2020 International Conference on Computational Performance Evaluation (ComPE),
North-Eastern Hill University, Shillong, Meghalaya, India. July 2–4, 2020

Characterizing Social Media Contents for Regulating Hate Crimes and Cyber Racism against Marginalized and Dalits In India

Amita Charan
Department of Commerce,
Janki Devi Memorial College (University of Delhi),
New Delhi, India
amitacharan@jdm.du.ac.in

Jitendra Kumar Verma
Amity School of Engg. & Technology,
Amity University Haryana,
Gurgram (Manesar), India
jitendra.verma.in@ieee.org

***Abstract*—This article is an attempt to identify the role of social media in regulating hate crimes and cyber racism in India. The freedom of speech is given to the citizens in many countries but a few disturbing elements are misusing this freedom and suppressing the voice of marginalized and Dalits. It is observed that the Fake news contents, hate crimes and cyber racism are increasing in India. The vigilance of internet contents and digital media is in nascent phase and need to be governed in the light of human rights and freedom of speech [1].**

One side network of haters and fake news bibliophiles are expanding their network in digital media at a faster rate. On the other side cohort of advocates dealing in International agreements, Human Rights and International Laws are establishing a sound framework for raising voice of marginalized on all possible decisive platforms. User generated contents are future of social media therefore, sketching the characters and identifying cyber crimes in advance must be an integral feature of the programming on social networking sites. This article is intended to identify overall impact of hate crimes on the society in general and on the Dalit and marginalized sections more specifically. The article will suggest some proactive measures for characterizing social media contents in order to regulate these contents by designated authorities or popular social networks.

***Index Terms*—Hate Speech, Objectionable contents, Characterizing contents, Social media, Hate crimes.**

I. Introduction

The comprehensive judicial system of any country is expected to maintain sanity, transparency and tranquility. Unfortunately, recent collegium system reflects trivial dominance and impact of bureaucratic regime on judicial decisions in many developed and developing nations. Criminal mindsets are indulged in spreading hate, vandalism and conflicting contemplations against minority groups and marginalized sections in the society [2]. It is very difficult to identify the root causes of violent extremism and racial crimes in the world [3]. It is observed that Online Hate Speeches are quite often used to dehumanize both marginalized groups and pathetic individuals [4] [5]. A few narcissistic leaders are using digital platforms to mislead majority that too either in the absence of strong opposition or by maligning the image of a race, caste, weak opposition, political representatives, human right activists and minority representatives [1]. The media especially interactive and broadcasting paid media is denigrated to play a divisible role is spreading political ideologies, religious fallacy and malicious paraphernalia during a sensitive time zone, emergency situations (or before election) and epidemics for creating its immediate, robust and deleterious impact on citizens. The print, broadcasting and digital media can play an affirmative role in injecting ethical values, moderating the characters and controlling the hate mongers through various digital tools but unfortunately the representation of marginalized is negligible in the media decisions of many countries including India. Private networks are least bothered to verify the hate contents and malignant characters. YouTube and video contents are future of the digital media and will expand further as per the report of Media and Outlook 2019. Twitter, Facebook and Tick Tock are quite popular among Indian youth where mostly contents are posted by the users and reviewed by the international operators. Many NRIs are living outside the India and posting hate contents. Therefore, the characters must be globally identified from all three modes audio, video and text based networks. In case, these contents are often found as objectionable, provoking, abusive and sometimes depressive the regulatory network or designators must report about the cyber criminals. The mainstream media, corporate and political connects are even polarizing the election process and voters opinion [6] . Due to language and cultural diversity this task would be little tedious and demand regular updates in India and in diversified population. User reporting system can help in random collection of the characters and hate contents.

A few regulations, bills and acts for the protection of freedom of speech, rights of subjugated groups are unexpectedly amended, discarded or drafted in a short span in India and in many other nations. UN Human Right Council is in favour of timely monitoring hate contents [7] [8]. Social activists and leaders are raising their voices against fascism and undue amendments. States are also obliged to adopt various measures to ensure the environment for freedom of speech and equality of citizens [9] in digital age. Criminalizing hate speeches [10] is very important as it may invade the importance of various tribes, languages, DNA segments, races, heritages,

978-1-7281-6644-5/20/$31.00 © 2020 IEEE

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

history and cultures that had been a landmark in the human development process. For example, various tribes (like Jarva, Kol, Santhal and Mundas) are protected in the tribal zones by Indian government in Adman, Jharkhand, Rajasthan, Assam, Nagaland, Meghalaya and Madhya Pradesh so that tourists can not harm and contaminate their native characteristics in natural settings. But various groups are spread hate contents against these tribes and pose them as non-Indians. These are ancient native tribes and our historical roots are connecting chains in our development and civilizations. We cannot afford to lose these tribes at any cost. The hate crimes against race, caste, religion [11] and women through digital media tools are common in many countries like South Africa, Japan, China, United States, France, India, United Kingdom and Germany.

II. Role of media, Freedom and Measures for Regulating Hate Speeches in the past

Various offline media and broadcasting tools were misused for Nazi propaganda during World War II especially radio [12]. Print media particularly newspaper and posters were also popular during and after the World War II. Later, the arena of electronic communications drastically changed as many social media platforms gradually emerged in the mid and late nineties. Henceforth, online hate speeches and online racism [13] gradually appeared in different countries at the same time at different platforms [14].

Nevertheless, print, audio visual and interactive media all are used for dissemination of abhorrence contents in the past. Moreover, audio visuals, live telecasts, digital displays and kiosks are more sensational, speedy, noticeable and powerful modes in recent times because of speed, user interloping, visibility and low cost [1]. When a user is insane, less educated, ignorant, unilingual, discrete and adolescent his/her cognitive ability and perception could be easily influenced by the counterfeited contents. Sometimes people disseminate racial and bitter contents against Muslims, Blacks and racial communities in public or in closed groups without guesstimating the colossal repercussions of the fake contents and hate speeches [15]. The future of media is hidden in interactive media and user generated contents on all possible platforms. The available researches suggest that hate based crime and hate based violence have increased significantly [16] in the entire world and must be countered immediately [17].

The United Nation World Conference against Racism, Racial discrimination, Xenophobia, Islamophobia and Related Intolerance was a major milestone in this area. Many other countries are taking preventive measures for hate speeches in a legitimate manner. In order to tackle discrimination and hate speeches Broadcasting Act 1990 was proposed by UK. In UK Ofcom regulator has legal obligation to promote diversity, inclusion and plurality. In order to tackle discrimination and hate speeches Istanbul process was initiated in 2011. Ofcom and EHRC issued a guideline in 2015 for improving the fairness in diversity. In Germany Interstate Broadcasting Treaty was proposed to protect human dignity and ethical standards. In Italy AGCOM was empowered for the proceedings against

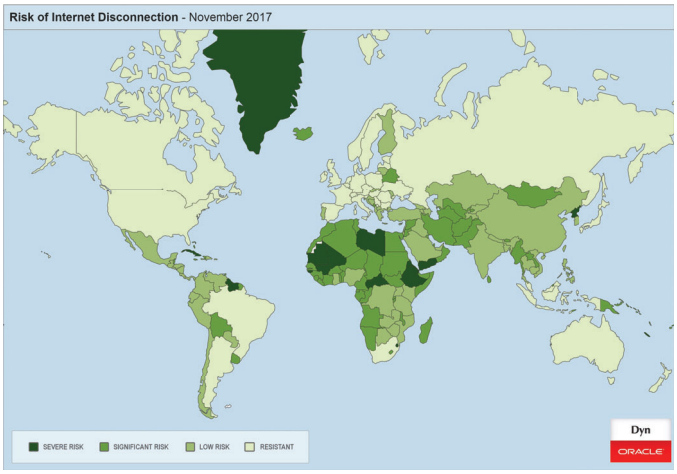

Figure 1: The Migration of Political Internet Shutdowns [21]

broadcasters and agencies. In Poland Broadcasting Council is established for prevention of racial discrimination and other similar issues [9]. 'Words Heal the World' is a non-profit organization functioning in UK and empowering youngsters for handling with different types of extremists. Japan implemented Hate Speech Law in 2016 in response to increasing attack on Korean students. China is restricting online contents and applications through various agencies. China's internet censorship and AI algorithms are more extensive as compared to other nations. Banning internet or social media is not a factual solution as people may resist, and Software Freedom and Law center is reporting real time geographical statistics [18]. International Network Against Cyber Hate (INACH) counter the online discrimination through a sound network of 18 organizations [19]. Insafe is another European network operating in 27 countries and is responsible for safe use of internet. Inhope manage safe hotline around the world for handling discrimination. Isolating a country from digital connectivity is like implementing internet curfew in digital age. UNDP issued guidelines and framework on prevention of violent extremism (PVE). Twitter is also observed by various researches for spreading cyber racism against blacks [15]. India also faced many internet shutdowns recently by November 121 shutdowns were reported and 52 of them were in J&K and 50 were in Rajasthan. India topped the list of internet shutdown in the world in 2019 [18]. India is included in five geo-political risky countries as per the report of Eurasia Group's 2020 report. The Delhi violence and riots has provided another level to the hate contents and crime against marginalized in India. Minorities and Dalits are prime target of these hate crimes in cyber space. 218 cases of hate crimes were reported in 2018 in India [20]. According UN more than 17.5 million Indians are living in other countries, suppose they will get the same treatment in these nations then it would be very difficult for India to handle the situation.

The reach and frequency of broadcasting, social media and other on line sources are much wider as compared to print media. The YouTube is capturing maximum contents and share

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

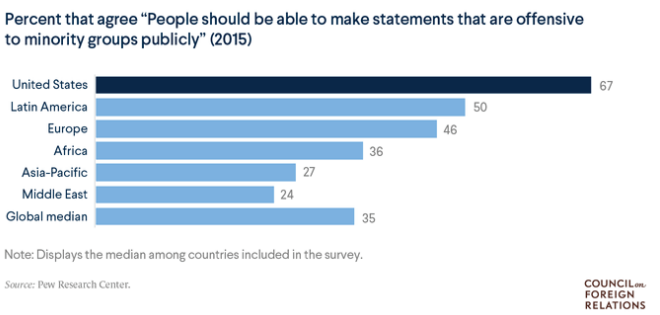

Figure 2: Hate Speech on Social Media: Global Comparisons [22]

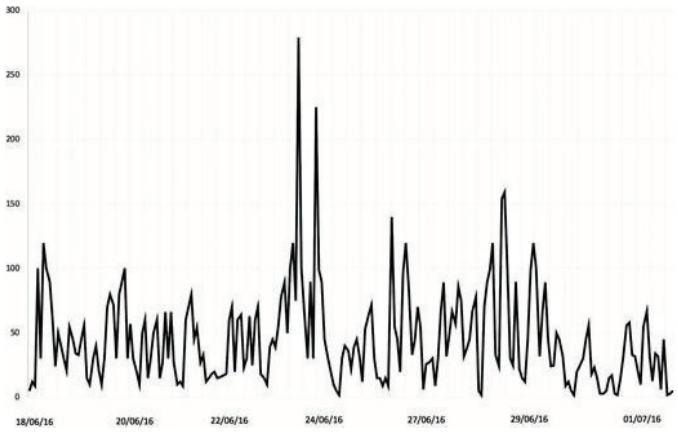

Figure 3: Increase in online hate speech leads to more crimes against minorities [28]

in the interactive media. A survey was conducted in different countries on the hate speeches in 2015 and many people agreed that they should be able to make offensive statements against minority groups.

Nevertheless, the infrastructural and procedural growth of digital literacy [23] is slightly poor in developing countries and only two-third of the world is accessing internet [24]. Marginalized section has limited accesses to digital resources and internet networks [23]. Socio economic barriers disconnected them from IT education and digital evolvement. Further, social networks, artificial intelligence and algorithms amplify opinions and facilitate the quick spread of hate speech against marginalized and Dalits [25] rather than protecting their rights [26]. The social media governance [1] is a narrower concept which is expected to magnify very soon but it is all in the hands of elite class. Hence guidelines, tutorials and pocket guides can help the social media users. The dominance of elite class is reflected everywhere in the management, expansion and directives of social media tools [27] especially in YouTube, Facebook, Twitter, TiK Tock and WhatsApp. The owners, sponsor, distributors, hosting companies have their own motives for investments and corporate ventures in various countries. All three are expanding their investments in India.

Increase in online hate speeches leads to more crimes

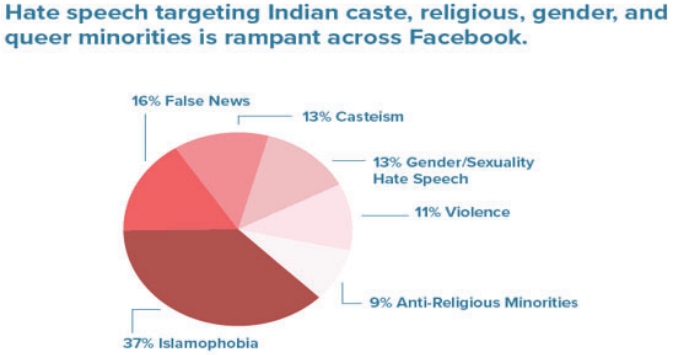

Figure 4: Facebook India–Towards A Tipping Point Of Violence Caste And Religious Hate Speech [32]

against marginalized as shown in the graph. Discriminatory thoughts are one of the main causes of hate crimes and hate speeches. Hate mongering [28] nature of social media is observed [29] by the regulators and technologists but there is a lack of efficient and transparent internet governance. It is catalyzing sponsored cybercrimes, spawned hoax and rapidly provoking folks for violence [3]. Unfortunately Facebook is favourite for hate mongers [30] because of its lucid, user friendly features and liberal security. Facebook communities and closed groups are quite popular among hate mongers especially among homophobic, anti-immigrant, anti-muslim [31], racists, extremists etc. computer keyboards, robots and programs will be more contaminated than explosives in the near future. Indian and South Asian population is very diversified because of climate, origin, dynasties and language. Therefore, this is very sensitive populated zone for such kind crimes. The graph reflects the increase and decrease in the hate crimes against minorities.

III. Judicial System, Reforms and Political Influence Connectivity to Hate Speeches

A large number of online users are facing hate speeches [29] in the world therefore; many countries are amending their broadcasting acts and framing guidelines. The notion for establishing a strong judicial system is based on two principles equality and justice. This idea of "moral disarmament" was initially proposed in the League of Nations by Poland in 1931; it was first brought into the United Nations in 1947 by the Soviet Union. The Durban world conference was a landmark in resolving discrimination and racial issues. By the time judiciaries were systematized in various countries under colonial rule various groups, dynasties and dominions were empowered and flourished in different countries. The inequality is one of the root causes of lopsided decisions and favoritism. Moreover, internet age is spreading the things and contents very fast through Twitter, Line, WhatsApp, Facebook etc., where the freedom of speech is protected by many nations on digital platforms [33]. Nevertheless, ignorant users also spread these contents in a speedy manner without checking the facts. For example, China restricts various external social

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

media tools and browsers in their country but developed on messenger. Digital deception is damaging the real democracy by creating misconception about a policy, party, report, event or person.

Article 19 of Human Right considers various grounds for discriminatory hate need to be tackled such as origin, colour, race, language, birth, gender, religion, disability and migration etc. There are two broad categories of hate speeches where a few that must be prohibited and a few that must be protected [34]. Furthermore, UN Human Rights committee also addressed to the Internet and other mobile technologies and recommended the states to foster the independence and easy access to the individuals [35]. In India judicial bodies and court system both function on the basis of information and inputs received from the majority of people and their elected representatives, precedents, religious philanthropy and faiths, past experience and various other factors in nation. Indian constitution is one of the best in the world and democracy is empowered at each step. Digital dilemma and NeGP is empowering private players to control pan India biometrics and MSO through private networks.

Majority in any country is prejudicedby belief system,values, cultural succession/invasion, leadership, ideology of ruling party, social stratification, education system and various other external and internal sources. There are empirical evidences that judiciary and administration treat aboriginal and non-aboriginal people in an entirely different way. They mislead and limit the information therefore, common man is deprived of real information and takes hasty decisions while voting and electing the representatives for their respective states. Henceforth, technology is also creating misconceptions and invading the factual, accurate, scientific and historical information.

IV. Literature Review

The circulation and printing of hate contents are absolutely banned in India under different Laws. Indian Penal Code (IPC) and Code of Criminal Procedure (CrPC) also suggest the ways to regulate and define the limits of freedom of speech in our country on various platforms. On the other Code of Criminal Procedure empower the government under section 95 to circulate, completely prohibit and regulate hate contents for discouraging communal violence and unrest. Promoting such violence or unrest is also punishable under Sections 124A, 153A, 153B, 292, 293 or 295A of the IPC. IPC Section 153(A) prohibits the promotion of enmity between different groups. IPC Section 153(B), 505(2), 295(A) specify the punishments related to publication, circulation and intentionally offending the religious sentiments of people by words, signs or otherwise. Section 66A of IT Act deals with objectionable and provocative contents. The political connections and increasing influence of powerful people during election is observed therefore, restriction on the freedom of speech is imposed and recommended in the recent past. The use of user friendly medium is also observed in the circulation of hate contents henceforth, it is important to characterize and regulate Online hate speech (audio or video) against dalits and minorities. In most of the cases ambiguity is created by certain influential people for diverting the vote bank and polarization of election. Unfortunately, these episodes caused severe damage to humanity and economy in long run as the stiffness among two groups remains same even after a decade. Therefore, identification of disseminating contents and characters both are important for protection of national wealth and image as a secular body [36]. Prime Minister Narendra Modi safeguarded India's human rights record (PTI, 2016), mentioning the Indian Constitution as a real holy book which provides freedom of expression. The Anual report of the US Commission for International Religious Freedom (USCIRF) reported that the religious freedom in India was on a negative trajectory in 2015 and spritual tolerance had declined while violations of religious freedom had increased. When the Prime Minister said, "800 million of my countrymen may exercise the freedom of franchise once every five years, but all the 1.25 billion of our citizens have freedom from fear which they exercise every moment of their lives", this was clearly not the case for the Bengali poet Srijato Bandyopadhay. In March 2017, a police complaint was lodged against him for posting a 12 line poem on social media that allegedly hurt Hindu religious sentiments. The poem, titled 'Abhishaap' ('Curse'), was posted on Facebook on the day CM of a state was sworn (IANS Feeds, 2017). Recently a senior journalist is also facing sedation charges and demands freedom of speech for independent media. OTT platforms like WhatsApp the contents are shared in heavy volumes and more than 200 million monthly active users are using this platform. Many users are not trained to handle social media tools while posting such contents. Therefore, an online tutorial is required to regulate the cybercrimes and hate contents on various social networks. One, unauthentic message can invoke violence against Dalits and marginalized section in minutes. Handling and regulating these online contents through and advanced technology or artificial intelligence is utmost important at the global and national navel. The AI based tools can sense the conspiracy in a proactive manner before and violent attack by mob or criminal goons.

V. Hate Speeches used by in Indian MP/MLAs

Representatives are elected by people in this largest democracy and it is assumed that their behaviour must be parliamentary behaviour towards the countermen. The tables below present the MP/MLA/Contestant involvement in hate speech cases. In case these representatives will use the hate contents the followers will also follow the similar behaviour. Therefore, a political consensus must be developed by all in prominent parties and the ruling parties may issue relevant guidelines after specified intervals in order to regulate these crimes.

VI. Method

The secondary data is used from twitter especially posted by Indians and around 100 messages are randomly searched using twitter filters which contain the word 'Muslims' and 'Dalit

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

Table I: Presenting Infamous Global Cases of Hate Contents

	Survey	Target and result	Country and Year
1	PeaceTech Lab's (PTL) data portal for monitoring and reporting online hate speech in South Africa [37]	It was observed that the increase in hateful posts was 26% in merely in two weeks in South Africa	South Africa, 2019
2	FBI [38]	The number of hate crimes increased by 17%, 58.1% of 7,106 hate crimes including online contents as reported in 2017 were related to racism or ethnicity	US, 2018
3	Pew Research Center titled Global Restrictions on Religion Rise Modestly in [39]	Harassment by Government Official and use of social hostilities involving religious minorities is increased and reported in ten countries i.e., Syria, Nigeria, Iran, India, Israel, Yemen, Russia, Afghanistan, Palestine and Egypt	10 countries including India, 2017
4	ENAR Ireland Report [40] [41]	According to ENAR Ireland's report approximately 111 cases of online hate speech including racial contents were filed using their iReport mechanism in June 2017	Ireland, 2017
5	Police in England and Wales [42]	The number of hate crime offences known as online crime was increased from 1,148 to 1,605 in between 2016-17. The biggest increase was observed in racially motivated online crimes. It was increased by 210 offences (up 29%) from 717 to 928	England and Wales, 2017-18
6	OSCE has been collecting figures from member countries [43]	A noticeable rise from 19 to 30 % was observed in one year, against Indian Muslims community which constitutes around 1.2-billion minority population in India	Europe, 2017
7	FBI [44]	Media personnel and research analysts reported that the events like the 2016 US Presidential election, the 'refugee crisis' in Europe, and the Brexit referendum have resulted in a worrying escalation of racist hate speech and racist incidents	US, 2017
8	Cardiff University's HateLab project collected Twitter and police recorded crime data [45] [46] [47] from London	An AI based algorithm method may help the police in preventing the cybercrimes including hate speeches, this may be a proactive approach in identifying the hate contents	London, 2016
9	Walters, Brown, & Wiedlitzka [48] [49]	It was observed that the incidence of hate crimes based on	Great Britain, 2016
10	Toronto police services Statistical Report [50] [51]	The report specified the incidence of 28.2% of hate crimes and discrimination were identified and enacted on the basis of race, ethnic origin, and nationality	Canada, 2016
11	Australian Hate Crime Network [52]	Hate crime against races is increasing including islamophobic contents on the various networks	Australia, 2016
12	Ukrainian organization [53]	It was observed that around 50% of negative commentary Was directed at nationalists and 17% at Supporters of the Party of Regions, 15% at supporters of communism of the comments directed towards people of different nationality	Ukrain, 2016
13	South African Human Rights Commission (SAHRC) [54]	Approximately 62 deaths were reported targeted migrants, refugees, and others viewed as foreigners. SAHRC said that hate speech cases on social media increased to 22% of matters investigated, compared to 3% in the same period the previous year.	South Africa, 2015
14	Reported by Junko Kotani [55]	Zaitoku-kai group marched on the roads towards Korean Residential colonies in Tokyo and Osaka. The groped filmed and uploaded the rally videos on internet for spreading hate.	Japan, 2013
15	Observer Research Foundation in March [56]	India has witnessed many cases of large-scale communal violence against Dalits and religious minorities, notably Uttar Pradesh in 2013 known AS Muzaffarnagar riots. In this incident more than 50 deaths and over 75,000 people displaced from their homes. Role of political parties was traced in invoking violence after spreading videos through Facebook videos.	India, 2013
16	Routhland International [57]	Increase in hate crimes was reported and laws were enforced to regulated hate crimes	Sydney, 2012
17	Observer Research Foundation in March [58]	Northeast exodus in the year 2012. Up to 50,000 citizens belonging to the Northeast displaced from their living places and back to the North-eastern states because of hate speeches circulated through SMSs, MMSs and Facebook posts.	India, 2012

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

Table II: Sitting MP/MLAs with hate speech cases registered against them [59]

Hate Speech Cases Against MP/MLAs	
BJP	27
Telangana Rashtra Samiti	6
All India Majlis-e-Ittehadul Muslimmen	6
Samajwadi Parti	5
Telugu Desam Party	4
Independents	3
Indian National Congress	3
Other Parties	12

Table III: Candidates with hate speech cases [11]

Hate Speech Cases Against MP/MLAs	
BJP	97
Independents	54
Indian National Congress	28
Bahujan Samaj Party	28
Samajwadi Party	25
Telangana Rashtra Samiti	15
Shivsena	14
All India Majlis-e-Ittehadul Muslimmen	12
Janata Dal	10
Other Parties	116

during Delhi riots (DATED 23 Feb – 2 March). The datasets are divided in five categories (contents posted directly and the counter speeches are included) i.e., Both, Muslims, Political or other social organization, Others etc. Summary is presented in Table IV.

This is data gathered only from one network using two characters. Nevertheless, if a program or artificial intelligence tool can trace the contents from all the networks it would be easier to regulate the hate contents and investigate any big conspiracy in advance. Therefore, it is important to develop an integrated program that may track the on line characters and narratives used by the people, group and authorities against Dalit and Minorities in India on various social networks. On the other side powerful political parties are using a paid army of people for spreading hate contents and invoke violence in sensitive zones [60] in a planned manner. Recent attacks in different states are notices immediately before or after the election. Henceforth, it is important to involve three bodies in observation cybercrime cells, social media and election commission.

Table IV: Tweets analyzed with specific contents using filter

Category	% of total characters reviewed (excluding special character and space total 1000)	Frequency
Both	0.054	54
Muslim	0.022	22
Political and other social organization	0.017	17
Upper caste	0.015	15
Other	0.015	15

VII. Suggestion and model for further researches

Above sited literature review and narratives suggest to develop an integrated web based real time software that may display the hate content statistics state wise and location wise. Further the unidentified and verified bodies hate contents against Muslims and Dalits must be observed regularly. A cybercrime officer in every state can smell the conspiracy in advance if something objectionable is floating on the web. The data must be divided on the basis of the Social Network or OTT used for spreading the hate contents. The characters and narratives may be classified on the basis of aggressive, moderate or insignificant. The linkage of hate contents and the real violence may help in identifying the actual culprits and their political links. Suppose a political party may found as a real accused than it must be debarred to fight election in that states at least for two consecutive terms.

VIII. Conclusion

The society and marginalized are facing consequences of hate speeches in Delhi Riots, Cow related Violence, Violence against police staff and medical staff, Violence against students and teachers. Many families are displaced or stopped their work after the violence. The violence against sanitation worker is also increasing in the last few years in the various parts of our country. These hate seeds are nurtured and now converted into big trees in the future. There seeds will spread in every state and in every village. If it is not controlled by the police department, government and judiciary at this point now people will start killing weak people in herd or mob. Therefore, it is important to investigate and identify the perpetrators in advance before a large section will lose their lives, employment or valuables. It is a large loss to the image of a nation, governance and economy as a whole. Life of people or security staff cannot be compensated by money or any other non-monetary compensation. Nonetheless, it is national loss and defaming our nation in the world. Incitement to religious and communal hatred has been an issue of serious concern in India as it is heinous crime. The increasing violence against Dalits and Minorities during elections causing communal tensions must be strictly observed and penalized. Hate speech against religious minorities must be avoided or discouraged by all the political social leaders in order to protect every community during sensitive time. Election commission must take a note on such episodes and can bar a person or party found indulged in such incidences or clashes, which have claimed many lives. The paid media is mostly regulated by upper caste majorly by news 18 group and times group. Therefore, strict guidelines must be imposed on paid channels for regulation of hate contents while they use social media platform using their official IDs. Digital contents flashed in the news are manipulated by psychic people and adolescent youth is spreading it at a faster speed. Provoking a young blood for violence is much easier as compared to others. Henceforth, representation of young minds from minority and suppressed can play a vigilant role as a social media volunteers

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

in various groups and communities. Employing youth for surveillance in every area for reporting the online hate contents is urgently required.

As a future direction of work, we propose to develop a framework to report online hate crimes to the cyber crime departments of the country like India which may also identify the concerned law of the land violated to facilitate a cyber officer for registering complaints.

REFERENCES

[1] T. Gelashvili, "Hate speech on social media: Implications of private regulation and governance gaps," 2018.

[2] "Increase in online hate speech leads to more crimes against minorities," https://phys.org/news/2019-10-online-speech-crimes-minorities.html, (Accessed on 08/15/2020).

[3] M. Mirchandani, "Digital hatred, real violence: Majoritarian radicalisation and social media in india," *ORF Occasional Paper*, vol. 167, pp. 1–30, 2018.

[4] "Secretary-general's message to the un core group meeting: Ending hate speech against lgbt people in social and traditional media - freedom of speech vs. license to hate — united nations secretary-general," https://www.un.org/sg/en/content/sg/statement/2019-09-24/secretary-generals-message-the-un-core-group-meeting-ending-hate-speech-against-lgbt-people-social-and-traditional-media-freedom-of-speech-vs-license-hate, (Accessed on 08/15/2020).

[5] "Facebook's hate-speech rules collide with indian politics - wsj," https://www.wsj.com/articles/facebook-hate-speech-india-politics-muslim-hindu-modi-zuckerberg-11597423346, (Accessed on 08/16/2020).

[6] "Voter manipulation on social media now a global problem, report finds — techcrunch," https://techcrunch.com/2019/09/26/voter-manipulation-on-social-media-now-a-global-problem-report-finds/, (Accessed on 08/15/2020).

[7] "Europe's refugee and migrant crisis in 2016. in numbers — World Economic Forum," https://www.weforum.org/agenda/2016/12/europes-refugee-and-migrant-crisis-in-2016-in-numbers, (Accessed on 08/15/2020).

[8] I. P. Karolewski and R. Benedikter, "Europe's refugee and migrant crisis: economic and political ambivalences," *Challenge*, vol. 60, no. 3, pp. 294–320, 2017.

[9] "Responding to 'hate speech' with positive measures: A case study from six EU countries," ARTICLE 19, Tech. Rep., 2018.

[10] "The origins of "hate speech" — the heritage foundation," https://www.heritage.org/civil-society/commentary/the-origins-hate-speech, (Accessed on 08/15/2020).

[11] B. Abraham, "Trends in religion-based hate speech. digital empowerment foundation," 2017.

[12] T. McGonagle *et al.*, "The council of europe against online hate speech: Conundrums and challenges," in *Expert paper. Belgrade: Council of Europe Conference of Ministers responsible for Media and Information Society*, 2013.

[13] "How social media impacts on online racism and hate speech," https://www.rte.ie/brainstorm/2017/1107/918190-how-social-media-impacts-on-online-racism-and-hate-speech/, (Accessed on 08/15/2020).

[14] J. Banks, "Regulating hate speech online," *International Review of Law, Computers & Technology*, vol. 24, no. 3, pp. 233–239, 2010.

[15] M. Maxwell, "Rage and social media: The effect of social media on perceptions of racism, stress appraisal, and anger expression among young african american adults," 2016.

[16] M. Williams, "The connection between online hate speech and real-world hate crime," *Retrieved February*, vol. 11, p. 2020, 2019.

[17] I. Gagliardone, D. Gal, T. Alves, and G. Martinez, *Countering online hate speech*. Unesco Publishing, 2015.

[18] "India leads the world in internet shutdowns - reuters," https://in.reuters.com/article/india-citizenship-internet/india-leads-the-world-in-internet-shutdowns-idINKBN1YO1WR, (Accessed on 08/15/2020).

[19] E. Keen, "Mapping study on projects against hate speech online," *Council of Europe*, 2014.

[20] "218 cases of hate crimes reported in india in 2018, up tops chart: Amnesty — Business Standard News," https://www.business-standard.com/article/pti-stories/218-cases-of-hate-crimes-reported-in-india-in-2018-up-tops-chart-amnesty-119030500956_1.html, (Accessed on 08/15/2020).

[21] D. Belson, "The migration of political internet shutdowns," *Last modified November*, vol. 27, p. 2017, 2017.

[22] Z. Laub, "Hate speech on social media: Global comparisons," *Council on Foreign Relations*, vol. 7, 2019.

[23] "Why it's time to prioritise digital literacy – world wide web foundation," https://webfoundation.org/2017/09/why-its-time-to-prioritise-digital-literacy/, (Accessed on 08/15/2020).

[24] A. Antonio and D. Tuffley, "Digital literacy in the developing world: a gender gap," *The Conversation*, vol. 8, pp. 1–3, 2014.

[25] "65% of hate crimes against dalits: Amnesty - The Hindu," https://www.thehindu.com/news/national/65-of-hate-crimes-against-dalits-amnesty/article26440412.ece, (Accessed on 08/15/2020).

[26] "Hate Crimes in 2018, Reveals 'Halt the Hate' Website — Amnesty India," https://amnesty.org.in/news-update/over-200-alleged-hate-crimes-in-2018-reveals-halt-the-hate-website/, (Accessed on 08/15/2020).

[27] "Hate Speech Explained: A Toolkit," ARTICLE 19, Tech. Rep., 2015.

[28] "Increase in online hate speech leads to more crimes against minorities," https://phys.org/news/2019-10-online-speech-crimes-minorities.html, (Accessed on 08/15/2020).

[29] "The dynamics of hate speech and counter speech in the social media_english," https://cihr.eu/wp-content/uploads/2018/10/The-dynamics-of-hate-speech-and-counter-speech-in-the-social-media_English-1.pdf, (Accessed on 08/15/2020).

[30] "Hate watch: Facebook a favourite of hate mongers? — CJP," https://cjp.org.in/hate-watch-facebook-a-favourite-of-hate-mongers/, (Accessed on 08/15/2020).

[31] "Facebook in india is drowning in anti-muslim hate speech," https://www.vice.com/en_in/article/mb8xxb/facebook-in-india-is-drowning-in-anti-muslim-hate-speech, (Accessed on 08/15/2020).

[32] "Facebook India–Towards A Tipping Point of Violence Caste and Religious Hate Speech," https://www.equalitylabs.org/facebookindiareport, (Accessed on 08/15/2020).

[33] "Propaganda and freedom of the media," https://www.osce.org/files/f/documents/b/3/203926.pdf, (Accessed on 08/15/2020).

[34] "Ohchr — prevention and punishment of the crime of genocide," https://www.ohchr.org/EN/ProfessionalInterest/Pages/CrimeOfGenocide.aspx, (Accessed on 08/15/2020).

[35] M. O'Flaherty, "Freedom of expression: article 19 of the international covenant on civil and political rights and the human rights committee's general comment no 34," *Human Rights Law Review*, vol. 12, no. 4, pp. 627–654, 2012.

[36] A. Jakubowicz *et al.*, "Alt_right white lite: trolling, hate speech and cyber racism on social media," *Cosmopolitan Civil Societies: An Interdisciplinary Journal*, vol. 9, no. 3, p. 41, 2017.

[37] "Monitoring and Analysis of Hateful Language in South Africa Report #5 — PeaceTech Lab," https://www.peacetechlab.org/south-africa-report-5, (Accessed on 08/15/2020).

[38] "2017 Hate Crime Statistics Released — FBI," https://www.fbi.gov/news/stories/2017-hate-crime-statistics-released-111318, (Accessed on 08/15/2020).

[39] P. R. Center, "Global restrictions on religion rise modestly in 2015, reversing downward trend," 2017.

[40] L. Michael and S. O'Curry, "Reports of racism in ireland: 13th & 14th quarterly reports of ireport. ie." 2017.

[41] E. Siapera, E. Moreo, and J. Zhou, "Hate track: Tracking and monitoring online racist speech," Irish Human Rights and Equality Commission, Tech. Rep., 2018.

[42] J. Flatley, "Hate crime, england and wales, 2017/18," *London: Home Office*, 2018.

[43] "Osce - odihr — hate crime reporting," https://hatecrime.osce.org/, (Accessed on 08/15/2020).

[44] "2017 hate crime statistics — fbi," https://ucr.fbi.gov/hate-crime/2017/topic-pages/victims, (Accessed on 08/15/2020).

[45] M. L. Williams, P. Burnap, A. Javed, H. Liu, and S. Ozalp, "Hate in the machine: anti-black and anti-muslim social media posts as predictors of offline racially and religiously aggravated crime," *The British Journal of Criminology*, vol. 60, no. 1, pp. 93–117, 2020.

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.

[46] S. Ozalp, M. L. Williams, P. Burnap, H. Liu, and M. Mostafa, "Antisemitism on twitter: Collective efficacy and the role of community organisations in challenging online hate speech," *Social Media+ Society*, vol. 6, no. 2, p. 2056305120916850, 2020.
[47] L. Andrews, *Facebook, the Media and Democracy: Big Tech, Small State?* Routledge, 2019.
[48] M. Walters, R. Brown, and S. Wiedlitzka, "Causes and motivations of hate crime," *Equality and Human Rights Commission research report*, vol. 102, 2016.
[49] ——, "Preventing hate crime: emerging practices and recommendations for the effective management of criminal justice interventions," 2016.
[50] B. Leber, *Police-reported hate crime in Canada, 2015*. Canadian Centre for Justice Statistics, 2017.
[51] A. Armstrong, "Police-reported hate crime in canada, 2017," *Juristat: Canadian Centre for Justice Statistics*, pp. 1–29, 2019.
[52] "Australian Hate Crime Network - The University of Sydney Law School," https://www.sydney.edu.au/law/our-research/research-centres-and-institutes/australian-hate-crime-network.html, (Accessed on 08/15/2020).
[53] A. Kravchuk, "Hate crimes and incidents in ukraine," https://gay.org.ua/publications/hatecrime2018-e.pdf, (Accessed on 08/15/2020).
[54] "2008 hate crime survey," https://www.humanrightsfirst.org/sites/default/files/FD-081103-hate-crime-survey-2008.pdf, 2008, (Accessed on 08/15/2020).
[55] J. Kotani, "Proceed with caution: Hate speech regulation in japan," *Hastings Const. LQ*, vol. 45, p. 603, 2017.
[56] J. Singh, "Communal violence in muzaffarnagar," *Economical & Political Weekly*, vol. 51, no. 31, pp. 94–101, 2016.
[57] N. Hall, A. Corb, P. Giannasi, and J. Grieve, *The Routledge international handbook on hate crime*. Routledge, 2014.
[58] S. Bhaumik *et al.*, "'look east through northeast': Challenges and prospects for india," Tech. Rep., 2015.
[59] "Bjp mps/mlas have most declared police complaints of hate speech against them —," https://www.indiaspend.com/bjp-mpsmlas-have-most-declared-police-complaints-of-hate-speech-against-them-58167/, (Accessed on 08/15/2020).
[60] "Hate speech: How far can political parties in india be made accountable for their digital propaganda," https://scroll.in/article/921340/how-far-can-political-parties-in-india-be-made-accountable-for-their-digital-propaganda, (Accessed on 08/15/2020).

Authorized licensed use limited to: University of Plymouth. Downloaded on September 29,2024 at 09:53:00 UTC from IEEE Xplore. Restrictions apply.