Journal of Information Systems & Information Technology (JISIT) *Vol. 4 No. 1 2019*
ISSN: 2478-0677 *1-9*

A Measurement Study on Racist Hate Speech in Twitter using Tweet Binder

Firthows Hassan Ahamed Shibly

Department of Arabic Language,
Faculty of Islamic Studies & Arabic Language,
South eastern University of Sri Lanka.
shiblyfh@seu.ac.lk

Abstract. The role of social media is in our daily lives is immense. It has many positive sides as well as negative sides especially in the society and human behaviors. Some people are trying to use the social media for tarnishing the routine setup of the society. Some people use these websites for anti-social behaviors including cyber stalking, cyber bullying, trolling, harassments and hate speech. At present, social media websites have started creating serious efforts to control racist hate speech. But, most of them are still facing challenges to come up with an efficient solution. The aim of this research is to explore and measure the racism hate speeches in Twitter. Predictive research method of quantitative studies was applied to carry out this research. Since it is a technological based research, Tweet Binder analytical tool was used to analyze the data. For this research work, the researcher use the dataset for racist hate speech distributed via data world which consists 517 racist hate speeches. Simple random sampling method was used to test the data. As a results, it found that all formats of tweets including text, replies, retweet, pictures and links have most number of racist hate speeches. Especially replies and retweets have highest number of hate speeches. It is highly recommended to Twitter and other social media websites to implement strict policies and mechanisms to control hate speeches to control them to create a peaceful social media environment.

Keywords: Hate Speech, Racism, Social Media, Twitter

1 Introduction

Social media and its social networking websites are very familiar names to almost all people around the world and most of us are using it for several purposes. Social networking websites have done a paradigm shift in many fields. The role of social media on education, culture, self-development, language, relationships and exchange information are immense now a days. Anyway it has many negative sides and budding issues especially in the society and human behaviors. Some people are trying to use the social media for tarnishing the routine setup of the society. Some people use these websites for anti-social behaviors including cyber stalking, cyber bullying, trolling, harassments and hate speech.

The exponential growth of social media such as Twitter and community forums has revolutionized communication and content publishing, but is also increasingly exploited for the propagation of hate speech and the organization of hate-based activities [1]. The term 'hate speech' was formally defined as 'any communication that disparages a person or a group on the basis of some characteristics (to be referred to as types of hate or hate classes) such as race, colour, ethnicity, gender, sexual orientation, nationality, race, or other characteristics' [2].

Research on information security and safety movements in social media has grown continually in the last decade. A particularly main research work is detecting and preventing the use of abusive language in social networks and other websites. A number of recent studies have been published on this issue on identifying cyber-bullying, the detection of hate speech [3] which was the topic of a recent survey [4], and the detection of racism in user generated content.

Motivated and encouraged by these explanations, this research makes two major contributions to the research of social media hate speech measurement. First, the researcher conduct a data collection to measure the racist hate speech on the social media, in order to identify the most common racism hate speeches. Second, the researcher propose proper mechanisms and techniques to reduce or stop such hate speeches in social media to ensure a secure usage of social networking sites.

With this background, it is not surprising that most existing efforts are motivated by the impulse to detect and eliminate hateful messages or hate speech [5]. These efforts mostly focus on specific manifestations of hate, like racism [6]. While these activities are very important, they do not attempt to provide a big picture of the problem of hate speech in the current popular social media systems. Specifically providing a broad understanding about the root causes of online hate speech was not main focus of these prior works. Consequently, these prior works also refrain from suggesting broad techniques to deal with the generic offline hate underlying online hate speech [7].

In this research, the researcher take an initial step towards better understanding about the racism hate speech by measuring its usage in social media in order to control it by social media developers to make sure a peaceful media and to avoid unnecessary issues which may lead cyber war among communities, people, users and other stakeholders.

2 Problem Statement

Hate speech is available in several communities for a very long time. In democratic societies, some people trust that racism views should be suppressed today. The problem arise when one person or a group of people use some words that they think are protected racism speech and some people listen and understand the same words believes it is hate speech.

Recent work has highlighted the repercussions of online hate. The relationship between hate speech and violence has been evidenced in history. Hate speech can be a key factor for some serious problems among the communities. More importantly, racism hate speeches are quit dangerous and quickly heat up people towards disasters. Hate crimes are known to inflict harm to society by instilling fear not only in the victim, but in his/her broader community [8].

At present, social media websites have started creating serious efforts to control racist hate speech. But, most of them are still facing challenges to come up with an efficient solution. Some developers are far ahead with their techniques, but still no one has found the concrete solution that will stop racist hate speech on social media.

It has been found that existing solutions for measuring and detecting racism hate speech are not effective. Therefore there is a need for new methods, which would do the work more effectively and efficiently.

The problem in this work can be described as follows:
“What are the common racism hate speeches in social media and how to effectively measure the racism hate speech to control them for creating a peaceful cyber environment?”

3 Objectives

The aim of this research is to explore and measure the racism hate speeches to propose a controlling mechanism to stop such hate speeches. Effective methods to be found in the area of hate speech since it is an emerging topic in social and professional issues of Information Technology. The researcher include an examination of racism hateful speech in Twitter which is a popular social networking website in the world and most hate speeches have been tweeted recent past.

4 Literature review

A number of researchers have examined the detection of flames and virulent messages in social media as well as the spread of hateful messages in the dark web forums [9]. Raymond A. Franklin, author of the Hate Directory 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". It is very important to find the classifications of hate speech as well. Karim says it 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 [10].

In 1997, the use of machine learning was proposed to detect classes of abusive messages. Cyberbullying has been studied on numerous social media platforms, e.g., Twitter [3] and YouTube. Other work has focused on detecting personal insults and offensive language [3].

A recent study, supported by UNESCO, reviews the growing problem of online hate speech with the advent of internet from a legal and social stand point.They pointed out that platforms like Facebook and Twitter have primarily adopted only a reactive approach to deal with hate speech reported by their users ,but they could do much more. These platforms have access to a tremendous amount of data that can be correlated, analyzed, and combined with real life events that would allow more nuanced understanding of the dynamics characterizing hate speech online".

Based on these literatures, it can be identified that there are several researches have been done and still many researches are ongoing to protect social media users. While racism hate speeches are not new to the world but, hate speech measurement is a recent area. Measuring hate speech has become an important part for analyzing and controlling before any serious incident happen. Manually sorting these text on the social media is seen as a massive work and difficult to scalable at all.

5 Methodology

Predictive research method of quantitative studies has been applied to carry out this research. Since it is a technological based research, predictive web analytics was used to execute this research properly. Web analytics is the data science practiced to collect, measure, analyze, and report on web data to better understand and optimize people's online user experience and which is comprised of website content quality and usability. Predictive analytics involves extracting data from existing data sets with the goal of identifying trends

and patterns. These trends and patterns are then used to predict future outcomes and trends [11]. The development of predictive web analytics computes statistical probabilities of upcoming online events.

6 Data

For this research work, the researcher use the dataset for hate speech distributed via data world which were uploaded by Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. The dataset features 517 racist hate speeches. To prepare the data set, the above authors begin with a hate speech lexicon containing words and phrases identified by internet users as hate speech, compiled by Hatebase.org. Using the Twitter API they searched for tweets containing terms from the lexicon, resulting in a sample of tweets from 33,458 Twitter users. They extracted the time-line for each user, resulting in a set of 85.4 million tweets. From this corpus they then took a random sample of 25000 tweets containing terms from the lexicon and had them manually coded by Crowd Flower (CF) workers. Therefore the dataset is highly validated. A simple random sample method was used to collect the data from the dataset. Simple random sampling permits the sampling error to be measured and decreases selection bias. An important advantage is that it is the best straightforward technique of probability sampling in sample selection methods.

As per the sample size calculation given below, 60 samples were selected for analyzing the data. The sample calculation was done with

$$x=Z(^{c}/_{100})^{2}r(100-r)n=^{Nx}/_{((N-1)E^{2}+x)}E=\mathrm{Sqrt}[^{(N-n)x}/_{n(N-1)}]$$

From the sample data, only 23 hate speech words were used more than 500 total tweets in the research period and those 23 words have been exhibited in results section.

7 Data Analysis

There are many analyzing tools available to measure the twitter tweets. The Researcher use tweet binder web analytic tool to measure the sample data to find the history of such hate speeches to fulfill the research problem and objectives. Table1 is shown the different analyzing tools and its features to justify the tweet binder as our data analysis tool.

Table 1 : Web analytic *tools of twitter*

Name of the web analytic tool	Advantages	Disadvantages
Twitter Analytics	The site is divided accordingly into 3 different categories: Tweets, followers and Twitter Cards.	It provides personal tweet analysis only.
Hootsuite	With a Hootsuite account, you can launch marketing campaigns, schedule posts in advance, identify and grow audiences on Twitter, create custom Twitter reports, track hashtags, mentions, Twitter lists and much more.	It's a freemium tool, meaning that you can use the free plan if you have fewer than five social profiles to manage.
Buffer	Users who tend to Tweet in short bursts no longer have to miss out on valuable followers with this browser-based app.	It provides personal tweet analysis only.

Foller.me	Foller.me is a Twitter analytics application that gives us rich insights about any public Twitter profile.	It is a beta version.
Tweet Binder	It provides many features including Twitter analytic report, Instagram analytic reports, solutions for event an summits and custom projects.	Free version is limited to 7 days analysis.
Key Hole	Keyhole is an analytics-only service that is great for tracking Tweets in real-time.	It is not a free version.

8 Results & Findings

As a result of racist hate speech in twitter, tweet binder web analytic tool helps us to produce total tweets, text tweets, replies, retweets and links/pictures.

8.1 Total tweets

According to the Figure 1, there are eleven speeches out of twenty four data from the dataset. The word Redneck was tweeted highest number of times during the research period. It describes Americans who live in rural areas.

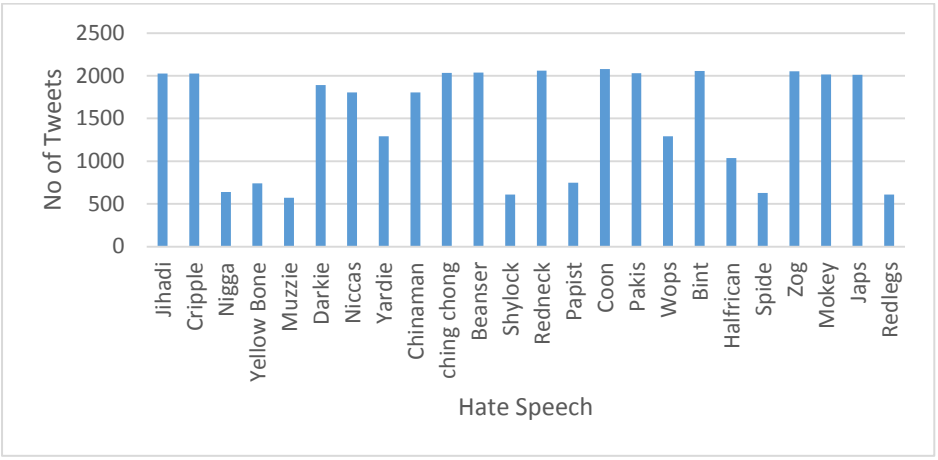


Fig. 1. Total Tweets

8.2 Text Tweets

Based on the Figure 2, the research analyzed the total number of text tweets of hate speeches and found that “Niccas” was the most number of tweets during the study period. “Nicca” is used in Buddhism at large. It is noted that “Beaner” and “Pakis” are next to “Niccas” in text tweets.

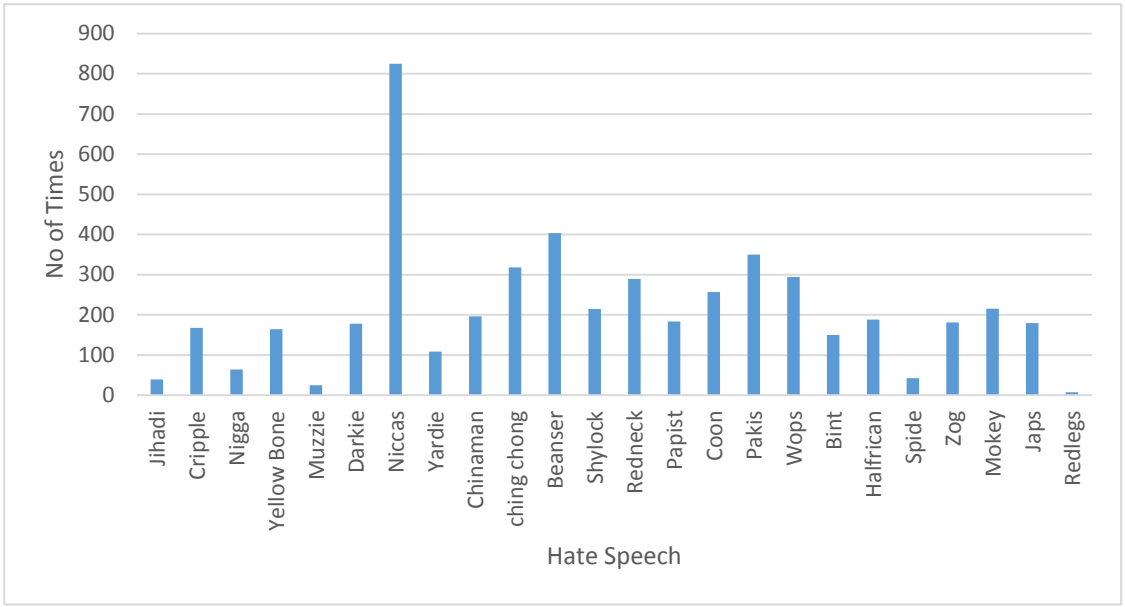
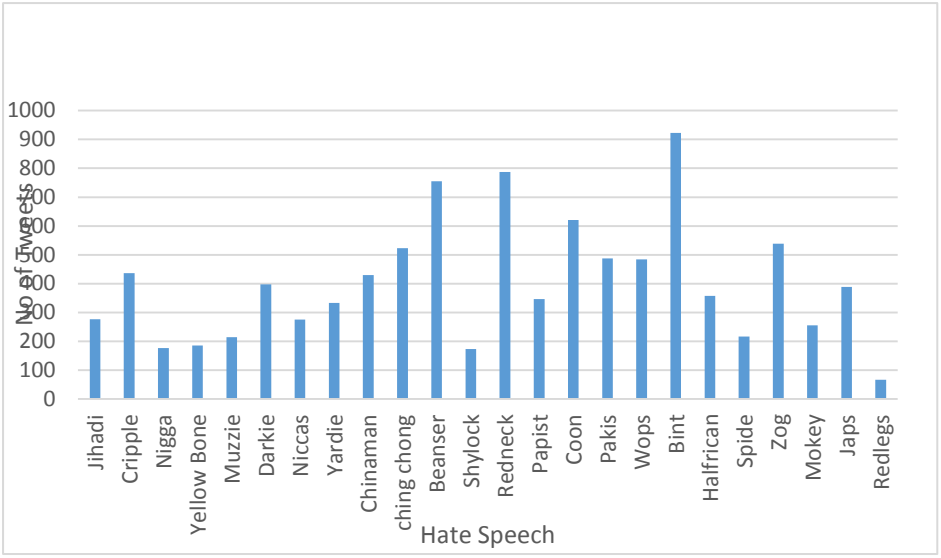


Fig. 2. Text Tweets

8.3 Replies

According to Figure 3, we can notice that "Bint" has got highest number of replies and



"redneck" and "Beaner" next to it.

Fig. 3. Replies

8.4 Retweets

According to figure 4, we can find that “Jihadi” has got most number of retweets during the research period. “Japs” and “darkies” are next to it

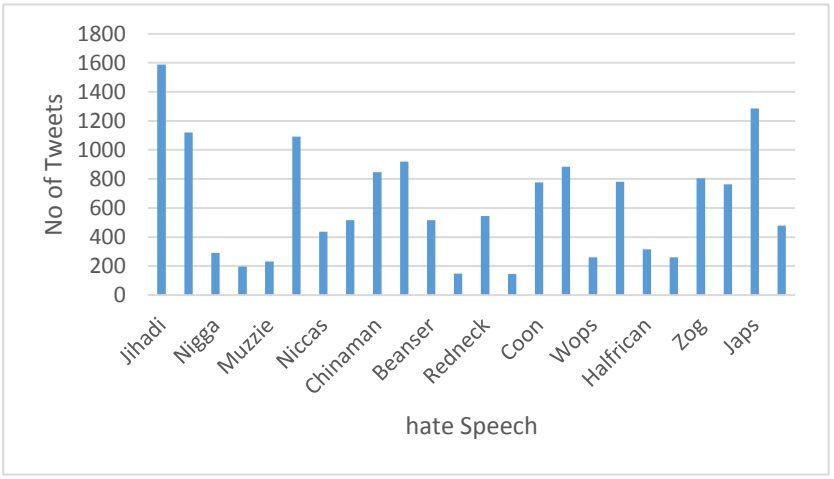


Fig. 4. Retweets

8.5 Links/Pics

It is important to study about sharing the external links and pictures in twitter to measure the trend and records. According to figure 5, “Mokey” has been used highest number of times as a link and/or picture. It is noted that most of the hate speech were used less than five hundred times in this category. There may be an issue to analyze pictures that difficult to measure the words or hate speeches which are in picture formats. But, proper mechanism should be found to analyze it accurately to control hate speech in picture formats.

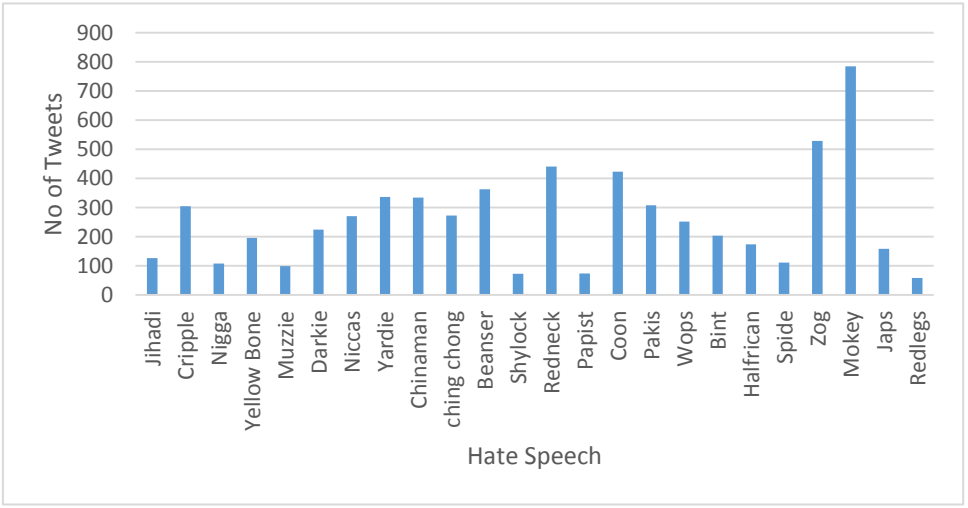


Fig. 5. Links/Pics

8.6 Hate Speech in twitter

According to figure 6, we can find the overall analysis of racist hate speech in twitter from the sample dataset during the research period. Text tweets and links/pics has been used lesser number in tweets when we compare them with replies and retweets. Twitter hate speech policies should focus on replies and retweets to control it.

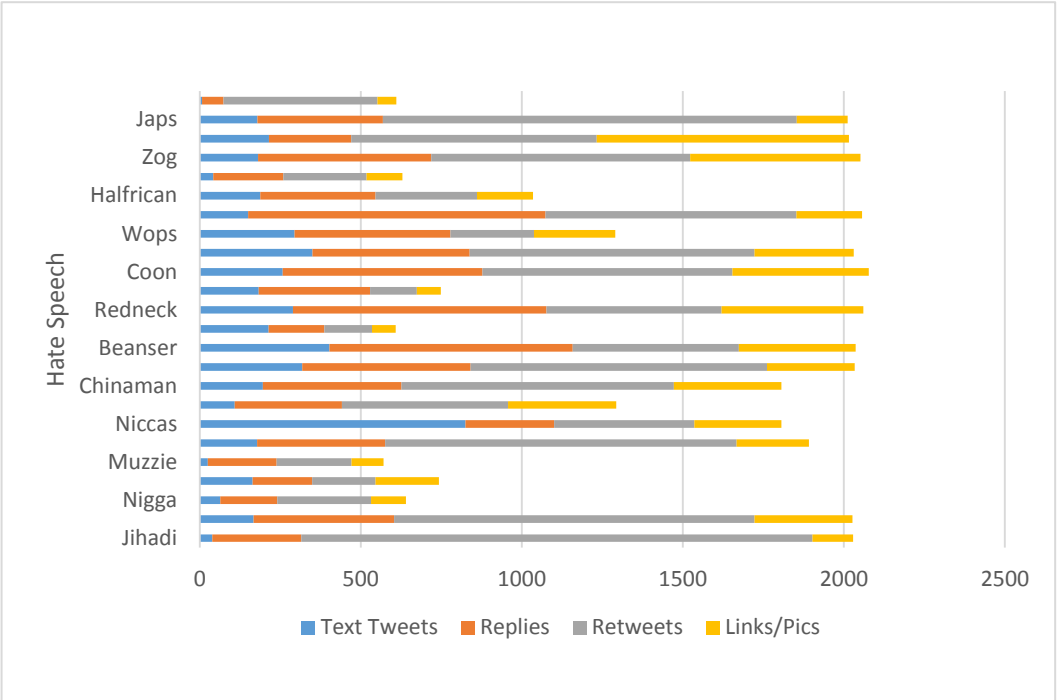


Fig. 6. Hate Speech in Twitter

9 Conclusions & Recommendations

In this study, the researcher analyzed the racist hate speech in twitter by using tweet binder web analytic tool. While this study delivers only a glimpse of the racist hate speech to measure the usage of hate speeches and suggests controlling methods to twitter developers and policy makers. Users also should get the idea about the seriousness of hate speeches and aware about to report such speeches in social media.

Results of this study elaborates that racist hate speeches are simply posted by users without any barriers. Most of the racist hate speeches collected by crowd flower are still posting by users. Those hate speeches didn't target a particular citizen or community people. It attacks all pacific regions and religions specially Muslims, Christians and Buddhists.

Our results also demonstrate how hate speech can be used in different modes: it can be directly posted to a person or group targeted by text, replies, retweets, links or picture formats. It is important to measure more carefully the people who use hate speech, focusing both on their individual characteristics and motivations and on the social structures they are embedded in [12]. It is also revealed that text, links and picture based are being used less

than replies and retweets. It is a must that detecting and controlling techniques should be enforced by developers.

10 Limitations & Future Researches

This research's results and findings should be interpreted with carefulness due to its limitations. First, the web analytic tool used in this research is a limited version. It provided seven day reports and the data were tested from 31st October to 7th November. But, It helped the researcher to find out important decisions and recommendations for the stakeholders of social media.

Second, this research used the data set which was available in data world which were uploaded by Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber [22]. The researcher only used the particular data set so the results and findings are based on them only.

In a future researchers, there are so many works to be done. The auto detection methods can be studied to recommend the social media top officials to control the racist hate speeches. Also an efficient reporting techniques should be studied. Such techniques may reduce the problems created by hate speeches. Auto delete option also will be a useful study once we identified the hate speech data dictionary.

11 References

[1] Pinkesh Badjatiya, Shashank Gupta, Manish Gupta, Vasudeva Varma, ACM WWW'17 Companion, Perth, Western Australia, Apr 2017.

[2] John T. Nockleby, Hate Speech. In Encyclopedia of the American Constitution (2nd ed., edited by Leonard W. Levy, Kenneth L. Karst et al., New York: Macmillan, 2000), pp. 1277-1279.

[3] P. Burnap and M. Williams, "Cyber Hate Speech on Twitter: An Application of Machine Classification and Statistical Modeling for Policy and Decision Making", Policy & Internet, vol. 7, no. 2, pp. 223-242, 2015.

[4] Anna Schmidt and Michael Wiegand., A Survey on Hate Speech Detection using Natural Language Processing. In Proceedings of the EACL-Workshop on Natural Language Processing for Social Media (SocialNLP). Valencia, Spain, 2017. Pages 1–10.

[5] Agarwal.S, N. Mittal, and A. Sureka, "A scientometric analysis of 9 ACM SIGWEB cooperating conferences," ACM SIGWEB Newsletter, no. Autumn, pp. 1–15, 2016.

[6] Chaudhry, I. "'Not So Black and White,'" Cultural Studies ↔ Critical Methodologies, vol. 16, no. 3, pp. 296–304, 2016.

[7] M. Mondal, L. A. Silva, D. Correa, and F. Benevenuto, "Characterizing usage of explicit hate expressions in social media," New Review of Hypermedia and Multimedia, vol. 24, no. 2, pp. 110–130, 2018.

[8] Craig, K. M, Retaliation, fear, or rage: An investigation of African American and White reactions to racist hate crimes. Journal of Interpersonal Violence, 1999, 138–151.

[9] 9Abbasi. A, Affect intensity analysis of Dark Web forums", in 5th IEEE International Conference on Intelligence and Security Informatics, 2007, pp. 282-288.

[10] Cohen-Almagor R. Countering hate on the Internet. Annual review of law and ethics. 2014, pp. 29;22:431-43.

[11] NGDATA. What is Predictive Analytics? Definition and Models," [Online]. Available: https://www.ngdata.com/what-is-predictive-analytics/. [Accessed: 04-Jan-2019].

[12] Thomas Davidson, Dana Warmsley, Michael Macy, Ingmar Weber, "Automated Hate Speech Detection and the Problem of Offensive Language," Proceedings of ICWSM 2017, 2017.