Title
Extracting Cognitional and Behavioral Information from Online Discussion Forum

Author
Tachaiya, Jakapun

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Extracting Cognitional and Behavioral Information from Online Discussion Forum

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Computer Science

by

Jakapun Tachaiya

March 2022

Dissertation Committee:

Professor Michalis Faloutsos, Chairperson
Professor Kevin M Esterling
Professor Vagelis Papalexakis
Professor Eamonn Keogh

The Dissertation of Jakapun Tachaiya is approved:

Committee Chairperson

University of California, Riverside

Acknowledgments

First and foremost, I would like to express my gratitude to my most fascinating advisor, Professor Michalis Faloutsos, for guiding me through the Ph.D. journey. His patience, support, and guidance play a key role in shaping my thought and perspective in tackling problems. Before meeting him, I solve problems like sailing a boat with no compass. There is no guarantee to reach a destination. Now, I learned the most important lesson is to have a clear goal and outcome in every step of the process. He does not only support me academically but also guides me in the other aspects of life. He was concerned about my well-being and health before, during, and after the pandemic. I genuinely feel that he care for all of his students, and I am grateful to have you as my advisor.

I appreciate all my committee members, Prof. Kevin M Esterling, Prof. Vagelis Papalexakis, and Prof. Eamonn Keogh, for their time and valuable research recommendations to complete my dissertation. Also, I would like to thank Prof. Kevin M Esterling for his insightful comments from a political science perspective. I have fun discussing a multitude of topics in our past collaborations with you.

I am very thankful to my family. Suporn Tachaiya and Narumon Tachaiya, my father and mother, are always concerned about my well-being. No matter what happens to me, I know they are always there to help me. Their everlasting support and unconditional love help me go through this journey. I am grateful to have them as my parents.

I also would like to thank my friends and lab-mates, Joobin Gharibshah, Ahmad Darki, Risul Isram, Omar Rokon, Sri Shaila G, Masud, Ben, and Arman. This incredible group of people supports me along the journey. It has been fun talking and working with you during the last five years. Also, I would like to thank Joobin Gharibshah, who guided me necessary skills, tips, and techniques to research since day one. I could not ask for better lab-mates.

I also want to express my appreciation to the Royal Thai Government who granted me a full scholarship to support this journey financially. Without this funding, I don't think I would be standing here as I am today. I am thank you for the opportunity you granted me.

Finally, I would like to allocate this paragraph to thank everyone that has come into my life, good or bad. They teach me and shape me to become a better version of myself. I am grateful, and thank you.

ABSTRACT OF THE DISSERTATION

Extracting Cognitional and Behavioral Information from Online Discussion Forum

by

Jakapun Tachaiya

Doctor of Philosophy, Graduate Program in Computer Science
University of California, Riverside, March 2022
Professor Michalis Faloutsos, Chairperson

How can we recognize users' cognition and identify behaviors in an online discussion forum? We see that online discussion forums constitute an untapped opportunity for understanding cognitional and behavioral information. Mining this publicly and freely available information can significantly benefit analysts as it helps reveal trends, behavior, and even bad actors. This thesis aims to answer the following problems. First, we identify and characterize thread-centric behaviors where the key novelty lies in an unsupervised model to recognize behaviors without requiring prior forum knowledge. The model reveals some fascinating abusive behaviors appearing in the forum. Second, we develop an aspect-based sentiment analysis model, a powerful state-of-the-art transformer-based model to detect sentiment toward specific aspects in posts. The model also helps quantify the effect of the real-world event on users' sentiment in the online forum. Third, we develop a stance detection model to recognize the user's position toward topics of interest and quantify the correlation of sentiment and stance conditioning to the events. Our finding on the relationship between sentiment and stance redefines how an analyst perceives this cognitional information. The

contribution of our work can be summarized in threefold: (a) collect, analyze, and profile thread-based behaviors, (b) detect sentiment toward specific topics in response to real-world events, and (c) infer cognitional information and understand the relationship of sentiment and stance at the events. We see our systematic approaches and tools as a significant step towards cognitional and behavioral understanding in online discussion platforms.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

How can we recognize users' cognition and identify behaviors in an online discussion forum? This overarching problem is the main drive behind this thesis.

The social media platform or online discussion forums become a part of everyday life. People reach out to these platforms for various reasons, starting from exchanging information, interacting with others, and even openly venting their frustration. Hence, such forums and media capture how people perceive or have opinions toward miscellaneous issues in society. This work makes use of freely and publically available datasets on online discussions such as Reddit to a) quantify users' cognition in terms of sentiment and stance and b) recognize common and abnormal online behaviors. Here, we specifically focus on the problem of analyzing textual input from posts and threads made by users as it can represent users' cognition and activity toward the issues.

To answer the main question, we essentially address the following sub-questions. a) "How can we recognize behaviors without prior knowledge of the dataset," b) "How can we quantify user's cognitions on both sentiment and stance toward the topic of interest?," and c) "Do we observe the correlation between sentiment and stance," We rely on online discussion forums to answer the above sub-questions. The desired outputs are as follows: a) clusters of different behaviors separated by characteristics, b) cognitive measuring models that can capture sentiment and stance toward specific aspects, and c) a statistical model to measure the correlation between sentiment and stance.

Studying the online discussion forums accompanies a set of challenges. First, data in an online platform are unstructured and occasionally contain some typing mistakes. Since we have to deal with user-generated content, cleaning and fixing text have to be done correctly before proceeding to the other steps. Second, there is no publicly available labeled dataset fitting our task. It challenges us to create or use adaptation techniques to prepare a new dataset for the specific models. Third, length and ambiguity in a post require us to handle those issues carefully with natural language techniques.

There are limited works for the problems as described above. We group prior efforts into two main categories: a) characterizing and understanding online behavior, and b) detecting and identifying sentiment and stance on online users. First, a few studies studies [98, 35] use machine learning and statistical models to detect behavioral trends and anomalies in social media platforms, such as Facebook and Reddit. Second, there are many works [74, 47] trying to capture overall sentiment and stance on text. Here, we focus on the aspect-based method, where we associate sentiment and stance to the specific topic or keyword in the post. Some of the relevant studies include an effort [30], which applies aspect-based sentiment analysis with a neural network. The recent trend of the model shift toward a transformer-based approach, as we can find in [104]. Apart from detecting techniques, researching communities try to understand the relationship between sentiment and stance. On one side, [37, 17] suggests the use of sentiment as one of the key features to detect stance. On the other side, the work [18] suggest otherwise where they propose sentiment cannot proxy stance.

The main contributions in this thesis are three. First, we let online behaviors emerge with an unsupervised method. We propose a comprehensive unsupervised clustering approach with a powerful visualization component to reveal distinct behaviors on forums. Second, we propose systematical methods to quantify sentiment and stance towards the aspect of interest. Third, we redefine the correlation between sentiment and stance by anchoring them at the event.

Our key results can be summarized in the following points.

1) We identify patterns and anomalies following a thread-centric angle. We find that the overwhelming majority of threads exhibit log-normal and ephemeral properties where they are short-lived and contain only one post. We also detect abusive threads that exhibit "Search Engine Optimization" properties containing a large amount of incoherent text and many URL links to a few sites.

2) We develop a systematic approach to detect sentiment toward specific aspects in response to the events. We customize and synthesize state-of-the-art methods to optimize aspect-based sentiment analysis and stance detection using the unstructured data of political discussion forums to achieve high accuracy to 74% of accuracy in three classifications. The model's performance can increase to 81.1% if we only examine short posts with less than 23 words.

3) We capture cognitional information and redefine the correlation between sentiment and stance by anchoring them at the event. We proposed a comprehensive approach to detect and quantify the change in cognitional information for both sentiment and stance. In addition, we find that observing sentiment and stance on aspects at the critical event. An against stance is a reasonable proxy for gauging Negative sentiment. 70% of all posts with Against stance towards an aspect is likely to have a negative sentiment towards it for the vast majority of the days.

1.1 Road map

This dissertation consists of the following chapters. Chapter 2, Rthread, capture and analyze thread-centric behavior with the matrix decomposition in an unsupervised manner presented in ASONAM 2020[95]. Chapter 3, RAFFMAN, introduces a systematic approach to quantify sentiment with aspect-based sentiment analysis as well as to detect the evolution of sentiment change addressed in ICWSM2021 [94]. Chapter 4, SentiStance, present an approach to stance detection, specifically in political discussion. We also explore the correlation between sentiment and stance by anchoring them at the event shown in ASONAM 2021 [96]. Chapter 5 concludes the work.

The overarching goal is to provide a comprehensive and flexible approach to extracting users' cognition and detecting online behavior in the discussion forums. We see our work as practical tools and methods to bridge the gap in understanding the overwhelmed realm of information in online forums.

Chapter 2

Rthread: A Thread-Centric Analysis of Security Forums

2.1 Introduction

In this section, we want to identify and analyze interesting behaviors of threads in computer security forums without a need for prior knowledge in the forum's data. Our goal here is to conduct an in-depth thread-centric analysis of online discussion forums. Several recent works have shown that there is a plethora of valuable information we can exploit in these forums [39, 84, 98]. To a large degree, the information is fascinating, because of the wide spectrum of users that engage in these forums. They range from benign users that mainly discuss tips and tools for how to protect themselves from cyber attacks, all the way to hackers who sell hacking tools and services in exchange for money.

We propose, Rthread, a comprehensive thread-centric analysis approach with unsupervised co-clustering and powerful visualization capabilities. Our approach is: (a) **comprehensive**: it combines 92 features that span three types of features, including temporal, behavior and text; (b) **unsupervised**: it does not rely on training data and can uncover unexpected phenomena; and (c) **interpretable:** it provides an intuitive and visual interpretation of the resulting clusters. Our key results can be summarized in the following points:

1. **We develop a comprehensive soft co-clustering approach.** We opt for soft co-clustering using the extensive set of features mentioned above. Our co-clustering does **feature selection and clustering simultaneously** by identifying the most appropriate set of features per user cluster. In addition, we develop a powerful visual way to capture the essence of each cluster, as shown in Fig. 2.3 for the Offensive Community forum.

2. **We identify clusters with surprising behaviors**. Our unsupervised approach categorizes threads into clusters with different behaviors, which we outline in Table 2.3. Among them, we identify two surprising clusters of threads. First, we find a cluster of "SEO" threads, which contain a large amount of incoherent text and many URL links to a few sites. Second, we identify "hidden" threads, which require users to register and post a reply to see the content of the post that initiated the thread.

3. **We identify persistent thread properties.** We find six properties of threads that follow a log-normal distribution with parameter values that are persistent over many years and comparable across many forums as we see in section 2.2.

	Year	Users	Posts	Threads
Wilders Security	2013-2016	2,213	81,561	28,661
Offensive Community	2013-2016	5,413	24,856	3,542
Kernelmode	2010-2018	1,442	25,024	3,144
Greysec	2015-2018	433	6,969	1,239
Garage4hackers	2010-2018	873	7,697	2,096
StresserForum	2017-2018	763	7,065	704
Raidforum	2015-2018	28,731	214,239	33,322
Safeskyhacks	2013-2018	7,379	26,842	12,892

Table 2.1: The collected security forums.

2.2 Dataset and Persistent Properties

Here, we discover fundamental and persistent thread-centric properties using capabilities from our platform.

Data. We study eight security forums which contain data ranged between 2010 and 2018 with the total of 47,000 users, 400,000 posts and 85,000 threads shown in Table 2.1. The data comes from two main sources, our automated crawler and Cambridge Cybercrime Centre [3]. WilderSecurity [15] and Kernelmode [7] are considered to be *white-hacker forums* attracting IT professionals. By contrast, Offensive community [8], Garage4hackers [4], and Raidforums [11] are mainstream *dark forums*, where people often share tools and knowledge for hacking into systems. The rest of the forums, Greysec [5] Stresserforums [14] and Safeskyhacks [12] are in an in-between grey area.

We consider six thread-centric features: (i) *the number of new threads per day*, (ii) *the number of active threads per day*, (iii) *the thread lifespan*, (iv) *the number of active days in a thread*, (v) *the number of posts in a thread*, and (vi) *the number of users in a thread.*

Persistent log-normal distributions over the years and across forums.

(a) Original data plot

(b) The fitted log-normal

Figure 2.1: Many thread properties exhibit log-normal distribution which is persistence in its parameters and over time in several forums. Showing the distribution of the number of active days of a thread in CCDF (log) for Wilder Security.

Five of the above features (except the number of active threads per day) exhibit a heavy-tail distribution especially pronounced in the large mainstream forums such as `Kernelmode` and `WilderSecurity` shown in Fig. 2.1. The CCDF of thread-centric features from (i) to (vi) in a log-log scale can be fitted with the log-normal distribution:

$$X = e^{\mu+\sigma Z} \tag{2.1}$$

where Z is standard normal variable, μ is a location parameter and σ is a scale/shape parameter.

Interestingly, the distribution parameters are fairly stable across years for each forum with a variance less than 0.04 for μ and 0.01 for σ

Type	Description	#features
Temporal	Temporal feature capture thread properties in time domain, e.g. lifeSpan, #activeDays and dayEntropy.	3
Behavioral	Behavioral feature tell how users interact within threads through posts, e.g. #posts, #users, threadDistribution, userEntropy, and userEngagement.	24
Textual	Describing the content of the threads and their posts: #words, #characters, #lines, #URLs and #email, including intention/topic related features, e.g. asking words, thanking words etc.	65

Table 2.2: Overview of the 92 features used in clustering.

2.3 Unsupervised Thread clustering

We propose a comprehensive and systematic way to cluster threads into different categories in an unsupervised learning fashion. We consider two clustering methods here.

a) K-Means, a standard unsupervised algorithm [44] that partitions threads into distinct non-overlapping clusters where each thread belongs to only one cluster. We mainly use K-means as a reference.

b) The soft co-clustering with Sparse Matrix Regression or SMR method [27]. It allows an overlapping co-clusters membership meaning that thread can belong to more than one cluster.

Our future work will consider more techniques, including a hierarchical and AutoEncoder-based [103] clustering.

Features. We use a total of **92 features** as shown in Table 2.2 that can be grouped into behavioral, temporal and content related. Most features are self-explanatory on the features' names shown in Fig. 2.3 Below is a list of critical features in detail.

- lifeSpan; the number of days between the first and the last post in a thread.
- #activeDays; the number of days that a thread generates at least one post.
- dayEntropy; the average entropy of the post distribution on threads in the day which the threads are active.
- userEntropy; the entropy of a distribution of posts made by each user on a thread.
- userEngagement; the number of days between the first post and the last post made by users.
- userAppearance; the number of unique days which users make a post in a thread.
- threadDistribution; the distribution of posts in a thread which captures if the thread is equally distributed, front-loaded or back-loaded.

Clustering algorithms. We assume that K-Means is a well known algorithm, so we will only discuss the soft co-clustering approach, SMR. Given a matrix X of threads with 92 features, the soft co-clustering via SMR can be posed as the minimization of the loss function [27]:

$$||X - AB^T||_F^2 + \lambda \sum_{i,k} |A_{ik}| + \lambda \sum_{j,k} |B_{jk}| \tag{2.2}$$

where A and B are matrices of size I x K (threads x clusters) and J x K (features x clusters), respectively. Matrix A is a result of the thread co-clustering algorithm where it shows a cluster to which the threads belong. Matrix B is a byproduct of SMR where it reflects features' clusters. Value in both matrices A and B which we will refer to as **Intensity** value is not a zero or one but can be any value in between. K is a parameter that determines the number of clusters, and parameter λ controls how we calculate the relevance of a thread for each co-cluster. As we increase λ, we get sparser results, namely, fewer threads per cluster.

We experimented with $\lambda = 0.1$, 0.2 and 0.4 and we selected $\lambda = 0.1$, which works well here.

Clustering inclusivity. In soft co-clustering, the algorithm allows overlapping members: each thread can belong to more than one cluster. The algorithm provides the **Intensity** value for each thread and cluster pair, which captures how strongly related is the thread with that cluster in matrix A and feature with that cluster in matrix B. The higher the number, the stronger relation is.

To assign threads to clusters, we use the **Intensity Threshold**: only threads with Intensity value above the Intensity Threshold will be included in that cluster. In more detail, we compare (and normalize) the Intensity of each thread with respect to the maximum observed Intensity across all threads for that cluster, thus the threshold becomes a percentage of the highest observed Intensity for that cluster. We evaluated the following values 5%, 10%, 20% and 40% of the maximum threshold. The higher number of threshold result fewer the member in each cluster. Note that the same reasoning applies for assigning features to a cluster. We use the same 20% threshold for assigning features to clusters, which gives good results here.

In this paper, we consider three algorithms:

1. K-means using the full set of 92 features.

2. SMR with a 20% Intensity Threshold

3. K-means-42 using the subset of 42 features that have intensity value more than 20% Intensity Threshold from matrix B in SMR.

The third algorithm was introduced to answer the following question: would K-means perform better, if we select the more discriminating features that we identify with our SMR algorithm? Also note, that with soft co-clustering, a thread can belong to multiple clusters. To compare SMR with K-Means, we associate each thread with the cluster for which the thread has the highest Intensity value.

A. Evaluating the clustering. Our main goal here is to profile thread-centric behavior without prior knowledge of the forums. Hence, we don't have any labeled dataset to rely on and the standard clustering evaluation strategies like accuracy and F1-score cannot be employed here. We turn to the average Silhouette coefficient [83] to evaluate the clustering quality. The coefficient measures how similar is each thread to its assigned cluster compared to other clusters. Its value ranges from -1 to 1, and the higher the co-efficient value the better the clustering is. We measure the average Silhouette coefficient for each forum as a function of the number of target clusters as shown in Fig. 2.2, which we discuss below.

A.1. Selecting the right number of clusters. This is a key question in every clustering problems [105]. For now, this parameter is provided by the end-users, which empowers them to tailor the query to the question of interest. In Fig. 2.2, the knee of the

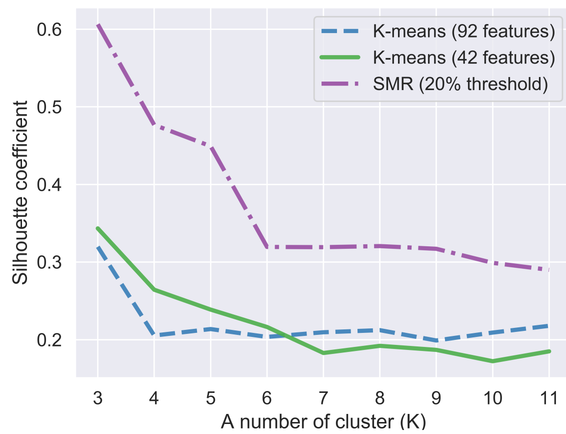


Figure 2.2: Silhouette coefficient between a number of cluster and methods from K=3 to K=11. It shows that soft co-clustering perform almost two times better than K-means.

curve appears between 4-6 clusters, which is the range that we used.

A.2. Soft co-clustering outperforms the K-Means algorithm. From our experiments in Fig. 2.2, we find that the SRM co-clustering has almost double the Silhouette coefficient of compared to both K-mean algorithms (using 92 and 42 features). The poor performance of K-Means could be partially attributed to the large number of features. To address this, we identify a "better" set of features with higher discriminatory capability, namely, 42 features that have Intensity value more than 20% in SMR (K-Means-42). However, this did not improve the results: K-Means-42 does not exhibit consistent or statistically-significant improvement as shown in Fig. 2.2.

B. A Visual and Intuitive cluster analysis: To facilitate the interpretation of the clustering results, we propose the use of color-coded table as shown in Fig. 2.3. In this plot, we calculate the mean value of each feature over all the threads for each cluster. Dark blue indicates low values, while dark red indicates high values. We demonstrate the power of the visualization in Fig. 2.3, where we show the clustering of Offensive Community for four clusters. In a figure, on the top left corner, we see dark blue, which suggests that

Type	Description
Ephemeral	One post and live for one day.
Long-lived	Long lifespan and high # of active days.
Hidden	Hide some part of their contents.
Long-post with URL	Threads with posts containing URLs and high # of words.
SEO	Threads with posts with repeating URLs and high # of incoherent words.

Table 2.3: The different types of clusters identified by our unsupervised learning methods.

threads in cluster 1 have low number of users, posts, lifespan and active days. This cluster represents the large "low activity" threads, which is aligned with the skewed distribution of the section 2.2. Similarly, cluster 4 consists of long-lived threads with many user contributors. This group corresponds to the "heavy hitter" threads at the tail of the skewed distribution of the previous section.

2.4 Identifying interesting clusters

Here, we apply our approach on our security forums in order to provide an indication of the types of results we could derive. Specifically, we used our soft co-clustering approach with 20% Intensity Threshold, 0.1 λ and K = 4 on forums.

We identify groups of threads with distinctive behaviors, as we show and define in Table 2.3. We discuss each type of cluster and its behavior below.

a. Identifying "Ephemeral" threads. For every forum, our clustering identifies a large cluster of primarily short lived threads, which one could have anticipated given the the skewed distribution in the size of the threads in section 2.2. We use the term

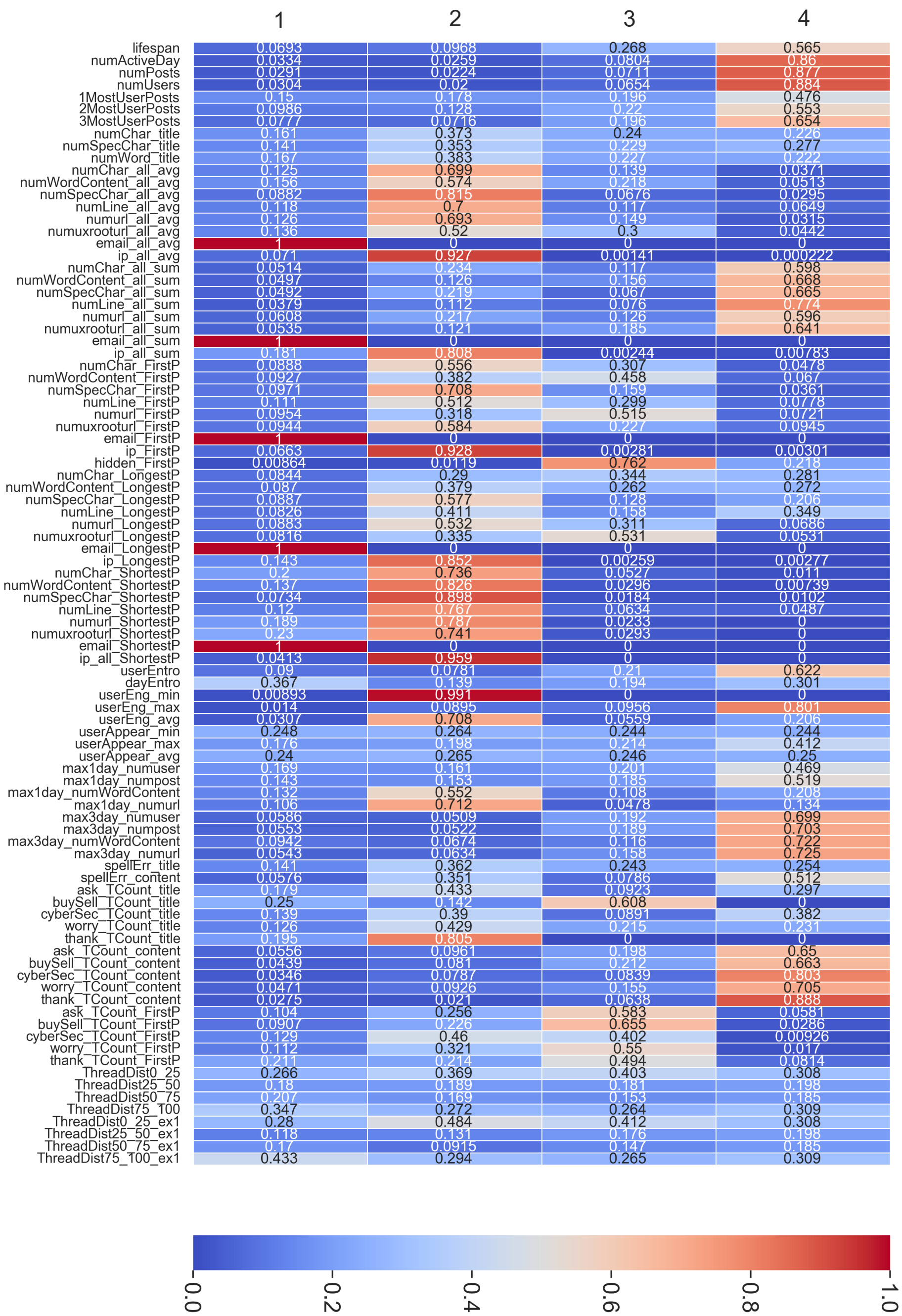


Figure 2.3: Cluster Visualization for in Offensive Community: The color-coded average feature value per cluster captures the differences among the clusters visually and intuitively. This clustering is derived by using SMR (20% Intensity threshold and K=4). We find a) cluster 1 displays "ephemeral" behavior, b) cluster 2 is recognized as "Long-post with URL," c) cluster 3 is "hidden" threads, and d) cluster 4 is "long-lived."

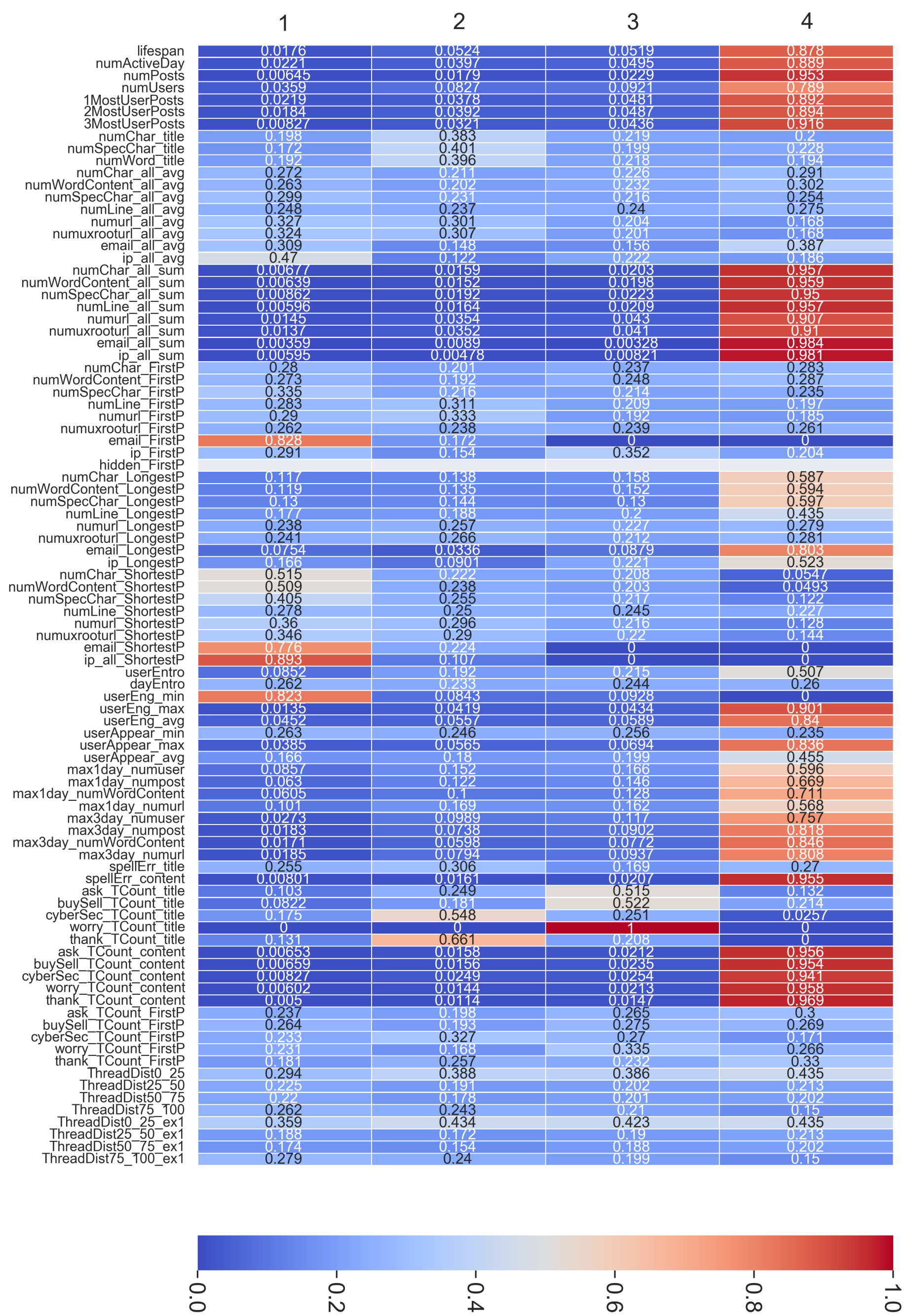


Figure 2.4: Cluster visualization for Wilders Security with SMR (Intensity threshold 20% with K=4). We find a) clusters 1,2, and 3 exhibit "ephemeral" behavior as the majority of threads, and b) cluster 4 is "long-lived."

“ephemeral” to refer to a threads with only one post. We find clusters dominated by such “ephemeral” threads in all forums. These clusters can be observed by our mostly blue colors in the top part of the plot, shown in cluster 1 for Offensive Community in Fig. 2.3 and cluster 1,2 and 3 for Wilders Security in Fig. 2.4.

b. Identifying “Long-lived” threads. Some of the emerging clusters seem to be dominated by threads of long life-span (difference between first and last post). In fact, some of these threads span four years! Most of these threads are sharing information, discussion of technologies and announcements. We are able to recognize Long-lived clusters, which can be observed by our color-coded tables, shown in cluster 4 for Fig. 2.3 and Fig. 2.4. Long-lived threads appear in almost every forum, but as a small percentage of the total number of threads, which is aligned with the skewed distribution seen in section 2.2.

c. Identifying “Hidden” threads. In Offensive Community, we found a cluster of threads that hide their content. These threads are always initiated by a post that requires the viewer to register as a member in the forum and post a reply to see the hidden content. It is natural to assume that this technique hides the content from an automated crawler, which will, most likely, not perform the unlocking behavior. In more detail, this cluster consists of 30 threads all of which are initiated by such a “hidden” post. Most of the replies are short “thank you” posts. The short first post, the keywords in the post, and the short “thank you” replies, are the characteristics of the threads, which our algorithm used to form the cluster.

What do these threads hide? Intrigued, we investigated 30 of these “hidden” threads. We responded with a post, and we got access to the hidden information. We

Type	Note	#posts
Hacking tool	Rooting Android and a key-logger	3
Hacking tutorial	Range from server vulnerability to phishing for credit cards	8
Illegal dist	Games & other software	4
Selling/buying	Rooted accounts, websites and shell scripts for hacking	2
Boasting	Bragging about their hacking success	3
Benign tutorial	Web & Windows app's tutorials	3
Benign tool	Web & Windows plugins and tools	4
Sharing info	News & tips in computer security	3

Table 2.4: The types of "hidden" threads.

found several questionable content, including hacking tutorials, hacking tools and illegal distributions of cracked software, as we list in Table 2.4. Also, one of those posts is a boasting post about their achievement of hacking into some well-known systems, such as Google's Morocco server in 2013.

d. Identifying "Long-post with URL" threads. In some clusters, we saw threads containing a moderate number of words and URLs in their posts. We use the term "Long-post with URL" to describe such threads. Most of these threads are sharing news and some information with one or more hyperlinks. These hyperlinks point to a source of news, an image file or a file-sharing sites. In our analysis, we find "Long-post with URL" clusters in three forums, Garage4hackers, Safeskyhacks and Offensive Community (cluster 2 in Fig. 2.3).

e. Identifying "SEO" threads. In Safeskyhacks, we identified a cluster of threads, which we suspect engage in Search Engine Optimization (SEO) boosting. Specifically, we find that cluster 3 of Safeskyhacks is identified as a "Long-post with URL" type.

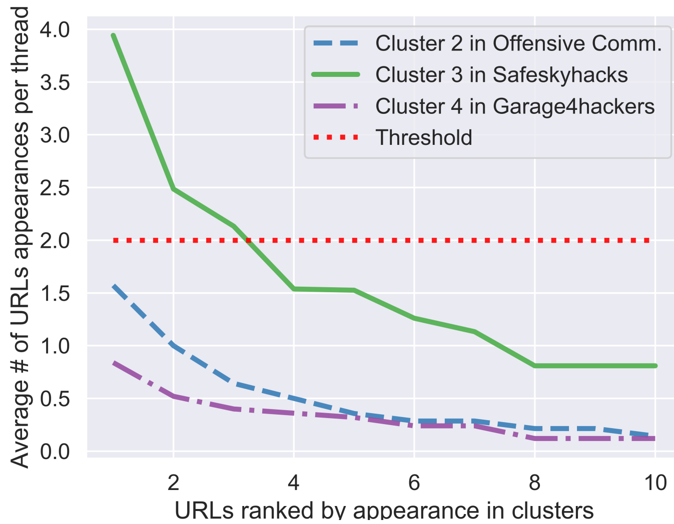


Figure 2.5: The Repetition Index of most referred URLs in three forums with "Long-post with URL" behavior. The index is significantly higher for Cluster 3 of SafeSkyHacks forum, which indicates its "SEO" behavior (linking to the same URL more than two times on average from each thread in the cluster). The red horizontal line indicates an empirically derived threshold that can be used in identifying such clusters.

On closer inspection though, we find that it is different from the other clusters of the same type of other forums. Most of its threads have one post, which contains approximately 1k - 2k words and 5k - 10k characters. Upon further inspection, these posts contain *"a large amount of out-of-context text"*. The structure of these posts follows a repetitive pattern: three of four paragraphs, separated by the same image, and one or more URL links. All embedded hyperlinks in a thread typically point to the same website. Though our initial thought is to consider these threads as spamming, the text around the links was not related to the link. In most cases, the text consists of random excerpts from books or manuals, without any attempt to persuade the reader to click on the link!

How can we distinguish "SEO" from "Long post with URL" threads? Interestingly, there is a total of 4,560 URLs in the 173 threads in this cluster. Those URLs only point to a small group of websites, although they usually point to different pages within

Sites	**IP Location - Host**	**#URLs**
ateasegames.com	London - Hydra Comm. Ltd.	682
elitegamersclub.com	Virginia - Amazon.com Inc.	430
goo.gl	Amsterdam - Google LLC	369
legalaidreform.org	San Jose - Websitewelcome	266
rindfleisch.reisen	Hong Kong - Host Europe Gmbh	264

Table 2.5: The most referred sites in the "SEO" cluster.

the same website, possibly to look less obvious. This lead us to define the Repetition Index as the number of times that a given site is linked by the threads of a cluster divided by the number of threads in the cluster. We show the result of these for three different "Long post with URL" clusters in Fig. 2.5. For example, the most highly cited site, aleasegames.com, in cluster 3 of Safeskyhacks (green line) is cited 3.2 times on average from each thread in that cluster! Also, the most highly cited site in Safeskyhacks point to specific gaming website whereas, in Offensive Community, it points to YouTube. These observations support our hypothesis that these threads are engaging in "SEO" boosting.

Which are the sites that benefit from "SEO" threads? We list the top five most highly-linked sites from the "SEO" threads in table 2.5. The top site is a gaming site, aleasegames.com, and it is pointed-to from 682 places in the cluster. Moreover, one of those highly referred site, *elitegamersclub.com*, are selling their domain name. Note that goo.gl is Google's URL shortening service. Some of those goo.gl URLs are hosting downloadable zip files, which could be malicious, and we intend to analyze this in more detail in future work.

2.5 Discussion

In this section, we further discuss the features that one can use to capture activities in a forum. The plethora of features is can be both a blessing and a curse.

A. Sparse Matrix Regression as a feature selection. In section 2.3, we show that Sparse Matrix Regression works optimally on threads clustering on forums. It can also be utilized to cluster features because the algorithm simultaneously clusters both threads and features in matrices A and B, at the same time. Other feature selection techniques can be used instead of SMR to select the optimal set of features. We merely use a byproduct of SMR we acquired from the previous step. The intensity value of matrix B from SMR in equation 2.2 can be used to determine the significance of features on clusters which it can serve as a feature selection. In Fig. 2.6, we show the top 42 features ranked by intensity in Offensive Community. We found that lifeSpan, dayEntropy and hiddenFirstPost are the most informative features on this forum.

Offensive C.	Garage4hackers	Kernelmode	Safeskyhack	Wilder S.
lifespan	#char_title	dayEntro	dayEntro	#Char_title
dayEntro	userEntro	userEntro	#Char_ShortestP	dayEntro
hidden_FirstP	#word_LongestP	#char_title	#Word_avg	userEntro
#char_title	dayEntro	#spellErr_title	#Char_title	#worryBOW_title
userEntro	#char_FirstP	lifespan	#worryBOW_title	#askBOW_FirstP
#3MostUserPosts	#char_LongestP	#word_title	#buySellBOW_title	#word_title
#word_title	#word_FirstP	userEng_max	#cyberSecBOW_title	#word_avg
#worryBOW_title	#posts	#askBOW_title	#spellErr_title	#char_avg
max1day_#post	#char_avg	#activeDay	#word_title	#worryBOW_FirstP
#spellErr_title	#askBOW_title	#word_LongestP	#askBOW_FirstP	#cyberSecBOW_title

Table 2.6: Top 10 features for each forum ranked by the intensity value.

B. Do the most informative features are shared consistently across different forums?

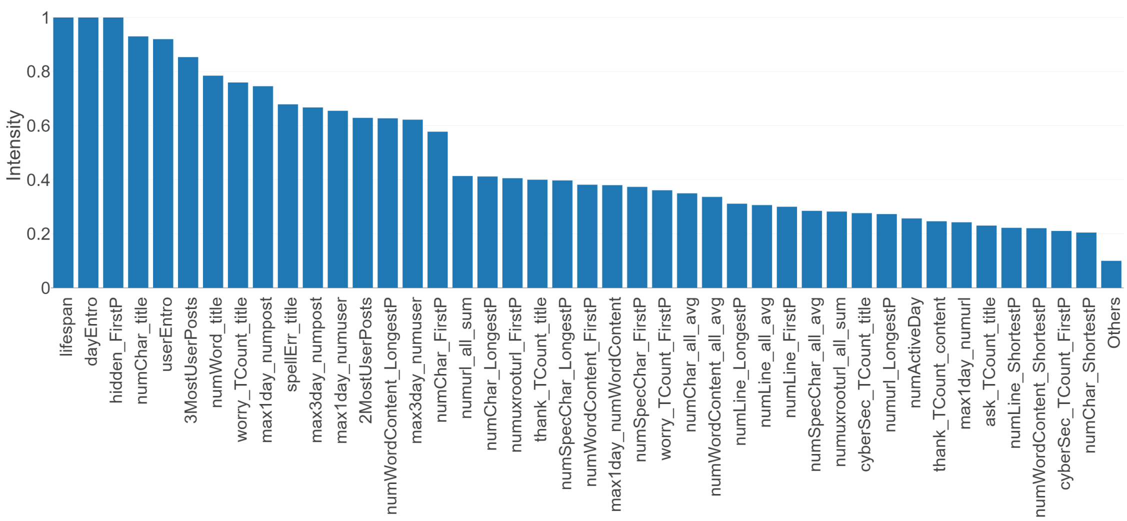


Figure 2.6: Top 42 features in Offensive Community ranked by normalized intensity value. The normalized intensity value for each feature is an average from the result of experiment of SMR with K (a number of clusters) from 3 to 12.

To answer this question, we extracted the top most informative features from matrix B in SMR on five different forums shown in Table. 2.6. We found that #characterOnTitle and dayEntropy are present in all forums. #wordOnTitle and userEntropy appear in four out of five forums. However, some top informative features in one forum might not always be observed in other forums. For example, the hiddenPost feature is unique to Offensive Community forum that has the distinctive capability that allows people to hide their posts.

C. Does including features with limited informative value can hurt the performance of the algorithms? We conduct the experiment to compare the silhouette coefficient in clusters and the number of features ranked by their informative properties in the previous step on five different forums shown in Fig. 2.7. We find that the higher number of features, the lower silhouette coefficient in K-means clustering is. This corresponds to the

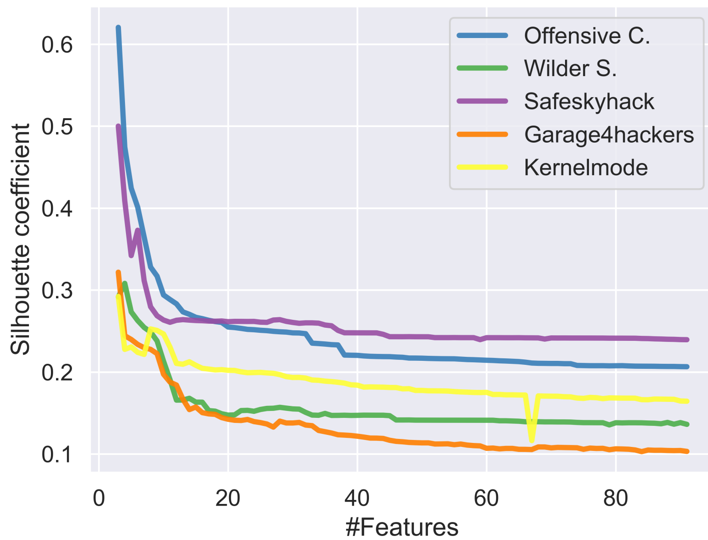


Figure 2.7: Silhouette coefficient from the experiment on the number of features used in K-means clustering experiment. We observe the same trend in five different forums that adding non-informative features in algorithms like K-means results in lesser internal clustering quality observed in the Silhouette coefficient.

finding of studies in feature selection methods [53] which proves that adding non-informative features can reduce the effectiveness of the model and add uncertainty to the predictions. Specifically, algorithms such as linear regression and K-means are shown to be susceptible to non-informative features.

2.6 Related Work

We briefly discuss two categories of relevant research. An extensive listing is not possible due to space limitations, but we will provide it in a subsequent full version of this work.

a. Analyzing computer security forums. Most works in this domain focus on finding function, intention, product, and services in posts. The [78, 29] make use of hand labeling data and NLP techniques in supervised classification to get function and intention of posts as well as a name and a price of product and service in a post. CrimeBB [75] is arguably the first security forum repository, which also reports on high-level trends, such as a number of threads, posts, and users. Some [84, 34, 39] uses data in forums with NLP techniques to predict a cyber attack.

b. Analyzing trends and anomalies in social media. Online social media, like Twitter and Facebook, have been studied extensively. For example, a few recent studies studies [98, 35, 57] use machine learning and data mining to detect behavioral trends and anomalies in social media platforms, such as Facebook and Reddit. There as well, several features exhibit a heavy tail distribution, similarly to our observations. Other studies focus in identifying group of users with similar behaviors. Many [101, 22] use community detection techniques to extract a group of key users with similar behavior.

2.7 Conclusion

We propose, Rthread, a comprehensive unsupervised co-clustering approach with visualization capabilities. Our approach provides a systematic and in-depth thread-centric analysis of online forums using We consider 92 thread features that span three groups: (a) temporal, (b) behavioral, and (c) content related. We also propose a visualization method to aid the interpretation of clusters in an intuitive way. First, we find that many properties follow a log-normal distribution, which is persistent across several forums and over time. Second, we show how our approach can identify classes of threads with similar behavior, revealing some unanticipated thread behaviors.

This preliminary work shows significant promise as a building block towards fully harnessing the wealth of information in online forums. Its unsupervised nature is a significant advantage, as it can explore and detect behaviors that we are not anticipating.

Chapter 3

RAFFMAN: Measuring and Analyzing Sentiment in Online Political Forum Discussions with an Application to the Trump Impeachment

3.1 Introduction

How can we assess the emotional affect toward the impeachment of Donald Trump among the users of an online discussion forums?

The more general question is how we can detect the evolution of the affect or

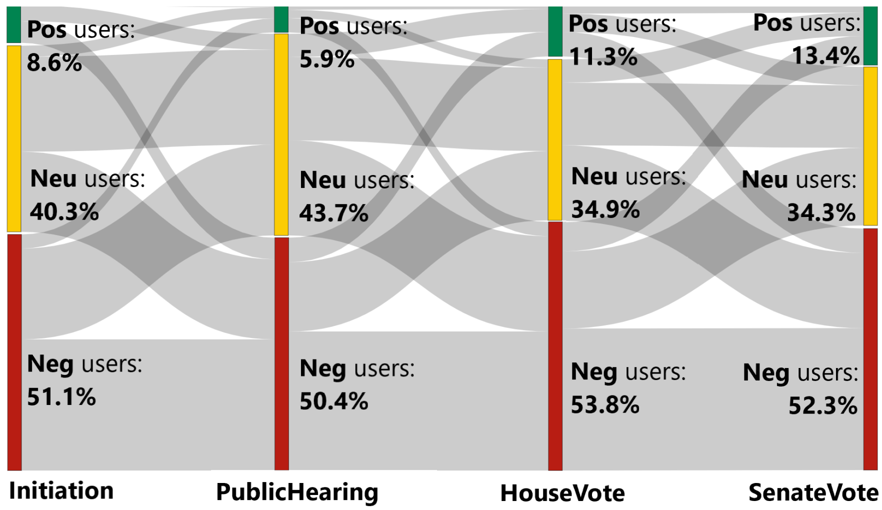


Figure 3.1: Our systematic approach in action: The evolution of users' affect on Reddit towards concept "Trump" on the significant dates (which coincide with changepoints) during the impeachment process. We observe that: (a) more than 50% of the users express negative affect towards Trump, and (b) the impeachment seems to have increased polarization toward concept "Trump" as neutral users decreased from 40.3% to 34.3%. Upon further investigation, we find that 12.1% of users flip-flop from negative to positive or vice versa.

sentiment of users in an online forum towards a **concept** (a person or an idea) in response to external real-world **events**. A major political event such as an impeachment is a concept that can evoke emotional affect in users that can manifest in discussions within a forum as the official proceedings unfold. The challenge is that a complex political event such as Trump's impeachment has many different concepts or aspects that can occur within a discussion, and the discrete political stages that unfold over time can change individuals' emotional affect toward those aspects.

Emotional "affect" is central to the field of political and social psychology. Human cognition connects emotional affect labels to objects such as concepts or ideas [59]. The affect label itself is *not necessarily a preference or opinion* regarding the object; even a supporter of an idea could express frustration or disappointment in series of posts. In the social sciences, however, the study of public opinion is not limited only to stance or preferences. Instead, these affect labels, in the form of positive and negative emotions, impact individuals' cognition and beliefs that are relevant to that object [64]. For example, negative affect toward an object can lead individuals to more closely attend to threatening information [25].

When users engage in online discussion forums, their messages contain latent affect or sentiment that we are able to connect to specific concepts or ideas. Because of the relationship between affect and information processing, having a flexible method to assess affect will be important to researchers who wish to understand how users learn about political information on social media and in online discussion forums [54]. We should clarify that the term "affect" can include the full range of human emotions, such as fear, anxiety, happiness, etc. Sentiment maps these emotions into **three categories**: positive, neutral and negative. Because of this mapping, we use "sentiment" and "affect" interchangeably with this relationship understood.

Given the context above, the problem we address here is inherently complex and difficult. The input to the problem is a concept, an event or group of events, and online forum data, and the desired output is the nature of the users' emotional affect toward that concept and its change over time. We want to model and quantify: (a) the intensity of user engagement, which is caused by the event, and (b) the change of the users' affect polarity, that is, their sentiment, towards concepts.

The problem introduces several challenges. First, we need a keyword expansion method to capture the complex set of aspects related to the political process of interest.

Second, we need a time series statistical method that will allow us to make inferences about the discrete events that drive user engagement. Third, we need to accurately recover sentiment or emotional affect that is connected to aspects under discussion during the discrete events. Fourth, we need to handle the unstructured nature of forum data, in that the posts can vary in lengths, threads become discussions and it is difficult to follow the discourse. Finally, we need to consider that user participation varies over time, and that some forums allow for anonymous users.

There has been limited work on the problem as framed here. We can group related work in two main streams. First, quantifying and measuring studies in forum are appeared in [45] which analyzes general properties and characteristics of 4chan. [85] detect changes related to real-world events. Second, sentiment analysis on the web is mostly used on social media data like Twitter [102]. Aspect-based sentiment analysis, a finer task of sentiment analysis, is usually implemented in the domain of user product reviews [102]. We revisit previous work in the related section at the end.

Contribution: We propose, RAFFMAN[1], a systematic approach to measure the change in users' affect towards a complex concept in response to real-world events in online discussion forums. Our approach consists of the following key components: (a) filtering, (b) detection of change, and (c) sentiment analysis. We adapt, customize and synthesize state-of-the art methods to optimize aspect-based sentiment analysis using the unstructured data of political discussion forums, including (a) developing a keyword expansion method that can identify and filter for different aspects of a complex event, (b) adapt a time series statistical method to identify important discrete stages that compose the complex event,

[1] Acronym not explained here for anonymity purposes.

and (c) optimizing aspect-based sentiment analysis to political discussion. We show that our approach yields up to 74% of accuracy in our three-class classification (positive-neutral-negative). The classification accuracy increases to 81.1% if we only examine short posts with less than 23 words.

We validate and showcase our approach using data from the online discussion forums 4chan and Reddit, where users provide an untapped wealth of information on people's thoughts and sentiments in response to current events: there are 1,000 and 5,000 new posts per minute respectively to 4chan and Reddit. Since we investigate the effect of President Trump's impeachment, we focus on politically-oriented sub-forums. In the remainder of this paper, we will use the terms Reddit and 4chan to refer exclusively to Reddit's r/politics, and 4chan's /pol sub-forums. We collect 32M posts that occurred during the impeachment process between September 2019 to February 2020. We provide an overview of the key results below:

- **The user engagement doubles at significant stages.** We found that, during significant events such as the "House vote" for 4chan and the "committee public hearing" for Reddit, the number of posts related to the topic doubled as shown in Fig. 3.2. This indirectly increases our confidence for our keyword selection and thread filtering methods.

- **Reddit users are more engaged with impeachment compared to those of 4chan.** The percentage of impeachment related posts on Reddit (51.8%) is much higher than that on 4chan (13.3%). This suggests that Reddit users were more concerned about the impeachment as an event. For example, during the "House vote" event, 95% of posts in Reddit engaged to the topic compared to only 38% in 4chan.

- **More than half of all posts have negative affect toward the concept "Trump" throughout the impeachment.** We find that more than 50% of all posts exhibit negative affect toward aspect "Trump" in both Reddit and 4chan. This is true for the majority of the the 6-month impeachment period, namely for 83% of the days for Reddit and 98% of the days for 4chan.

- **The impeachment events increased the divergence of user affect for concepts "Trump," "Impeachment" and "Pelosi."** We find that around 6% of neutral users change their affect to either negative or positive on the key aspects "Trump" and "Impeachment" in Reddit. We also observe a similar divergence for "Pelosi," the Speaker of the House and a key figure in the impeachment saga.

- **The impeachment increased the negative affect towards the concept "Pelosi."** We find a 7.9% increase in users with negative affect and 6.6% decrease in users with positive affect between the two events "House vote" and "Senate vote" in Reddit.

Our work can be seen as a building block to harness the untapped potential of online discussion forums. We argue that effective ways to analyze such forums can provide valuable information on: (a) what resonates with people, as can be seen by increases in the engagement, and (b) how people feel about events and prominent people. Detecting deliberate misuse and social engineering is an important next step in this line of work.

Forum	Posts	Threads	Users
Reddit - politics (total)	10.5M	149K	509K
Reddit - politics (filtered)	5.4M	62K	392K
4chan - pol (total)	16.9M	411K	-
4chan - pol (filtered)	2.2M	38K	-

Table 3.1: Our datasets from Reddit and 4chan over a span of six months from September 2019 to February 2020. The term filtered refers to posts and threads that we identify as related to the event the impeachment of President Trump in Step 1 of our approach. Anonymity in 4chan prevents us from having the number of unique users.

3.2 Background and Datasets

Our work focuses on online discussion forums. We have collected data from two forums, Reddit and 4chan, over a span of six months during the impeachment period between September 2019 and February 2020. We discuss and present our datasets with their basic statistics in Table 4.2.

1. Reddit. We use Reddit a well-known text-based discussion forum with eponymous users. We select the "politics" subreddit (/politics/) because it is directly related to our main focus. The "politics" subreddit contains a large pool of 6.5M registered users with roughly 100k daily posts. The users that select into these forums, along with their posts, can function as a convenience sample for social science research [32] and serve as an interesting population of direct interest regarding online engagement. To collect this subreddit data, we use the archiver service, Pushshift (pushshift.io)[2], that collects every post made in the main Reddit site and makes that data publicly available for academic purposes.

2. 4chan. We use 4chan, which is considered a fringe alt-right forum, as an interesting contrast to Reddit. On 4chan, users do not need to create an account to use the

[2]https://pushshift.io/

platform. As a result, most users remain anonymous while posting comments in the forum. We focus on the "politically incorrect" subforum (/pol/). This is the most active subforum in 4chan with an average of 150k daily posts as reported by 4stats.io. 4chan does not make their data publicly available and it routinely deletes data in the forum. Here, we collect data from a community-run archiver 4plebs (4plebs.org), [3] which crawls and archives all the activity from 4chan and makes it publicly available.

3. Ground-truth for aspect-based sentiment analysis. We have access to a gold standard benchmark data set for aspect-based sentiment analysis (ASBA) obtained from the NLP workshop SemEval (Semantic Evaluation); however this benchmark data is mostly in the domain of restaurant reviews or laptop reviews which is not matched to our task in political discussion. To remedy that drawback, we create our own benchmark dataset using the existing posts in both Reddit and 4chan. We use two groups of annotators (a) five general annotators from Amazon's Mechanical Turk platform and (b) three political unbiased experts in the scientific field. The annotators labeled each post with sentiment toward a given aspect. The final label is produced by using a two-round majority vote approach from (a) and (b) to get a balanced and unbiased training set shown in Table 3.2. We assess our annotated data by using the Fleiss-Kappa coefficient on the two groups of annotators in Table 3.3. We observe the highest agreement in all aspects from experts. These results showcase the benefit of using politically unbiased experts in the ABSA annotation tasks.

4. Concepts, events, aspects, and keywords. Our goal is to study the effect of an event on user sentiment toward a concept. We use the term concept or aspect to

[3] https://4plebs.org/

Aspect	Negative		Neutral		Positive	
	Train	Test	Train	Test	Train	Test
Trump	412	102	406	102	413	103
Impeachment	184	46	183	46	184	46

Table 3.2: Our ground-truth dataset with more than 2K posts for concepts "Trump" and "Impeachment" using: (a) Mturkers, and (b) experts.

Aspect	Mturk	Experts
Trump	0.453	0.583
Impeachment	0.372	0.691
All	0.433	0.601

Table 3.3: Assessing the annotator agreement using the Fleiss-Kappa coefficient on ground-truth for aspect-based sentiment analysis.

refer to a person or an idea, and we can use a set of keywords to describe that concept. For example, "Trump" as an aspect can be referred to with keywords such as "Trump," "Potus," "Donald," ...etc. We explore the following aspects in this paper:

- "Trump" : Donald Trump is the 45th president of the United States from the Republican party.
- "Impeachment" : Impeachment is a U.S. constitutional process to remove government official from the office.
- "Pelosi" : Nancy Peslosi is the Speaker of the House, a leading figure of the opposition Democrat party.
- "QAnon" : QAnon or Q is a far-right conspiracy theory.
- "Goodell" : Roger Goodell is the current American football league Commissioner (an aspect that should not be related to impeachment that we use below for a placebo test of our methods).

Simply put, an event is also a concept that can be "defined" by a set a keywords.

As always, the scenarios can be more complex in practice. For example, the impeachment of Trump is a complex process that is composed of discrete political stages, where each stage can can span multiple days. The *New York Times* lists the following as major stages for Trump's impeachment:

- "Initiation" : Sep 24 2019, the Speaker of the House announced a formal impeachment inquiry.
- "Articles of Impeachment" : Dec 11-13 2019, Committee voted to approve two articles of impeachment.
- "House vote" : Dec 18 2019, House passed the two articles of impeachment.
- "Senate trial" : Jan 29-31 2020, Senators questioned and rejected for any new witnesses or documents
- "Senate vote" : Feb 5 2020, Senate rejected both articles of impeachment against Trump.

3.3 Overview of Our Approach

Our approach provides a method to systematically quantify sentiment in online forums consisting of three major steps, which we outline below.

3.3.1 Step 1: Identifying Related Activity

Given a small set of keywords that are known to be relevant to an event of interest, we want to capture related activities in a forum without requiring specific domain knowledge.

This step consists of (a) expanding a set of initial keywords, and (b) identifying related posts and threads in the forum.

a. Keyword expansion. We utilize an iterative embedding-based approach to expand a set of initial keywords. The key design elements of this approach are as follows: a) We use two similarity expansions, one in the word-word space and one in the post-post space, (b) we use an iterative approach in each of these expansions, and (c) we provide a flexible ranking of the identified words to meet the user needs. Specifically, in order to implement the keyword expansion step, we take following phases:

Phase 1: Domain representation. We represent words and posts of forums in an m-dimensional embedding space with the Word2Vec method [65].

Phase 2: Word-space expansion. We expand the initial set of keywords by adding relevant words iteratively.

Phase 3: Post-space expansion. We identify posts that are similar to the set of posts that contain the relevant words from the previous step.

Phase 4: Result Processing. We extract and rank the keywords from the posts of the previous step, based on several metrics like word-word similarity, post-post similarity and TF-IDF which is based on importance and relevancy. *Similarity score* is calculated by the average of cosine similarity in a Word2Vec embedding space [65] between the initial set and the expended set. Then, a subset of ranked keywords that passes a threshold of similarity will represent an expansion set. This threshold varies depending on the task of interest.

We implement our keyword expansion techniques on the initial event-keywords

Word	Similarity	Word	Similarity
impeachment	1.000	dismiss	0.405
trump	1.000	contempt	0.403
censure	0.568	inquiry	0.403
bush	0.482	prosecute	0.387
trial	0.475	speaker	0.386
judiciary	0.454	resort	0.385
acquit	0.443	remove	0.381
perjury	0.437	cloture	0.380
resolution	0.430	evidence	0.373
witness	0.428	constitution	0.364

Table 3.4: Top 20 similar words acquired from initial event-keywords, "Trump" & "Impeachment" with keyword expansion techniques trained with data from Wikipedia. The higher the score the greater the similarities between that word to an initial keyword set.

known to be related to Trump's impeachment, namely "Trump" and "Impeachment," on Wikipedia pages that contain those words. We selected Wikipedia to expand the event-keywords set because it is external to our forums and so prevents bias in the event-keywords expanded set that could occur if we used our forums' specific posts. The results of our keyword expansion technique are shown in Table 3.4.

b. Identifying related threads and posts. A key step in our approach is to identify the threads that relate to the event and concepts of interest, which we achieve as follows.

i. Identifying related posts. We label a post as related if it contains keywords in any part-of-speech obtained from the previous step. In our experiment, we select only keywords with *similarity score* more than 0.4, which yields 13 unique keywords. We discuss the selection of this value below.

ii. Identifying related threads. We label a thread as related if the title of the thread contains selected keywords or the percentage of related posts are more than the *post-relevance threshold*. In our case, we use 30% as a threshold which we justify below.

Threshold selection. We set the value of our two thresholds, 0.4 similarity and 30% post percentage, using the elbow method [50] by comparing the quantity of related posts and threads obtained with different parameter settings. We identify overall related posts and threads shown in Table4.2, which are (51.8%, 41.6%) and (13.3%, 9.4%) of the total posts and threads in Reddit and 4chan, respectively. The parameters for similarity score and post percentage can be varied depending on the goal of the experiment task and one's preference in the trade-off between too much or too little inclusivity. Higher threshold values yield more posts and threads in exchange for possibly including more posts and threads that are unrelated to the concept and event of interest.

3.3.2 Step 2: Detecting Engagement Change

To identify real-world stages of Trump's impeachment event that impact a forum's engagement activity with respect to our concepts, we turn to changepoint algorithms that can detect significant changes in time series data. Specifically, we choose Pruned Exact Linear Time (PELT) [51], a parametric algorithm that can (a) detect changes and (b) rank them by maximizing its log-likelihood of mean and variance of the time series. In our case, we use a daily number of related posts containing our expanded set of keywords acquired from Step 1 as our time series data. We choose to model the number of posts rather than the number of threads or users because posting reflects the base activity of engagement in forums where users post in response to a topic of interest.

We apply the PELT algorithm to the daily number of related posts to get a list of

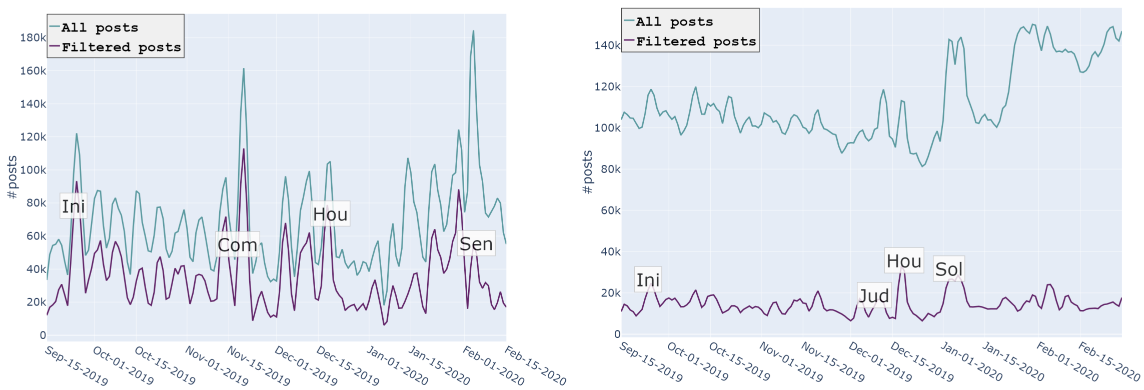
Reddit		4chan	
Changepoint	Stage	Changepoint	Stage
1. 11/18/2019	11/18/2019-11/21/2019: Committee public hearings.	1. 12/19/2019	12/18/2019: House voted to pass the two articles of impeachment.
2. 12/19/2019	12/18/2019: House voted to pass the two articles of impeachment.	2. 01/03/2020	01/03/2020: Trump announced the death of Iranian general (unrelated to impeachment).
3. 02/05/2020	02/05/2020: Senate vote (acquitted).	3. 09/24/2019	09/24/2019: The initiation of impeachment.
4. 09/24/2019	09/24/2019: The initiation of impeachment.	4. 12/09/2019	12/04/2019-12/09/2019: Judiciary committee hearings.

Table 3.5: Correlated real-world stages ranked by changepoint algorithm (PELT) on a significant increase of a number of posts on the impeachment of Trump.

dates ranked by significance. We then compare changepoints with the real-world events in our domain with a window of 1-2 days to accommodate asynchronous activity that occurs just after the event itself. Given (a) a daily number of related posts in forums and, (b) a list of real-world stages of the impeachment of Donald Trump obtained from the *New York Times*[4], we identify the most impactful real-world events, listed in Table 3.5. On Reddit, the “Committee public hearings” is the most statistically significant changepoint compared to the “House vote” on 4chan. Interestingly, we also find a non-related event to the “impeachment” in 4chan on January 03, 2020. This changepoint emerges from the increase in 4chan activities of the keyword “Trump,” in response to the announcement by Trump himself of the assassination of Iranian general Qasem Soleimani.

As a robustness check, we specifically look into the impeachment concept where we only expand one initial keyword, “Impeachment,” and identify the expanded impeachment keywords with Step 1. With this filtering, we find that “House vote” becomes the most statistically significant changepoint on both Reddit and 4chan. Also, the Soleimani changepoint now becomes non-significant on 4chan because it is not directly related to the impeachment event itself.

[4]https://www.nytimes.com/interactive/2019/us/politics/what-is-impeachment-process.html



(a) Reddit

(b) 4chan

Figure 3.2: The temporal view of the posts in forums over a span of 6 months. These figures are labeled with significant changepoints that correlated to the real-world stages in Table 3.5. The purple line represent the amount of related posts filtered with our approach from Step 1. The green line is the total number of posts made daily in each forum. We observe that: (a) user engagement doubles at significant stages (b) Reddit users are more engaged with the impeachment than 4chan users.

Validation of an expanded set of keywords. With the changepoint detection algorithm, Fig. 3.2 plots the number of related posts acquired from Step 1 with the most significant changepoints from Table 3.5. We see the correlation of an increase in an activity of engagement on the topic of interest with the significant changepoints on both forums. This verifies our filtering techniques and expanded set of keywords from Step 1.

3.3.3 Step 3: Aspect-based Sentiment Analysis

To determine users' affect toward a concept, we use aspect-based sentiment analysis (ABSA), a subtask of sentiment analysis in the natural language processing field. While traditional sentiment analysis captures the overall sentiment in text, ABSA aims to detect the corresponding sentiment towards a specific aspect, which in our application are keywords. That is, ABSA can associate specific (negative, neutral and positive) sentiment with different aspects in the same post.

BERT [36], a recent language model from Google, outperforms other traditional techniques like neural networks [48] in many NLP tasks including sentiment analysis because it has the ability to capture the context around words. While there are many variations of ABSA with BERT, we choose [104] as our implementation due to a simplicity while yielding reasonable accuracy when compared to a very complex model like [81].

ABSA consists of two main subtasks: (a) detecting aspects in a sentence, and (b) determining a sentiment associated with an aspect.

a. Detecting aspects. To determine which word is an aspect in the sentence, ABSA employs IOB (Inside-outside-beginning), a common tagging techniques for an NLP task such as POS (part-of-speech) tagging [100] and NER (name-entities-recognition) tagging [38]. However, since we focus on specific aspects such as "Trump" and "Impeachment," we use our expanded set of keywords, the 13 unique keywords from Step 1, as aspects in ABSA.

b. Determining sentiment associated with aspects. ABSA aims to classify a text with respect to a given aspect into the three different classes of polarity (negative, neutral and positive). BERT implements ABSA using a sequence-pair classification task. First, we transform our posts into tokens with a corresponding format. Let x represent BERT embedding sequences:

$$x = [CLS]a_1, ..., a_m[SEP]t_1, ..., t_n[SEP] \tag{3.1}$$

where $a_1, ..., a_m$ are tokens of an aspect, $t_1, ..., t_n$ are tokens of words in a post, $[SEP]$ is separation token, and $[CLS]$ is a special token that can represent the whole embedding sequence. Second, we feed these embedding sequences into the BERT model $h = BERT(x)$. Third, $h[CLS]$ that represents the last hidden representation of embedding sequence is an input to a softmax layer for a sequence-pair classification task which generate the probability of each sentiment's polarities which we show with p.

$$p = softmax(W \cdot h[CLS] + b) \tag{3.2}$$

where $W \in R^{3\times768}$ (weights of our embedding sequence for each polarity on BERT), $b \in R^3$, $p \in [0,1]^3$, 3 is the number of polarities (negative, neutral and positive), 768 is a default length of embedding sequence on BERT. Finally, $argmax(p)$ returns the classification result.

To maximize our ABSA task, we experiment with different language models using the same testbed:

- **NLTK+VADER**: traditional rule-based sentiment analysis that captures the overall sentiment of a post. Stopword removal is performed during the prepossessing of a post.
- **BERT-baseline**: Original pre-trained model and fine-tuned with our ABSA dataset.
- **BERT-custom**: Post-trained model with review data from Yelp and Amazon reviews and fine-tuned with our ABSA dataset [36].
- **XLNet**: A larger language model that claims a better performance over state-of-the-art BERT [107].

All models (except the NLTK+VADER model) are pre-trained model which we fine tune them for our own specific task. We evaluated their performance with 5 fold cross validation. The results of our classification are shown in Table 3.6. BERT-custom is able to achieve a

Model	All posts		Short posts	
	Accurcy	F1	Accurcy	F1
NLTK + Vader	51.1%	50.0%	56.7%	54.4%
XLNet	74.6%	74.3%	75.2%	75%
BERT baseline	70.7%	70.7%	75.9%	75.4%
BERT custom	74.3%	74.4%	**81.1%**	80.9%

Table 3.6: The summary of result of accuracy and F1 scores on aspect-based sentiment classification with our political dataset where short posts contain less than 23 words. For NLTK+VADER, we use a traditional rule-based sentiment analysis model to determine overall sentiment of a post.

competitive result with a larger model like XLNet because training from the reviews data transfers to our political forum dataset. BERT takes less time in this classification task and yields higher accuracy on short posts than XLNet, so we choose BERT-custom to associate sentiment with aspects.

Shorter post lead to higher (81.1%) classification accuracy. We investigate if the length of the post affects the accuracy of our ABSA model with BERT-custom performs. We compare between 234 short posts that contain less than 22 words, which is at 50 percentile of post length distribution, and 242 long posts that have more or equal than 23 words. Our BERT-custom model achieves 81.1% of classification accuracy on the set of short posts compared to 67.6% on the set of long posts. We conjecture that the longer posts may provide the user the ability to ramble and even mix discussion topics, which could affect the classification accuracy. Upon manual investigation, one post in Reddit not only uses many cursed words regarding “Trump,” doubts the so called fair “Trial” procedure but it is also strongly in favor of “Impeachment.” This shows longer posts introduce and mix several arguments and discussions and even appear self-contradicting at times.

Determining user's affect from her posts. We now want to identify the sentiment of a user towards a concept based on posts at given time. Intuitively, the process will aggregate the sentiment of the posts of that user during an appropriate interval of observation. Following our event-driven approach, we introduce the *temporal focus* parameter, windowInterval, to be the number of days around a given date, during which we collect the user posts for that concept. We then use a majority vote on the sentiment of these posts toward an aspect to determine the user's affect on that time period. Note that, if there is a tie between any two categories, we assign the user's affect as neutral. Finally, to increase our confidence in the outcome, we can require a minimum number of posts that a user has to have in that time period in order to be included in the report.

3.4 Case Study: The Impeachment

In this section, we study the dynamics in user affect related to the impeachment of President Trump, which consists of several related stages, as a case study on the political subforums of Reddit and 4chan. To recap, our case study involves concepts (a) "Trump," (b) "Impeachment," (c) "Pelosi" and (d) "Qanon" captured by 13 event-keywords shown in Table 3.4. We also investigate an unrelated aspect, "Goodell" representing Roger Goodell, the current NFL commissioner, as a placebo test of our methods that can help to show our analysis is indeed specific to the political impeachment process. We expect to see significant changepoints in user engagement in the first four impeachment aspects but not in the Goodell aspect.

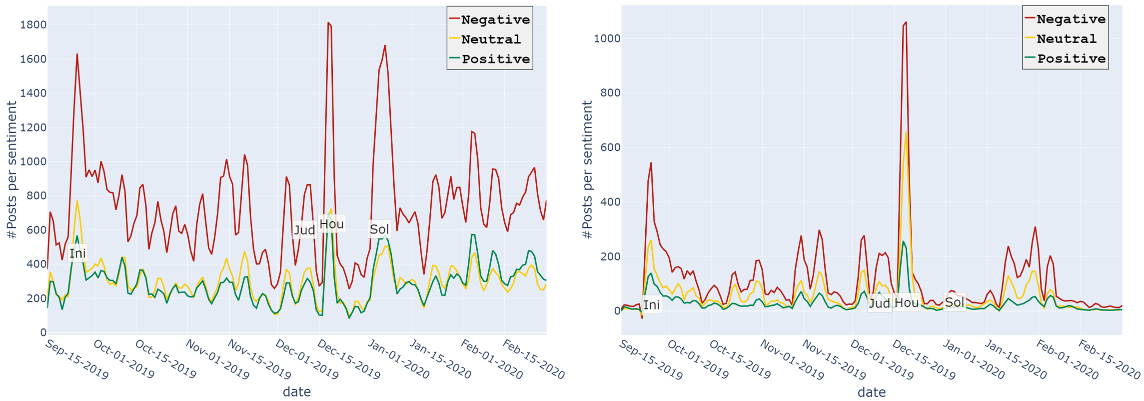
3.4.1 Identifying Engagement Changes

a. User engagement correlates with impeachment related events. Our changepoint detection algorithm of user engagement identifies spikes that coincide with key stages of the impeachment process as identified by the *New York Times*. This observation acts as an indirect validation that our keyword selection and thread filtering follow the user activity adequately. Interestingly, in 4chan, we do not observe significant activity change on the stages "Articles of impeachment" and "Senate trial."

b. User engagement doubles in reaction to impeachment-related events. We find that there is a 218.7% increase in impeachment-related posts to "Committee public hearing" in Reddit, which is shown in Fig. 3.2a. We also see a 186.7% increase of such posts during the "House vote" in 4chan, as shown in Fig. 3.2b. The increase in the most significant changepoint in both forums is calculated by comparing the average of a two-day window before the change point and the peak of the changepoint itself.

c. Reddit users are more engaged regarding impeachment compared to those of 4chan. The trends in Fig. 3.2 show that posts made in Reddit are two times more likely to be a post about the impeachment and Trump compared posts made in 4chan. This observation shows that Reddit users were more engaged over the impeachment of Trump. Furthermore, the impeachment related posts dominate even more during significant external stages. This again is more pronounced for Reddit. For example, we observe the highest percentage of the related posts to the total of all daily posts at "House vote" for both Reddit and 4chan, equal to 95% and 38% of posts respectively.

This contrast in engagement coincides with the different lifespan of threads in



(a) "Trump" concept. (b) "Impeachment" concept.

Figure 3.3: The temporal view of the number of posts per sentiment acquired from our ABSA BERT-custom model on 4chan in a span of 6 months. These figure are labeled with significant changepoints that correspond with the real-world stages in Table 3.5. We observe that: (a) 50% posts about Trump are negative at 98.8% of the days (b) find a spike in posts only on "Trump" at (2) Jan 03 2020, the death of Soleimani.

Reddit and 4chan. 4chan has significantly shorter lifespan of threads than Reddit, since it is known to regularly delete their threads due to its infringed nature. Most threads live up to 3.9 minutes (median) and the fastest thread to expire was around 28 seconds [24]. With those properties, users are less likely to follow conversations in specific threads and post which make a topic of interest diverse.

3.4.2 Identifying User Affect Changes

We assess how user affect towards a concept evolves over time in response to external events.

a. More than half of all posts are negative toward the aspect "Trump." We find that more than 50% of all related posts are negative in Reddit and 4chan on the aspect "Trump." We observe this trend at 83% and 98.8% of the days in the 6-month impeachment period, with respect to Reddit and 4chan. The peak of negative posts on "Trump" reaches 61% of all related posts on 4chan on January 03 where he announced the death of Soleimani, which is an event that we identify in our changepoint analysis, but is only tangentially related to impeachment (that is, a newsworthy event that possibly diverted public attention away from the impeachment proceedings). Given the lesser coherence of discourse found in 4chan, it is not surprising that our changepoint method appears to work better in Reddit.

Interesting, during February 5, 2020 in 4chan, positive posts (26.4%) outpace neutral posts (19.9%) on "Trump," a results shown in Fig. **??**a, a result we do not observe in Reddit. This change corresponds to the "Senate Vote" to acquit all articles of the impeachment on Donald Trump. This is no surprise for 4chan, the forum known for alt-right and supporting Trump.

b. The impeachment increased the polarization of user affect for the aspects "Trump," "Impeachment" and "Pelosi." Polarization [56] of a user's affect occurs when users tend to change their sentiment from neutral to become either more positive or more negative. We find that neutral affect among users is decreased by 6% on "Trump" and 6.4% on "Impeachment" when compared to the start of impeachment process, "Initiation," and the end of the process "Senate vote." We also observe a similar polarization for "Pelosi," the Speaker of the House, and a vocal critic of President Trump as the neutral users are also decreased by 1.2% at "House vote" and "Senate vote," two events where we have enough users' affect on "Pelosi" to draw a conclusion.

Although most users are anonymous in 4chan, we try to gauge polarization by comparing the percentage of posts per sentiment expressed during the same interval above. We find that there is a 3.8% decrease in neutral posts on "Trump," an 8.9% decrease on

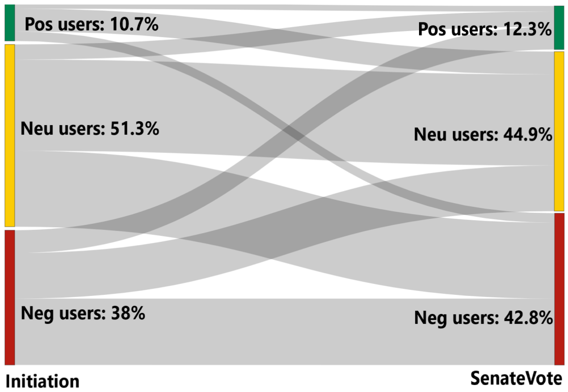


Figure 3.4: The evolution of users' affect towards concept "Impeachment" between events: "Initiation" and "Senate vote" on Reddit. We observe that: (a) 12.8% of users flip-flop from negative to positive or vice versa, and (b) a increase in polarization as neutral decreased from 51.3% to 44.9%.

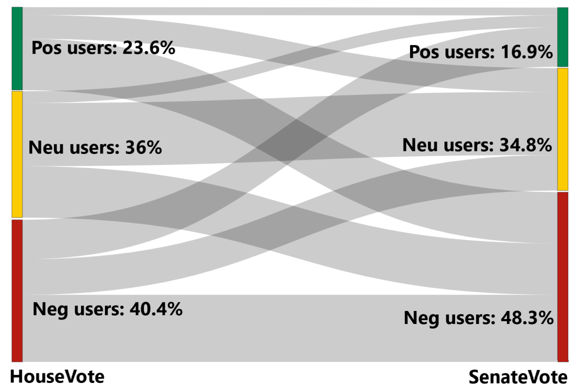


Figure 3.5: The evolution of users' affect on Reddit toward an concept "Pelosi" in the significant dates "House vote" and "Senate Vote" in the impeachment process. We observe that: (a) 25.8% of users flip-flop between negative to positive, and (b) the impeachment increased polarization slightly towards concept "Pelosi."

"Impeachment" and a 10.5% decrease on "Pelosi."

c. The impeachment process increased negativity towards the aspect "Pelosi." In Reddit, we find a 7.9% increase in users with negative affect and a 6.6% decrease in users with positive affect toward Pelosi between the two significant events, "House vote" and "Senate vote" where we have enough number of users for conclusive analysis. This shows how user's affect toward "Pelosi" develop in response to the impeachment.

d. Concepts "QAnon" and "Goodell" and placebo test. QAnon or Q is a far-right conspiracy theory that was created online by a user with the name Q. This theory claims that "a cabal of Satan-worshiping pedophiles running a global child sex-trafficking ring is plotting against President Donald Trump, who is battling them" according to Wikipedia. Reporting suggests that many Trump supporters and the president himself are sympathetic to this idea. We observe some "QAnon" engagement and changepoints during the impeachment period, but they are not aligned with the impeachment events on either Reddit or 4chan.

As a placebo test, we also consider the concept "Goodell," which represents Roger Goodell, the NFL Commissioner. We include this aspect as a placebo test to ensure that our methods not only show relevant engagement, but they also do not pick up on irrelevant engagement. We observe that Goodell's appearance on the forum is limited and also his name does not show any increased engagement aligned with the impeachment events.

3.5 Discussion

Here, we discuss the scope and limitations of our work. **a. Emotions, Sentiment and Stance.** RAFFMAN is designed to detect sentiment in discussions about specific aspects that compose discrete stages of an event. Sentiment is of significant emerging interest to social scientists. The use of automated sentiment analysis is only just beginning to emerge in social science; for example [16]. Sentiment is closely related to emotion, and the study of emotion is a vast field in the social sciences. It has long been established that emotions are central to cognition and the information processing that informs individuals' preferences. For example, [63] is a highly cited review of the field of emotions and politics; [61] is a more recent overview of the methodological considerations in quantifying emotions and the impact of emotions; [42] and [25] are highly cited applications demonstrating the role of emotions in persuasion research; and [26] is a highly cited application investigating the role of emotions in information processing. We argue that RAFFMAN provides a set of tools that would help to advance the study of the role of sentiment in information processing and opinion formation on events discussed in social media settings, in parallel to emotions in the social sciences, and so will make a strong contribution.

Stance detection differs from sentiment analysis in that it is an NLP task to infer the preferences of individuals in favor, against or neither towards the aspect. Some [67, 19] have studied the relation between stance and sentiment, showing sentiment by itself is not enough to detect a person's stance; however, sentiment can be used as one of the important features to detect a stance.

b. Dataset size and Google's BERT. BERT [36] is a state of the art language model that we use here. BERT has transfer learning capability and has proven to be very effective in providing good accuracy with fewer labeled datasets in many classification tasks. A recent stufy [82] shows that BERT works well in an image classification task with around 1,000 labeled datasets. Another project[5] uses only 500 labeled datasets to do sentiment analysis on IMDB movie reviews with BERT and was able to yield 83% classification accuracy. These sutdies show that our 2,000 labeled dataset is ample enough to be used on the ABSA task with BERT transfer learning.

c. Who could use our tool in practice, and how? RAFFMAN is a powerful tool to gauge user affect towards any concept that users discuss online. Sentiment is of interest to itself to social scientists, and in addition our methods could be adapted to the study of individual engagement and information processing in online settings. The additional power lies in that: a) it uses organically derived responses, reflecting genuine engagement, b) it collects opinions *in vivo*, namely at the time that different events take

[5]https://blog.insightdatascience.com/using-transfer-learning-for-nlp-with-small-data-71e10baf99a6

place, and c) it can identify the sentiment evolution at the level of individuals, and not simply in the aggregate. The latter, of course, requires that the forum uses permanent user-names, like Reddit.

From a practical point of view, a user can specify: (a) the forum, (b) a time interval, (c) the concept as a group of keywords, and (c) the events of interest as a group of keywords. The outcome could be a plot as shown in Fig. 3.1: the identified engagement spikes, and a the evolution of the user sentiment between several time-points of interest.

d. The potential impact of our tool. The value and impact of our tool could be quite broad. The potential users could span a wide range: (a) politicians and political advisors, (b) policy makers, (c) marketing firms, and (d) social science researchers. We think that the last group could derive immediate and significant benefits (as enthusiastically argued by the political scientist in our team). In particular, RAFFMAN will enable social scientists to test hypotheses about the role of sentiment in opinion formation that occurs over time in response to social media engagement, which is a core interest in the fields of political psychology, communication and public opinion.

e. How can we detect and account for bots and manipulation? In this work we do not investigate social engineering or deliberate campaigns in our forums. Such misuse from individuals or foreign state actors has been observed in other social media and has sparked national debates. One could argue that this kind of behavior may be less prevalent in our two forums as they attract significantly fewer views compared to, say, Twitter. However, we have manually identified a few cases of such parasitic behavior. In Reddit, users named "GoldyTSA" and "OriginalWorldliness" exhibit a spamming behavior where they posted the exact posts, 43 and 47 times respectively, using a cursed word regarding Trump on the day of "House Vote."

In our future work, we intend to investigate such phenomena and: (a) develop techniques to identify misbehaving users and bots, and (b) quantify the effect of these behaviors on the forum discussions. In that effort, we will leverage the vast literature on detecting fake users and accounts and our own experience in identifying parasitic behavior in online commenting platforms, such as Disqus (disqus.com).

f. How representative is the data? This is the usual concern in every data-driven study. We argue that for the purposes of political discourse our data represents a reasonable case study to illustrate our methods. First, we use two different discussion forums with significantly different appeal and focus (Reddit and 4chan). Second, we consider a substantial amount of time (6 months) with a total of 32M posts during a significant event in U.S. political history. Naturally, for events of smaller magnitude, the intensity of the engagement will be lower, but our approach should provide accurate results.

Furthermore, although we focus on political events and discussions here, our method could apply more generally to other forums and other domains, such as discussions on sports, business, health, entertainment etc.

3.6 Related Work

We summarize related works in the following general areas.

Forum analysis. There are several studies trying to understand general activities in web-based discussion forums. [45], [72] and [97] work on understanding properties, trends and characteristics of forums like ephemerality, heavy-tail and anonymity on posts, threads and users. Some focus on specific tasks in forums [62], [69] [88] and try to identify main actors like hacker users, depressed users and influential users using a variety of techniques including linguistics, behavioral modeling on user activities and graph-based approaches. The most relevant study to our paper [85] explores new emerging words and trends from the concept, "Covid-19," to see how the engagement and topics evolve over time, but they do not study the user sentiment towards these topics.

Sentiment analysis on the web. Most studies in sentiment analysis focus on the area of product reviews [102] or social media like tweets [28]. Although [74] and [47] analyze forum data, they mostly use a base sentiment classification model to capture overall polarity on the text. Aspect-based sentiment analysis (ABSA) is rarely used on forums. Some of the relevant studies include an effort [30], which applies ABSA using a neural network on reviews of scientific papers to quantify the reviewers' sentiment towards an aspect like originality, and a recent effort [104] that focuses on training the BERT approach for different domains using knowledge transfer.

3.7 Conclusion

The key contribution of our work is RAFFMAN, a systematic approach to quantify the change of forum user affect towards a concept in response to real events. Our approach consists of three phases: (a) identifying the related posts, (b) detecting changes in engagement, and (c) conducting sentiment analysis. These three components work synergistically to quantify user sentiment towards a concept is response to a complex event, which could consist of many sub-events. To quantify sentiment, we customize and synthesize state-of-the art methods to classify posts into three categories: positive, neutral and negative. We show that RAFFMAN achieves a classification accuracy of 81.1%, if we focus on posts with less than 23 words and up to 74% of accuracy with all posts.

Chapter 4

SentiStance: Quantifying the Intertwined Changes of Sentiment and Stance in Response to an Event in Online Forums

4.1 Introduction

How can we quantify the combined effect of an event on sentiment and stance in online forums? This is the question we address in this work.

Online forums provide an unprecedented opportunity for answering many social and political science questions. Forums are a publicly available source of human thought and emotions that are both: (a) vast, and (b) largely unconstrained. Unlike tweeter, forums

Stance\Sent	Negative	Neutral	Positive
Against	-7.1%	-1.8%	-1.2%
None	-0.4%	+5.5%	+1.3%
Favor	+0.2%	+1.2%	+2%

(a) Reddit: aspect “Pence”, Jan 6th 2020

Stance\Sent	Negative	Neutral	Positive
Against	**+33.1%**	+0.7%	+4.9%
None	-0.6%	-5.6%	-4%
Favor	-5.2%	-2.6%	**-20.8%**

(b) Parler: aspect “Pence”, Jan 6th 2020

Table 4.1: Our Delta-SentiStance Table (ΔSST) intuitively captures the effect of an event towards an aspect for both sentiment and stance. Here, we show the effect of the Insurrection on January 6, 2020, for aspect “Mike Pence.” Reddit users become more Positive appreciating his certification of the vote. On the contrary, Parler users become upset with a significant increase in Negative sentiment and Against stance, which aligns with the “Hang Mike Pence” campaign that emerged in the forum.

have advantages: a) they allow for the extensive presentation of ideas as posts are not limited in length, and b) they can capture discussions among the participants, therefore forcing longer exchanges and going deeper into thoughts, issues, and feelings. As a result, the analysis of forums promises to capture information that traditional forums (polls and surveys) may not have been able to obtain.

Sentiment and Stance: We focus on both sentiment and stance, and argue that they are equally important (and distinct) for understanding the user's state of mind. In fact, we find that we need to: (a) study them in conjunction, and (b) in connection to relevant events. Stance and sentiment are elements of a vector, describing the state of mind of a user towards an aspect, and the elements of that vector often will be responsive to the circumstances and real-world events.

In each post within a forum thread, users will take a stance as either in *Favor,* *None,* or *Against* some aspect[1], which is the user's position or preference regarding the aspect. At the same time, users will hold some sentiment toward the same aspect that can be of *Positive*, *Neutral*, or *Negative* affect.[2] The existence of a functional relationship or correlation between sentiment and stance is not obvious and is a topic of scientific debate. For example, knowing that a forum user has a Positive stance toward former Vice President Pence need not be predictive of one's sentiment in a post about Pence, since that could depend on the circumstances at that time. We provide context regarding the US elections and its politics in the next section.

An event might change one's stance or sentiment toward an aspect, either jointly or independently. For example Pence's decision to certify the election results on January 6, 2021 might matter differently for supporters of (then) President Donald Trump than it did for supporters of (now President) Joe Biden. As a result, conditioning on the class of events that create winners and losers relevant to the aspect could induce a correlation between stance and sentiment, since one is likely happy that one's side won and one is likely unhappy when one's side lost at the event. However, some events might not create clear winners or losers and so may not impact affect or attitudes. For example, it is not clear whether one side "won" the insurrection on January 6th.

We want to quantify the effect of an event on sentiment and stance in online forums. The input is a set of posts, events, and aspects, and we want to understand how the events affected the sentiment and stance of the users towards the aspect. Knowing

[1] An aspect can be a person, an object, an institution, or an idea. Often, the terms aspect and concept are used interchangeably.

[2] While the term "affect" can include the full range of human emotions, such as fear, anxiety, or happiness, sentiment maps these emotions into three categories: Positive, Neutral and Negative. Because of this mapping, we use "sentiment" and "affect" interchangeably.

changes in the joint distribution of sentiment and stance regarding an aspect, rather than in just the two marginal distributions, is important for two reasons.

First, stance and sentiment are both important elements in public opinion research. First, understanding the preferences of citizens in a democracy is critical to much of public opinion research. But emotional sentiment or "affect" toward political aspects, such as ideas, proposals or government officials, is emerging as another core factor in the field of public opinion and political psychology [59]. Research studies have shown that effect toward an aspect can impact how individuals process information relevant to that aspect [64], and in particular, negative affect toward an object can lead individuals to more closely attend to information [25].

Second, understanding how events affect both sentiment and stance can provide insights into the intertwined dynamics of these aspects, user groups and events. In current politics, there are some aspects for which citizens have fixed or stable opinions, such as toward the U.S. Republican Party or President Donald Trump. This might be because of hardened ideological or partisan positions [40, 55], or because of fixed personality traits [77]. But other aspects might be more responsive to the circumstances surrounding events, and knowing that might be relevant to those interested in understanding processes of persuasion [66].

There has been limited literature on the problem as we frame it here. We group the related works into two main categories. First, there are political studies in online forums [98, 72, 21] which analyze properties and measure the essential statistics on Reddit, 4chan, and Parler, respectively. Second, there are efforts that focus on sentiment analysis and stance detection on the web and platforms other than online forums. These works [20, 68] provide a basis for understanding the relationship between sentiment and stance and the measuring models. The research community seems split. Some efforts [37, 17] suggest the use of sentiment as a key feature to infer stance. By contrast, a recent study [18] questions the use of sentiment as a proxy to the stance.

We propose SentiStance, a systematic approach to quantify the impact of *events* on user *stance* and *sentiment* towards an *aspect* in an online forum. We also apply our approach to real data to understand the dynamics between stance, sentiment and real-world events. Our approach consists of methods for quantifying, and visualizing the change of sentiment and stance as shown in Table 4.1. From a technical point of view, we adopt and customized state-of-the-art approaches to optimize aspect-based sentiment analysis and stance detection on unstructured data in political discussion forums. Our approach yields up to 74% accuracy in our three-class classification of stance (Against-None-Favor) and sentiment (Negative-Neutral-Positive) using a thorough evaluation approach with established benchmarks and our own domain-specific validation set.

As an additional contribution, we conduct a real world study using 7.5M posts from three forums (Reddit, 4chan, and Parler) between November 2020 to January 2021. First, we observe interesting behavioral phenomena that vary between different forums. Second, we attempt to reconcile the aforementioned debate as to whether sentiment and stance are correlated by introducing the idea of conditioning to events. The key observations can be captured in the following points.

a. Parler and 4chan users want to "Hang Mike Pence" while Reddit

users appreciate his role during the Insurrection. The violent Insurrection on January 6th, 2020 created very different reactions towards Mike Pence. A total of 53% (11k posts) and 36% (1.7k posts) posts with Negative sentiment and Against stance on Parler and 4chan aligned with the "Hang Mike Pence" movement. Conversely, Reddit users asked "Pence" to invoke the 25th Amendment with 28% (2.2k posts) of Neutral sentiment and None stance. This is captured visually with the use of the Delta-SentiStance Table, which is shown in Table 4.1 and we explain in detail later.

b. Insurrection does not change the general opinion towards "Donald Trump." Despite Trump's role in inciting the insurrection, perceptions towards "Trump" on Reddit and 4chan did not change. Although the posting activity increased by roughly 150%, there is minimal change in the percentage of overall sentiment and stance in the forums.

c. A significant and related event can cause the correlation of sentiment and stance to intensify. We see that significant events induce a strong relationship between sentiment and stance in Parler. During the 2020 US Election, the "Stop the steal" campaign induced up to an 87.6% correlation of sentiment and stance, with the overwhelmingly Positive sentiment and Favorable stance towards "Donald Trump."

d. An against stance is a reasonable proxy for gauging Negative sentiment. We see that 70% of all posts with Against stance towards an aspect also have a negative sentiment towards it for the vast majority (84.5%) of days and aspects in our observations.

4.2 Background and Datasets

Our work focuses on online discussion forums as a case study. We have collected data from three forums, Reddit, Parler, and 4chan, over a span of three months during the US 2020 election period between November 2020 and January 2021. We present the key statistics of our data in Table 4.2.

1. Reddit. We use Reddit, a popular text-based discussion forum with eponymous users. We select the "politics" subreddit (/politics/) because it contains a large pool of registered users actively discussing politics. The users along with their posts can serve as a convenience sample for social science research. To collect this data, we use the archiver service Pushshift (https://pushshift.io/) that collects every post made on the main Reddit site and makes that data publicly available for academic purposes.

2. 4chan. We use 4chan, which is considered to be a fringe alt-right forum. On 4chan, users do not require an account to use the platform. Hence, most users remain anonymous while posting comments in this forum. We focus on the "politically incorrect" subforum (/pol/) which is the most active subforum in 4chan. 4chan does not make their data publicly available and it routinely deletes data in the forum. So we turn to a community-run archiver 4plebs (https://4plebs.org/), which crawls and archives all the activity from 4chan and makes it publicly available.

3. Parler. We use Parler, a microblogging social network platform that is popular among right-wing extremists and conspiracy theorists. This platform saw a massive increase in new users following Twitter's ban of President Donald Trump on January 8, 2021. Un-

Forum	Posts		Threads		User
	Total	Av.	Total	Av.	Total
Reddit	1.4M	15k/day	64k	0.7k/day	261k
4chan	1.3M	14k/day	218k	2.3k/day	NA
Parler	4.8M	53k/day	1.7M	19k/day	545k

Table 4.2: Our datasets from Reddit, 4chan, and Parler with the total number and the average per day over a span of three months from November 2020 and January 2021. Anonymity in 4chan prevents us from identifying a number of unique users.

Dataset	Against	None	Favor
1. SemEval-2016	2,356	1,231	1,183
2. Multi-Target-2017	2,009	1,217	1,650
3. Forum-2020	240	184	83

Table 4.3: Candidate training sets to train and evaluate our stance detection model in political domain.

surprisingly, this site was shut down a few days later by Amazon after it was found to be inciting violence and spreading misinformation of the sort that led to the attack on the US capitol on January 6, 2021. An invaluable effort [21] managed to collect Parler data before it got shut down and they have made it publicly available.

4. Training and validation sets for targeted stance detection model. We have selected two standard datasets, SemEval-2016 [68] and Multi-Target-2017 [89] to train and evaluate our stance detection model. These datasets were collected during the 2016 presidential election which coincides with our focus in the case study. The summary statistics of these datasets are shown in Table 4.3. Although both sets contain examples in the political discussion domain and were extracted from Twitter, they are different in nature and characteristic. SemEval-2016 includes other topics in the dataset such as abortion, climate change, and atheism. Importantly, SemEval-2016 lacks targeted keywords in some of the posts. On the other hand, the Multi-Target-2017 set contains targeted keywords in the political domain.

To evaluate our approach, we create Forum-2020 an evaluation dataset from our online forum. Specifically, we randomly select posts from Reddit and use three politically-unbiased annotators with advanced degrees, who were briefed with extensive standard guidelines. The annotators are given aspects and assign stance labels for each post. The final label is produced by using a majority vote approach. The annotator agreement exhibits a Fleiss-Kappa coefficient of 0.72, which is indicative of substantial agreement.

5. Aspects and keywords. Our goal is to measure user sentiment and stance toward an aspect conditioning on a real-world event. Recall that we use the term aspect and concept interchangeably to refer to a person, entity or idea. A challenge is that one aspect can be referred to by a set of keywords. For example, the aspect "Democrat" can be referred to with keywords "Democrat", "Dem," "The left", . . . , etc. Identifying all the keywords for an aspect is a challenge in its own right but it goes beyond the scope of this paper: we use well known techniques, which we further validate and curate manually [**?**] Here, we explore the following **six aspects**:

- "Trump": Donald Trump is the 45th president of the United States from the Republican party.
- "Pence": Mike Pence is the 45th vice-president of the United States from the Republican party
- "Republican": The main right-wing political party.
- "Democrat": The main left-wing political party.
- "Pelosi": Nancy Pelosi is the Speaker of the House, a leading figure of the opposition

Democrat party.

- "QAnon": QAnon or Q is a far-right conspiracy theory.

We focus on the following **four events** during the three months surrounding the 2020 US Election.

- "Election": The 59th presidential election, Nov 3 2020.
- "Stop the steal": The right-wing campaign aimed to overturn the US 2020 election result. This misinformation campaign started on Nov 7, 2020 which later led to the Insurrection at the Capitol.
- "Insurrection": On Jan 06 2021, an attempt to overturn Donald Trump's defeat in the election by attacking the US Capitol. Vice-president Mike Pence certified the electoral vote in favor of Joe Biden.
- "Inauguration": Jan 20 2021, the inauguration day of Joe Biden, as the 46th President of the United States.

4.3 Overview of our Approach

We develop SentiStance, an approach to systematically quantify the effect of an event on sentiment and stance in online forums. Our approach consists of three major components: (a) quantifying sentiment, (b) quantifying stance, and (c) visualizing the information in an insightful way.

4.3.1 RAFFMAN: Detecting sentiment

To quantify user sentiment, we use RAFFMAN method that was introduced in our previous chapter. RAFFMAN associates a sentiment (Negative, Neutral, and Positive), for a post towards an aspect. We choose ABSA because we want to detect the corresponding sentiment towards a specific aspect rather than capturing overall sentiment like traditional sentiment analysis. Our approach modifies and combines state-of-the-art BERT [36] [93] with an effective sentence-pair classification method [104]. Note that most previous efforts have evaluated these approaches using Twitter data, which is arguably easier to handle than lengthy posts.

We present a brief overview of the approach. It consists of two main subtasks: (a) detecting aspects in a post, and (b) determining a sentiment associated with the aspect. For each aspect, we reduce the first subtask (a) into identifying the right set of keywords that relate to the aspect. We utilize our prior work's technique in keyword expansion to determine the related set of keywords.

This way we identify all the posts that contain aspects keywords. Clearly, we could miss posts that refer to an aspect without clearly using the appropriate keywords (e.g. "Agreed, he is an idiot"), and for that one would need sophisticated modeling of a discussion, which we will consider in future work.

For the second subtask, we determine sentiment toward an aspect in each post with a sequence-pair classification technique which can be achieved by transforming a post into embedding sequences with an appropriate format.

The RAFFMAN model has been shown to exhibit good performance in the political

Training set	Testing set	Accuracy
SemEval-2016	SemEval-2016	75.7%
	Multi-Target-2017	52.1%
	Forum-2020	57.1%
Multi-Target-2017	SemEval-2016	61.1%
	Multi-Target-2017	78.5%
	Forum-2020	69.6%
Domain adapted set	**Forum-2020**	**73.3%**

Table 4.4: The effect of the training set on the accuracy of StanceMeth using the Forum-2020. The highest performance comes from using the domain adapted set. SemEval-2016 and Multi-Target-2017 are Twitter datasets, and thus, the reported accuracy is less indicative for forum data.

domain [94] with 74.3% accuracy overall, and 81.1% accuracy when applied on short posts (less than 23 words).

4.3.2 Target-based stance detection

We develop a stance detection method that can overcome the challenges of our application domain that we listed earlier. As we mentioned earlier, the idiosyncrasies of this space require non-trivial adaptation to even deploy prior methods [52] without a guarantee that their performance will carry over in our domain. We focus on the basic variation of stance detection to detect a stance or support of a user toward a specific aspect which can be categorized with a label in the set of classes (Against, None, Favor). We utilize the same technique we described in RAFFMAN. Namely, we use an embedding, with which we capture the stance as expressed on the text around the concept using a neural network at its core. We also evaluate two datasets in the political domain and experiment on various techniques to optimize the stance detection task.

Types of training sets. Due to the lack of an established benchmark, identifying a forum-specific ground truth is a challenge. To address this problem, we evaluate the suitability of two public Twitter-specific data sets, SemEval-2016 [68] and Multi-Target-2017 [89]. We show how our algorithm performs for different training and testing datasets. The results are shown in Table 4.4. We see that SemEval-2016's accuracy significantly drops when tested across sets. The decrease in performance occurs because SemEval-2016 allows the absence of aspect keywords in the post, meaning that the model needs to infer the aspect or target using only other available words on a post. This inference hinders the model performance when compared with the Multi-Target-2017 where all of the set contains an aspect keyword. Hence, the model will only need to associate the stance keyword to the specific aspect in the post without requiring an aspect word inference. This finding corresponds to [86] which compared and evaluated stance detection benchmark datasets.

Employing domain adaptation to increase accuracy.

To improve our stance detection model, we experimented with many techniques such as incorporating our embedding vector with other features such as a number of words per post, and TF-IDF [79]. At the end of this exploration, we find and use a domain adaptation technique, within the broader space of transfer learning. We described the process below. Essentially, we use the Multi-Target-2017 dataset as our source domain and we create a domain-adapted target set, which we use to train our forum-specific model. Our process is as follows:

1. We train an initial model with Multi-Target-2017.

2. We deploy the initial model on 100K Reddit posts randomly selected from our Reddit dataset.

3. We create a high-confidence domain-specific ground truth *HiConfSet* dataset

Stance\Sent	Negative	Neutral	Positive
Against	25.9%	4.8%	7.1%
None	5.5%	12%	2.9%
Favor	21.2%	8.1%	12.5%

Table 4.5: SentiStance table (SST) captures the percentage of posts in each pair-class of sentiment and stance toward the aspect at a given interval. Here we show the values in Reddit regarding the aspect "Trump" on January 6, 2021.

by selecting 3,000 posts uniformly randomly from each class with a classification confidence above 90%.

4. We create a domain-adapted model by training with our high-confidence HiConfSet data from the previous step.

Our domain-adapted model exhibits improved performance. We evaluate our model using the Forum-2020 dataset of Table 4.4. The model shows a 3.7% increase in classification accuracy in the political discussion domain compared to the model without domain adaptation. We will use this improved model for the remainder of this paper.

4.3.3 C. Visualizing and Identifying Changes

A useful method should help a user to extract knowledge by making it intuitively easy to observe phenomena. This is particularly important when there are many views and many intertwined factors. We propose the following solutions.

1. SentiStance Table (SST). We propose to use contingency tables to observe the "joint distribution" between sentiment and stance. SST will show, for each day, each forum, and each aspect, how many posts fall under the pair-class categories shown in Table 4.5. The table captures the daily interaction and connection between sentiment and stance in 9 pair-class combinations.

Forum	Term	%Post	INCR	SSCHG	NEG	NEU	POS	AGA	NON	FAV
4chan	Donald Trump	2.1%	1.5X	-	-	-	-	-	-	-
	Mike Pence	0.4%	3.0X	20%	$\uparrow$	$\downarrow$	-	$\uparrow$	$\downarrow$	-
Reddit	Donald Trump	2.4%	1.5X	-	-	-	-	-	-	-
	Mike Pence	0.6%	3.0X	10%	-	-	-	$\downarrow$	-	-
	Republican	0.5%	2.0X	-	-	-	-	-	-	-
Parler	Donald Trump	2.0%	1.0X	-	$\uparrow$	-	$\downarrow$	-	-	$\downarrow$
	Mike Pence	0.5%	3.0X	20%	$\triangle$	-	$\downarrow$	$\triangle$	$\downarrow$	$\triangledown$
	Qanon	0.1%	3.0X	-	-	-	-	-	-	$\downarrow$
	Republican	0.2%	2.0X	-	$\uparrow$	-	-	$\triangle$	-	$\downarrow$

Table 4.6: SentiStance Summary Table (SSST) captures the change between the percentage of sentiment and stance for multiple forums and aspects using colors and symbols for ease of exposition. Here we focus on the Insurrection on January 6, 2020 event. The symbol $(-)$ implies $<10\%$ change, while $\uparrow$ and $\downarrow$ indicate the increase and decrease within 10% - 20%, and $\triangle$ and $\triangledown$ represent change $>20\%$. The smallest change was for "Trump" and "Republican" on Reddit and 4chan despite a doubling of posts at the event, which implies a "hardened" position towards these aspects.

2. Delta-SentiStance Table (ΔSST). We propose to use an offset table, as shown in Table 4.1 to display the change of sentiment and stance at an event. Intuitively, this is equivalent to taking the delta change of aspect between sentiment and stance at the event and compared it with a historical average. In our study, we use the average throughout three months period. The values show the increase in green color or decrease in red color in the percentage of the posts. The goal is to enable a user to quickly gauge the effect even with a quick glance with the use of colors. The numbers can then provide more detail if the user wants to investigate things further.

3. SentiStance Summary Table (SSST). We propose a compact way to gauge the change across forums and aspects. Intuitively, we want to show an informative overview of multiple Delta-SentiStance Table. In the SentiStance Summary Table (SSST), we summarize the change of sentiment and stance for each forum at several events. We show a part of the SSST in Table 4.6 at the Insurrection, due to space limitations. We use colors and symbols to capture the effect of the event on stance and sentiment towards aspects in each forum. The table include the following information:

- Percentage of posts (%Post) captures the number of posts containing aspects of interest at the events. We report as the percentage of the total number of posts in the three months in each forum shown in Table 4.2.
- Post increase (INCR) measures the increase in the number of posts that refer to an aspect at the event. We report this number relative to the average number of posts that refer to the aspect per day on each forum.
- SentiStance Change (SSCHG) is the maximum change in the daily percentage of each pair-class of sentiment and stance at a given event. The SSCHG percentage shown in SSST is the highest change of the 9 pair classes. This number is also used to detect significant events.
- Sentiment (NEG, NON, POS): shows the percentage change of each sentiment class responding to the event in which we compare the percentage of sentiment at the event and the average percentage of sentiment for each aspect over three months of observing period.
- Stance (AGA, NON, FAV) shows the percentage change of each stance class responding to the event. Similarly to sentiment, we compare the percentage of stance at the event against the average for each aspect.

Using the SentiStance Summary Table. We provide some guidelines as to how a user can use this more complex table.

First, the user can start by looking at the colored boxes which indicate a change. Among these changes, the use can identify if changes for an aspect are consistent across forums. Second, the user can examine if one or more aspects exhibit similar changes. Finally, the user can consider if the changes were associated with change in the posting activity for relevant to that aspect as captured by INCR and SSCHG.

4. Aspect-Centric Evolution Table (ACET). A complementary view is to see how different events affect one or more aspects. We propose to use the Aspect Centric Evolution Table (ACT) as shown in Table 4.7 for aspects "Trump," "Pence," "Democrat," and "Republican." Due to space limitations, we don't show all forums for all aspects. The table shows the class with the most dominant change both sentiment and stance for each event with colors to enhance usability. The value of this table is that we can viscerally and quickly gauge: (a) which events affect an aspect, and (b) which aspects seem to be affected similarly from the same events. Here, we see that Parler and 4chan turn negative (Negative, Against) towards "Pence."

It is also interesting to observe that Parler turns negative towards the aspect "Republican" in the same way they turn against "Pence." By contrast, aspect "Trump," the republican leader, does not see a similarly negative effect. In other words, people get *angry* towards "Pence" and "Republican" but not "Trump." A tempting inference here is that Mike Pence represents the establishment republican, unlike Donald Trump who transcends the term republican.

4.4 Case Study: Phenomena and Observations

In this section, we showcase the usefulness of our approach in practice focusing on the six aspects and three events that we described in the Background section. We focus on two

Term	Forum	Election			StopTheSteal			Insurrection			Inauguration		
		Sentiment	Stance		Sentiment	Stance		Sentiment	Stance		Sentiment	Stance	
Donald Trump	Reddit	NEG	AGA	↓	–	–	–	–	–	–	–	–	–
	4chan	POS	FAV	↑	–	–	–	–	–	–	–	–	–
	Parler	–	–	–	POS	FAV	△	POS	FAV	↓	–	–	–
Mike Pence	4chan	NEU	NON	↑	–	–	–	NEG	AGA	↑	–	–	–
	Parler	–	–	–	–	–	–	NEG	AGA	△	–	–	–
Democrat	Reddit	NEU	NON	△	–	–	–	–	–	–	–	–	–
	4chan	NEG	AGA	↓	–	–	–	–	–	–	–	–	–
Republican	Reddit	NEG	AGA	↓	–	–	–	NEG	AGA	↑	–	–	–
	Parler	–	–	–	–	–	–	NEG	AGA	↑	–	–	–

Table 4.7: The Aspect-Centric Evolution Table (ACET) displays the evolution of sentiment and stance across multiple events. Due to space limitations, we show only a subset of aspects and forums. We observe that events Election and Insurrection had a significant effect on several aspects, while StopTheSteal and Inauguration much less so. Similarly, aspect Trump was mostly affected by the event Election. Interestingly, the aspect "Republican" seems to be more similar to aspect "Pence" than the concept "Trump" at the Insurrection event.

large categories: (a) event-driven user behavior, (b) posting behavior, and (c) the correlation between sentiment and stance.

Caveat: The results we will report are only relevant for political discussions and even then they could be influenced by the forums and the type of aspects that we examine. In other words, a forum focused on soccer or financial policy may not exhibit similar behavior. We discuss the generalization of the approach in the next section.

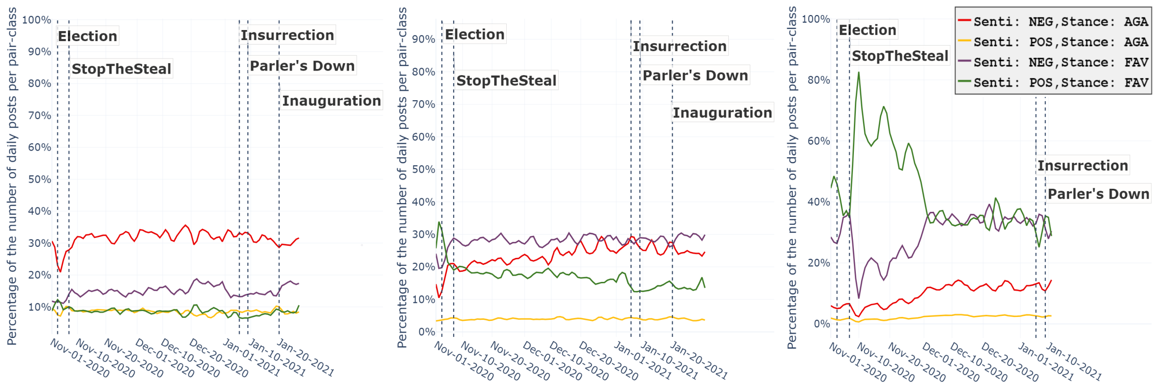
4.4.1 Identifying Event-Driven User Behavior

1. Parler and 4chan users want to "Hang Mike Pence," while Reddit users appreciate his role during the Insurrection. Users across forums exhibit widely different behaviors for the same concept. This observation is fairly apparent in several of the tables of the previous section. For completeness, we can revisit to Table 4.7 that focuses on the class with the biggest change. Different forums exhibit significantly different responses to different events. In fact, some events exhibit no significant change due to an event, while others respond with resentment or favorably. Focusing on aspect "Pence" at the Insurrection, we can see more detailed analysis in Table 4.1. The far-right groups on Parler become resentful towards "Pence" with an increase of 33.1% in Negative sentiment and Against stance. By contrast, Reddit users appreciate his role that day and there is a decreast of Negative-Against by 7.1.% with increases in the four pairwise combinations of Neutral or Positive classes with the maximum increase of 5.5% in Neutral-None class.

2. Sentiment and stance toward "Trump" change at the Election event but don't change at the Insurrection. The first indication of this observation can be seen in Table 4.7. Only the election creates a significant change in the senti-stance for aspect "Trump" for Reddit and 4chan. While for Parler, the StopTheSteal and Insurrection events creates Positive-Favourable change. Intrigued, we want to investigate this further. We plot the time series evolution of a subset of the pair-class in Fig. 4.1 for each forum. We observe that for Reddit and 4chan the change is mostly at the Election and even then it is relatively small. By contrast, Parler exhibits wildly varying behavior: there is a huge spike of roughly doubling the Positive-Favorable pair-class at the StopTheSteal event. Furthermore, the effect of this event seems to lingers for roughly two months! Surprisingly, the Insurrection has no effect on the aspect "Trump".

The above analysis strongly suggests that one could use this kind of analysis to: (a) characterize a forum and its users, and (b) detect unusual emerging phenomena and the groups affiliated with them.

3. Aspect profiling: Events do not affect tangentially-related aspects. We can use the effect of the events to characterize an event. In the Aspect-Centric Evolution Table in Table 4.7, we see that each event differs from every other event with respect to its



(a) Donald - Reddit (b) Donald - 4chan (c) Donald - Parler

Figure 4.1: The temporal view of the SentiStance framework on aspect "Trump." Each line represents the percentage of a pair-class of sentiment and stance (a subset shown for clarity). Note that on Jan 10, Parler was shut down. Most events do not affect the users predisposition toward "Trump". The exception is the initiation of the StopTheSteal campaign, which has huge impact but only on Parler.

effect on the reported aspects.

Although not-shown due to space limitations, aspects "QAnon" and "Pelosi" are not significantly affected by any of the events on all three forums. We attribute this to the fact that the users do not find these aspects relevant to the events.

4.4.2 Are Sentiment and Stance Correlated?

Can sentiment be used as a proxy of stance? We revisit here this question that has divided the community with some studies in favor [37, 17] and others against [18]. The latter and most recent study argues sentiment and stance are unrelated, but they report results averaging across many disparate aspects, and across large time intervals. We revisit the question with a new twist: we observe aspects individually and anchor them to events.

1. Does a significant event intensify the correlation between sentiment and stance? We observe that some events can strengthen the correlation between senti-

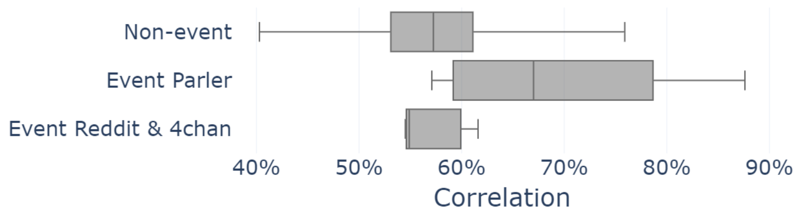


Figure 4.2: Box plots compare the daily correlations between the significant events and non-events. We find that significant events on Parler intensify the correlation between sentiment and stance. We observe the highest correlation for the aspect "Trump" during the "Stop the steal" campaign.

ment and stance. Such events would have to satisfy following condition: the event must be related to the aspect and have a clear-cut outcome corresponding to the stance. The following example captures the intuition: the Positive sentiment of the fans of a football team will reveal their stance when their team wins. By contrast, an "indifferent" event, such as a change in the rules of the game will reveal as easily team preferences.

First, we evaluate this idea by applying χ^2 test on the daily sentiment and stance on each aspect in our forums. We find that 95.6% of the days reject the null hypotheses with a p-value$<\alpha$, with α=0.05, which suggests that sentiment and stance are not statistically independent.

We then calculate the daily sentiment and stance correlation at days without a significant event and at days with significant events as shown in Fig. 4.2. Note that sentiment and stance were completely independent the observed value would be 33.3% given the existence of three classes. In the figure, we see that days without significant events exhibit an average of 58%. The correlation at significant events is 55% for Reddit and 4chan and 68% for Parler. The key observation is that significant events can sometimes induce a substantial correlation between sentiment and stance, but not always.

Intrigued, we want to investigate why Reddit and 4chan do not exhibit the intensified correlation in contrast to Parler. Let's focus on the Insurrection event. The

Insurrection is a significant event for "Pence" in all forums, where the correlation values are 54.9%, 54.5%, and 67% in 4chan, Reddit, and Parler, respectively. Parler showed a significant increase in Negative sentiment and Against stance. 4chan displayed a similar trend to Parler but at a lower scale and a mix of Negative sentiment and None stance. Reddit shows a small increase in Neutral sentiment and None stance. This shows that the Insurrection fails to induce the majority of the community on Reddit and 4chan to respond to the event, perhaps because it is not clear, which side "won" at the Insurrection, a complex event that could be interpreted in many ways.

2. Is sentiment a good predictor of stance? In general, sentiment is not a good predictor of stance. We reach this conclusion by finding the accuracy of sentiment to predict stance, which can be calculated by collapsing the daily SentiStance Table 4.5 column-wise to get a percentage of each stance per particular sentiment. A good predictor can be indicated with the high percentage of posts diagonally in the table. Some aspects such as "Democrat" and "Trump" yield more than 80% accuracy for at least 89% of the observed days on Parler, for Negative and Positive sentiment respectively but they got a low predictor accuracy in other forums. We conjecture that, under the right conditions on the dynamics among aspect and event, sentiment can be a good predictor of the stance.

3. Is stance a good predictor of sentiment? We observe that an Against stance can imply Negative sentiment relatively accurately, but other pair-classes are less strongly correlated. We perform a similar process as above by collapsing the SentiStance table row-wise to gauge the stance prediction accuracy for sentiment. We see that when people are Against an aspect, they are also likely to accompany Negative sentiment toward

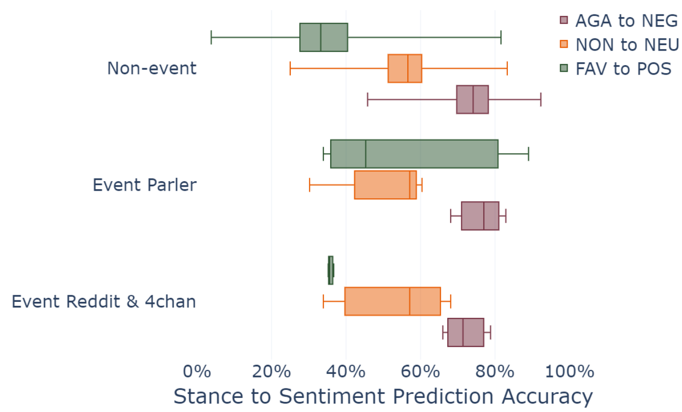


Figure 4.3: Stance as a predictor of sentiment: Each color represents the accuracy distribution for each class of stance for days without event (all forums) and days with significant events. An Against stance predicts a Negative sentiment most accurately irrespective of the presence of an event.

that aspect. The figure shows again the correlation of Stance to Sentiment accuracy for (a) days without a significant event for all forums, (b) days with significant events for Parler, and (c) days with significant events for Reddit and 4chan. The Against to Negative correlation is the highest (red boxes with average range 71-75%), with None to Neutral correlation being second (orange color and average range 55-57%), and Favorable to Positive correlation being last (green color with average range 33-44%).

4.5 Discussion

Will SentiStance generalize to other online forums and domains? Yes. Our approach is an ML-based framework that will generalize to other domains as long as it is trained with appropriate data. Naturally, the accuracy will need to be assessed for different domains or even specific concepts. For example, a concept may be mentioned scarcely in a forum, or specific terms may have multiple and ambiguous meanings.

Here, we provide evidence for the capabilities of SentiStance at detecting sentiment and stance on the political discussion in forums. It can also apply to other domains but slight adjustments would be needed. For example, posts in the sports domain might use different terms to express their stance and sentiment regarding their favorite football teams. Domain-specific training sets will certainly improve the model detection accuracy.

How our finding can generalize to other domain? From our case study on the political domain, the intensification of the correlation and the predictability between sentiment and stance driven by events are intuitive. We see SentiStance as a starting point in understanding the relationship between sentiment and stance. With more experiments in different domains, with different aspects, users, events and time periods, we can generalize our findings and provide a broader perspective on the relationship between sentiment and stance across other topics driving user engagement on social media.

4.6 Related Work

We summarize related works in the following general areas.

Political analysis on online forums. There are several studies of activities in politically-oriented online forums. The works [91, 72, 21] analyze and characterize Reddit, 4chan and Parler, respectively. They also measure vital statistics such as the distribution of posts and users which serve as a basic step on further analysis. The work in [46] compares Parler and Twitter during the Insurrection where they observe similarity in trends and language use. The study [73] shows how fact-checking on the validity of information can influence users on online forums. Other studies [80, 85] focus on the connection between political views and other issues, such as COVID-19.

Sentiment analysis and stance detection on the web. Many papers explore the interconnection between sentiment and stance. These works [20, 90, 87, 68] compare and contrast the practical use of sentiment and stance on different models and available datasets. Despite the difference of sentiment and stance, some works [37, 17] suggest the use of sentiment as one of the key features to detect stance. The work [18] suggest otherwise where they propose sentiment cannot proxy stance, although that finding does not condition on events as we do in SentiStance.

4.7 Conclusion

The key contribution of our work is SentiStance, a systematic approach to measure and quantify the change of sentiment and stance towards an aspect in online forum. Our approach consists of two components: (a) measuring, and (b) visualizing the evolution of sentiment and stance. These two components work harmoniously to quantify the sentiment and stance as a pair class towards an aspect in response to real-world events. We develop comprehensive models to classify posts into three categories in two dimensions: (Negative-Neutral-Positive), and (Against-None-Favor), respectively to sentiment and stance. We show that SentiStance achieves a classification accuracy of up to 74% for sentiment and stance. We see SentiStance as a systematic approach that can shed light in understanding the relationship and the connection between sentiment and stance in the community.

Chapter 5

Conclusions

Our thesis proposes and develops a systematic suite of methods to recognize users' cognition and identify behaviors in an online discussion forum. We develop a comprehensive approach to a) recognize, detect and characterize both common and abnormal behaviors in online forums and b) quantify and redefine the understanding of users' cognition in terms of sentiment and stance. Our approaches have the following main advantages: a) we develop comprehensive tools to detect and analyze online behavior, b) our tools can quantify users' cognition in terms of sentiment and stance with respect to the specific topic or keyword in the post and c) we show the sentiment and stance can be correlate conditioning to the critical event.

We see our work as an important step that can enable many new research directions including: a) detecting emerging users' cognition in a different domain, b) reevaluating how sentiment and stance correlate, and c) monitoring online activity in the online discussion forum in a structured and meaningful way.

Bibliography

[1] 4chan statistic. https://4stats.io/.

[2] 4plebs. https://4plebs.org/.

[3] Cambridge cybercrime centre. https://www.cambridgecybercrime.uk/.

[4] Garage4hackers. http://garage4hackers.com.

[5] Greysec. https://greysec.net/.

[6] Impeachment process. https://www.nytimes.com/interactive/2019/us/politics/what-is-impeachment-process.html.

[7] Kernelmode. https://www.kernelmode.info/forum/.

[8] Offensive community. http://www.offensivecommunity.net.

[9] Profanity checker. https://pypi.org/project/profanity-check. Accessed: 2021-10-30.

[10] Pushshift. https://pushshift.io/.

[11] Raidforums. https://raidforums.com/.

[12] Safeskyhacks. http://www.safeskyhacks.com/.

[13] soleimanideath. https://www.whitehouse.gov/briefings-statements/remarks-president-trump-killing-qasem-soleimani.

[14] Stresserforum. https://www.stresserforums.co/.

[15] Wilders security. http://www.wilderssecurity.com.

[16] Nicholas Adams-Cohen. Policy change and public opinion: Measuring shifting political sentiment with social media data. *American Politics Research*, 48, 06 2020.

[17] Abdulrahman Al-Ghadir, Aqil Azmi, and Amir Hussain. A novel approach to stance detection in social media tweets by fusing ranked lists and sentiments. *Information Fusion*, 67:29–40, 10 2020.

[18] Abeer AlDayel and Walid Magdy. Assessing sentiment of the expressed stance on social media. *Socinfo 2019*, 2019.

[19] Abeer AlDayel and Walid Magdy. Assessing sentiment of the expressed stance on social media. *CoRR*, abs/1908.03181, 2019.

[20] Abeer ALDayel and Walid Magdy. Stance detection on social media: State of the art and trends. *Information Processing and Management*, 58(4):102597, Jul 2021.

[21] Max Aliapoulios, Emmi Bevensee, Jeremy Blackburn, Barry Bradlyn, Emiliano De Cristofaro, Gianluca Stringhini, and Savvas Zannettou. An early look at the parler online social network. *ICWSM 2021*, 2021.

[22] T. Anwar and M. Abulaish. Ranking radically influential web forum users. *IEEE Transactions on Information Forensics and Security*, 10(6):1289–1298, June 2015.

[23] James R. Ashford, Liam D. Turner, Roger M. Whitaker, Alun Preece, and Diane Felmlee. Assessing temporal and spatial features in detecting disruptive users on reddit. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 892–896, 2020.

[24] Michael S. Bernstein, Andrés Monroy-Hernández, Drew Harry, Paul André, Katrina Panovich, and Gregory G. Vargas. 4chan and /b/: An analysis of anonymity and ephemerality in a large online community. In Lada A. Adamic, Ricardo Baeza-Yates, and Scott Counts, editors, *ICWSM, 2011*. The AAAI Press, 2011.

[25] Ted Brader. Striking a Responsive Chord: How Political Ads Motivate and Persuade Voters by Appealing to Emotions. *American Journal of Political Science*, 49(2):388, apr 2005.

[26] Ted Brader, Nicholas Valentino, and Elizabeth Suhay. What triggers public opposition to immigration? anxiety, group cues, and immigration threat. *American Journal of Political Science*, 52:959 – 978, 10 2008.

[27] Rasmus Bro, Evangelos E. Papalexakis, Evrim Acar, and Nicholas D. Sidiropoulos. Coclustering—a useful tool for chemometrics. *Journal of Chemometrics*, 26(6):256–263, 2012.

[28] Caetano, Almeida, and Marques. Characterizing politically engaged users' behavior during the 2016 us presidential campaign. In *2018 IEEE/ACM (ASONAM)*, 2018.

[29] Andrew Caines, Sergio Pastrana, Alice Hutchings, and Paula J. Buttery. Automatically identifying the function and intent of posts in underground forums. *Crime Science*, 7(1):19, Nov 2018.

[30] Souvic Chakraborty, Pawan Goyal, and Animesh Mukherjee. Aspect-based sentiment analysis of scientific reviews. *ACM/IEEE 2020*, Aug 2020.

[31] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 1217–1230, 2017.

[32] Alexander Coppock. Generalizing from Survey Experiments Conducted on Mechanical Turk: A Replication Approach, jul 2019.

[33] Alexander Coppock and Oliver A. McClellan. Validating the demographic, political, psychological, and experimental results obtained from a new source of online survey respondents. *Research & Politics*, 6(1):205316801882217, jan 2019.

[34] Ashok Deb, Kristina Lerman, Emilio Ferrara, Ashok Deb, Kristina Lerman, and Emilio Ferrara. Predicting Cyber-Events by Leveraging Hacker Sentiment. *Information*, 9(11):280, nov 2018.

[35] Pravallika Devineni, Danai Koutra, Michalis Faloutsos, and Christos Faloutsos. If walls could talk: Patterns and anomalies in facebook wallposts. In *Proceedings of the 2015 IEEE/ACM International Conference on*, ASONAM '15, pages 367–374, New York, NY, USA, 2015. ACM.

[36] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. *CoRR*, abs/1810.04805, 2018.

[37] J. Ebrahimi, D. Dou, and Daniel Lowd. A joint sentiment-target-stance model for stance classification in tweets. In *Conf. on Computational Linguistics 2016*, 2016.

[38] Finke, Rose, Grenager, Trond, Manning, and Christopher. Incorporating non-local information into information extraction systems by gibbs sampling. ACL '05, 2005.

[39] Joobin Gharibshah, Tai Ching Li, Maria Solanas Vanrell, Andre Castro, Konstantinos Pelechrinis, Evangelos E. Papalexakis, and Michalis Faloutsos. Inferip: Extracting actionable information from security discussion forums. In *Proceedings of the 2017 IEEE/ACM Conference on*, ASONAM '17, pages 301–304, New York, NY, USA, 2017. ACM.

[40] Paul Goren. Party identification and core political values. *American Journal of Political Science*, 49, 2005.

[41] Maarten Grootendorst. Bertopic: Leveraging bert and c-tf-idf to create easily interpretable topics., 2020.

[42] Kimberly Gross. Framing persuasive appeals: Episodic and thematic framing, emotional response, and policy opinion. *Political Psychology*, 29:169 – 192, 04 2008.

[43] Anna Guimaraes, Oana Balalau, Erisa Terolli, and Gerhard Weikum. Analyzing the traits and anomalies of political discussions on reddit. *Proceedings of the International AAAI Conference on Web and Social Media*, 13(01):205–213, Jul. 2019.

[44] J. A. Hartigan and M. A. Wong. Algorithm AS 136: A K-Means clustering algorithm. *Applied Statistics*, 28(1):100–108, 1979.

[45] Gabriel Emile Hine, Jeremiah Onaolapo, Emiliano De Cristofaro, Nicolas Kourtellis, Ilias Leontiadis, Riginos Samaras, Gianluca Stringhini, and Jeremy Blackburn. Kek, cucks, and god emperor trump: A measurement study of 4chan's politically incorrect forum and its effects on the web, 2017.

[46] Hitkul, Avinash Prabhu, et al. Capitol (pat)riots: A comparative study of twitter and parler. 2021.

[47] Hsu, Hsu, and Tseng. Discovering negative comments by sentiment analysis on web forum. *World Wide Web*, 22(3):1297–1311, 2019.

[48] Huang, Ou, and Carley. Aspect level sentiment classification with attention-over-attention neural networks. *CoRR*, abs/1804.06536, 2018.

[49] Peter Jachim, Filipo Sharevski, and Paige Treebridge. Trollhunter [evader]: Automated detection [evasion] of twitter trolls during the COVID-19 pandemic. In *NSPW*, pages 59–75. ACM, 2020.

[50] David J. Ketchen and Christopher L. Shook. The application of cluster analysis in strategic management research: An analysis and critique. *Strategic Management Journal*, 17(6):441–458, 1996.

[51] Killick, Rebecca, Fearnhead, Paul, and I.A. Eckley. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association*, 107:1590–1598, 12 2012.

[52] Dilek Küçük and Fazli Can. Stance detection: A survey. *ACM Comput. Surv.*, 53(1):12:1–12:37, 2020.

[53] Max Kuhn and Kjell Johnson. *Applied Predictive Modeling.* 01 2013.

[54] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, Michael Schudson, Steven A. Sloman, Cass R. Sunstein, Emily A. Thorson, Duncan J. Watts, and Jonathan L. Zittrain. The science of fake news. *Science*, 359(6380):1094–1096, 2018.

[55] Matthew Levendusky. The partisan sort: How liberals became democrats and conservatives became republicans. *Bibliovault OAI Repository*, 01 2009.

[56] Matthew S. Levendusky. Why do partisan media polarize viewers? *American Journal of Political Science*, 57(3):611–623, jul 2013.

[57] Tai-Ching Li, Joobin Gharibshah, Evangelos E. Papalexakis, and Michalis Faloutsos. Trollspot: Detecting misbehavior in commenting platforms. *CoRR*, abs/1806.01997, 2018.

[58] Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. Roberta: A robustly optimized bert pretraining approach, 2019.

[59] Milton Lodge and Charles~S. Tabor. The Automaticity of Affect for Political Leaders, Groups, and Issues: An Experimental Test of the Hot Cognition Hypothesis. *Political Psychology*, 26(3):455–482, 2005.

[60] Luca Luceri, Silvia Giordano, and Emilio Ferrara. Detecting troll behavior via inverse reinforcement learning: A case study of russian trolls in the 2016 us election. *Proceedings of the International AAAI Conference on Web and Social Media*, 14(1):417–427, May 2020.

[61] Kennet Lynggaard. Methodological challenges in the study of emotions in politics and how to deal with them. *Political Psychology*, 40, 10 2019.

[62] M. Macdonald, R. Frank, J. Mei, and B. Monk. Identifying digital threats in a hacker web forum. In *2015 IEEE/ACM International Conference (ASONAM)*, pages 926–933, 2015.

[63] George Marcus. Emotions in politics. *Annual Review of Political Science - ANNU REV POLIT SCI*, 3:221–250, 06 2000.

[64] George E. Marcus, Michael MacKuen, and W. Russell Neuman. Parsimony and Complexity: Developing and Testing Theories of Affective Intelligence. *Political Psychology*, 32(2):323–336, apr 2011.

[65] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781, 2013.

[66] William Minozzi, Michael Neblo, Kevin Esterling, and David Lazer. Field experiment evidence of substantive, attributional, and behavioral persuasion by members of congress in online town halls. *Nat. Acad. of Science*, 112, 03 2015.

[67] Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *CoRR*, abs/1605.01655, 2016.

[68] Saif M. Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transaction*, 17(3), 2017.

[69] Munger, Tyler, Zhao, and Jiabin. Identifying influential users in on-line support forums using topical expertise and social network analysis. 2015.

[70] Gergely Palla, Imre Derényi, Illés Farkas, and Tamás Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 07 2005.

[71] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 2010.

[72] Antonis Papasavva, Savvas Zannettou, Emiliano De Cristofaro, Gianluca Stringhini, and Jeremy Blackburn. Raiders of the lost kek: 3.5 years of augmented 4chan posts from the politically incorrect board, 2020.

[73] Deven Parekh, Drew Margolin, and Derek Ruths. Comparing audience appreciation to fact-checking across political communities on reddit. WebSci '20, 2020.

[74] Park, Beck, Fletche, Lam, and Tsang. Temporal analysis of radical dark web forum users. In *ASONAM, 2016*, 2016.

[75] Sergio Pastrana, Daniel R. Thomas, Alice Hutchings, and Richard Clayton. Crimebb: Enabling cybercrime research on underground forums at scale.

[76] Juan Manuel Pérez, Juan Carlos Giudici, and Franco M. Luque. pysentimiento: A python toolkit for sentiment analysis and socialnlp tasks. *CoRR*, abs/2106.09462, 2021.

[77] Thomas Pettigrew. Social psychological perspectives on trump supporters. *Journal of Social and Political Psychology*, 5:107–116, 03 2017.

[78] Rebecca S. Portnoff, Sadia Afroz, Greg Durrett, Jonathan K. Kummerfeld, Taylor Berg-Kirkpatrick, Damon McCoy, Kirill Levchenko, and Vern Paxson. Tools for automated analysis of cybercriminal markets.

[79] Anushka Prakash and Harish Tayyar Madabushi. Incorporating count-based features into pre-trained models for improved stance detection. In *3rd NLP4IF Workshop*, December 2020.

[80] Ashwin Rao, Fred Morstatter, et al. Political partisanship and anti-science attitudes in online discussions about covid-19. *Journal of Medical Internet Research*, 2020.

[81] Alexander Rietzler, Sebastian Stabinger, Paul Opitz, and Stefan Engl. Adapt or get left behind: Domain adaptation through bert language model finetuning for aspect-target sentiment classification. *arXiv preprint arXiv:1908.11860*, 2019.

[82] Miguel Romero, Yannet Interian, Timothy D. Solberg, and Gilmer Valdes. Training deep learning models with small datasets. *CoRR*, abs/1912.06761, 2019.

[83] Peter Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.*, 20(1):53–65, November 1987.

[84] Anna Sapienza, Alessandro Bessi, Saranya Damodaran, Paulo Shakarian, Kristina Lerman, and Emilio Ferrara. Early warnings of cyber threats in online discussions. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 667–674. IEEE, 2017.

[85] Leonard Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. "go eat a bat, chang!": An early look on the emergence of sinophobic behavior on web communities in the face of COVID-19. *CoRR*, abs/2004.04046, 2020.

[86] Benjamin Schiller, Johannes Daxenberger, and Iryna Gurevych. Stance detection benchmark: How robust is your stance detection? *Künstl Intell (2021)*, 2020.

[87] Hendrik Schuff, Jeremy Barnes, Julian Mohme, Sebastian Padó, and Roman Klinger. Annotation, modelling and analysis of fine-grained emotions on a stance and sentiment detection corpus. Asso. for Com. Linguistics 2017, 2017.

[88] Shrestha, Anu, Spezzano, and Francesca. Detecting depressed users in online forums. 2019.

[89] Parinaz Sobhani, Diana Inkpen, and Xiaodan Zhu. A dataset for multi-target stance detection. In *European Chapter of ACL: Volume 2*, April 2017.

[90] Parinaz Sobhani, Saif Mohammad, and Svetlana Kiritchenko. Detecting stance in tweets and analyzing its interaction with sentiment. Asso. for Com. Linguistics 2016, August 2016.

[91] Ahmed Soliman, Jan Hafer, and Florian Lemmerich. A characterization of political communities on reddit. HT '19, 2019.

[92] Ahmed Soliman, Jan Hafer, and Florian Lemmerich. A characterization of political communities on reddit. In *Proceedings of the 30th ACM Conference on Hypertext and Social Media*, HT '19, page 259–263, New York, NY, USA, 2019. Association for Computing Machinery.

[93] Chi Sun, Luyao Huang, and Xipeng Qiu. Utilizing BERT for aspect-based sentiment analysis via constructing auxiliary sentence. *NAACL 2019*, abs/1903.09588, 2019.

[94] Jakapun Tachaiya, Joobin Gharibshah, Kevin E. Esterling, and Michalis Faloutsos. RAFFMAN: measuring and analyzing sentiment in online political forum discussions with an application to the trump impeachment. In Ceren Budak, Meeyoung Cha, Daniele Quercia, and Lexing Xie, editors, *Proceedings of the Fifteenth International AAAI Conference on Web and Social Media, ICWSM 2021, held virtually, June 7-10, 2021*, pages 703–713. AAAI Press, 2021.

[95] Jakapun Tachaiya, Joobin Gharibshah, Evangelos E. Papalexakis, and Michalis Faloutsos. Rthread: A thread-centric analysis of security forums. In *2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1–5, 2020.

[96] Jakapun Tachaiya, Arman Irani, Kevin M. Esterling, and Michalis Faloutsos. Sentistance: quantifying the intertwined changes of sentiment and stance in response to an event in online forums. In Michele Coscia, Alfredo Cuzzocrea, Kai Shu, Ralf Klamma, Sharyn O'Halloran, and Jon G. Rokne, editors, *ASONAM '21: International Conference on Advances in Social Networks Analysis and Mining, Virtual Event, The Netherlands, November 8 - 11, 2021*, pages 361–368. ACM, 2021.

[97] S. Thukral, H. Meisheri, T. Kataria, A. Agarwal, I. Verma, A. Chatterjee, and L. Dey. Analyzing behavioral trends in community driven discussion platforms like reddit. In *(ASONAM) 2018*, pages 662–669, 2018.

[98] Sachin Thukral, Hardik Meisheri, Tushar Kataria, Aman Agarwal, Ishan Verma, Arnab Chatterjee, and Lipika Dey. Analyzing behavioral trends in community driven discussion platforms like reddit. *CoRR*, abs/1809.07087, 2018.

[99] Michele Tomaiuolo, Gianfranco Lombardo, Monica Mordonini, Stefano Cagnoni, and Agostino Poggi. A survey on troll detection. *Future Internet*, 12(2), 2020.

[100] Kristina Toutanova, Dan Klein, Christopher D. Manning, and Yoram Singer. Feature-rich part-of-speech tagging with a cyclic dependency network. NAACL '03, page 173–180, USA, 2003. Association for Computational Linguistics.

[101] Nguyen Vo, Kyumin Lee, Cheng Cao, Thanh Tran, and Hongkyu Choi. Revealing and detecting malicious retweeter groups. In *Proceedings of the 2017 IEEE/ACM International Conference*, ASONAM '17, pages 363–368. ACM, 2017.

[102] C. Wang, M. Day, C. Chen, and J. Liou. Temporal and sentimental analysis of a real case of fake reviews in taiwan. In *(ASONAM) 2017*, pages 729–736, 2017.

[103] Junyuan Xie, Ross B. Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis. *CoRR*, abs/1511.06335, 2015.

[104] Hu Xu, Bing Liu, Lei Shu, and Philip S. Yu. Bert post-training for review reading comprehension and aspect-based sentiment analysis. In *ACL 2019*, jun 2019.

[105] Mingjin Yan. Methods of determining the number of clusters in a data set and a new clustering criterion. 01 2005.

[106] Yinfei Yang, Steve Yuan, Daniel Cer, Sheng yi Kong, Noah Constant, Petr Pilar, Heming Ge, Yun-Hsuan Sung, Brian Strope, and Ray Kurzweil. Learning semantic textual similarity from conversations, 2018.

[107] Zhilin Yang, Zihang Dai, Yiming Yang, Jaime G. Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. Xlnet: Generalized autoregressive pretraining for language understanding. *CoRR*, abs/1906.08237, 2019.

[108] Savvas Zannettou, Tristan Caulfield, Emiliano De Cristofaro, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. Disinformation warfare: Understanding state-sponsored trolls on twitter and their influence on the web. In *WWW (Companion Volume)*, pages 218–226. ACM, 2019.

[109] Savvas Zannettou, Tristan Caulfield, William Setzer, Michael Sirivianos, Gianluca Stringhini, and Jeremy Blackburn. Who let the trolls out?: Towards understanding state-sponsored trolls. In *WebSci*, pages 353–362. ACM, 2019.