2018 European Intelligence and Security Informatics Conference

Online Monitoring of Large Events

Johan Fernquist and Lisa Kaati
Decision Support Systems
Swedish Defence Research Agency
Kista, Sweden
firstnamne.lastname@foi.se

***Abstract*—In this paper, we describe an approach that can be used to monitor activity online that concerns large events. We propose six different tasks that can be used separately or in combination. The different tasks include analyzing messages from various actors, understanding the impact of messages to receivers, studying online discussions, analyzing hate and threats directed towards people and threats towards the execution of the large event and finally if there are any ongoing influential operations directed towards the general public.**

To illustrate how the approach can be used, we provide some examples of the different steps when monitoring online environments a few months before the Swedish general election in 2018.

***Index Terms*—monitoring, social media, intelligence, threat, hate, elections**

I. Introduction

Monitoring online environments for security reasons becomes more and more important due to the increased use of social media. However, monitoring online environments are difficult since it usually requires a lot of manual resources to provide a functional and useful analysis. One of the reasons for this is the large amounts of available data and the difficulties to decide what kind of analysis that should be done and what digital environments the analysis should focus on.

There are many different technologies that can aid analysts in their work. One of the most well-known fields is called sentiment analysis and opinion mining where the goal is to analyze peoples opinions, sentiments, evaluations, attitudes, and emotions from written language [12]. However, many other approaches and technologies can be used. Examples of other technologies that can be used for intelligence analysis are described in [3].

Social media provides researchers as well as law enforcement with large amounts of data and the ability to get insights into how different digital groups express themselves, evolve and respond to events in the real world. Computer-assisted technologies are nowadays not only used by computer scientist but also by social scientist [1]. However, in many aspects, social media is challenging to analyze. In many cases, it requires social scientists and computer scientists to work together. It is also the case that many technologies need labeled training data to perform reasonably well and this requires human resources.

This research is financed by the Swedish Civil Contingencies Agency.

Several different analysis methods have been used when studying elections and online activity. Researchers at the Oxford Internet Institute have developed one such approach. The approach includes analysis of what kind of information that is shared on social media and how bots are used to manipulate public opinion. A study describing the Swedish general election on Twitter is described in [7]. The study measures the amount of junk news shared on social media on Twitter before the Swedish election. The results show that the amount of shared junk news is higher than any other European country that has been studied.

Another approach that also focuses on Twitter and elections is described in [10]. Three elections are studied with the focus to predict the outcome of elections in India, Pakistan, and Malaysia. The study concludes that sentiment analysis using machine learning was the most accurate predictor of election outcomes of the methods used.

A. A method to monitor social media

In this paper, we describe a method to monitor online discussions, reactions, and impact in conjunction with events that take place in the real world. One of the benefits of using techniques for analyzing social media is that it can be used to make the analysis more objective and also provide support on what human analysts should focus on.

A systematic approach to monitoring online aspects could be beneficial for many researchers, analysts and data scientists. First of all, it provides a framework for what to study that, of course, can be extended. Secondly, it makes comparisons of the results possible. This means that in some cases it would be possible to compare the results between different events in the same country and perhaps also compare the results between different countries. One of the reasons for developing this method was the lack of a structured way to use technology to monitor digital aspects of large events among Swedish law enforcement.

We propose six different task for analyzing online environments that can be used separately or in combination. The different tasks include analysis of messages from different actors, the impact of such messages on the receivers, how to study online discussions, analyzing hate and threats directed towards people as well as towards the execution of the event and if there are any ongoing influential operations directed towards the general public. To illustrate how the method can

978-1-5386-9400-8/18/$31.00 ©2018 IEEE
DOI 10.1109/EISIC.2018.00015

be used, we provide some examples of how the different steps are used when monitoring some different online environment before the Swedish general election in 2018.

B. Outline

In Section II, we introduce and motivate six steps that can be used when monitoring and analyzing online aspects of an event. Both methodological and technological aspects are described. In Section III we show some examples of how each of the six steps was used to monitor the Swedish general election. We do not provide a full analysis of the Swedish election, the results presented here should be seen as examples of how the suggested steps can be used in a real case. Finally, Section IV provides a discussion of the method and some directions for future work.

II. Monitoring Large Events Online

The aim with this work is to describe an approach to monitor several different aspects of an event by analyzing editorial material from various actors, user-generated comments that can be seen as a reaction to the editorial material, online discussion forums, and Twitter. Each source needs to be analyzed independently due to the differences in how and why the data was produced and the format of the data. The approach is divided into six different tasks. Each task can be performed on a number of different digital environments. The tasks we suggest are:

1) Editorial messages
2) Editorial impact
3) Online discussions
4) Hate and threats
5) Threats towards execution of the event
6) Influential operations

Each task is described in detail below.

A. Step 1: Editorial messages

In conjunction with any significant event, different kinds of news media will inform as well as influencing people through the editorial material. There are many different ways to influence an audience with editorial material. Producing and spreading junk news and fake news is perhaps the most well-known way to influence the public. However, the editorial material does not have to be direct lies or untruthful - the actual choice of what editorial material to publish influences the readers and in some cases also the public opinion.

To get an understanding of what kind of editorial messages that are spread we suggest to focus on analyzing the editorial material that gets the largest spread and what kind of messages that are conveyed.

When studying editorial messages and the actual spread of editorial material, data from a number of different alternative media sources (within a certain time frame) can be used. Not only the data needs to be collected, but also information about the number of times each article has been shared needs to be collected. By selecting the n-most shared articles for each source (where n is the number of articles that should be analyzed) a selection of relevant articles that can be manually studied is obtained. This approach will only provide the articles that got most frequently shared. The number of articles that are selected for a manual analysis depends on the human resources.

B. Step 2: Editorial impact

The editorial impact or how the public responds to the editorial material is of interest to study since it can provide information about public opinion. To get an understanding of the impact of the editorial material feelings expressed by user-generated comments on the editorial material can be studied. This kind of analysis provides an estimation of the reactions on the editorial material.

Studying emotions in the commentary fields is one way to reveals the impact of editorial messages. Emotions are an important window for insight into one's thoughts and behaviors. Suggestions of emotions to study are anxiety and anger. Even though violence is not an emotion, it can still be relevant to study since it is a behavior that can be seen as a reaction to the emotions that we study. Anyone can show anxiety and anger, but different individuals may have quite different levels of anger in response to one and the same situation. These individual differences in how we respond to a situation make us unique. For that reason, emotions can be seen as part of an individual's characteristic traits, just like the personality, or perhaps as part of someone's personality. An interesting difference is that emotions are expressed almost exclusively in response to something in the environment.

To study emotions, a keyword-based analysis of text can be used. There are many limits using dictionary-based approaches. For example, it is not possible to detect linguistic nuances such as irony or sarcasm. Another drawback of using dictionaries is that words are context dependent. One word can have very different meanings, depending on how and where it is used. To create good dictionaries, each dictionary containing the keywords related to each category that is analyzed should be constructed by human experts (psychologist and computer scientist). In order to improve the coverage of the dictionaries, a word embedding can be used to suggest complementary terms to the experts. This can be done by simply computing the 15 nearest neighbors in the embedding space to each term in the dictionaries. For each term suggestion, the expert has the choice to either include or reject the term suggestion.

When studying editorial impact the comments of the most shared articles for the different news sites can be used. The comments corresponding to each article can be aggregated to one text and the number of occurrences of the words from each dictionary can be counted. To obtain an objective interpretation of the result, the results should preferably be presented both normalized (the percent of words from the dictionary) and the absolute value.

C. Step 3: Online discussions

Discussions that take place in both online discussion forums and commentary fields are relevant to monitor in relation to

large events. The discussions in the commentary fields are usually linked to the corresponding editorial material while the discussions in more traditional discussion forums typically are freer and can be invoked by a user.

For each of the commentary field and discussion forum that is studied, a top list of the most used words (after removing stop words) can be created. By using the top 10 of the most frequent words and count how often they occur in every comment the comments containing most frequent words can be selected. The idea behind this is to get comments that are representative of each digital environment. By letting the computer select a number of comments instead of a human analyst, the hope is that the analysis becomes more objective. The selected comments can be analyzed further manually but also be used to illustrate common discussions.

Another approach that can be used to study online discussions is to use topic models [2] to automatically detect topics. However, the results using topic models depends heavily on the data, something that is discussed in [15].

D. Step 4: Hate and threats

In conjunction with any specific events, public persons might be exposed to threats as well as too hateful comments. Monitoring and detecting hate and threat towards public figures is therefore important to get an adequate situational awareness and provide the right kind of security.

When studying hate and possible threats towards public figures several different methods can be used. However, it is important to note that detecting hateful comments and threats is a very challenging task. Several attempts to automatically detect hate messages in online environments have been made. One approach is described in [17] where machine learning coupled with template-based features to detect hate speech. Another approach is described in [18] where various types of linguistic features for detecting threats of violence is investigated. Approaches that consider hate directed towards individuals are described in [9], [13] and in [4].

E. Step 5: Threats towards the execution of the event

Some large events such as elections will always be exposed to risks when it comes to the execution. Therefore it is important to monitor signals of certain narratives related to threats of the execution with the aim to detect online discussions that might have an impact on the execution.

To detect threats towards an event, some understanding of what kind of threats to look for is needed. Threats can be related to things such as material and critical personnel necessary for executing the event, threats towards organizations working with the event and rumors regarding the actual execution of the event. A risk- and vulnerability analysis needs to be developed together with relevant experts.

One approach to detecting threats is to use a set of dictionaries representing the different threats. Each dictionary corresponds to a signal and a combination of signals indicate a threat. A signal could, for example, be words related to voting or words related to cheating. All texts that contain a combination of the two signals could be of interest for further investigation.

By extracting all texts (e.g. comments and tweets) containing combinations of signals a manual analysis can be performed. This is perhaps not the most efficient way to detect threats towards the execution of an event but it can at least restrict the amount of data that human analysts need to read.

F. Step 6: Influence operations

In conjunction with large events, there might be attempts to influence the public in different ways. Attacks that are coordinated with the aim to impact attitudes, behaviors, or decisions are commonly called influence operations or information operations. Usually, the use of the terminology implies that a foreign state coordinates the attack. Influential operations were for example debated during the American presidential election campaign in 2016 where Twitter revealed that Russia-linked accounts used Twitter to post automated material about the election.

Using automatic profiles in social media (bots) to influence and make an impact on peoples opinions and feelings is one way of influencing and for that reason it is of interest to analyze how and if bots are used to influence the online discussion in relation to the event.

There are many different definitions of ”bots”. In [8] bots have been defined as executable software that automates the interaction between a user and content or other users. In [6] a typology of bots is presented that extends the definition of bots used in [8] to include fake identities used to interact with ordinary users on social networks (sock puppets and trolls).

One way to study bots is to restrict the study only to consider the social platform Twitter. There are many different methods for detecting bots on Twitter. In [16] it is suggested that between 9% and 15% of active Twitter accounts are bots. Several different machine learning approaches to detect bots have been tried, and random forest is the classification algorithm that has proven to give the best performance for bot detection for the supervised problem when several different classifiers have been tested [5], [11], [14], [16].

III. A Study of the Swedish Election

We have used the six different steps described in the previous section to study the online preamble related to the Swedish general election that takes place in September 2018. This section provides some sample results from our study.

The dataset that we use in our analysis consists of a number of different digital sources: both partisan media, discussion forums, and Twitter. The partisan media that we study have editorial material (articles) and commentary fields. The different domains that we included in our analyzing are listed in Table I. We only consider data that was generated between March 1st and May 31, 2018.

As mentioned before, some examples of how an analysis of the different sources can be done are provided in the rest of this section. A complete analysis using the six steps described in this paper will be presented after the Swedish general election.

Domain	Description
Gatorna	A information portal for different Swedish autonomous milieus. Contains posts from different autonomous milieus and comments.
Samhällsnytt	A partisan media source with editorial material and user generated comments.
Nordfront	An partisan media that is related to the political party The Nordic resistance movement (NMR). The website contains editorial material and user generated comments.
Familjeliv	A Swedish discussion forum with an underlying focus on family. We study the sub-forum about Swedish politics.
Flashback	Sweden's largest discussion forum, we study the sub-forum about Swedish politics.
Twitter	Data related to the Swedish election and Swedish politics, including hashtags such as #svpol and #valet2018.

TABLE I
THE DIFFERENT DOMAINS INCLUDED IN OUR DATA SET WE USED WHEN STUDYING THE SWEDISH ELECTION

A. Editorial messages and impact

To illustrate how editorial messages and the responses to the editorial messages (step 1 and step 2) can be studied, we have chosen to exemplify it with the partisan nationalistic media Nordfront. The most shared articles during the time period and the corresponding emotions in the commentary field are presented in Table II. The emotions are presented in relation to each other if only anger is present in the comment field then anger is 100% while if both anger and anxiety are expressed equally often then anger and anxiety are expressed 50% each.

The most shared articles on Nordfront are about immigrants. This is not a surprise since the site is hosted by the Nordic resistance movement - a national socialistic political party. The first article is about a trial where five boys were convicted for a series of robberies. The second article is about a dog that got raped and died in a refugee camp in Greece. The article is a Swedish translation of an article from `www.voiceofeurope.com` (where it got almost 70 000 shares). Voice of Europe is a far-right website that reports stories from Europe that are negative for immigrants and the European Union.

The third most shared article is about a speech given by a former Swedish prime minister. The speech is interpreted by the editorial of Nordfront to be very immigration friendly and to encourage immigration to Sweden. A remade picture of the former Swedish prime minister where he is colored black accompanies the article. The former prime minister and his speech are not mentioned positively.

The reactions to the three different articles differ. In the first article, the comments expressed anger with comments that trials are unnecessary and instead someone should just put a bullet in their necks. The comments regarding the article about the dead dog in the refugee camp expressed anger but also violence. The comments regarding the article about the speech from the former Swedish prime minister expressed both fear, anger, and violence. How the distribution can be illustrated is shown in Figure 1. The violence is directed towards the former Swedish prime minister while fear is expressed regarding where the Swedish society is heading and who is in charge. It is worth noticing that there are very few comments to the articles and therefore the results should be interpreted with that in mind.

B. Online discussions on Flashback

We have chosen to exemplify analyzing online discussion (step 3) with an example from the discussion forum Flashback.

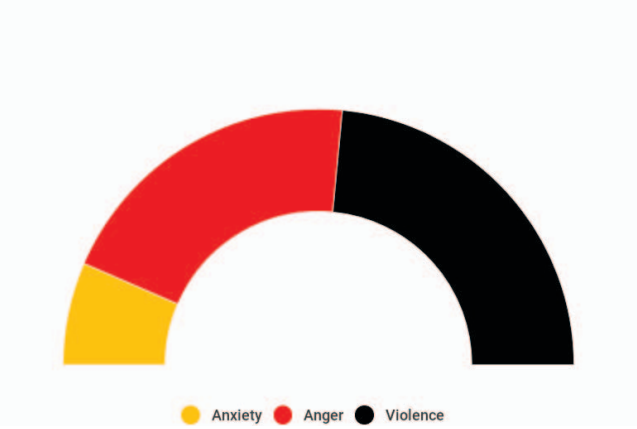

Fig. 1. Emotions and expressions of violence in the commentary field.

Flashback is a Swedish discussion forum consisting of 126 subforums. In our study, we focus on the subforum Flashback: Politik inrikes (domestic politics). Flashback is one of Sweden's largest discussion forums with the purpose of facilitating freedom of speech.

The ten most used terms (translated) in the Flashback domestic politics forum are:

- SD (Swedish democrats - political party)
- Sweden
- AFS (Alternative for Sweden - Swedish political party)
- year
- people
- S (Social democrats - Swedish political party)
- see
- party
- Swedish
- think

Using the most frequent words to select representative comments we get a set of comments that all discuss the Sweden Democrats (SD) or Alternative for Sweden (AFS). The quotes are all about immigration which is not a surprise since both SD and AFS are immigration critic political parties that focus on immigration. Using the most frequent words gives us an understanding of what kind of discussions that are most common in the digital environment. By selecting some comments to analyze in more detail, the aim is that the chosen comments are more objective than if they were manually selected. Hopefully, they are also more representative comments

Article title (translated)	Shares	Anxiety	Anger	Violence
The Africans behind the youth robbery in Nacka	641	0 %	100 %	0 %
Dog raped to death at refugee accommodation	238	0 %	50 %	50 %
Reinfeldt: The future comes from Africa	228	13 %	40 %	47 %

TABLE II
TOP 3 MOST SHARED ARTICLES FROM NORDFRONT AND THE DISTRIBUTION OF FEELINGS FOR EACH ARTICLE.

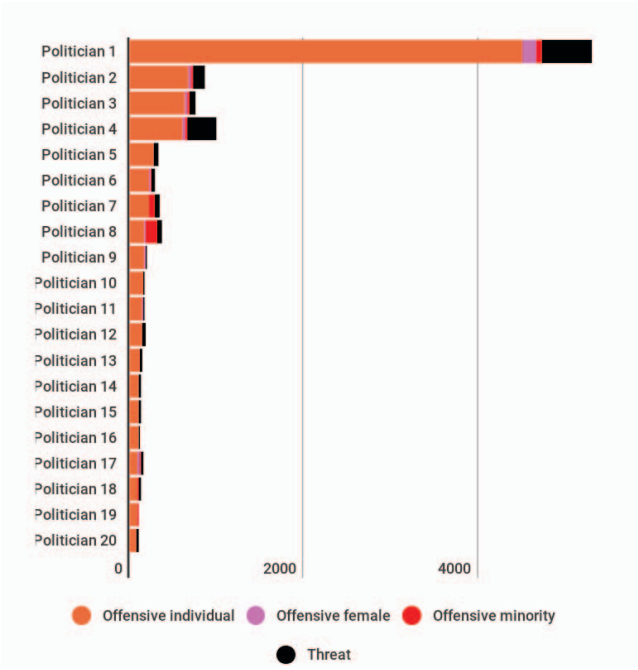

Fig. 2. The 20 most hated politicians in our study.

for the studied environment than if they were chosen randomly.

C. Hate and threats

To study hate and threats towards politicians, we analyze all data included in our study (see Table I) and measure hate and threats towards Sweden's 349 parliamentarians.

We use approaches similar to the one described in [9]. We have divided hate into three different categories: offensive comments directed towards an individual, offensive comments directed at women and offensive comments directed at minorities. We also study threats towards an individual.

Each category is represented by a dictionary of terms, as exemplified in Table III. In our study, we use Swedish data, but to ease understanding we have translated some of the words into English.

The hate intensity of the 20 most hated politicians is shown in Figure 2. As can be seen in the figure, one of the politicians has more hateful comments than all the rest of the politicians. It is also clear that most of the comments are hate directed towards the individual with comments regarding the looks or the intelligence of the politicians.

D. Threats towards the Swedish general election

To get a deeper understanding of the possible threats towards the execution of the Swedish general election we have collaborated with different experts. To detect threats towards the execution of the election, we use a dictionary-based approach where each dictionary correspond to a possible threat signal. The signals that we monitor in the case of the Swedish election are:

- disinformation regarding the election procedure
- discussions about voting
- discussions about ballot papers
- discussions about polling stations
- discussions about the counting of votes
- discussions about possible actions towards the election

As mentioned before, each signal consists of a dictionary with keywords. Our approach is to extract all comments that contain specific combinations of keywords from the different categories and manually study the comments.

In our analysis, we found that there are discussions (in the discussion forums and the commentary fields) about the voting procedure and in particular about how some political parties in previous elections have been cheating in the voting procedure. Some discussions mention cheating with the ballot papers and other cheating during the counting of the votes.

E. Influence operations and political bots

To get an understanding of the extent that political bots influence the Swedish election, we have collected data from all accounts that discuss the Swedish politics and the Swedish election (in our case this means that they use hashtags such as #valet2018 and #svpol in their communication).

The collected dataset consists of 406 163 tweets from 37 294 accounts collected between March 1st and July 31, 2018. Since we are not particularly interested in accounts that are connected to news sites that are automatic but are not trying to influence discussions, a white-list of known accounts that we have manually classified as genuine is created. When only consider genuine accounts and bots, the results show that around 7% (2 060) of the accounts are identified as bots according to our classification model [5]. The bots produced around 13% (49 409 tweets) of the content related to the Swedish election.

A number of tweets that we analyzed belong to accounts which, during the period of study, were suspended or deleted as a result of violating the end-user agreements. If we make the somewhat extreme assumption that all suspended/deleted accounts are automated, then 18% (6 822) of accounts are automated, and 17% (68 136) of the content is produced by automated accounts. There are many reasons for an account to

Category	Sample terms (ENG)	Sample terms (SWE)
Offensive individual	stupid, ugly, idiot	cepe, ful, idiot
Offensive women	whore, slut, bitch	hora, subban, slampan
Offensive minority	jew, nigger, assboy	jude, neger, stjärtpojke
Threat	kill, cut, stab	döda, kniva, hugga

TABLE III
DIFFERENT CATEGORIES OF HATE WITH SOME EXAMPLE WORDS.

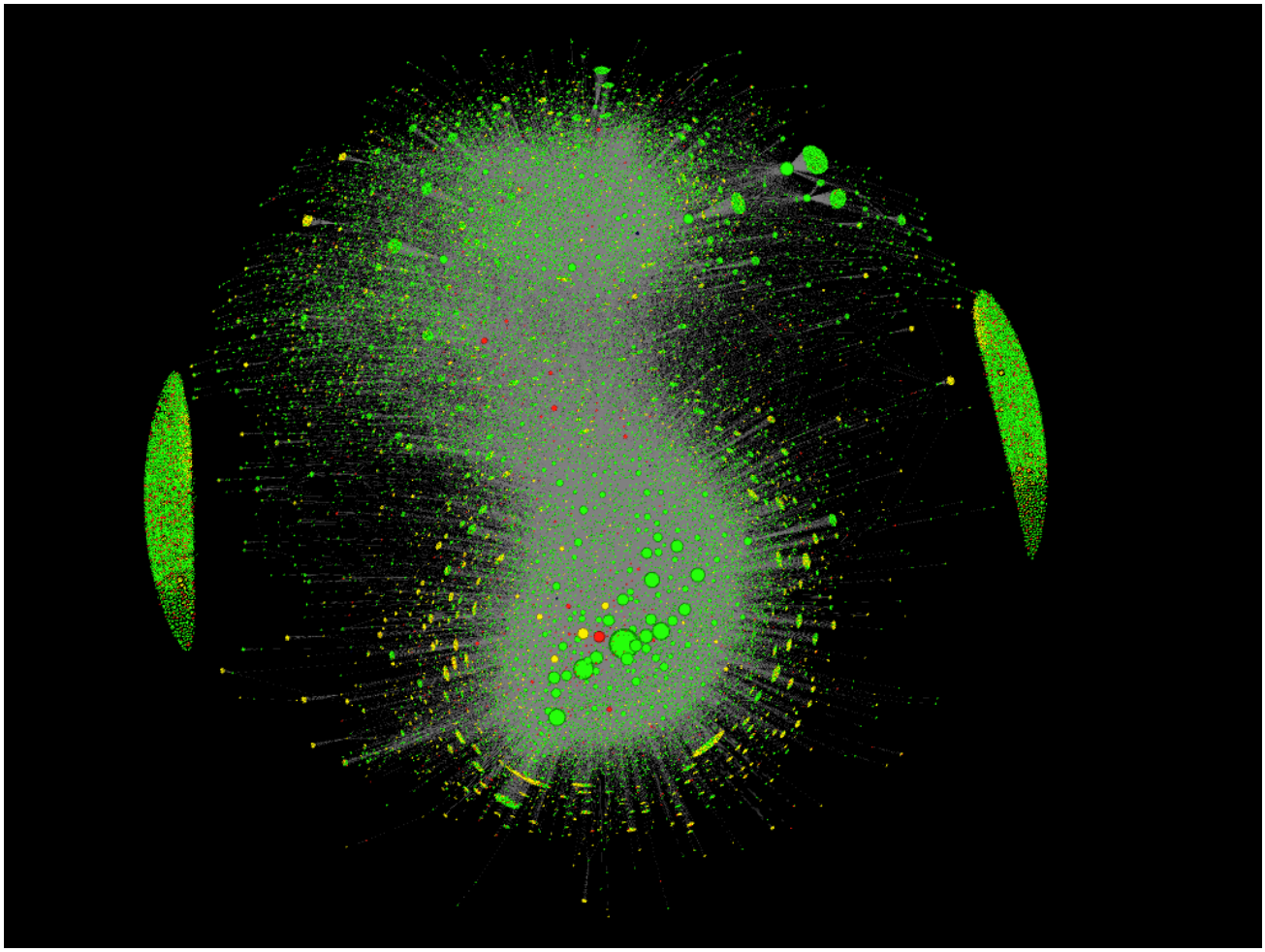

Fig. 3. Network of Twitter users discussing the Swedish election.

get suspended, one reason is that Twitter's terms of conditions are violated in some way.

To get an understanding of the impact of the different users that are tweeting about the Swedish election we visualized a network of users involved in discussions about Swedish politics. The result can be seen in Figure 3 where each node in the figure corresponds to a user. An edge between two users indicates that one of the users (or both) has retweeted tweets from the other. The size of the node corresponds to the number of outgoing edges, i.e., the bigger the node, the more users are retweeting that user. The color indicates whether the user is classified as genuine (green), bot (red) or suspended/deleted (yellow).

As can be noted in Figure 3 there are two clusters of users. By a manual inspection, we noticed that these two clusters could be divided into two clusters. Manual content analysis revealed that one cluster contains content with general political discussions. The other cluster contains content concerning migration and the negative consequences of migration. In the area between the two larger clusters, there is a set of accounts that are retweeted by accounts from both clusters. These accounts consist mostly of news media that position themselves as politically neutral. There are also clusters of nodes without edges on both sides in the figure. These are users who are tweeting but they do not get any retweets.

IV. Discussion and Future Work

In this paper, we have suggested six steps for monitoring digital environments in relation to large events. The proposed steps may be adapted to suit monitoring of any large event where it is of interest to analyze different digital milieus and their reactions. Many of the techniques that we have suggested are keyword-based, but we acknowledge that machine learning technologies usually perform better when analyzing data. However, since it requires training data the number of human resources that are needed to annotate data is in many situations not possible. Our suggestions for how to monitor digital environments and the technologies we have used to exemplify each step should be seen as a way to assist analysts and law enforcement agencies in their work.

We have illustrated how the six different steps can be used when analyzing online activity related to the Swedish general election. Based on the results from our analysis on Twitter data that is related to Swedish politics and the election, we notice that a lot of the discussions are related to immigrant criticism.

The analysis we present in this paper is not final. Instead, the examples provided are only present to illustrate how the different steps of method can be executed. For future work, we plan to release a full study on the Swedish election.

References

[1] B. Batrinca and P. C. Treleaven. Social media analytics: a survey of techniques, tools and platforms. *AI & SOCIETY*, 30(1):89–116, Feb 2015.

[2] D. M. Blei, A. Y. Ng, M. I. Jordan, and J. Lafferty. Latent dirichlet allocation. *Journal of Machine Learning Research*, 3:2003, 2003.

[3] J. Brynielsson, A. Horndahl, C. Kaati, L Mårtenson, and P. Svenson. Development of computerized support tools for intelligence work. *In Proceedings of the 14th International Command and Control Research and Technology Symposium (14th ICCRTS).*, 2009.

[4] M. ElSherief, V. Kulkarni, D. Nguyen, W. Y. Wang, and E. M. Belding. Hate lingo: A target-based linguistic analysis of hate speech in social media. *CoRR*, abs/1804.04257, 2018.

[5] J. Fernquist, L. Kaati, and R. Schroeder. Political bots and the swedish general election. *IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2018.

[6] R. Gorwa and D. Guilbeault. Understanding bots for policy and research: Challenges, methods, and solutions. *CoRR*, abs/1801.06863, 2018.

[7] F. Hedman, F. Sivnert, B. Kollanyi, V. Narayanan, L.-M. Neudert, and P. N. Howard. News and political information consumption in sweden: Mapping the 2018 swedish general election on twitter. Data Memo 2018.3. Oxford, UK: Project on Computational Propaganda, November 2018.

[8] P. N. Howard, S. Woolley, and R. Calo. Algorithms, bots, and political communication in the us 2016 election: The challenge of automated political communication for election law and administration. *Journal of Information Technology & Politics*, 15(2):81–93, 2018.

[9] T. Isbister, M. Sahlgren, L. Kaati, M. Obaidi, and N. Akrami. Monitoring targeted hate in online environments. *arXiv preprint arXiv:1803.04757*, 2018.

[10] K. Jaidka, S. Ahmed, M. M. Skoric, and M. Hilbert. Predicting elections from social media: a three-country, three-method comparative study. volume 0, pages 1–21. Routledge, 2018.

[11] K. Lee, B. David Eoff, and J. Caverlee. Seven months with the devils: A long-term study of content polluters on twitter. 01 2011.

[12] B. Liu. *Sentiment Analysis and Opinion Mining*. Synthesis Lectures on Human Language Technologies. Morgan & Claypool Publishers, 2012.

[13] B. Pelzer, L. Kaati, and N. Akrami. Directed digital hate. *IEEE International Conference on Intelligence and Security Informatics (ISI)*, 2018.

[14] M. Singh, D. Bansal, and S. Sofat. Who is who on twitter–spammer, fake or compromised account? a tool to reveal true identity in real-time. *Cybernetics and Systems*, pages 1–25, 2018.

[15] J. Tang, Z. Meng, X. Nguyen, Q. Mei, and M. Zhang. Understanding the limiting factors of topic modeling via posterior contraction analysis. In E. P. Xing and T. Jebara, editors, *Proceedings of the 31st International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 190–198, Bejing, China, 22–24 Jun 2014. PMLR.

[16] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini. Online human-bot interactions: Detection, estimation, and characterization. *CoRR*, abs/1703.03107, 2017.

[17] W. Warner and J. Hirschberg. Detecting hate speech on the world wide web. In *Proceedings of the Second Workshop on Language in Social Media*, LSM '12, pages 19–26, Stroudsburg, PA, USA, 2012. Association for Computational Linguistics.

[18] A. Wester, L. vrelid, E. Velldal, and H. L. Hammer. Threat detection in online discussions. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 66–71, San Diego, USA, 2016.