Measuring Online Affects in a White Supremacy Forum

Léo Figea
Dept. of Information Technology
Uppsala University
Uppsala, Sweden
leo.figea@telecom-bretagne.eu

Lisa Kaati
Dept. of Information Technology
Uppsala University
Uppsala, Sweden
lisa.kaati@it.uu.se

Ryan Scrivens
International CyberCrime Research Centre
School of Criminology
Simon Fraser University
Burnaby, Canada
rscriven@sfu.ca

***Abstract*—Since the inception of the World Wide Web, security agencies, researchers, and analysts have focused much of their attention on the sentiment found on hate-inspired web-forums. Here, one of their goals has been to detect and measure users' affects that are expressed in the forums as well as identify how users' affects change over time. Manual inspection has been one way to do this; however, as the number of discussion posts and sub-forums increase, there has been a growing need for an automated system that can assist humans in their analysis. The aim of this paper, then, is to detect and measure a number of affects expressed in written text on Stormfront.org, the most visited hate forum on the Web. To do this, we used a machine learning approach where we trained a model to recognize affects on three sub-forums: *Ideology and Philosophy*, *For Stormfront Ladies Only*, and *Stormfront Ireland*. The training data consisted of manual annotated data and the affects we focused on were racism, aggression, and worries. Results indicate that even though measuring affects is a subjective process, machine learning is a promising way forward to analyze and measure the presence of different affects on hate forums.**

I. Introduction

It is widely acknowledged that people around the world are increasingly using the Internet to connect with each other, and so too are violent right-wing extremists and those who subscribe to far-right beliefs [1]. Indeed, the Internet has enabled members of the radical Right to recruit new members, enhance existing collectives, and create new online-shared identities. A primary outlet for these threatening activities has been on one of the most notorious online communities: Stormfront [2].

In the mid-1990s, the Internet became accessible to the general public, and amongst the first websites to surface online was Stormfront, a theme-based discussion forum [3]. Launched in 1996 by white supremacist and ex-felon Donald Black, Stormfront was arguably the first "hate site" and is still one of the most visited and well-known sites of its class, with online membership increasing each year. For instance, the site has gained nearly 300,000 members since its 20th anniversary in 2015, and members have posted of over 1,000 messages each day [4].

In essence, Stormfront paved the way for white nationalists to stake their claim in cyberspace, encouraging others to promote discussions on Jewish conspiracy, Black inferiority, and government atrocities, for example [5]. Within this virtual community, much of the sentiment tends to revolve around a key element: "White supremacy," which promotes the Aryan race as the elite of all races and the targeting of "racial enemies" [1].

Scholars have generally agreed that right-wing extremists use the Internet to build a sense of unity and collective identity. This is done through online interactions with other like-minded individuals, othering those who do not meet the standard (i.e., anyone who is non-White) and depicting the White race as victims of cultural genocide [6]. On Stormfront, for example, such radical discussions are preserved by the site's countless links to community discussion boards, news stories, and other radical sites [3]. Scholars suggest that right-wing extremists use this tactic for a number of reasons, which includes: (1) efforts to maintain a legitimate and even justifiable image of a hate movement; (2) build their collective identity and, by extension, develop and maintain extremist ideologies, and; (3) recruiting new members, to name but a few. Several studies have analyzed these driving components on right-wing extremist forums, yet much of the research has been largely descriptive [7], [8] [9], counted the hyperlinks posted within and between sites [3], [10], and assessed the online formation of far-right communities [11], [4], [1]. Much is still unknown about how to identify extremists or radical content on the Web, especially using a machine learning method on a large scale. Sentiment analysis and affect analysis have been useful means of addressing this challenging task, but this area of research remains very underdeveloped [12].

A. Outline

This paper is outlined as follows: Section II outlines research that is related to our study. In Section III we discuss the different affects that we identify and measure. In Section IV we describe the experimental setup and the results from our experiments reported in Section V. Section VI includes the results of the study while Section VII includes a discussion of the results. Finally, conclusions and directions for future work are presented in Section VIII.

II. Related work

Emotions/affects play an important role in people's decision making process and influences their perceptions as well.

978-1-5090-3865-7/16/$31.00 ©2016 IEEE

However, recognizing affects requires emotional intelligence - a skillset that machines do not have. This problem is described in [13] where a machine learning approach assesses eight classes of emotions in human behavior. In this work, we are interested in recognizing affects in written text - something that is even more difficult to recognize than affects in real life behavior. This is known as affect analysis, a category of sentiment analysis. In affect analysis, measuring emotions that are expressed as text has proven to be useful in measuring the presence of violence and hatred on online extremist forums. For example, Abbasi and Chen [12] employed affect analysis on 16 U.S. supremacist and Middle Eastern extremist group discussion forums [12]. In their work, the authors used an affect lexicon where each term had an intensity score. The intensity assigned to each term in the lexicon depended heavily on the context within the forums. Each term was assigned a probability of being used in the proper context by sampling a number of occurrences of each term to determine the number of times the feature appeared in an irrelevant context. Here, affects that were measured were violence and hate affects. While the study demonstrated how affect analysis can measure the presence of violence and hatred on extremist forums, the researchers did not measure affects on forums as a whole, nor on individual forum users specifically.

Chen [14] developed an automated approach to explore sentiment and affect analysis on two Dark Web forums that related to the Iraqi insurgency and Al-Qaeda, employing a rich textual feature set on 500 sentences from each of the forums. The purpose of this successful project was to identify the presence and intensities of four affects in forum postings (i.e., violence, anger, hate, and racism), and this was done through support vector regression (SVR). SVR can predict continuous affect intensities while benefitting from the robustness of support vector machines (SVM), which is one of the classifiers that we used in this study.

Another related study, one which evaluated sentiment (positive and negative) rather than affects to explore user's online activity in extremist-based web-forums, is described in [15]. In this work, the focus was on detecting radical authors in online forums. The authors developed an algorithm called "Sentiment-based Identification of Radical Authors (SIRA)" to quantify an individual's online activity that may be deemed as "extreme" based on their collective posts within a discussion forum. Specifically, sentiment analysis was used to identify the most radical users across approximately 1 million posts and 26,000 unique users found on four Islamic-based web-forums. Several characteristics of each user's postings were examined, including their posting behavior and the content of their posts. The content was analyzed using Parts-Of-Speech (POS) tagging, sentiment analysis, and the SIRA algorithm, which accounted for a user's percentile score for average sentiment score, volume of negative posts, severity of negative posts, and duration of negative posts.

The purpose of the current study is to develop classifiers that can identity three unique affects found within the written communications on a hate-based web-forum. Affects include racism, aggression, and worries, all of which are important to monitor, both for security agencies and researchers. Indeed, we continue to see a growing need to understand how users' affects changes over time and increase our understand of how particular events effect online communities. On an individual level, affect analysis can be used to gain a deeper understanding of group polarization and how users change their online behavior when they become members in a discussion forum. It is also of interest to identify forum users who exhibit a high level of a particular affect(s) in their online communication as well as how this changes over time.

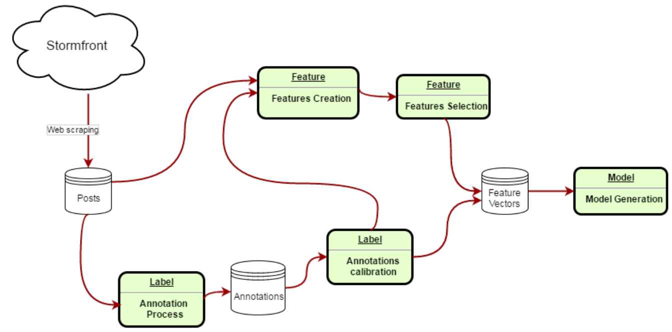

Fig. 1. The workflow of the experiments

III. Measuring Affects

Affects can be understood as the experience of a feeling or emotion. As noted previously, the affects we consider in this work are affects that we believe are important to monitor and understand from an intelligence and research perspective. The three affects that we consider are:

- Aggression towards an out-group (i.e., aggression directed towards other communities such as visible minorities, immigrants, Jews, etc.)
- Racism (i.e., racist insults and references to race)
- Worries (i.e., expressions of concern, fear, or worry about the current situation or the future)

Measuring these affects in the above mentioned environment is a difficult and very subjective task. In an effort to gather the necessary data to create our models, domain experts annotated posts with the aim of detecting the presence of three different affects. The scale was between 0 and 7, with 0 indicating no presence of the affect and 7 indicating a high presence of the affect. To get consistent results, the annotation process was limited to a small amount of annotators and was monitored. Each post was annotated by three separate annotators. Due to the highly subjective nature of annotation, many of the posts were annotated significantly differently by two annotators. These posts were removed from the training set, and the final annotation of a post was set to the the mean of the annotations by the different annotators.

The scale that we used in the annotation of the affects was used to obtain some of the features and to observe the performance of the model during the experiments. In the end,

TABLE I
THE DIFFERENT SUB-FORUMS USED IN THE EXPERIMENTS.

Name	Description	No of posts	No of annotations
Ideology and Philosophy	Foundations for White Nationalism	35808	226
For Stormfront Ladies Only	Sugar and spice, and everything nice	47502	158
Stormfront Ireland	White Nationalist issues in Ireland	37451	160

each post was labeled as either expressing a high or low level of affect in a binary scale. If a post for a certain affect was, on average, annotated above two, then it was considered an expression of a high level of the affect, otherwise it was considered an expression of a low level of the affect. After removing the posts that were annotated differently by the annotators, we had a total of 300 annotated posts in our sample. The level of the affects was used to identify some of the features that we employed in the classification.

IV. EXPERIMENTAL SETUP

In this section we describe the experimental setup. Fig. 1 shows the workflow of how the experiments were conducted. First, data was collected and stored in a database. The data was annotated and the annotations calibrated. Features were then created based on the annotations and important features that contributed to the classification (for each classifier) were selected and feature vectors that included the result of the annotations were stored in a database. Lastly, a model was generated. For the experiments, we utilized three different classifiers: Random Forest, support vector machine (SVM) and Adaboost.

A. Dataset

For the experiments we used data from three of Stormfront's sub-forums: *Ideology and Philosophy*, *For Stormfront Ladies Only* and *Stormfront Ireland*. Table I provides a short description of each unique sub-forum. Since the number of annotated posts showing high presence of affects is much smaller than the number of posts showing a low presence, we have a dataset that is unbalanced. Unbalanced datasets is always a problem in machine learning since the algorithms usually try to obtain the highest accuracy possible. However, accuracy is not always the most appropriate measurement according to its definition. The pre-processing of the data included removing URLs, stop words (the most common words in a language) and punctuation. The number of misspellings were counted (since the number of misspelling is one of our features) and stemming and lemmatization were then conducted.

B. Features

Deciding which features should be included in the classification is an important task and requires an in-depth understanding of the domain under investigation. The features we used can be divided into two classes: (1) data dependent features and (2) data independent features. The data dependent features are those that rely heavily on the dataset while the data independent features are less dependent on the dataset.

The data independent features we used include the LIWC 2015 categories [16], the number of misspelling, the number of words, three different expert knowledge dictionaries (i.e., words related to worries, racism and aggression), and Part-Of-Speech (POS) tags. Linguistic Inquiry and Word Count (LIWC) is a text analysis tool that counts words in psychologically meaningful categories. Based on the relative frequency of words from the different categories, it is possible to create a profile of the person who wrote the text, such as how much they used words from the different LIWC categories (which is presented in percentage). LIWC has been tested and evaluated in a number of different studies [17], [18]. Examples of categories that are present in LIWC are positive and negative emotions, personal pronouns (i.e., first person plural, second person plural, third person plural, first person singular, etc.) and cognitive processes, such as insight, certainty, tentative and inhibition. Table III shows the the different dimensions of LIWC and the categories included in each dimension. We used LIWC and the psychologically meaningful categories to gain an understanding of how they would effect the results of measuring the different affects.

The data dependent features included the 100 most frequent words found in the posts with a high presence of a given affect (i.e., level 6-7), and the 100 words that had the highest difference between the frequency of occurrence in posts with a high presence of an affect and the frequency of occurrences in posts with a low level of an affect. Table II shows the categories of features, the number of features, and a short description of the features we used.

C. Evaluation

We report the results using confusion matrices in which we present the number of true positives, false negatives, true negatives, and false positives (see Table IV). In our evaluation we included a number of statistics that are derived from the confusion matrices. The measures include accuracy and recall where accuracy is defined as: $\frac{TP+TN}{TP+FP+TN+FN}$ and recall as $\frac{TP}{TP+FN}$.

V. RESULTS

In this section, we provide the results that are derived from the two sets of features (i.e., data dependent + data independent and only data independent) and the three different classifiers. Since the input dataset was relatively small, a computationally expensive all but one (ABO) cross validation was considered. ABO cross validation trains as many models as number of records on the entire dataset, leaving one record out each time for testing. The low level of affect are

TABLE II
THE DIFFERENT FEATURES USED TO MEASURE THE AFFECTS.

Feature category	**No of features**	**Description**	**Data dependent**
LIWC	73	Psychologically meaningful categories from LIWC 2015	no
Frequent words	100	Frequent words from the high affect posts	yes
Difference words	100	Words with different frequency in high and low level of affect posts	yes
Part of speech tags	23	Lexical items with similar grammatical properties	no
Expert knowledge dictionaries	3	Dictionaries with words related to racism, worries and anger	no
Misspelling	1	Frequency of misspelling	no
Word count	1	Total number of words	no

TABLE III
THE DIFFERENT LIWC DIMENSIONS AND THE CATEGORIES THAT ARE INCLUDED IN EACH DIMENSION.

LIWC Dimensions	**Categories included**
Summary variables	Analytic, clout, authentic, tone
Language metrics	Words per sentence, words bigger than six letters, dictionary words
Function words	Pronouns, articles, prepositions, negations, auxiliary verbs, common adverbs, conjunctions, quantifiers
Other grammar	Regular verbs, adjectives, numbers, comparatives, interrogatives,
Affect words	Positive emotion, negative emotion
Social words	Family, friends, female referents, male referents
Cognitive processes	Insight, cause, certainty, differentiation, tentativeness, discrepancies
Perceptual processes	Seeing, hearing, feeling
Biological processes	Body, health/illness, sexuality, Ingesting
Drives and Needs	Affiliation, achievement, power, reward focus, risk/prevention focus
Time orientation	Past focus, present focus, future focus
Relativity	Motion, space, time
Personal concerns	Work, leisure, home, money, religion, death
Informal speech	Swear words, netspeak, fillers, assent, nonfluencies

TABLE IV
A CONFUSION MATRIX USED FOR REPORTING THE RESULTS.

	Predicted class	
Actual class	True Neg. (TN)	False Pos. (FP)
	False Neg. (FN)	True Pos. (TP)

overrepresented in the dataset, and the SMOTE technique [19] is used to balance the dataset by creating new records for the underrepresented class (i.e., high presence of the racism affect for instance). The SMOTE technique was only used on the training set within the loop, and tests were conducted on real records in an effort to reduce bias in the evaluation of the model.

A. Experiment 1: Data dependent features

In experiment one, we used all available features and three different classification algorithms: Random Forest, SVM and Adaboost [20]. The results are shown in Table V. The base classifiers we used was classification trees and the R ada package was used for the implementation. The R ada package provided means for estimating the optimal number of base classifiers, and we followed these suggestions. The features were ranked according to their Mahanalobis relevance estimate, which is shown in Table VI.

TABLE V
CONFUSION MATRIX FOR THE AFFECTS IN EXPERIMENT 1.

	Low	**High**	**Classifier**	**Accuracy**	**Recall**
Racism	257 11	10 21	Random Forest	92.98%	67.74%
Racism	256 12	2 29	SVM	95.32%	93.55%
Racism	257 11	13 18	Adaboost	91.97%	58.06%
Worries	271 11	13 6	Random Forest	92.03%	31.58%
Worries	279 3	5 14	SVM	97.34 %	73.68 %
Worries	256 26	10 9	Adaboost	88.04%	47.37%
Aggression	284 7	6 11	Random Forest	95.78%	64.71%
Aggression	285 6	4 13	SVM	96.75%	76.47 %
Aggression	262 29	6 11	Adaboost	88.64%	64.71%

TABLE VI
IMPORTANT FEATURES FOR THE CLASSIFICATION OF THE DIFFERENT AFFECTS EXPERIMENT 1.

Racism	**Worries**	**Aggression**
black	immigr	immig
race	cherish	isnt
white	advoc	migrant
jew	threat	conflict
adolf	bash	riot
Religion (LIWC)	riot	scotland
exploit	boundari	destroy
protest	aside	aside
pro	non	islam
orang	collaps	thousand
zionist	preced	preced
invent	protest	Anger (LIWC)
See (LIWC)	object	Aggression (expert list)

B. Experiment 2: Data independent features

Experiment two included the same setup as experiment one; however, we used only data independent features to gain an understanding of how the data dependent features effected the result. The results are shown in Table VII. For the data independent features, the rank according to their Mahanalobis relevance estimate is shown in Table VIII.

TABLE VII
CONFUSION MATRIX FOR THE AFFECTS IN EXPERIMENT 2.

	Low	High	Classifier	Accuracy	Recall
Racism	230 38	13 18	Random Forest	82.94%	58.06%
Racism	224 44	12 19	SVM	81.27%	61.29%
Racism	228 40	14 17	Adaboost	81.94%	54.84%
Worries	252 30	7 12	Random Forest	87.71%	63.16%
Worries	244 38	6 13	SVM	85.38 %	68.42 %
Worries	249 33	6 13	Adaboost	87.04%	68.42%
Aggression	276 15	6 11	Random Forest	93.18%	64.71%
Aggression	254 37	4 13	SVM	86.69%	76.47 %
Aggression	263 28	6 11	Adaboost	88.96%	64.71%

TABLE VIII
IMPORTANT FEATURES FOR THE CLASSIFICATION OF THE DIFFERENT AFFECTS EXPERIMENT 2.

Racism	Worries	Aggression
Relig (LIWC)	Anger (LIWC)	Anger (LIWC)
See (LIWC)	Aggression (Expert)	Aggression (Expert)
word count	word count	Power (LIWC)
They (LIWC)	Racism (Expert list)	Risk (LIWC)
possesive (POS)	Power (LIWC)	Money (LIWC)
Anger (LIWC)	See (LIWC)	Focus Past (LIWC)
Aggression (Expert)	fullstop (POS)	Negemo (LIWC)
Adj (LIWC)	Focus Past (LIWC)	fullstop (POS)
Affiliation (LIWC)	Negemo (LIWC)	word count
preposition (POS)	possesive (POS)	Racism (Expert)
fullstop (POS)	Sad (LIWC)	existential (POS)
Drives (POS)	pronoun (POS)	foreign (POS)
Power (LIWC)	preposition (POS)	They (LIWC)

VI. AFFECTS IN WHITE SUPREMACY FORUMS

When applying the models to the sub-forums, it is possible to measure the affects and analyze changes over time. Figure 2 shows the level of racism, Figure 3 the level of aggression, and Figure 4 the level of worries on the three different forums. As it can be seen in the figures, the level of racism, aggression, and worries vary on the different sub-forums. For example, the *Stormfront for Ladies Only* sub-forum has a low level of all the affects, while the *Ideology and Philosophy* sub-forum has the highest level of racism and the *Stormfront Ireland* sub-forum has the highest level of aggression and worries (after 2004).

VII. DISCUSSION

The results are twofold: (1) using a variety of features (both data dependent and data independent) provide a superior classification result and, (2) SVM outperforms all other classifiers in the study (i.e., Random Forest and Adaboost). Similarly, the use of SVM to classify affects was found to be the appropriate choice in [21]. It can be considered a spatial approach of machine learning. Conversely, feature selection is less effective when using SVM compared to Adaboost and Random Forest. Adaboost and Random Forest apply a bootstrapping over the features, all of which leads to inaccurate results if uninformative features are sampled. However, the algorithms usually compensate by iterating over large training data and putting weight on misclassified records, as is the case with Adaboost. In this study, the limited size of the dataset is a factor explaining why SVM obtains the most accurate results. Our results indicate that the affect of racism is easier to recognize than affects of aggression and worries, and the affect of aggression is easier to recognize than worries.

Previous work has obtained meaningful results in classifying affects using data dependent features, such as character N-

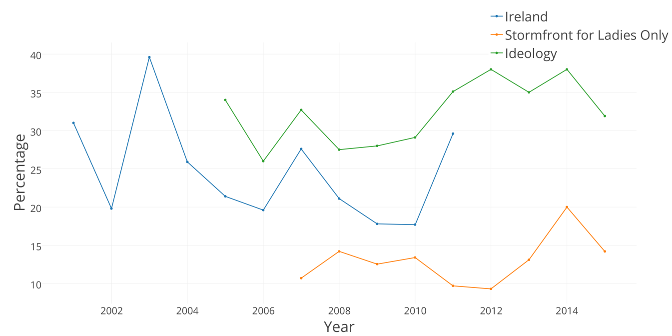

Fig. 2. The level of racism on the different forums

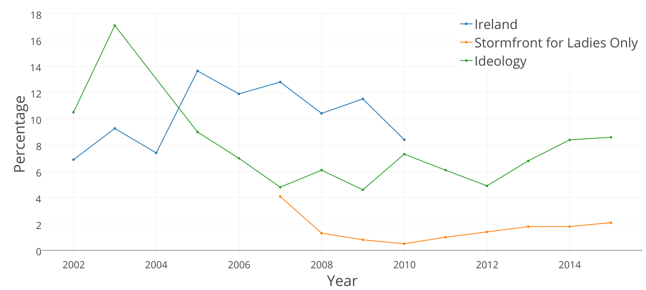

Fig. 3. The level of aggression on the different forums

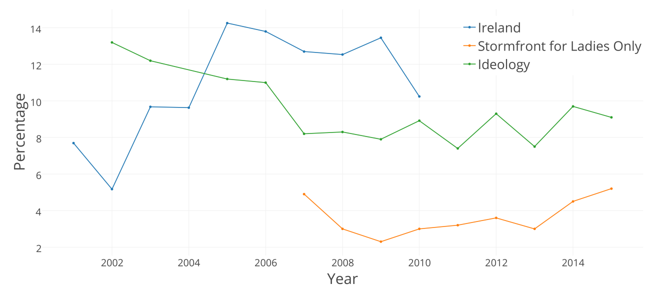

Fig. 4. The level of worries on the different forums

grams, word N-grams and root N-grams [14]. In experiment two we used only data independent features, such as the psychological meaningful categories from LIWC and dictionaries developed by domain experts, and we still obtained a classification accuracy between 80% and 93% and recall between 61% and 76%. The importance of the different features for each affect, as well as the differences between the importance of features, provides some insight into the expression of affects. Together, racist words that are derived from the category "Religion" (i.e., terms related to religion such as church, Muslim, etc.) and the category "See" (e.g., view, saw, seen) are important with the use of third person plural (e.g. they, them). The use of third person plural words, as was suggested in [18], indicates that the group is defining itself to a large degree by the existence of an oppositional group and, as is pointed out in [22], prejudice and discrimination are based on psychological distinction between "us" versus "them". For the worries affect, words related to anger, racist words, and power words (e.g., bully, superior) are important classification features. For aggression, words related to anger, power words, and risk words (e.g. danger, doubt), and money words (e.g. audit, cash and owe) are important for the classification. A more detailed analysis, one which addresses why these features are important for the classification process, is beyond the scope of this paper but would be of great interest to investigate in future research.

VIII. Conclusion and Future Work

The experiment highlights that machine learning, and in particular the use of SVM as a classifier, is a way forward when classifying affects on white supremacy sub-forums. Not surprisingly, data dependent features provide a more successful result than using only data independent features. On the machine learning theory level, the feature selection was based on an approximation of Kullback Lieber distance, a heuristic of the informativity of individual features. To improve the results for future work, the features could be considered as sets and the relevance of the sets of features could be considered. On another note, the most difficult part of this exercise is to attain a large amount of training data. The process of annotating the level of affects is very subjective, especially when considering affects on hate forums where the entire context is extremely high in affects. How we interpret subjective affects such as worries, aggression, and racism is very much dependent on our own background and the knowledge of the domain that is considered. Due to the subjectiveness of the affects that we considered and even more the context where the posts were published, the annotation process was harder than we expected. Further, since we removed all posts that were annotated much differently by two annotators, we did not obtain a dataset that was as large as we initially planned. A direction for future work is to put more emphasize on the annotation and have a more structured process to annotate more posts. In our study, the annotation was done on the entire posts and not on selected sentences as in [14]. For future work, a combination of both strategies may provide new insight into this growing field of inquiry.

References

[1] M. Whine, "Cyberspace a new medium for communication, command, and control by extremists," *Studies in Conflict and Terrorism*, vol. 22, no. 3, pp. 231–245, 1999.

[2] M. Potok, "From the ground up," *New York Times*, 2005.

[3] V. Burris, E. Smith, and A. Strahm, "White supremacist networks on the internet," *Sociological Focus*, vol. 33, no. 2, pp. 215–235, 2000.

[4] K. Thompson, "Watching the Stormfront: white nationalists and the building of community in cyberspace," *Social Analysis*, vol. 45, no. 1, pp. 32–52, 2001.

[5] B. Levin, "Cyberhate: A legal and historical analysis of extermists use of computer networks in america," *American Behavioral Scientist*, vol. 45, no. 6, pp. 958–988, 2002.

[6] B. Perry and R. Scrivens, "White pride worldwide: Constructing global identities online," *J. Schweppe and M Walters (eds.), The Globalisation of Hate: Internationalising Hate Crime? New York: Oxford University Press*, pp. 65–78, 2016.

[7] S. Anahita, "Blogging the boarders: virtual skinheads, hypermasculinity, and heteronormativity," *Journal of Political and Military Sociology*, vol. 32, no. 1, pp. 143–164, 2006.

[8] L. Back, "Aryan reading Adorno: cyber-culture and twenty-first century racism," *Ethnic and Racial Studies*, vol. 24, no. 4, pp. 628–651, 2002.

[9] L. Bowman-Grieve, "Exploring stormfront: A virtual community of the radical right," *Studies in Conflict and Terrorism*, vol. 31, no. 11, pp. 989–1007, 2009.

[10] M. A. Wong, R. Frank, and R. Allsup, "The supremacy of online white supremacists - an analysis of online discussions by white supremacists," *Inf. & Comm. Techn. Law*, vol. 24, no. 1, pp. 41–73, 2015.

[11] M. E. Duffy, "Web of hate: a fantasy theme analysis of the rhetorical vision of hate groups online," *Journal of Communication Inquiry*, vol. 27, no. 3, pp. 291–312, 2003.

[12] A. Abbasi and H. Chen, "Affect intensity analysis of dark web forums," in *Proceedings of the 5th IEEE International Conference on Intelligence and Security Informatics*, 2007.

[13] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, pp. 1175–1191, 2001.

[14] H. Chen, "Sentiment and affect analysis of dark web forums: Measuring radicalization on the internet," in *Intelligence and Security Informatics, 2008. ISI 2008. IEEE International Conference on*, 2008, pp. 104–109.

[15] R. Scrivens, G. Davies, R. Frank, and J. Mei, "Sentiment-based Identifiation of Radical Authors (SIRA)," in *the 2015 IEEE ICDM Workshop on Intelligence and Security Informatics.*, 2015.

[16] J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count (LIWC): A text analysis program." New York: Erlbaum Publishers, 2001.

[17] Y. Tausczik and P. J.W., "The psychological meaning of words: LIWC and computerized text analysis methods," *Journal of Language and Social Psychology*, vol. 29, no. 1, March 2010.

[18] J. W. Pennebaker and C. K. Chung, "Computerized text analysis of al-Qaeda transcripts," in *The Content Analysis Reader*, K. Krippendorf and M. A. Bock, Eds. Sage, 2008.

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "Smote: Synthetic minority over-sampling technique," *J. Artif. Int. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002.

[20] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *In Proceedings of the thirteenth International Conference on Machine Learning.*, 1996, pp. 148–156.

[21] H. Cho and K. J. Lee, "Automatic affect recognition using natural language processing techniques and manually built affect lexicon," *IEICE Tran. Information Systems*, vol. E89, no. 12, 2006.

[22] J. W. Pennebaker and C. K. Chung, "Language and social dynamics," in *Technical Report 1318*. University of Texas at Austin, 2012.