which indicates that overall tweets of a dataset have no negativity or less negativity (racist). From the analysis, it may be perceived

that there exists some amount of racism in India. The interpretation is based on observed results from the racism detection approach, which may not depict the actual scenario of racism in India. Finally, we can conclude that results are encouraging in nature towards racism prediction; further development in the direction may lead to the building of an efficient system for the prediction and controlling of racism in India.

The case of physical violence reveals that the problematic act influences the generation of online hate content. Problematic groups use social media sites to plan and coordinate their acts. Social media users react to problematic incidents immediately, sooner than others. Hatred against the problematic incident is represented through different related emotions. The score value for emotions such as fear and anger is more than the other emotions and reveals hatred. As there are more tweets with negative sentiments, almost thrice the positive sentimental tweets, the outcome supports the presence of hatred against problematic incidents.

CHAPTER 6

SYNTHESIS, IMPLICATIONS AND RECOMMENDATIONS

6.1 Introduction

The previous chapter analyzes and reports individualistic, community-related, ethnic, and physical violence expressions as case studies using a qualitative approach. Each case study reflects the nature of hatred based on different protected characteristics such as gender, religion, and race. Apart from protected identities, a case study on the problematic incident which leads to the expression of hatred is analyzed and reported. This chapter tries to integrate the results of different studies involved in research.

The content of the chapter is structured as follows. Section 6.2 synthesizes the findings of quantitative and qualitative approaches. The implications of the study are discussed in section 6.3. In section 6.4, the recommendations are made to improve the control of online hate content. Section 6.5 concludes the chapter.

6.2 Synthesis of the Study

Synthesis is a process of evaluation of available information from the field of study to generate composite knowledge. It is a collective pulling of information from multiple sources to answer a research question. Research synthesis draws generalizations from research by integrating empirical researches (Hedges and Cooper, 2009). Synthesis of findings in research is associated with a review and summarization of the concerned study. It supports the identification of areas with agreements or discrepancies in a phenomenon. As a single study cannot cover the breadth of research, the synthesis is spawned over multiple studies to accumulate wider concepts (Mosteller and Colditz, 1996).

As discussed in chapter 3, the synthesizing of research can be attained through quantitative and qualitative study approaches. The synthesis of findings that are obtained by applying these approaches in chapter 4 and chapter 5 are discussed in this section. This section provides modeling variables and a synthesis of findings.

Understanding and Modelling Variables

As mentioned earlier, the research question is “How to improve the quality of online hate content control in India?”. To answer this research question, both quantitative and qualitative approaches are adopted. The quantitative approach is employed to identify the relationships among the variables (gender hate content, religious hate content, racist

hate content, disability hate content, and online hate content) derived from the literature and the relationship between the variables and their constituent components. The qualitative approach is applied to secondary data (Twitter content) to ensure the association between the variables and their components and to identify the additional components of variables and their relation with the components. Both the approaches are employed in series, one after the other to avoid dependency issues. The quantitative study identifies the relationship among the components and the qualitative study ensures that relationship and reveals the new variables. The rest of the section discusses the connectedness of the approaches used and their comparison towards the control of online hate content.

As mentioned earlier, the quantitative approach is used to analyze the relationship among the different expressions such as individualistic, community-related, ethnic, health-related, and global and between the expressions and their components. The quantitative analysis is performed in three steps. The first step identifies the relationship between the variables and their components, the second step is used to identify the relationship among the variables and the third step analyzes the policy perspectives. The policy perspectives are analyzed separately to identify and recommend the findings for improving the control of online hate content in India.

On the other hand, the qualitative study is performed in a layered manner. First, the study is used to analyze Twitter content in terms of advanced sentiments including different emotions to identify the relationships between the individualistic expressions and their components. Second, the study performs the analysis of online news content to identify the relationship between community-related expression and its components and to identify new variables. The third layer analyzes the Twitter content for basic sentiments to ensure the relationship between the ethnic expressions and their components and to identify new variables. Fourth, the study identifies the relationship between global expressions (online hate content) and their components through advanced sentiments.

It has been observed that the steps of both quantitative and qualitative approaches associate the people, process, and technology together. The association indicates that though the studies are different, their underlying structure is similar. The similar nature

of studies allows them to conduct under the same conceptual framework. Therefore, the findings of both approaches can be associated and validated easily during the triangulation process.

The analysis from a quantitative study shown that the individualistic, community-related, ethnic, and health-related expressions could influence the global expressions (Online hate content). It is also shown that the terrorist incidents, extremism, freedom of speech, and communal nature contribute to global expressions. The qualitative studies revealed that personal experience, a stance of superiority, skin color and fright of health influences online hate content through gender-based, religious, racist, and disability hate content respectively.

From the literature review and quantitative analysis, different information such as generation, propagation, and imparting of hate content has been identified. Online hate content will be generated more by the persons who experienced hatred, having an attitude, lack of religious knowledge, a stance of superiority, ethnic unfamiliarity, the fright of health, ignorance, extremist behavior, communal nature, watching stereotypes, differentiating skin color and observing terrorist incidents.

To validate identified variables from quantitative study and identify the new variables, different case studies are conducted using social media data. Different variables which have been identified from the case studies are shown in Table 6.1. In Table 6.1, social movements represent gender-based hate, honor-based violence represents religious hate, racist incidents represent racism, and Trigger incidents represent global hate incidents. These identified variables are presented under different cases below.

Case i: Social movements, honor-based violence, racist incident, and trigger incident

Expression of hatred is common in all the cases of Twitter social media and other online data analysis. However, the expression of hatred is triggered by different incidents. In the case of social movements, the hatred is expressed as a response to sexual abuse faced by one gender from other. In an honor-based violence case, the hatred is exhibited to retain the honor of the family or community. With the racist incident, the hatred is backed by superiority over others. In case of a trigger event, hatred is expressed against the offenders for their violent acts.

Table 6.1: Variables and practices identified from social media data analysis

Variables/Themes	Social Movements (MeToo & Temple entry)	Honor-based Violence	Racist Incident	Trigger Event (Violent Act)
Hatred	√	√	√	√
Violence	√	√		√
Harassment	√		√	
Insult	√	√		
Religion	√	√		√
Gender	√	√		
Race			√	√
Age	√	√		
Skin Color			√	
Men		√		√
Women	√	√		
Health risk	√		√	
Depression	√		√	
Suicide	√	√	√	
Policies	√	√	√	
Campaigns	√			√
Government	√	√		√
Personal experience of hatred	√		√	
Feeling of superiority		√	√	

Case ii: Social movements, honor-based violence, and trigger incident

In this case, the identified variables such as violence, religion-based incidents, and the government’s presence are discussed.

Violence: The violence is observed in the analysis of data for social movements, honor-based violence, and trigger incident. As violence is mostly physical activity, it harms more to the human being. Though the violence takes place differently in different cases or scenarios, the impact of violence is harmful.

Religion-based incidents: The occurrence of this kind of incident is common in all social movements, honor-based violence, and trigger incident cases. Though the temple entry social movement concerning gender identity, also involves religious rituals and rises the issues related to religious sentiments of the citizens. According to the honor-based study, the violence is exhibited by higher community people for protecting the

honor of their religion or family. In case of violence or trigger incidents, the violence is exhibited at an extremely severe level.

The presence of government: In social movements case, the temple entry movement is initiated by the government but the citizen-initiated movement (MeToo) demands the government's presence for the control of sexual abuse and related acts. The incidents of honor-based and physical violence both require the presence of government to reduce the acts like honor killing, terrorism, etc.

Case iii: Social movements, honor-based violence, and racist incident

Committing suicide and demanding policies are the common variables resulted from the analysis of data on Social movements, honor-based violence, and racist incident cases.

Committing suicide: Concerning social movements MeToo, often, the victims of sexual harassment commits suicide. In case of an honor-based incident, the committing of suicide is by the adults involved in love affairs. They commit suicide because of obstruction to their marriage and difficulties they face after the honor-based act against them. As the racist incidents are mostly with minorities and based on physical appearance, the victims of racism feel bad and, often attempt suicide.

Demanding policies: The policies are necessary to maintain public order. The analysis of tweets and other online content revealed that policy is a commonly occurred word in case studies data. In the case of social movements, the policies are required for fighting against sexual abuse, retaining access rights, and security. The khap panchayats are involved in honor-related incidents. Mostly, these panchayats decide in favor of higher communities by neglecting the rights of minorities and the persons involved in love affairs. To control khap like panchayats the presence of policies is required. Often, racism is exhibited against the outside nationalists. The appropriate policies could keep migrants safe in the present country.

Case iv: Social movements, and honor-based violence

Insult, gender identity, age-related acts, and women are the common variables in social movements and honor-based violence cases. Twitter content of these cases has shown that there is a discussion involving these phrases.

Insult: Insult results when a person is disrespected and shamed. The feelings of insult are bitter. In the case of social movements, the people are insulted by abusing, denying permissible rights, and inequality among different gender and communities. Honor-based violence incident insults the people involved by interfering in their personal life.

Gender identity: As the social movements are initiated to bring equality among the different gender and religious sentiments, the occurrence of gender words is prevalent. Mostly, the honor-based incidents are related to marriages and end with the honored killing of a family member.

Age-related acts: As the MeToo movement initiated against the harassment mostly on adults, age is a common factor in this analysis. On the other side, temple entry is restricted based on the age and gender of a person, hence its occurrence is common. The occurrence of the word age in social media data indicates that the age of a person decides the rights.

Women: The word is common in the data sets of both case studies as the victims are mostly women. To cope with the situation, the empowerment of women is essential.

Case v: Social movements, and racist incident

Harassment, health risk, depression, and the personal experience of being hatred are the commonly occurred variables in the data sets of social movements, and racist incident case studies.

Harassment: Harassment is a common word and results from multiple incidents. The MeToo movement has raised the issue of harassment like sexual abuse against the perpetrators of the offense. The victims of racism also suffer from harassment.

Health risk: In case of any unwanted incident, often, health issues are reported. The people exchanged the content involving the term health risk on social media. Concerning the case studies of social movements and racist incidents, the health risk here is mostly a mental health issue.

Depression: Depression is a kind of health-associated issue and exhibits the mental status of a person. Often, the victims of both sexual harassment and racism face depression.

Personal experience of being hatred: The analysis of both social media and racist incident data sets revealed that the persons who are exposed to hatred incidents express hatred against others in the future. Identifying and educating the victims of hatred is essential to avoid future incidents of similar kinds.

Case vi: Honor-based violence, and racist incident, and racist incident

The feeling of superiority is a common variable that occurred in the data sets of both honor-based violence, and racist incident case studies. The analysis reveals that in honor-based violence, people express hatred or violence by thinking that their community is superior to others. Similarly, a racist incident also involves violence with the assumption that one's race is superior to others.

Case vii: Social movements, and trigger incident

Campaigns: Campaign is a common term that occurred in social movements, and trigger incident data sets. This indicates that the organization of some of the campaigns to educate the people is required to reduce future incidents of violence. The violent actors use social media to campaign and recruit the people who are ready to work for their groups.

Case viii: Honor-based violence, and trigger incident

The variable *men* commonly occurred in honor-based violence, and trigger incident case studies. This indicates that due to the patriarchal nature mostly, the men exhibit violence or hatred towards women. The trigger or violent incidents like terrorism involve mostly the men to create fear in the environment.

Case ix: Racist incident, and trigger incident

Race is a common term discussed with both racist and trigger incidents data sets. This indicates that racist hate is expressed by considering the race of a person. After the terrorist incident also people try to express hatred against the perpetrators based on race or religion.

Case x: Racist incident

Skin color is a variable that is identified from the racist incident. Often, people express hatred based on the physical appearance of a person. Mostly, the victims of racism are people of color.

6.2.1 Synthesis of Survey Findings

The survey analysis is made for the identification of different determinants and the policy perspective of online hate content. The initial analysis revealed the different determinants and contributing variables to online hate content. The policy perspective analysis provides different suggestions to reduce online hate content. The synthesis of both analyses is presented in Table 6.2.

Table 6.2: Synthesis of survey findings

Determinants perspectives for online hate content	Policy perspectives for online hate content
An attitude towards gender differences causes online hate content	Both governmental and non-governmental policy attributes are important to reduce online hate content.
An attitude towards religious sentiments causes online hate content	
An attitude towards race/ethnic cultures causes online hate content	Strengthening of legislation, increasing penalties, and campaigns by the government reduce online hate content
An attitude towards disabilities causes online hate content	
Alienated, inclination and personal experience of hatred are the contributors to gender-based hate content	Campaigns by private sectors, civil societies, and efforts by social media reduce online hate content
The stance of superiority, ignorance, and religionizing the nation are the contributors to religion-based hate content	Hate content can be reduced by taking precautions during the online presence of an individual
Clothing style, inherited nature, peer pressure, stereotype, the unfamiliarity of ethnicity, personal experience, skin color, and poverty are the contributors to racist hate content	Hate content can be reduced by spending minimum time in a virtual world
Ignorance, envious, the fright of health, and being able-bodied are contributors to disability hate content	
Communal nature, terrorist attack, freedom of expression, and extremism are other contributors to online hate content.	

The integration of results from the survey analyses reveals that the attitude towards gender difference, religious sentiments, ethnic cultures, and disabilities are the determinants of online hate content. The analysis also reveals that these attitudes are

influenced by different variables separately based on the protected characteristics. Though the government and non-government policy attributes are required to reduce hate content, the non-governmental policy attributes are more important than the government policy attributes. With this information on hatred influencing variables and the policy perspectives, the mitigation of hate content may be possible.

6.2.2 Synthesis of Gender and Religious case Findings

Two different case studies are conducted to analyze social media data on gender-based and religious hate content. As a part of the gender-based hate content study, social movements such as MeToo and Temple entry cases are considered. The synthesis of findings from both studies is shown in Table 6.3.

Table 6.3: Synthesis of gender-based and religion-based case findings

Case of gender-based hate content	Case of religion-based hate content
The persons who were exposed to hatred in the past may be perpetrators of hate in the future	The persons with an intention to religionize the nation often, exhibit hatred on others
The social media users expressed that hatred and violence are common on the platforms	The social media users expressed that hatred and violence are common on the platforms
Attitude towards both gender differences and religious sentiments are expressed on the platforms	Attitude towards both gender differences and religious sentiments are expressed on the platforms
Health risks, depression, and suicidal attempts are observed from the data analysis results	The suicidal attempts are observed from the data analysis results
Harassment and insulting acts are observed from the analysis results	An insulting act is observed from the analysis results

The expression of hatred on social media based on gender and religion is common. The personal experience of being exposed to hate content and an intention to religionize the nation could be the influencing factors to online hate content. There are the possibilities of health issues, depression, and suicidal attempts with the victims of hatred. The studies revealed that harassment and insulting acts are also taking place on online platforms.

6.2.3 Synthesis of Racist and Physical Violence Cases Findings

In this section, the synthesis of the findings from two different studies such as racist and, physical violence incidents are presented. Racist and, physical violence incidents are global incidents as they are not associated with a particular nation. The synthesized results of these studies are shown in Table 6.4.

Table 6.4: Synthesis of racist and physical violence incidents case findings

Case of a racist incident	Case of physical violence incident
The skin color or physical appearance of a person is an influencer of hate content	An incident of terrorist act influences the hate content over social media
Racism is a more harassing act	Terrorism is a more violent act
Health risks, depression, and suicidal attempts are observed from the data analysis results	Mostly, men are involved in the activities
Superior feelings about the race may influence hatred	Requires immediate interference by the government
Policies related to immigration are essential to control hate content	Campaigns are organized for recruiting persons to carry out violent acts

Often, people will perceive the association of one's nationality based on the physical appearance of that person. This perception concerning skin color could be a cause of racism. A terrorist act is one of the causes of online hate content. Sometimes, the victims of the incidents may face health issues and suicidal attempts. Appropriate policies and immediate government interference could reduce the magnitude of online hate content.

6.2.4 Synthesis of Survey and All Cases Findings

The synthesis of the results the of survey and different case studies is shown in Table 6.5. The attitude towards gender and religion are drivers of online hate content and are supported by survey, case of gender and case of religion results. Attitude towards race driver of online hate content is supported by survey and case of racism results. Similarly, the other driver of online hate content attitude towards disability is supported by survey results only.

Experience of hatred is a contributing factor to gender hate content and is supported by survey and gender case studies. Religionizing the nation attribute supports the religious

hate content and is resulted from survey and case of religion. Skin color attribute contributes to racism and supported by survey and case of racism results. Similarly, the terrorist act contributes to online hate content and supported by survey and the case of physical violence results. In this way, the findings of case studies are weaved with the findings of different case studies.

Table 6.5: Integration of findings

Finding	Survey	Case of gender	Case of religion	Case of racism	Case of physical violence
Attitude towards gender	√	√	√		
Attitude towards religion	√	√	√		
Attitude towards race	√			√	
Attitude towards disability	√				
Experience of hatred	√	√			
Religionizing nation	√		√		
Skin color	√			√	
Terrorist act	√				√

Collectively, the current research provides different determinants of online hate content, policy-related attributes to reduce hate content, possible health issues from hate content, and the actions required to control hate content.

6.3 Implications

In this section different implications of the research to government and online service providers, citizens, researchers, and theory are presented.

6.3.1 Implications for the Government and Online Service Providers

The Indian Government is already working for the control of fake news and digital hate content. In this regard, the government authorities conducted several meetings with the concerned Information and Communication Technology industrial giants such as Microsoft, Facebook, and Google. Recently the incidents such as mob lynchings fuelled by fake news took place in several Indian states. From the synthesis of results, we know the predictors of online hate content along with their contributing variables. Therefore, instead of making only the strict rules to avoid hate content propagation on the Internet, some precautions can be taken to control it by preventing the occurrence of its

contributors indirectly. Some of the possible actions to avoid the occurrence of indicators of predictors are briefed in the following paragraphs.

The discussion on gender-based hate revealed that alienated/divorced, inclination/attitude, and personal experience are the contributors to its generation. To avoid divorce among the couples, the couples can be well educated on the possible harms of divorce on them as well as their kids. In this way, the hate content can be controlled by avoiding divorce among the couples to a certain extent. As the attitude of a person is one of the contributors to gender hate and is essential to shape that properly. This can be achieved by providing education on attitude shaping to youngsters at the school level. Once if a person is a victim of online gendered hate content, then that person's perception towards the perpetrator/sex of a perpetrator may be violent. This may lead to exhibit hate against them. Therefore counseling/educating the victim for persuading him/her to avoid propagation of hate content as revenge is essential. To control harassment against feminists such as threats of rape, the do-it-yourself approach can be used (Jane, 2016).

The discussion on religious hate content ensures that stance of superiority, ignorance, religionizing the nation, and clothing style of a person are its contributors. Some people will perceive that their religion is superior to all other religions and start hating other religious people. Here it is essential to educate them on the equality of religions and their relationships. Ignorance being the lack of knowledge about other religions leads to a misperception about other religious traditions, activities, and programs. This can be reduced by conducting interreligious competitions and making them public to participate by any religious persons. Some extremist groups may intend towards religionizing the nation. Identifying these groups and educating them on de-religionizing the nation may help to reduce the impact of this nature. The clothing style of a person contributes least to the religious hate content but can be reduced by providing awareness among the population that the freedom of wearing is like freedom of expression. Collectively, religious hate content can be controlled by analyzing the activities of perpetrators such as information they post, share, and like of online communities (Awan, 2016). With the help of these actions against hate content towards religion, the generation of online hate content can be reduced.

The discussion on racist hate content reveals that inherited nature, peer pressure, stereotypes, the unfamiliarity of ethnicity, personal experience, skin color, and poverty are its contributing variables. The family of a person is responsible for his/her inherited nature. Every family has to be instructed for providing a good environment for their kids so that they inherit good nature. Peer pressure is a result of being in groups and these groups may be related to work, schools, or any other environment. By forming some monitoring bodies over the groups the absence of peer pressure can be ensured among the members of the group. Stereotypes are more powerful mechanisms than the written or graphical communication to spread propaganda. This effect can be reduced by blocking the circulation of harmful stereotypes and propagating videos that show integrity among the population irrespective of tradition, ethnicity, religion, etc.

The unfamiliarity of ethnicity is due to less awareness of other traditions. Ethnic familiarity can be achieved by providing awareness information on ethnicities among the citizen. As seen in the gender-based hate content, the victim of racism is disappointed and might plan for revenge against the other ethnic groups or perpetrators of the racism in the future. Persuading the victim to forget the incident and concentrate on the future may reduce racism. Concerning skin color, it seems that the racism is within the colored persons also. Persuading citizens to accept the natural color of human beings and to treat them well may reduce hate content. Poverty is a global issue that can be controlled by providing at least the minimum wage of employment/work to every citizen. If the concerned authorities such as public and private administrations adopted these approaches, racism can be reduced to some extent.

According to the studies on disability hate content, ignorance, envious, the fright of health, and being able-bodied are the contributors to disability hate content. Ignorance like with religious hate content, here mentions the lack of knowledge on disability. Citizens can be educated on disabilities so that they start caring for persons with disabilities. Envious is a jealousy nature that harms others without any reason. A citizen can be encouraged to live without jealousy. Fright/fear of health is another contributor to disability hate content. This can be reduced by educating the people on kinds of diseases that are communicable and non-communicable. Being proud of the able-bodied is another reason for the disability hate content. This can be reduced by letting

the people know that no one is permanently able-bodied, all are temporarily abled. These are some appropriate actions to control and manage disability hate content generation.

There are other direct contributors to the generation of online hate content. One of the contributors to online hate content is the communal nature of a person. With the identification of persons of this nature, they can be nurtured to remove communal nature from them. The occurrence of a terrorist incident is the major contributor to online hate content generation. After the terrorist incident, the flow of content over social media can be monitored by concerned authorities to identify the perpetrators of the hate content. A counter-attack or other actions can be taken to control online hate content. Freedom of expression is another reason to online hate content with the least contribution. As it cannot be removed completely but can be restricted to a certain extent by implementing some more laws like section, 295(A) of the Indian penal code (IPC). This section of IPC punishes hatred expression exclusively on religion with imprisonment of a maximum of three years or fine or both (Indian Penal Code, 1860; Law Commission of India, 1971). Extremist activities are carried out usually in groups. Avoidance of framing the groups and educating the framed groups can result in reduced online hate content.

The majority of the above-discussed approaches to control online hate content are required to work in collaboration with the government, online service providers, and citizens. These are not like framing and implementing the rules from administrative views. They are difficult to adopt but results appropriately. The online service providers can act as the campaigning platforms and content moderators to control hate content.

6.3.2 Implications for Citizens

With increased online hate content on social media, an individual and the communities are getting hurt and losing confidence in life leading. Terrorist activities are also coordinated using social media and other Internet services. The interconnected acts of fake news, terrorism, and online hate content degrade the social life of the citizen resulting in reduced social health of the country. The safety of its citizens is vital to every nation. The identified contributors to hate content can be used by the citizens themselves to shaping their behavior. As an example, the efforts by an individual can

be made to forget the incidents of hate content exposures in the past. Similarly, by referring to other hate content contributors and taking self-corrective actions, the citizens may be safe in the online world.

The research revealed that *online precaution* is an essential variable to reduce hate content. Therefore, the citizens can take utmost care during their presence on online platforms (Costello and Hawdon, 2020). For example, they can avoid establishing a relationship with unfamiliar persons and discontinue the existing relationships with unknown persons online. Similarly, other online activities can be carried out with caution. Hence, the exposure to harmful hate content could be reduced online.

Limited online presence is another variable that causes a reduction in exposure to online hate content. In the present cyber days, it is difficult to have a limited presence online but could result in the least exposure to online hate content. For the betterment of their lives, the citizen should adopt this feature.

6.3.3 Implications for Management

Digital content plays an important role in shaping the business of an organization. The content which flows on social media affects the capital of the firm. The social capital of a firm can be modeled with the activities of platforms such as Wikipedia, blogs, and search engines (Paniagua and Sapena, 2014). Hate content at the workplace affects the mental health of the victim (an employee) and in the market demote the products. Ultimately, these effects may degrade the status of the company and reduce the number of customers.

Better governance can be achieved by serving online to the citizen. Hate content will affect the performance of governance by creating groups among the public and hating each other. Therefore, it is essential to monitor the generation of hate content and take corrective actions to control the diffusion of hate content. Every business organization can designate a person to take care of online content related to company business. In case of any negative reviews about the company or products, the designated person should analyze and identify the facts. If there exist issues with the products, the designated person should accept the fact and assure the customers of better service in the future. If the reviews are identified as fake, it indicates that some persons are

expressing hatred on products to take revenge. In this case, the designated employee should counter these negative reviews by counterspeech.

The management of an organization can emphasize identified contributors to online hate content for policy-making for the organization. For example, the management can monitor and avoid peer pressure on an individual and making unwanted groups. On the other hand, the management can conduct campaigns to educate employees and shape the behavior concerning the identified contributing variables to hate content.

6.3.4 Theoretical Implications

The theoretical framework designed and discussed for analysis of hate content in this study can be a basis for researchers in the field. As the different predictors of online hate content and their reinforcing contributors are identified, these contributing variables can be used by the researchers to frame and test their research models concerning online hate content. The predicted influencing factors act as drivers of online hate content and reveal the attitude of an individual towards different protected characteristics. The social structure-social learning theory (Akers, 2009) deals with the behavior of an individual and could be used to explain the identified hate content drivers in the study.

The conceptual model is tested with the opinions of respondents concerning the nature, forms, and policy perspectives of online hate content. Hence, the results are encouraging and could update the domain knowledge of the research field. The General Theory of Crime (GTC) depicts that the lack of self-control causes the production of online hate content (Gottfredson and Hirschi, 1990). This theory also explains extremist acts in society. Therefore, this can be used to explain the superior feeling about the religion influencing factor of hate content.

6.4 Recommendations for Improving the Digital Content Governance

Regulating digital hate content is very much essential to maintain the dignity and decorum of the citizens. One of the common tools used to control online hate content is blocking the messages or user accounts (Mathew et al., 2018). On the other hand, social media platforms are also taking several actions to combat hate content but unfortunately, they are not so effective.

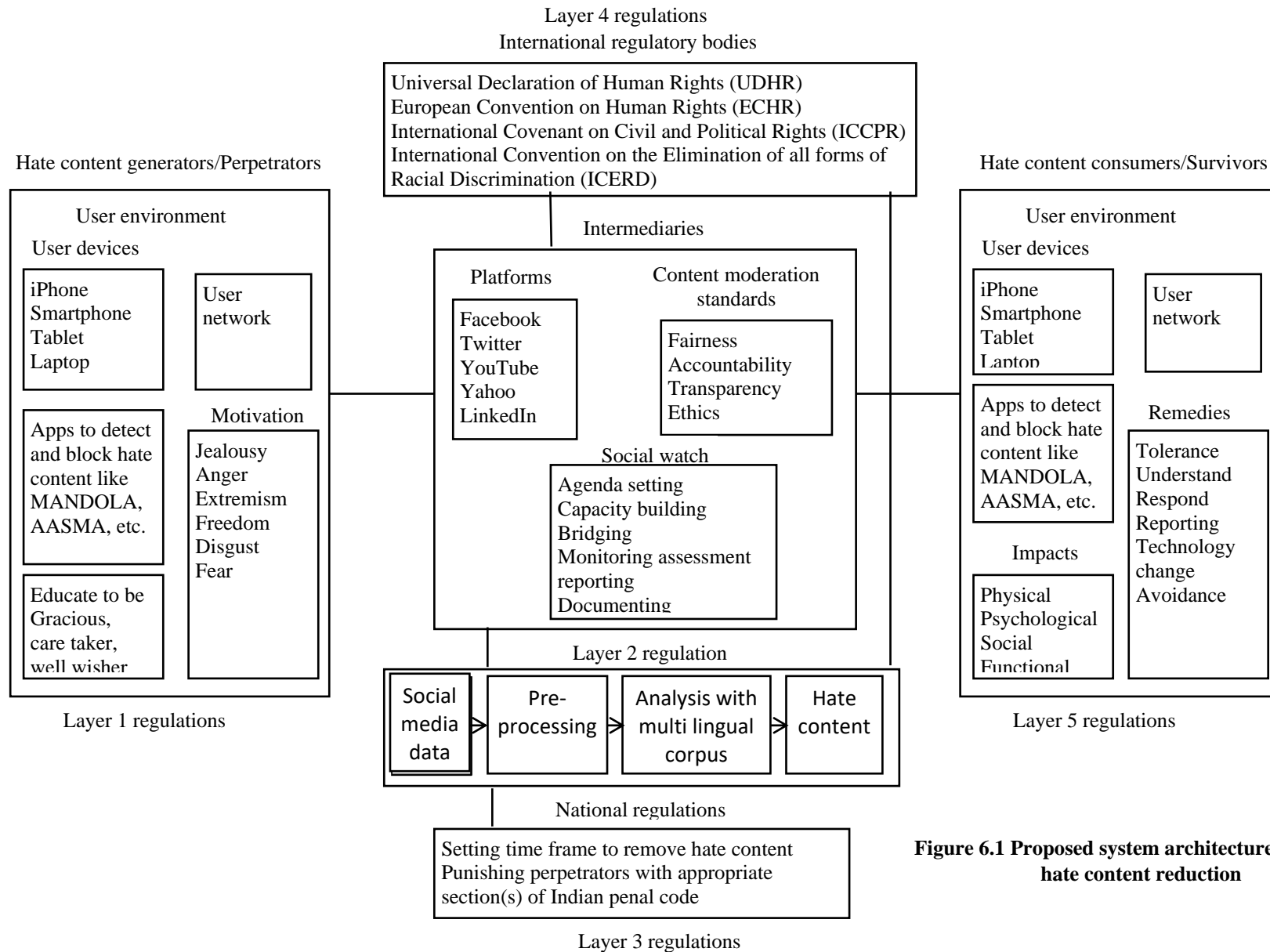


Figure 6.1 Proposed system architecture for digital hate content reduction

Therefore, it is necessary to adopt an appropriate system architecture to control online hate content. With the information gathered from the synthesis of results, a system architecture for the purpose is developed and shown in Figure 6.1. As the online world involves a variety of stakeholders such as users, intermediaries/social media, and national and international governmental authorities, the control of hate content by regulating at one level is difficult. Therefore, an architecture is proposed to control digital hate content at different levels or layers such as users, intermediaries, national and international levels.

Layer 1 regulations: Users access the online world by operating from a mobile edge environment. When users expose to the online world, prevalently, they face online threats such as harassment, abuse, hatred, and fake news. The user devices such as iPhones, smartphones, tablets, and laptops are the generators as well as consumers of digital hate content. A mobile edge network is equipped with a router and connected to different user devices. These networks assist user devices by providing computing and caching capabilities in the user environment (Wang et al., 2017). Generation of hate content is motivated by characteristics such as jealousy, anger, sexual desire, extremism, freedom, and disgust.

The hate content can be controlled by adopting some policies at the user level. The hate content detection and blocking tools or apps such as MANDOLA and AASMA can be installed on user devices or networks. On the sender/perpetrator side, these tools first check the information to be sent by the user for the existence of hatred. If the hatred exists with the information, using predefined policies, appropriate actions can be taken to block outgoing information. Apart from restricting the sender to send hatred messages, the mobile environment can be used to educate the sender to possess gracious, caretaking, and well-wishing characteristics.

At the receiver/survivor side, the tools check the information received by the user for the presence of hatred. If the hatred exists with the information, using predefined policies, appropriate actions can be taken to block the information. As the impact of hate content depends on the perception by the receiver, the mobile environment can also be used to educate the receiver on tolerance, understanding, responding, reporting, avoidance, and technology change.

Layer 2 regulations: This layer of regulation is adopted at the intermediaries' level. The regulations at this level can be strengthened by penalizing intermediaries for weak actions against hate content. The platforms through which the hate content may be shared are Facebook, Twitter, YouTube, and Linked In, Instagram, etc.

As a part of the regulation, social media sites should possess fairness, accountability, transparency, and ethics. Fairness can be achieved through sharing best practices and addressing legal issues by the platform. Accountability deals with processing complaints on content moderation and establishing fines and other measures to control hate content. Transparency can be maintained by providing guidelines for reporting complaints and data sharing with others. Ethics involves codes of conduct to address harmful content and other related issues.

The task of agenda-setting prioritizes the issues of hate content and supports community involvement in addressing issues. Capacity building supports relevant discussions, solutions, and best practices for addressing the issues. Awareness-raising campaigns, live broadcasts on the impacts of hate content, and regulations updates can be used to bridge the information gap. Regular monitoring of content on the platform and assessing for its quality is to be supported by the respective media. The amount and nature of digital content can be maintained in the form of documents for analysis purposes.

Layer 3 regulations: The regulations in this layer will be formed by the national government authorities. The governance authorities or public institutes can establish centers for monitoring digital hate content. These centers can collect data from intermediaries or social media, pre-process it, analyze using multi-lingual corpus and report the nature of the content. If the content exhibits hostile behavior, it will be blocked or deleted, and actions will be initiated to punish the perpetrator.

As part of the regulation, based on the harmfulness of the content, the time limit for deleting the content will be set. Currently, in India, the time limit for deleting harmful content is 24-48 hours after the occurrence. Punishing the perpetrators can be based on the harmfulness of the content and its targeted purpose. The sections of IPC such as section 195 required to punish the perpetrators.

Layer 4 regulations: The international legal frameworks can be used to form the regulations in this layer. As the differences in national jurisdiction exist, these regulations in isolation cannot punish hate-content perpetrators. The actions against the harmful content at this layer can be initiated by international legal bodies such as UDHR, ECHR, ICCPR, and ICERD.

6.5 Concluding Remarks

Synthesizing results from different studies and applying hate content regulations may reduce the online hate content generation in the future. As an implication, the results can be useful to act on hate content and its control in the future. Several policy recommendations are made to punish perpetrators and educate the citizen on the impacts of online hate content such as the degraded social health of the country.

CHAPTER 7

CONCLUSION

7.1 Introduction

The research attempted to address the issue of improving digital hate content regulations in India. The summarization of the different studies involved in the research is made in this chapter. The research question and objectives are revisited to brief the purpose of the study. The chapter emphasizes the key findings from the research and significant contributions to the practitioners, researchers, and society at large. In the end, the limitations of the research and future suggestions for the research are highlighted.

This chapter is organized as follows. Section 7.2 summarises the findings of different studies. The research question and objectives are revisited in section 7.3. In section 7.4, the research question is answered. Section 7.5 highlights key findings from synthesizing the results of different studies. In section 7.6, the significant contributions of the research are discussed. The limitations of the research are outlined in section 7.7. Section 7.8 suggests future possibilities in the research direction. In the end, section 7.9 concludes the research work.

7.2 Summary of Research

The research findings are summarized according to the research objectives. In this section, the results of different studies such as quantitative and qualitative which are discussed in chapters 4 and 5 are summarized. To attain the third objective of the research, the outcome of the study on existing national and international frameworks for controlling online hate content is also summarized. Apart from the empirical and existing framework studies, the section also summarizes the results of integration which are obtained from combining multiple studies.

For a better understanding of the research, the findings are summarized according to the four objectives of the research. Firstly, the quantitative study to analyze and identify the relationships among the different categories of hate content and their relationships with contributing components (RO1) is summarized. Secondly, the findings of qualitative studies which are conducted to identify the contributing components to different categories of hate content and understand their influence on online safety (RO2) are summarized. Thirdly, national and international frameworks of digital hate

content control and a quantitative study to analyze policy perspectives (RO3) are summarized. Finally, some suggestions are made to improve hate content regulations in India (RO4).

Hate Expressions and Their Contributing Components

Research objective 1: To study the nature and forms of digital hate content in the Indian context.

Hate expressions are used to express hatred against the protected characteristics such as gender, religion, race, and disability. Each hatred expression is contributed by some attributes which are referred to as contributing components. To identify the relationship among the hate expressions and their components, a quantitative analysis is performed with the help of a questionnaire research instrument, which is constructed based on literature and expert opinions.

As discussed in chapter 4, the data has been collected through both offline and online modes. Data collection in both modes covers different parts of the country. For an online collection of data, online services like Facebook, Twitter, and E-mail are used. After the collection and integration of data, it has been processed and analyzed using software developed in R. A model for hate content analysis is constructed using structural equation modeling. The model is evaluated in two different parts-measurement and structural models. The findings related to hate expressions and their components are summarized as follows.

- As the responses are sought on the existence of online hate content, the percentage of expression about the existence of online hate content by female respondents is more than the male respondents.
- To measure the quality of responses, the respondent's online presence is analyzed. During analysis, it has been observed that almost 50 percent of respondents will be online for 10 to 40 hours per week.
- As smartphones are convenient and affordable, around 80 percent of respondents use a smartphone to connect with the Internet for their online activities. Around 15 percent of respondents use laptops for the purpose.

- The analysis of the structural part of the structural equation model revealed that gender-based or individualistic expressions could influence the generation of online hate content. The study also revealed that the percentage of gender hate influence towards the online hate content is 26.
- The analysis of the structural part of the structural equation model revealed that religious or community-related expressions could influence the generation of online hate content. The percentage of religious hate influence towards the online hate content is 23.
- The analysis has shown that racism or ethnic expressions could influence the generation of online hate content. The study also revealed that the percentage of racism influence towards the online hate content is 21.
- The study has shown that disability-based or health-related expressions could influence the generation of online hate content. The percentage of disability-based hate influence towards the online hate content is 11.
- The analysis of the measurement part of the structural equation model revealed that the communal nature of a person contributes to online hate content. The percentage of contribution is 33.
- The occurrences of problematic incidents are prevalent. The analysis has shown that the problematic incident like terrorist attack contributes to online hate content by 48 percentage.
- Freedom of expression is a fundamental right of an individual globally. It has been identified that the freedom of expression is a contributor to online hate content by a percentage of 22.
- Extremism possesses extreme views about religion, nation, or politics. The analysis revealed that extremism contributes to online hate content by a percentage of 40.
- The person with an alienated status is a divorced one. The results show that these persons may receive or express hatred based on gender differences. Hence, the alienated status of a person contributes to gender hate content or individualistic expression.

- The inclination refers to an attitude of a person. The person who possesses an attitude expresses hatred against gender. So, the inclination contributes to gender hate content.
- Experiencing hatred changes the behavior of an individual. The person who experiences hatred in the past could express hatred towards gender and race. Therefore, the personal experience of an individual contributes to gender and racist hate content.
- Every individual will feel that their religion is superior to others. The person who has a stance of superiority about the religion expresses hatred against the other religious persons. Hence, the stance of superiority contributes to religious hate content.
- Ignorance refers to a lack of knowledge about something. The person with a lack of knowledge about the culture, tradition, and language of others expresses religious hatred. Similarly, the person who has a lack of knowledge about health issues expresses hatred against the persons with disabilities. So, ignorance contributes to religious and disability hate content.
- Feeling proud about the nation and associating with religion refers to religionizing the nation. This association increases nationalism based on religion and encourages fighting against anti-national and religious activities. The persons with this nature may express hatred against other religious sentiments. Therefore, religionizing the nation contributes to religious hate content.
- The particular clothing style of a person may not be comfortable to others and could attract hatred from other persons. So, the clothing style of a person contributes to religious hate content.
- A person may inherit the racist behavior from the antecedent. The person who inherits this behavior could express racist hate content. Hence, the inherited nature of a person contributes to racist hate content.
- While working in a group, a person may be subjected to pressure from others to do something which may or may not be acceptable. The person who is subject to this pressure could express hatred against the race of others. So, peer pressure contributes to racist hate content.

- Extremist groups represent their views to followers through stereotypes. They also present videos that encourage someone to act against other ethnicities. The person who is exposed to the stereotypes of this nature could express racist hate content. Hence, the stereotype contributes to racist hate content.
- For an individual, it is difficult to be familiar with the ethnicity, culture, language, and other regional traditions of all races in the world. A person who is unfamiliar with other ethnic cultures could express hatred against other races. Therefore, the unfamiliarity of ethnicity contributes to racist hate content.
- There is no uniformity in the skin colors of persons globally. The people may be of different colors like white, black, brown, etc. This ununiform skin color could influence the hatred against other races. So, the skin color of a person contributes to racist hate content.
- Social capital varies from person to person. A person may be rich, poor, or middle class. The poverty of a person could attract racist hate content from others. Hence, poverty contributes to racist hate content.
- To safeguard persons with disabilities, the governments present some beneficial schemes. The implementation of these schemes may initiate jealousy among able-bodied persons towards persons with disabilities. The persons with envy could present hatred against persons with disabilities. Therefore, the envious contributes to disability hate content.
- Some disabilities may be due to illness of a person such as asthma. This illness may create fear among able-bodied persons and could influence the hatred against persons with disabilities. Hence, the fright of health contributes to disability hate content.
- Like skin color, persons with different abilities live in the world. Often, some able-bodied persons may feel proud about their abilities in comparison to persons with disabilities. These persons could express hatred against persons with disabilities. So, the feeling of being able-bodied contributes to disability hate content.

Digital Hate Content and Online Safety

Research objective 2: To understand the influence of digital hate content on online safety.

The digital hate content is impacting the safety of online users. While surfing over the Internet, the users can understand that how safe they are based on their exposure to problematic content. To attain this objective, four different case studies on racism, religious hate, gender hate, and hate triggering incidents are conducted with online data such as the content of Twitter social media and online news articles.

i. A study on gender and online hate content

Social reforms initiatives attempt to empower disadvantaged people in a society. The purpose of this study is to identify the impact of social-reform initiatives on gender-based hate content generation. Twitter sentiments for government and citizen-initiated participations are analyzed. The score and contributing words for each emotion are examined, and a comparison of both the reform participation is made. The summarization of the findings is made below.

- Citizen-initiated reform participation shares more hate content than government-initiated reform participation. The score of emotions like anger, disgust, and sadness is more in the citizen-initiated social reform movement than the government-initiated.
- The sharing of victim's issues on social media alerts others and encourages them to fight against the perpetrators. So, the sharing of online hatred information with high magnitude by citizen-initiated reform movements may result in reduced hate content in the future.
- The victims of hatred expressed their feelings and dislike about the perpetrators. A person who had been the victim of hatred in the past (personal experience) has expressed more hate content about the perpetrators.

ii. A study on religion and online hate content

Honor-based hate content is predominantly generated from family hate content and may affect mankind. Honor is perceived about family and religion. In the Indian context, analysis of multiple resources such as literature articles, the news reported articles and social media sites pertinent to honor-based hate content is less. Therefore, with the purpose to identify and understand the influencing factors and emotions of honor-based hate content, the case study is conducted. Following are the findings of this case study.

- In India, marrying a person against family members' ideologies is observed as one of the causes to express honor-based hate. This hatred is expressed with a perception that by punishing the person, the respect of the family or community, or religion can be retained.
- Neglecting the traditions of a family or religion is also one of the causes to express honor-based hate. The elders of the family feel that every member associated should follow the traditions of the religion.
- The stance of superiority influences to express hatred against the other religious persons. The members associated with a particular religion feel that their religion is superior or greater than all other religions and support extremist activities.

iii. A study on racism and online hate content

The occurrence of problematic behavioral incidents is prevalent globally, in India, media often interpret it as towards people of color or blacks. During the discussion about these incidents both online and offline, some groups consider that India is not free from racism and others view it as misinformation processing. Therefore, to evaluate and determine the possible scenarios towards racism in India, this case study is conducted. The findings of the case study are reported below.

- Racism is a global issue and targets communities based on their race or ethnicity. The results of racism analysis reported the possibilities of racism in Indian social media.
- The results of racism analysis also revealed that apart from the racist tweets, there exist some non-racist and neutral tweets. The existence of neutral and non-racist tweets indicate that some social media users are trying to neutralize and some are trying to mitigate the issue of racism,
- From the analysis of tweets, it has been identified that skin color is one of the causes of racist hate content expression. This indicates that basically, racism is based on the physical appearance of a person.

iv. A study on physical violence and online hate content

The problematic act damages the target and seeds the fear in the neighborhood often come under hate incidents. Social media sites are used for planning and coordinating

problematic acts. The problematic act is a trigger event that influences hatred feeling. The case study analyzes the aftermath of a problematic incident in the southern part of the Asian continent from Twitter content to understand the influence of problematic incidents in hate content expression. Following are the findings of the case study.

- Emotions can be used to express the attitude and opinion of an individual. The hatred against the problematic act is expressed through different emotions. Fear and anger emotions exhibit a high degree of emotions than the other.
- The analysis of results reported more negative tweets than positive tweets, almost thrice the positive tweets.
- After the problematic incident, citizens used to share their opinions over social media sites. So, the problematic incident is observed as one of the causes to express hatred against the act or the community of the actors.

Policy Frameworks

Research objective 3: To study policies and regulations addressing digital hate content

Policies and regulations are important to control the generation and spreading of hate content. To attain this objective different national and international policy frameworks are studied and the respondents' opinions are analyzed for policy perspectives.

i. A study on select national and international frameworks for hate content control

Legal frameworks are the tools to combat hate content. To address the issue of hate content control, selected national and international frameworks are studied. The summarization of the findings is made below.

- The international legal frameworks on hate speech reveal that all the framework laws on free speech are almost the same except ICERD and slightly different on hate speech.
- The constitutional and legal articles on hate speech from select different countries reveal that some nations are flexible and some are nonflexible in acting against hate speech. The flexibility to address the issue may lead to ununiformed laws against hate speech globally.

- In India, if there is any disrespect by a citizen or a group towards another citizen based on the protected characteristics, the existing laws try to punish the perpetrators.
- Article 19 of the Indian constitution provides the right to every citizen on freedom of speech and expression with the constraints to preserve morality, public interest, or decorum. There are several IPCs under different sections to punish perpetrators of hate content in India.

ii. Analysis of respondents' opinions for hate content control

Apart from studying different frameworks, the respondents' opinions are also analyzed for policy perspectives to attain the objective of policy regulations. The findings from the respondents' opinions analysis are reported below.

- Hate content can be controlled by the efforts of both governmental and non-governmental organizations. The efforts and policy criteria of both organizations are important.
- The variables strengthening legislation, increasing penalties, and campaigns by the government are the influencers to governmental policy attributes.
- The non-governmental policy attribute is influenced by the variables campaigns by private sectors, campaigns by civil societies, and efforts by social media.
- The self-caring of an individual by being precocious and limited time online could influence hate content reduction.
- The non-governmental communities- civil societies, private sectors, and intermediaries are more important to reduce digital hate content.

7.3 Revisiting Research Question and Objectives

In this section, the research question and objectives are revisited to justify their formation through the relevant information.

7.3.1 Research Question Revisited

Research Question: How to improve digital hate content regulations in India?

The research question is formed based on the following information.

- Apart from being the most populous country on the planet, India has multiple religions, cultures, ethnicities, and languages. The increased diversities from different backgrounds could result in dislike feelings among the people.
- As reported by the different organizations, religious hate crimes are increasing yearly in India. Mostly, the victims of these crimes are the minority religious people.
- Like religious hate crimes, often, crimes against women are also reported in India. These crimes include dowry, kidnapping, rape, assault, and insult.
- Based on the age of the people, the crimes are increasing in India. Mostly, the senior citizens are the victims of these crimes.
- Like offline crimes, cybercrimes are also increasing in India. Electronic and communication technologies are used for these kinds of crimes. These crimes are victimizing people on different levels such as economic, social, personal, and cultural.
- The ranking of the global gender gap index is low and requires to improve to attain goal 5 (gender equality) of the United Nation's sustainable development goals. The global gender equality index is based on several equal opportunities in the domains such as politics, education, employment, and healthcare.
- The rank for the social hostilities index of India is high at the global level. The measurement of this index is based on religious hostilities. The improvement with this index may attain goal 16 (peace and justice strong institutions) of the United Nation's sustainable development goals.
- The global terrorism index ranking for India is high. This index is evaluated based on terrorism and related activities in the country. Controlling this rank may contribute to goal 16 of UN's SDG.
- Recently, the dissemination of fake news on social media is gaining popularity. According to a global report, India is also one of the countries which worries about fake news as a weapon to destabilize the system.
- The ranking for the global cybersecurity index indicates how to secure the cyber activities of the country. The rank of this index for India is moderate and needs to be improved.

- The 4G LTE penetration coverage in India is increasing and has a good ranking among the other countries. This may encourage to increase in the use of mobile devices by people.
- The literature survey is carried out on hate content against different protected characteristics such as gender, religion, race, and disability. Apart from this, the different regulatory frameworks are also reviewed. The review revealed that there is limited research on digital hate content analysis and governance in India.
- At the time of framing the research question, a search for the information on hate content over social media limiting to the Indian geographical area is made. The search resulted in some information about the topic specified.

Addressing of the research question

The research question, how to improve digital hate content regulations in India is addressed in the study. It has been observed that gender, religious, racist, and disability hate content are the influencers of online hate content. With the emphasis on these influencers, while framing the hate content regulations, it may be possible to strengthen existing hate content regulations in India. As a remedy, the exclusive sections under IPC for gender, racist, and disability hate content regulations can be drawn like section 195 of IPC for religious hate content control.

The influencers of online hate content are in turn influenced by many attributes such as attitude, religionizing the nation, skin color, and fright of health. The contribution of these attributes to hate content generation can be minimized by educating the citizen. Countering hate content to neutralize its effect is a challenging task for policymakers. This can be addressed with the following approaches.

Do-it-yourself strategy to counter hate content: In this approach, instead of initiating the legal fight against the perpetrators the victim will initiate the actions such as urging the perpetrators' parents to act against their wards.

Speech versus speech: More speeches can be delivered by popular persons to nullify the effect of hate content/speech (Citron and Norton, 2011). Some of the hate content neutralizing expressions observed from the case study are “If you hate a person, then you are defeated by them”, “Smile more than you cry, give more than you take, and

love more than you hate” and “Stop being racist, your words spark racism which will create a fuss amongst others”.

Comments favoring victims’ community: A discussion group can be formed to comment in favor of the victim’s group. This discussion may stop the further expression of hate against the minorities (Miškolci et al., 2020).

7.3.2 Research Objectives Revisited

The first objective of the research is to study the nature and forms of digital hate content in the Indian context. This objective is attained by studying the ecosystem of hate content. From the literature review and quantitative analysis, different information such as generation, propagation, and imparting of hate content have been identified. Online hate content will be generated more by the adults who lie in the age span of 18-24 years old.

The different attributes such as the experience of hatred, having an attitude, lack of religious knowledge, a stance of superiority, ethnic unfamiliarity, the fright of health, ignorance, extremist behavior, communal nature, watching stereotypes, differentiating skin color, and observing terrorist incidents are observed as causes for hate content generation. On the other hand, hate content may be generating due to the hatred behavior towards gender, religion, race, and disability.

Social networking sites are a common platform through which the diffusion of online hate content takes place. During the data collection, it has been observed that late evening more people will be online and busy with social networking sites. As more people will be online, this may be the time of the day users tries to spread hate content. The online hate content can be generated through posting/tweeting, sharing/retweeting, and liking the information on social media sites.

The second objective is to understand the influence of digital hate content on online safety. To understand the influence of hate content on online safety, digital hate content is studied in different environments and layers. Different case studies (environments) such as a case study on gender hate content, religious hate content, racism, and hate content backed by triggering incidents are conducted.

Initially, the gender-based hatred analysis is made with different emotions by considering Twitter content on different topics. This study revealed that online users may get exposed to hatred against gender identity or incidents of hatred. Secondly, the religious hatred analysis is made considering the content of research and news articles along with the Twitter content. With this study, it has been understood that while surfing online, an individual may be exposed to religious hate incidents. Third, for racism, the analysis is made for basic sentiments using Twitter content. This study has shown that for an individual, during online presence, there are possibilities of exposure to racist hate content or the discussions on that. Fourth, the analysis is made for different emotions with the Twitter content on hatred triggering events. This study has shown that after a problematic incident the online users will be exposed to hatred against the incident and the communities to which the persons behind the incident belong. From the case studies, it has been observed that online users are not free from exposure to hate content or related incidents.

The third objective is to study policies and regulations addressing digital hate content. To attain this objective different international and national hate content regulation frameworks are studied. Different frameworks are compared in terms of regulations for free and hate speeches. The views of respondents about online hate content control are collected and analyzed for policy perspectives. The information on regulatory frameworks and analysis of user responses are useful in recommending hate content regulations.

The fourth objective is to bring out recommendations for improving hate content regulations in India. As the hate content is subjective, sometimes, there will be a dilemma in differentiating hate and non-hate content. For effective implementation of policies, first, the boundary between hate and non-hate content must be cleared. In other words, the hate content has to be defined clearly. As the nature of online hate content is not constant, its control requires a collective effort from global governments, Intermediaries, and citizens. Based on studies carried out, the following recommendations have been made.

- The punishment is necessary to mitigate online hate content. The perpetrator of the hate content should be punished heavily so that the person may think more before the next act related to online hate content.
- Section 153(A) of IPC punishes hatred expression against any identifiable group by imprisonment of a maximum of 3 years or fine or both. With this punishing policy, politically and economically influencing people may get exempted from imprisonment and may enjoy freedom of expression by paying only a fine. Therefore, to have uniform punishment for every citizen who expresses hate content and to reduce similar instances by others in the future, section 153(A) of IPC can be altered to have only imprisonment option.
- Like section 153(A), section 295(A) of IPC can be modified with only imprisonment option to maintain uniform punishment and reduce future occurrences of hate content.
- If a division of citizens is made based on gender, it results in larger groups than any other identity-based divisions such as religion, race, and disability. Therefore, a separate section under IPC to punish the perpetrator of hate content against gender can be included.
- Online social networks and the Internet, should be given some freedom for the moderation of content flowing through their networks. The activities of these organizations can be monitored by some concerned Government authorities to ensure the proper use of rights given to them.
- Educate the citizen over different religions, genders, cultures, traditions, ethnicities, languages, age groups, and abilities through media, campaigns, and specific programs.
- Make children citizens of the world by educating them on the equality among the different identities at schools.
- As the hate content based on different identities influences the generation of online hate content, the effective combating of online hate content can be done at the individual identity level.

- As the non-governmental organizations such as civil societies, private sectors, and intermediaries are important to reduce digital hate content, they should be given the freedom to mitigate through educating the people and blocking the information

7.4 Key Findings from the Synthesis of the Studies

The research involves multiple studies with different data sets. The mixed-method approaches such as qualitative and quantitative are applied for the analysis of data. The study highlights the following outcome as key findings.

- The laws are necessary to combat the expression of hate content against the marginalized communities but the existing laws for the purpose are less effective.
- The laws to combat hate content across the globe are varying from nation to nation. The varying nature of laws causes difficulties in combating hate content globally.
- The transnational nature of online hate content causes implementation problems for laws to combat it.
- There is a necessity for specific sections under IPC to punish the generators of gender, racist and disability hate content like religious hate content.
- In the Indian context, there is limited research on the nature of online hate content and its control.
- The hate content expressed over the Internet based on gender identity positively influences the generation of online hate content.
- The hate content expressed over the Internet based on religious identity positively influences the generation of online hate content.
- The hate content expressed over the Internet based on race/ethnic identity positively influences the generation of online hate content.
- The hate content expressed over the Internet based on disability identity positively influences the generation of online hate content.
- Non-governmental organizations such as civil societies, private sectors, and intermediaries are important to reduce digital hate content than governmental organizations.

- The victims of hatred feel helpless and bad about the offenders and the system in which they hated. This personal experience of hatred influences them to exhibit hatred about the offenders in the future.
- Often, the people feel proud about their religion and perceive that religion as superior to others. This stance of superiority positively influences online hate content.
- Skin color is one of the constituent components for the physical appearance of a person and positively influences online hate content.
- The lack of knowledge about the transmission of diseases creates fright of health and this fear positively influences online hate content.

7.5 Significant Research Contribution

Based on the key findings presented in the previous section, this section discusses the contribution of the research. The significant contributions of the study are discussed in different domains such as theoretical, practical, and methodological.

Theoretical Contribution

The theoretical contributions of the research are discussed as follows.

- Initially, the theoretical framework is designed and discussed concerning existing literature and theories. Based on the analysis and evaluation of the model with multiple methods and datasets, the model is reduced to incorporate only prominent components of the framework. The reduced framework model can be a basis for researchers in the field of digital content regulations.
- After analysis of the data, different influencing components of online hate content are identified. The identified influencing components such as individualistic expressions (gender-based hate), community-related expressions (religious hate), ethnic expressions (racist hate), and the health-related expressions (disability hate) contribute as basic components to broaden the existing literature and theory on online hate content. Specifically, these drivers of online hate content reveal the attitude of an individual could be explained with social structure-social learning theory (Akers, 2009).

- The synthesis of the study revealed that personal experience (being a victim of hatred in the past) is a positively influencing factor to gender-based hate content. This strengthens the existing literature on gender-based hate content influencers.
- The General Theory of Crime (GTC) depicts that the lack of self-control causes the production of online hate content (Gottfredson and Hirschi, 1990). This theory also explains extremist acts in society. In the present study, the stance of superiority about own religion is observed as a positive influencing factor to religious hate content. The superior feeling about the religion is extremist and contributes to the theoretical knowledge domain of hate content.
- The skin color component strengthens the existing literature of racist content influencers as it positively influences racism. The researchers who are working on racism analysis can emphasize skin color during their analysis.
- For disability hate, the fright of health is observed as an influencing component. The future studies, while analyzing disability hate content, this component can be considered as impactful.

Practical Contribution

- The study identified the presence of limited sections for controlling hate content against different identities under IPC. This information can be used while updating IPC concerning hate content control.
- The study has shown that non-governmental organizations such as civil societies, private sectors, and intermediaries (social media) are important to reduce digital hate content. If the policies to control online hate content are framed by emphasizing these organizations, hate content controllability could be increased.
- The study identified that strengthening legislation, increasing penalties, and campaigns by the government are associated with hate content reduction. The quality of hate content control policies can be improved by incorporating this information while framing policies.
- The identified influencing components of hate content can be used to educate people. Educating people about the influencers and related information can reduce the generation of hate content.

- The product online reviews are backed by intention to take revenge, brand love, and rewards. The reviews which are motivated by revenge and rewards are not actuals, they are fake reviews. Further, the escalated revenging may lead to dislike or hatred against the brands. Therefore, the control of hatred could reduce the fake reviews on brands which in turn make justice to sincere organizations.

Methodological Contribution

- The mixed-method approach improves the quality of research. The study adopted a quantitative approach followed by a qualitative approach for strengthening the outcome. The qualitative approach is applied in a layered manner. Once the social media data is analyzed for advanced sentiments involving different emotions and next the social media data is analyzed for basic sentiments such as positive, negative, and neutral. This layered approach helped in identifying different influencing components of online hate content. Therefore, the analysis approach adopted in the study could be followed by researchers in the future to arrive at an appropriate solution.
- The data for quantitative analysis is collected through both online and offline modes. While collecting data in online mode, the emphasis is given to time. During the initial days of data collection, it has been observed that most of the people will be online in the late evening. Therefore, during subsequent days, the people are contacted during the late evening through social media platforms for data collection. This resulted in increased responses from online users. This method of data collection can be adopted by future researchers to increase the quantity and quality of responses as people will be free from their routine works in the evening.
- The secondary data for the study is collected through social media platforms and online news sites. The data collection through social media platforms is made immediately after or during the trigger event. This time of data collection resulted in more impactful tweets. For future researches also this approach may be appropriate and impactful.

7.6 Limitations and Scope of Future Works

In the previous sections, the contributions and the implications of the research are discussed. This section highlights the following limitations of the study.

- The hatred can be expressed based on different identities of persons. The current study considered the analysis of hate content based on gender, religion, race, and disability identities only. Based on the analysis, the study identified that individualistic, community-related, ethnic, and health-related expressions are the influencers of online hate content. There is a possibility of the existence of other influencers to online hate content.
- Social media platforms are the common sources of online data. Though there exist several social media platforms, only Twitter is used as a source of data collection for case studies. The generalization of the outcome would have been increased by collecting data from other sources also.
- The limitation on obtaining old relevant tweets forces the study to use a moderate amount of Twitter content for analysis of hate content. Hence, obtained results might not be more accurate to resemble the real scenario of hate content.
- For the existence of online hate content, there could be technical and non-technical causes. The current study emphasized more on non-technical influencers rather than technical influencers. So, the security issues involving technical skills are less addressed.
- For the analysis of online hate content through social media platforms, only one or at most two trigger events are considered. The incorporation of more trigger events could increase the generalizability of the research outcome.
- The survey respondents from some of the states in India are more. The data analysis and findings lacks the reason behind it.

In the future, the study can be enhanced to overcome the above limitations through the following suggestions.

- By limiting hate content analysis based on gender, religion, race, and disability, the study identified different influencers of online hate content. In the future, the study can be extended to incorporate other protected characteristics such as age, sexual orientation, language, and national origin. This may result in an increased number of influencers of online hate content.
- Other than Twitter, there are several social media platforms. The use of social media varies from person to person. Some may use Twitter, some other may use Facebook,

and so on. To gather the views of persons on hate content with different backgrounds, the data collection and analysis can be performed with other social media platforms such as Facebook, Linked In, etc.

- With free accounts, Twitter provides only the limited percentage of total tweets available at a given time on a particular topic. The collection of tweets can be increased by obtaining paid membership with social media platforms. Future studies can consider this facility to increase the generalization of the outcome.
- In the future, the studies can analyze different technical issues involving hardware and software components associated with online hate content.
- More trigger events can be considered for data collection and analysis to increase the generalizability of the results.
- An in-depth analysis of the data could be made to identify the reason behind the increased responses from some of the states in India.

7.7 Concluding Remarks

The study tried to answer the question, how to improve digital hate content regulations in India? With the reviews from hate content regulations, it has been observed that the regulations on hate content expression in India still require some sections under IPC. The identified influencers of online hate content are important to reduce and control online hate content. The amount of gender, religious, racist, and disability hate content can be reduced by controlling their contributors. The control of emotions such as anger, disgust, fear, and sadness may result in reduced hate content.

REFERENCES

- Acuna, R. F. (2005). The Illusive Race Question & Class: A Bacteria That Constantly Mutates. *JSRI Occasional Paper #59*, The Julian Samora Research Institute, Michigan State University, East Lansing, Michigan, <https://jsri.msu.edu/upload/occasional-papers/oc59.pdf>.
- Agur, C., Subramanian, R., & Belair-Gagnon, V. (2015). Interactions and policy-making: Civil society perspectives on the multi-stakeholder internet governance process in India. *Internet Policy Observatory Center for Global Communication Studies Annenberg School for Communication University of Pennsylvania*. 1-32.
- Akers, R. (2009). Social learning and social structure. New York: Routledge.
- 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.
- Alathur, S., Ilavarasan, P. V., & Gupta, M. P. (2016). Determinants of e-participation in the citizens and the government initiatives: Insights from India. *Socio-Economic Planning Sciences*, 55, 25-35.
- Alhaboby, Z. A., al-Khateeb, H. M., Barnes, J., & Short, E. (2016). 'The language is disgusting and they refer to my disability': the cyberharassment of disabled people. *Disability & Society*, 31(8), 1138-1143.
- Alkiviadou, N. (2019). Hate speech on social media networks: towards a regulatory framework? *Information & Communications Technology Law*, 28(1), 19-35.
- Allegra, E., Di Pietro, R., La Noce, M., Ruocco, V., & Verde, N. V. (2011). Cross-border co-operation and education in digital investigations: A European perspective. *digital investigation*, 8(2), 106-113.
- Almosaed, N. (2004). Violence against women: a cross-cultural perspective. *Journal of Muslim Minority Affairs*, 24(1), 67-88.
- Amedie, J. (2015). The Impact of Social Media on Society. *Advanced Writing: Pop Culture Intersections*. 2
- Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual review of psychology*, 53(1), 27-51.

- Arun, C.. & Nayak, N. (2016). Preliminary Findings on Online Hate Speech and the Law in India. *Berkman Klein Center Research Publication*, (2016-19).
- Asaduzzaman, M., & Virtanen, P. (2016). Governance theories and models. *Global Encyclopedia of Public Administration, Public Policy, and Governance*. New York, USA: Springer.
- Askins, K. (2009). 'That's just what I do': Placing emotion in academic activism. *Emotion, Space and Society*, 2(1), 4-13.
- Asquith, N. (2015). Honour, violence and heteronormativity. *International Journal for Crime, Justice and Social Democracy*, 4(3), 73-84.
- Athar, A., Khan, M. S., Ahmed, K., Ahmed, A., & Anwar, N. (2011). A fuzzy inference system for synergy estimation of simultaneous emotion dynamics in agents. *Int. J. Sci. Eng. Res*, 2(6), 35-41.
- Awan, I. (2016). Islamophobia on social media: A qualitative analysis of the Facebook's Walls of Hate. *International Journal of Cyber Criminology*, 10(1), 1-20.
- Awan, I. (2017). Cyber-extremism: Isis and the power of social media. *Society*, 54(2), 138-149.
- Awan, I., & Blakemore, B. (2012). *Policing cyber hate, cyber threats, and cyberterrorism*. Ashgate: Farnham.
- Awan, I., & Zempi, I. (2016). The affinity between online and offline anti-Muslim hate crime: Dynamics and impacts. *Aggression and violent behavior*, 27, 1-8.
- Baer, S. (2009) 'Dignity, liberty, equality: a fundamental rights triangle of constitutionalism', *University of Toronto Law Journal*, Vol. 59, No. 4, pp. 417-468.
- Baines, P. R., & Shaughnessy, N. J. O. (2014). Al-Qaeda messaging evolution and positioning 1998–2008: Propaganda analysis revisited. *Public Relations Inquiry*, 3(2), 163-191.
- Baird, Z. (2002). Governing the Internet-Engaging Government, Business, and Nonprofits. *Foreign Affairs*. 81(6), 15-20.

- Banerjee, M., Harrell, Z. A., & Johnson, D. J. (2011). Racial/ethnic socialization and parental involvement in education as predictors of cognitive ability and achievement in African American children. *Journal of youth and adolescence*, 40(5), 595-605.
- Banton, M. (1992). The nature and causes of racism and racial discrimination. *International Sociology*, 7(1), 69-84.
- Barlow, C., & Awan, I. (2016). “You need to be sorted out with a knife”: the attempted online silencing of women and people of Muslim faith within academia. *Social Media+ Society*, 2(4), 1-11.
- Bar-Tal, D. (2007). Socio psychological foundations of intractable conflicts. *American Behavioral Scientist*, 50(11), 1430-1453.
- Bastiaensens, S., Vandebosch, H., Poels, K., Van Cleemput, K., Desmet, A., & De Bourdeaudhuij, I. (2014). Cyberbullying on social network sites. An experimental study into bystanders’ behavioural intentions to help the victim or reinforce the bully. *Computers in Human Behavior*, 31, 259-271.
- Bates, C. (1995). *Race, Caste and Tribe in Central India: the early origins of Indian anthropometry*. Edinburgh: University of Edinburgh, Centre for South Asian Studies.
- Beckman, L., Hagquist, C., & Hellström, L. (2013). Discrepant gender patterns for cyberbullying and traditional bullying—An analysis of Swedish adolescent data. *Computers in Human Behavior*, 29(5), 1896-1903.
- Benesch, S. (2014). Defining and diminishing hate speech. *Freedom from hate: State of the world’s minorities and indigenous peoples 2014*, 18-25.
- Berry, N., Lobban, F., Belousov, M., Emsley, R., Nenadic, G., & Bucci, S. (2017). #WhyWeTweetMH: understanding why people use Twitter to discuss mental health problems. *Journal of medical Internet research*, 19(4), e107.
- Bertram, L. (2016). Terrorism, the Internet and the Social Networks Advantage. *Journal for Deradicalization*, 7, 225-252.

- Best, M. L. (2004). Can the internet be a human right?. *Human Rights & Human Welfare*, 4(1), 13.
- Bhandari, M. K., & Bhatt, M. N. (2012). Hate Speech and Freedom of Expression: Balancing Social Good and Individual Liberty. *The Practical Lawyer*, January S-5.
- Bhuiyan, S. I. (2011). Social media and its effectiveness in the political reform movement in Egypt. *Middle East Media Educator*, 1(1), 14-20.
- Bianchi, C. (2014). Slurs and appropriation: An echoic account. *Journal of Pragmatics*, 66, 35-44.
- Bieda, D. (2015). Cyberspace: A Venue For Terrorism. *Issues in Information Systems*, 16(3), 33-42.
- Blaya, C. (2019). Cyberhate: A review and content analysis of intervention strategies. *Aggression and violent behavior*. 45(2), 163-172.
- Bloom, S. S. (2008). *Violence against women and girls: A compendium of monitoring and evaluation indicators*. MEASURE Evaluation, Carolina Population Center.
- Bonilla-Silva, E. (1997). Rethinking racism: Toward a structural interpretation. *American sociological review*, 465-480.
- Boyd, D., Ryan, J., & Leavitt, A. (2011). Pro-self-harm and the visibility of youth-generated problematic content. *ISJLP*, 7, 1.
- Braithwaite, J., Coglianese, C., & Levi-Faur, D. (2007). Can regulation and governance make a difference?. *Regul. Gov.*, 1(1), 1-7.
- Breen, D., & Nel, J. A. (2011). South Africa-A home for all: The need for hate crime legislation. *South African Crime Quarterly*, 38, 33-43.
- Brewer, G., Roy, M., & Smith, Y. (2010). Domestic violence: The psychosocial impact and perceived health problems. *Journal of Aggression, Conflict and Peace Research*. 2(2), 4-15.

- Brown, R. P., Baughman, K., & Carvallo, M. (2018). Culture, masculine honor, and violence toward women. *Personality and social psychology bulletin*, 44(4), 538-549.
- Brudholm, T. (2010). Hatred as an attitude. *Philosophical Papers*, 39(3), 289-313.
- Burke, M. & Kraut, R.E. (2016). The relationship between Facebook use and well-being depends on communication type and tie strength. *Journal of Computer-Mediated Communication*, 21(4), 265-281.
- Burnap, P., Williams, M. L., Sloan, L., Rana, O., Housley, W., Edwards, A., Knight, V., Procter, R., & Voss, A. (2014). Tweeting the terror: modeling the social network's reaction to the Woolwich terrorist attack. *Social Network Analysis and Mining*, 4(1), 206.
- 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.
- Campbell, J. (1981). Misogyny and homicide of women. *Advances in Nursing Science*, 3(2), 67-86.
- Caplan, R., Hanson, L., & Donovan, J. (2018). Dead reckoning navigating content moderation after “Fake News”. *Data & Society*, <https://apo.org.au/node/134521>.
- Carroll, W. K., & Ratner, R. S. (1995). Old unions and new social movements. *Labour/Le Travail*, 195-221.
- Caruso, R., & Schneider, F. (2013). The brutality of Jihadist terrorism. A contest theory perspective and empirical evidence in the period 2002–2010. *Journal of Policy Modeling*, 35, 685–696.
- Casas, J. A., Del Rey, R., & Ortega-Ruiz, R. (2013). Bullying and cyberbullying: Convergent and divergent predictor variables. *Computers in Human Behavior*, 29(3), 580-587.
- Chatfield, A. T., Reddick, C. G., & Brajawidagda, U. (2015, May). Tweeting Propaganda, Radicalization, and Recruitment: Islamic State Supporters Multi-

- Sided Twitter Networks. Proceedings of the 16th Annual International Conference on Digital Government Research, Phoenix, Arizona, USA, 239-249, ACM.
- Catholic Charities. (2008). Poverty and Racism: Overlapping Threats to the Common Good. *ORIGINS-WASHINGTON-*, 37(34), 537.
- Chatterjee, D. (2016). Gendering ISIS and Mapping the Role of Women. *Contemporary Review of the Middle East*, 3(2), 201–218.
- Chaudhry, I. (2016). “Not So Black and White” Discussions of Race on Twitter in the Aftermath of # Ferguson and the Shooting Death of Mike Brown. *Cultural Studies? Critical Methodologies*, 16(3), 296-304.
- Chen, Y., Zhou, Y., Zhu, S., & Xu, H. (2012, September). Detecting offensive language in social media to protect adolescent online safety. In *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on Social Computing (SocialCom)* (pp. 71-80). IEEE.
- Cheung, C. M., Chiu, P. Y., & Lee, M. K. (2011). Online social networks: Why do students use facebook?. *Computers in human behavior*, 27(4), 1337-1343.
- Cholin, V. S., Vijayakumar, J. K., & Murthy, T. A. V. (2003). Resource Sharing in the Digital Environment in India: Role of INFLIBNET. *Library Hi Tech News*, 20(3).
- Chua, V., Mathews, M., & Loh, Y. C. (2016). Social capital in Singapore: Gender differences, ethnic hierarchies, and their intersection. *Social Networks*, 47, 138-150.
- Citron, D. K., & Norton, H. (2011). Intermediaries and hate speech: Fostering digital citizenship for our information age. *BUL Rev.*, 91, 1435.
- Cockbain, E., & Bowers, K. (2019). Human trafficking for sex, labour and domestic servitude: how do key trafficking types compare and what are their predictors?. *Crime, Law and Social Change*, 72(1), 9-34.
- Cohen-Almagor, R. C. (2011). Fighting Hate and Bigotry on the Internet. *Policy & Internet*, 3(3), 1-28.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Lawrence Erlbaum Associates, 2nd edition.

- Cohen, R. (2014). Regulating Hate Speech: Nothing Customary About It. *Chi. J. Int'l L.*, 15, 229.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Academic Press. New York.
- Conway, M. (2007). Terrorism and internet governance: Core issues. In *Disarmament Forum* (Vol. 2007, No. 3, pp. 23-34). United Nations.
- Corazza, M., Menini, S., Cabrio, E., Tonelli, S., & Villata, S. (2020). A multilingual evaluation for online hate speech detection. *ACM Transactions on Internet Technology (TOIT)*, 20(2), 1-22.
- Corcoran, K. E. (2015). Thinkers and feelers: Emotion and giving. *Social Science Research*, 52, 686-700.
- Correa, T., Hinsley, A. W., & De Zuniga, H. G. (2010). Who interacts on the Web?: The intersection of users' personality and social media use. *Computers in human behavior*, 26(2), 247-253.
- Cossman, B., & Kapur, R. (1997). Secularism's Last Sigh: The Hindu Right, the Courts, and India's Struggle for Democracy. *Harv. Int'l. LJ*, 38(1), 113-170.
- Costello, M., Hawdon, J. (2020). Hate Speech in Online Spaces. In: Holt T., Bossler A. (eds) *The Palgrave Handbook of International Cybercrime and Cyberdeviance*. Palgrave Macmillan, Cham. https://doi.org/10.1007/978-3-319-78440-3_60.
- Council of Europe. (1997). <https://rm.coe.int/CoERMPublicCommonSearchServices/DisplayDCTMContent?documentId=0900001680505d5b>, Accessed 16 February 2017.
- Crawford, J. (1998). Media, stereotypes and the perpetuation of racism in Canada. <https://etad.usask.ca/802papers/crawford/crawford.PDF>.
- Criminal Justice and Public Order Act (1994). https://www.legislation.gov.uk/ukpga/1994/33/pdfs/ukpga_19940033_en.pdf, Accessed date: 17 April 2018.

- Culnan, M. J., McHugh, P. J., & Zubillaga, J. I. (2010). How large US companies can use Twitter and other social media to gain business value. *MIS Quarterly Executive*, 9(4).
- d'Anjou, L., & Van Male, J. (1998). Between old and new: social movements and cultural change. *Mobilization: An International Quarterly*, 3(2), 207-226.
- Deloughery, K., King, R. D., & Asal, V. (2012). Close cousins or distant relatives? The relationship between terrorism and hate crime. *Crime & Delinquency*, 58(5), 663-688.
- Denning, D. E. (2000). Cyberterrorism: The logic bomb versus the truck bomb. *Global Dialogue*, 2(4), 29-37.
- Deol, S. S. (2014). Honour killings in India: A study of the Punjab State. *International Research Journal of Social Sciences*, 3(6), 7-16.
- Derman-Sparks, L., Higa, C. T., & Sparks, B. (1980). Children, race and racism: How race awareness develops. *Interracial Books for Children Bulletin*, 11(3/4), 3-15.
- Dharmapala, D., & Garoupa, N. (2004). Penalty enhancement for hate crimes: An economic analysis. *American Law and Economics Review*, 6(1), 185-207.
- Dhavan, R. (2007). Harassing Husain: Uses and Abuses of the Law of Hate Speech. *Social Scientist*, 35(1/2), 16-60.
- Djuric, N., Zhou, J., Morris, R., Grbovic, M., Radosavljevic, V., & Bhamidipati, N. (2015, May). Hate speech detection with comment embeddings. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 29-30). ACM.
- DNA. (2017). Daily News and Analysis, Honour Killing. <http://www.dnaindia.com/topic/honour-killing>, accessed 20 September 2017.
- Dunn, L. C. (1975). *Race, science, and society*. Columbia University Press.
- Dwivedi, S. (2017). Sexual Objectification of Females in Bollywood Rap and Item Songs. *International Journal of Humanities & Social Science Studies*, 3(6), 399-406.

- ECHR-European Convention on Human Rights. (1950). http://www.echr.coe.int/Documents/Convention_ENG.pdf, Accessed 06 February 2017.
- Edwards, S. S. (2017). Cyber-grooming young women for terrorist activity: Dominant and subjugated explanatory narratives. *Cybercrime, organized crime, and societal responses* (pp. 23–46). Cham: Springer.
- Eisler, R. (2015). Can international law protect half of humanity? A new strategy to stop violence against women. *Journal of Aggression, Conflict and Peace Research*, 7(2), 88-100.
- Esmail, S., Darry, K., Walter, A., & Knupp, H. (2010). Attitudes and perceptions towards disability and sexuality. *Disability and Rehabilitation*, 32(14), 1148-1155.
- Farkas, J., & Neumayer, C. (2017). ‘Stop Fake Hate Profiles on Facebook’: Challenges for crowdsourced activism on social media. *First Monday*, 22(9).
- Faqir, F. (2001). Intrafamily femicide in defence of honour: the case of Jordan. *Third world quarterly*, 22(1), 65-82.
- Ford, J., & Blumenstein, L. (2013). Self-control and substance use among college students. *Journal of Drug Issues*, 43, 56–68.
- Fu, M. H., Peng, C. H., Kuo, Y. H., & Lee, K. R. (2012, June). Hidden community detection based on microblog by opinion-consistent analysis. In *Information Society (i-Society), 2012 International Conference on* (pp. 83-88). IEEE.
- Gagliardone, I., Gal, D., Alves, T., & Martinez, G. (2015). *Countering online hate speech*. Unesco Publishing. Paris.
- George, C. (2016). Regulating" Hate Spin": The Limits of Law in Managing Religious Incitement and Offense. *International Journal of Communication*, 10(18), 2955-2972.
- Gill, A. (2006). Patriarchal Violence in the Name of Honour. *International Journal of Criminal Justice Sciences*, 1(1).

- Gill, A. K., Begikhani, N., & Hague, G. (2012, March). Honour-based violence in Kurdish communities. In *Women's Studies International Forum*, 35(2), 75-85, Pergamon.
- Gill, A.K. & Brah, A. (2014). Interrogating cultural narratives about honour-based violence. *European Journal of Women's Studies*, 21(1), 72-86.
- Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.
- Gitari, N. D., Zuping, Z., Damien, H., & Long, J. (2015). A lexicon-based approach for hate speech detection. *International Journal of Multimedia and Ubiquitous Engineering*, 10(4), 215-230.
- Glenn, C. L. (2015). Activism or “slacktivism?”: Digital media and organizing for social change. *Communication Teacher*, 29(2), 81-85.
- Goli, S., Singh, D., & Sekher, T. V. (2013). Exploring the myth of mixed marriages in India: Evidence from a nation-wide survey. *Journal of comparative family studies*, 44(2), 193-206.
- Goolsby, R., Shanley, L., & Lovell, A. (2013). *On cybersecurity, crowdsourcing, and social cyber-attack*. Office of Naval Research, Arlington, VA.
- Goswami, R., & Kumari, M. (2013). Impact of Indian Cinema on adolescents: a sociological study of Jaipur-Rajasthan, India. *International Research Journal of Social Sciences*, 2(7), 19-32.
- Gottfredson, M. R., & Hirschi, T. (1990). *A general theory of crime*. Stanford, CA: Stanford University Press.
- Gross, E. (2001). Legal Aspects of Tackling Terrorism: The Balance Between the Right of a Democracy to Defend Itself and the Protection of Human Rights. *UCLA J. Int'l L. & Foreign Aff.*, 6, 89.
- Guiora, A., & Park, E. A. (2017). Hate speech on social media. *Philosophia*, 45(3), 957-971.
- Gupta, C. (2009). Hindu women, Muslim men: Love Jihad and conversions. *Economic and Political Weekly*, 44(51), 13-15.

- Gupta, A., Lamba, H., Kumaraguru, P., & Joshi, A. (2013, May). Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web* (pp. 729-736). ACM, Rio de Janeiro, Brazil.
- Haciyakupoglu, G., Hui, J. Y., Suguna, V. S., Leong, D., & Rahman, M. F. B. A. (2018). Countering Fake News: A Survey of Recent Global Initiatives. Policy report, Nanyang Technological University.
- Hollomotz, A. (2013). Disability, oppression and violence: Towards a sociological explanation. *Sociology*, *47*(3), 477-493.
- Halperin, E., Russell, A. G., Dweck, C. S., & Gross, J. J. (2011). Anger, hatred, and the quest for peace: Anger can be constructive in the absence of hatred. *Journal of Conflict Resolution*, *55*(2), 274-291.
- Halperin, E., & Gross, J. J. (2011). Intergroup anger in intractable conflict: Long-term sentiments predict anger responses during the Gaza war. *Group Processes & Intergroup Relations*, *14*(4), 477-488.
- Hamlin, J. E. (1988). The misplaced role of rational choice in neutralization theory. *Criminology*, *26*(3), 425-438.
- Hanes, E., & Machin, S. (2014). Hate crime in the wake of terror attacks: Evidence from 7/7 and 9/11. *Journal of Contemporary Criminal Justice*, *30*(3), 247-267.
- Hannon, F. (2007). *Literature review on attitudes towards disability*. National Disability Authority. https://www.ucd.ie/t4cms/0048-01%20NDA_public_attitudes_disability_2006_literature_review.pdf.
- Hardaker, C., & McGlashan, M. (2015). “Real men don’t hate women”: Twitter rape threats and group identity. *Journal of Pragmatics*, *91*, 80-93.
- Hargrave, A. M., Livingstone, S., & Livingstone, S. M. (2009). *Harm and offence in media content: A review of the evidence*. 2nd Revised edition, Intellect, Bristol, UK.

- Harlow, S. (2012). Social media and social movements: Facebook and an online Guatemalan justice movement that moved offline. *New Media & Society, 14*(2), 225-243.
- Harris, S. T., Oakley, C., & Picchioni, M. (2013). Quantifying violence in mental health research. *Aggression and Violent Behavior, 18*(6), 695-701.
- Hartnack, C. (2012). Roots and Routes: The Partition of British India in Indian Social Memories. *Journal of Historical Sociology, 25*(2), 244-260.
- Haythornthwaite, C. (2005). Social networks and Internet connectivity effects. *Information, Community & Society, 8*(2), 125-147.
- Hayes, B. E., Freilich, J. D., & Chermak, S. M. (2016). An exploratory study of honor crimes in the United States. *Journal of Family Violence, 31*(3), 303-314.
- Hayes, B. E., Mills, C. E., Freilich, J. D., & Chermak, S. M. (2018). Are honor killings unique? A comparison of honor killings, domestic violence homicides, and hate homicides by far-right extremists. *Homicide studies, 22*(1), 70-93.
- Heady, F. (2001). *Public Administration, A Comparative Perspective*. CRC Press.
- Hedges, L. V., & Cooper, H. (2009). Research synthesis as a scientific process. *The handbook of research synthesis and meta-analysis, 1*.
- Heerdink, M. W., Koning, L. F., van Doorn, E. A., & Van Kleef, G. A. (2019). Emotions as guardians of group norms: Expressions of anger and disgust drive inferences about autonomy and purity violations. *Cognition and emotion, 33*(3), 563-578.
- Heyman, S. J. (2008). Hate speech, public discourse, and the First Amendment. in I. Hare & J. Weinstein, eds., *Extreme Speech and Democracy*. New York: Oxford Univ. Press. 158-181.
- Henry, J. S. (2009). Beyond free speech: novel approaches to hate on the Internet in the United States. *Information & Communications Technology Law, 18*(2), 235-251.
- Holbrook, D., & Taylor, M. (2019). Terrorism as process narratives: a study of pre-arrest media usage and the emergence of pathways to engagement. *Terrorism and political violence, 31*(6), 1307-1326.

- Hosseini, A. S., and Zohdi, E. (2016). The kite runner and the problem of racism and ethnicity. *International letters of social and humanistic sciences*, 74, 33-40.
- Hurmerinta-Peltomäki, L., & Nummela, N. (2006). Mixed methods in international business research: A value-added perspective. *Management International Review*, 46(4), 439-459.
- Hutcherson, C. A., & Gross, J. J. (2011). The moral emotions: A social–functionalist account of anger, disgust, and contempt. *Journal of personality and social psychology*, 100(4), 719.
- IAMAI. (2018). Internet users in India expected to reach 500 million by June. *The Economic Times*, Feb. 20, 2018.
- ICA. (2017). Assessing Russian activities and intentions in recent US elections, office of the director of national intelligence, national intelligence council, January 6, 2017, https://www.dni.gov/files/documents/ICA_2017_01.pdf.
- ICCPR-International Covenant on Civil and Political Rights (1976). <http://www.ohchr.org/Documents/ProfessionalInterest/ccpr.pdf>, Accessed date: 6 February 2017.
- ICERD-International Convention on the Elimination of All Forms of Racial Discrimination. (1969). <http://www.ohchr.org/Documents/ProfessionalInterest/cerd.pdf> Accessed 08 February 2017.
- Ichheiser, G. (1943). Structure and dynamics of interpersonal relations. *American sociological review*, 302-305.
- Indian Penal Code. (1860). <http://www.vakilno1.com/bareacts/IndianPenalCode> Accessed 17 April 2018.
- Iosifidis, P. (2002). Digital convergence: Challenges for European regulation. *Javnost-The public*, 9(3), 27-47.
- Ishengoma, F. R. (2013). Online Social Networks and Terrorism 2.0 in Developing Countries. *International Journal of Computer Science and Network Solutions*, 1(4), 1-12.

- Ismail, M. M. B., & Bchir, O. (2015). Insult detection in social network comments using possibilistic based fusion approach. In *Computer and Information Science* (pp. 15-25). Springer International Publishing.
- Jain, A., Lata, P., Goyal, A. R., Khandelwal, S., & Jain, G. (2015). Socio-cultural impact of film celebrities on teenagers: an empirical study. *International Journal of Indian Culture and Business Management*, 11(3), 308-322.
- Jane, E. A. (2016). Online misogyny and feminist digilantism. *Continuum*, 30(3), 284-297.
- Jang, S. M., & Kim, J. K. (2018). Third person effects of fake news: Fake news regulation and media literacy interventions. *Computers in Human Behavior*, 80, 295-302.
- Jay, T. (2009). Do offensive words harm people?. *Psychology, public policy, and law*, 15(2), 81.
- Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing. *Journal of Politeness Research. Language, Behaviour, Culture*, 4(2), 267-288.
- Joffres, C., Mills, E., Joffres, M., Khanna, T., Walia, H., & Grund, D. (2008). Sexual slavery without borders: trafficking for commercial sexual exploitation in India. *International Journal for Equity in Health*, 7(1), 1-11.
- Johnson, K. R., & Ingram, J. E. C. (2012). Anatomy of a modern-day lynching: The relationship between hate crimes against Latina/os and the debate over immigration reform. *NCL Rev.*, 91, 1613.
- Johnson, N. F., Leahy, R., Restrepo, N. J., Velasquez, N., Zheng, M., Manrique, P., ... & Wuchty, S. (2019). Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature*, 573(7773), 261-265.
- Jones, T. (2013). The significance of skin color in Asian and Asian-American communities: Initial reflections. *UC Irvine L. Rev.*, 3, 1105.
- Jubany, O., & Roiha, M. (2016). Backgrounds, experiences and responses to online hate speech: a comparative cross-country analysis. *Barcelona: Universitat de*

Barcelona,<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.734.2403&rep=rep1&type=pdf>

- Jyoti, D. (2017). The seed of racism that fuelled Greater Noida attacks has roots in Bollywood. *Opinion, Hindustan Times*, New Delhi.
- Kant, A. (2021). Digital India - Where Knowledge Is Strength. Available at <http://niti.gov.in/digital-india-where-knowledge-strength>, accessed 5 July 2021.
- Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. *Business horizons*, 53(1), 59-68.
- Kapoor, K. K., Tamilmani, K., Rana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: past, present and future. *Information Systems Frontiers*, 20(3), 531-558.
- Kapur, R. (1996). Who Draws the Line? Feminist Reflections on Speech and Censorship. *Economic and Political Weekly*, 31(16/17), WS15-WS30.
- Karyotis, G. (2007). Securitization of Greek Terrorism and Arrest of the Revolutionary Organization November 17'. *Cooperation and Conflict*, 42(3), 271-293.
- Kasakowskij, T., Fürst, J., Fischer, J., & Fietkiewicz, K. J. (2020). Network enforcement as denunciation endorsement? A critical study on legal enforcement in social media. *Telematics and Informatics*, 46, 101317.
- Kaur, R. (2010). Khap panchayats, sex ratio and female agency. *Economic and Political Weekly*, 14-16.
- Khan, R. (2007). Honour-Related Violence (HRV) in Scotland: A cross-and multi-agency intervention involvement survey. *International Journal of Criminology*, 1-8.
- Khan, R., & Cooke, D. J. (2013). Measurement of sibling violence: A two-factor model of severity. *Criminal Justice and Behavior*, 40(1), 26-39.
- Khoo, C. S., & Johnkhan, S. B. (2018). Lexicon-based sentiment analysis: Comparative evaluation of six sentiment lexicons. *Journal of Information Science*, 44(4), 491-511.

- Kirti, A., Kumar, P., & Yadav, R. (2011). The face of honour based crimes: Global concerns and solutions. *International Journal of Criminal Justice Sciences*, 6(1/2), 343.
- Kovacs, A., Padte, R. K., & Shobha, S. V. (2013). Don't Let It Stand!": An Exploratory Study of Women and Verbal Online Abuse in India. New Delhi: Internet Democracy Project, 2013.
- Kowalski, R.M., Giumetti, G.W., Schroeder, A.N. and Reese, H.H. (2012), Chapter 14 Cyber Bullying Among College Students: Evidence from Multiple Domains of College Life. Wankel, L. A. and Wankel, C. (Ed.) *Misbehavior Online in Higher Education (Cutting-Edge Technologies in Higher Education, Vol. 5)*, Emerald Group Publishing Limited, Bingley, pp. 293-321.
- Krishnan, K. (2018). Honour crimes in India: An assault on women's autonomy. <https://www.aljazeera.com/indepth/opinion/honour-killings-india-assault-women-autonomy-180314090856246.html>, Access date: 19 April 2019.
- Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The lancet*, 360(9339), 1083-1088.
- Kudlick, C. J. (2003). Disability history: Why we need another "other". *The American Historical Review*, 108(3), 763-793.
- Kumar, A. (2012). Khap panchayats: A socio-historical overview. *Economic and Political Weekly*, 59-64.
- Kumar, D. (1982). Racial Discrimination and Science in Nineteenth Century India. *The Indian Economic & Social History Review*, 19(1), 63-82.
- Kumar, R. (1997). *The history of doing: An illustrated account of movements for women's rights and feminism in India 1800-1990*. Zubaan.
- Kwate, N. O. A., Meyer, I. H., Eniola, F., & Dennis, N. (2010). Individual and group racism and problem drinking among African American women. *Journal of Black Psychology*, 36(4), 446-457.
- Kwok, I., & Wang, Y. (2013, July). Locate the Hate: Detecting Tweets against Blacks. In *Proceedings of 27th AAAI Conference on Artificial Intelligence*, 1621-1622.

- Lankala, S. (2006). Mediated Nationalisms and 'Islamic Terror': The Articulation of Religious and Postcolonial Secular Nationalisms in India. *Westminster Papers in Communication & Culture*, 3(2), 86-102.
- Lash, C. L. (2017). Defining "American" in the context of immigration: a case study of Helping Hands Elementary. *Ethnic and Racial Studies*, 40(6), 871-890.
- Law Commission of India. (1971). <http://lawcommissionofindia.nic.in/1-50/report42.pdf> Accessed 17 April 2018.
- Lawler, E. J., Thye, S. R., & Yoon, J. (2000). Emotion and group cohesion in productive exchange. *American Journal of Sociology*, 106(3), 616-657.
- Lawler, E. J. (2001). An affect theory of social exchange. *American journal of sociology*, 107(2), 321-352.
- Lawrence, F. M. (1994). The punishment of hate: Toward a normative theory of bias-motivated crimes. *Michigan Law Review*, 93(2), 320-381.
- Levy, B. L., & Levy, D. L. (2016). When love meets hate: The relationship between state policies on gay and lesbian rights and hate crime incidence. *Social science research*, 61, 142-159.
- Liebman, C. S. (1983). Extremism as a religious norm. *Journal for the Scientific Study of Religion*, 22(1), 75-86.
- MacCarthy, M. (2019). A Consumer Protection Approach to Platform Content Moderation. *Fundamental Rights Protection Online: the Future Regulation of Intermediaries*, Edward Elgar, <http://dx.doi.org/10.2139/ssrn.3408459>.
- Macdonald, S. J., Donovan, C., & Clayton, J. (2017). The disability bias: understanding the context of hate in comparison with other minority populations. *Disability & Society*, 32(4), 483-499.
- Mahmood, S. (2013). Online social networks and terrorism: Threats and defenses. In *Security and privacy preserving in social networks* (pp. 73-94). Springer, Vienna.

- Manmathan, M. R. (2013). Temple as the Site of Struggle: Social Reform, Religious Symbols and the Politics of Nationalism in Kerala. *Advances in Historical Studies*, 2(2), 57.
- Masucci, M., & Langton, L. (2017). Hate Crime Victimization, 2004-2015. Washington, DC: US Department of Justice, Bureau of Justice Statistics. https://nccpsafety.org/assets/files/library/Hate_Crime_Victimization_2004-15.pdf.
- Mathew, B., Saha, P., Tharad, H., Rajgaria, S., Singhanian, P., Maity, S. K., ... & Mukherjee, A. (2019, July). Thou shalt not hate: Countering online hate speech. In *Proceedings of the international AAAI conference on web and social media* (Vol. 13, pp. 369-380).
- Matsuda, M. J. (2018). Public response to racist speech: Considering the victim's story. In *Words that wound*, 17-51, Routledge.
- McCarthy, J. D., and Zald, M. N. (1977). Resource mobilization and social movements: A partial theory. *American journal of sociology*, 82(6), 1212-1241.
- McCarthy, M. (2017). 'What kind of abuse is him spitting in my food?': reflections on the similarities between disability hate crime, so-called 'mate' crime and domestic violence against women with intellectual disabilities. *Disability & Society*, 32(4), 595-600.
- McDevitt, J., Levin, J., & Bennett, S. (2002). Hate crime offenders: An expanded typology. *Journal of Social Issues*, 58(2), 303-317.
- McGonagle, T. (2017). "Fake news" False fears or real concerns?. *Netherlands Quarterly of Human Rights*, 35(4), 203-209.
- McKim, C. A. (2017). The value of mixed methods research: A mixed methods study. *Journal of Mixed Methods Research*, 11(2), 202-222.
- McLaughlin, H., Uggen, C., & Blackstone, A. (2017). The economic and career effects of sexual harassment on working women. *Gender & Society*, 31(3), 333-358.
- McVeigh, R. (2004). Structured ignorance and organized racism in the United States. *Social Forces*, 82(3), 895-936.

- Meetoo, V., & Mirza, H. S. (2007, May). “There is nothing ‘honourable’ about honour killings”: Gender, violence and the limits of multiculturalism. In *Women's Studies International Forum*, 30(3), 187-200, Pergamon.
- Miaou, S. P., Lu, A., & Lum, H. S. (1996). Pitfalls of using R2 to evaluate goodness of fit of accident prediction models. *Transportation Research Record*, 1542(1), 6-13.
- Miller, J. L., & Garran, A. M. (2017). Racism in the United States: Implications for the helping professions. *Springer Publishing Company, LLC*.
- Mills, C. E., Freilich, J. D., & Chermak, S. M. (2017). Extreme hatred: Revisiting the hate crime and terrorism relationship to determine whether they are “close cousins” or “distant relatives”. *Crime & Delinquency*, 63(10), 1191-1223.
- Miro-Llinares, F., & Rodriguez-Sala, J. J. (2016). Cyber hate speech on Twitter: Analyzing disruptive events from social media to build a violent communication and hate speech taxonomy. *International Journal of Design & Nature and Ecodynamics*, 11(3), 406–415.
- Mirchandani, M. (2018). Digital Hatred, Real Violence: Majoritarian Radicalisation and Social Media in India, *Observer Research Foundation Occasional Paper # 167*.
- Miškolci, J., Kováčová, L., & Rigová, E. (2020). Countering hate speech on Facebook: The case of the Roma minority in Slovakia. *Social Science Computer Review*, 38(2), 128-146.
- Modha, S., Majumder, P., Mandl, T., & Mandalia, C. (2020). Detecting and visualizing hate speech in social media: A cyber-Watchdog for surveillance. *Expert Systems with Applications*, 161, 113725.
- Mohammad, S. M., & Turney, P. D. (2020). Nrc emotion lexicon. National Research Council, Canada. 2013; 2.
- Mohammad, S. M. (2012). From once upon a time to happily ever after: Tracking emotions in mail and books. *Decision Support Systems*, 53(4), 730-741.
- Molina-Azorín, J. F. (2011). The use and added value of mixed methods in management research. *Journal of Mixed methods research*, 5(1), 7-24.

- Mondal, M., Silva, L. A., Correa, D., & Benevenuto, F. (2018). Characterizing usage of explicit hate expressions in social media. *New Review of Hypermedia and Multimedia*, 24(2), 1-21.
- Moon, B., & Alarid, L. F. (2015). School bullying, low self-control, and opportunity. *Journal of Interpersonal Violence*, 30, 839–856.
- Morrow, D. F. (2001). Older gays and lesbians: Surviving a generation of hate and violence. *Journal of Gay & Lesbian Social Services*, 13(1-2), 151-169.
- Mossie, Z., & Wang, J. H. (2020). Vulnerable community identification using hate speech detection on social media. *Information Processing & Management*, 57(3), 102087.
- Mosteller, F., & Colditz, G. A. (1996). Understanding research synthesis (meta-analysis). *Annual review of public health*, 17(1), 1-23.
- Mulligan, J. (2018). Stop hate crime-Religion. Police and crime commissioner, North Yorkshire. Accessed from <https://www.northyorkshire-pcc.gov.uk/for-you/victims/hate-crime-report/stop-hate-crime-religion>.
- Munger, K. (2016). Tweetment effects on the tweeted: Experimentally reducing racist harassment. *Political Behavior*, 39(3), 629-649.
- Nair, N. (2013). Beyond the ‘Communal’ 1920s: The Problem of Intention, Legislative Pragmatism, and the Making of Section 295A of the Indian Penal Code. *The Indian Economic & Social History Review*, 50(3), 317-340.
- Nakagawa, K., & Arzubaga, A. E. (2014). The use of social media in teaching race. *Adult Learning*, 25(3), 103-110.
- Ngulube, P., Mathipa, E. R., & Gumbo, M. T. (2015). Theoretical and conceptual frameworks in the social and management sciences. *Addressing research challenges: making headway for developing researchers*, 43-66.
- Nobata, C., Tetreault, J., Thomas, A., Mehdad, Y., & Chang, Y. (2016, April). Abusive language detection in online user content. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 145-153). International World Wide Web Conferences Steering Committee.

- Noguchi, Y. (2006). Tracking Terrorists. <http://www.washingtonpost.com/wp-dyn/content/discussion/2006/04/11/DI2006041100626.html>, Accessed 06 March 2018.
- Office of General Counsel (2009). Hand book of free speech issues. The California State University.
- O'hara, R. E., Armeli, S., Scott, D. M., Covault, J., & Tennen, H. (2015). Perceived Racial Discrimination and Negative-Mood-Related Drinking Among African American College Students. *Journal of studies on alcohol and drugs*, 76(2), 229-236.
- Olteanu, A., Castillo, C., Boy, J., & Varshney, K. R. (2018). The Effect of Extremist Violence on Hateful Speech Online. *arXiv preprint arXiv:1804.05704*.
- Onwuachi-Willig, A. (2018). What About# UsToo?: The Invisibility of Race in the# MeToo Movement. *The Yale law journal forum*, 104-120.
- Opotow, S., & McClelland, S. I. (2007). The intensification of hating: A theory. *Social Justice Research*, 20(1), 68-97.
- Pace, E. (2011). Religion as communication. *International Review of Sociology*, 21(1), 205-229.
- Pandya, H., & Pandya, H. (2011). Racial discrimination and human trafficking in India: Challenges ahead. *International Journal of Humanities and Social Science*, 1(6), 97-121.
- Paniagua, J., & Sapena, J. (2014). Business performance and social media: Love or hate?. *Business horizons*, 57(6), 719-728.
- Parajuli, J. (2007). A Content Analysis of Selected Government Web Sites: a Case Study of Nepal. *Electronic Journal of E-government*, 5(1), 87-94.
- Patry, M. (2013). India: Digital freedom under threat?. From https://www.indexoncensorship.org/wp-content/uploads/2013/11/india_digital-freedom-under-threat.pdf, Accessed date: 17 April 2018.

- Patton, D. U., Eschmann, R. D., & Butler, D. A. (2013). Internet banging: New trends in social media, gang violence, masculinity and hip hop. *Computers in Human Behavior*, 29(5), A54-A59.
- Pauwels, L. J. R., & Svensson, R. (2017). How robust is the moderating effect of extremist beliefs on the relationship between self-control and violent extremism? *Crime & Delinquency*, 63(8), 1000–1016.
- Penal Code of Poland (1997). https://www.unodc.org/res/cld/document/pol/1997/penal_code_html/Poland_Penal_Code_1997.pdf, Accessed date: 17 April 2018.
- Pew Research Center. (2015). *Latest Trends in Religious Restrictions and Hostilities*. http://assets.pewresearch.org/wpcontent/uploads/sites/11/2015/02/Restrictions2015_fullReport.pdf, Accessed date: 17 April 2018.
- Platts, A., Griesbach, D., & Mackay, F. (2019). One Scotland: Hate Has No Home Here: Consultation on amending Scottish hate crime legislation: Analysis of responses, A Report, June 2019.
- Public Order Act (1986). http://www.legislation.gov.uk/ukpga/1986/64/pdfs/ukpga_19860064_en.pdf, Accessed date: 17 April 2018.
- Qin, J., Zhou, Y., & Chen, H. (2011). A multi-region empirical study on the internet presence of global extremist organizations. *Information Systems Frontiers*, 13(1), 75-88.
- Quayle, E., & Taylor, M. (2011). Social networking as a nexus for engagement and exploitation of young people. *Information Security Technical report*, 16, 44-50.
- Radhakrishnan, S. (October 2018). India glows with #MeToo. <https://www.thehindu.com/sci-tech/technology/internet/india-glows-with-metoo/article25239194.ece>, access date: 20 November 2018.
- Rahman, O., Fung, B., & Yeo, A. (2016). Exploring the Meanings of hijab through online comments in Canada. *Journal of Intercultural communication research*, 45(3), 214-232.
- Rainie, H., Anderson, J. Q., & Albright, J. (2017). *The future of free speech, trolls, anonymity and fake news online*. Washington, DC: Pew Research Center.

- Rajagopal, K. (2018). 'Protecting marriages across castes', <https://www.thehindu.com/opinion/op-ed/protecting-marriages-across-castes/article22466351.ece>, Accessed date: 20 April 2019.
- Razavi, A. H., Inkpen, D., Uritsky, S., & Matwin, S. (2010, May). Offensive language detection using multi-level classification. In *Canadian Conference on Artificial Intelligence* (pp. 16-27). Springer Berlin Heidelberg.
- Roulstone, A., Thomas, P., & Balderston, S. (2011). Between hate and vulnerability: unpacking the British criminal justice system's construction of disablist hate crime. *Disability & Society*, 26(3), 351-364.
- Ruane, K. A. (2014). Freedom of speech and press: Exceptions to the First Amendment. Congressional Research Service.
- Reji, S. K. (2021). Infusion of information and communication technology (ict) in gujarat rural development with e-governance. *Media, Culture and Society*, 22, 24.
- Rempel, J. K., & Burris, C. T. (2005). Let me count the ways: An integrative theory of love and hate. *Personal Relationships*, 12(2), 297-313.
- Retzinger, S. M. (1991). Shame, anger, and conflict: Case study of emotional violence. *Journal of Family Violence*, 6(1), 37-59.
- Reynolds, K., Kontostathis, A., & Edwards, L. (2011, December). Using machine learning to detect cyberbullying. In *Machine learning and applications and workshops (ICMLA), 2011 10th International Conference on* (Vol. 2, pp. 241-244). IEEE.
- Ridgeway, C. L. (2011). *Framed by gender: How gender inequality persists in the modern world*. Oxford University Press.
- Rivas-Drake, D., Hughes, D., & Way, N. (2009). A preliminary analysis of associations among ethnic-racial socialization, ethnic discrimination, and ethnic identity among urban sixth graders. *Journal of Research on Adolescence*, 19(3), 558-584.
- Roberts, N., & Everton, S. F. (2011). Strategies for Combating Dark Networks. *Journal of Social Structure*. 12, 1-32.

- Roberts, C., Innes, M., Williams, M. L., Tregidga, J., & Gadd, D. (2013). Understanding who commits hate crimes and why they do it. *Welsh Government Social Research*, Number: 38/2013.
- Ross, B., Rist, M., Carbonell, G., Cabrera, B., Kurowsky, N., & Wojatzki, M. (2017). Measuring the reliability of hate speech annotations: The case of the european refugee crisis. *arXiv preprint arXiv:1701.08118*.
- Rowley, J. (2008). Understanding digital content marketing. *Journal of marketing management*, 24(5-6), 517-540.
- Rudra, K., Sharma, A., Ganguly, N., & Ghosh, S. (2016, August). Characterizing communal microblogs during disaster events. In *Advances in Social Networks Analysis and Mining (ASONAM), 2016 IEEE/ACM International Conference on* (pp. 96-99). IEEE.
- Saha, K., Chandrasekharan, E., & De Choudhury, M. (2019, June). Prevalence and psychological effects of hateful speech in online college communities. In *Proceedings of the 10th ACM Conference on Web Science* (pp. 255-264).
- Saidi, F., Trabelsi, Z., Salah, K., & Ghezala, H. B. (2017). Approaches to analyzing cyber terrorist communities: Survey and challenges. *Computers & Security*, 66, 66-80.
- Sample, R. J. (2003). *Exploitation: What it is and why it's wrong*. Rowman & Littlefield.
- Sanchez, G. (2013). PLS path modeling with R. *Berkeley: Trowchez Editions*, 383, 2013.
- Sangsuwan, K. (2013) 'Balancing freedom of speech on the Internet under international law', *NCJ Int'l L. & Com. Reg.*, 39, 701.
- Sasireka, G., Balamurugan, S., Gnanasekaran, D., & Gopalakrishnan, S. (2011). Use of E-resources in digital environment among engineering institutions in Tamil Nadu (India): an empirical study. *European Journal of Scientific Research*, 60(3), 326-333.

- Sawyer, R. & Chen, G.M. (2012). The Impact of Social Media on Intercultural Adaptation. *Intercultural Communication Studies*, 21(2), 151-169.
- Schaffner, B. F., MacWilliams, M., & Nteta, T. (2016). Explaining white polarization in the 2016 vote for president: The sobering role of racism and sexism. In *Conference on the US Elections*.
- Scheff, T. J. (2003). Shame in self and society. *Symbolic interaction*, 26(2), 239-262.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. 2nd Edition, Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey London.
- SCJ,. (2018). The Supreme Court judgement. http://www.governancenow.com/files/Judgement_27-Mar-2018.pdf, Accessed date: 20 April 2019.
- Seglow, J. (2016). Hate Speech, Dignity and Self-Respect. *Ethical Theory and Moral Practice*, 19(5), 1103-1116.
- Serper, M.R. (2011). Aggression in schizophrenia. *Schizophrenia bulletin*, 37(5), 897-898.
- Sharma, A. & Sengupta, H. (2020). COVID-19 has accelerated India's digital reset. World Economic Forum, <https://www.weforum.org/agenda/2020/08/covid-19-has-accelerated-india-s-digital-reset>, Accessed 10 March 2021.
- Shensa, A., Escobar-Viera, C. G., Sidani, J. E., Bowman, N. D., Marshal, M. P., & Primack, B. A. (2017). Problematic social media use and depressive symptoms among US young adults: A nationally-representative study. *Social Science & Medicine*, 182, 150-157.
- Shtatland, E. S., Kleinman, K., & Cain, E. M. (2002). One more time about R2 measures of fit in logistic regression. *NESUG 15 proceedings*, 15, 222-226.
- Siapera, E., & Viejo-Otero, P. (2021). Governing Hate: Facebook and Digital Racism. *Television & New Media*, 22(2), 112-130.
- Simons, R. N. (2015). Addressing Gender-Based Harassment in Social Media: A Call to Action. *iConference 2015 Proceedings*, 1-6.

- Sitlhou, H., & Punathil, S. (2014). Research in progress: Racial Discrimination and Violence against Northeasterners and the Bezbaruah Committee Report.
- Smalls, C. (2009). African American adolescent engagement in the classroom and beyond: The roles of mother's racial socialization and democratic-involved parenting. *Journal of Youth and Adolescence*, 38(2), 204-213.
- Social, W. A. (2017). Digital in 2017 global overview. *January, dostupno na: <https://www.slideshare.net/wearesocialsg/digital-in-2017-global-overview> (24.04. 2017.)*.
- Sommer, J., Iyican, S., & Babcock, J. (2019). The relation between contempt, anger, and intimate partner violence: A dyadic approach. *Journal of interpersonal violence*, 34(15), 3059-3079.
- Sood, S. O., Churchill, E. F., & Antin, J. (2012). Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, 63(2), 270-285.
- Spertus, E. (1997, July). Smokey: Automatic recognition of hostile messages. In *AAAI/IAAI* (pp. 1058-1065).
- Srivastava, G. (2000). *Women's Higher Education in the 19th Century*. Concept Publishing Company, New Delhi.
- Srivastava, N. (2009). Bollywood as national (ist) cinema: Violence, patriotism and the national-popular in Rang De Basanti. *Third Text*, 23(6), 703-716.
- Sternberg, R. J. (2003). A duplex theory of hate: Development and application to terrorism, massacres, and genocide. *Review of general psychology*, 7(3), 299-328.
- Stevens, B. (2011). Structural barriers to sexual autonomy for disabled people. *Hum. Rts.*, 38, 14.
- Stevens-Watkins, D., Perry, B., Harp, K. L., & Oser, C. B. (2012). Racism and illicit drug use among African American women: The protective effects of ethnic identity, affirmation, and behavior. *Journal of Black Psychology*, 38(4), 471-496.

- Stoycheff, E., Liu, J., Wibowo, K.A. & Nanni, D.P. (2017). What have we learned about social media by studying Facebook? A decade in review. *New Media & Society*, 19(6), 968-980.
- Sundar, N. (2004). Teaching to hate: RSS'pedagogical programme. *Economic and Political Weekly*, 1605-1612.
- Tashakkori, A., & Creswell, J. W. (2007). The new era of mixed methods. *Journal of mixed methods research*. 1(1), 3-7.
- Tatum, B. D. (2001). Defining racism: Can we talk. *Race, class, and gender in the United States: An integrated study*, 100-107.
- Tejada, J. J., & Punzalan, J. R. B. (2012). On the misuse of Slovin's formula. *The Philippine Statistician*, 61(1), 129-136.
- The Constitution of India. (2007). <http://lawmin.nic.in/coi/coiason29july08.pdf> Accessed 09 February 2017.
- The Constitution of the Republic of Poland (1997). http://www.wipo.int/wipolex/en/text.jsp?file_id=194980, Accessed date: 17 April 2018.
- The Hindu. (2018), “#UsToo-on India's #MeToo moment”, <https://www.thehindu.com/opinion/editorial/ustoo/article25195406.ece>, access date: 20 November 2018.
- Thiara, R. K., Hague, G., & Mullender, A. (2011). Losing out on both counts: disabled women and domestic violence. *Disability & Society*, 26(6), 757-771.
- Thomas, P. (2011). 'Mate crime': ridicule, hostility and targeted attacks against disabled people. *Disability & Society*, 26(1), 107-111.
- Thrasher, J., & Handfield, T. (2018). Honor and violence. *Human Nature*, 29(4), 371-389.
- Tichy, L. L., Becker, J. V., & Sisco, M. M. (2009). The downside of patriarchal benevolence: Ambivalence in addressing domestic violence and socio-economic considerations for women of Tamil Nadu, India. *Journal of Family Violence*, 24(8), 547-558.

- TOI. (2018). Honour killings: More than 300 cases in last three years. <https://timesofindia.indiatimes.com/india/honour-killings-more-than-300-cases-in-last-three-years/articleshow/65908947.cms>, Access date: 19 April 2019.
- Tornberg, A., & Tornberg, P. (2016). Muslims in social networks discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context and Networks*, 13, 132-142.
- Torok, R. (2013). Developing an explanatory model for the process of online radicalization and terrorism. *Security Informatics*, 2(6), 1-10.
- Tripathi, V. (2017). Youth violence and social media. *Journal of Social Sciences*, 52(1-3), 1-7.
- Tseis, A. (2009). Dignity and speech: The regulation of hate speech in a democracy. *Wake Forest L. Rev.*, 44, 497.
- UAE Anti-discriminatory Law. <http://www.emirates247.com/news/government/uae-antidiscriminatory-law-bans-hate-speech-promotion-of-violence-2015-07-22-1.597389> Accessed 09 February 2017.
- UDHR-Universal Declaration of Human Rights. (1948). http://www.ohchr.org/EN/UDHR/Documents/UDHR_Translations/eng.pdf, Accessed 06 February 2017.
- Uemura, R. (2017). Minority statuses and positive views of the country in the United States: an investigation of nativity/immigrant status, and racial/ethnic and religious backgrounds. *Ethnic and Racial Studies*, 40(6), 931-950.
- Ullmann, S., & Tomalin, M. (2020). Quarantining online hate speech: technical and ethical perspectives. *Ethics and Information Technology*, 22(1), 69-80.
- Vartanian, L. R., Thomas, M. A., & Vanman, E. J. (2013). Disgust, contempt, and anger and the stereotypes of obese people. *Eating and Weight Disorders-Studies on Anorexia, Bulimia and Obesity*, 18(4), 377-382.
- Vishwanath, J., & Palakonda, S. C. (2011). Patriarchal ideology of honour and honour crimes in India. *International Journal of Criminal Justice Sciences*, 6(1/2), 386.
- Walker, J. (2013). *Canadian Anti-hate Laws and Freedom of Expression*. Library of Parliament, Canada.

- Walter-Brice, A., Cox, R., Priest, H., & Thompson, F. (2012). What do women with learning disabilities say about their experiences of domestic abuse within the context of their intimate partner relationships?. *Disability & Society*, 27(4), 503-517.
- Walters, M., Brown, R., & Wiedlitzka, S. (2016). Causes and motivations of hate crime. *Equality and Human Rights Commission*, Research Report 102, Manchester, <https://tandis.odihhr.pl/bitstream/20.500.12389/22282/1/08618.pdf>.
- Wang, S. J. (2007). Measures of retaining digital evidence to prosecute computer-based cyber-crimes. *Computer standards & interfaces*, 29(2), 216-223.
- Wang, S., Zhang, X., Zhang, Y., Wang, L., Yang, J., & Wang, W. (2017). A survey on mobile edge networks: Convergence of computing, caching and communications. *IEEE Access*, 5, 6757-6779.
- Warner, W., & Hirschberg, J. (2012, June). Detecting hate speech on the world wide web. In *Proceedings of the Second Workshop on Language in Social Media* (pp. 19-26). Association for Computational Linguistics.
- Weedon, J., Nuland, W. & Stamos, A. (2017). Information operations and Facebook. *version, 1*, 27.
- Weimann, G. (2010). Terror on facebook, twitter, and youtube. *The Brown Journal of World Affairs*, 16(2), 45-54.
- Weimann, G. (2014). *New terrorism and new media*. Washington, DC: Commons Lab of the Woodrow Wilson International Center for Scholars.
- Weiss, T. G. (2000). Governance, good governance and global governance: conceptual and actual challenges. *Third world quarterly*, 21(5), 795-814.
- Wendell, S. (1989). Toward a feminist theory of disability. *Hypatia*, 4(2), 104-124.
- Werlin, H. H. (2003). Poor nations, rich nations: A theory of governance. *Public Administration Review*, 63(3), 329-342.
- West, L. M., Donovan, R. A., & Roemer, L. (2010). Coping with racism: What works and doesn't work for Black women?. *Journal of Black Psychology*, 36(3), 331-349.

- Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration. *MIS Quarterly*, 33(1), 177-195.
- Wikan, U. (1984). Shame and honour: A contestable pair. *Man*, 635-652.
- Wilkinson, B., & Barclay, J. (2011). *The Language of Jihad: Narratives and Strategies of Al-Qa'ida in the Arabian Peninsula and UK Responses*. Royal United Services Institute, London.
- Williams, M. L., & Burnap, P. (2015). Cyberhate on social media in the aftermath of Woolwich: A case study in computational criminology and big data. *British Journal of Criminology*, 56(2), 211-238.
- Williams, M. L., Burnap, P., & Sloan, L. (2017). Crime sensing with big data: The affordances and limitations of using open-source communications to estimate crime patterns. *The British Journal of Criminology*, 57(2), 320-340.
- Wodak, R., & Reisigl, M. (1999). Discourse and racism: European perspectives. *Annual Review of Anthropology*, 28(1), 175-199.
- Wolfe, K. (2018). Narrative Form and Agency in #MeToo. *Student Research Submissions*, 227, https://scholar.umw.edu/student_research/227, accessed date: 22 November 2018.
- Wolny, W. (2016, September). Knowledge gained from twitter data. In *Computer Science and Information Systems (FedCSIS), 2016 Federated Conference on* (pp. 1133-1136). IEEE.
- Women, V. A. (2002). Effects on reproductive health. *Outlook*, 20(1), 1-8.
- Yadav, B. (2009). Khap panchayats: Stealing freedom?. *Economic and Political Weekly*, 16-19.
- Yadav, A. (2018). Countering hate speech in India: Looking for answers beyond the Law. *ILI Law Review*, 2.
- Yousafzai, G. J., & Khan, Q. (2011). Rudeness, Race, Racism and Racialism in EM Forster's "A Passage to India". *Dialogue*, 6(1), 76.

APPENDIX

Questionnaire:

National Institute of Technology Karnataka-Surathkal

Perceptions and Experiences of Hate Content through Online Social Media

As a part of a research project undertaken at NITK Surathkal, I am looking for the opinions of people about their perspectives and experiences with the hate content in India. Thank you in advance for sparing the time to contribute. There are no right or wrong answers, and these are your opinions that are of most importance. Hope that you will indeed contribute to this survey so that your opinions can make a difference and influence policy changes. For the convenience of answering, some of the terms are defined below.

Gendered hate content: The content which is expressed against the gender identity with a feeling that their gender is superior to others.

Religious hate content: This content is expressed against the person or a group, which belongs to a religion other than the perpetrators' religion.

Racist hate content: The racist content is expressed against the person or the group based on ethnicity or traditional culture.

Disability hate content: This content is expressed against persons with disabilities by the feeling of being able-bodied.

Perpetrator: The person who expresses hate against others.

NAGANNA CHETTY

nsc.chetty@gmail.com

QUESTIONNAIRE

Please mark anyone's answer for the following questions from 1 to 4.

	Age under 18 years[1]	Age 18-24 years [2]	Age 25-34 years [3]	Age 35-44 years [4]	Age 45-54 years [5]	Age above 54 years[6]
1. Which age group expresses more hate on the Internet?						

	10 th [1]	12 th [2]	UG [3]	PG & Above[4]	Others[5]
2. Which educational level expresses more hate on the Internet?					

	Income ≤ 25000 [1]	Income from 26000 - 5 lakh [2]	Income > 5 lakh [3]
3. Which income group expresses more hate on the Internet?			

	0-3 hrs./ week [1]	4-6 hrs./ week [2]	7-9 hrs./ week [3]	10-20 hrs./ week [4]	21-40 hrs./ week [5]	Above 40 hrs./week [6]
4. On average, how many hours in a week do you use the Internet?						

Please indicate your answer of how strongly you accept the frequent occurrence of hate content on the following categories through social media in India.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
5. Gendered hate content is more in social media of India					
6. Religious hate content is more in social media of India					
7. Racist hate content are more frequent					
8. Hate content on the disabled are more frequent					

Please indicate your acceptance level with the following statement about hatred expression.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
9. The communal nature of a person leads to the expression of hate on the Internet					
10. A terrorist attack on any public event/gathering leads to hate expression on the Internet					
11. The occurrence of hate content leads to terrorist activities					
12. Freedom of expression supports hate content generation on the Internet					
13. Extremism is one of the causes for hate content occurrence on the Internet					

Please mark in relevant boxes your acceptance of the existence of gendered hate content in India with the following statements.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
14. Expression of hate is common on/by the divorced person					
15. An attitude of a person leads to hate					
16. Hate experienced person expresses more hate					

Show your acceptance of the occurrence of religious hate in India with the following statements.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
17. Hate is because of the superior feeling about own religion					
18. Hate is because of a lack of religious knowledge					
19. Hate is because of the purpose of religionizing the nation					
20. Hate is because of the different religious clothing style					

Please indicate your acceptance of the existence of racism in India in relevant boxes with the following statements.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
21. Racial hate is because of inherited nature from parents					
22. Racial hate is because of the pressure from colleagues/friends					
23. Racial hate is because of exposure to stereotypes on racism					
24. Racial hate is because of unfamiliarity about ethnicity/common tradition					
25. Hate experienced person usually expresses racial hate					
26. Racial hate is because of the skin color of a person					
27. Racial hate is because of the poverty					

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]

28. Disability hate is because of lack of knowledge					
29. Disability hate is because of jealousy on disabled person's benefits					
30. Disability hate is because of the fear of communicable diseases from disabled persons					
31. Disability hate is because of being proud as able-bodied (Not disabled)					

Provide your acceptance for the following set of statements.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
32. Online gendered hate leads to suicide attempts by the victim					
33. With gendered hate on Internet, victims shift physically from one location to another					
34. Online gendered hate leads to change of name by the victim					
35. With gendered hate on Internet, victims withdraw their online presence					
36. Religious hate undermines the religion of the victim					
37. With online religious hate, people feel bad about their religion					
38. Religious hate weakens the inter-religious relationship					

Indicate your acceptance of the following set of statements.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
39. Racism forces non-racists to behave like racists					
40. Racism destroys the social capital of the victims					
41. Racism reduces the quality of education and employment to the target group					
42. Hate content affects an entire group or class of people					
43. More disability hate content is generated by the people close (friends /caretakers /neighbors) to the victim					
44. Disability hate content discusses the nature of the victim's disability					

45. Disability hate affects the physical and mental health conditions of the victim					
--	--	--	--	--	--

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
46. Strengthen legislation relating to hate content expression will reduce hate content					
47. Increased penalties for hate content expression will reduce hate content					
48. Government-led information campaign through social media on hate content expression will reduce hate content					
49. Private sector-led information campaign through social media on hate content expression will reduce hate content					
50. Civil Society led information campaign through social media on hate content expression will reduce hate content					
51. Social media companies are required to monitor and delete those accounts use for hate content generation					
52. Encouraging people to be careful during online presence as a personal responsibility will reduce hate content					
53. The limited presence of people with social media and other online services will reduce hate content					

Please indicate your online presence with the following statements.					
	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
54. I use social networking sites (such as Facebook and Twitter) frequently					
55. I use blogging sites (such as WordPress and Blogger) frequently					
56. I use photo sharing sites (such as Instagram and Flickr) frequently					
57. I use video sharing sites (such as YouTube) frequently					
58. I use e-mail frequently					
59. I watch films frequently					

60. I Play games frequently					
61. I do online shopping frequently					
62. I search for information frequently					
63. I use the Internet for other activities					

Do you/your friends/your relatives have faced any online hatred harassment? Provide your response with the following categories of hate.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
64. I/my friends/my relatives harassed with gendered hate content					
65. I/my friends/my relatives harassed with religious hate content					
66. I/my friends/my relatives harassed with racist hate content					
67. I/my friends/my relatives harassed with hate content on disabled					
68. I/my friends/my relatives harassed with sexually-oriented hate content					

Have you been blocked by any blocking mechanisms while searching for information on the Internet? Please provide your acceptance with the following blocking mechanisms.

	Strongly disagree [1]	Disagree [2]	Unsure [3]	Agree [4]	Strongly agree [5]
69. I was blocked by personal computer level security					
70. I was blocked by organizational level security					
71. I was blocked by national authorities (Government) using internet service providers (ISPs)					
72. I was blocked by national authorities at backbone networks for security					

73. Please tick your age group.

- 15-19
- 20-29
- 30-39
- 40-49
- More than 49

74. How do you describe your gender?

- Male

- Female
 - Transgender (Other)
75. Please provide your physical characteristic (Optional).
- Able-bodied
 - Disabled
76. Please provide your skin color/complexion (Optional).
- White
 - Black
77. Please provide your nationality.
- Indian
 - Foreigner
78. Please specify your native state or union territory:
79. Please provide your resident location.
- Urban area
 - Rural area
80. With which religion do you identify yourself?
- Hindu
 - Islam
 - Buddhism
 - Christian
 - Sikhism
 - Others
81. What sort of device do you use most often to access the Internet?
- Simple phone
 - Smartphone
 - Tablet
 - Laptop
 - Others
82. Please tick the type of your organization.
- Public sector
 - Private sector
83. Your highest qualification is :
84. Please add anything else you would like to mention about hate content expression in India.
-

Once again, thank you very much for participating in this important survey.

BIO-DATA

Naganna Chetty

School of Management

National Institute of Technology Karnataka,

Surathkal, Srinivasnagar - 575025

Ph: +91 7760621885

Mail ids: nsc.chetty@gmail.com / sm16f03.naganna@nitk.edu.in

WORK EXPERIENCE

Assistant Professor | Mangalore Institute of Technology and Engineering,
Moodabidri Year: 2014 - 2016

Assistant Professor | Galgotias University, Greater Noida
Year: 2012 - 2014

Assistant Professor | Quantum School of Technology, Roorkee
Year: 2009 - 2012

Assistant Professor | Graphic Era Institute of Technology, Dehradun
Year: 2007 - 2009

Lecturer | BLDEA's College of Engineering, Bijapur
Year: 2003 - 2007

Lecturer | MBES's College of Engineering, Ambajogai
Year: 2001 - 2003

Lecturer | NES Polytechnic Gulbarga
Year: 1999 - 2001

EDUCATIONAL QUALIFICATION

Ph.D. Research Scholar | National Institute of Technology Karnataka, Surathkal
Research Area: Information Systems Year: 2016 - Present

Master of Technology (M.Tech.) | K.B.N. College of Engineering, Gulbarga
Specialization: Computer Science & Engineering Year: 2004 - 2006
University: Visveswaraya Technological University, Belgaum

Bachelor of Engineering (B.E.) | P.D.A. College of Engineering, Gulbarga
Specialization: Computer Science & Engineering Year: 1995 - 1999
University: Gulbarga University, Gulbarga

PUBLICATIONS ACCEPTED

1. Chetty, N., & Alathur, S. (2018). Hate speech review in the context of online social networks. *Aggression and Violent Behavior*, 40 (May-June), 108-118.
2. Chetty, N., & Alathur, S. (2019). Racism and Social Media: A Study in Indian Context. *International Journal of Web Based Communities*, 15(1), 44-61.
3. Chetty, N., & Alathur, S. (2019). Digital hate content reduction with mobile edge computing: An architecture. *Digital Communications and Networks*, 6(2), 217-222.
4. Chetty, N., & Alathur, S. (2019). Honour, hate and violence in social media: insights from India. *International Journal of Web Based Communities*, 15(4), 315-326.
5. Chetty, N., Alathur, S. & Kumar, V. (2021). Technology Assisted Social Reforms and Online Hate Content: Insights from India, *Brazilian Archives of Biology and Technology* (**In Press**).
6. Chetty, N., & Alathur, S. (2021). Identity-Based Online Hate Content: Empirical Analysis. *International Journal of Cyber Behavior, Psychology and Learning (IJCBL)*, 11(4), 1-24.
7. Chetty, N., & Alathur, S. (2019, April). Trigger Event and Hate Content: Insights from Twitter Analytics. In *2019 International Conference on Advances in Computing and Communication Engineering (ICACCE)* (pp. 1-5). IEEE.
8. Chetty, N., & Alathur, S. (2020). Policies to Mitigate Select Consequence of Social Media: Insights from India. *ICT Systems and Sustainability: Proceedings of ICT4SD 2019, Volume 1, 1077*, 351-359.

Other Co-authored Publications During Ph.D.

1. Chetty, N., & Alathur, S. (2020). 2019-nCoV Disease Control and Rehabilitation: Insights from Twitter Analytics. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-4). IEEE.
2. Chetty, N., & Alathur, S. (2020, December). Developing Indian Smart Cities: Insights from Social Media. In *International Working Conference on Transfer and Diffusion of IT* (pp. 209-218). Springer, Cham.

3. Andrews, D., Alathur, S., & Chetty, N. (2020). International Efforts for Children Online Safety: A Survey. *International Journal of Web Based Communities*, 16(2), 123-133.
4. Alathur, S., Kottakkunnummal, M., & Chetty, N. (2021). Social media and disaster management: influencing e-participation content on disabilities. *Transforming Government: People, Process and Policy*. <https://doi.org/10.1108/TG-07-2020-0155>,
5. Chetty, N., & Alathur, S. (2020). Mitigation of Disability Hate Content. *Solid State Technology*, 5601-5616.
6. Pai, R., Chetty, N., & Alathur, S. (2020). Impact of COVID-19 on Individuals Mental Health and Preventive Health Behaviors: A Conceptual Framework, *International Journal of Medical Engineering and Informatics (In Press)*.
7. Andrews, D., Alathur, S., & Chetty, N. & Kumar, V. (2020, October). Child Online Safety in Indian Context. In *2020 5th International Conference on Computing, Communication and Security (ICCCS)* (pp. 1-4). IEEE.
8. Andrews, D., Alathur, S., & Chetty, N. (2020, December). Child Online Safety Intervention Through Empowering Parents and Technical Experts: Indian Context. In *International Working Conference on Transfer and Diffusion of IT* (pp. 662-673). Springer, Cham.
9. Chetty, N., & Alathur, S. (2021, October). Climate Change and COVID19 Metaphors: Environmental Consciousness in social media. In *2021 6th IEEE International Conference on Computing, Communication & Security* held at Las Vegas, USA during 04-06 October 2021.
10. Chetty, N., Alathur, S., Andrew, D. & Kumar, V. (2021, October). Computational Analysis of Online Hate Content using Cognitive - AI. In *2021 6th IEEE International Conference on Computing, Communication & Security* held at Las Vegas, USA during 04-06 October 2021.

PUBLICATIONS UNDER REVIEW

1. Chetty, N., & Alathur, S. Online Hate Content Research in Two Decades: A Bibliometric Review and Analysis.
2. Chetty, N., & Alathur, S. Digital Hate Content and Disability: Insight from Twitter Analysis.

3. Chetty, N., & Alathur, S. Impact of Disability Acts on Full Inclusion: Insights from Social Media Analytics.
4. Chetty, N., & Alathur, S. Toxic Content During COVID-19: Insights from Social Media.

EXTRA-CURRICULAR ACTIVITIES

Participated in Industrial visits and workshops held at GEIT Dehradun, QST Roorkee, GU Greater Noida, MITE Moodabidri, and NITK Surathkal.

AWARDS AND RECOGNITIONS

1. Received the best paper award for the paper entitled “2019-nCoV Disease Control and Rehabilitation: Insights from Twitter Analytics” in *2020 5th International Conference on Computing, Communication and Security* held at IIT Patna, India.
2. Ph.D. research work is accepted and presented as a poster in UNESCO’s Information and Communication Technologies for Development 2020 (ICT4D2020) Non-Conference held on 15-17 September 2020 at Royal Holloway, University of London, UK. <https://ict4d2020.org/posters-chetty/>.

PERSONAL PROFILE

Name: : Naganna Chetty
Gender: : Male
Date of Birth : 10 December 1975
Marital Status : Married
Address : 16/98, Anugraha, Udayanagar, Mukka, Mangalore - 575025
Nationality : Indian

DECLARATION:

I hereby declare that the above-mentioned information is correct up to my knowledge and I bear the responsibility for the correctness of the above-mentioned particulars.

Date: 08 March 2022

Place: Surathkal

Naganna Chetty