Does Outrage Signal Cyber Attacks?: Predicting "Bad Behavior" from Sentiment in Online Content

Kristy Hollingshead and **Bonnie J. Dorr** and **Adam Dalton**
Florida Institute for Human and Machine Cognition
Ocala, Florida
{kseitz,bdorr,adalton}@ihmc.us

Meg Barton
Leidos, Inc.
Arlington, Virginia
{marguerite.r.barton}@leidos.com

Abstract

We demonstrate that it is possible to leverage big data in the form of tweets and linked webpages to find expressions of sentiment that signal "bad behavior" such as cyber attacks. We hypothesize that expressions of "outrage" (high intensity, negative affect sentiment) against an organization in public data may be predictive of cyber attacks for two reasons: 1) threat actors may be motivated to launch an attack based on anger/discontent, and 2) outrage associated with an organization or industry may increase the likelihood of that organization or industry being victimized by threat actors (i.e., as a form of "vigilante justice"). We measure sentiment in online content and determine trends in public emotion and their correlation to trends in cyber attacks, as reported in Hackmageddon. We demonstrate that dimensions of sentiment, as afforded by our use of the Circumplex model of emotion, do yield correlations to reported cyber attacks, but differ dependent upon the domain of the data. Thus the use of this technique requires careful analysis for optimal application.

1 Introduction

The idea that emotion directly causes behavior is both intuitive and pervasive in the psychological literature (e.g., (Baumeister et al. 2007; Loewenstein et al. 2001; Russell 2003)). This is especially true for negative emotions; the concept that fear can cause a fight-or-flight response or that anger can cause aggression is widely accepted (Baumeister et al. 2007). Furthermore, within the field of criminology, Agnew's (Agnew 1992) General Strain Theory (GST) asserts that "strain" (i.e., stress) produces negative emotions that lead to criminal behavior (Ganem 2010). In particular, GST purports that anger is conducive to criminality as it incites individuals to act out aggressively, creates a desire for retaliation or revenge, and lowers inhibitions. As an alternative to GST, (Baumeister et al. 2007) propose a feedback model of emotion, wherein emotions tend to result from behavior and provide feedback into cognitive control of future behavior. However, even in this feedback model of emotion, the authors acknowledge that intense emotion can "bypass rational analysis to influence behavior directly," often with negative consequences.

Given the literature linking negative emotion, particularly strong negative emotions such as anger, with maladaptive and criminal behavior, we hypothesize that expressions of "outrage" against an organization may be predictive of cyber attacks for two reasons: 1) threat actors may be motivated to launch an attack based on anger/discontent, and 2) outrage associated with an organization or industry may increase the likelihood of that organization or industry being victimized by threat actors (i.e., as a form of "vigilante justice").

While *vigilantism* is defined in a variety of different ways, the act may be broadly characterized as "the handling of a grievance by unilateral aggression" (Black 2014; Haas 2010). Hacktivism, the phenomenon whereby political and social activists engage in civil disobedience via computer networks (Lohrmann 2016), is in some cases a type of *cyber vigilantism* (Coleman 2011), which has become increasingly common in recent years. Cyber vigilantes and hacktivists leverage the relative anonymity of the internet and their technical prowess to enact "retributive actions" in the cyber realm against "identified wrongdoers" (Smallridge, Wagner, and Crowl 2016). When the public discusses organizations in strong negative terms, we suspect an increase in the probability that those organizations will be identified as "wrongdoers" by threat actors and targeted for vigilante action. Hence, we hypothesize that public outrage associated with an organization or industry might increase the likelihood of that organization or industry being targeted

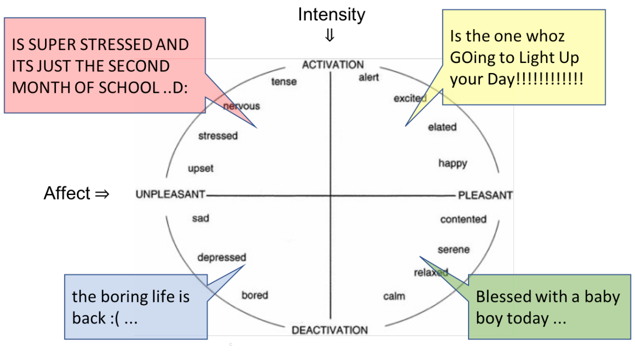

Intensity ⇓
ACTIVATION
IS SUPER STRESSED AND ITS JUST THE SECOND MONTH OF SCHOOL ..D:
Is the one whoz GOing to Light Up your Day!!!!!!!!!!!!
tense
alert
nervous
excited
stressed
elated
upset
happy
Affect ⇒
UNPLEASANT
PLEASANT
sad
contented
depressed
serene
relaxed
bored
calm
DEACTIVATION
the boring life is back :(...
Blessed with a baby boy today ...

Figure 1: Examples in the Circumplex Model of Sentiment.

by attackers acting as cyber vigilantes. Thus, if we could measure outrage, we might be able to predict such attacks.

The notion of outrage as a sensor is unconventional and novel, relying on inputs that relate to social behaviors and interaction as found in big data—as opposed to the more-conventional, hardware-related sensors typically used to predict attacks. In this work, we demonstrate a newly developed "Outrage Sensor" that calculates emotion based on the two-dimensional Circumplex Model of Sentiment (Posner, Russell, and Peterson 2005). Figure 1 shows example sentences from (Preoţiuc-Pietro et al. 2016), placed in each quadrant of the model according to the sentence's emotional score as calculated by their data-driven implementation of the Circumplex model:

- Upper right: Yellow callout indicates happy/elated post (e.g., *Is the one whoz GOing to Light Up your Day!!!!!*).
- Upper left: Red callout indicates upset/stressed post (e.g., *IS SUPER STRESSED AND ITS JUST THE SECOND MONTH OF SCHOOL*). Note: We define the notion of *outrage* within this quadrant.
- Lower left: Blue callout indicates depressed/bored post (e.g., *the boring life is back :(*).
- Lower right: Green callout indicates relaxed/contented post (e.g., *Blessed with a baby boy today*).

As shown in Figure 1, the Circumplex model represents emotion as the two dimensions of *affect* (shown on the X axis), where positive is pleasant and negative is unpleasant, and *intensity* (shown on the Y axis), where positive is active and negative is deactive. This is well grounded in the psychological literature but is treated in a novel way in this work, having been adapted to computational linguistics with the cyber domain.

The rest of the paper is organized as follows: In Section 2, we discuss related work that motivates and informs our research. The details of our technical approach are described in Section 3, while Section 4 describes the data used to test our hypotheses. Section 5 presents our results, leading to the conclusion that signals of sentiment do yield correlations to cyber attacks, but the strength of these signals depends on the nature of the domain and the vocabulary.

2 Background and Related Work

We hypothesize that negative emotion expressed by threat actors or associated with a target has a causal relationship with cyber attacks, either by provoking the attack or increasing the likelihood of a target being victimized. We capture this expressed negative emotion using *sentiment analysis*.

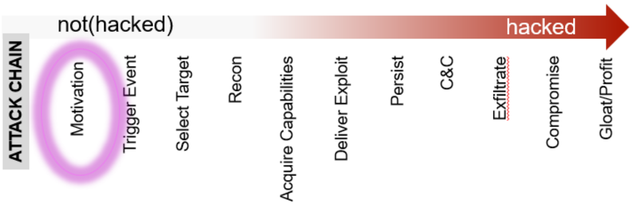


Figure 2: Detecting outrage as a motivation, early in the Cyber Attack Chain.

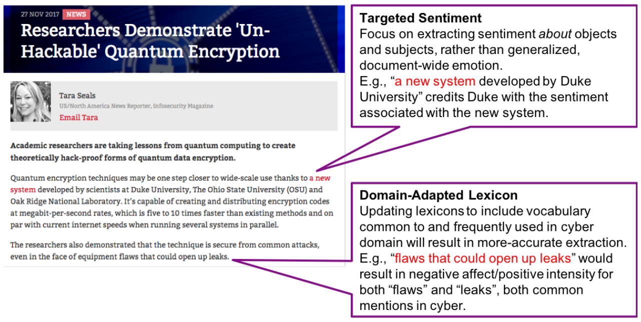
Multiple studies have evaluated whether sentiment analysis can effectively contribute to the prediction of various phenomena, from stock market prices (Bollen, Mao, and Zeng 2011; Nguyen, Shirai, and Velcin 2015), to movie sales (Mishne and Glance 2006), to election results (Bermingham and Smeaton 2011). Of further interest, work from Shaw et al. (Shaw et al. 2013) suggested that negative sentiment captured in digital communications could be used as a psycholinguistic indicator of insider risk.

We measured sentiment in tweets and webpages linked from tweets to see if trends in public emotion *correlate* to trends in cyber attacks. We do not specifically focus our sentiment analysis on the online communications of threat actors. Rather, by casting a wide net, we expect to capture both the sentiment of threat actors along with everything else we gather (relevant to Hypothesis 1) and general negative sentiment around an organization or industry (relevant to Hypothesis 2). More precisely, our outrage analysis captures the brewing discontent that precedes the attack event (Hutchins, Cloppert, and Amin 2011) (see Figure 2), with the assumption that there is a building up of outrage and discontent on the part of the attackers, with respect to a targeted entity, that gives rise to the decision to carry out the attack.

2.1 Sentiment Lexicons

Our research on Outrage as a signal has required significant refinements to improve the utility and specificity of emotion detection. One challenge concerns the nature of sentiment analysis which relies on *lexicons* for emotion classification. Our initial experiments were with the data-driven lexicon[1] from (Preoţiuc-Pietro et al. 2016), containing about 2,000 entries, each with machine-learned weights for the two dimensions of affect and intensity. Analysis revealed that this lexicon often triggered an "outrage signal" on words that are neutral in this domain, and commonly used in cyber-

[1] Lexicon available for download at http://lexhub.org/data_sets/22. While our work is English-based, it is potentially extensible, with these lexicons already extended to Spanish, and several other emotion-based lexicons extended to German, Finnish, Italian, and British English.



27 NOV 2017 NEWS

Researchers Demonstrate 'Un-Hackable' Quantum Encryption

Tara Seals
US/North America News Reporter, Infosecurity Magazine
Email Tara

Academic researchers are taking lessons from quantum computing to create theoretically hack-proof forms of quantum data encryption.

Quantum encryption techniques may be one step closer to wide-scale use thanks to a new system developed by scientists at Duke University, The Ohio State University (OSU) and Oak Ridge National Laboratory. It's capable of creating and distributing encryption codes at megabit-per-second rates, which is five to 10 times faster than existing methods and on par with current internet speeds when running several systems in parallel.

The researchers also demonstrated that the technique is secure from common attacks, even in the face of equipment flaws that could open up leaks.

Targeted Sentiment
Focus on extracting sentiment *about* objects and subjects, rather than generalized, document-wide emotion.
E.g., "a new system developed by Duke University" credits Duke with the sentiment associated with the new system.

Domain-Adapted Lexicon
Updating lexicons to include vocabulary common to and frequently used in cyber domain will result in more-accurate extraction.
E.g., "flaws that could open up leaks" would result in negative affect/positive intensity for both "flaws" and "leaks", both common mentions in cyber.

Figure 3: Example of the need for cyber-adapted lexicons.

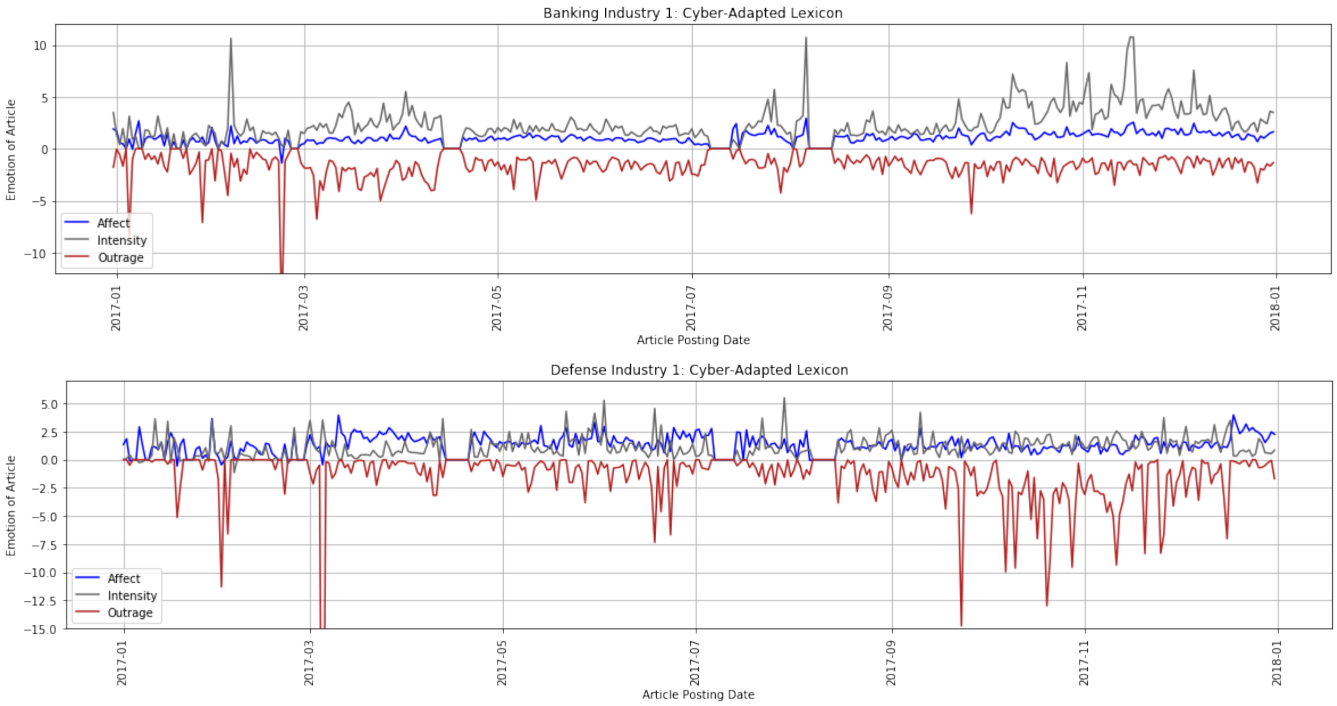


Figure 4: Timelines of outrage expressed in online content around organizations in the banking (top) or defense (bottom) industries using a cyber-adapted data-driven lexicon. Note differences in patterns across domains yet using the same lexicon.

related content. Therefore, we created an updated version of our original lexicon (which we refer to as “Adapted”), to be more cyber-specific by removing about 5% of the entries that are specific to the cyber domain and triggered outrage detection, such as the “@” symbol, which in informal genres signifies a mention or reply and an increased intensity, but in the formal genres studied here, does not carry the same increased intensity. Thus the original lexicon (“Orig Lex”) subsumes the cyber-adapted lexicon.

In addition to the data-driven lexicons (Preoţiuc-Pietro et al. 2016), we also experimented with the well-known Affective Norms for English Words (ANEW) lexicon (Bradley and Lang 1999), a lexicon of words rated for sentiment by human subjects. Its extension, which we term ANEW+, developed by (Warriner, Kuperman, and Brysbaert 2013), includes more than 15,000 lexical items in comparison to the 2,000 items in the data-driven lexicons. Two potential drawbacks to using this ANEW+ lexicon are that (1) it is based on formal genre of language and contains words that are extremely unlikely to appear in cyber-attack data such as “aardvark,” and (2) the weights for each word were not learned from online content as the original lexicon was. One benefit to using the ANEW+ lexicon is an increased coverage of lexical content, although this increased coverage may result in a weaker signal.

To demonstrate the effects of using different lexicons to detect outrage in a document, Figure 3 shows an example of a webpage with negative affect and positive intensity. The callout on the bottom right of the figure gives an example of how including cyber-specific vocabulary in the sentiment lexicons results in more-relevant output from the outrage sensor. The callout on the top right of the figure demonstrates a future improvement to the sensor, where sentiment will be attributed to specific objects rather than document-wide as it currently stands. The example shows that “Duke University” will be credited with the sentiment associated with the “new system” discussed in the article.

Another major challenge for building an outrage sensor was the use of an *emotion lexicon* not designed for noisy text surrounding the meaningful content. We return to this point in Section 5, which provides a comparison of results using three different lexicons.

3 Methodology

We take as input large datasets of online messages, including tweets, news articles, and other webpages, and poll for targeted entities. We then perform sentiment analysis on any message mentioning an entity of interest, where an “entity of interest” might be a person or business, such as “Brian Krebs” and “Securus,” as shown in Section 5.1, or it might be more abstract like an industry, such as “defense” or “banking,” as shown in Section 5. Any messages that express outrage, which we have defined within the Circumplex model as any sentiment of negative affect and positive intensity, are stored and used to create a timeline of outrage for each entity. Each point along the timeline represents a day’s worth of online content relating to the entity. Figure 4 shows the timelines of different emotion dimensions (affect, intensity, and outrage) detected in online content related to

Banking (top) or Defense (bottom) Industries in 2017.

Table 1 shows the prevalence of outrage detected in the online content collected for each of four entities or domains of interest. Each of the three lexicons (original, cyber-adapted, and ANEW+) detected some level of emotion—non-zero affect or non-zero intensity—in 100% of the documents. The percentage of analyzed documents with detected outrage varies substantially across entity (subject domain) and lexicon.

From these timelines of outrage, we then detect *anomalous* periods of outrage, which are in turn used to predict cyber attack events. While we could have predicted cyber attacks each time outrage is detected, or each time outrage levels exceed some threshold, we hypothesized that cyber attacks would be related to *anomalous* periods of outrage—times when outrage suddenly peaked or, potentially, suddenly dropped. In order to detect and flag anomalous behavior, we apply a generalizable, online anomaly detection system (Wei et al. 2005) based on timeseries bitmaps (Kumar et al. 2005). With such a generalizable technique as its foundation, this anomaly detection system can take as input any behavior over time—such as communication time of day, or message topic, or sentiment—and automatically analyze the behavior for anomalous data points. We used this system to detect "unexpected" levels of outrage in a timeseries.

The identified anomalous periods of time—for example, when outrage was anomalously high, or there was an anomalous gap in the timeseries—was then used to predict an attack. An anomalous time period of outrage, for example, could signal a shift in general sentiment towards an organization, preceding or inciting an attack. We calculate the accuracy of our prediction by comparison to publicly-known cyber attacks against several public targets, described below.

4 Data

The ground truth of cyber attacks used herein is the Master List of daily cyber attack statistics reported by Hackmageddon[2] during 2017. There are 951 cyber attacks listed by Hackmageddon for 2017; for each of the reported attacks, Hackmageddon reports the date of the attack, the author of the attack if known, and the target and type of the attack. We aim to use extracted outrage signals to predict these reported cyber attacks.

The data used to predict cyber attacks in this study was collected from the contents of webpages linked from tweets collected between January 2017 through December 2017. These tweets were collected based on a set of more than 2,000 keywords related to cyber attacks, including "*DDoS*" (distributed denial-of-service), "*breach*," and "*hacked*." Webpages linked from the tweets were scraped and then filtered based on keywords relevant to specific industries such as *banking* or *defense* industries, or for mentions of specific entities such as *Krebs* or *Securus*. For each of these documents, the content was analyzed for sentiment as described in Section 3. The total number of documents and length of time analyzed is shown in Table 2.

Entity/	**% Outrage Detected**		
Domain	**Orig Lex**	**Adapted**	**ANEW+**
Krebs	94.9%	97.9%	97.5%
Securus	77.2%	90.6%	93.2%
Banking	5.6%	99.9%	70.1%
Defense	3.5%	100%	73.4%

Table 1: Outrage prevalence as % of analyzed documents containing outrage as detected by 3 different lexicons.

Subject	**Time Period**	**# Docs**
Krebs	Aug-Sept 2016 (1mo)	9,849
Securus	Jan-May 2018 (5mo)	1,468
Banking	Jan-Dec 2017 (1yr)	97,572
Defense	Jan-Dec 2017 (1yr)	45,520

Table 2: Number of documents analyzed and time period per entity.

5 Results & Discussion

We ran experiments demonstrating the Pearson correlation coefficients between sentiment extracted from online content related to banking and defense industries, and the cyber events reported by Hackmageddon. We also conducted several case studies, to more deeply explore the potential of using Outrage as a sensor of bad behavior.

The results in Table 3 provide a comparison of Pearson correlations between the different dimensions of emotion provided by the Circumplex model (affect, intensity, and our calculated outrage), and the cyber events reported on Hackmageddon. We compare all three lexicons, from the original, data-driven lexicon from (Preoţiuc-Pietro et al. 2016), to our cyber-adapted lexicon, to the extended ANEW+ lexicon (Warriner, Kuperman, and Brysbaert 2013), for online content related to two different industries of banking and defense. Interestingly, anomalous outrage had the highest correlation for both industries, although the correlation values differed significantly across lexicons.

5.1 Case Studies

The following case studies provide the opportunity to delve deeper into the utility and challenges of using extracted outrage to predict bad behavior.

Our first case study involves Brian Krebs, an investigative journalist specializing in cyber attacks, who experienced a known DDoS attack on krebsonsecurity.com on 20 Sept 2016. Ten thousand tweets and news articles linked from tweets containing the word 'Krebs' were analyzed for sentiment, and a timeline of outrage up to the day of attack was generated, shown in Figure 5. In the timeline, there are two sharp peaks of outrage detected by the original lexicon, a few days before August 30 and again a few days before Sept 20, as well as a period of high outrage around Sept 12. After further analysis, the Sept 12 articles were discovered to be about an arrest based on a Krebs expose, with the word 'attack' (from cyber attack) contributing heavily to the detection of 'outrage,' providing an example of the potentially

[2] https://www.hackmageddon.com/2017-master-table/

Banking Industry			Defense Industry		
Lexicon	**Emotion Dim.**	r (Corr)	**Lexicon**	**Emotion Dim.**	r (Corr)
Original	affect	-0.031	Original	affect	-0.069
	intensity	-0.042		intensity	0.004
	outrage	-0.026		outrage	0.067
	anomalous outrage	0.022		anomalous outrage	**0.129**
Cyber-Adapted	affect	0.064	Cyber-Adapted	affect	-0.100
	intensity	0.093		intensity	-0.073
	outrage	0.073		outrage	0.065
	anomalous outrage	0.092		anomalous outrage	-0.040
Extended ANEW+	affect	0.084	Extended ANEW+	affect	-0.048
	intensity	-0.063		intensity	0.041
	outrage	-0.089		outrage	0.069
	anomalous outrage	**0.138**		anomalous outrage	0.111

Table 3: Dimensions of emotion correlated to cyber events (Pearson's r).

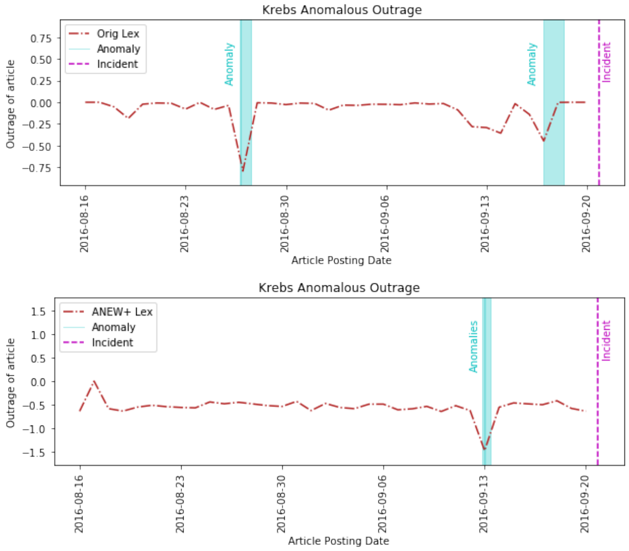


Figure 5: Timeline of outrage expressed in online content about Brian Krebs weeks prior to a DDoS attack on his website. Magenta dotted line (far right) indicates attack date; cyan shaded areas indicate anomalous periods of outrage.

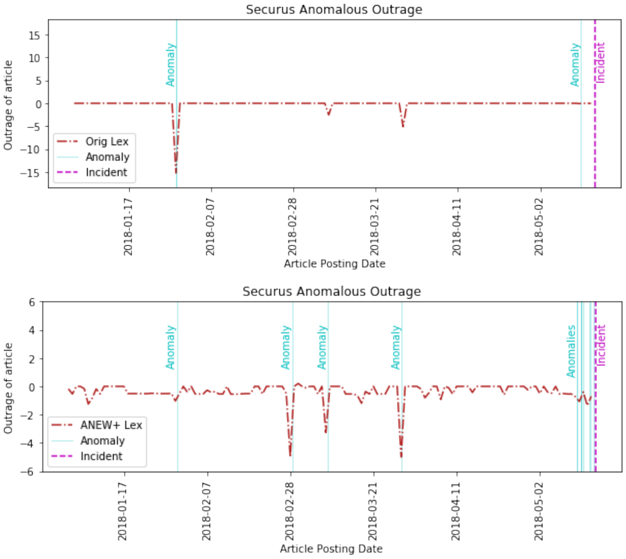


Figure 6: Timelines of outrage expressed in online content about Securus months prior to the breach of its data. Magenta dotted line (far right) indicates attack date; cyan shaded areas indicate anomalous periods of outrage.

spurious but relevant signal here. Note that the largest period of anomalous outrage (shaded cyan area) actually occurs at a period with a sudden drop in detected outrage.

In a second, more recent case study, on 16 May 2018, a hacker released thousands of logins for Securus, a company that buys phone location data from major telecom companies and then sells it to law enforcement. This incident was a different class of attack from the Krebs Case Study, with a lower-profile subject (Securus) but an attack that affects more of the public as login details were released. In Figure 6 we see that outrage was detected in far fewer documents, yet there are still several peaks and anomalous periods of outrage. Note that the ANEW+ lexicon produced many more instances of anomalous periods of outrage as compared to the original lexicon (top), or as compared to the same lexicon for a different subject/domain (left).

6 Conclusions & Future Work

While the correlations seen in Table 3 are clearly not strong ones, there were noteworthy findings. The correlations of the different dimensions of emotion differ across lexicons, even changing from negative to positive as is the case for intensity in the original versus the cyber-adapted lexicons. Thus, whereas Outrage was our original hypothesis as signalling bad behavior such as cyber attacks, we actually found that it was something a bit more primitive than the compositional elements that contribute to outrage. We found that intensity and affect, if treated independently, yield correlations that depended on the nature of the domain (the dataset) and the vocabulary (the lexicons used). Ultimately, we determined the underlying Circumplex model of emotion to be useful for discovering the utility of other features (like intensity and affect) with respect to certain domains (or datasets) and vocabulary (or lexicons). However, this signal is not a "one-size-fits-all" technique, but rather depends on the context and underlying resources. As a general technique, we need careful analysis to understand how to apply it, with respect to the domain and vocabulary.

Future work will examine a targeted sentiment paradigm, where the detected emotions are specific to an entity (e.g., a potential victim) or an event, rather than simply measuring the overall emotion expressed within an entire document. Additionally, in order to provide more detail on impending attacks, we will train separate models for each attack type and target domain, enabling the sensor to provide more specific and potentially more accurate information.

Acknowledgements

This research is supported by the Office of the Director of National Intelligence (ODNI) and the Intelligence Advanced Research Projects Activity (IARPA) via the Air Force Research Laboratory (AFRL) contract number FA875016C0114. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of ODNI, IARPA, or the U.S. Government.

References

Agnew, R. 1992. Foundation for a general strain theory of crime and delinquency. *Criminology* 30(1):47–88.

Baumeister, R. F.; Vohs, K. D.; DeWall, C. N.; and Zhang, L. 2007. How emotion shapes behavior: Feedback, anticipation, and reflection, rather than direct causation. *Personality and Social Psychology Review* 11(2):167–203.

Bermingham, A., and Smeaton, A. 2011. On using Twitter to monitor political sentiment and predict election results. In *Proceedings of the Workshop on Sentiment Analysis where AI meets Psychology (SAAIP 2011)*, 2–10.

Black, D. 2014. *The Social Structure of Right and Wrong*. Academic Press.

Bollen, J.; Mao, H.; and Zeng, X. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2(1):1–8.

Bradley, M. M., and Lang, P. J. 1999. Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical report, The Center for Research in Psychophysiology, University of Florida.

Coleman, E. G. 2011. Anonymous: From the lulz to collective action.

Ganem, N. M. 2010. The role of negative emotion in general strain theory. *Journal of Contemporary Criminal Justice* 26(2):167–185.

Haas, N. E. 2010. *Public Support for Vigilantism*. Ph.D. Dissertation, Leiden University.

Hutchins, E. M.; Cloppert, M. J.; and Amin, R. M. 2011. Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. *Leading Issues in Information Warfare & Security Research* 1(1):80–106.

Kumar, N.; Lolla, V. N.; Keogh, E.; Lonardi, S.; Ratanamahatana, C. A.; and Wei, L. 2005. Time-series bitmaps: a practical visualization tool for working with large time series databases. In *Proceedings of the 2005 SIAM Data Mining Conference*, 531–535. SIAM.

Loewenstein, G. F.; Weber, E. U.; Hsee, C. K.; and Welch, N. 2001. Risk as feelings. *Psychological Bulletin* 127(2):267–286.

Lohrmann, D. 2016. Understanding new hacktivism: Where next for hackers with a cause?

Mishne, G., and Glance, N. 2006. Predicting movie sales from blogger sentiment. In *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, 155–158.

Nguyen, T. H.; Shirai, K.; and Velcin, J. 2015. Sentiment analysis on social media for stock movement prediction. *Expert Systems with Applications* 42(24):9603–9611.

Posner, J.; Russell, J. A.; and Peterson, B. S. 2005. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology* 17(3):715–734.

Preoţiuc-Pietro, D.; Schwartz, H. A.; Park, G.; Eichstaedt, J.; Kern, M.; Ungar, L.; and Shulman, E. 2016. Modelling valence and arousal in Facebook posts. In *Proceedings of the 7th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, 9–15. San Diego, California: Association for Computational Linguistics.

Russell, J. A. 2003. Core affect and the psychological construction of emotion. *Psychological Review* 110(1):145–172.

Shaw, E.; Payri, M.; Cohn, M.; and Shaw, I. R. 2013. How often is employee anger an insider risk (I)? Detecting and measuring negative sentiment versus insider risk in digital communications. *Journal of Digital Forensics, Security and Law* 8(1):39–72.

Smallridge, J.; Wagner, P.; and Crowl, J. N. 2016. Understanding cyber-vigilantism: A conceptual framework. *Journal of Theoretical & Philosophical Criminology* 8(1):57–70.

Warriner, A. B.; Kuperman, V.; and Brysbaert, M. 2013. Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods* 45(4):1191–1207.

Wei, L.; Kumar, N.; Lolla, V. N.; Keogh, E. J.; Lonardi, S.; and Ratanamahatana, C. A. 2005. Assumption-free anomaly detection in time series. In *Proceedings of the 17th International Conference on Scientific and Statistical Database Management (SSDBM)*, volume 5, 237–242.