Proceedings of the 52nd Hawaii International Conference on System Sciences | 2019

A Social Network Analysis (SNA) Study On Data Breach Concerns Over Social Media

Naga Vemprala
The University of Texas at San Antonio
Naga.Vemprala@utsa.edu

Glenn Dietrich
The University of Texas at San Antonio
gdietrich@utsa.edu

Abstract

In the current era of digital devices, the concerns over data privacy and security breaches are rampant. Understanding these concerns by analyzing the messages posted on the social media from linguistic perspective has been a challenge that is increasing in complexity as the number of social media sites increase and the volume of data increases. We investigate the diffusion characteristics of the information attributed to data breach messages, first based on the literary aspects of the message and second, we build a social network of the users who are directly involved in spreading the messages. We found that the messages that involve the technicalities, threat and severity related security characteristics spread fast. Contrary to conventional news channels related posts on social media that capture wide attention, breach information diffusion follows a different pattern. The messages are widely shared across the tech-savvy groups and people involved in security-related studies. Analyzing the messages in both linguistic and visual perspective through social networks, researchers can extract grounded insights into these research questions.

1. Introduction

Data breaches have been one of the most popular discussion topics over the last decade. There are various companies and many customers that are affected because of the data breaches at commercial firms including BJ's Stores, Target, Home Depot, the Office of Personnel and Management (OPM), Equifax, Anthem, Coca-Cola are to name a few. The Pew research showcase finds that 90% of Americans feel they have lost control of their personal data [1]. Many scholarly studies were conducted around data breaches in many different domain interests. Some of the studies focused on making policy changes, stricter laws [2], estimating cost of data breaches [3], educating users about security [4], [5], estimating the loss of market value post breach [24], [25] and on the after effects of data breaches by analyzing the responses of data breach victims [7] and by tagging the psychological theories to the breach situation [6]. The earlier studies in the Information Systems space focused on analyzing the outcome of the data breach using event studies [24], [25] and processed the responses of victims by surveying them using a set of well-phrased questions around the situation. However, there is not much research available on the data breach information/discussion encompassing public reactions stated over social media. Analyzing the first-hand opinions from a wide range of victim-groups on social media facilitates in bringing valuable insights about public reactions to the post-breach situation. These insights help the organizations and lawmakers to protect the reputation and mitigate the losses incurred through the data breach. Specifically, in this study we address the following research questions:

RQ1: What characteristics of a breach message makes the information diffuse widely across the social network?

RQ2: Does the presence of evidence in the form of media files, URLs and videos, drive information diffusion?

RQ3: Who among the social media users exchange wide amounts of breach messages?

To answer these questions, we collected the social media messages about the data breaches from Twitter.com. Twitter data is extensively used in research because of its popularity and extensive reliance by various user groups [26]. We studied the prior literature on this subject to retrieve and analyze key aspects about the victim-groups. Based on the findings from earlier studies [27], data breach related keywords and the survey questions used, we have retrieved 16 keywords that have been used in previous studies. To find the contextually similar words and to extract the words for the study, the words were converted into a meaningful numeric form as a real-valued vector and we represented them on a scale. For this reason, we have used the word

URI: https://hdl.handle.net/10125/60155
ISBN: 978-0-9981331-2-6

embedding [19] (The vector representation of words using real numbers) to represent semantically connected words together. The numerically extracted words that are contextually similar in meaning to the 16 words we previously selected, allows us to form the basis for building the features for our model. We used the Negative Binomial Regression, a statistical method employed by the prior study to logically conclude which set of features significantly contributed to the information diffusion. The first two research questions of our study can be answered using the results from our statistical methods. Finally, we have segregated the users who shared the breach information on social media and built a Social Network of the users. This provides a visual representation in answering our third research question.

From a technical and methodological standpoint, our study is novel and first of its kind building the social network on breach message information. We provide three important contributions through our study:

1) A highly accurate and broadly demographic sample: Utilized social media data in real-time for capturing user responses to the data breaches avoiding surveying users.
2) Representation sample: Using word embedding for extracting contextually similar words from the entire set of user generate messages not only looks at the meaning of the words but also the context a word represents. This approach is more robust than using a linguistic dictionary as a reference to extract similar meaning words from the dictionary.
3) Quick analysis on user characteristics who involve in message diffusion through social networks backs the evidence provided by statistical results on the dissemination characteristics of a message.

2. Theoretical model and Hypothesis

Our study broadly builds on the domain aspects of data breaches and systems security, two streams of technical research from the existing literature in social networking analysis and the sparse neural networks. The sub-sections provide detailed explanation about prior theoretical and methodological techniques utilized in the prior studies and finally helps in building hypothesis is derived based on the theoretical background.

2.1. Data Breach studies

With the increased number of data breaches, there has been a huge uproar among all the user groups exchanging information. Especially with the social media becoming more and more available through many channels, the concerns are equally rising at an unprecedented rate. Bulgurcu et.al. studied the phenomena of tightening security controls and educating employees and users on the data security. [8] These measures improve the overall security within an organization and reduce the effect for a data breach. Some of the researchers focused on the ethical aspects of the breach and studied how the organizations should consider ethical aspects of managing customer's private information [9]. These studies are more focused towards identifying the root causes and how the future data breaches could be avoided. Victim-groups are asked to respond to a set of pre-defined survey questions aligned in the direction of research questions. These responses are further analyzed for understanding the user perception towards capturing the user responses through a set of survey questions. These surveys are not only costly but also poses the risk of building a non representational sample.

Moreover, these studies suffer a high degree of margin-of-error [20] due to various unaccounted factors including users subject matter knowledge. This happens even after considering heterogeneous user's groups while collecting the survey data. To address the issue of addressing representational sample, Chakraborty et al., collected data using Qualtrics, from a random sample of user groups segregated by gender and age. [21]. With the help of the tool provided by Qualtrics, the responses for the survey questions were captured. One limitation of the study is about the concentration area of shopping experience of various users depending on the recent data breach trends. The problem of limited sample size applicable to one domain area could be addressed if a diverse sample with heterogeneous users is considered for the study. This is only possible with the first-hand data collected from social media. Social Media data is one such platform where users communicate at will on various topics. The research interests of users also cover vast interest groups. Mining the responses of users on data breach topics provided valuable information. In this study we collected the data breach related information from Twitter.com for mining the information, which provides an alternate solution of capturing survey responses and we extracted the user interests by reading through their public profile pages.

2.2. Neural Networks (Word2Vec)

The first step of the research methodology is to find the contextually similar words for the given set of input keywords. We used the prior literature for building a word dataset that was as the input to query for the contextually and semantically similar words. This

allows us to build the features (message characteristics) of data breach messages. The contextually similar words are used for establishing the features for the regression model, which is explained in the next sub-section. Advancements in the linguistic studies enabled scholars to retrieve the contextual words around a given query word. Converting words into numeric format helps in processing text and representing it in a space vector based on the similarity. This process of retrieving numeric word vector in low-dimension space is known as word embedding. Word2Vec is a neural network based algorithm [10] that takes words as inputs by converting the words into word embeddings. Using these word embeddings that are plotted with similar words, a given input word can be queried to retrieve the contextually similar words around it. For this the second version of the algorithm is used. There are two versions of Word2Vec, Continuous Bag of Words (CBOW) [11] and Skip-gram model [10]. CBOW takes a set of words to get the missing contextual word, while the skip-gram model takes an input word to fetch the set of contextual words around the given input word, utilizing the text corpus for training. In our study, we used a set of machine learning, including the Skip-gram model and data analytics tools to dynamically learn word embeddings from the ongoing supply of fragmented social media messages. We performed text pre-processing to filter the unrelated information that induce the noise contained in these messages. This eliminates certain messages that are not related to the data breach. E.g: Username containing breach which is used in a tweet. Some announcement about cybersecurity training, posts that contain taglines with "breach" related keywords etc. were also eliminated from the study. We created the dimensions in tweets using shallow (single layer network) neural networks technique called Word2Vec, which maps a word to the target (target is also a word). Word2Vec focuses on shared meaning of the words rather than on links. For example, two words are connected if their use of their concepts overlap. It is an unsupervised machine learning model that does not require any labels and can learn the weights of words as word vector representations incorporating most of the semantically rich information. These vectors can then be used to find similarities between words to get relations.

2.3. Negative Binomial Regression (NBR)

Regression analysis is the most popular method used in research. However, mere multiple linear or multivariate regression analysis with Ordinary Least Squares (OLS) method is not a one fit solution for many problems. Especially when the dependent variable is a continuous variable which in the context of Twitter, it is the retweet count, then log-linear regression is usually employed. In analyzing the information diffusion considering the sentiments carried in the message, Stieglitz and Dang-Xuan [16] utilized NBR with retweets count as the dependent variable and sentiment, hashtags, and URL, and followers are independent variables. A similar analysis is carried out by Lee, Agrawal and Rao [17] to estimate the message diffusion on rumors during the Boston Bombing. We adapted the same statistical method for capturing the significant message characteristics as the assumptions from previous studies holds valid for our study. Along with the URL, the number of hashtags we have used, the message characteristics like stress, public opinion, anxiety, efficacy, threat, hacking related and company-specific discussions have also been as features. We ran the Negative Binomial Regression using SPSS software.

2.4. Social Network Analysis

Social network analysis (SNA) is a well-developed methodology that uses network graphics to illustrate social structures among related subjects. A SNA network graph consists of nodes (also known as vertices) and relations (also known as edges). The nodes can be individual or group entities in a network [12] while edges are the connections between nodes and can be either uni-or bi-directional (dyadic). These notions were initially introduced by sociologists and subsequently broadly applied to multiple disciplines, such as information science [13], marketing [14], psychology [15], among others. Based on these network constructs, a variety of properties such as centrality, diversity, cohesion, and equivalence can be measured and adopted to summarize or predict the evolution of the network and the performance of individual entities in the Exploring Untapped Markets by Tracing Through Weak Links network. In this study, we use nodes to represent similar group of users and the communication between these groups are represented using the edges.

2.5. Hypothesis

Our research plan consists of setting up a dictionary of words that convey the various information opinions during a data breach. We have studied literature from the perspective of customers who are also the victim and from the viewpoint of the firms in protecting the systems [22] [23]. The major concerns are broadly related to stress, anxiety, opinion about company image or a news, self efficacy in addressing the problem, threat and severity associated with the breach due to information hacking and loss of privacy. We have established our

search keywords based on the words related to these aspects. Using these keywords, we extract the semantically similar words that are used within the entire set of messages. These words help in identifying message characteristics. Not all characteristics of a message play the catalyst role in information propagation. Identifying these keywords that are significant in information propagation and visualizing groups of users based on their interests contributes to the information diffusion, is very helpful. To build a network, we need to establish a relationship between the aspects of the message. We use the prior literature studies and the survey questions to frame the initial set of words to draw the semantically related words used in the overall set of data breach related message. The research model consists of literature reviews we studied to determine the features extraction, regression and graph visualization.

The user profiles provide interesting insights on the interests they share. Running a classification algorithm to segment user profiles into multiple groups based on the words they used in the profiles helps in establishing a user network. Based on the initial data we collected, it is evident that majority of the profiles are interested in security, technology, news, social aspects and other general interests. Based on this foundation we have framed our hypothesis as below:

H1: Stress, anxiety and opinion about company should have positive effect on the information dissemination.

H2: Evidence based messages (messages with URLs and more hashtags) contribute more positively to information dissemination

H3: Security and Technology related interest groups creates a dense social network on information sharing and positively contributes for information dissemination

Running a negative binomial regression using the features we framed helps in answering the first two hypotheses, where we obtain the coefficients for the regression model with the retweets as dependent variable on a logarithmic scale. If the coefficients values are positive and large they contribute more towards information diffusion. The social network built using the user profiles helps in answering the last hypotheses. If a dense network is built around technology and security focused groups and more links are established, then we can conclude that they contribute more towards information diffusion.

3. Research Methodology

The research methodology for validating the retrieval of message characteristics that contributes to the information diffusion and visualizing various user groups involved in sharing messages constitutes of primarily six phases as shown in the figure 1. starting from collecting social media data (step (a) in the figure) to visualizing the social networks of user groups (step (f) in the figure). The steps in the methodology of implementation are represented in clockwise direction in the figure 1. All the six phases of the methodology are explained below from sub-sections 3.1 for (a) through 3.6 for (f).

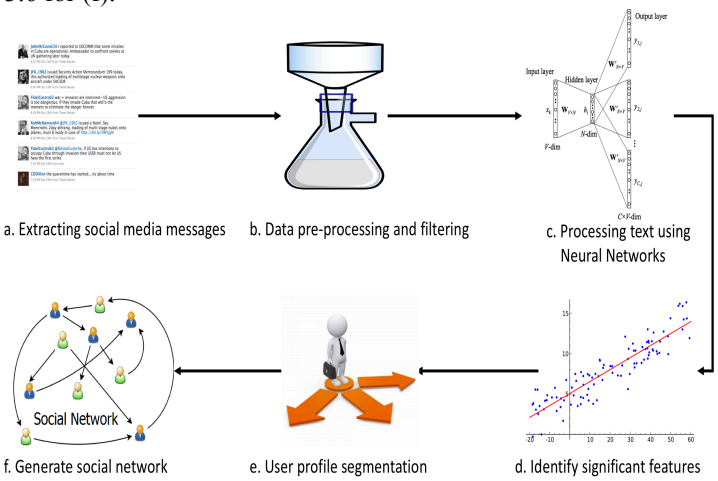

Figure 1. Context diagram

3.1. Data Collection

Data collection is the first phase of our methodology. To validate and test our hypothesis we have collected the user tweets from social media website Twitter. The dataset covers a period of about 35 days of tweets from May 1st, 2018 till May 21st, 2018. The total number of tweets collected are 73,036 using various search keywords and hashtags. The breach data includes discussions around Facebook, Commonwealth Bank a major Canadian bank data breach and Coca-Cola breach. We used a combination of search criteria, using both the hashtags and search keywords. To make the data collection exhaustive, we have used multiple keywords other than just "data breach" as a single keyword and considered 7 hashtags in conjunction with the search keywords. The below table shows the list of search keywords and the hashtags used.

Table 1. Search Keywords for data collection

Search Keywords	data breach, breach, security, hack
Hashtags	#databreach #cybersecurity #data #infosec #cybercrime #breach #hacking

3.2. Data Pre-processing and Cleaning

The next phase of our methodology includes data cleaning and filtering unnecessary records. As it is with every text mining technique, the quality of output depends on the quality of input used for training the algorithm. So we have removed all the unnecessary special characters, URLs, hashtags, mentions, common and most frequent stop words like the prepositions, and other attribute fields like @RT for retweet. Some of the tweets we collected are not related to the data breach. Examples of which include the sarcastic messages tagged with #breach hashtag and class training notifications tagged with hashtags #cybersecurity. These are identified using multiple search conditions.

3.3. Word2Vec

The third phase of our methodology involves extracting contextually similar words. For this we use Word2Vec, a single layer neural network. Word2Vec is an unsupervised machine learning algorithm. We have adapted the tensorflow (a Python package) implementation of Word2Vec to capture results. We pass the input sentences as words by tokenizing the sentences to words. The Word2Vec converts the high dimensional representation of words and sentences into low dimensional word embeddings. The output of running the algorithm is the closest words of the given word based on the input text. The input text is the set of tweets related to data breach and the query word. If the input word is "stress", then based on all the tweets passed to the algorithm, the closest words associated with "stress" are "reaction", "effect", "risk", "bug", "learn", and "response". These are the 6 words closest to "stress", as they either appear before the word or after it in the semantic context. The window used here is 3 before and 3 after, which makes the algorithm to return 6 closely associated words. Taking too many close words is not ideal as there is a high probability of considering overlapping words and it adds up to noisy data. The below table shows all the words semantically associated to input word.

Table 2. Search Keywords for data collection

Query word based on prior literature	Semantically close words
stress	reaction, effect, risk, bug, learn, vulnerable
public	opinion, detail, social, prepare, reveal, key, comprehend
anxiety	worry, sensitive, personal, perplex, report, compromise
efficacy	regulate, security, account, password, protect, legal
threat	discovery, massive, missing, million, severe, upset
hack	breach, attack, stole, discover, share, avoid
company	facebook, equifax, uber, arronbank, commbank, cambridgeanalytica

3.4. Estimation methods

We utilized regression analysis to validate hypothesis (1) and (2). We derive the keywords from the entire set of tweets about data breaches, by querying the model using the words collected from literature review based on prior studies. These derived keywords form the basis for features of the model. We search for the presence of these words in each and every tweet message and mark for its presence using binary values of 1 and 0 for presence and absence of the keywords respectively. We have also considered the presence of URLs in the message, which is usually an image, video file or a news article link, conveying the evidence for the message posted or shared and the number of hashtags in each tweet. With these being independent variables, we have taken the retweets count as the dependent variable for our regression model. With the retweets count being a continuous integer variable, we have used the negative binomial regression model as below:

$$\log(\text{retweet count}) = \beta 0 + \beta 1*\text{url} + \beta 2*\text{hashtags} + \beta 3*\text{stress} + \beta 4*\text{public} + \beta 5*\text{anxiety} + \beta 6*\text{efficacy} + \beta 7*\text{threat} + \beta 8*\text{hack} + \beta 9*\text{company} + \epsilon$$

3.5. User profile segmentation

Segregating users based on their interests provides insights about the user groups who are involved in sharing the messages. For this we have relied upon the estimation parameters. We retrieved the users who posted or shared the messages that contain the significant keyword parameters. E.g: If "stress" turns out to be a significant parameter, we capture all the users who posted messages related to stress. Using the

username from the tweet post, we retrieved the profile description from Twitter. We ran the clustering algorithm [18] on the text description to retrieve topics that interest social media users who tweet on data breach messages. Table 3. shows all the topic words generated from the user profiles.

Table 3. Interesting topics of social media users

User Group	Topic Words
Security Awareness	analytics, cybersecurity, malware, safety, privacy, solutions, security, pentest
Technology Focus	tech, infrastructure, jobs, bigdata, blockchain, finance, nosql, bigdata
News Related	reporter, news, politics, bloomberg, article, breaking, events, broadcast
Social Work	social, country, fight, activist, controversy, consumers, cloud, people
Other interests	markets, women, business, professional, fun, network, wallets, casual

3.6. Social Network Analysis (SNA)

The last phase of our methodology is to create a visualization of the category of users who share the messages across social media. To gain a deeper insight into the information diffusion, we build a social network using the groups the users are clustered into. We retrieve all the messages that contain the significant characteristics in a message along with the information of user who posted, replied or shared the message. We did not build a directional graph as we are interested in finding communication between user groups but not the direction of communication. Users are the nodes of the graph and the communication link is the edge between the nodes in the graph. Edges of equal weight are formed between user1 and user2, if a message posted by user1 or user2 is replied to or shared by user2 or user1 respectively. We color code the type of user groups in the social network for clear visualization.

5. Results

We summarize the results of regression in Table 4. The first column represents all the variables in our mode including the intercept and the coefficients are present in the estimate. The higher the value of coefficient, the more is the contribution of the variable in the overall model. Also, Z-value provided in the table 4 provide valuable insights about the error. The higher the magnitude of Z-value there is less error in the overall model due to this variable and it has to be considered in the overall model. Standard error is captured in the third column with the Z Value in 4th which could also provide an estimate of standard deviation. The exponent of estimate is provided in the last column. The highly significant variables of the model are intercept, URLs, and Threat. The next significant variable is Stress. All these 4 factors are contributing positively which conveys that the information diffusion takes place with the presence of stress and threat in the messages.

Table 4. Regression results

User Group	Estimate	Std.Err	Z Value	Exp(β)
Intercept	0.854	0.229	3.729	2.349***
URLs	1.026	0.181	5.668	2.790***
Hashtags	0.745	5.602	0.133	-
Stress	0.303	0.109	2.756	1.353**
Public	5.781	4.314	1.340	-
Anxiety	0.232	0.124	1.880	1.261
Efficacy	0.278	0.165	1.683	1.320
Threat	-0.469	0.121	-3.891	0.625***
Hack	-0.732	0.441	-1.661	0.481
Company	0.474	0.246	1.928	1.606

Based on the above statistical results, the updated model turns out to be as below:

$$\log(\text{retweet count}) = \beta 0 + \beta 1*\text{url} + \beta 3*\text{stress} + \beta 5*\text{anxiety} + \beta 6*\text{efficacy} + \beta 7*\text{threat} + \beta 8*\text{hack} + \beta 9*\text{company} + \epsilon$$

The figure 2., shows interaction between various user groups. There is wide amount of communication between technology focused groups and security related experts. While there are significant contributions made by the news channels and personnel from media, the other community users were not widely discussed their

posts. There are users who responds to some of the data breach messages have wide range of interests that are not specific to breaches alone. Interestingly the users who are more concerned toward society in general responded to significant amount of posts. Thus, visualizing the posts with significant message characteristics by user groups provide interesting insights into the information diffusion pattern.

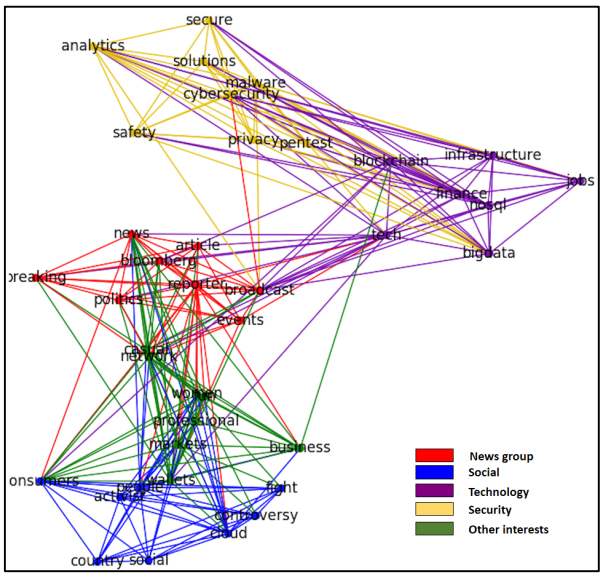

Figure 2. Social Network of various user groups

6. Conclusion and Limitations

Our research provides a novel approach of finding the message characteristics that facilitates information diffusion. Especially, addressing data breach scenarios through this study helps organizations and governments officials two folds by not only providing insights into the kind of messages that are vulnerable in the social media platforms and the type of user groups that they need to address through communication in order to mitigate the costs through data breaches and helps in reestablishing lost reputation.

We have utilized the prior studies for creating a list of query words for capturing the semantically similar words from the social media messages. This list could not be considered as exhaustive. Adding additional query strings to the search criteria could bring in more valuable insights and could bring in other significant features from the data. Even though we have considered an automated manner of retrieving the characteristics of breach messages through text mining techniques and unsupervised machine learning, we did not perform the longitudinal analysis using the message posts from various social media users. Restricting the posts through certain users by time period helps in performing this analysis. We plan to include this study in our next extended work on this data. We have considered 35 days of data breach messages as a test bed for validating our study, increasing the volume and timing of these messages toward breach situation and then segregating the messages by industry could bring in different insights. This is another perspective which we would like to add to follow-on studies.

7. References

[1] Kedmey, D. (2014, November 12). 9 in 10 Americans Feel They've Lost Control of Their Personal Data. Time. from http://time.com/3581166/privacy-personal-data-report/

[2] Romanosky, S., Telang, R., & Acquisti, A. (2011). Do data breach disclosure laws reduce identity theft?. Journal of Policy Analysis and Management, 30(2), 256-286.

[3] Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. ICIS 2006 Proceedings, 94.

[4] Belsis, P., Kokolakis, S., & Kiountouzis, E. (2005). Information systems security from a knowledge management perspective. Information Management & Computer Security, 13(3), 189-202.

[5] Choi, N., Kim, D., Goo, J., & Whitmore, A. (2008). Knowing is doing: An empirical validation of the relationship between managerial information security awareness and action. Information Management & Computer Security, 16(5), 484-501.

[6] Barnes, C. M., & Van Dyne, L. (2009). I'm tired': Differential effects of physical and emotional fatigue on workload management strategies. Human Relations, 62(1), 59-92.

[7] Elhai, J. D., & Hall, B. J. (2016). Anxiety about internet hacking: Results from a community sample. Computers in human behavior, 54, 180-185.

[8] Bulgurcu, B., Cavusoglu, H., & Benbasat, I. (2010). Information security policy compliance: an empirical study of rationality-based beliefs and information security awareness. MIS quarterly, 34(3), 523-548.

[9] Culnan, M. J., & Williams, C. C. (2009). How ethics can enhance organizational privacy: lessons from the choicepoint and TJX data breaches. Mis Quarterly, 673-687.

[10] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and

phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119).

[11] Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. ICLR Workshop.

[12] Burt, R. S., Kilduff, M., & Tasselli, S. (2013). Social network analysis: Foundations and frontiers on advantage. Annual review of psychology, 64, 527-547.

[13] Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. Journal of information Science, 28(6), 441-453.

[14] Ansari, A., Stahl, F., Heitmann, M., & Bremer, L. (2018). Building a social network for success. Journal of Marketing Research.

[15] Laninga-Wijnen, L., Ryan, A. M., Harakeh, Z., Shin, H., & Vollebergh, W. A. (2018). The moderating role of popular peers' achievement goals in 5th-and 6th-graders' achievement-related friendships: A social network analysis. Journal of Educational Psychology, 110(2), 289.

[16] Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. Journal of management information systems, 29(4), 217-248.

[17] Lee, J., Agrawal, M., & Rao, H. R. (2015). Message diffusion through social network service: The case of rumor and non-rumor related tweets during Boston bombing 2013. Information Systems Frontiers, 17(5), 997-1005.

[18] Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of machine Learning research, 3(Jan), 993-1022.

[19] Turian, J., Ratinov, L., & Bengio, Y. (2010, July). Word representations: a simple and general method for semi-supervised learning. In Proceedings of the 48th annual meeting of the association for computational linguistics (pp. 384-394). Association for Computational Linguistics.

[20] Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of management information systems, 12(4), 5-33.

[21] Chakraborty, R., Lee, J., Bagchi-Sen, S., Upadhyaya, S., & Rao, H. R. (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.

[22] Patel, A., Qassim, Q., & Wills, C. (2010). A survey of intrusion detection and prevention systems. Information Management & Computer Security, 18(4), 277-290.

[23] Choi, B. C., Kim, S. S., & Jiang, Z. (2016). Influence of firm's recovery endeavors upon privacy breach on online customer behavior. Journal of Management Information Systems, 33(3), 904-933.

[24] Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. International Journal of Electronic Commerce, 9(1), 70-104.

[25] Acquisti, A., Friedman, A., & Telang, R. (2006). Is there a cost to privacy breaches? An event study. ICIS 2006 Proceedings, 94.

[26] Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—sentiment of microblogs and sharing behavior. Journal of management information systems, 29(4), 217-248.

[27] Cavusoglu, H., Mishra, B., & Raghunathan, S. (2004). The effect of internet security breach announcements on market value: Capital market reactions for breached firms and internet security developers. International Journal of Electronic Commerce, 9(1), 70-104.