MIS Quarterly

RESEARCH ARTICLE

THE OPM DATA BREACH: AN INVESTIGATION OF SHARED EMOTIONAL REACTIONS ON TWITTER[1]

Eric Bachura and Rohit Valecha
Department of Information Systems and Cyber Security, Alvarez College of Business,
University of Texas at San Antonio, San Antonio, TX, U.S.A. {eric.bachura@utsa.edu} {rohit.valecha@utsa.edu}

Rui Chen
Department of Information Systems and Business Analytics, Ivy College of Business,
Iowa State University, Ames, IA, U.S.A., {ruichen@iastate.edu}

H. Raghav Rao
Department of Information Systems and Cyber Security, Alvarez College of Business,
University of Texas at San Antonio, San Antonio, TX 78249 U.S.A. {hr.rao@utsa.edu}

This paper investigates the shared emotional responses of Twitter users in the aftermath of a massive data breach, a crisis event known as the Office of Personnel Management (OPM) data breach of 2015. This breach impacted the lives of several million individuals due to the exposure of sensitive and personally identifying information. We take a data exploration approach to analyzing over 18,000 tweet messages of the ensuing discussion that took place after public notification that the breach had occurred. The resulting analysis reveals that although the emotions of anxiety, anger, and sadness may initially appear erratic, at an aggregate level, the public display of these emotions corresponds to the situational awareness of the breach event. Further, our analysis finds that this relationship extends to the sharing of emotions, indicating that those participating in the conversation congregate around a sense of shared emotional experience. Finally, an in-depth analysis of the ensuing dialogue identifies the most salient conversational drivers of these emotions, revealing breach concepts most significantly related to each emotion. Based on the results, we present propositions that draw from this analysis to inform emotional response characteristics that emerge over the duration of such crisis events. The results of this study can inform organizational practices and policy making in the context of response to crisis events such as data breaches.

Keywords: Data breach, emotional response, situation awareness, OPM hack, Twitter, data-driven research, crisis management, word-to-vector, W2V

Introduction

Data breaches represent a significant risk faced by organizations that involve unauthorized access to data, often personal information, facilitated by a wide variety of means (e.g., hacking, phishing, insider threat activity, etc.) (Culnan & Williams, 2009). According to Verizon's *Data Breach Investigation Report* (Verizon, 2017) personal data represents the most frequent type of data that is compromised as a result of data breaches. Although reports vary on the frequency and volume of data breaches, there is widespread agreement that the threat of data breaches is increasing.

Prior research in this area has focused on a variety of topics: trends in data breaches (Garrison & Ncube, 2011; Holtfreter & Harrington, 2015), practitioner-oriented prevention and mitigation techniques (Brown, 2016; German, 2016) and the economic costs of data breaches (Acquisti et al., 2006; Layton & Watters, 2014). These studies have examined data breaches from the perspectives of technology, organizational victims, and attack or attacker characteristics (Goode et al., 2017). While

[1] Sirkka Jarvenpaa was the accepting senior editor for this paper. Heng Xu served as the associate editor.

there is substantial anecdotal evidence about the emotions and stresses experienced by the victims of a data breach (Guynn, 2020), research on the nature of the emotions of people reacting to large-scale data breaches is scarce, as is research that examines shared emotional responses of people to large-scale data breaches on social media. Gaining an understanding of the emotional perspective in IT security research is important because emotions spur reactions among people that, in turn, impact their social transmission of information (Berger, 2011), which could have implications for security management. For instance, anecdotal information shows that different emotions could influence organizational information security.[2] Studying shared emotions unveils the fundamental response of members of the concerned public and informs our understanding of ensuing actions. Examining shared emotions can specifically help to improve understanding of social media usage behavior (Chmiel et al., 2011; Jalonen, 2014) as well as in the context of data breaches.[3]

The Office of Personnel Management (OPM) data breach of 2015 (Chaffetz et al., 2016) impacted millions of government personnel, their families and friends, as well as anyone these personnel provided as references for work, education, and character aspects of the background investigation process used to vet clearances (Adams, 2016). OPM notified the nation about the crisis on June 4, 2015, and for two months, there was intense discussion on social media (e.g., Twitter) until the OPM director eventually resigned. To set a clear scope to the current research, we focus on Twitter to understand the fluctuations and transitions of shared emotions. Twitter has received increased attention from researchers interested in capturing the attitudes and sentiment of social activity (Bang et al., 2021).

In this paper, we examine the following research questions in the aftermath of the large-scale OPM data breach:

RQ1: *How do negative emotions, expressed on social media platforms such as Twitter, change over time in the context of the data breach?*

RQ2: *What are the effects of negative emotions on the social transmission of information online?*

RQ3: *What are the key characteristics of the data breach that are associated with the shared negative emotions expressed in the online discussion?* We identify three stages of situational awareness and use measures of key negative emotions within and across these stages. We then explore important characteristics of the response to data breaches in the different stages.

The rest of the paper is organized as follows. First, we present background on the OPM data breach, discuss related research, and briefly present theoretical perspectives. We then present and apply the concept of situational awareness. Next, we introduce the research data, analysis, and results. Finally, we develop research propositions and end with a discussion and the conclusion.

Background and Literature Review

In this section, we discuss social response to the breach event, emphasizing the role of change and emotions. We first present a background on the 2015 OPM data breach. Then, we review the literature on data breaches, change, emotions, and social media. Finally, we present a discussion of situational awareness theory and its utility in this study through an examination of the key events of the 2015 OPM data.

2015 OPM Data Breach

The 2015 OPM data breach stands as one of the most impactful data breaches in terms of PII that was compromised with over 5.6 million fingerprint records, 4.2 million personnel files, and 21.5 million individual background investigation files stolen (Chaffetz et al., 2016). It is important to note that the millions of individual background investigation files consist of data that can be found on what is known as an SF-86 (https://www.opm.gov/forms/pdf_fill/sf86.pdf). This form is designed to capture a broad and deep range of personal information. This information goes well beyond what is normally considered traditional PII, such as social security number, home address, phone numbers, and email addresses. These forms contain information regarding friends and family members of the individual requesting the clearance. As a result, the number of people impacted by this data breach extends to a broader audience after the event. This is demonstrated in FBI Director James Comey's remark regarding the OPM data breach: "My SF86 lists every place I've ever lived since I was 18, every foreign travel I've ever taken, all of my family, their addresses. So it's not just my identity that's affected. I've got siblings. I've got five kids. All of that is in there" (Chaffetz et al., 2016, p. iii).

Notification about the OPM data breach to the nation occurred on June 4, 2015. The two months following the announcement saw a series of news events revealing increasing details about the size and scope of the data that was compromised as well as the events that lead to the data breach. The end of this two-month period following the initial announcement was marked

[2] https://www.securityindustry.org/2020/12/15/the-impact-of-emotions-on-corporate-crisis-management/

[3] https://www.todayslegalcyberrisk.co.uk/partner-news/psychology-and-data-breaches-the-emotional-impact-of-privacy-violations/

by the resignation of OPM Director Katherine Archuleta following the acknowledgment that 21.5 million records of background investigation data were compromised (Chaffetz et al., 2016). Because of the significance of the OPM hack, we use this breach as the context for the current research and investigate the emotional responses that followed.

Data Breaches: Information and Emotion

Data breaches inevitably bring changes to the equilibrium between the defensive and offensive paradigms of cyberspace. Incidents of data breaches signal an adverse change in data security protocols and processes. They also indicate the increase of risk in the use of the internet for personal and business purposes (Berinato, 2020). A major data breach can serve as a societal signal that prior perceptions and trust placed in authorities and organizations associated with the breach have been violated. In response to these violations of trust and perception, changes in expectations can be driven by both information and emotion, creating new perceptions about the world (e.g., a move towards the perception that cyberspace is less secure).

Emotional response is the normal and natural reaction to change (James et al., 1998). Emotions are experienced in response to changes in our lives and are driven by situations requiring change (Goldsworthy, 2005). Breach incidents will lead concerned individuals to react to the changed environment with emotional reactions. Negative emotions may surface in response to concerns over breach details, potential attacker motivations, and incident management.

In the aftermath of data breaches, social networking sites such as Twitter can provide a platform for communication. This can be used to convey crisis-related information as well as emotional information, while also allowing the public to share or retweet what is conveyed (Austin et al., 2012; Rao et al., 2020; Venkatesan et al., Forthcoming). In 2016, Hao and Dai suggested, "little is known about how people view security breaches on social media platforms" (2016, p. 857). Since then, a few studies have analyzed data breaches using the social media perspective. Novak and Vilceanu (2019) have examined discourses of Twitter users in the aftermath of the 2017 Equifax breach, while others have investigated the sharing of software vulnerability information (Syed et al., 2018). A review of additional literature examining data breaches, seen in Table 1, illustrates the lack of studies fully exploring the emotional characteristics of responses to a data breach.

We know that people react to disruptive change with a rollercoaster of emotions, as described by a variety of *change curve* models (Elrod & Tippett, 2002). Change curve models have been used in the change management field as a template to examine the impact of change (Shoolin, 2010) and organizational change (Goodman & Loh, 2011). Indeed, practitioners have utilized these models in a variety of contexts, including business (Belyh, 2020) and cybersecurity (Valentine, 2017). Change curve models transition from normalcy through some form of disruption and then to a redefined state of normalcy (Elrod & Tippett, 2002). Data breaches serve as disruptions in trust, privacy, and organizational operations. Thus, an understanding of change curve models implies that there is an understudied element to data breaches: emotions in reaction to the disruptive changes that breaches affect. This paper explores this understudied element of data breaches.

Stages of Emotional Response: Situation Awareness Theory

Before we can explore the emotional responses to data breaches and the changes they force on us, we must first explore the data to seek to identify boundaries for when changes are occurring or when they are perceived to occur. To do this, we take guidance from Endsley's situation awareness (SA) theory (1995). Using SA theory, we can categorize the public's awareness of breach information as it changes over time. This will allow our data exploration to determine whether emotions change, in aggregate, in accordance with the changing information landscape. SA theory describes our orientation with the information landscape as being categorized into one of three stages: perception, comprehension, and projection. This orientation is the result of information processing or the cognitive reaction to stimuli. Because cognitive functions are heavily involved in emotional responses (Izard et al., 1984), being able to categorize crisis information using SA theory will inform any understanding of the overall pattern of emotional responses to a crisis event (Oh et al., 2011).

Categorizing OPM Events into SA Stages

The two months following public notification of the OPM data breach saw numerous news events revealing increasing details about the size and scope of the data that was compromised. These news events also provided clarification and contextualization of the events surrounding and leading up to the data breach. Toward the end of this period, OPM Director Katherine Archuleta acknowledged publicly that 21.5 million records of background investigation data were compromised and submitted her resignation the following day. These news events can be understood from the perspective of the three temporal stages of SA theory.

We qualitatively assess the content of key events (as detailed in the congressional report and summarized in Table 2) to define the boundaries of each stage chronologically as they are represented in the content of the news events (Chaffetz et al., 2016).

Table 1. Data Breach Literature Sample

Study	Area of focus (organizations, data subjects, industry, etc.)	Sources of data (design or organic)	Context	Key findings
(Goode et al., 2017)	Organizations	Longitudinal field study of Sony customers	Modified assimilation-contrast model	Contribute to the research on data breaches and service failure by demonstrating the impacts of compensation on customer outcomes
(Angst et al., 2017)	Industry	IT Panel data	Model the heterogeneity in the likelihood of a breach	Deeper integration of security into IT-related processes and routines leads to fewer breaches
(Culnan & Williams, 2009)	Organizations	Two high-profile data breaches experienced by two U.S. companies, ChoicePoint and TJX	Illustrate arguments for enhancing organizational level privacy programs based on ethical reasoning	Contributes to the dearth of prior organizational-level privacy research
(Liu et al., 2020)	Organizations	Sample of 504 U.S. higher-education institutions over a four-year period	How an important IT governance mechanism affects the likelihood of cybersecurity breaches	Findings highlight the tradeoff between granting autonomy and flexibility in the use of information systems and enforcing standardized, organization-wide security protocols
(Aliyu et al., 2020)	Organizations	Survey of international citizens of the U.S.A.	This study seeks to determine the factors that individuals, particularly those outside the U.S.A. deem important when considering providing information for EHRs	The intent of international citizens to provide personal health information depends on trust, risk, privacy, and perceived benefits
(Sarkar et al., 2020)	Organizations	Prominent professional healthcare groups: physicians, nurses, and support staff	Conducted an exploratory qualitative study to identify different attitudes toward ISP violations	The substantial effect of professional subculture on ISP violations in organizations and insights for researchers and managers that may be used to improve overall ISP compliance
(D'Arcy et al., 2020)	Industry	Publicly available data on firms' data breach incidents	Elaborate relationships between aspects of a firm's CSP and the likelihood of experiencing a data breach	Results suggest that firms that are noted to have poor CSP records (i.e., CSP concerns) are no more likely to experience a data breach
(Syed, 2019)	Organizations	Twitter postings related to the 2014 Home Depot data breach	Taxonomy of data breach frames and subframes and the related reputation threats	Reputation threats vary for intentional, accidental, and victim data breach frames
Current paper	Emotions	Twitter postings related to 2015 OPM data breach	Explore emotional response characteristics following a data breach	Emotions are strongly aligned with situational awareness levels; dominant emotions correspond to message sharing activity; breach concepts identified

Table 2. Key Information Events Grouped by SA Stages

Date	Event	Stage
June 4, 2015	Public notification occurs	Perception
June 8, 2015	US-CERT determines information stolen contained many sources of PII, news outlets relay information to the public citing anonymous sources	
June 12, 2015	OPM announces second data breach where security clearance data may have been compromised	
June 16, 2015	OPM director attends hearing before the House Committee on Oversight and Government Reform	Comprehension
June 24, 2015	OPM chief information officer (CIO) testifies before congress	
July 9, 2015	OPM issues press release regarding background investigation data compromise	
July 10, 2015	OPM Director Katherine Archuleta resigns	Projection

These events are recognized as the major events in the OPM breach timeline (Sternstein & Moore, 2015; Bisson, 2015).[4] The first three key events, occurring on June 4, June 8, and June 12, provide the initial information elements regarding the data breach.

From these events, the public learned that the organization suffering the data breach was the Office of Personnel Management, that federal employees were the primary group of individuals affected, and that security clearance data was the type of data that was breached. These first three events provide the initial contextualization of the incident and thus can be used to frame the initial perception of the data breach. Together these events provide the fundamental elements that are necessary before observers can develop a comprehensive understanding of the situation.

The fourth event on June 16 introduces an organizational representative to the public and represents the start of the comprehension stage. The comprehension stage is facilitated through the use of congressional hearings (Govinfo, 2017). The fifth event on June 24 represents further hearing testimony. The sixth event is an organizational report that represents the victim organization's comprehension of the data breach. These three events collectively represent the comprehension stage by providing elaboration and detail that the key events corresponding to the perception stage lacked.

The seventh event represents the start of the projection stage as the OPM director completed her testimony in front of congress and submitted her resignation. The resignation of the chief representative of the OPM served as a change effected within the offending organization. This marked the beginning of organizational action and the start of the projection stage of the news events, answering questions such as what should happen now that the event has been understood. The resulting groupings of these events into SA stages is shown in Table 2 (Chaffetz et al., 2016).

Study Design and Analysis

The growing abundance of datasets generated as a result of organic behaviors and activity is recognized as a valuable opportunity for researchers to develop insights and theory through a process of intensive data exploration using powerful analytic techniques (Berente et al., 2019; Xu et al., 2020). Using a range of analytic techniques, we conduct exploratory research on Twitter data to gain an understanding of how a data breach drives the emotions of those reacting to it.

The Dataset

Within a day of the public announcement that OPM had been breached, the hashtag #OPMHack started being used on Twitter. The nature of Twitter as a social dialogue medium presents an opportunity to conduct sentiment and content analysis of tweet messages regarding the OPM data breach. The time frame to bound the dataset is from the point of initial notification up until the end of the month containing

[4] This data was retrieved from the congressional report (Chaffetz et al., 2016) but can also be found in a number of OPM timeline resources (Sternstein & Moore, 2015; Bisson, 2015)

the resignation of OPM Director Katherine Archuleta. This provides an opportunity to measure fluctuations in tweet message characteristics as they change over time in response to breach-related events (e.g., breach details, leadership announcements, resignations, etc.).

The dataset was purchased from a third-party vendor of Twitter data and consisted of 18764 tweet messages posted on Twitter from June 4, 2015, to July 31, 2015. All messages contained the hashtag #OPMHack. A summary view of daily tweet counts, coinciding key events, and SA boundaries is illustrated in the graph presented in Figure 1.

The investigation of social response to the OPM breach in this paper is divided into 3 phases: Phase 1 explores the emotional content of messages, identifying key emotions in each stage and how these emotions change over time. Phase 2 examines the effect of emotions on social media sharing (retweets). Phase 3 explores breach-related characteristics that are associated with emotions. The three-phased & Krüger, 2011; Verma et al., 2011).

Text analysis techniques such as those represented by the Linguistic Inquiry and Word Count tool (LIWC) (Tausczik & Pennebaker, 2010) or the Natural Language Toolkit libraries (NLTK) (Bird et al., 2009) are used for the analysis of emotional measures (Brown et al., 2013; Yin et al., 2014) and are known for reliability and validity (Bantum & Owen, 2009; Pennebaker et al., 2015). LIWC calculates scores of the emotional dimensions of anxiety, anger, and sadness as a proportion of words that fall within a validated psychometric category.

For the temporally bound dataset used in this study, where text segments are time stamped, the sentiment analysis can help determine whether any emotional dimensions stand out as clear signals among the collection of emotions expressed over that time and within the defined group. We use sentiment scores to understand how the emotional signals change over the course of a crisis event.

Phase 1

In this phase, we examine the negative emotions of anxiety, anger, and sadness expressed by Twitter users in the aftermath of the OPM hack. These emotions were the dominant emotions, implying that they were the most expressed in our initial data analysis. We used multiple sentiment scoring techniques to identify the highest scoring emotions (see the Appendix for more detail), and used SA stages to break the chronological development of OPM reactions into three stages (perception, comprehension, and projection). Thus, this phase focuses on RQ1—How do negative emotions, expressed on

investigation is bounded within situation awareness stages of perception, comprehension, and projection. A depiction of the three-phased process and the overall data exploration process is illustrated in Figures 2 and 3, respectively. A detailed discussion of the data analysis process can be found in the Appendix.

Measuring Emotion in Tweets

The core phenomenon of interest in our data exploration is emotional expression in response to a data breach event. To facilitate this, the sentiments of communications within social media messages need to be captured. Sentiment analysis combines natural language processing and text analysis to extract emotion information from the text (Liu, 2010; Stieglitz & Krüger, 2011). It is an established method for measuring what people feel, capturing a range of opinions, emotions, and language characteristics. It has been used in the context of a variety of crisis events (Biever, 2010; Lee et al., 2015; Stieglitz

social media platforms such as Twitter, change over time in the context of the data breach? Findings will inform the development of research propositions on the emergence and transition of data breach emotions.

Analysis

To test for the peak of an emotion across stages, a Games-Howell post hoc was performed in SPSS. We used the Games-Howell post hoc because of its ability to analyze the statistical significance of differences among multiple groups (Games & Howell, 1976). To analyze and compare the emotions of anxiety, anger, and sadness in the same stage (and drawing from the same data), we utilized ANOVA with repeated measures and the Bonferroni test.

Statistical Results

Comparing anxiety levels across the different chronological crisis stages revealed a statistical significance in the difference in anxiety between the perception and comprehension stages and the perception and projection stages (Table 3). Comparing anxiety to other emotions in the SA perception stage reveals that anxiety is significantly higher than the other emotions in this stage. Together, this demonstrates that anxiety peaks in the perception stage, where it is also the dominant emotion. Likewise, our results show that anger is the dominant emotion in the comprehension stage but that the peak of anger in the comprehension stage is not significant (Table 4). Finally, sadness peaks in the SA projection stage, but its difference from anger in that stage is not significant, meaning we cannot determine if it or anger is the dominant emotion of that stage (Table 5).

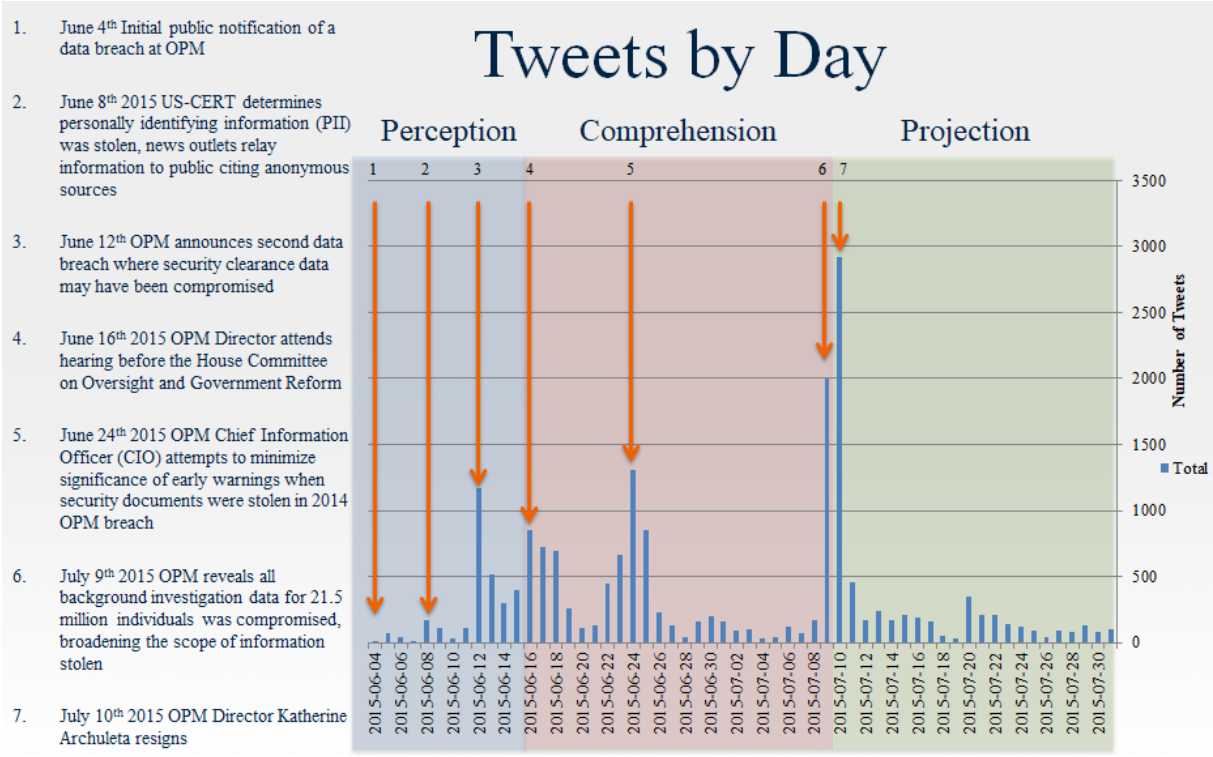


Figure 1. Tweet Messages by Day

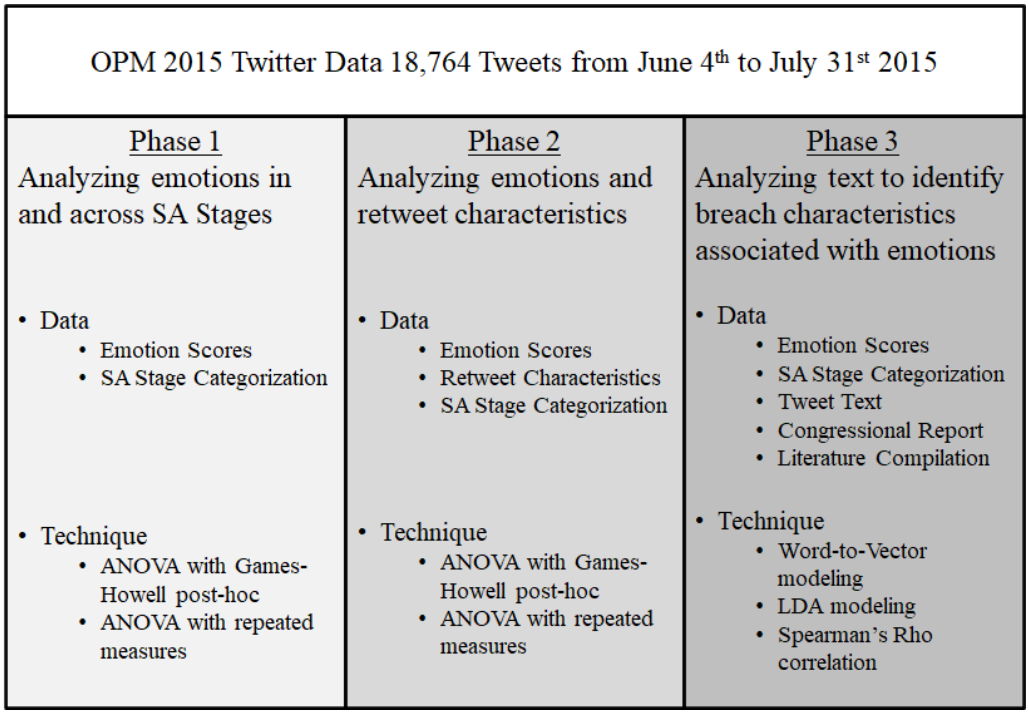


Figure 2. Phases of Data Exploration

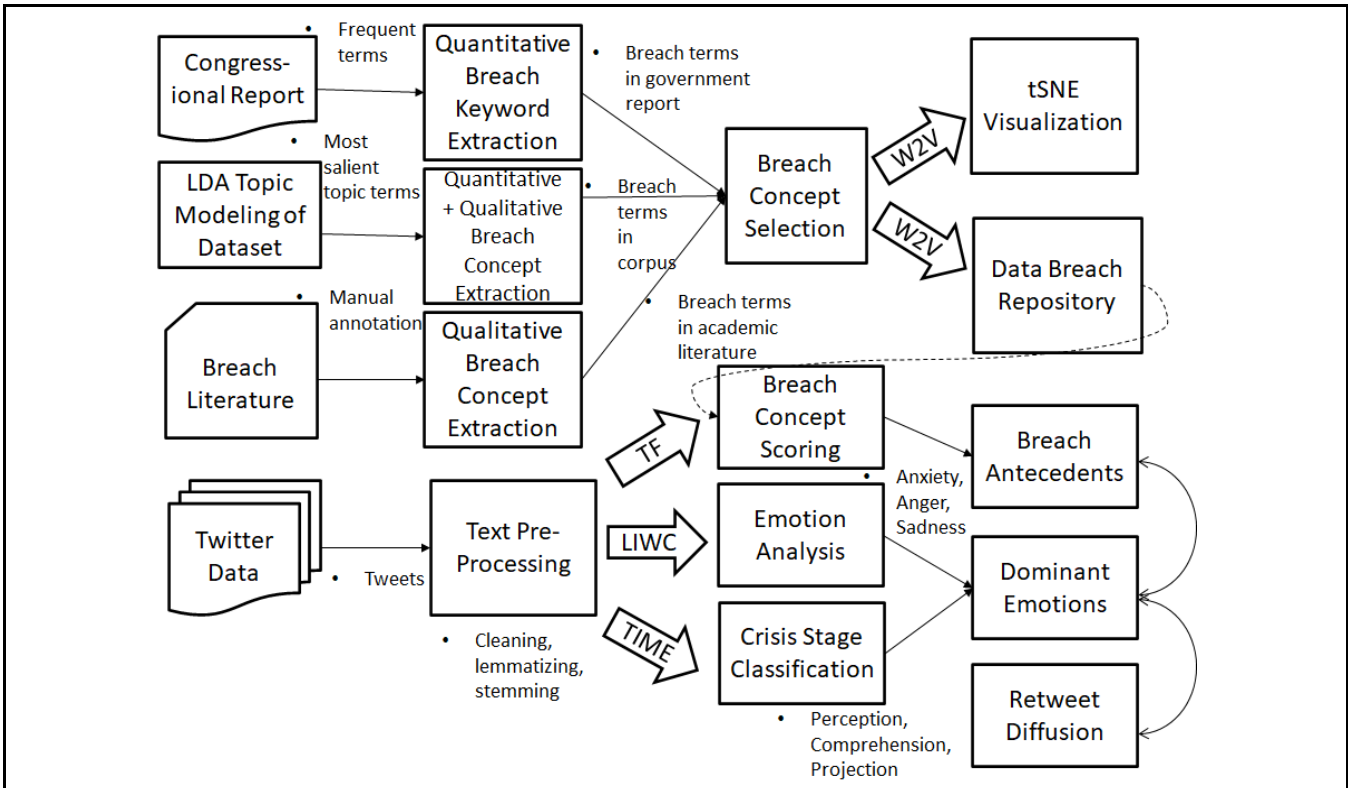


Figure 3. Overall Data Exploration Process

Table 3. Tests of Anxiety

Anxiety: *Games-Howell (across stages)*						
(I) SA stage	**(J) SA stage**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% conf. intervals**	
					LL	**UL**
Perception	Comprehension	0.367	0.038	0.000	0.277	0.456
Perception	Projection	0.359	0.039	0.000	0.267	0.451
Perception stage: *Repeated measure (within stage)*						
(I) Emotion	**(J) Emotion**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Anxiety	Anger	0.128	0.044	0.010	0.024	0.233
Anxiety	Sadness	0.453	0.042	0.000	0.353	0.554

Table 4. Tests of Anger

Anger: *Games-Howell (across stages)*						
(I) SA stage	**(J) SA stage**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Comprehension	Perception	0.089	0.040	0.063	-0.004	0.182
Comprehension	Projection	0.098	0.031	0.004	0.026	0.170
Comprehension stage: *Repeated measure (within stage)*						
(I) Emotion	**(J) Emotion**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Anger	Anxiety	0.328	0.024	0.000	0.271	0.384
Anger	Sadness	0.377	0.024	0.000	0.320	0.435

Table 5. Tests of Sadness

Sadness: *Games-Howell (across stages)*						
(I) SA stage	**(J) SA stage**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Projection	Perception	0.337	0.034	0.000	0.258	0.416
Projection	Comprehension	0.300	0.029	0.000	0.232	0.368
Projection stage: *Repeated measure (within stage)*						
(I) Emotion	**(J) Emotion**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Sadness	Anxiety	0.242	0.031	0.000	0.169	0.316
Sadness	Anger	0.020	0.035	1.000	-0.064	0.104

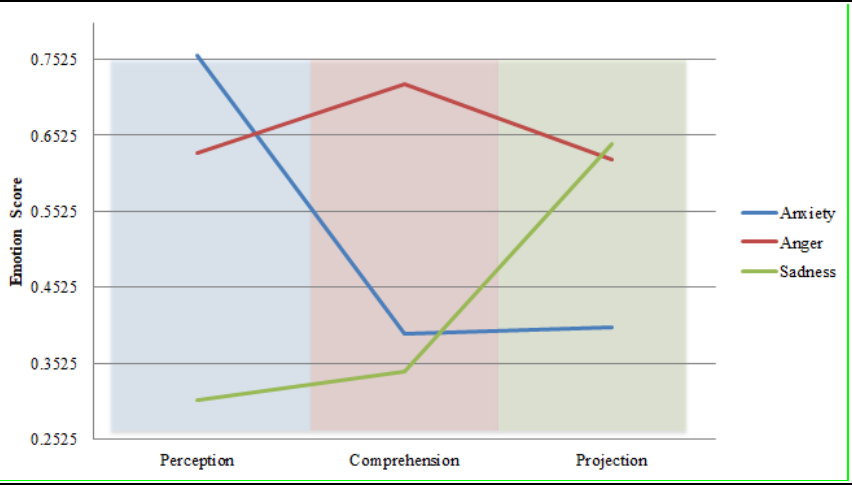


Figure 4. Means Plot of Emotion Scores

The analysis of anxiety, anger, and sadness reveals that there are emotion-specific responses according to crisis stages in post-breach situations. An examination of the means plot for each emotion across the three SA stages provides a visual illustration (Figure 4).

Proposition Development

In the analysis of the OPM Twitter data, we see the following: (1) Anxiety is significantly higher in the perception stage than in the comprehension ($p < 0.01$) and projection ($p < 0.01$) stages. (2) Anxiety is significantly higher than anger ($p < 0.01$) and sadness ($p < 0.01$) in the perception stage. The first finding suggests that anxiety dominates the early stage of a crisis (chronological peak). Prior research has shown that there is a correlation between uncertainty and anxiety (Berenbaum et al., 2008) and that uncertainty is a necessary condition for anxiety of any kind (Dugas et al., 2005). In the early stages of a crisis, information serves to confirm that a crisis has occurred but a lack of details can foment uncertainty about things such as the scope, scale, and consequences. As a result, in comparison with later stages, we found that uncertainty is highest in the perception stage. Thus, as described by research, uncertainty likely facilitates the higher anxiety in the perception stage.

The second finding suggests a within-stage dominance of anxiety (emotional peak). While anxiety is at the peak of its development, the perception stage does not offer fertile ground for the growth of anger or sadness. The lack of comprehension and early nature of the breach inhibits the introspection and retrospection that would facilitate other feelings such as anger or sadness. In the case of data breaches, Twitter users may not feel angry when the insights of the breach and exact impacts are yet to be confirmed. Further, the concerned public is less likely to experience sadness in this stage because the consequences of a breach that would evoke sadness are not yet fully understood or anticipated. Considering the above, we propose that in the context of a data breach:

P1.1: *Anxiety is **(a)** dominant (i.e., more prevalent than anger and sadness) among the shared emotions in the perception stage, and **(b)** anxiety is also more prevalent in the perception stage than the comprehension and projection stages.*

Analysis results show that (1) anger is marginally significantly higher in the comprehension stage than in the perception stage ($p < 0.10$) and significantly higher than in the projection stage ($p < 0.01$), and (2) that anger is significantly higher than anxiety ($p < 0.01$) and sadness ($p < 0.01$) in the comprehension stage. The first finding suggests a between-stage dominance of anger (chronological peak). Anger is an emotion that involves a perception that trusts and beliefs have been violated (Grégoire & Fisher, 2008). As an example, people get angry after failed service encounters and may find it difficult to forgive the provider (Grégoire et al., 2009). Over time, understanding the negligence that led to the data breach also becomes publicly understood and reported, resulting in psychological reactance and higher anger levels, as compared to other SA stages (Brehm & Brehm, 2013; Roeckelein, 2006). Individuals are also likely to engage in the attribution of blame to certain people or processes (Su, 2014). Such attribution will likely evoke the individual's emotional reactions of anger. Compared with other stages, the comprehension stage provides concerned individuals a context to understand how the focal organization has failed to provide data security. Such information is unavailable in the perception stage and its value in triggering anger then diminishes in the projection stage because perceived violation and blame attribution are already set in the comprehension stage.

The second finding suggests a within-stage dominance of anger (emotional peak). The feeling of anger is a powerful survival tool that provides a surge of energy, making people feel more in charge rather than vulnerable or helpless (Pratt, 2014). In contrast, sadness emerges when people exhaust other emotional responses (e.g., anger) that offer them control in fighting the adverse consequences that are caused by the change. Lacan (1977) noted that people resort to anger in order to protect themselves from or cover up other vulnerable feelings, which may include sadness—a feeling of hopelessness and emptiness. As a result, we expect sadness to be less developed when individuals are still occupied by anger in the comprehension stage. Further, because of the high level of certainty that comes from a comprehensive understanding of the situation, the lack of anxiety would be evident in this stage. In accordance, we propose:

P1.2: *Anger is **(a)** dominant (i.e., more prevalent than anxiety and sadness) among the shared emotions in the comprehension stage, and **(b)** anger is also more prevalent in the perception stage than in the comprehension and projection stages.*

Our analysis reveals that (1) sadness is significantly higher in the projection stage than in either the perception stage ($p < 0.01$) or comprehension stage ($p < 0.01$), and (2) sadness is significantly higher than anxiety ($p < 0.01$) but not anger ($p > 0.10$). The first finding suggests a between-stage dominance of sadness (chronological peak). Prior research has demonstrated that a holistic appraisal of negative events is associated with sadness (Baumeister et al., 2001; Tiedens & Linton, 2001). Compared with other stages, the projection stage may have greater intensity of sadness because introspection and retrospection are more prevalent in this stage than reactance. Concerned people can deeply reflect on the changes because of the breach event. They may feel sad because they have a connection with what is lost (Pratt, 2014). In the data breach context, the changes may result in the loss of trust in the government or organization entrusted with their valuable personal information.

The second finding gives partial support to a within-stage dominance of sadness (emotional peak). While sadness grows in the projection stage, anxiety disappears because uncertainty is largely absent in the projection stage. Anger, however, demonstrates a prolonged exhibition, coexisting with sadness in this stage. It is possible that, although this stage is characterized by the first evidence of the public's successful projection of their will onto the breach event (with the resignation of Archuleta following public outcry), it is still insufficient to satiate the broader anger about failed leadership. Sadness is characterized by feeling a lack of control, such as hopelessness (Cherry, 2019) and its occurrence implies that one has depleted defensive emotional responses such as anger, which is part of our instinct for protection and preservation (van Warmerdam, 2016, p. 1). Considering the above, we propose:

P1.3: *Sadness **(a)** is an emerging dominant emotion (i.e., more prevalent than anxiety and anger) in the projection stage, but **(b)** its emergence is slower to give way to anger in the comprehension stage than other dominant emotions are in their stages.*

Phase 2

The social sharing of emotions (SSE), where people share emotional experiences with others, is an important source of interpersonal interaction (Hidalgo et al., 2015). Rimé (2009) has pointed out that SSE is a mechanism that individuals engage in to reduce their state of cognitive dissonance and perhaps recover from an emotional experience. The phenomenon of the sharing of emotions resulting from an emotional event becomes particularly important in social media because of its reach and the speed at which communications spread. In this phase, we examine the sharing of emotions in the post-breach social response in the form of retweets. Thus, this phase responds to

RQ2—What are the effects of emotions on the social transmission of information online? We analyze the Twitter data to understand how the dominant emotions analyzed in Phase 1 are associated with message sharing through retweet activity.

Emotion plays a significant role in influencing the behavior of concerned individuals (Deng & Poole, 2010; Stein et al., 2015; Tsai & Bagozzi, 2014). The literature on SSE (Christophe et al., 2008; Curci & Bellelli, 2004; Hidalgo et al., 2015) discusses factors that shape emotion-sharing behaviors (e.g., higher emotional intensity leads to more sharing frequency) (Berger, 2011). However, prior studies have not investigated whether the dominance of emotions in the situational awareness stages may affect information sharing differently. That is, does a dominant emotion receive more sharing than other nondominant emotions according to the SA stage it is in? This is an important question because, as prior literature has pointed out (Christophe et al., 2008), the sharing of (negative) emotions with others elicits emotional reactions in receivers, thus triggering the potential for collective action, for example, through mobilization (Stürmer & Simon, 2009).

Analysis

In this phase, we examine the sharing patterns among a collected set of emotions resulting from the OPM data breach. That is, we compare retweet measures of tweets across the different crisis stages (perception, comprehension, and projection) and compare retweet measures of anxious, angry, and sad tweets in each stage. We follow the same statistical analyses and methods as described in Phase 1. This helps us answer the question of whether the dominant emotion in each stage will be shared differently than nondominant emotions.

Statistical Results

The analysis of tweet sharing suggests that retweet patterns of dominant emotions differ from that of nondominant emotions. Retweeting behavior for anxiety-laden tweets show statistically significant and higher mean retweets in comparison to other emotions in the perception stage, while also demonstrating a statistically significant higher mean retweet in the perception stage than in the other stages (Table 6). Likewise, retweeting behavior for anger-laden tweets also demonstrate statistically significant and higher mean retweets both in comparison to retweets of other emotionally laden tweets in the comprehension stage and to the retweets of anger-laden tweets in other stages (Table 7). However, the retweeting behavior of sad tweets is nondominant in the projection stage, showing a significant but lower mean retweet behavior in comparison to angry tweets, and a mixture of marginally and significant results in demonstrating higher mean retweet behavior in the projection stage in comparison to other stages (Table 8).

These results indicate that the emotion of tweets, and whether they are dominant or nondominant, differently affect the social sharing of the information. Particularly noticeable is the high retweet of anxiety in the perception stage and its drop-off in the comprehension stage. This is visualized in a means plot (Figure 5).

Proposition Development

Our results show the following: (1) Tweets laden with language indicating anxiety are retweeted significantly more in the perception stage than in the comprehension ($p < 0.01$) and projection ($p < 0.01$) stages, and (2) anxious tweets are retweeted significantly more than angry ($p < 0.01$) and sad ($p < 0.01$) tweets in the perception stage. The perception stage is characterized by uncertainty, the cognitive equivalent of anxiety (Cushman & Kovacic, 1995; Grupe & Nitschke, 2013). We see from the analysis that anxiety-laden retweets dominate those of anger and sadness in the perception stage. This reinforces the clear relationship between uncertainty and anxiety (Gentes & Ruscio, 2011) as uncertainty is dealt with by active discussion (i.e., retweeted messages). Such discussion may also indicate that anxiety in the perception stage increases action-related behaviors such as sharing of information (Berger & Milkman, 2012).

We attribute the finding that anxiety-laden tweets receive more sharing in the perception stage than in other stages to the need for *information search* (Wang et al., 2010, 2012). Uncertainty is mitigated by information seeking. In sharing anxious tweets about OPM (e.g., one is worried about the breach but lacks information to get questions answered), Twitter users spread words about the breach in the hope that increased message coverage will lead to the provision of additional information (Oh et al., 2018). The need for information searching is at its peak in the perception stage, which explains why the observed effect is strongest therein.

We attribute the finding that anxiety-laden tweets receive more sharing than anger or sadness in the perception stage to the fact that anxiety is so closely associated with uncertainty and the information-seeking behavior of users unifies around the desired reduction of uncertainty. That is, users are more likely to trigger reactions (e.g., retweeting) when their message expresses a shared uncertainty and resulting anxiety than they are if it expresses other emotions in this stage. Based on the results and above logic, we propose:

P2.1: *The sharing (retweet) of anxious tweet messages is **(a)** higher than the sharing of tweets involving other emotions in the perception stage, and **(b)** higher in the perception stage than it is in other stages.*

Table 6. Tests of Retweets of Anxious Tweets

Anxiety RTS: *Games-Howell (across stages)*						
(I) SA stage	**(J) SA stage**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Perception	Comprehension	14.282	0.969	0.000	12.01	16.55
Perception	Projection	14.335	0.969	0.000	12.06	16.61
Perception stage: *Repeated measure (within stage)*						
(I) Emotion	**(J) Emotion**	**Mean Difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Anxiety	Anger	13.648	0.982	0.000	11.297	15.999
Anxiety	Sadness	14.161	0.975	0.000	11.827	16.496

Table 7. Tests of Retweets of Angry Tweets

Anger RTS: *Games-Howell (across stages)*						
(I) SA stage	**(J) SA stage**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Comprehension	Perception	2.545	0.213	0.000	2.05	3.04
Comprehension	Projection	1.988	0.217	0.000	1.48	2.50
Comprehension stage: *Repeated measure (within stage)*						
(I) Emotion	**(J) Emotion**	**Mean difference (I-J)**	**Std. error**	**Sig.**	**95% Conf. intervals**	
					LL	**UL**
Anger	Anxiety	3.179	0.180	0.000	2.747	3.611
Anger	Sadness	3.249	0.179	0.000	2.820	3.677

Table 8. Tests of Retweets of Sad Tweets

Sadness RTS: *Games-Howell (across stages)*						
(I) SA stage	**(J) SA stage**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Conf. Intervals**	
					LL	**UL**
Projection	Perception	0.232	0.104	0.064	-0.01	0.48
Projection	Comprehension	0.423	0.081	0.000	0.23	0.61
Projection stage: *Repeated measure (within stage)*						
(I) Emotion	**(J) Emotion**	**Mean Difference (I-J)**	**Std. Error**	**Sig.**	**95% Conf. Intervals**	
					LL	**UL**
Sadness	Anxiety	0.406	0.077	0.000	0.221	0.591
Sadness	Anger	-0.838	0.149	0.000	-1.194	-0.481

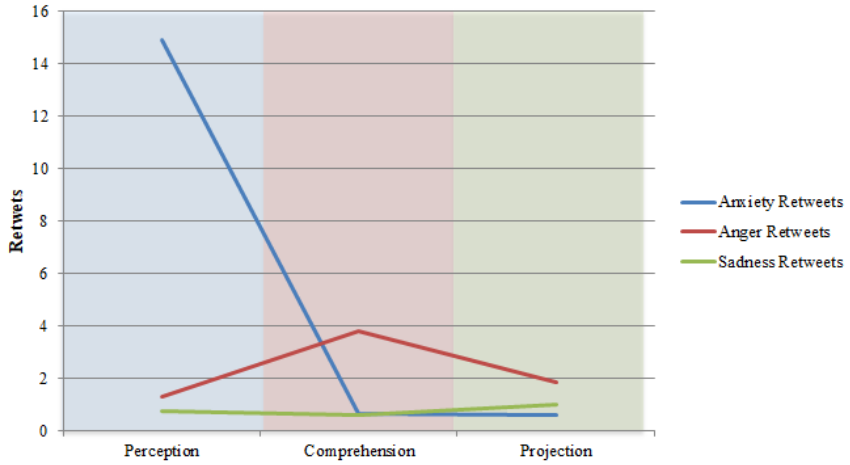


Figure 5. Means Plot of Retweets

Similar results are seen with respect to anger in the comprehension stage. (1) Angry messages have significantly more retweets in the comprehension stage in comparison to the perception ($p < 0.01$) and projection ($p < 0.01$) stages. (2) Angry messages in this stage are retweeted significantly more than anxious ($p < 0.01$) and sad ($p < 0.01$) messages.

The connection between anger and conflict resolution is theorized by evolutionary psychology (Sell et al., 2009). Drawing on this research, we suggest that the fact that anger-laden tweets receive more sharing in the comprehension stage than in other stages can be ascribed to *getting needs met* (Sell et al., 2009), an end goal of conflict resolution. By sharing tweets that voice anger, Twitter users help get their shared requests heard by policy makers. The desire to get needs met is strongest in the comprehension stage because individuals comprehend loss induced by the data breach. This informs an understanding that they have needs to be fulfilled. In contrast, in the perception stage, concerned individuals have insufficient information about the breach to develop an understanding of their needs. Therefore, getting their needs met is less of a concern. The projection stage relates to the period when the concerned public anticipates some type of resolution and there is less necessity for needs to be met. Further, as shown in Figure 4, both anxiety and sadness are low in the comprehension stage, rendering them weak effects that may make it less likely that they will trigger reactions such as retweeting. Accordingly, we propose:

P2.2: *The sharing (retweet) of angry tweet messages is **(a)** higher than the sharing of tweets involving other emotions in the comprehension stage, and **(b)** higher in the comprehension stage than it is in other stages.*

Finally, in interpreting sadness in the projection stage, we find weaker results and nondominance. (1) Sad tweets in the projection stage are retweeted significantly more than they are in the comprehension ($p < 0.01$) and perception ($p < 0.1$) stages. The differences are smaller than seen for other emotions and their dominant stages and, in the case of the perception stage, the significance is only marginal. (2) An interesting finding is that although the retweeting behavior for sadness-laden tweets in the projection stage is significantly higher than for anxiety-laden ones ($p < 0.01$), it is *lower* in comparison to anger-laden tweets ($p < 0.01$).

We attribute the finding that sadness-laden tweets receive more sharing in the projection stage than other stages to the need to *restore normalcy* (Elrod & Tippett, 2002). Smith (2019) suggests that to return to a state of normalcy, it is imperative for one to express sadness following a crisis. The restoration of normalcy is most salient in the projection stage because it happens only after the breach has been dealt with in the perception and comprehension stages. It is also reasonable to expect the restoration of normalcy to involve the prevention of similar crises in the future. In the OPM case, Twitter users were found to retweet sadness-laden tweets about the lack of post-breach improvement in cyberdefense by the OPM. Through retweeting actions, users expressed sadness at what the OPM data breach implied about organizational leadership.

As a result, although we might have expected sadness-laden tweets to receive more sharing than both anxiety and anger in the projection, the data exploration has revealed this is not the case. While anxiety is clearly not a relevant emotion in the projection stage and does not drive people to retweet, anger still dominates sharing behavior. Although anger sharing is on the decline, as shown in Figure 5, we suspect that the sharing of sadness is limited due to its slower emergence as a dominant emotion in the projection stage, as discussed in Phase 1 and Proposition 1.3. As a result, we propose:

P2.3: *In the projection stage, the sharing (retweet) of sad tweet messages is **(a)** higher than in other stages, and **(b)** occurs at an increasing rate in comparison to other emotions.*

Phase 3

In this phase, we explore those breach-related variables that are associated with the dominant emotions in the awareness stages. Roseman commented, "appraisals of events directly influence emotions" (2001, p. 68), indicating that how we perceive an event drives the emotions we express in relation to it. To date, the literature on emotion appraisal models has identified a range of variables that shows a strong connection with emotions, such as unexpectedness, situational state, motivational state, probability, problem type, and anticipated effort (Roseman, 2001; Scherer, 1982; Smith & Ellsworth, 1985). Yet little is known in the data breach literature on what variables may link to the presence of breach emotions (Syed, 2019; Syed & Dhillon, 2015). This phase answers RQ3—What are the key characteristics of the data breach that are associated with the shared negative emotions expressed in the online discussion?

Analysis

Because the data used in this study is a collection of Twitter messages, it lends itself to word vector analysis. To achieve this, we (1) built a word-to-vector (W2V) model, (2) compiled a list of data breach terms, (3) queried the W2V model using these terms to arrive at a tweet-level value for each breach term,[5] and then (4) compared the resulting values to the emotion scores. The following sections describe each of these steps in more detail, followed by a discussion of the results.

[5] We use "word" and "term" interchangeably when referring to the W2V

Word-to-Vector Derived Analysis

Building the word-to-vector model: W2V analysis uses a word embedding approach to convert the linguistic structure and context of words within a corpus into a vector space (Mikolov et al., 2013a). This learning process has the benefit of encoding the linguistic characteristics represented by a language as well as the contextual nuances of a particular corpus (Mikolov et al., 2013b). W2V representations are ideal for understanding not only how the language of a particular corpus is structured but also how any given word used within that corpus is related to other words. For example, W2V modeling can enable the understanding of semantic similarities, concept properties, and phrase similarity (Erk, 2012).

To create the W2V model, we performed cleaning steps that included the removal of punctuation, tokenization of words, lower-casing of words, and the removal of medium specific noninformative tokens (such as retweet indicators and URLs). The eventual model captures both syntactic and semantic word usage. This process was completed using the genism library in Python (Rehurek & Sojka, 2010) and the total vocabulary size of the model was 9,100 words. We used a vector dimensionality of 300 (Mikolov et al., 2013b) since there is evidence that such dimensionalities improve the model representation of subtle semantic relationships (Mikolov et al., 2013a). Prior to using the model, we performed a manual inspection of the model space through several visualizations. One of these visualizations, annotated with our qualitative remarks, can be found in the Appendix in the form of a t-SNE reduction (Figure 8).

Identifying breach concepts: Following W2V model development, we curated a list of breach terms through three different methods. In the first method, we reviewed breach and security-related literature to identify common terms. To accomplish this, we found several relevant studies in the literature using relevant keywords such as "breach," "security," and "social media." Then, we hired a graduate research assistant to extract breach-related concepts from these studies. Concepts that were closely related were combined into clusters, e.g., malware and infiltrate became represented by the concept of security failure reasons. Multiple coauthors reviewed and verified the concepts gathered through this method.

We also extracted breach-related concepts from the congressional report regarding the OPM data breach (Chaffetz et al., 2016). We developed a PDF parsing process that extracted all text content, performed a similar cleaning and tokenization of that content as described in the W2V corpus preparation, and then counted the frequency of terms. Once counted, we reviewed the top occurring terms for inclusion, ignoring common stop-words and other syntactical-glue words with little semantic value. These words were then compared to those obtained from the literature review, with unique words not already present included in the overall list.

Finally, we used an unsupervised data analytic technique to extract topical terms from the overall corpus, so that they might be reviewed for inclusion in the list. To achieve this, we performed a topic-modeling analysis of the total set of tweets using Latent Dirichlet Allocation (LDA) (Blei et al., 2003). The top salient terms from each topic were reviewed for inclusion in the overall list of terms. This completed the compilation of breach-related concepts, resulting in 360 total terms. A sampling of these concepts is reported in Table 9.

Analyzing concepts using W2V: With the breach concept list and the W2V model both completed, we analyzed each tweet in an iterative process. We first cleaned tweets in a manner identical to the cleaning process performed for the W2V. Then each tweet was split into its set of words. Then, for each breach concept, the W2V model was queried for the top 100 words by similarity (e.g., words that the model considered close matches semantically) and their associated similarity score (weight). This was compared to the tweet word set, with the similarity values for all occurring words summed up and divided by the total number of words in the tweet.

The resulting value represented the tweet's content representation for the given breach concept. Because the W2V model is built specifically for this corpus, the similarly occurring words in the corpus are built into the model, resulting in the ability to score concepts as variables even when the keyword itself is not present. A correlation analysis of these scores against the emotion scores previously obtained was done with Spearman's Rho using R's cor.test package.

Results and Proposition Development

Breach concepts and anxiety in the perception stage: From the W2V model, we identify the top 10 concepts that have the highest association with anxiety in the perception stage. These 10 concepts were collapsed into three clusters. Table 10 lists the concepts and their correlations with anxiety in the perception stage, as well as corresponding words that describe each concept, relevance in tweet messages, a summary of each concept, and issues represented.

The first issue, leadership incompetency, concerns the role of top leadership at the national level (such as governments and unions) and administration level, including the intelligence community. It also elaborates on incompetency in managing infrastructure and performance issues dealing with cyberdefense. The second issue, breach framing, involves understanding the initial context of the data breach.

Table 9. Sample Breach-Related Concepts from Literature, Congressional Report, and LDA Topic Model

Source	Sample concepts	Concept clusters
(Chaffetz et al., 2016; Chatterjee et al., 2019) & LDA topic model	Unlawful, access, sensitive, background, clearance	Breach framing
(Chaffetz et al., 2016; Syed & Dhillon, 2015) & LDA topic model	Malware, infiltrate, legacy, compromise	Security failure reasons
(Confente et al., 2019) & LDA topic model	Penalty, complaint, blame	Breach warning
(Chaffetz et al., 2016; Confente et al., 2019)	Restore, customize, investigate, monitor, audit	Recovery
(Chakraborty et al., 2016) & LDA topic model	Casualty, stolen, scam	Victim
(Chaffetz et al., 2016; Syed, 2019)	Human manipulation, system vulnerability, technology failure	Accident

Note: *Chaffetz et al. 2016 is the congressional report

Table 10. Exploration of Anxiety in Perception Stage

Concept and correlation	W2V query sample	Tweet relevance	Summary	Concept clusters
Email (0.49***)	Server, insecure	Email server, Insecure server, cover up	Incompetency of top leadership in storing email data on insecure server	Leadership incompetency
Performance (0.45***)	Mud, server	Cyberdefense, Security, Detection, Protection	Performance issues dealing with failure of cyberdefense	
History (0.37***)	Recruiting, nation, largest	US, UK, Patriot, Government, China (Gov), State	Nation states' and government's responsibility for data hack	
Admin (0.36***)	Surfaces, top	Intelligence Community, Fed employees	Administration delays in contacting employees	
Threat (0.36***)	Vulnerability	Dedicated adversary, Ever evolving threat	U.S. government and people facing adversary and continuous threats	Breach framing
Behavior (0.28***)	Destroying, monumental	Suspicious act, Risky actions, Considering the hack unserious	Breach action threatening American spies and putting them at risk	
Fire (0.27***)	Scalise, place	Hacking federal employees' data, Making key files public	Breach actions including key files becoming fully public	
Scary (0.25***)	Terrifying, worrying	Big scary deal, Terrifying and infuriating, Worry and security problem, SF-86 getting leaked	Breach details are terrifying and pose a matter of worry for affected individuals	
Extortion (0.24***)	Pressurize	Put pressure, Obtain secrets, Convert targets into assets	Hackers may pressure victims for secrets by blackmail/social engineering	Personal impact
People (0.18***)	Impacted, ppl, million, Americans, affected	Affected individuals, Employees, Union People, Contractors, Experts, Hackers	Data of government employees in hands of hackers	

This is characterized by breach-related descriptions such as the vulnerabilities exposed, breach actions including destruction caused, and the terror of the breach. The third issue, personal impact, identifies the consequence of the breach. In this issue, there is a discussion over who are the people affected (Americans, government employees, and millions) and how they are affected (e.g., blackmail and social engineering).

We found higher levels of anxiety in tweets directed towards the top leadership, including political parties/candidates, federal agencies, and government officials. The incompetency of top management is captured in the W2V concepts of "email," "performance," "history," and "admin" (administration). These include tweets concerning email security, tweets about the incompetence of the office, as well as tweets relating to insecurities of the government:

- …our Intelligence Community worries about #OPMhack & how our government got caught #PantsDown
- Sacrificing cyber defense for cyber offense leads to huge vulnerabilities & insecurity in data

As these tweets suggest, when the public grasped more about the situation, with respect to the management practices and the handling of the breach, they contemplated the role of management in the data breach. Feeling that the leadership is incompetent likely drove the concerned public to have less confidence in the data security at OPM. As a result, the public may have become more uncertain about the security posture and management practices of government leaders, thus contributing to an increase in anxiety. In contrast, people may have less uncertainty regarding a breach if they have strong confidence in the leadership, which would result in lower anxiety. Therefore, we propose:

P3.1.1: *Leadership incompetency is positively associated with anxiety in the perception stage*

The analysis shows higher levels of anxiety in tweets searching for details and framing the initial perception of the breach. This framing of the breach is captured in the W2V concepts of "threat," "behavior," "fire," and "scary." There are tweets related to destruction caused by the breach, warnings of threat, and breach attributes:

- … hacking of federal employees' data
- #OPMhack unfortunately the key files must become fully public …
- … Shameful that the federal govt cant keep data secure
- #OPMhack … is terrifying and infuriating in equal measure

Such tweets providing breach framing give an initial indication that a breach event has occurred and hint at the scale of it, but they lack specificity and detail. For example, in the initial announcement (June 4) of the OPM hack, only 28 words out of the 962-word announcement revealed breach details. Thus, while it confirmed that a crisis had occurred, the absence of comprehensive information about the breach likely created uncertainty. This uncertainty, in turn, may have fostered anxiety about the broad, potential impacts of the breach on society, ranging from cyberspace security to homeland security (Berenbaum et al., 2008). Thus, uncertainty is a necessary condition for anxiety of any kind (Dugas et al., 2005). This research in combination with our analytic results leads to the following proposition:

P3.1.2: *Breach framing (in terms of the degree of potential impact) is positively associated with anxiety in the perception stage*

The content analysis employed in this paper has identified two foci underlying the issue of the personal impact of the breach: (1) blackmail/extortion, dealing with how people can be pressured into disclosure, and (2) victimization (direct or indirect), focusing on the severity of the breach for people. Examples of these focuses can be seen in tweets like:

- … useful for applying pressure and turning targets into assets
- all the blackmail vulnerabilities of govt employees uncovered by govt clearance investigations are now in the hands of hackers
- #SF86 carries info on who the intel officer/contractors 'knew well'

The concerns about personal impact reflected in these tweets were captured in the W2V concepts of "extortion" and "people." While the initial breach information presumably triggered the concerned public to grow anxious about the broader impacts, individuals who may have been directly (e.g., victims) or indirectly (e.g., family members or friends) affected also would have developed anxiety since their involvement in the breach and their personal impacts would have not yet been confirmed. While the public may have believed that the lack of implementation of adequate safety measures by organizations led to the victimization of customers (Syed & Dhillon, 2015), they would have not yet had an opportunity to digest the full ramifications of the crisis event, thus providing a catalyst for uncertainty. Such uncertainty makes users anxious (Gentes & Ruscio, 2011; Gudykunst, 2005). Therefore, we propose:

P3.1.3: *Personal impact is positively associated with anxiety in the perception stage*

Breach concepts and anger in comprehension stage**:** Analysis of the W2V concepts and negative emotions expressed in the comprehension stage indicate that anger is expressed in relation to three issues: (1) lack of a cybersecurity culture, (2) lack of leadership transparency and accountability, and 3) poor security controls and configurations (see Table 11).

The analysis revealed that anger was positively and significantly correlated with discussions indicating perceptions that the data breach was the result of an organizational culture that did not take cybersecurity seriously leading up to the breach. Likewise, there was a significant positive correlation between anger and discussions that post-breach cybersecurity had been considered secondary to nonsecurity concerns or that the risk had been minimally considered and misrepresented. Together, these issues indicate that a perceived lack of cybersecurity culture was positively associated with anger, which can be seen in tweets such as:

- I think it'd be an awesome start if the Admin treated #OPMhack as a national security threat instead of a political problem
- DHS testifies #OPMhack is a "medium-to-high" threat on its ""severity scale.' Medium?
- Congratulations to the U.S. government, for proving that "a good offense is the best defense" is total BS. #OPMhack

These discussions were not limited to the context of the OPM organization but extended to the national level as well. Looking at the concepts that correlate with anger, the focal issues are those of the organizational cybersecurity culture and the treatment of the breach as a real security concern. Based on this, we propose:

P3.2.1: *A lack of cybersecurity culture is positively associated with anger in the comprehension stage*

The second issue of discussion that was associated with anger in the comprehension stage is that of leadership roles and accountability. The difference from the previous context is that these discussions were focused on the leader within the organization. Here, words such as "punishment," "responsible," and "disregard" characterize the discussion. Example tweets are:

- #OPMhack federal employee unions will make damn sure no one is held responsible for their incompetence #asUsual
- The storm of firings and accountability that will now sweep DC after #OPMhack will be an amazing - Who the hell am I kidding
- All I want is an apology from a leader re: #OPMHack, whether Archuleta or @POTUS. Why's that too much to ask? Enough blame and ass covering.

These examples demonstrate consistency in the focal issue: the lack of leadership accountability and transparency. In the response to OPM, the public was troubled when the leadership did not offer a clear pathway on how to proceed and what actions to take when they were most needed to clean up the mess. In the case of breaches, accountability should clearly reveal itself, and the responsible leaders should accept the consequences in a fast and firm manner. The White House, for example, stressed they would "hold agency heads accountable for managing cybersecurity risks… improve awareness and transparency of cybersecurity practices" (The White House, 2018). As discussed by Lowenstein and Lerner (2003), there is a relationship between anger and the attending focal issue. The positive significant correlation found between the associated issues and anger demonstrates a consistent theme. Thus, we propose:

P3.2.2: *A perceived lack of leadership transparency and accountability is positively associated with anger in the comprehension stage.*

In the third context of discussion that was associated with anger in the comprehension stage, we find the concepts of poor security controls and configurations. Unlike the previous issues, here the concern is more specific to the inappropriate use of technology within an organization. A focal phenomenon of interest for IS researchers, *information and cybersecurity* involves the appropriate use of security technologies, processes, and procedures to produce a secure technological environment. Example tweets indicate an awareness of these necessary realities for robust cybersecurity:

- Good to know that sensitive files cannot be encrypted due to antiquated technology! #idiots
- Intel agencies knew about vulnerability, so share much of the blame for #OPMHack
- #OPMhack when backdoors=barndoors Sacrificing cyber defense for cyber offense. Leads to huge vulnerabilities & insecurities

When concerned individuals found out that a poor job in implementing security controls and configurations contributed to the data breach, they become furious. Therefore, we propose:

P3.2.3: *Poor security controls and configurations are positively associated with anger in the comprehension stage.*

Table 11. Exploration of Anger in Comprehension Stage

Concept and correlation	W2V query sample	Tweet relevance	Summary	Concept clusters
Thank (0.11***)	Furious, outraged, awfulness, dumbass, dipshit	Thank you and thanks [sarcasm], Furious, Incensed, Appalled, Geez	Exasperation over breach events and security posture of organizations and government	Lack of cybersecurity culture
Sue (0.09***)	Fix, need	OPM hiring cybersecurity professionals, Fix their security issues, Incompetent liars	Organizations should maintain a proactive rather than reactive cybersecurity culture	
Foreign (0.11***)	Spies, concerns, relations, violate	Foreign threat to USA, Act of war, Cyber Pearl Harbor	The nation should take cybersecurity as seriously as it does physical security	
National (0.11***)	Threat	National threat, National security threat, Iraq level failures of cyber, Cyberthreat sharing	The nation should take cybersecurity violations as real threats to national security	
Crisis (0.10***)	Incompetence, defense, American	Huge vulnerabilities, “a good offense is best defense is total BS”, Stupidity of one horrible department	How the nation handles cybersecurity has serious implications for national defense	
Proposed (0.11***)	Question, disregard, punish	Where is MSM on these scandals, How hard is it to answer a yes or no question, [sarcasm] “No-ones fault!”, And still no one has been fired	Leaders should be held responsible for security failures	Lack of leadership transparency and accountability
United States (0.10***)	Outrage, distrust	Screws of corruption in federal government, Breach caused employees to distrust government	The manner in which leaders handle the response to data breaches impacts how competent those leaders are perceived	
Admin (0.09***)	Obama, incompetence, regime	No one in Obama admin responsible, Why can’t we fire OPM director,	Leaders emphasizing a perception management approach incite anger and outrage rather than reassurance	
Web (0.09***)	Targeting, scale	FBI discloses malware related to OPM hack, China targeting U.S.A., Hack dates back to December	Security controls and processes should accurately reflect risks	Poor security controls and configurations
Harm (0.16***)	System, antiquated, network, vulnerabilities	Network security, New vulnerabilities, Privileged access, U.S. system	Security environment and configuration should be state-of-the-art	

Table 12. Exploration of Sadness in the Projection Stage

Concept and correlation	W2V query sample	Tweet relevance	Summary	Concept clusters
Open (0.20***)	Resists, mismanagement	President needs fire her, chief shrugs off calls for her resignation	Dissatisfaction about the lack of action	Lack of agency and control in data breach response
Answer (0.16***)	Delayed	Failure to address these attacks results	Actions are overdue	
Justice (0.15***)	Shouldn't, responsibility	Should have been fired, allowing her to resign is wrong	Public opinion was ignored in breach response	
Archuleta (0.49***)	Katherine, resignation, repair	Director Resigns, what's worse than losing data? Losing trust in it	Demise of security guardianship, loss of trust	Decrease of citizen trust in secure digital operations
Director (0.41***)	Resigns, goodbye	OPM Dir Archuleta resigns	Demise of security guardianship	
Resign (0.40***)	Removal, director	Katherine Archuleta resigning today	Demise of security guardianship	
Resignation (0.38***)	Removal, director, destroys	Katherine Archuleta resigns, earning back trust of American public	Demise of security guardianship, loss of trust	
Chief (0.34***)	Resigns	Director Resigns in wake of epic #OPMHack	Demise of guardianship	
Survive (0.37***)	Committee, senate, lawmakers	Resignation doesn't address prevention of more data breaches, admin intends this to be "end of story," it'll take more than a resignation to move beyond the challenges, this is a symbolic move	Lack of improvement in cybersecurity strategies and practices	Lack of security governance
House (0.21***)	Appointment, legislate, fixes, leadership	USOPM was a political appointment, Do you trust #OPM to fix this with the same CIO? failure of military and Federal civilian leadership to anticipate threats	Lack of change in personnel management, lack of improvement in strategic planning	

Breach concepts and sadness in the projection stage: In Table 12, we summarize the findings of sadness in the projection stage. The analysis suggests three emerging issues associated with sadness: (1) lack of agency or control, (2) decrease in trust, and (3) lack of security governance.

In distinguishing differences in why some people react to the same event with sadness or anger, Levine notes that perceptions of agency and control differentiate those that are angry from those that are sad (1996). In his study, Levine found that even ordinary sadness is associated with the rumination and recall of prior hopes and expectations (1996). This could serve as an explanation not only for some of the potentially surprising topics associated with sadness in this stage (e.g., Archuleta being allowed to resign rather than being fired), but also for the mixed results seen in the previous phases regarding the differences in anger and sadness in the projection stage. As seen in Levine's work (1996), sadness and anger can both stem from the same event because of differences in how individuals interpret their level of agency and control, as well as their ability to reinstate prior beliefs and desires. However, the further one gets from a focal event that evokes these emotions, the more likely that sadness will emerge, due to the mounting experiential evidence of lack of control and agency. Beginning with the initial announcement of the breach in June, there was public demand that OPM Director Katherine Archuleta be fired. Yet after a prolonged period of ignoring these cries, the administration allowed Archuleta to resign rather than be fired. Thus, while earlier calls for Archuleta to be fired were expressions of anger, after the event passed, a realization emerged that this goal could not be reached. Such realizations have been shown to be associated with sadness (Levine, 1995; Smith & Lazarus, 1993; Stein & Levine, 1989).

In the OPM data breach, the lack of agency is captured in the W2V concepts of "justice," "answer," and "open." It is reflected in tweets that repeatedly discussed the slow and overdue resignation of Director Archuleta. Examples include:

- President needs to fire her, chief shrugs off call for her resignation
- US personnel chief resists calls to resign amid data hack
- #KatherineArchuleta shouldn't have resigned. She should've been fired

Therefore, we propose:

P3.3.1: *Lack of agency in a data breach response is positively associated with sadness in the projection stage*

The second issue focuses on trust. Citizens bestow trust in organizations when they hand over personal data. Given constant threats from hackers, individuals have become hesitant to surrender sensitive information. To this end, trust acts as a psychological assurance for citizens against likely threats (Jarvenpaa et al., 2004). Without this trusting belief, organizations will have difficulty in attracting the public to use their digitally enabled services. According to Gus Hunt of Accenture (2019), citizens must be able to trust that the government has ensured secure and safe digital operations: "today's federal IT modernization investments and efforts will be undermined by an erosion of public trust" (p. 4).

The public becomes sad when the perceived trustworthiness of the federal government dwindles. They are sad because the government appears unmatched to the skills of the hackers, leaving civilian data vulnerable to future attacks. This sadness is compounded with the realization that the government is unwilling to adequately hold those within the organization responsible for leadership failures, such as allowing Archuleta to resign rather than be fired following months of public outcry that was largely ignored. The decrease of citizen trust in secure government digital operations is captured in the W2V concepts of "Archuleta," "director," "resign," "resignation," and "chief." There is a wide spread of messages that cover this significant moment by speaking directly to the resignation and the erosion of trust:

- #BreakingNews: Director Resigns over #OPMhack
- What's worse than losing data? Losing trust in it
- #OPM has hard task of fixing failures & earning back trust of American public

Hence, we propose that:

P3.3.2: *Decrease of citizen trust in secure digital operations is positively associated with sadness in the projection stage*

Cybersecurity governance is the key to secured information assets and digital services. Security governance provides the solutions to organizations that are in dire need of coping with the challenges created by "the increasingly interconnected, information-intensive business landscape, legal pressures, and ongoing scrutiny to transparency and overall governance" (Korhonen et al., 2012, p. 1). It requires an organization to develop strategic directions, clear objectives, and high priorities to build a sound security system (Conner et al., 2004).

The public feels sad when an organization whose security was breached does not engage in corrective actions to build sound cybersecurity governance. The lack of security governance enhancement is captured in the W2V concepts of "survive" and "house." Tweets expressing this include:

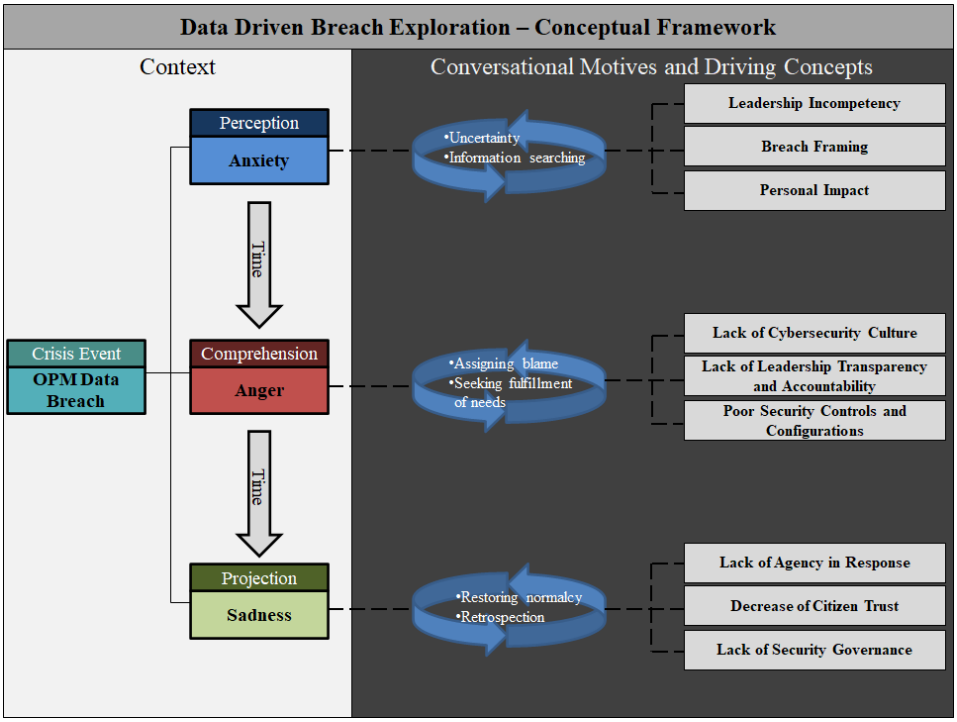
- Personnel chief's resignation doesn't address prevention of more data breaches
- Oddly enough I'm even more upset re #OPMHack after Archuleta resignation. Why? B/c Obama admin intends this to be "end of story." It's not.
- Failure of military and Federal civilian leadership to anticipate threats; adequately prepare

Therefore, we propose that:

P3.3.3: *The lack of enhanced security governance is positively associated with sadness in the projection stage*

Discussion

We analyzed organic tweets from June 4 to July 30, 2015, in a longitudinal research design over three time periods. We present a three-phased methodology: Phase 1 identified changes in the negative emotions of anxiety, anger, and sadness. Analysis from this phase revealed that anxiety is dominant in the perception stage, anger is dominant in the comprehension stage, and although sadness is not able to be described as dominant in the projection stage, it is an emerging emotion. Phase 2 examined the effect of emotions on social media sharing (retweets) and revealed that sharing of dominant emotion tweets is higher than sharing on nondominant emotion tweets. Phase 3 explored breach-related characteristics that are associated with the dominant emotion



Note: We acknowledge the invaluable guidance of the AE in suggesting this figure.

Figure 6. Conceptual Framework

This paper conceptualizes a breach as a change event, which results in a negative emotional response. This study makes several contributions. First, it explores large-scale data from Twitter in the aftermath of a breach event and reveals the negative emotions that dominate over the course of the crisis event. In addition, the study examines the characteristics of negative emotions, and confirms the changes in emotion dominance according to SA stage.

Second, it explores the breach characteristics within online conversations. It provides explanations for the association of breach concepts and post-breach sharing behavior through the perspective of emotions. It analyzes an under-addressed area of breach research—"breach emotions." By utilizing emotions for introspecting the question of "why" surrounding breach-related reactions on social media, this paper unpacks the "black-box models with simpler interpretable models" (Rai, 2020, p. 138).

Third, to understand breach-specific concepts strongly associated with emotion, we used a comprehensive data-driven approach. This included the triangulation between the informative sources of breach literature, social media data, and the congressional report to identify breach terms, and combined a qualitative review with data-driven analytic techniques such as unsupervised topic modeling (LDA) and term saliency scoring. Finally, the use of a W2V model to measure tweet-level relevancy for identified breach concepts allowed for a nuanced understanding of how these concepts are related to emotions. These data-driven approaches allowed this paper to bridge social media data and breach literature, providing insights into the role emotion plays in how we communicate and what we communicate about regarding breaches.

This is among the earliest studies to explore the emotional characteristics of responses to crisis events such as data breaches. Using a variety of analytic techniques, we identify an alignment between situation awareness stages and dominant emotions expressed in response to disruptive change and the conversational motives and concepts that result. These results indicate that an understanding of the state of *situational awareness* regarding an event, such as a data breach, can allow one to anticipate the likely emotions that will be expressed by those in that situation. This is of tremendous value for crisis management and data breach management, as it can inform management strategies and tactics. These results also identify conversational motives associated with these emotions, which serve as drivers of discourse among observers and stakeholders. Understanding these motives can inform effective communication strategies. Finally, these results demonstrate the value of data exploration in identifying the salient concepts associated with dominant emotions. These techniques could be used to identify analogous concepts that exist in other crisis situations. Figure 6 illustrates a

preliminary framework that we hope future studies might consider when evaluating crisis events, particularly data breaches and events that impact large groups of people.

The OPM breach resulted in shared emotions that were expressed by users on social media platforms. The shared emotions resulted in a communal social experience. Government relies on such shared emotional expressions to develop and implement public policy (Anderson et al., 2018). Our analysis can enable decision makers to devise productive strategies, avoid counterproductive strategies, to mitigate such negative emotions (Park et al., 2016). Organizational leadership should seek to understand perceptions of the organization as well as evaluations of its products and services (as is done today in many firms under the aegis of chief listening officers[6]). Such consideration will help foster a human-centered society and will enable convergence between cyberspace and physical space.

Prior literature has shown that a negative perspective of a change may result in its poor management (DiFonzo et al., 1994; Smeltzer & Zener, 1992). By understanding emotions resulting from a crisis, organizations and policy makers can develop best practices to handle crisis situations (Beaudry & Pinsonneault, 2010). For example, Caredda has identified three types of support in crisis situations, "Information and Communication, Emotional Support, and Guidance and Direction — all critical in the way that change can impact" (2020, p. 2). Our findings add to the existing literature by revealing data breach management insights.

Conclusion

IS literature has observed an increasing number of research studies on data breaches (Garrison & Ncube, 2011; Holtfreter & Harrington, 2015); Yet they do not investigate the emotional aspect of breach response. IS studies that do explore emotions have shed little light on their involvement in data breaches (Chin et al., 2003; Tsai & Bagozzi, 2014; Venkatesh, 2000; Yin et al., 2014). Further, while prior literature has examined the sharing of breach-related information, it is mostly limited to the mining of social media data for addressing what breach information is being shared: there is little theory-building introspection into how breach details shape people's reactions on social media (Dwivedi et al., 2020). This paper takes a data-driven approach to extracting key associations among data breach characteristics and emotions and between emotions and sharing behavior and develops propositions that can be used to guide future research (Kanavos et al., 2014).

Future studies could also explore privacy breaches, where data is taken with authorization but then used without permission. An example to this point is the Cambridge Analytica scandal, in which the collected data of Facebook users was later exploited for political analysis without users' awareness and consent. It would be interesting to explore the emotional responses associated with this event. Future studies could also explore post-breach intervention strategies, examining, for example, whether organizations can shorten the crisis stages through comprehension and pre-empting strategies or by anticipating projection (through compensation).

Acknowledgments

The authors wish to thank the SE, the AE, and the review team for their sustained guidance and critical comments that greatly improved the paper. This research was funded by the National Science Foundation under grants #1554480 and #1651060. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. An early version of this paper was selected as the best paper at AMCIS 2017.

References

Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. In *Proceedings of the International Conference on Information Systems*.

Adams, M. (2016, March 11). *Why the OPM Hack Is Far Worse Than You Imagine*. Lawfare. https://www.lawfareblog.com/why-opm-hack-far-worse-you-imagine

Aliyu, F., Wimmer, H., Powell, L., & Rebman, C. (2020). An international review and study on perceptions of security, adoption, and implementation of electronic health records. In *2020 Proceedings of the Conference on Information Systems Applied Research*.

Anderson, M., Toor, S., Rainie, L., & Smith, A. (2018). *Activism in the social media age*. Pew Research Center. https://www.pewinternet.org/2018/07/11/activism-in-the-social-media-age/

Angst, C. M., Block, E. S., D'Arcy, J., & Kelley, K. (2017). When do IT security investments matter? Accounting for the influence of institutional factors in the context of healthcare data breaches. *MIS Quarterly*, *41*(3), 893-916.

Austin, L., Fisher Liu, B., & Jin, Y. (2012). How audiences seek out crisis information: Exploring the social-mediated crisis communication model. *Journal of Applied Communication Research*, *40*(2), 188-207.

Bang, C., Lee, J., & Rao, H. R. (2021). The Egyptian protest movement in the Twitter Sphere: An investigation of dual

[6] https://www.socialmediatoday.com/content/meet-chief-listening-officer

sentiment pathways of communication. *International Journal of Information Management, 58,* Article 102328.

Bantum, E. O., & Owen, J. E. (2009). Evaluating the validity of computerized content analysis programs for identification of emotional expression in cancer narratives. *Psychological Assessment, 21*(1), 79-88.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323-370.

Beaudry, A., & Pinsonneault, A. (2010). The Other Side of Acceptance: Studying the Direct and Indirect Effects of Emotions on Information Technology Use. *MIS Quarterly*, *34*(4), 689-710.

Belyh, A. (2020, July 28). *Understanding the Kubler-Ross change curve*. Cleverism. https://www.cleverism.com/understanding-kubler-ross-change-curve/

Berenbaum, H., Bredemeier, K., & Thompson, R. J. (2008). Intolerance of uncertainty: Exploring its dimensionality and associations with need for cognitive closure, psychopathology, and personality. *Journal of Anxiety Disorders, 22*(1), 117-125.

Berente, N., Seidel, S., & Safadi, H. (2019). Research Commentary—Data-Driven Computationally Intensive Theory Development. *Information Systems Research*, *30*(1), 50-64.

Berger, J. (2011). Arousal Increases Social Transmission of Information. *Psychological Science*, *22*(7), 891-893.

Berger, J., & Milkman, K. L. (2012). What makes online content viral? *Journal of Marketing Research*, *49*(2), 192-205.

Berinato, S. (2020). That discomfort you're feeling is grief. *Harvard Business Review*. https://hbr.org/2020/03/that-discomfort-youre-feeling-is-grief

Bisson, D. (2015). *The OPM breach: Timeline of a hack*. Tripwire. http://www.tripwire.com/state-of-security/security-data-protection/cyber-security/the-opm-breach-timeline-of-a-hack/

Biever, C. (2010). *Twitter mood maps reveal emotional states of America*. New Scientist. https://www.newscientist.com/article/dn19200-twitter-mood-maps-reveal-emotional-states-of-america/

Bird, S., Klein, E., & Loper, E. (2009). *Natural language processing with Python: Analyzing text with the natural language toolkit*. O'Reilly Media.

Blei, D. M., Griffiths, T. L., Jordan, M. I., & Tenenbaum, J. B. (2004). Hierarchical topic models and the nested Chinese restaurant process. *Proceedings of the Advances in Neural Information Processing Systems 16* (pp. 17-33).

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, *3*, 993-1022.

Brehm, S. S., & Brehm, J. W. (2013). *Psychological reactance: A theory of freedom and control*. Academic Press.

Brown, C. R., Greitzer, F. L., & Watkins, A. (2013). Toward the development of a psycholinguistic-based measure of insider threat risk focusing on core word categories used in social media. *Proceedings of the Nineteenth Americas Conference on Information Systems*.

Brown, H. S. (2016). After the data breach: Managing the crisis and mitigating the impact. *Journal of Business Continuity & Emergency Planning*, *9*(4), 317.

Caredda, S. (2020, April 14). *Models: Kübler-Ross change curve*. Sergio Caredda. https://sergiocaredda.eu/organisation/tools/models-kubler-ross-change-curve/

Chaffetz, J., Meadows, M., & Hurd, W. (2016). *Committee releases year-long investigative report into OPM data breaches*. https://oversight.house.gov/release/committee-releases-year-long-investigative-report-opm-data-breaches/

Chakraborty, R., Lee, J., Bagchi-Sen, S., Upadhyaya, S., & Raghav Rao, H. (2016). Online shopping intention in the context of data breach in online retail stores: An examination of older and younger adults. *Decision Support Systems*, *83*, 47-56.

Chatterjee, S., Gao, X., Sarkar, S., & Uzmanoglu, C. (2019). Reacting to the scope of a data breach: The differential role of fear and anger. *Journal of Business Research*, *101*, 183-193.

Cherry, K. (2019). *The 6 types of basic emotions and their effect on human behavior*. Verywell Mind. https://www.verywellmind.com/an-overview-of-the-types-of-emotions-4163976

Chin, W., Marcolin, B. L., & Newsted, P. R. (2003). A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic mail emotion/adoption study. *Information Systems Research*, *14*(2), 189-217.

Chmiel, A., Sienkiewicz, J., Thelwall, M., Paltoglou, G., Buckley, K., Kappas, A., & Hołyst, J. A. (2011). Collective emotions online and their influence on community life. *PLoS ONE*, *6*(7), Article e22207.

Christophe, V., Delelis, G., Antoine, P., & Nandrino, J.-L. (2008). Motives for secondary social sharing of emotions. *Psychological Reports*, *103*(1), 11-22.

Confente, I., Siciliano, G. G., Gaudenzi, B., & Eickhoff, M. (2019). Effects of data breaches from user-generated content: A corporate reputation analysis. *European Management Journal*, *37*(4), 492-504.

Conner, F. W., Coviello, A. W., Sullivan, M., Hantman, H., Glynn, M., Luker, M., Lainhart, J. W., & Cullinane, D. (2004). *Information Security Governance: A call to action*. National Cyber Security Summit Task Force.

Culnan, M. J., & Williams, C. C. (2009). How ethics can enhance organizational privacy: Lessons from the Choicepoint and TJX data breaches. *MIS Quarterly*, *33*(4), 673-687.

Curci, A., & Bellelli, G. (2004). Cognitive and social consequences of exposure to emotional narratives: Two studies on secondary social sharing of emotions. *Cognition and Emotion*, *18*(7), 881-900.

Cushman, D. P., & Kovacic, B. (1995). *Watershed research traditions in human communication theory*. SUNY Press.

D'Arcy, J., Adjerid, I., Angst, C. M., & Glavas, A. (2020). Too good to be true: Firm social performance and the risk of data breach. *Information Systems Research*, *31*(4), 1200-1223.

Deng, L., & Poole, M. S. (2010). Affect in web interfaces: A study of the impacts of web page visual complexity and order. *MIS Quarterly*, *34*(4), 711-730.

DiFonzo, N., Bordia, P., & Rosnow, R. L. (1994). Reining in rumors. *Organizational Dynamics*, *23*(1), 47-62.

Dugas, M. J., Hedayati, M., Karavidas, A., Buhr, K., Francis, K., & Phillips, N. A. (2005). Intolerance of uncertainty and information processing: evidence of biased recall and interpretations. *Cognitive Therapy and Research*, *29*(1), 57-70.

Dwivedi, Y., Kar, A., Angelopoulos, S., & Rao, H. R. (Eds.). (2020). Call for papers: Theory building in is with big data-driven research. *International Journal of Information Management*. https://www.journals.elsevier.com/international-

journal-of-information-management/call-for-papers/big-data-theory-building

Elrod, P. D., & Tippett, D. D. (2002). The "death valley" of change. *Journal of Organizational Change Management*, *15*(3), 273-291.

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. *Human factors: The journal of the human factors and ergonomics society*, *37*(1), 32-64.

Erk, K. (2012). Vector space models of word meaning and phrase meaning: A Survey. *Language and Linguistics Compass*, *6*(10), 635-653.

Games, P. A., & Howell, J. F. (1976). Pairwise multiple comparison procedures with unequal n's and/or variances: A Monte Carlo study. *Journal of Educational Statistics*, *1*(2), 113-125.

Garrison, C., & Ncube, M. (2011). A longitudinal analysis of data breaches. *Information Management & Computer Security*, *19*(4), 216-230.

Gentes, E. L., & Ruscio, A. M. (2011). A meta-analysis of the relation of intolerance of uncertainty to symptoms of generalized anxiety disorder, major depressive disorder, and obsessive-compulsive disorder. *Clinical Psychology Review*, *31*(6), 923-933.

German, P. (2016). A new month, a new data breach. *Network Security*, *2016*(3), 18-20.

Goldsworthy, K. K. (2005). Grief and loss theory in social work practice: All changes involve loss, just as all losses require change. *Australian Social Work*, *58*(2), 167-178.

Goode, S., Hoehle, H., Venkatesh, V., & Brown, S. A. (2017). User compensation as a data breach recovery action: an investigation of the Sony PlayStation network breach. *MIS Quarterly*, *41*(3), 703-727.

Goodman, E., & Loh, L. (2011). Organizational change: A critical challenge for team effectiveness. *Business Information Review*, 28(4), 242-250.

Govinfo (2017). *About congressional hearings.* https://www.gpo.gov/help/about_congressional_hearings.htm

Grégoire, Y., & Fisher, R. J. (2008). Customer betrayal and retaliation: When your best customers become your worst enemies. *Journal of the Academy of Marketing Science*, *36*(2), 247-261.

Grégoire, Y., Tripp, T. M., & Legoux, R. (2009). When customer love turns into lasting hate: The effects of relationship strength and time on customer revenge and avoidance. *Journal of Marketing*, *73*(6), 18-32.

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and Anticipation in Anxiety. *Nature Reviews. Neuroscience*, *14*(7), 488-501.

Gudykunst, W. B. (2005). An anxiety/uncertainty management (AUM) theory of effective communication: Making the mesh of the net finer. In *Theorizing about intercultural communication* (pp. 281-322). SAGE.

Guynn, J. (2020). Anxiety, depression and PTSD: The hidden epidemic of data breaches and cyber crimes. *USA TODAY*. https://www.usatoday.com/story/tech/conferences/2020/02/21/data-breach-tips-mental-health-toll-depression-anxiety/4763823002/

Hao, J., & Dai, H. (2016). Social media content and sentiment analysis on consumer security breaches. *Journal of Financial Crime*, *23*(4), 855-869.

Hidalgo, C., Tan, E. S. H., & Verlegh, P. W. J. (2015). The social sharing of emotion (SSE) in online social networks: A case study in Live Journal. *Computers in Human Behavior*, *52*, 364-372.

Hofmann, T. (2013). *Probabilistic latent semantic analysis.* Available at http://arxiv.org/abs/1301.6705

Holtfreter, R. E., & Harrington, A. (2015). Data breach trends in the United States. *Journal of Financial Crime*, *22*(2), 242-260.

Hunt, G. (2019). *Cyber resilience: Building citizen trust.* Accenture. https://www.accenture.com/_acnmedia/PDF-98/Accenture-Cyber-Resilience-Building-Citizen-Trust.pdf#zoom=50

Izard, C. E., Kagan, J., & Zajonc, R. B. (1984). *Emotions, cognition, and behavior*. Cambridge University Press.

Jalonen, H. (2014). Social media and emotions in organisational knowledge creation. In *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems* (pp. 1371-1379).

James, J. W., Friedman, R., & Keeler, B. (1998). *The grief recovery handbook: The action program for moving beyond death, divorce, and other losses*. HarperPerennial.

Jarvenpaa, S. L., Shaw, T. R., & Staples, D. S. (2004). Toward contextualized theories of trust: The role of trust in global virtual teams. *Information Systems Research*, *15*(3), 250-267.

Kanavos, A., Perikos, I., Vikatos, P., Hatzilygeroudis, I., Makris, C., & Tsakalidis, A. (2014). Modeling retweet diffusion using emotional content. In *Proceedings of the IFIP International Conference on Artificial Intelligence Applications and Innovations* (pp. 101-110).

Kherwa, P., & Bansal, P. (2019). Topic modeling: A comprehensive review. *EAI Endorsed Transactions on Scalable Information Systems*, *7*(24), e2.

Korhonen, J., Hiekkanen, K., & Mykkänen, J. (2012). Information Security Governance. In M. Gupta, J. Walp, & R. Sharman (Eds.), *Strategic and Practical Approaches for Information Security Governance: Technologies and Applied Solutions* (pp. 53-66). IGI Global.

Lacan, J. (1977). Aggressivity in psychoanalysis. In *Écrits: A selection* (trans. A Sheridan, pp. 8-29). Norton.

Layton, R., & Watters, P. A. (2014). A methodology for estimating the tangible cost of data breaches. *Journal of Information Security and Applications*, *19*(6), 321-330.

Lee, J., Rehman, B. A., Agrawal, M., & Rao, H. R. (2015). Sentiment analysis of Twitter users over time: The case of the Boston bombing tragedy. In *Proceedings of the Workshop on E-business*

Levine, L. J. (1995). Young children's understanding of the causes of anger and sadness. *Child Development*, *66*(3), 697-709.

Levine, L. J. (1996). The anatomy of disappointment: A naturalistic test of appraisal models of sadness, anger, and hope. *Cognition and Emotion*, *10*(4), 337-360.

Liu, B. (2010). Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing* (2nd ed., pp. 627-666). Chapman and Hall/CRC.

Liu, C.-W., Huang, P., & Lucas, H. C. (2020). Centralized IT decision making and cybersecurity breaches: Evidence from U.S. higher education institutions. *Journal of Management Information Systems*, *37*(3), 758-787.

Lowenstein, G., & Lerner, J. (2003). The role of affect in decision making. In R. J. Davidson, K. R. Scherer, & H. H. Goldsmith

(Eds.), *Handbook of affective science* (pp. 619-642). Oxford University Press.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013a). Efficient estimation of word representations in vector space. Available at http://arxiv.org/abs/1301.3781.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. Available at http://arxiv.org/abs/1310.4546.

Mohammad, S. M., & Turney, P. D. (2013). *NRC emotion lexicon.* National Research Council, Canada.

Novak, A. N., & Vilceanu, M. O. (2019). "The internet is not pleased": Twitter and the 2017 Equifax data breach. *The Communication Review*, *22*(3), 196-221.

Oh, O., Agrawal, M., & Rao, H. R. (2011). Information control and terrorism: Tracking the Mumbai terrorist attack through Twitter. *Information Systems Frontiers*, *13*(1), 33-43.

Oh, O., Gupta, P., Agrawal, M., & Raghav Rao, H. (2018). ICT mediated rumor beliefs and resulting user actions during a community crisis. *Government Information Quarterly*, *35*(2), 243-258.

Park, E. H., Ramesh, B., & Cao, L. (2016). Emotion in IT investment decision making with a real options perspective: The intertwining of cognition and regret. *Journal of Management Information Systems*, *33*(3), 652-683.

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Available at https://repositories.lib.utexas.edu/handle/2152/31333

Pratt, K. (2014). *Psychology tools: What is anger? A secondary emotion.* HealthyPsych.Com. https://healthypsych.com/psychology-tools-what-is-anger-a-secondary-emotion/

Rai, A. (2020). Explainable AI: From black box to glass box. *Journal of the Academy of Marketing Science*, *48*(1), 137-141.

Rao, H. R., Vemprala, N., Akello, P., & Valecha, R. (2020). Retweets of officials' alarming vs reassuring messages during the COVID-19 pandemic: Implications for crisis management. *International Journal of Information Management*, *55*, Paper 102187.

Rehurek, R., & Sojka, P. (2010). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks*.

Rimé, B. (2009). Emotion elicits the social sharing of emotion: Theory and empirical review. *Emotion Review*, *1*(1), 60-85.

Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the space of topic coherence measures. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* (pp. 399-408).

Roeckelein, J. E. (2006). *Elsevier's dictionary of psychological theories*. Elsevier.

Roseman, I. J. (2001). A model of appraisal in the emotion system: Integrating theory, research, and applications. In K. R. Scherer, A. Schorr, T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research.* (pp. 68-91). Oxford University Press.

Sarkar, S., Vance, A., Ramesh, B., Demestihas, M., & Wu, D. T. (2020). The influence of professional subculture on information security policy violations: A field study in a healthcare context. *Information Systems Research*, *31*(4), 1240-1259.

Scherer, K. R. (1982). Emotion as process: Function, origin and regulation. *Social Science Information*, *21*, 555-570.

Sell, A., Tooby, J., & Cosmides, L. (2009). Formidability and the logic of human anger. In *Proceedings of the National Academy of Sciences*, *106*(35), 15073-15078.

Shoolin, J. S. (2010). Change management—Recommendations for successful electronic medical records implementation. *Applied Clinical Informatics*, *1*(3), 286-292.

Sievert, C., & Shirley, K. (2014). LDAvis: A method for visualizing and interpreting topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, (pp. 63-70).

Smeltzer, L. R., & Zener, M. F. (1992). Development of a model for announcing major layoffs. *Group & Organization Management*, *17*(4), 446-472.

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813-838.

Smith, C. A., & Lazarus, R. S. (1993). Appraisal components, core relational themes, and the emotions. *Cognition and Emotion*, *7*(3-4), 233-269.

Smith, K. (2019). *Why you should express your sadness*. PsychCentral. http://psychcentral.com/blog/why-you-should-express-your-sadness/

Stein, M.-K., Newell, S., Wagner, E. L., & Galliers, R. D. (2015). Coping with information technology: mixed emotions, vacillation, and nonconforming use patterns. *MIS Quarterly*, *39*(2), 367-392.

Stein, N. L., & Levine, L. J. (1989). The causal organisation of emotional knowledge: A developmental study. *Cognition and Emotion*, *3*(4), 343-378.

Sternstein, A., & Moore, J. (2015). *Timeline: What we know about the OPM breach (UPDATED)*. Nextgov. http://www.nextgov.com/security/2015/06/timeline-what-we-know-about-opm-breach/115603/

Stieglitz, S., & Krüger, N. (2011). Analysis of sentiments in corporate Twitter communication: A case study on an issue of Toyota. *Proceedings of the 22nd Australasian Conference on Information Systems*.

Stürmer, S., & Simon, B. (2009). Pathways to collective protest: calculation, identification, or emotion? A critical analysis of the role of group-based anger in social movement participation. *Journal of Social Issues*, *65*(4), 681-705.

Su, L. (2014). *Understanding psychological contract breach in the customer-firm relationship* [PhD diss., Iowa State University], Digital Repository. https://doi.org/10.31274/etd-180810-265

Syed, R. (2019). Enterprise reputation threats on social media: A case of data breach framing. *The Journal of Strategic Information Systems*, *28*(3), 257-274.

Syed, R., & Dhillon, G. (2015). Dynamics of data breaches in online social networks: Understanding threats to organizational information security reputation. In *Proceedings of the 36th International Conference on Information Systems*.

Syed, R., Rahafrooz, M., & Keisler, J. M. (2018). What it takes to get retweeted: An analysis of software vulnerability messages. *Computers in Human Behavior*, *80*, 207-215.

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, *29*(1), 24-54.

The White House. (2018). *National cyber strategy of the United States of America*. https://www.whitehouse.gov/wp-content/uploads/2018/09/National-Cyber-Strategy.pdf

Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology*, *81*(6), 973.

Tsai, H. T., & Bagozzi, R. P. (2014). Contribution behavior in virtual communities: Cognitive, emotional, and social influences. *MIS Quarterly*, *38*(1), 143-163.

Valentine, B. (2017). *Creating a culture of security through change management.* Security Intelligence. https://securityintelligence.com/creating-a-culture-of-security-through-change-management/

van der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579-2605.

van Warmerdam, G. (2016). *Understanding anger: Pathway to happiness.* https://www.pathwaytohappiness.com/anger/understanding-anger.htm

Venkatesan, S., Valecha, R., Yaraghi, N., Oh, O., & Rao, H. R. (Forthcoming). Influence in social media: An investigation of tweets spanning the 2011 Egyptian social movement. *MIS Quarterly*, *45*(4), 1679-1714.

Venkatesh, V. (2000). Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. *Information Systems Research*, *11*(4), 342-265.

Verma, S., Vieweg, S., Corvey, W. J., Palen, L., Martin, J. H., Palmer, M., Schram, A., & Anderson, K. M. (2011). Natural language processing to the rescue? Extracting "situational awareness" tweets during mass emergency. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media*. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/viewPaper/2834

Verizon. (2017). *2017 Data breach investigations report (No. 10)*. http://www.verizonenterprise.com/verizon-insights-lab/dbir/2017/

Wang, J., Xiao, N., & Rao, H. R. (2010). Drivers of information security search behavior: An investigation of network attacks and vulnerability disclosures. *ACM Transactions on Management Information Systems*, *1*(1), Article 3.

Wang, J., Xiao, N., & Rao, H. R. (2012). An exploration of risk information search via a search engine: Queries and clicks in healthcare and information security. *Decision Support Systems*, *52*(2), 395-405.

Xu, H., Zhang, N., & Zhou, L. (2020). Validity concerns in research using organic data. *Journal of Management*, *46*(7), 1257-1274.

Yin, D., Bond, S. D., & Zhang, H. (2014). Anxious or angry? Effects of discrete emotions on the perceived helpfulness of online reviews. *MIS Quarterly*, *38*(2), 539-560.

About the Authors

Eric Bachura is an assistant professor in the Department of Department of Information Systems and Cyber Security, Alvarez College of Business, University of Texas at San Antonio.

Rohit Valecha is an associate professor in the Department of Information Systems and Cyber Security, Alvarez College of Business, University of Texas at San Antonio.

Rui Chen is an associate professor in the Department of Information Systems and Business Analytics at Ivy College of Business, Iowa State University.

H. Raghav Rao is the AT&T Chair Professor in the Department of Information Systems and Cyber Security, Alvarez College of Business, University of Texas at San Antonio.

Appendix

Data Analysis Process

The data analysis used in this paper can be understood from an overall pipeline process that involved the steps of (1) data collection, (2) feature engineering, and (3) analysis. The details of these steps are described in the following subsections.

Data Collection

In this study, the Twitter data was the result of a purchase from a third-party vendor. The flat file data provided by the vendor was converted into a SQL database for the purposes of facilitating simpler data mapping and backup processes. This Twitter data included user profile information, tweet metadata, and tweet content. The data purchased was the same as data that can be obtained through dynamic webscraping of the Twitter website or through the API (at the time of this writing). The choice to purchase the dataset rather than conducting our own data collection effort was due to a desire to expedite the study and lack of tangible benefits to wait on API and collection limits to collect the same data ourselves. The Twitter data served as the focal dataset for the study.

Ancillary data used in this study includes a full copy of the lengthy congressional report and a compiled list of data breach research articles. These were collected so that we might extract both context-specific breach terms (congressional report) and general breach terms (research articles). The congressional report also served as a resource for identifying key events that occurred over the course of the data breach event (which were cross-verified by examining related news articles).

Feature Engineering

A variety of feature engineering techniques were applied to the data used in this study. Most of these techniques were applied to the focal data (Twitter data).

Sentiment scoring (and other psychometric features): Originally, multiple sentiment scoring techniques were applied to the dataset, including NLTK sentiment scoring and LIWC sentiment scoring. During the review process, we included NRC (Mohammad & Turney, 2013) sentiment scoring. We decided to use the LIWC approach. This decision was maintained when evaluating the NRC scores. All LIWC features (2015 dictionary) were mapped into the SQL dataset with scores conducted at the tweet level (not all features were used in this study).

Breach concept scoring: As part of the process for developing a list of breach concepts, we created an LDA topic model from the entire set of tweet text. Because of the unsupervised nature of LDA modeling, we derived the final model through an iterative hyper-parameter tuning process that varied the Alpha, Beta, and K (number of topics) parameters and compared their coherence scores. We chose the coherence score due to its utility as a performance metric for LDA model performance (Röder et al., 2015). The resulting 540 models revealed that K=10 was the best fit with the highest average coherence (0.62). A visual of the topic groupings by Jensen-Shannon divergence can be found in the appendix section containing supplemental tables and figures (Figure A1). The source of the visual was an interactive web app created using the LDAvis library available in Python (Sievert & Shirley, 2014).

The next source for breach concepts came from the ancillary data (congressional report and breach literature). These were assessed in two different ways. The congressional report was analyzed for term frequencies, with stop words removed. Breach literature was reviewed initially by a graduate assistant, tasked with extracting breach concepts from the literature. The authors then reviewed this list to group similar concepts together into a single more commonly used term representing all grouped concepts. Terms from both ancillary data sources were added to the SQL database as breach terms. In a manner similar to other sentiment scoring techniques, tweets were scored for each breach concept by querying a Word-to-Vector model built from the Twitter dataset. This model, described in greater detail in the main paper section for Phase 3, was able to be queried for each word in a tweet and its vector distance from a specific breach concept (restricted to the top 100 nearest terms). The model would return a weighted value for each word in a tweet. These were then summed and divided by the number of total words queried, giving a ratio value similar to the type of score one would receive from traditional psychometric scoring techniques that rely on dictionaries. In some ways, this approach to scoring is more granular because the dictionary represented by the vocabulary of the W2V model returns weighted values based on vector distance. Whereas a psychometric dictionary approach, such as LIWC, treats each word in the dictionary with the same weight. Additionally, using a W2V approach allows for the scores to be contextualized by the discussion on which the model was built. The final output from this process was the conversion of each breach concept into a variable that could be analyzed and correlated against the sentiment scores.

A Discussion of Alternative Analysis Choices

Despite our attempts to use a robust set of techniques and data analysis strategies throughout this study (not all of which were included in the final report), there are alternative options that we may have missed. However, we do wish to provide a brief discussion of alternative techniques that we either considered or employed but ultimately did not use in this final presentation of our work.

Alternative sentiment/emotion scoring: There are a variety of emotion and sentiment scoring techniques available to data analysts and researchers. Although we ultimately chose to use LIWC scores, we did explore alternatives such as the NRC emotion lexicon (Mohammad & Turney, 2013) and NLTK emotion scoring using R. In both cases, the differences in the dominant emotions scores were nonsignificant and no other emotion indicated a greater presence than those of anger, anxiety, and sadness. It is possible that other sentiment scoring techniques may emerge, particularly those built for online discussions, and we would recommend that any future work exploring emotions consider new and emerging techniques.

Alternative techniques for modeling the discussion: In our work, we used both W2V and LDA for modeling the discussion for breach scoring and breach concept extraction, respectively. Although LDA has been demonstrated to outperform other topic modeling techniques, resulting in tremendous popularity among researchers in machine learning (Kherwa & Bansal, 2018)[7], there are numerous other techniques available to researchers. These include derivatives of LDA, such as Hierarchical-LDA (Blei et al., 2004), and LDA alternatives such as Probabilistic Latent Semantic Analysis (PLSA) (Hofmann, 2013). Researchers may find that one of these alternative topic modeling techniques outperforms LDA for their particular dataset or modeling task. However, LDA has been shown to outperform other techniques in general (Kherwa & Bansal, 2018) and, in our study, it was able to yield coherent breach concepts as part of a 10-topic model.

Supplemental Tables and Figures

Figure A1 is a snapshot of a lightweight visualization app created for the LDA model of the Twitter dataset for this study. This was created using the LDAvis library created by Sievert and Shirley (2014). This library facilitates the creation of a web-based interface. Through this interface, pictured below, each topic can be inspected for its most salient terms, ranked by their topic relevancy and with bars indicating topical ownership of the term (within-topic frequency versus without-topic frequency). As a result of a careful hyperparameter selection process, which involved the creation of 540 total LDA models, the final best model resulted in topics with nearly complete ownership of the relevant terms. As part of our breach-concept identification and collection process, we examined each topic term set for breach-related terms. An interesting observation facilited by this interface is the intertopic distance mapping of the 10 topics (the distance between them is a Jensen-Shannon divergence measure). It became clear that there are three major topical groups, with one group being a cluster of 8 topics identified by the LDA model.

We used the word-to-vector (W2V) model to facilitate the breach-concept scoring and assess the structures and nature of the conversation characteristics represented in the Twitter dataset. W2V is a modeling of the syntactic and semantic structures that exist in the dataset (Mikolov et al., 2013a). The W2V model *learns* the rules of the conversation from both a formal and grammatical perspective (syntactic) as well as a meaning and conceptual perspective (semantic). W2V models are highly dimensional (potentially as high as the overall vocabulary, but typically between 300 and 1000 vector dimensions). To gain a visual understanding of the structures inherent to a W2V model, a dimension reduction needs to be employed. To this end, we chose t-SNE reduction which is a modified Stochastic Neighbor Embedding technique (van der Maaten & Hinton, 2008). This allows a reduction that balances global structure preservation and representation in the lower (destination) dimension. In other words, a t-SNE reduction preserves global structures that exist in the full dimensional space, post-reduction. Subsequently, we proceeded to label vector points associated with breach concepts that were significantly correlated with the emotion scores (see Phase 3 results for more on this analysis). The process revealed notable differences in the words of various parts of the reduction. As seen in Figure A2 below, there was a clear separation in the terms used more broadly by individual users and the terms used by media outlets. The connecting line of vector points between these two groups was largely syntactical in nature with low semantic value. It was also noted that one group contained many terms associated with blame (e.g., behavior, unethical, unlawful, crime, and bribe) while the other contained terms associated with breach details (e.g., fingerprint, secret, database, backdoor, and background). Interestingly, the emotions significantly correlated with terms were evenly spread between the two groups (Because of the number of points and labels, we were not able to include the labeled version of this plot in the paper, instead opting for the annotated one. However, a 4k image copy is available upon request).

[7] This reference is recommended to any reader interested in exposure to a variety of topic modeling techniques.

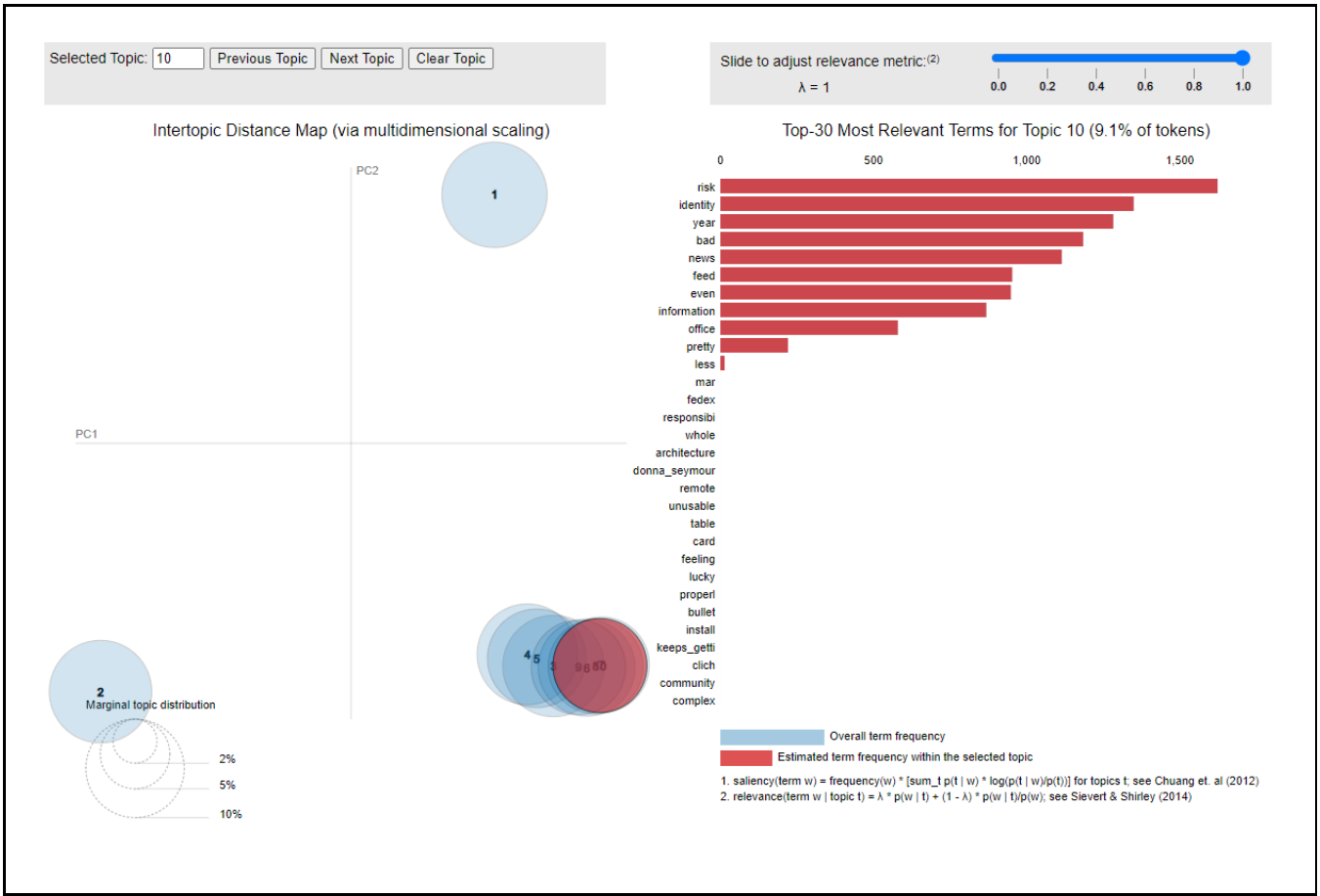


Figure A1. LDAvis Topic Model Screenshot

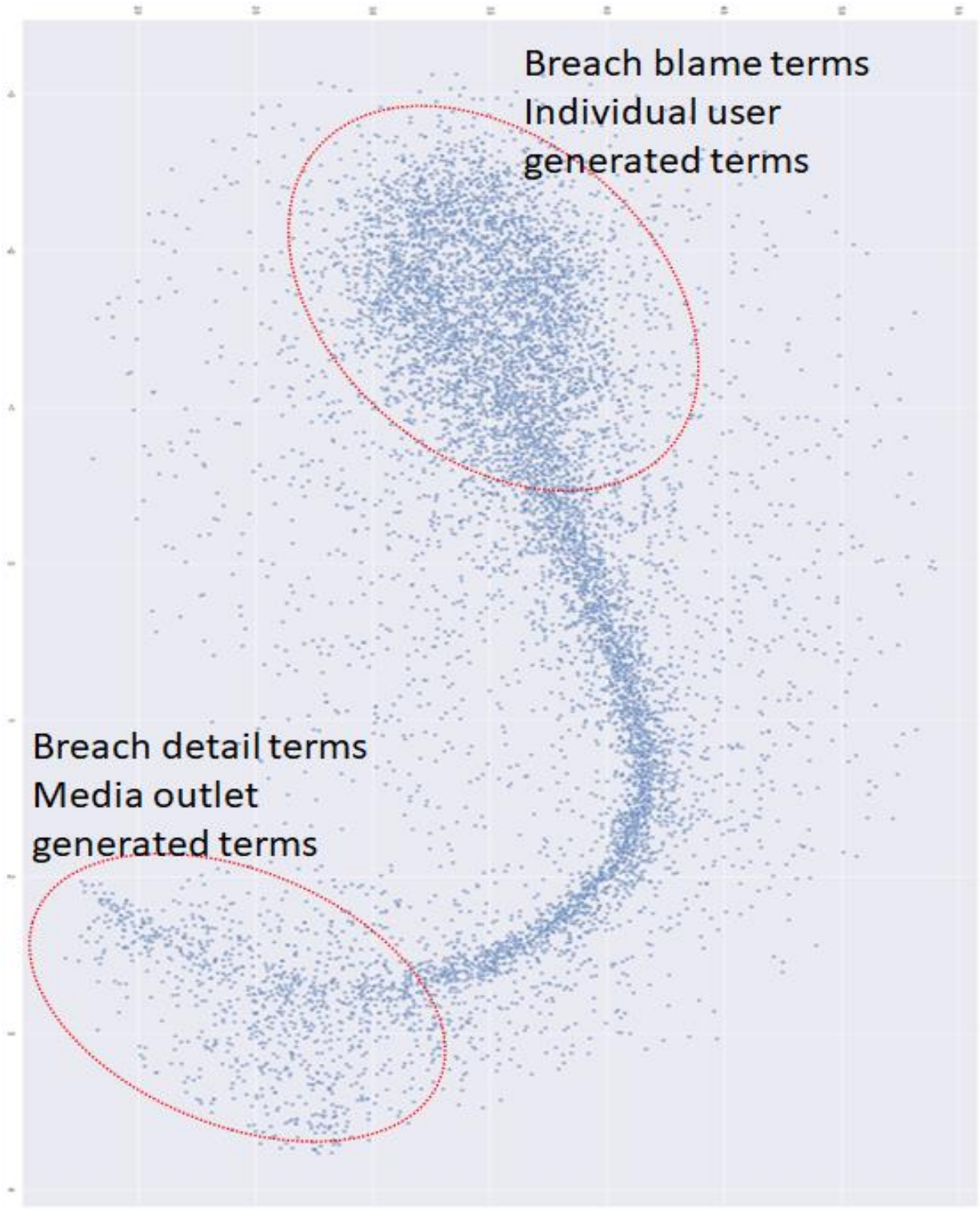


Figure A2. t-SNE Reduction of Entire W2V Model (no labels)