# ANALYZING SENTIMENT OF VIOLENT POLITICAL EXTREMISM CONTENT

SUBMITTED IN PARTIAL FULFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE


STAVROULA AFRODITI CHRISTOPOULOU
11412984

MASTER INFORMATION STUDIES
DATA SCIENCE
FACULTY OF SCIENCE
UNIVERSITY OF AMSTERDAM


2017-08-23

| | Supervisor | Second reader | Third reader |
|---|---|---|---|
| **Title, Name** | Dr.Stevan Rudinac | Pascal Mettes | Dr.Masoud Mazloom |
| **Affiliation** | UvA, FNWI, IvI | UvA, FNWI, IvI | UvA, FNWI, IvI |
| **Email** | S.Rudinac@uva.nl | P.S.M.Mettes@uva.nl | M.Mazloom@uva.nl |

UNIVERSITEIT VAN AMSTERDAM

# Analyzing Sentiment of Violent Political Extremism Content

Stavroula Afroditi Christopoulou
University of Amsterdam
Science Park 904 Amsterdam

evitaams@gmail.com

## ABSTRACT

In the century of vast technological developments and intellectual achievements, the emergence of numerous supporters of extreme doctrines is remarkable. In this thesis we show that sentiment analysis can be used for gaining insight into the structure and evolution of an online extremism forum. Additionally, we analyze gender differences and seasonal trends related to sentiment of user-generated posts.

## Keywords

sentiment analysis, extremism, forum, trend developments

## 1. INTRODUCTION

Over the years, technology has significantly changed the way that people interact and communicate. Internet inaugurated the era of instantaneous access to any source and unobstructed communication worldwide, but most importantly the freedom of expression and exchange of ideas.

Nowadays, a common way for people to communicate is through social platforms, having the opportunity to express how they feel or think, and disclosing how they react to different situations. This way not only they expose themselves, but also they get influenced by other people's opinions. Sentiment analysis, analyses text through the spectrum of emotions [22] and specifically aims to extract positive or negative opinions from unstructured text [23].

This thesis evolves around extracting and evaluating sentiment scores derived from the extremist forum Stormfront, where members from all over the world communicate and exchange ideas. Our collection contains 2 million extremist posts, associated with various meta data.

Considering that extremism is any political theory that favors immoderate uncompromising policies, an extremist group is a group of individuals whose values, ideals and beliefs differ significantly of what society deems as normal [3]. It is characteristic that they do not usually view themselves as being extremists, which results in sharing racist or violent ideas without hesitation through social networks.

In this thesis we show that sentiment analysis can be used for gaining insight into the structure and evolution of an online extremism forum. More specifically we investigate whether it is possible to make estimations about the extremeness of ideologies, demographics and trend developments, as well as gender differences and seasonal trends.

A zestful topic to explore is whether Stormfront evolves over time with an increase in number of posts. Furthermore, we assay whether it is possible to estimate the extremeness of ideologies, by analyzing the content of these posts.

Additionally, we investigate the proportion of the users that participate systematically in the forum, as well as the sentiment scores of the most active users. Months with the most positive or negative content can disclose information connected with important events or seasonal trends.

Gender differences will also be revealed in an in depth review of specific categories. The main research question of this thesis is defined as:

*Can sentiment analysis provide insight in the demographics and trend developments of an extremist forum?*

Since environmental temperature and seasonality usually affect human behaviour, it is interesting to investigate whether these factors affect the expression of ideologies, resulting to the first research sub question:

*Does seasonality affect the expression of extremeness?*

Furthermore, we will investigate the way that female members express themselves compared to the general population. This leads to the formulation of the second research sub question as:

*Do women extremists express themselves more negatively compared to the general population?*

The results of this study may be useful in capturing the evolution and structure of Stormfront. These results can also trigger further investigation on the impact that this evolution has in the wider society.

## 2. RELATED WORK

In this chapter relevant related work as well as different approaches will be discussed, including information concerning the research questions apposed.

Many years before internet has appeared, extremists have found their own ways to communicate and interact, through the past available technology. Strong relationships were created between them, focusing on the same ideology that they shared and their passion to distinct [17].

Followingly, after internet has been extensively known, they have developed their way of communication by forming their own forums, which offered them a more direct way of communication, and strengthened the relationship between them. One of these extremist forums, called Stormfront, is analyzed in this thesis. It is a social media platform which sidelong harbors its racist identity [6].

Jungherr's related research based on Twitter has shown that internet can be an important tool in everyone's hands, as it can spread information fast, and influence people's opinion [2]. In regards with extremism, it can also be a medium of political expression.

Sentiment analysis on political ideology content has been implemented formerly by Bakliwall et all in 2013, who aimed to analyze political tweets with an accurate classifier [10]. For that purpose two different subjectivity lexicons were used, one part-of-speech tagger (designed specifically for tweets), and one parser (Stanford parser). The identification of sentiment polarity of a word was approached through Subjectivity Lexicon [33], and SentiWordNet 3.0 [31]. Our sentiment analysis approach will be discussed in the next subsection.

### 2.1 Textual Analysis

Internet has introduced an epoch in which people share broadly their opinion through social media. It has also become a source of gathering available information concerning people's notion, due to the fact that a majority of people who participate in social networks, are willing to express their experiences, beliefs and sentiments about their every day lives (such as: the purchases that they make, events that they attend or choices that they will make in the future). This free expression of opinion has led to the free access in mass opinions. *Sentiment analysis* is a method that emits sentiment from text [21], and can give us insight whether the content of a text is positive or negative. In our research sentiment analysis will be used in order to analyze the beliefs and emotions that Stormfront users have.

Goncalves et al. provide an overview of eight methods for extracting sentiment from text [15]. There are machine-learning-based approaches, such as SVM, Maximum Entropy and Naïve Bayes and lexicon-based approaches, such as LIWC and POMS-ex [16].

Generally, there are two methods of sentiment analysis. The machine learning method, in which we train a classifier to categorize and distinguish the quality of a text as positive or negative. This method is automatic and it lacks of content comprehension [27]. And the lexicon-based method, in which the sentiment analysis is based on each lexicon (the lexical units that compose each lexicon), therefore the results of this analysis can vary according to lexicons' completion [24].

More specifically, lexicon-based approaches require a lexicon with predefined set of terms that determine sentiments [15], afterwards words are scored according to the amount of sentiment they evoke. In our research we are using the second method (lexicon-based approach).

Currently, the primary method used to extract sentiment from the Stormfront forum, is a lexicon-rule-based model, called VADER. VADER (Valence Aware Dictionary for Sentiment Reasoning) is a Python package, it is simple, intuitive and the results are easy to interpret. What makes it distinct from other tools is that it is quickly applied and has no need for extensive learning or training [19].

VADER lexicon consists of a number of words. Each word is assigned to a specific score which indicates the intensity of the sentiment of this word. When we process a text with VADER, it can identify if any word of the text exists in the lexicon, and it rates it with the respective score. Subsequently, it gives four overall scores to the text: positive, negative, neutral and compound. According to the proportion of these scores we can identify a text as positive, negative or neutral and see how positive, negative or neutral it is. (The compound score, is the sum of the positive, negative and neutral scores, ranging between -1 and +1). Furthermore, it is remarkable that VADER recognizes emoticons and adapts the scoring according to this sentiment-related information [8]. In our current research the calculation of sentiment analysis scores were derived through VADER and corroborated with TextBlob [7].

### 2.2 Seasonality

According to Kallek, seasonality is characterized by the stability in the appearance of an event, during a time line [5]. Another research based on human behaviour during different periods of the year, has apposed that according to environmental changes, people's behaviour changes as well, making them vulnerable to negative sentiments, crime, aggressive behaviour etc. It has been observed that during the summer months the aggressive behaviour was being increased, resulting to numerous murders, sexual assaults and attacks. A deeper insight in this research reveals that when the sun is shinning, people are becoming more aggressive [12].

In our research we are going to investigate whether summer months, or any other season, affect violent (extreme) behaviour. Since summer months and heat coexist with aggressive behaviour according to the aforementioned research [12], we expect that extremist users of Stormfront will have more negative scores in the content of their posts, during the summer season.

## 2.3 Gender language differences

Multiple researches have been conducted around the fact that there is difference in the way that female and male internet users express themselves, and participate in discussions.

One of these researches, which aimed to compare the participation of female and male students in two academic networks, has shown that female users participated less than male, and with shorter posts. A possible explanation was that female users did not receive positive feedback and felt less motivated to participate in such networks. In addition, it was revealed that female users express themselves differently, adopting a "male vocabulary" in order to feel appreciable and accepted by those male dominated communities [4].

According to Colley, women are using a more condescending and emotional language than men, who prefer to express themselves in a more strict and direct way, including aggressive language. Furthermore women tend to get involved in discussions about more feminine themes (such as relationships) [13].

Withal, according to Newman, women are using more polite expressions than men and are more conformed in societal commands. They are mostly focusing on their emotions when they communicate with others, compared to men, who usually communicate for more pragmatic reasons [25].

However, in contrast with the aforementioned researches, some other researches support that women tend to be more aware in remembering and expressing negative feelings, than men [11]. In this thesis we are focusing mostly on female extremist users. They participate in a special category in Stormfront, dedicated to women. This category is known to attract mostly female forum members.

Although women are usually considered to have less rights in male dominant societies and communities such as extremists, it is remarkable that they are accepted as "equal" members in these forums. A possible reason is that they are treated as the persons who create the next generations of extremists, therefore they are treated with respect [29].

In this research, language expression differences were approached by examining 2 million posts, using an automated sentiment analysis tool, called VADER. The aim of this exploration, is to detect whether there are differences in the way that female members express extremeness compared to the general population. This way we will gain insight in the structure of Stormfront and understand the role of women in extremist societies.

# 3. METHODOLOGY

This thesis methodology is a combination of qualitative and quantitative analysis. The basic components are apposed and explained below.

## 3.1 Data Collection

The data collection includes user posts from the extremist forum, Stormfront.

The available dataset of 5.3 GB, contains 2 million extremist posts. Computations were done with a DAS-4 supercomputer (fourth generation DAS) [18]. The dataset is broadly divided in 40 categories, associated with various metadata, such as the document id, the user name, whether it was a reply or an initial post, the topic and category of the post, the content of the post and the entities [26] [14] automatically extracted from the post.

People from all over the world take part in the discussions and these available metadata can be used to reveal hidden aspects of extremists' relationships and how like minded people communicate.

## 3.2 Statistical Analysis and Research hypothesis

This section will discuss the statistical analysis of the acquired sentiment scores and how the data will be analyzed for possible correlations. It is noteworthy that after calculating the sentiment scores of some data, aggregations were implemented by finding the mean compound sentiment scores of the respective data.

In order to examine whether there are correlations between our data, Pearson r correlation tests were conducted. Values can range between -1 and +1 and indicate the intensity of the linear relationship that two variables have ( 0 is considered as no correlation) [30].

# 4. EXPERIMENTAL RESULTS

The aforementioned research questions were supported by a series of experiments.

## 4.1 Demographics and trend developments

Firstly, in order to support whether sentiment analysis can provide insight in the demographics and trend developments of an extremist forum, sentiment analysis was conducted to extract the average sentiment score of posts that had been posted, for each month of each year in the data set.

The Pearson r correlation test between the average compound sentiment score of each month of each year and the timeline of the data, indicates the evolution of Stormfront users' sentiment over time and has revealed a strong downhill (negative) linear relationship of $p = -0.79578923$, which shows that users' sentiment score has become more negative over time (Figure 1).

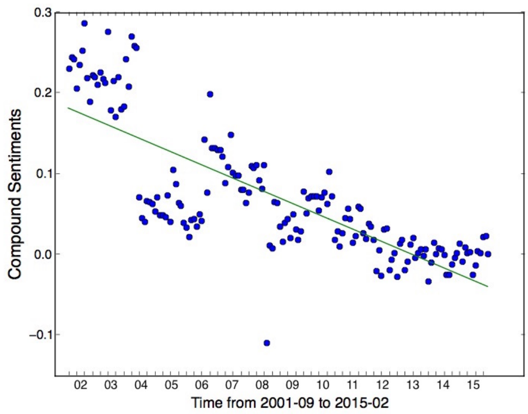


**Figure 1: Correlation between monthly average compound sentiment scores and time line over months/years**

Furthermore, the correlation between the average compound sentiment score of each month of each year with the number of posts of each month has also revealed a strong downhill (negative) linear relationship of p = -0.73970208 (Figure 2), showing that as the number of posts increase, the more negative they become. Therefore we can arrive to the conclusion that Stormfront has evolved over time, with an increasing number of posts and the content of these posts has become more negative.

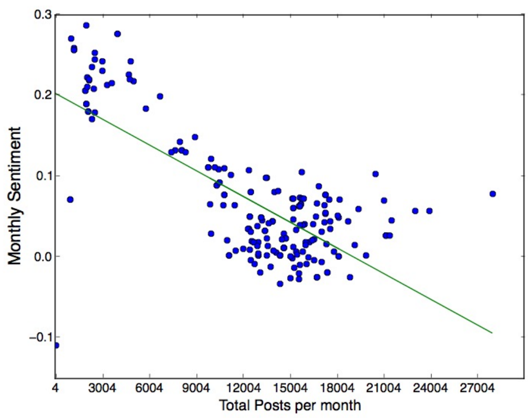


**Figure 2: Correlation between monthly average compound sentiment scores and total number of posts per month**

Subsequently, we aimed to explore the nature of the data set and discover how many users are systematically active and their respective sentiments. For that purpose, Stormfront users have been divided into 5 groups: lowest, low, medium, high and highest frequency level users, according to the total number of posts that they have contributed in Stormfront.

Overall, out of the 55.872 Stormfront users and 2 million posts, lowest frequency level users is the category of users that have participated in Stormfront only with 1 post, low frequency are the users with 1-10 posts, medium are the users with 10-200 posts, high frequency level are the users with 200-800 posts and highest frequency level users are the most active users with 800+ posts (Figure 3).

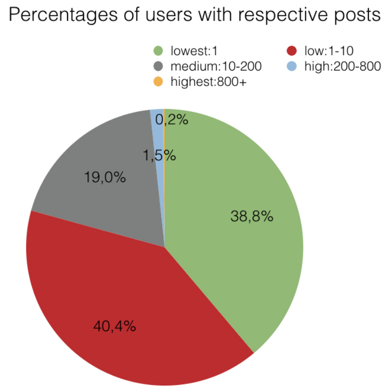


**Figure 3: Percentages of users with respective number of posts and usage frequency levels**

The above pie chart illustrates that the largest proportion of users who have contributed with 1 - 10 posts in Stormfront forum, was 40.4% (22580 users) with average compound sentiment score of those posts = 0.151, compared to the most active users that are significantly less, reaching the 0.2% of the total percentage (132 users) and have posted more than 800 posts per person, with average compound sentiment score of the posts = -0.030, which indicates more negative content.

Lowest frequency level users, those with only 1 post, hold the second biggest part of the pie 38.8% (21683 users) with average compound sentiment score = 0.16979. Followingly, moderate users with 10-200 posts, occupy 19% of the pie (10621 users), with average compound sentiment score = 0.07822, while high frequency level users with 200-800 posts, hold 1.5% of the percentage (856 users) with average compound sentiment score = 0.03521. (All user groups, posts range, number of users, percentages of users and average sentiment score of posts are provided in Appendix A).

Another experiment has revealed that no significant correlation could be found between the number of posts of every user and the average of the sentiment score of these posts per user (p = -0.01819576).

The traffic of Stormfront throughout the years is clearly increasing, starting with 27.145 total posts in 2002 (2001 data do not include all months), and growing to 161.324 total posts in 2014 (2015 data do not include all months) (Figure 4).

The following histogram shows the annual Stormfront traffic evolution, measured in number of yearly posts. The year with the highest traffic is 2009, with 222.111 posts. (Total numbers of monthly posts and yearly posts are provided in Appendix B).

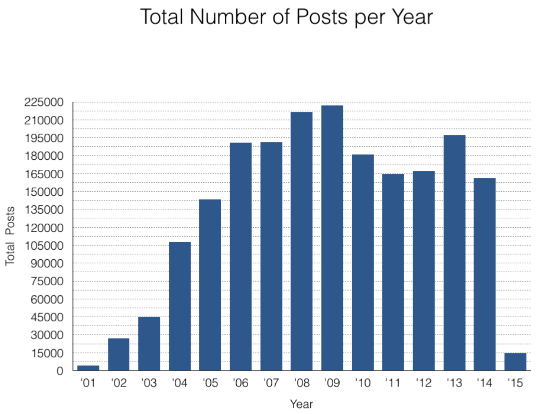


**Figure 4: Total number of posts per year**

Another interesting observation is that the aggregated monthly sentiment scores fluctuate widely over the years and tend to become more negative over time, forming the following Sentiment time line (Figure 5).

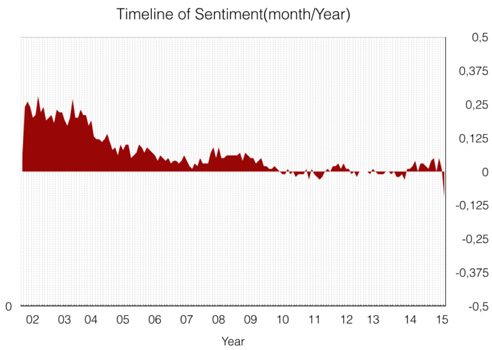


**Figure 5: Sentiment over time**

As research has shown, the months with the highest positive scores are March 2002 and June 2000 (with sentiment scores 0.2861 and 0.2052 respectively). The analysis of the automatically extracted entities reveals that the most discussed entities in both months, are Jews, Race and ethnicity in the United States Census, Aragorn[1], Peter the Great, Sieg Heil[2] and Sieg Fein, as well as Don.

Remarkable are the highest negative sentiment scores, which correspond to more negative content and the months are February 2015 and November 2013, with sentiment scores -0.1099 and -0.3395 respectively.

The most discussed entities within these two months were Sonderweg[3], Ahmad ibn Rustah[4],Rugii[5], Opel Omega and International court of Justice, for February 2015, as well as Jews, Race and ethnicity in United States Census, Iran, Gender and Adolf Hitler, for November 2013.

It is noteworthy that in January 2015 one of the most deadly and symbolic attacks occurred in France at the offices of the satirical magazine Charlie Hebdo in Paris, by Islamic terrorists [1]. Analyzing the most usual words in the content of the most active categories of February 2015, we can assume that this incident has affected the public opinion, contributing to one of the most negative monthly sentiment scores. Most usual words are: people, killed, murder, police, white, race, war, hate, equal, black, freedom, liberal, religion, free, national, right, kids.

Subsequently, Tf-idf measurement is applied in order to adduce which are the most important words that appear in one of the highest negative months, and do not appear in the rest of the months.

More specifically, Tf-idf shows which important words are included in November 2013, compared to all of the other months in our data set (the document corpus consists of the total content of each month's posts) [28]. The first 6 extracted words, descending from the word with the highest weight to the word with the lowest weight (importance), are the following: Canada, joins, move, felt, aggression, group, leading to assumptions about the content of the posts for this month.

What is remarkable is that the most active user of the whole dataset, appears as the most active user, within this month. He contributed 286 posts out of the total 14356 posts of November 2013, and his overall compound Sentiment score (derived from Sentiment Analysis at his overall posts) was -0.0045866 (Top 5 most active users, the number of their total posts, as well as the total compound Sentiment Scores of their posts, are cited in Appendix C ).

Furthermore, by gathering all posts that are connected with one entity and computing the average sentiment score of these posts we can gain insight on the extremeness of ideologies, by recognizing the most negative entities. Therefore, we can distinguish the entities that are associated with negative or positive sentiment and have a better insight on the content of this forum. Top-40 negative entities are shown in Appendix D (for example: Rosa Parks, Kermit Gosnell, Minstrel show, Miley Cyrus, Phil Robertson, Jack McMahon, School voucher etc.)

## 4.2 Seasonality

[1]fictional character from J. R. R. Tolkien's legendarium
[2]Nazi salute

[3]German Ideology and policy, Nazistic philosophy
[4]Persian explorer in the 10th century, important personality of Islam
[5]East Germanic tribe

In order to investigate whether seasonality affects the expression of ideas, firstly the average compound sentiment scores of the posts of every season were computed by the respective scores of the months which comprise each season. The results show that there is no considerable difference between seasons (Autumn = 0,0487107, Spring = 0,0435988, Winter = 0,045224 and Summer = 0,044475).

Secondly, the average sentiment scores of the posts that belong to each season of each year formulate the graph below which shows that the sentiment scores of all seasons fluctuate similarly, following approximately the same line over the years of our dataset (Figure 6).

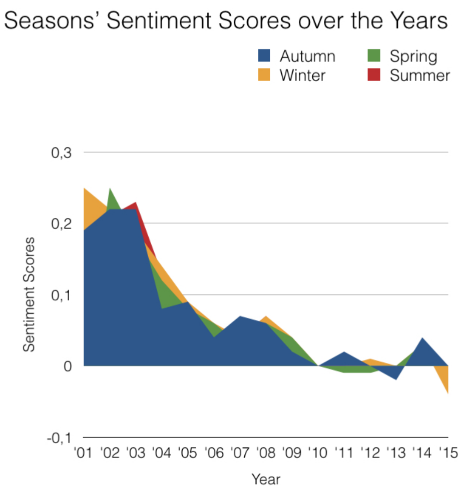


**Figure 6: Sentiment and seasonality**

### 4.3 Gender language differences

Analyzing the gender differences in expression of ideas, we compared a category known to attract mostly female forum members, "For Stormfront Ladies Only", with the general population, which consists of all the other categories of the Stormfront forum.

By extracting the content of posts of each category (women - general population) and applying sentiment analysis, the outcome of the aggregated compound sentiment score of the respective posts was 0.06120770 for women and 0.03346303 for the general population, revealing that women express themselves more positively than the rest of the population. The following word cloud shows the most frequent words in discussions on the "For Stormfront Ladies Only" category (Figure 7).

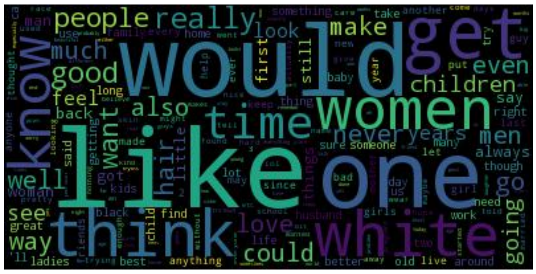


**Figure 7: Female users wordCloud**

The second word cloud refers to the general population category (Figure 8).

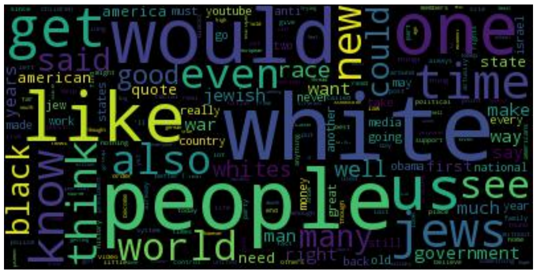


**Figure 8: General population wordCloud**

In both word clouds, the main (bigger) words are very similar, showing that both groups express themselves in a similar way. Female users' word cloud reveals that the words that appear most in their posts are: women, like, know, would, one, good, think, time and get. Respectively, in the general population word cloud, we see almost the same words. Terms that differ are: white, people, and world.

What is of great importance and discloses significant information, is that the slightly smaller and consequently less frequent words yet still prevalent enough to appear in the world clouds, reveal a lot about the essential differences in the forum discussions of these two groups.

Firstly, in the female users' word cloud, the terms of the specific category as compared to the general population are: hair, love, baby, proud, men, friends, life, better, child, home, racist, and feel. However, in the general population's word cloud these are: Jew, culture, never, hate, European, believe, hope, learn, little, white people, showing that women's posts often cover the topics commonly perceived as more feminine and talk about family, relationships, beauty tips and children. The conclusions are in line with the results presented in [32].

## 5. DISCUSSION AND CONCLUSION

The purpose of this research was to investigate whether sentiment analysis can be used for gaining insight into the structure and evolution of an online extremism forum, such as Stormfront. An additional aim was to explore the deeper gender differences and seasonal trends related to the sentiment of user-generated posts.

This research showed that a strong negative correlation was found between the monthly average compound sentiment scores of the posts over the specific time frame of our data, revealing that as time passes the posts are becoming increasingly negative.

Withal, another strong negative correlation between the average compound sentiment score of each month of each year over the number of monthly posts has shown that as the number of posts increase the more negative their content becomes. Therefore, there is indication that Stormfront has evolved over time, with an increasing number of posts and the content of these posts has become more negative.

According to Knigge the source of the problem of the intense extremism growth, can be identified in socio-political reasons. More specifically the numerous rapid changes in society, such as the mutation of political regimes, the dwindle of values, the decline in economy and more tangible examples such as the increase of unemployment or the rapid movement of masses of people to many countries, have contributed to the efflorescence of extremism [20]. This could explain the results of our analysis.

Another interesting finding of our research is that only 0.2% of the sample population corresponds to the highly active users, which is translated as 132 users. These users are contributing to the forum with negative posts. Furthermore, it is remarkable that 38.8% of all users, have participated in the forum with only 1 post, which raises questions about the way that the platform is used.

Furthermore, seasonality appears not to affect the expression of ideas, as the aggregated sentiment scores of the posts that belong to the seasons of each year fluctuate similarly, following approximately the same line over the years of our dataset. This leads to the conclusion that the expression of extreme ideas, is part of a deeply ingrained ideology rather than a temporary reaction [9].

Similarities and differences have emerged between female members and the general population. While the first impression was that both general population and women express similarly based on the word clouds that showed no significant difference, further investigation showed that women are more frequently involved in discussions about family, relationships, beauty and children. This is in agreement with having more positive scores derived through sentiment analysis compared to the general population scores.

This research can be improved with further investigation in some fields. For example, it would be interesting to investigate how Stormfront users react after various events, and whether such events have an interactive audience which could function as a mechanism for instigating subsequent reactions in physical world.

# 6. REFERENCES


[1] Charlie hebdo shooting. https://en.wikipedia.org/wiki/Charlie$_Hebdo_shooting, note = Accessed : 2017 - 08 - 15$.

[2] The digiactive guide to twitter for activism. http://www.digiactive.org/wp-content/uploads/digiactive$_twitter_guide_v1 - 0.pdf.Accessed : 2017 - 06 - 14$.

[3] The free dictionary by farlex. http://www.thefreedictionary.com/extremism. Accessed: 2017-06-02.

[4] Gender and democracy in computer-mediated communication. http://www.cios.org/EJCPUBLIC/003/2/00328.HTML. Accessed: 2017-06-14.

[5] How to handle seasonality. https://www.bjs.gov/content/pub/pdf/hhs-idasfcjts.pdf. Accessed: 2017-06-15.

[6] Stormfront.org. :https://www.stormfront.org/forum/. Accessed: 2017-07-07.

[7] Textblob: Simplified text processing. ://textblob.readthedocs.io/en/dev/. Accessed: 2017-06-16.

[8] Using vader to handle sentiment analysis with social media text. ://t-redactyl.io/blog/2017/04/using-vader-to-handle-sentiment-analysis-with-social-media-text.html. Accessed: 2017-06-16.

[9] I. Awan and B. Blakemore. Extremism, counter-terrorism and policing. pages 147–158, 01 2013.

[10] A. Bakliwal, J. Foster, J. van der Puil, R. O'Brien, L. Tounsi, and M. Hughes. Sentiment analysis of political tweets: Towards an accurate classifier. Association for Computational Linguistics, 2013.

[11] S. Ching and Y. Hsieh. Gender Differences in the Language for Emotions. *Department of Foreign Languages and Literature, National Cheng Kung University*, 2(1-4):89–97, 2007.

[12] E. G. Cohn. Weather and Crime. *The British Journal of Criminology*, 30(1):51–64, 1990.

[13] T. Z. B. M. H. M. K. M. P. H. Colley, A. Style and content in emails and letters to male and female friends. *Journal of Language and Social Psychology*, 23:369–378, 2004.

[14] W. W. Edgar Meij and M. de Rijke. In *In Proceedings of the fifth ACM international conference on Web search and data mining (WSDM '12).*

[15] P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha. Comparing and combining sentiment analysis methods. In *Proceedings of the First ACM Conference on Online Social Networks*, COSN '13, pages 27–38, New York, NY, USA, 2013. ACM.

[16] Z. Hailong, G. Wenyan, and J. Bo. Machine learning and lexicon based methods for sentiment classification: A survey. In *2014 11th Web Information System and Application Conference*, pages 262–265, Sept 2014.

[17] M. Hamm. American skinheads: The criminology and control of hate crime. 01 1993.

[18] C. d. L. R. v. N. J. R. F. S. C. S. Henry Bal, Dick Epema and H. Wijshoff. A Medium-Scale

Distributed System for Computer Science Research: Infrastructure for the Long Term. *IEEE Computer*, 49(5):54–63, 2016.

[19] E. Hutto, C.J. Gilbert. *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text (ICWSM-14)*. Eighth International Conference on Weblogs and Social Media.

[20] P. Knigge. The ecological correlates of right-wing extremism in western europe. *European Journal of Political Research*, 34(2):249–279, 1998.

[21] B. Liu. Sentiment analysis and subjectivity. *Handbook of natural language processing*, 2:627–666, 2010.

[22] B. Liu. Sentiment Analysis and Opinion Mining. *Synthesis Lectures on Human Language Technologies*, 5(1):1–167, 2012.

[23] G. P. D. C. A. K. Mike Thelwall, Kevan Buckley. Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, 61(12):2544–2558, 2010.

[24] C. Musto, G. Semeraro, and M. Polignano. A comparison of lexicon-based approaches for sentiment analysis of microblog posts. *Proceedings of the 8th International Workshop on Information Filtering and Retrieval co-located with XIII AI*IA Symposium on Artificial Intelligence (AI*IA 2014)*, pages 59–68, 2014.

[25] H. L. P. J. Newman M, Groom C.J. Gender differences in language use: An analysis of 14,000 text samples. *Discourse Processes*, 45(3):211–246, 2008.

[26] D. Odijk, E. Meij, and M. D. Rijke. Feeding the second screen: Semantic linking based on pages. In *10th Conference on Open Research Areas in Information Retrieval (OAIR '13)*, pages 9–16, 2013.

[27] B. Pang, L. Lee, and S. Vaithyanathan. Thumbs up?: Sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing - Volume 10*, EMNLP '02, pages 79–86, Stroudsburg, PA, USA, 2002. Association for Computational Linguistics.

[28] J. Ramos. Using tf-idf to determine word relevance in document queries.

[29] J. A. Schafer. Spinning the Web of hate: Web-based hate propagation by extremist organisations. *Journal of Criminal Justice and Popular Culture*, 9(2):69–88, 2002.

[30] P. Sedgwick. Pearson's correlation coefficient. *BMJ*, 345, 2012.

[31] F. S. Stefano Baccianella, Andrea Esuli. Sentiwordnet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. http://www.lrec-conf.org/proceedings/lrec2010/pdf/769$_Paper.pdf.Accessed$ : $2017 - 06 - 13$.

[32] I. G. Stevan Rudinac and M. Worring. Multimodal classification of violent online political extremism content with graph convolutional networks. In *Proceedings of ThematicWorkshops'17*, page 8, October 23-27 2017.

[33] T. Wilson, J. Wiebe, and P. Hoffmann. Recognizing contextual polarity in phrase-level sentiment analysis. In *HLT '05: Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 347–354, Morristown, NJ, USA, 2005. Association for Computational Linguistics.

# Appendices

### A. Proportion of users and sentiment

**Table 1: Pie chart detailed characteristics**

| UserGroups | PostsRange | NumberOfusers | PercentageUsers | AverageSentimentOfPosts |
|---|---|---|---|---|
| lowest | 1 | 21683 | 38.8% | 0.16979 |
| low | 1-10 | 22580 | 40.4% | 0.15133 |
| medium | 10-200 | 10621 | 19% | 0.07822 |
| high | 200-800 | 856 | 1.5% | 0.03521 |
| highest | 800+ | 132 | 0.2% | -0.03032 |

### B. Stormfront Traffic Over the Years

**Table 2: Number of Monthly Posts over the Years**

| Month | 01 | 02 | 03 | 04 | 05 | 06 | 07 | 08 | 09 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Jan | | 2115 | 3237 | 6681 | 10456 | 15589 | 16620 | 20472 | 23861 | 15110 | 17007 | 13494 | 16507 | 16226 | 14573 |
| Feb | | 1967 | 2053 | 7939 | 9750 | 15160 | 17511 | 15623 | 19330 | 15778 | 13066 | 14450 | 13742 | 17005 | |
| Mar | | 1955 | 2272 | 7389 | 10754 | 18151 | 16499 | 17216 | 18692 | 16388 | 15030 | 15251 | 16147 | 17537 | |
| Apr | | 2289 | 2480 | 7604 | 11407 | 15547 | 15802 | 15095 | 21457 | 15998 | 15573 | 15396 | 13935 | 17295 | |
| May | | 2467 | 3592 | 8282 | 10738 | 17218 | 15358 | 15822 | 22997 | 15192 | 14083 | 16681 | 17770 | 15613 | |
| Jun | | 1878 | 3916 | 8047 | 12455 | 15657 | 15152 | 15723 | 21086 | 17379 | 12651 | 14315 | 18078 | 12317 | |
| Jul | | 1976 | 2422 | 8915 | 10180 | 16430 | 13625 | 15168 | 21367 | 15351 | 13482 | 15000 | 19861 | 9916 | |
| Aug | | 2092 | 4700 | 9918 | 12311 | 18084 | 17171 | 17488 | 19121 | 12458 | 12485 | 12952 | 18814 | 10960 | |
| Sept | 895 | 1960 | 4771 | 10292 | 14184 | 13141 | 13864 | 17228 | 14706 | 12964 | 13362 | 11048 | 16705 | 12402 | |
| Oct | 1138 | 2503 | 4959 | 11180 | 13451 | 15912 | 16823 | 18035 | 14550 | 12910 | 12880 | 12695 | 15591 | 9859 | |
| Nov | 963 | 2978 | 4650 | 10786 | 13464 | 13173 | 15727 | 27955 | 12946 | 15539 | 12542 | 13534 | 14356 | 11500 | |
| Dec | 1128 | 2965 | 5732 | 10401 | 13998 | 16666 | 17189 | 20971 | 11998 | 15995 | 12290 | 12382 | 15977 | 10694 | |

**Table 3: Total number of Yearly posts**

| Years | TotalPosts |
|---|---|
| 2001 | 4.124 |
| 2002 | 27.145 |
| 2003 | 44.784 |
| 2004 | 107.434 |
| 2005 | 143.148 |
| 2006 | 190.728 |
| 2007 | 191.341 |
| 2008 | 216.796 |
| 2009 | 222.111 |
| 2010 | 181.062 |
| 2011 | 164.451 |
| 2012 | 167.198 |
| 2013 | 197.483 |
| 2014 | 161.324 |
| 2015 | 14.573 |

# Appendices

### C. Most Active Users

**Table 4: Most active users, total number of posts and Sentiment Scores**

| Users | Posts | SentimScores |
|---|---|---|
| revision | 19992 | -0,0045866 |
| SDY6401 | 12032 | 0,0049658 |
| ADAMANT | 11668 | 0,0224922 |
| kazan188 | 8990 | -0,0088649 |
| junkers88 | 8616 | -0,0089848 |

### D.Sorted entities

**Table 5: Top-40 entities associated with the most negative average compound sentiment scores**

| Entity | SentimentScore | Entity | SentimentScore |
|---|---|---|---|
| Rosa Parks | -0.9999 | Library of Alexandria | -0.9932 |
| Kermit Gosnell | -0.9999 | Mel Gibson | -0.9932 |
| Minstrel show | -0.9999 | Fiat money | -0.9834 |
| Miley Cyrus | -0.9999 | Fifth column | -0.9828 |
| Phil Robertson | -0.9999 | Kenya | -0.9823 |
| Jack McMahon | -0.9999 | Uganda | -0.9823 |
| School voucher | -0.9999 | Killas | -0.9812 |
| Blackface | -0.9999 | Norman Lear | -0.9651 |
| Abner Louima | -0.9999 | Lena Dunham | -0.9651 |
| Dora the Explorer | -0.9999 | Barbara Walters | -0.9651 |
| Don Lemon | -0.9999 | Harvey Weinstein | -0.9651 |
| Teach-in | -0.9967 | European American | -0.9573 |
| Greater Netherlands | -0.9962 | German Empire | -0.9547 |
| Uri Rosenthal | -0.9962 | Freemasonry | -0.933 |
| Julius Malema | -0.9962 | North West England | -0.9194 |
| Geert Wilders | -0.9962 | Neuro-linguistic programming | -0.9083 |
| African National Congress | -0.9962 | Reddit | -0.9062 |
| Carjacking | -0.9951 | Hurricane Katrina | -0.9052 |
| Apocalypto | -0.9932 | Skrewdriver | -0.8762 |
| Royal Navy | -0.9932 | James Traficant | -0.8577 |

### E. Seasonality and Extremeness

**Table 6: Pearson r Correlation scores, between General population sample (of 39 users) and Aggregated Compound Sentiment scores of Seasons**

| Season | CorrelationScore |
|---|---|
| autumn | 0,04807868 |
| winter | -0,05384657 |
| spring | -0,04180299 |
| summer | -0,07269894 |