Countering Online Hate Speech

* Pramod Bhatnagar

ABSTRACT

Latest data released by UK government says hate crime during the year 2016 -17 has increased by 29 per cent, a largest recorded rise in the six years. The Hate Directory (2010) had documented 170 pages of hate content that included websites, blogs, games, radio broadcasts, podcasts and racist friendly web-hosting services. Online Hate speech is an example how technologies with a transformative potential bring with them both opportunities and challenges. The Council of Europe (1997) defines hate speech as any expression of racism, xenophobia, anti-Semitism or aggressive nationalism and ethnocentrism which targets minorities and migrants. In radicalization process people come to believe that violence against others and even oneself is justified to protect their own group. Calvert (2008) says young people are particularity at risk of exposure to online hate as many treat the Internet as their 'home away from home'. Statistics Canada (2009: Juristat) reveals teens comprise the largest group of perpetrators of hate crimes offline in Canada – young males ages 12 to 17. This paper besides analyzing the ugly side of web technology development seeks to study the catalyzing effect various media platforms have had on the rapid growth of hate websites and stories and its impact. The study also analyzes how some ill-intentioned brains exploit technology to manage a safe haven. Global measures to contain the rising menace and their limitations have also come under study.

KEYWORDS: Hate, Web, Facebook, Internet, Discrimination, Communication, Directory

Facebook recently removed Britain First's Facebook page and those of its leaders, Paul Golding and Jayda Fransen, for repeatedly violating rules designed to stop the incitement of hatred against minority groups. A blog post said: "They have repeatedly posted content designed to incite animosity and hatred against minority groups, which disqualifies the pages from our service" (Reuters, 14 March 2018).

Meanwhile, new data released by the Home Office in UK shows reported hate crime during the year 2016 -17 has increased by 29 per cent. Police forces across England and Wales recorded almost 80,400 hate crimes in 2016-2017. This is the largest recorded rise in the six years since records began. Many people do not report hate crime to the police for a number of reasons – so the figures could well be even higher.

The Hate Directory (2010) had documented 170 pages of hate content that included websites, blogs, games, radio broadcasts, podcasts and racist friendly web-hosting services. The Simon Wiesenthal Center’s Digital Terrorism and Hate Project (2016) has so far traced over 14,000 potentially hateful websites, blogs, social networking pages and videos on video-sharing sites like YouTube, in order to help identify such threats.

Hailed as a groundbreaking interactive marketplace of ideas where anyone with the right hardware and software can set up a cyber-stall, the Internet has become an essential means for people to access information and services. Social interaction and communication, membership, using different search engines, openness, and collaboration among users are the fundamental characters of social media networks. However, the Net also offers a host of offensive materials – including hateful content – that attempt to inflame public opinion against certain groups (Eysenbach, 2008).

*Dr. Pramod Bhatnagar, Professor, Amity School of Communication, Amity University Haryana
Hate speech online is an example how technologies with a transformative potential bring with them both opportunities and challenges. U.S.-based Tolerance.org (2011) says that “prejudices are formed by a complex psychological process that begins with attachment to a close circle of acquaintances, or an ‘in-group’, such as a family. Prejudice is often aimed at ‘out- groups’ that are not included in the ‘in-group’ on the basis of certain shared characteristics. Most definitions of hate focus on the ways people are viewed as the ‘Other’.

In national and international legislation, hate speech refers to expressions that advocate incitement to harm, discrimination, hostility or violence based upon the targets being identified with a certain social or demographic group. The Council of Europe (1997) defines hate speech as any expression of racism, xenophobia, anti-Semitism or aggressive nationalism and ethnocentrism which targets minorities and migrants.

While the ill effects and impact of the hate sites loom large, their continuous growth is a matter of concern for us all. Who is responsible for their proliferation; what has been done to contain their monstrous effect; does the web technology encourage them or provides a convenient home to grow safe; what does the future hold and shall we be able to successfully combat the menace, are some of the questions that strike any sane mind.

This paper besides analyzing the seamy side of web technology development seeks to study the catalyzing effect various media platforms have had on the rapid growth of hate websites and stories and its impact. The study also analyzes how some ill intentioned brains exploit technology to manage a safe haven. Global measures to contain the rising menace and their limitations have also come under study.

Raymond A. Franklin, author of the Hate Directory (2010) defines hate groups who “advocate violence against, separation from, defamation of, deception about, or hostility towards others based on race, religion, ethnicity, gender, or sexual orientation”.

Hate speech relies on tensions, which it seeks to re-produce and amplify. The concept may extend also to expressions that foster a climate of prejudice and intolerance on the assumption that this may fuel targeted discrimination, hostility and violent attacks.

Karim (2003) saysit is usual that people transfer their fears and hatred to the ‘other’; the group is viewed less than human. Hate speech is always an attempt to marginalize and discriminate against particular individuals, groups and disadvantaged groups such as minorities. It is a tool to dehumanize and defame, and discriminate against target groups (Cohen Almagor, 2011).

In a wide spectrum, Benesch (2014) explained hate speech is almost in every community. Hence, she pronounced that hate speech is a catalyzing tool that magnifies the violence and hostilities over some particular individuals or groups.

Calvert (2008) says young people are particularity at risk of exposure to online hate as many treat the Internet as their ‘home away from home’. Statistics Canada (2009: Juristat) reveals teens comprise the largest group of perpetrators of hate crimes offline in Canada –young males ages 12 to 17. The report warns given the particular vulnerability of youth, it is vitally important to engage them as early as possible in discussions about online cultures of hate and hateful content.

In radicalization process people come to believe that violence against others and even oneself is justified to protect their own group. According to McCauley and Moskalenko (2008) the base is made up of sympathizers who support and share the group’s ideals though not actively involved. Next come that identify themselves strongly with the group and participate in its everyday activities. Finally, activists push the group towards more radical positions and extreme actions.

Researchers say people have ability to justify even actions they would normally consider wrong. The effect is particularly powerful online, where consequences are less apparent.
Amon (2010) says hate groups function as substitute families for vulnerable, disenfranchised youth. Members of hate groups such as skinheads act as mentors to youth frequently coming from family backgrounds with a lack of love and affection and traditional moral or religious values.

The Alberta Hate and Bias Crime and Incidents Committee report (2007) reveals that the average hate group member’s opinion becomes more extreme over time. People identify more closely with a group if the group appears to be isolated or under external threat. It is common for soldiers to consciously sacrifice their lives to save other members of their platoon. Portraying opposing groups as inhuman makes it easier to justify any action against them. In World War II the Japanese were portrayed in a heavily caricatured style in American propaganda and roughly half of American soldiers favored exterminating the Japanese nation after the war was over. In Canada, hate groups frequently target Sikh, Muslim and Jewish groups based on their religious affiliation.

McCauley & Moskalenko (2008) say identity seeking is a natural part of adolescence. Taken to its extreme, this can provide a toe-hold for hate mongers. “Anomie” is a state of mind in which family or cultural values appear worthless. Youth with anomie will seek a group or cause that gives those values, an identity and a surrogate family.

According to Turpin-Petrosino (2002) when economic and social problems – or the perception of such problems – are combined with a threat to one’s identity, people become vulnerable to messages of hate.

Hate groups of all kinds are highly skilled at identifying those youth most likely to be vulnerable to their message. Positive and supportive attitudes toward hate groups are influenced by word-of-mouth recommendations from personal acquaintances; criminal hateful actions are also reinforced through word-of-mouth, aided by new media such as email and social networking site (Ibid.)

HATEFUL IDEOLOGIES

Fundamental to all ideologies of hate is the idea of a target group designated as ‘Other’ (Meddaugh, 2009). The Other is portrayed as both inferior and threatening. A 2002 study of racist chat rooms found that participants responded less severely to the possibility of material threats and more viscerally to scenarios that imagined the ‘Other’ harming the integrity or purity of the group, particularly through interracial sex or marriage (Glaser, 2002).

According to Kim (2006) an essential element of hate ideology is the notion that the group has fallen from its once-glorious past that is portrayed as the fault of either the Other or of members of the group whom ‘Other’ fooled or subverted. Victimhood is central to ideologies of hate. A widely-reprinted article by David Duke, a former Grand Wizard of the Knights of the Ku Klux Klan, claimed that Whites were the true victims of slavery.

This idea of a special sanction manifests itself in two ways. The idea of an upcoming final conflict tells the group will defeat its enemies and retake its rightful place. The special status of the group elevates its defeats to the level of martyrdom. This constantly recurring theme in hate material is an effective tool to radicalize supporters (Amon, 2010).

FALL OUT OF HATE

Hate speech seriously violates the human rights of individuals and humiliates their stature, which then leads the victim to varied psychological and troubled situations. Exposure to online bigotry and hate can have serious effects like making people feel uncomfortable or unwelcome and can cause stress, anxiety and depression. Besides, it also terrorizes, isolates and degrades the self-confidence of individuals (Leets, 2002, Downs and Cowan, 2012). Nielsen (2002) and Parekh (2006) opined hate speech that targets group can also lead to creation of discrimination and fear among the members of the victimized group that also potentially leads to withdrawal.
of participation and expressing the ideas in their community. Hence, the fear and silence of the victims pushed them to take aggressive actions and violations.

Nemes (2002) laments social media platforms are becoming safe haven for disseminating hate speeches. Mainly, after the coming of different mobile technologies and increased use of social media outlets the pace of hate speech is aggravating.

A study by Tynes et al. (2008) found that one in five participants in moderated chat rooms was exposed to negative remarks about a racial or ethnic group.

**FACILITATING HATE GROWTH**

An unprecedented global reach and anonymity have made Internet a convenient tool for extremists. They can now do a safe packaging of hatred, raise funds, and recruit members. As the technology improves, so does the quantity and sophistication of extremist websites.

The Simon Wiesenthal Center’s Digital Hate and Terrorism project (2011) has identified over 14,000 problematic websites, forums, blogs, and social media postings. Some hate groups have developed their own versions such as New Saxon, “a Social Networking site for people of European descent” produced by the American Neo-Nazi group National Socialist Movement. On Facebook some extremists groups include Al Shabab Mujahideen, Storm front, National Socialist Life, Libertarian National Social Movement, Aryan Guard, FARC, Hamas, Hezbollah, Faloja Forum, Support Taliban and scores of anti-Israel sites. According to the Simon Wiesenthal Center the spread of online hate has so far outpaced efforts to remove them. Common are “cultures of hatred”: communities in which racism, misogyny and other prejudices are normalized.

Sites which host user-created content, such as eBaum’s World, New grounds and 4chan, make a virtue out of being offensive. Here this free-floating hate develops into a widespread animosity towards Asians in general and Chinese in particular. That even led to the production of films within the game that mocked Chinese players by drawing on classic Asian stereotypes such as the “China doll” character and the idea that Chinese kidnap pet cats and dogs to eat (though there are no pets within the game) as well as organized attacks within the game on characters perceived to be “Chinese.”

Hate speech online can be itinerant. Even when content is removed, it may find expression elsewhere, possibly on the same platform under a different name or on different online spaces. If a website is shut down, it can quickly reopen using a web-hosting service with less stringent regulations or via the reallocation to a country with laws imposing higher threshold for hate speech. (Citron & Norton 2017).

**MEDIA PLATFORMS AND HATE SITES**

**WEBSITES:** With blogging platforms easily available, hate sites have proliferated. Some sophisticated ones mimic popular commercial websites and many offer audiovisual material and discussion forums with professional-looking layout and graphics. Some like the Al-Fateh specifically appeal to youth and children, with content that promotes suicide terrorism.

**SOCIAL NETWORKING:** Online social networks encourage group interaction and strengthen members’ connections. In Media Smarts’ 2001 study young North Americans in a Wired World seven per cent of students surveyed reported having encountered hate material by accident, compared to five percent who said they had found it on purpose. Sites such as Facebook and Twitter are used to join and recruit group and help youth interested in hate find friends and mentors important in the radicalization process.

**COMMON TECHNIQUES**
McNamee et al. (2010) say a variety of techniques work to reinforce hate group identity, reduce external threats and recruit new members. Many hate sites call supporters to educate their friends, families and communities about the “real” truth.

Hate sites adopt many of their markers of credibility – quoting from old editions of the Encyclopedia Britannica, or selectively citing articles from reputable sources such as the Wall Street Journal. Gerstenfield and Grant (2003) say misinformation through denial, pseudo-science, hero narratives and promoting false sense of nationalism are other ways to promote hate campaigns. An appeal to religion can be an even stronger tool for building solidarity.

BRAND MAKING

The hate groups continue using symbols such as the Nazi swastika to ‘brand’ its message. The Anti-Defamation League (2001) warns that hate symbols are more than mere signs: “These symbols are meant to inspire a sense of fear and insecurity and give haters a sense of power and belonging, and a quick way of identifying with others who share their ideology”. Dehumanization is the basic mechanism of radicalization to successfully promote and justify annihilation of a particular group.

Overt hate sites actively promote hatred towards other groups while showcasing racist propaganda or offering hate-based communities online. Cloaked hate websites perpetuate hatred through more implicit and deceptive messages, masquerading as legitimate sources of opinion or information. Young people are vulnerable as research shows many lack the ability to critically evaluate cloaked hate sites (Daniels, 2008).

Gagliardon (2014) says Social media outlets do expect users to follow self-regulation while inviting them to post images, videos, audios and other contents. Failure in self-regulation by users leads to a media outlet becoming a channel for hate and intolerance (Mutsvairo, 2016). Commaerts (2009) and Harris et al. (2009) feel social media is relatively free from any censorship and users do whatever they want in online platforms. According to Henry (2009) and Mahoney (2013) social media allows users to write detailed messages, different options such as re-tweeting, commenting, sharing and linking contents to other platforms. This way, their messages reach a broader audience within a short time. One big issue is anonymity. In Facebook and Twitter it is possible even to create account in others’ names or they can use fake account with a fake name (Eichhorn, 2001). Anonymity on the social media outlets is contributing to the extremists disseminating contaminated messages through hate speech (Nemes, 2002).

CONTAINING THE MENACE

According to Eltis (2012), Perry and Olsson, (2009) some social media, though they have rules and regulations to control hate speeches and other acts, are not regulating the norms effectively. There is also little serious editorial policy and gate keeping.

While people world over are aware of the harmful impact of online spread of hate and polluting of minds, world governments seem to have done little to combat the onslaught. The Internet's speed and reach also make it difficult for governments to enforce national legislation in the virtual world. Hope (2006) says legislation so far has failed in adequately capturing the broad scope and complicated, disputed nature of online hate. According to Fong et al. (2005) current protocols for dealing with online hate have proven inadequate at managing hateful content and providing educational opportunities. Guichard (2009) too, feels criminal legislation and formal policies have had limited success in addressing the complex issues.

Citizenship education focuses on preparing individuals to be informed and responsible citizens through the study of rights, freedoms, and responsibilities and has been variously employed in societies emerging from violent conflict. Subrahmanyam (2008) says while preserving the ability of modern technology for education and social development, it is important to have policy to counter harmful, negative and distracting uses of electronic media. Andrea Slane (2007) identifies sites like Franklin’s Hate Directory that hosts a hate hotline.
The U.S.-based Simon Wiesenthal Center and the Anti-Defamation League recommend that Internet users alert them to any online hate, so they can try to get the offending material removed.

Another way is to report it to the service or site that hosts the content. Many Internet Service Providers remove hate content hosted on their servers once they are made aware of it. But many hate groups in the United States operate where there are almost no legal limitations on hate speech. The 1948 Universal Declaration of Human Rights (UDHR) contains the right to equal protection under the law in Article 7: "All are entitled to equal protection against any discrimination in violation of this Declaration and against any incitement to such discrimination". The UDHR also states that everyone has the right to freedom of expression, which includes "freedom to hold opinions without interference and to seek, receive and impart information and ideas through any media and regardless of frontiers".

Though UDHR was decisive in setting a framework and agenda for human rights protection, the declaration is non-binding.

The International Covenant on Civil and Political Rights (ICCPR) is important and comprehensive when addressing hate speech and contains the right to freedom of expression in Article 19 and the prohibition of advocacy to hatred that constitutes incitement to discrimination, hostility or violence in Article 20. In the ICCPR, the right to freedom of expression is not an absolute right. It can legitimately be limited by states under restricted circumstances.

ICERD

The International Convention on the Elimination of All Forms of Racial Discrimination (ICERD) of 1969 has also implications for conceptualizing forms of hate speech. It asserts in Article 4, paragraph (a), that state parties: "Shall declare as an offence punishable by law all dissemination of ideas based on racial superiority or hatred, incitement to racial discrimination, as well as all acts of violence or incitement to such acts against any race or group of persons of another color or ethnic origin, and also the provision of any assistance to racist activities, including the financing thereof . . . ."

The Committee on the Elimination of Racial Discrimination has actively addressed hate speech in its General Recommendation 29, in which the Committee recommends state parties to: "(r) Take measures against any dissemination of ideas of caste superiority and inferiority or which attempt to justify violence, hatred or discrimination against descent-based communities; (s) Take strict measures against any incitement to discrimination or violence against the communities, including through the Internet; (t) Take measures to raise awareness among media professionals of the nature and incidence of descent-based discrimination;"

The American Convention on Human Rights describes limitations on freedom of expression in a manner similar to the ICCPR in Article 19 (3).

PRIVATE SPACES

Schomerus (2012) states various online interactions on social networking platforms transcend national jurisdictions and have also defined hate speech and have measures to respond to it. A user who violates the terms of service, the content posted may be removed, or its access may be restricted to be viewed only by a certain category of users.

SOCIAL MEDIA RESPONSES

Yahoo!’s terms of service prohibit the posting of "content that is unlawful, harmful, threatening, abusive, harassing, tortuous, defamatory, vulgar, obscene, libelous, invasive of another's privacy, hateful, or racially, ethnically or otherwise objectionable". Twitter alerts its users that they "may be exposed to content that might be offensive, harmful, inaccurate or otherwise inappropriate, or in some cases, postings that have been
mislabeled or are otherwise deceptive”. YouTube’s terms of service read: "We encourage free speech and defend everyone's right to express unpopular points of view. But we do not permit hate speech: speech which attacks or deems a group based on race or ethnic origin, religion, disability gender, age, veteran status and sexual orientation/gender identity. Facebook elaborates "Facebook removes hate speech, which includes content that directly attacks people based on their: race, ethnicity, national origin, religious affiliation, sexual orientation, sex, gender or gender identity, or serious disabilities or diseases". Microsoft policy for mobile phones prohibits applications that "contain any content that advocates discrimination, hatred, or violence based on considerations of race, ethnicity, national origin, language, gender, age, disability, religion, sexual orientation, status as a veteran, or membership in any other social group."

Online shaming is a form of Internet vigilantism in which targets are publicly humiliated using technology like social and new media. Proponents of shaming see it as a form of online participation that allows hacktivists and cyber-dissidents to right injustices. Critics see it as a tool that encourages online mobs to destroy the reputation and careers of people or organizations that made perceived slights.

Online shaming frequently involves the publication of private information on the Internet (called doxing), which can frequently lead to hate messages and death threats being used to intimidate that person.

Daniels (2008) recommends digital literacy skills development that enables young people to critically deconstruct images produced by hateful media and provides an effective way of understanding multiple perspectives, in turn reducing racism, sexism and homophobia.

One most crucial digital literacy skill for combating hate is to know how to verify sources. Flanagin (2010) opines hate groups put considerable effort into faking their sites legitimate. These include a dot-org Web address, quotes and citations from other sources, claims of expertise like “doctor” and “institutes”, and an appealing and professional design. Teaching youth to think critically about the media they consume will prepare them to distinguish between overt and cloaked hate.

(http://mitpress.mit.edu/books/full_pdfs/Kids_and_Credibility.pdf)

CONCLUSION

The issue of addressing social media hate speech is a pressing concern. Educating, creating awareness, discouraging hate speech intolerance, punishing the most extreme and dangerous forms of hate speech in front of law are the most important measures. This must be done without compromising the fundamental human rights to free speech (Cohen Almagor, 2011).

Banks (2010) says the internet gives an opportunity of anonymity and mobility which can enable the extremist groups to disseminate hate speech and this becomes abstract and is beyond the realms of national and international law enforcement. Banks holds as these technologies create problem, they must create solution too. Banks believes the integration of different systems such as law, technology, education, awareness and guidance can immensely contribute to the reduction of dissemination of social media hate speech without affecting the freedom of expression.

Tsesis (2001) recommended that the sites with hate speech contents must be filtered and blocked. Besides, the regulations should be implemented in a highly serious manner. Governments should not allow the users of internet to spread hateful contents on social media which can potentially foment intolerance and discrimination.

Gagliardone (2015) is very skeptical about the legal response to hate speech. Hate speech laws might simply enforce the given natural norms of a decent society since laws always enforce the customs of the dominant groups that control the content of the land.
Social media hate speech being a comparatively recent issue, there is some vagueness about the enforcement of legal measures against it. A recent report by US based Hewlett Foundation says various studies about the “extent to which disinformation shared on social media has any effect on citizens’ political beliefs,” have conflicting findings as the studies use different definitions of what misinformation and polarization are. It is also not clear how “digital media can contribute to polarization and disinformation in more unstable, hybrid regimes and non-democratic regimes. Also “sometimes the differences between rumors, false information, misleading information, and hyper-partisan information are blurry”. The report recommends more research into how exposure to disinformation affects people’s viewpoints on various issues, their overall interest in politics, and their trust in institutions (Laura Hazard Owen, 21 March 2018).

REFERENCES


**REFERENCE REPORTS**


40) Article 19 (3) if the ICCPR.

41) Article 54 of the Charter of Fundamental Rights of the European Union.


43) Committee on the Elimination of Racial Discrimination, General Recommendation 29

44) Hate Base, Hate speech statistics, http://www.hatebase.org/popular

45) Human Rights Committee. General Comment no. 11, Article 20: Prohibition of Propaganda for War and Inciting National, Racial or Religious Hatred


48) Interview: Andre Oboler, CEO, Online Hate Prevention Institute, 31 October 2014.

49) Interview: Drew Boyd, Director of Operations, the Sentinel Project for Genocide Prevention, 24 October 2014.


52) Interview: Laura Geraghty, No Hate Speech Movement, 25 November 2014.