A VISUAL MONITORING TECHNIQUE BASED ON IMPORTANCE SCORE AND TWITTER FEEDS

Zhenhuan Sui, David Milam, and Theodore Allen

The Ohio State University
1971 Neil Avenue, 210 Baker Systems
Columbus, Ohio 43210

ABSTRACT

As social media is becoming more popular, researchers have begun applying text analytics models and tools to extract information from these social media platforms. Many of the text analytics models are based on Latent Dirichlet Allocation (LDA), a topic model method. But these models are often poor estimators of topic proportions. In this paper, we propose a visual monitoring technique based on topic models, a point system, and Twitter feeds to support passive monitoring and sensemaking. The associated "importance score" point system is intended to mitigate the weakness of topic models. The proposed method is called TWitter Importance Score Topic (TWIST) monitoring method. TWIST employs the topic proportion outputs of tweets and assigns importance points to present trending topics. TWIST generates a chart showing the important and trending topics that are discussed over a given time period. We illustrate the methodology using two cyber-security field case study examples.

1 INTRODUCTION

Zaman, Herbrich, Gael, and Stern (2010) described some of the many possible uses of Twitter and other social media. For example, companies and research institutes are using social media to predict how events will be received based on the thoughts and feelings users are posting. Some Hollywood production companies are using Twitter to predict how a movie will perform. Its use may have also helped improve how particular movies have done in the box office (Britt 2015). Here, we explore the use of Twitter-based analysis methods for improving sensemaking and monitoring. The specific case study examples that we use relate to improving the situation-awareness of system administrators in cyber security contexts. Cyber-security is a growing field of study due to the growing use of data collection and the use of newer internet enabled devices. Therefore, this paper will investigate through examples of the connection between cyber-security and social media, in particular twitter, in addition to their individual importance.

Yang and Counts (2010) studied twitter and used name association and key word identification to track the speed with which tweets travel through accounts and the paths that these tweets take. Their study finds that mentioning an individual on twitter indicates that a tweet will have more diffusion in terms of speed and number of users viewing and reposting said tweet. Zaman, Herbrich, Gael, and Stern (2010) presented a method for predicting the spread of information in a social network using retweets as positive feedback and lack of retweets as negative feedback. The number of retweets can be used as an important indicator in the prediction model for social events and changes. Zaman, Fox, and Bradlow (2014) used a Bayesian approach to develop a probabilistic model for the evolution of retweet counts. Their model successfully predicted the final total number of retweets through the time-series path of retweets. In our examples, we use retweet counts as an indicator of importance and our point-based "importance score" can be viewed as an approximate estimate of retweet counts.

Building on Latent Dirichlet Allocation (LDA), Allen, Xiong, and Afful-Dadzie (2015) proposed subject matter expert refined topic (SMERT) for probabilistic clustering of texts to permit experts or users

to edit the topics using knowledge about the system or their own needs. SMERT and LDA estimate the proportion of words in the overall corpus on each topic. As a special case of LDA, SMERT potentially incorporates "high-level" inputs from a subject matter expert to adjust the topics and clusters by zapping or boosting words in the topic definitions. Allen, Vinson, Raqab, and Allam (2013) applied the SMERT model to course evaluation analysis. Using Pareto charts, this method helped to screen out less effective feedback and allow researchers to focus on relevant and important information.

Topic models and SMERT have shown promise for creating intuitive summaries of bodies of text. But there are issues with estimation and in particular topic proportions are often poorly estimated and fail to capture what is new temporarily in the topic proportions. Therefore, this article proposes a visualization and point-base system designed to help users with sense-making of twitter feeds. The goals of this paper are to overcome the reported estimation issues from the SMERT models and demonstrate the value of the point system in relation to Twitter monitoring. To illustrate the problems with SMERT and LDA and the advantage of the proposed TWitter Importance Score Topic (TWIST) modeling, we seek to show improved correlation with retweet counts of the point system. In Section 2, we describe a motivating example relating to cyber vulnerabilities and describe the need for interpretation. In Section 3, we will review SMERT models. In Section 4, we propose the point system associated a visualization methods, which could aid in many twitter-related sense-making cases. In Section 5, we return to the cyber-security case studies and illustrate the application of the proposed methods. Finally, we summarize our findings and suggest opportunities for future research.

2 CYBER VULNERABILITY EXAMPLE

To demonstrate the new TWIST method, we use a case from 2014 relating to cyber security. During 2014, there were several major cyber vulnerabilities that became public knowledge. Most notably was the vulnerability commonly known as the Heartbleed. The Heartbleed vulnerability was made public knowledge on April 1, 2014. This vulnerability resulted from a lack of bounds in memory allocations for operating systems. The vulnerability and notification allowed for large amounts of information to be stolen from any susceptible computer. Upon this disclosure many hackers made use of the vulnerability before a patch could be created. As a result the number of attacks on a large Midwest institution's computers increased by approximately 400% in the month of April as shown in *Figure 1*.

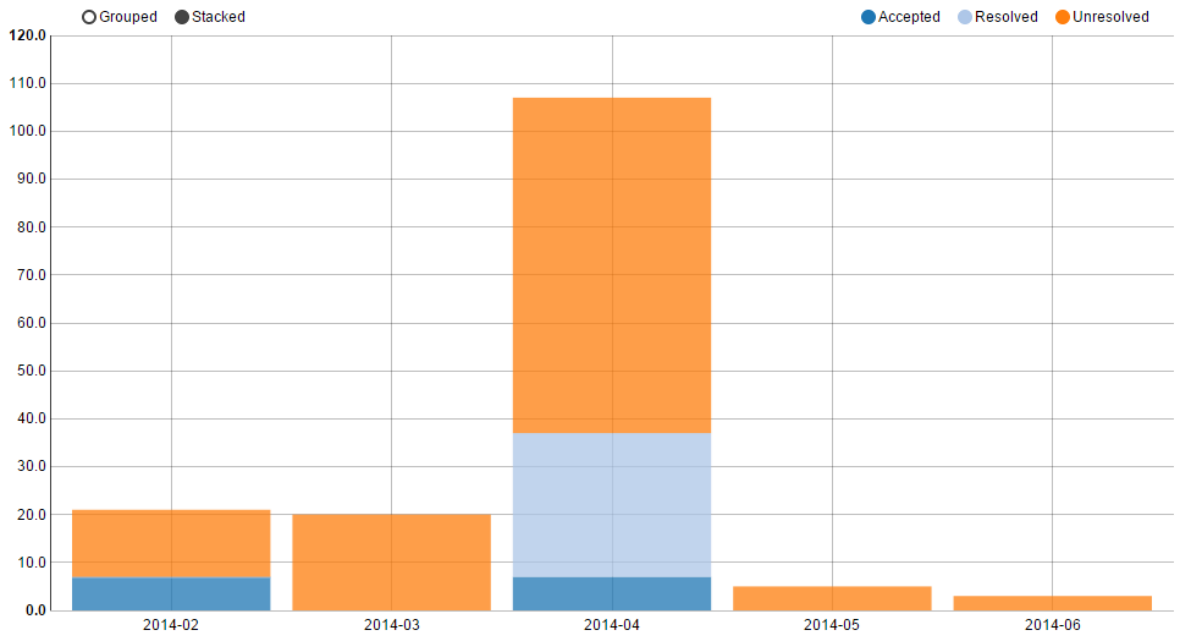

Figure 1: Known computer intrusions for a large Midwest organization in 2014.

No doubt, some system administrators at the Midwest organization knew about heartbleed after the announcement but many did not. Yet, all observed the spike in attacks as detected using the intrusion detection system (IDS). The IDS generally intercepts only a fraction of all attacks so likely some were missed and all administrators needed to perceive the vulnerability and understand its cause. This is the objective of our proposed methods in this article, i.e., to improve situation awareness at all times by synthesizing Twitter feeds into an intuitive chart.

As another example, we consider the November 2014 attack on the Sony Corporation by, reportedly, North Korea. This was a well-publicized attack that received a large amount of media attention. These famous cyber events raise discussions on social media platforms. The proposed methods seek to identify and interpret the events in both cases.

3 SUBJECT MATTER EXPERT REFINED TOPIC

Allen, Xiong, and Afful-Dadzie (2015) proposed subject matter expert refined topic (SMERT) for probabilistic clustering of texts. Both are "topic" models with the topics being clusters of words in the documents associated with fitted multivariate statistical distributions. In practice, not all of the distribution is relevant to the user and the topics can be represented by ordered lists of words which users often find interpretable. SMERT is a generalization of Lantent Dirichlet Allocation (LDA) methods.

SMERT generalizes LDA in that it incorporates input from a Subject Matter Experts (SMEs) or ordinary users. The method derives the main topics with a body of documents and estimate what portion of the text corresponds to each topic. In SMERT, the distribution is in equation (1). The distribution is fit using collapsed Gibbs sampling which is a form of Markov Chain Monte Carlo. Collapsed Gibbs is an iterative process where the topic assignments and distribution are modified. The topic assignments converge to the samples from the new distribution and are then used for estimations for the topics and proportions. Below you will find equation 1, or how fitting the distribution.

$$P(w, z, \theta, \emptyset|\alpha, \beta) = P(\theta|\alpha)P(z|\theta)P(\emptyset|\beta)P(w|z, \emptyset) \qquad ((1))$$

$$\propto \prod_{d=1}^{D} Dir(\theta_d|\alpha) \times \prod_{d=1}^{D} \prod_{n=1}^{N_d} Mult\left(z_{d,n}\middle|\theta_d\right) \times \prod_{t=1}^{T} Dir(\emptyset_t|\beta) \times \prod_{d=1}^{D} \prod_{n=1}^{N_d} Mult\left(w_{d,n}\middle|\emptyset_{z_{d,n}}\right)$$

where $Dir(\theta|\alpha) = \frac{1}{B(\alpha)} \prod_{t=1}^{|\alpha|} \theta^{\alpha_t - 1}$

and $Mult(x|\theta, n) = \frac{n!}{\prod_{k=1}^{K} x_k!} \prod_{k=1}^{K} \theta_k^{x_k} \rightarrow Mult(x|\theta, 1) = \prod_{k=1}^{K} \theta_k^{x_k}$.

$w_{d,n}$ is the n^{th} word in document *d*. "N_d" is the number of words in that specific document, not the entire body of text. "N_d" is the number of words in that document. The number of words that are created in the dictionary is denoted as *"WC"*. The *WC*-dimensional random vector $\emptyset_t$ represents the probability that randomly selected words are assigned to each pixel in the topic indexed by $t = 1 \ldots T$. After applying collapsed Gibbs sampling, the sampled topic assignments ($z_{d,n}$) permit estimation of the topic definitions (β). The most common words in each topic are often the most relevant outcome. The topic proportions can also be estimated (θ) but the estimates are often less accurate.

4 TWITTER IMPORTANCE SCORE TOPIC MONITORING METHOD

4.1 Notations and Assumptions

In this section, we define additional notations and assumptions for the proposed TWIST method. Consider a finite number of text document with *L* sentences and each sentence is signified as s_l, where $l = 1, \ldots, L$.

Our method is based on the SMERT method but it could be based on LDA only. In either case, the derived topics are denoted t_i, $\forall i \in I$, where I is the set of topic indices. Within each topic, the words are ordered as w_{ij}, $\forall j \in$ J and J is the set of word indices. P_{il} is the estimated mean posterior probability (what the Gibb sampling generates) that sentence l falls in the topic t_i. A set of documents is called a corpus and q is the number of top words in each topic that are studied by the subject matter expert. The default is q = 10 words for each topic (top 10). Also, the predicted score or importance number is the *PS*.

4.2 The Proposed TWitter Importance Score Topic (TWIST) Monitoring Method

The TWitter Importance Score Topic Monitoring Method is as follows.

TWitter Importance Score Topic (TWIST) Monitoring Method
Step 0. Select Twitter content to follow and create a corpus of tweets from the relevant time period.
Step 1. Run LDA on the corpus.
Step 2. Loop over each topic t_i, $\forall i \in I$
  Step 2.1 Loop over each word w_{ij}, $\forall j \in the\ first\ q\ words\ in\ the\ topic$, zap w_{ij} if w_{ij} does not make sense. otherwise, boost it. End loops.
Step 3. Run SMERT without sorting topics using the high-level boosts and zaps.
Step 4. Loop over each tweet (sentence) s_l with property m, $V_{im} = \sum_{l \in m} P_{il}$
Step 5. Loop over each topic $m \in M$, rank V_{im} from largest to smallest
Step 6. Select *N* largest values of V_{im}, $V_{nm} = V_{im}$, where $n = 1, \dots, N$
Step 7. For all $m \in M$ and $n = 1, \dots, N$,
  Step 7.1. If count of topic i, $C_i = 1$, assign predicted score $PS_{1im} = C_{1n}$, where $n = 1 \dots N$, $C_{11} > C_{12} > \cdots > C_{1N}$.
  Step 7.2. If count of topic i, $C_i = 2$, assign predicted score $PS_{2im} = C_{2n}$, where $n = 1 \dots N$, $C_{21} > C_{22} > \cdots > C_{2N}$.
  Step 7.3. If count of topic i, $C_i \neq 1$ or 2, assign predicted score $PS_{im} = 0$.
Step 8. $S_m = \sum_i (PS_{1im} + PS_{2im} + PS_{im} + PS)$, where PS is the constant predicted score for all $m \in M$. End loops.
Step 9. Plot PS_{1im}, PS_{2im}, PS_{im}, PS in a column chart with short phrase extracts from the topic definitions as labels.

For *Step 4*, $V_{im} = \sum_{l \in m} P_{il}$ means sum all of the probabilities for the same property. Here, the property includes examples like different months, years, or even days. For SMERT, normally 20 topics are selected as outputs. In Step 6, among the 20 topics, normally, $N = 5$, or the top 5 topics are selected in the TWIST monitoring method in most cases. In the method, predicted scores are normally either equal to predicted numbers or proportional to the predicted numbers.

5 CASE STUDIES

In this section, two cyber Tweet examples are studied using the TWIST monitoring method. The first case study relates to the heartbleed vulnerability from 2014. The second relates to the Sony Hack and, to a lesser extent, the shellshock vulnerability which occurred in the same time period. In this section, these examples are presented to show how the TWIST monitoring method could react to new topics. Admittedly, both case studies could have been studied together but we wanted to evaluate the level of generality of TWIST and explaining them separately is simpler.

The same 15 Twitter broadcasters were analyzed for the purposes of both studies (*Step 0*). These users were found by searching for the twitter users who have a reputation for being cyber-security analysts. Also, a combination of individuals and organizations/groups were found to ensure there wasn't a bias based on the goal of the twitter user. The twitter sources (usernames) in the following examples are:

Mathewjschwartz, Neilweinberg, Scotfinnie, Secureauth, Lennyzeltser, Dangoodin001, Dstrom, Securitywatch, Cyberwar, Jason_Healey, FireEye, Lancope, Varonis, DarkReading, RSAsecurity, and Mcafee_Labs.

5.1 Heartbleed

5.1.1 Background

Heartbleed is a security vulnerability in the windows software package. As discussed earlier, the vulnerability was made public knowledge in April 2014 with wide ranging impacts. Many attacks were attempted before a patch could be applied to remove the vulnerability. The heartbleed received a large amount of publicity due to the severity and number of people it impacted (millions).

5.1.2 Data and Computation Results

Figure 2 shows the total number of retweets for the first six months of 2014 which will be compared later to the chart this new method generates. Notice that the retweet counts correlate with the known intrusion counts confirming that retweet counts often relate to important events.

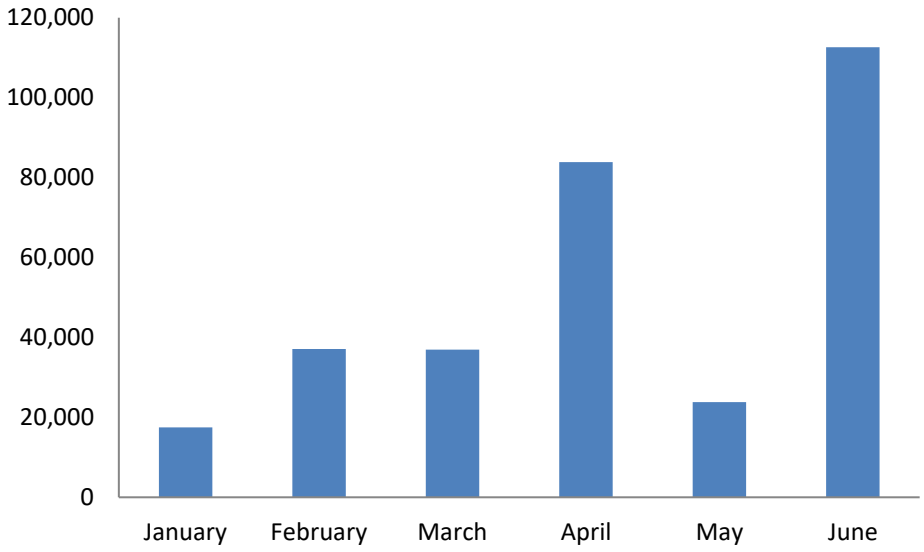

Figure 2: Retweets for January to June 2014 with the heartbleed announcement in April.

Next, we applied the remaining steps in the TWIST method. Below are the topics that SMERT created based upon the tweets and zapping any unwanted words. Steps 1-3 involve applying SMERT. We zapped heartbleed in February and March because we know from our expertise that there were no tweets about heartbleed until April after it was publically announced. The developed topics were then identified by words and the words. Then, we manually translated the word lists into (hopefully) interpretable topics with the results in *Table 1*.

Table 1: SMERT topics which were interpreted manually incorporating the highest frequency words.

Number	Topics
1	Jason Healey and cyberwar, among others, tweeted with a moderate following tweeted in a few months about cyber security.
2	RSA security, among others, tweeted about its own products and an event called the archer summit with a small following.
3	Dangoodin001, among others, retweeted topics from many different months, without much following on Twitter.
4	Cyberwar, among others, tweeted about Eric Snowden and the NSA in multiple months
5	MacAfee Lab, Darkread, and dstrom, among others, tweeted about network security it multiple months
6	Dangoodin001 and Darkread, among others, tweeted about the heartbleed with a moderate following on twitter in particular during April.
7	Security watch and dangoodin001, among others, tweeted about apps and passwords with a moderate following on twitter.
8	Lancop tweeted about its own company in particular during February and March with a low number of retweets.
9	Mathewjschwartz and Darkread, among others, tweeted about the target breach and information security with a low number of retweets.
10	Lennyzeltser and security watch, among others, retweeted topics and specifically at a neiljrubenk on twitter.
11	Mathewjshwartz and dangoodin001, among others, tweeted with a high number of retweets in multiple months.
12	RSA security and Darkread, among others, tweeted about data security in multiple months
13	Fire eye, among others, tweeted about information security, malware, and threats with a moderate number of retweets particularly in April and May.
14	Varoni, among others, tweeted about information security and data privacy in multiple months.
15	Varoni, among others, tweeted about big data and security in multiple months with a low number of retweets.
16	Scotfinnie and security watch among others tweeted about Microsoft windows with a low number of retweets
17	Cyberwar and Dangoodin001 among others tweeted about thanking others in multiple months
18	Varoni and Darkread, among others, tweeted about social media and information security in multiple months.
19	McAfee Lab, among others, tweeted about security stories and particularly to twitter users davemarcu and Slashdot in multiple months with a low number of retweets.
20	Darkread and Varoni, among others, tweeted about information security and Darkread in particular during April and June.

For both examples we use $N = 5$. Also, for steps 4-8, the predicted score is $PS = 10{,}000$. If the topics are unique among all the 6 month, a predicted score of 65,000, 52,000, 39,000, 26,000, 13,000 is assigned if the topic ranks No. 1 to 5 respectively. If the topics appear twice among all the 6 month, assign predicted score of 15,000, 12,000, 9,000, 6,000, and 3,000 if the topic ranks No. 1 to 5 respectively. *Figure 3* shows the predicted scores with a breakdown by topics (*Step 9*).

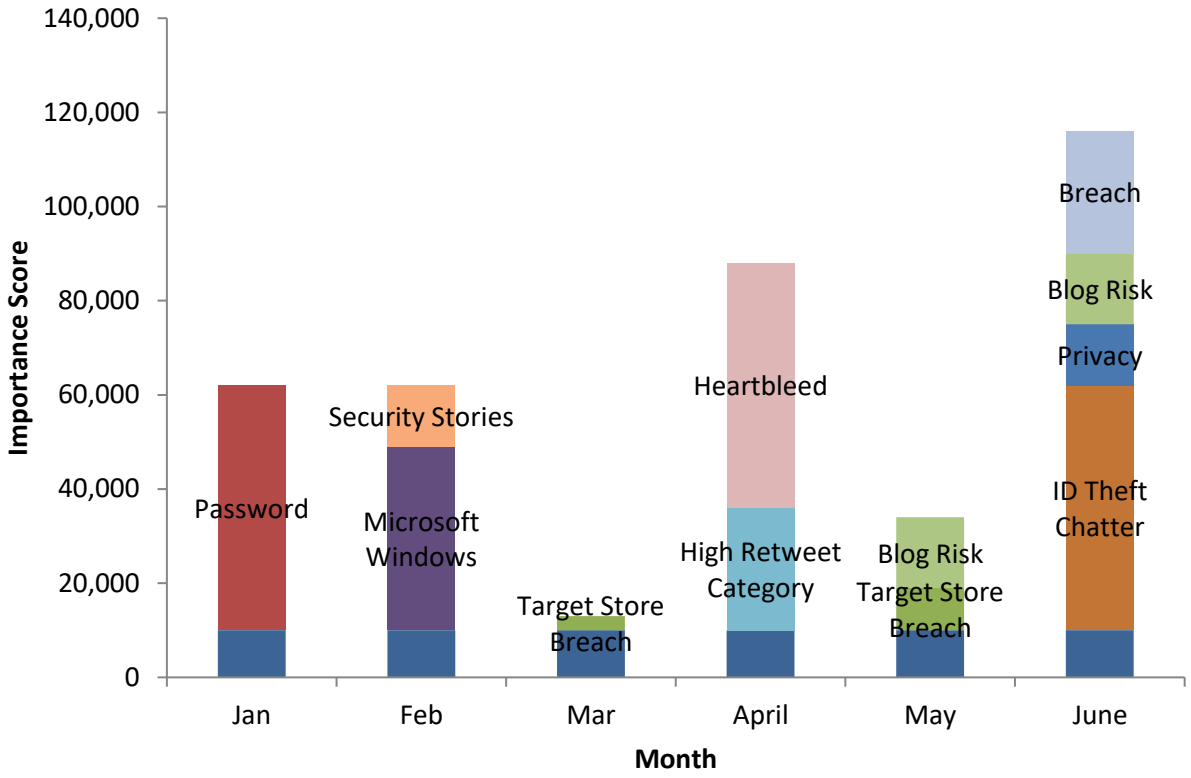

Figure 3: Heartbleed example predicted score breakdown by topic.

5.1.3 Discussion

As can be seen in the figure above, April is characterized by the heartbleed event and the high retweet category. This makes sense, as the heartbleed event would cause a few particular announcements and updates to be highly retweeted. This characterizes April as a month which is abnormal and focuses on the heartbleed event (as we now know is correct from the IDS data in *Figure 1*)

January and February both have slightly elevated PS and predicted retweet numbers as well. The January password focus and February story and system update focus may result in part from the Target credit card theft in December of the previous year, and an increased focus on cyber security. The target theft involved many peoples credit card information being stolen and was a major event for many individuals who may not think of cyber security very often.

The month of June also had a large number of points associated with it. June seemed to have a large amount of discussion associated with breaches of security resulting in theft of personal information and privacy issues. However, this system did a good job of predicting the real retweet numbers. The month of April was clearly dominated by discussions of the Heartbleed vulnerability which is exactly what an IT professional would want to know about if they did not know already.

5.2 Shellshock and the Sony Hack

5.2.1 Background

Shellshock is a security bug in Unix Bash Shell. It was disclosed on September 24, 2014. Many web server deployments use Bash to process web requests. Therefore, the bug could cause potential vulnerability issues

to execute arbitrary commands and allow attackers acquiring unauthorized access to hosts. This bug can be compared to the Heartbleed bug in severity as it could potentially compromise millions of unpatched hosts.

The Sony Hack is another interesting example as it aroused more public attention. However, it is probably less relevant to local system administrators. On November 24, 2014, Sony released a movie called *The Interview*, which is about North Korea and their leader's dictatorship. Therefore, North Korea attacked Sony's online system and hacked Sony employees' personal data. Both of these two events attract active discussions on social media.

5.2.2 Data and Computation Results

The data set is also from the 16 Twitter accounts as in the Heartbleed example, but the data in these two examples are from July to December in 2014.

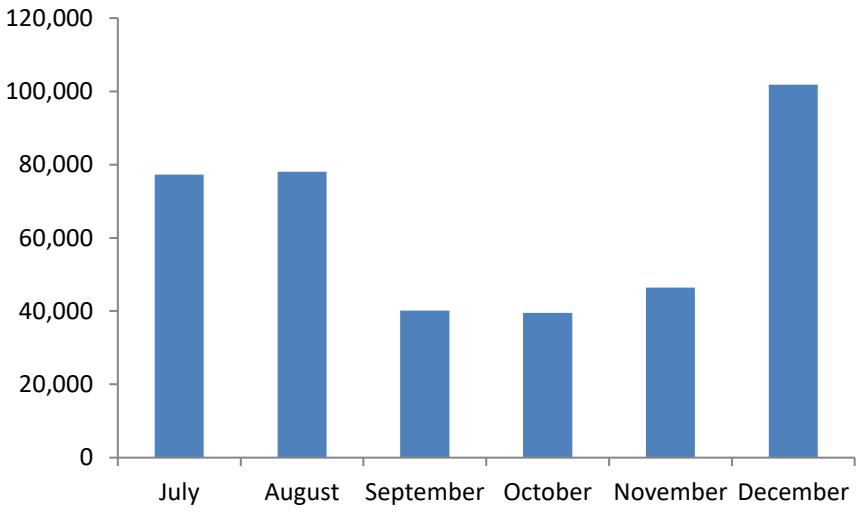

Figure 4: Retweet numbers from the period involving shellshock and the sony hack.

Figure 4 shows the total retweet number across the 16 accounts. Different from expectations about Shellshock and the Sony Hack events, the total retweet numbers in September and November are not very high compared to other months.

Table 2: SMERT topics for the second case study during the shellshock and Sony hack period.

Number	Topics
1	Lancop tweeted about information security and cyber security for companies during July, August, October, and November with high number of retweets.
2	Lennyzelts, varoni and nealweinberg tweeted about new malware tool in October and December with high number of retweets.
3	Varoni tweeted about big data, information security, and data privacy in July and August with high number of retweets.
4	Mathewjschwartz, darkread and scotfinni tweeted about malware breach for Apple during September to November with high number of retweets.
5	Dangoodin001 and lennyzelts tweeted about year 2014 in August and December with high number of retweets.
6	Cyberwar and jasonhealei tweeted and retweeted about new things on internet during July, August and November with high number of retweets
7	Mathewjschwartz, cyberwar and dangoodin001 tweeted and retweeted about the Sony Hack during December with high number of retweets.
8	Dstrom, mathewjschwartz and cyberwar tweeted and retweeted about great reading and look during September and November with high number of retweets.
9	Secureauth tweeted and retweeted about security authenticity during September and October with high number of retweets.
10	Securitywatch tweeted and retweeted about online ID security protection during October and November with high number of retweets.
11	Jasonhealei tweeted and retweeted about cyber attack and National Security Agency (NSA) during September and October with high number of retweets.
12	Securitywatch, dangoodin001 and mathewjschwartz tweeted and retweeted about apps on mobile device during July, August and November with high number of retweets.
13	Fireeye tweeted and retweeted about information security during July, August and October with high number of retweets.
14	Dangoodin001 tweeted about thank and questions during July, November and December with high number of retweets.
15	Darkread and dstrom tweeted about cloud data breach and security during July, October, and November with high number of retweets.
16	Rsasecur tweeted about blog, sharing security and RSA summit event during September and December with high number of retweets.
17	Cyberwar, darkread, and mathewjschwartz tweeted about the new bug Shellshock and potential attack during August to October with high number of retweets.
18	Rsasecur tweeted about cyber security threat detection in RSA during October and November with high number of retweets.
19	Mcafeelab tweeted about malware attack and new phishing threat report during July and December with high number of retweets.
20	Varoni tweeted about information security and password hack during July and August with high number of retweets.

After zapping the unwanted words, SMERT output 20 topics as in *Table 2*. Topics 7 is about the Sony Hack and Topic 17 is about Shellshock. In this example, the parameters and importance scores are assigned as the same values from the previous example. Then using the TWIST monitoring method, Figure 5 shows the predicted scores with a breakdown by topic for the Shellshock and the Sony Hack example.

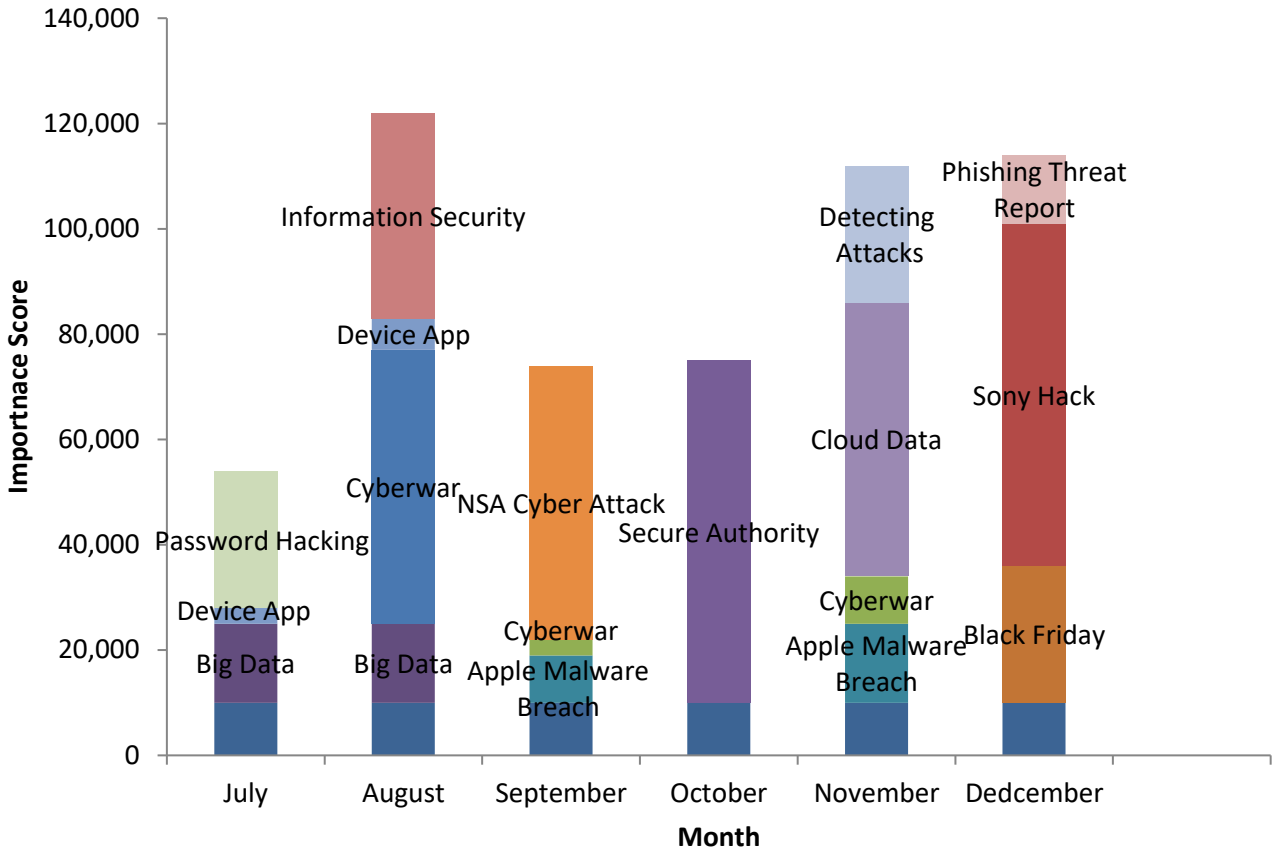

Figure 5: Shellshock and the Sony Hack Predicted Scores Breakdown by Topics.

5.2.3 Discussion

In Figure 4, July, August and December have a higher total retweet number than other months. The reason that July and August have a higher total retweet number may be from discussion of the Unix Bash Shell (shell shock) on social media. Shellshock did not receive its name until September however. The reason that December has a higher total retweet number may be because the Sony Hack happened in late November. Although it aroused active discussions on social media in November, the total retweet does not react to this accident very sensitively due to the late time of the month. But the total retweet number of December behaves as we would expect.

Figure 5 shows that the predicted scores from the TWIST monitoring method are more sensitive to the social events than the real total retweet number. There is a peak in the August predicted score and the breakdown of topics for August has shown that the social media users have observed a new information security issue. As discussed in the last paragraph, the bug was just not named as Shellshock yet. The shellshock bug being referred to consistently over the time frame means it does not show up as clearly using this method however.

6 CONCLUSIONS AND FUTURE RESEARCH

In this article, we proposed the TWitter Importance Score Topic (TWIST) method to aid in monitoring and sensemaking. We illustrated the application of TWIST to two data sets related to cyber security. In the first case, the TWIST method explained at a glance the large uptick of cyber intrusions during the month of April 2014. The chart clearly shows that the uptick corresponded to the heartbleed vulnerability.

Similarly, for case study 2, the Shellshock vulnerability is also readily apparent. Another relevant occurrence (the Sony Hack) is clearly visible. In both case studies, the so-called "importance score" correlated highly with the numbers of retweets providing confirmation that the TWIST method generates relevant information.

TWIST leverages the explanatory capability of Twitter while simplifying the outputs into a single screen. This can potentially save reading streams from tens or hundreds of content generators. We have explored the application of TWIST in the cyber security domain. Yet, a number of topics remain for future study.

First, TWIST can be compared with alternatives including methods based on more repeatable estimation procedures than collapsed Gibbs sampling. Second, TWIST can be made more automatic. Instead of including manually generated labels in Step 9, auto generation can be investigated. Also, TWIST based on the simpler LDA may be sufficient without human high-level data generation and the complications of SMERT. Third, the validation of TWIST could be explored with simulated numerical examples and the related statistical properties can be evaluated. Finally, domains outside of cyber security can be studied. These might relate, e.g., to sentiment analysis and the interests of populations relating to marketing or military conflicts.

REFERENCES

Allen, T. T., H. Xiong, and A. Afful-Dadzie 2015. "A directed topic model applied to call center improvement." *Applied Stochastic Models in Business and Industry* (preprint online).

Allen, T.T., S.M. Vinson, A. Raqab, and Y. Alam 2013. "Using SMERT to Identify Actionable Topics in Student Feedback." *Integrated Systems Engineering Technical Report 2013*.

Britt, R. 2015. *How you and 'The Rock' Turned His Movie Around*. Retrieved from http://www.marketwatch.com/story/how-hollywood-is-using-social-media-to-tell-if-a-movie-will-be-a-hit-Accessed June 19, 2015.

Shah, D. and T. Zaman 2010. "Community detection in networks: The leader-follower algorithm". *arXiv preprint arXiv:1011.0774*.

Yang, J. and S. Counts 2010. "Predicting the Speed, Scale, and Range of Information Diffusion in Twitter". *ICWSM*, *10*: 355-358.

Zaman, T., E. B. Fox, andE. T.Bradlow 2014. "A Bayesian approach for predicting the popularity of tweets." *The Annals of Applied Statistics*, *8*: 1583-1611.

Zaman, T. R., R. Herbrich, J. Van Gael, and D.Stern 2010. "Predicting information spreading in twitter." In *Workshop on computational social science and the wisdom of crowds, nips* 104: 17599-601.