CYBER BUZZ: EXAMINING VIRALITY CHARACTERISTICS of CYBERSECURITY CONTENT in SOCIAL NETWORKS

Thomas Shields, Hannah Li, Peter Lebedev
Georgia Tech Research Institute

Josiah Dykstra
National Security Agency

The Internet is a rich environment for information to spread rapidly and widely. The ability of cybersecurity content to achieve virality in social networks can be useful for measuring security awareness, policy adoption, or cybersecurity literacy. It may also reveal new and emerging cybersecurity events. Virality in online social networks can be characterized and measured many ways and have different causes. Leveraging existing research in social network virality measurements, we calculate and analyze virality measurements and correlations on an anonymized Reddit dataset, examining overall trends and characteristics of individual cybersecurity forums (subreddits). We reproduce content-based virality prediction algorithms and assess their performance, then introduce additional features beyond post title, including time of day, to improve prediction accuracy to ~71% for each of the virality scores. We examine the intersection of the virality facets to reveal correlations about the content and times when cybersecurity content is most viral.

INTRODUCTION

Every day, billions of messages are shared online across email, social media, and online forums. The ability for any message or story to diffuse rapidly across the Internet stems from global interconnectedness and the ease of information sharing. Among them, cybersecurity content – including breaking news, products and service release, and security vulnerabilities – is spread among security professionals and the general public.

There are many uses for understanding what goes viral. Viral cybersecurity content can be important for raising security awareness or countering widespread cybersecurity threats. The ability to measure cybersecurity content's various forms of virality can be important for characterizing cybersecurity awareness, informing policy, or assessing cybersecurity risk. Insights about the content or other characteristics of viral messages could aid the early detection or optimize release of information to achieve stronger cybersecurity outcomes.

The topic of virality within social media has been widely studied from the perspectives of psychology, sociology, advertising, and most recently, misinformation modeling (Brown et al., 2007; Yang and Counts, 2010; Del Vicario et al., 2016; Kim, 2018). There is sparse research about the virality of cybersecurity information. One study by Horawalavithana et al. (2019) considered mentions of security vulnerabilities and found that "while more security vulnerabilities are discussed on Twitter, relevant conversations go viral earlier on Reddit."

Reddit describes itself as "the front page of the internet" and is the sixth most popular website in the United States, ahead of Twitter and LinkedIn (Alexa, 2020). Reddit is organized into individual communities called "subreddits," each of which is self-regulating and devoted to a particular topic. There are more than sixty-one subreddits devoted to cybersecurity. We chose Reddit as one example of a social network, and for the active cybersecurity community known to use Reddit. Reddit data is commonly used for scientific research. For example, Lakkaraju, McAuley, and Leskovec (2013) explored how various factors affected the popularity of images on Reddit, including the title, community, and time of posting.

Other work has investigated the identification of security content in social media, without considering its virality. Mittal et al. (2010) presented a system to discover and analyze cybersecurity threats and vulnerabilities. Sabottke, Suciu, and Dumitras (2015) went further to detect real-world exploits using social media. This approach considered the volume of Twitter messages about vulnerabilities, but did not consider virality characteristics. In 2017, Sapienza et al. presented a method to detect cyber threats based on activity of cybersecurity experts. Sapienza et al. (2018) also mined Twitter text of security experts to detect cyber threats. Bose et al. (2019) applied unsupervised machine learning to detect novel and developing cyber events on Twitter, additionally adding weighted keywords and influential users. These results highlight a gap in understanding and predicting what and when cybersecurity information becomes viral.

Guerini, Strapparava, and Ozbal (2011) (hereafter GSO) proposed a basic framework for characterizing generic text virality in social networks. They introduced four metrics of virality that capture its many facets: appreciation, buzz and spreading, raising-discussion, and controversiality. They achieved moderate success predicting each metric using a Digg dataset from the submission title alone.

Since there is no universal definition or consensus on what virality is or what "going viral" means, we use the four measurable, definable characteristics from GSO.

In this work, we implement the GSO approach against data from Reddit, explore additional features, and discover that timing features improve prediction accuracy. We examine the intersection of the virality facets and find varying correlations between them within Reddit data, particularly across different subreddits. In particular, we compare the differences in the four virality metrics between general news, cybersecurity content and overall Reddit.

METHOD

Dataset

We acquired 181 million Reddit submissions (and their associated 540 million comments) across all subreddits from April–May, 2019 from the open source Pushshift Reddit data dumps (Baumgartner et al., 2020).

Virality Scoring

We computed the four virality metrics from GSO on all 181 million Reddit posts according to the procedures below. This produced a new dataset for analysis consisting of 181 million anonymous records with title, timestamp, and virality metrics. Figure 1 shows one example post and its virality metrics.

Appreciation (A). GSO computed appreciation based on the number of "diggs" a submission received. Diggs are produced by a user's vote on a submission, so to compute Appreciation of a Reddit submission, we used the submission's score, which is equal to the difference between upvotes and downvotes from users.

Buzz & Spreading (Buzz). GSO computed *Buzz* from the number of unique users (*NUC*) commenting on a submission. We compute it in exactly the same way.

Raising-discussion (RD). GSO computed $RD = (NC_L / NC_T) * NUC$, where NC_L is the number of low-level comments (i.e., replies to other comments) and NC_T is the number of top-level comments directly on the submission. We compute it the same way.

Controversiality (C). GSO computed $C = min(A,B)/max(A,B)$ where A is the highest number of positive votes on a comment on the submission and B is the highest number of negative votes on a comment on the submission. Since the Reddit data did not include the breakdown between positive and negative votes on a comment, we set A to the highest score on a comment and B to the lowest score.

Viral Thresholds. GSO dubbed a submission "viral" according in a given metric if the score for that metric exceeds a particular threshold. We use the same thresholds: A >= 100, Buzz >= 100, RD >= 50, and C > .9.

Figure 1. Example Reddit post with Appreciation=237, Buzz & Spreading = 131, Raising-discussion = 501.33, Controversiality = 0.81. https://reddit.com/r/btc/comments/bly4e5/binance_hacked/

Data Subset Identification

To characterize how cybersecurity news and discussion differs from the rest of Reddit, we manually identified names of cybersecurity subreddits (Table 9) to analyze each virality metric on. We also identified lists of subreddits associated general news (Table 10), as well as the most recent list of Reddit's default subreddits (default-subreddits, 2017) to serve as comparison points. We performed the following procedures on each of these subsets, as well as the entire set as a whole. The general news subreddits encompassed 2.6 million submissions, the default subreddits encompassed 16.7 million submissions, and the cybersecurity subreddits encompassed 99,692 submissions.

Virality Correlation Analysis

In order to establish a baseline characterization of the Reddit platform as a whole, we examined the correlation between the four virality metrics. We computed the Pearson correlation coefficient between each pair of metrics.

We examined the "overlap" between each pair of metrics by computing the percentage of submissions with a viral score in one metric also having a viral score in the second metric. E.g., the overlap between Appreciated stories and Buzzed stories is the percentage of stories with an Appreciation score > 100 that also have a Buzzed score > 100.

We also compute correlation and overlap for each of the subsets identified: cybersecurity, general news, and default subreddits.

Virality Temporal Analysis

To further understand how the virality metrics characterize Reddit content, we examined the temporal nature of each metric over all of Reddit, cybersecurity posts, general news posts, and default subreddit posts. Specifically, we examined how these scores change based on the submission time over the course of a day by plotting the ultimate summed scores of submissions posted during each hour. To account for changes in submission volume over the course of a day, we examine the volume-normalized and score-normalized relative score per-hour for each virality metric.

Predicting Virality

In order to understand the effects and consequences of cybersecurity news and information, we developed a model to predict virality metric scores for new Reddit posts. We leverage the same prediction technique used by GSO, but also considered the time of day the post occurred.

To prepare the data for the prediction task, we select an equal number (40,000) of viral and non-viral submissions for each metric, resulting in four datasets of 80,000 submissions each. The datasets include only the data necessary for the prediction task: title, timestamp, and virality metric score – and thus are anonymized and do not include Reddit user information.

Prediction Calculation

Guerini et. al. (2011) encoded the submission titles to feature vectors by using the popular bag-of-words approach. A shortcoming of this approach is its inability to support submissions whose titles contain words not previously seen in the training data. Because of this, we leveraged Fasttext (fastText, n.d.) to produce an *embedding* of the title to use as

input to the classifier. We used the open-sourced ag news fastText model ("Supervised models · fastText", n.d.).

We trained four Support Vector Machines (SVMs) for the same classification tasks as GSO: one SVM per virality metric. The input was the embedding of the submission title from fastText. The binary label was whether the virality metric exceeded the corresponding threshold. We used scikit-learn's SVM classifier, SVC (scikit-learn developers, 2019). Next, we extended our feature vector by including two pattern-of-life features: hour of the day and day-of-the-week. These features were scaled from the 0-24 range and 1-7 range respectively to the -1 to 1 range in order to have the same scale as the fastText features. We trained four new SVMs, using the same parameters as before, on the data with these two new features.

For all SVMs, we used the parameters in Table 11 and trained and evaluated using the standard train-test-split paradigm with a test size of 30%. Additionally, we further evaluated the SVMs performance against a randomly selected, balanced-label set of posts from cybersecurity subreddits and separately against a random, balanced set from general news subreddits.

RESULTS

Correlation Analysis

Our analysis shows that the Buzz and Raising-discussion scores, Appreciation and Buzz scores, and Appreciation and Raising-discussion scores exhibit strong correlation across all of Reddit, while the other score pairings had little-to-no correlation.

	All Reddit	General News	Cybersecurity	Default Subreddits
A / Buzz	0.3028	0.5200	0.4109	0.2828
A / RD	0.2130	0.4478	0.4140	0.1966
A / C	0.0226	0.0258	0.0346	0.0181
Buzz / RD	0.6830	0.9035	0.8184	0.7113
Buzz / C	0.0727	0.0824	0.2715	0.0337
RD / C	0.0498	0.0634	0.1826	0.0223

Table 1: Correlation between metrics

The overlap analysis (Table 2) reveals Buzz is, in general, a strong predictor of Appreciation and even stronger predictor of Raising-discussion across all of Reddit.

	A	Buzz	RD	C
A	-	2.83%	10.32%	0.31%
Buzz	87.47%	-	96.91%	0.44%
RD	52.19%	15.84%	-	1.37%
C	11.55%	0.53%	9.97%	-

Table 2: Averaged metric overlap for Reddit overall

However, the overlap analysis finds that in news-focused subreddits (Table 3), Buzz is a near-perfect predictor (99%) of Appreciation and Raising Discussion. That is, a news story submitted to these subreddits that achieves at least one hundred unique commenters (i.e., is Buzzed) will nearly always receive at least a net score of 100 (i.e., will be Appreciated) and have a high low-level comment to high-level comment ratio (i.e., Raises Discussion). Across all of Reddit the Buzz-Appreciation overlap was 87.5% and the Buzz-Raising Discussion overlap 96.9%. Across the default subreddits (Table 4), they were 91% and 97% respectively.

	A	Buzz	RD	C
A	-	7.10%	17.50%	0.50%
Buzz	99.41%	-	99.98%	0.53%
RD	73.00%	30.12%	-	2.94%
C	20.04%	1.52%	28.17%	-

Table 3: Averaged metric overlap in news subreddits

	A	Buzz	RD	C
A	-	7.61%	13.56%	0.31%
Buzz	91.66%	-	97.37%	0.32%
RD	73.68%	43.94%	-	1.39%
C	9.78%	0.85%	8.05%	-

Table 4: Averaged metric overlap in default subreddits

Additionally, the inverse overlap was also higher than normal. Appreciated stories in news subreddits were Buzzed 7% of the time, compared with only 2% overall and 7% in default subreddits. Appreciated stories in news subreddits overlapped with Raised Discussion 17% of the time, compared with 10% overall and 13% in default subreddits.

Our results show that stories in cybersecurity subreddits (Table 5) did *not* reflect the trends of general news stories, with overlap analysis placing cybersecurity subreddit stories in line with the average overall overlap. Only the Buzz-Raising Discussion overlap differed significantly, matching the general news value at 100%.

	A	Buzz	RD	C
A	-	0.90%	14.16%	0.45%
Buzz	88.00%	-	100.0%	0.00%
RD	50.65%	3.66%	-	3.22%
C	6.67%	0.00%	13.33%	-

Table 5: Averaged metric overlap in cybersecurity subreddits

Temporal Analysis

Our results find that all four virality scores still have temporally varying behavior, even after normalized against submission volume. Figure 2 shows the normalized submission volume by-hour for all four categories.

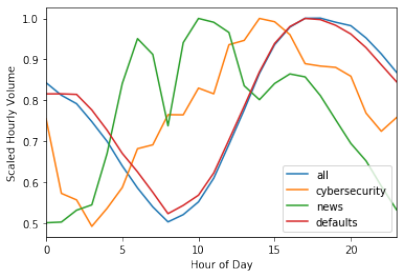

Figure 2: Scaled Hourly Submission Volume by Category

We use this normalized submission volume to normalize the summed hourly score for each virality metric for each category; shown in Figures 3-6.

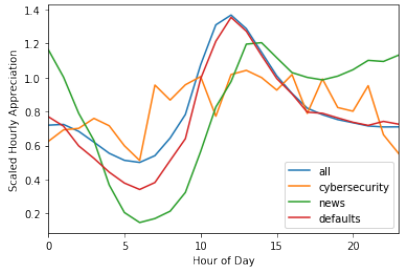

Figure 3: Scaled Hourly Appreciation Scores by Category

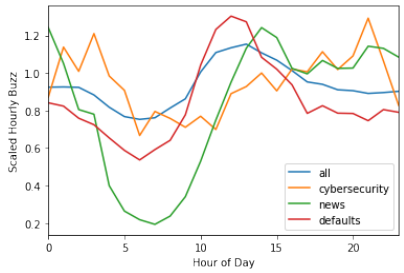

Figure 4: Scaled Hourly Buzz Scores by Category

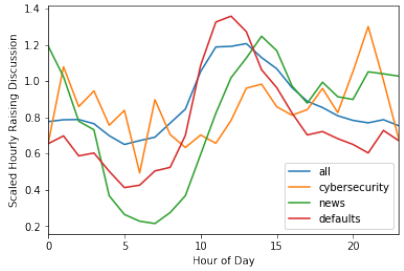

Figure 5: Scaled Hourly Raising Discussion Scores by Category

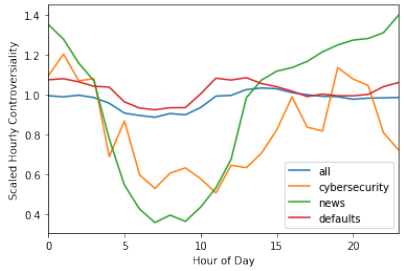

Figure 6: Scaled Hourly Controversiality Scores by Category

These figures show that all four virality metrics have temporal features across all four categories, except for controversiality, where only the news and cybersecurity categories exhibit temporally varying behavior. News posts exhibit the strongest, most consistent temporal behavior across all four categories, and the temporal behavior matches closely the hourly volume behavior. Cybersecurity posts follow general news most closely, but are less sensitive to the early hours than general news data.

Prediction

Content-only Prediction Performance. The average accuracy of the four content-only features virality metric SVMs was 56.75%.

Metric	Precision	Recall	F1-Score	Accuracy
A	0.58	0.56	0.56	0.56
Buzz	0.60	0.57	0.58	0.57
RD	0.62	0.58	0.58	0.58
C	0.56	0.56	0.56	0.56

Table 6: Content-only SVM performance

Content+Time Prediction Performance. The average accuracy of the four content and time features virality metric SVMs was 71.25%.

Metric	Precision	Recall	F1-Score	Accuracy
A	0.73	0.73	0.73	0.73
Buzz	0.85	0.71	0.73	0.71
RD	0.77	0.71	0.72	0.71
C	0.85	0.70	0.73	0.70

Table 7: Content+Time SVM performance

Cybersecurity Prediction Performance. The content+time SVM performed significantly more poorly when evaluated only on submissions from the cybersecurity subreddits, barely outperforming random guess with an average accuracy of 50.75%. This performance was consistent with performance on news related content in general, which averaged 50% accuracy.

Metric	Precision	Recall	F1-Score	Accuracy
A	0.51	0.51	0.49	0.51
Buzz	0.52	0.52	0.52	0.52
RD	0.51	0.51	0.50	0.51
C	0.49	0.49	0.49	0.49

Table 8: Content+Time SVM performance – cybersecurity posts

DISCUSSION

Virality comes in many flavors. Our analysis of the popular content aggregation site Reddit reveals virality analysis and prediction cannot be solved with a one-size-fits-all solution, because virality characteristics vary wildly across subreddits.

Further, our results suggest that while GSO claimed that content drives virality may hold true for the Digg network, it is only a small part of the story on Reddit, with time-based features equally if not more important in the prediction of virality scores.

Our temporal analysis shows that general news and cybersecurity discussions begin earlier in the day (Figure 2). It's not clear whether there is a single explanation for this or independent explanations. Further analysis could investigate the time between when the post was submitted to Reddit (which is the time used in our analysis) and the average time of the post's content, which could reveal possible explanations of this phenomenon.

Our temporal analysis further shows that cybersecurity posts clearly have a unique temporal footprint for each virality metric. Further analysis could explore what types of cybersecurity content are posted at the high and low scoring hours to discover whether cybersecurity events, news, or discussions are affecting virality metrics.

The correlation and overlap analysis of the virality metrics reveals general Reddit dynamics in addition to cybersecurity

specific ones. In general, regardless of topic or subreddit, Buzz is an extremely strong predictor of two other virality measures (Appreciation and Raising Discussion). The increased strength of this association in news content suggests consumers of news value content (via Appreciation) when it warrants discussion. The relative lack of increase in this association for cybersecurity news has implications for cybersecurity policy and awareness, such as the quality of cybersecurity content posted or level of cybersecurity awareness amongst the consumers.

However, the fact that stories in cybersecurity subreddits did have higher Buzz-Raising Discussion overlap than average content suggests that cybersecurity stories generate richer, more involved conversations. Further research could investigate the nature of these comment threads.

The differences in correlation and overlap of virality metrics in cybersecurity content from typical news content suggests cybersecurity content has unique characteristics warranting further study. The inferior performance of the prediction models on cybersecurity content seems to confirm this.

In the future, we plan to incorporate characteristics of influential users. GSO explicitly argued that virality was a phenomenon of the content being spread, rather than to the influencers who spread it. Yang and Counts (2010) found that the rate with which a user is mentioned is a strong predictor for information diffusion. We also plan to improve virality prediction with a larger training set and more robust classifiers, expanding the feature set to include URLs and other attributes.

While we did not attempt to distinguish between misinformation and real information for this work, we plan to explore that in the future, perhaps leveraging our temporal analysis to aid in discovering misinformation.

TABLES

Table 9. List of cybersecurity subreddits

Androidhacking	darknetdiaries	MalwareAnalytics
blackhat	databreach	Malwarebytes
Codexploitcyber	DataBreaches	4chanExploitables
computerforensics	digitalmunition	BlackHatExploits
ComputerSecurity	hacking	ExploitDev
computerviruses	Hacking_Tutorials	SocialEngineering
cryptosecurity	HackWareNews	viruses
cyber	HowToHack	networkingsecurity
CyberForensics	Malware	ethicalhacking
cyberlaws	netsec	GreyHatHacking
cybersecurity	netsecstudents	hacking101
cyber_security	Ransomware	Hacking_Tricks
CyberSecurityFans	RictaNews	KaliLinux_Hacking
CyberSecurityGroup	threatconnect	Python_AND_Hacking
Cyber_Security_News	websecurityresearch	pythonhacking
CyberSleuth	HackingNews	romhacking
CybSecTIA	Infosec	hardwarehacking
darknet	InternetHacks	Hacking_Cracking
security	ironhack	WiFihacking
	ITComputerSecurity	WiiHacking
	LatestHackingNew	WebApplicationHackin

Table 10. List of general news subreddits:

news	business	politics	UpliftingNews
worldnews	Economics	geopolitics	

Table 11. SVM parameters:

Parameter	**Value**
C (regularization)	1.0
Kernel	Linear
Degree	3
Gamma (kernel coefficient)	Auto

ACKNOWLEDGEMENTS

This material is based upon work supported by the Defense Technical Information Center (DTIC) under Contract No. FA8075-16-D-0005. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the Defense Technical Information Center (DTIC), Department of Defense, or U.S. Government.

REFERENCES

Alexa. (2020). *Top Sites in the United States*. https://www.alexa.com/topsites/countries/US

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M., & Blackburn, J. (2020). The Pushshift Reddit Dataset. *arXiv preprint* arXiv:2001.08435.

Bose, A., Behzadan, V., Aguirre, C., & Hsu, W. H. (2019, August). A novel approach for detection and ranking of trendy and emerging cyber threat events in Twitter streams. In *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining* (pp. 871-878).

Brown, J., Broderick, J., & Lee, N. (2007). Word of mouth communications within online communities: Conceptualizing the online social network. In *Journal of Interactive Marketing* 21 (3). 2-20.

default-subreddits (2017) Retrieved from https://web.archive.org/web/20170101000339/https://www.reddit.com/

Del Vicario, M., Bessi, A., Zollo, Fl, Petroni, F., Scala, A., Caldarelli, G., Stanley, H.E., & Quattrociocchi, W. (2016). The spreading of misinformation online. In *Proceedings of the National Academy of Sciences 113 (3).,* 554-559.

fastText. (n.d.). Retrieved from https://fasttext.cc/

Supervised models · fastText. (n.d.). Retrieved from https://fasttext.cc/docs/en/supervised-models.html

Guerini, M., Strapparava, C., & Ozbal, G. (2011, July). Exploring text virality in social networks. In *Fifth international AAAI conference on weblogs and social media.*

Horawalavithana, S., Bhattacharjee, A., Liu, R., Choudhury, N., O. Hall, L., & Iamnitchi, A. (2019, October). Mentions of Security Vulnerabilities on Reddit, Twitter and GitHub. In *IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 200-207).

Kim, J.W. (2018) They liked and shared: Effects of social media virality metrics on perceptions of message influence and behavioral intentions. In *Computers in Human Behavior 84*, 153-161.

Lakkaraju, H., McAuley, J., & Leskovec, J. (2013) What's in a name? understanding the interplay between titles, content, and communities in social media. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media.*

Mittal, S., Das, P. K., Mulwad, V., Joshi, A., & Finin, T. (2016, August). Cybertwitter: Using twitter to generate alerts for cybersecurity threats and vulnerabilities. In *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)* (pp. 860-867). IEEE.

Sabottke, C., Suciu, O., & Dumitraș, T. (2015). Vulnerability disclosure in the age of social media: exploiting twitter for predicting real-world exploits. In *24th USENIX Security Symposium* (pp. 1041-1056).

Sapienza, A., Bessi, A., Damodaran, S., Shakarian, P., Lerman, K., & Ferrara, E. (2017, November). Early warnings of cyber threats in online discussions. In *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (pp. 667-674). IEEE.

Sapienza, A., Ernala, S. K., Bessi, A., Lerman, K., & Ferrara, E. (2018, April). Discover: Mining online chatter for emerging cyber threats. In *Companion Proceedings of the The Web Conference 2018* (pp. 983-990).

scikit-learn developers. (2019). sklearn.svm.SVC. Retrieved from https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

Yang, J. & Counts, S. Predicting the Speed, Scale, and Range of Information Diffusion in Twitter (2010). In *Proceedings of the Fourth International Conference on Weblogs and Social Media (ICWSM)*. 355-358.