

Leveraging the Implicit Structure within Social Media for Emergent Rumor Detection

Justin Sampson, Fred Morstatter, Liang Wu, Huan Liu
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
699 S. Mill Ave.
Tempe, AZ 85287-8809

ABSTRACT

The automatic and early detection of rumors is of paramount importance as the spread of information with questionable veracity can have devastating consequences. This became starkly apparent when, in early 2013, a compromised Associated Press account issued a tweet claiming that there had been an explosion at the White House. This tweet resulted in a significant drop for the Dow Jones Industrial Average. Most existing work in rumor detection leverages conversation statistics and propagation patterns, however, such patterns tend to emerge slowly requiring a conversation to have a significant number of interactions in order to become eligible for classification. In this work, we propose a method for classifying conversations within their formative stages as well as improving accuracy within mature conversations through the discovery of implicit linkages between conversation fragments. In our experiments, we show that current state-of-the-art rumor classification methods can leverage implicit links to significantly improve the ability to properly classify emergent conversations when very little conversation data is available. Adopting this technique allows rumor detection methods to continue to provide a high degree of classification accuracy on emergent conversations with as few as a single tweet. This improvement virtually eliminates the delay of conversation growth inherent in current rumor classification methods while significantly increasing the number of conversations considered viable for classification.

1. INTRODUCTION

Social media is an expansive and varied source of social conversation and interaction data. As social media continues to pervade our culture, it has become an attractive platform to facilitate the spread of information to a large group of people. This is an invaluable feature for the communication of important and time-sensitive information, commercial advertising, and the creation and maintenance of communities. Unfortunately, these properties, combined with the potential for automation, allow for the rapid creation and dis-

semination of rumors. Rumors spread through social media have the ability to create wide-reaching confusion, distrust, and chaos. Additionally, they are easily spread through the social network. For example, a rumor spread by a compromised Associated Press account on Twitter resulted in an estimated \$136.5 billion drop on the Dow Jones Industrial Average¹. Unfortunately, while this is an example of a particularly well crafted attack using rumors, this situation is far from isolated. To complicate the problem, repairing damage may not be as simple as refuting the information through a reputable source. In addition, if a rumor gains enough attention to become viral it may also become more difficult to derail. Therefore, it is necessary to identify a rumor as early as possible to provide the best possible chance for countermeasures.

The ability to detect viral rumors in their formative stages is imperative in order to take early action. However, current state-of-the-art detection methods employ a pruning process where conversations with less than a certain threshold of tweets are removed from the dataset. Unfortunately, pruning is a necessary step for several reasons. First, graph propagation methods and time series-based methods cannot be effectively employed on emergent conversations as there is not enough usable network and statistical data generated until several posts have been made. Second, small conversations suffer from data sparsity issues that add noise to otherwise effectively classifiable data. Therefore, a significant problem exists where it is important to perform classification early enough to allow for timely detection while reducing the noise inherent to early conversation data. Fortunately, conversations do not exist in a vacuum. Disjointed conversations, those without explicit links created through reply chains, that share similar topics of discussion may contain important context and propagation data. These can be leveraged in order to provide proper classification of smaller conversations as well as improving the discriminative quality of the features of larger conversations.

We formulate the problem as follows: given a conversation M with m_i messages such that $i \in \{1, \dots, |M|\}$ in chronological order and some classification function, $F(M_t) = \alpha$, where t is a detection deadline and α is the resulting classification accuracy, it is understood that $F(M_1) < F(M_t) < F(M_{|M|})$. Now, suppose we can find some related conversation N with the same properties and $N_i < M_1$. Is it possible to leverage the additional information in N such that $F(N_i + M_t) > F(M_t)$? Additionally, it is important to know at what point,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CIKM'16, October 24-28, 2016, Indianapolis, IN, USA

© 2016 ACM. ISBN 978-1-4503-4073-1/16/10...\$15.00

DOI: <http://dx.doi.org/10.1145/2983323.2983697>

¹<http://www.cnbc.com/id/100646197>

if any, is $F(M_{|M|}) > F(N_i + M_t)$ as intuitively the approximation $N_i + M_t$ may not perform as well as M_M . In line with explicit links, those that are explicitly created as a function of conversation responses, we define implicit links as links that join conversation structures based on some similarity but are not inherent to the social media structure. Implicit links can be used to reduce the noise inherent to emergent conversation structure by employing additional information from related conversations which can significantly improve the rate at which new rumor conversations are detectable. Related conversation statistics can also be used to improve discriminative features within established rumors for additional classification accuracy. However, this is also a challenging proposition as it is not trivial to decide what data should be connected, what augmenting data source should be applied, what steps should be taken to reduce cross contamination between classes, and what types of links provide the best discriminating features. The main contributions of this paper are:

- We design a conversation-centric approach to the gathering of rumor data along with curated ground truth;
- We identify and analyze implied links, their distributions, and their impact upon detecting rumors within their emergent stages; and
- We propose two methods for the discovery of implicit links including: Hashtag Linkage and Web Linkage.

2. DATA SETS AND DATA COLLECTION

We study the problem of emergent rumor detection using two credibility-based datasets. The first was initially created by Castillo et al. [3] in order to discover features that could be used to classify the credibility of news and has since become a benchmark dataset for rumor detection. This dataset is made up of 280 separate groups of tweets created by using a set of keywords tailored to gather tweets related to individual news events. Keyword-based search methods result in a dataset which has characteristics such as failing to incorporate either conversation structure or important contextual information that would be found within reply tweets, an exaggerated conversation lifecycle of 29.57 hours, and low content variability within discovered tweets. These characteristics are optimal for studying the feasibility of various methods of rumor detection upon heavily duplicated news data. Though the tweets in each group are loosely associated, to avoid confusion between datasets we will refer to all groups of tweets associated with a specific label as a conversation.

Keyword-based news datasets provide a basis for credibility assessment based on viral data but are unable to provide valuable insight useful for credibility assessment at the conversation level. Studying the effects of conversation structure, reply content, and user engagement is necessary for early detection of rumors as the lag time inherent to news reporting can slow detection speed. To combat this, we propose a new dataset that automatically captures conversation trees related to rumors. Snopes² provides a professional rumor repository and verification services. While professional annotation is not applicable to real-time rumor detection applications due to the significant investigation time required

²<http://www.snopes.com>

Table 1: Statistics of the Datasets

Statistic	Castillo	Snopes
Users	48,224	39,535
Tweets	53,147	90,779
Conversations	280	19,397
Rumors	133	9,397
Non-Rumors	147	10,000
Avg. Conversation Time	29.57 hrs	18.33 hrs
Avg. Duplication	54.3%	9.5%
Conversation Length ≥ 10	76.7%	4.5%

for each rumor, known and suspected rumors can be leveraged to discover active rumor-based conversation threads. Often on Twitter, when it is difficult to verify the truthfulness of a conversation, users will mention Snopes in an effort to ask for help, implore other users to verify their statements, or provide known facts related to a rumor. We use this Snopes mentioning trait to gather rumor conversations from the twitter search stream using keywords: snopes, #snopes, snopes.com, and @snopes. Each Snopes mention acts as a ‘seed’ tweet that is used to reconstruct the conversation structure using Twitter Search API functionality. Dataset statistics can be seen in Table 1.

3. DETECTING EMERGENT RUMORS

For each given tweet in a conversation, we extract a feature vector, \hat{t} , using the set of features proposed by Ma et al [9]. Conversation feature vectors, used for classification, can then be built using z -score normalized average of the tweet features with links within a given conversation,

$$\theta_i = \frac{t_i - \mu_{t_i}}{\sigma_{t_i}}, \quad (1)$$

where θ_i is the i th feature in the aggregate conversation feature set θ , t_i is the average of the i th feature for all conversations, μ_{t_i} is the mean, and σ_{t_i} is the standard deviation. Therefore, we can group all conversation vectors into the set, R . In order to improve discriminative quality, we create several types of implicit links which we define as underlying contextual clues, or similarities, between otherwise unconnected conversations. For example, if conversation A is discussing event C and we can determine that conversation B is also discussing event C then we can claim that there is an implicit link between these conversations. This is beneficial as the additional user and content information help to reduce noise that is caused by outlier discussions that have too few connected responses. To accomplish this, we need an appropriate transformation that can accommodate various types of implicit links. We express this as follows:

$$R' = \{\theta \in R \mid \frac{1}{N}(\theta + \sum_{\theta \in R} \mathbb{1}_{\theta}(I)\theta) > 0\}, \quad (2)$$

where I is any defined implicit link discovery formula, N is the number of discovered links, and $\mathbb{1}_{\theta}$ is an indicator function defining whether the implicit link I exists in θ . Additional constraints to the set provide filtering that reduces those conversations without any sloping feature values. In

effect, this transformation discards any conversations which will not show change over time and are also unable to be linked to other gathered conversations.

3.1 Implicit Links

Implicit links are any form of similarity that can be discovered between otherwise unlinked conversations.

3.1.1 Hashtag Linkage

The use of hashtags has become very prevalent within social media communication. Since hashtags are a general abstraction, it is simple to find links within tweets. As a result, it is difficult to determine the extent to which an implicit link formed using hashtags should impact the extracted conversation features.

3.1.2 Web Linkage

Implicit links formed with web links are rare but with much stronger evidence for similarity. In this case, a conversation link is formed when two unlinked conversations use the same web link. This is a very strict requirement which enforces similarity down to the individual page level of a web domain. Intuitively, this requirement could be loosened to allow links to form between conversations who provide links to the same domain, or link to similar web content types. However, we adhere to a strict web linking policy in order to quantify the impact of the extreme accuracy of exact web links at the expense of additional general link discovery.

4. RESULTS

In order to verify the efficacy of implicit linking techniques, we examine the problem through three distinct experiments. The first experiment involves performing classification using the proposed linking methods on both datasets without performing pruning. This allows us to study how linking methods relate to standard classification methods. The second experiment involves the use of a detection deadline to determine the implications of the addition of external data and at what point, if any, the algorithm will obtain optimal performance. The third experiment involves classification of pruned datasets with several classifiers to understand the relation between linking methods, separability, and classification performance.

4.1 Zero Pruning Classification

In contrast to standard rumor detection, a prime advantage of implicit linking to support rumor detection is the ability to classify datasets without removing data that would otherwise contribute an undue amount of noise. To show this, we perform classification using a linear Support Vector Machine (SVM) and 10-fold cross validation. The experiment was run ten times for each method and the results were averaged. We report accuracy, precision, recall, and F1 score for each method. First, all 280 conversations were classified within the Castillo dataset. As can be seen in Figure 1, standard classification fails to capture discriminating patterns when a pruning strategy is not employed. However, with the addition of implicit links our classification measurements rise using all three methods. In particular, hashtag linking shows an averaged 30% improvement in all metrics when compared to unlinked performance. Web linking was also able to achieve a small improvement in discriminative performance. However, this is acceptable as the extremely

duplicated nature of the Castillo news dataset contains very low variability within conversations, resulting in very few web link matches. The same procedure was repeated for all 19,397 conversations within the Snopes dataset. Following the same pattern as the Castillo data, unpruned classification again fails when implicit linking is not used to mitigate the problem. In addition, the larger dataset results in a severe drop in recall. Fortunately, both linking methods provide ample improvement in classification performance.

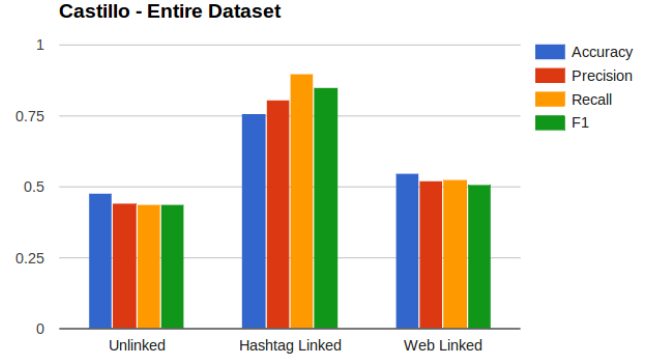


Figure 1: Classification on the Castillo dataset without the removal of small conversations results in near-random classification performance. Hashtag linking performs well while Web linking improves performance but is hampered by difficulty in finding links within the dataset.

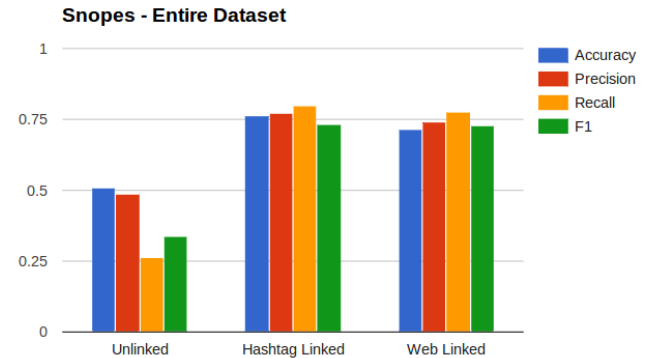


Figure 2: Similarly to the Castillo dataset, classification on the Snopes dataset performs poorly pruning. All linking methods perform well due to the wide range of variability with real conversation data.

4.2 Early Detection and Detection Deadline

By the problem formulation given in Section 1.1, the previous experiment was able to show that the addition of conversation data from conversation N into M was able to achieve significantly improved performance. In order to determine the extent of improvement over time, we perform classification with the addition of a detection deadline. In the initial state, the detection deadline, t , begins at hour 0, the first tweet is exposed for classification. Therefore at time $t = 0$ classification is performed using $F(N_t + M_1)$ where N is any implicitly linkable conversation with tweet timestamps $t \leq M_1$ if using implicit linkage and M_1 other-

Table 2: Classification without Linkage

Castillo Dataset 215 Convs 45% Rumors	Accuracy	Precision	Recall	F1
DT	0.5471	0.5253	0.5318	0.5230
DSTS	0.5351	0.5189	0.4964	0.5014
Hybrid SVM	0.5897	0.5852	0.5055	0.5360
RBF	0.6617	0.6633	0.5518	0.5857
Snopes Dataset 848 Convs 55% Rumors				
DT	0.8197	0.8282	0.8514	0.8389
DSTS	0.6995	0.7105	0.7758	0.7403
Hybrid SVM	0.7895	0.7950	0.8387	0.8148
RBF	0.8428	0.8646	0.8492	0.8562

wise. At each successive hour deadline, all tweets in M_t and N_t may be used for classification such that all timestamps are less than t .

Following this procedure, at the initial detection deadline the unlinked Snopes dataset was able to correctly classify about 350 tweets while the implicitly linked data performed twice as well at 700 tweets. Likewise, the Castillo implicitly linked data outperformed unlinked classification but with a smaller margin. However, as the detection deadline increased and additional data was able to be incorporated into N , the number of correctly classifiable instances grows significantly for the initial six hours for both implicitly linked datasets. Additionally, counter to intuition, while the improvement gain becomes negligible as time increases, degradation of performance due to incorporation of estimated conversation data does not take place within the first ten hours. This indicates that as a conversation grows, the performance gain from estimated data is not completely removed while the error introduced becomes negligible. While counter-intuitive, this result can be explained by the classification process using averaged aggregate statistics. This also suggests that, using a principled matching process, data from external conversations can improve classification results in the general case.

4.3 Pruned Classification

Optimally, implicit linking should be applied without performing pruning. However, it is necessary to perform classification using current state-of-the-art methods to allow for comparison with previous work. Classification was performed using several state-of-the-art rumor classification methods including a decision tree (DT), Dynamic Series-Time Structure (DSTS) [9], Hybrid DSTS with SVM classifier, and RBF kernel method. Following previous work, only conversations with 10 or more tweets were used and classification was performed using 10-fold cross validation. First, both datasets were classified without applying any form of implicit linking. This gave us a classification accuracy of 66.2% and 84.3% using the RBF kernel for the Castillo and Snopes datasets respectively as shown in Table 2. For both datasets, the RBF kernel is capable of providing improved discrimination between classes by projecting the feature vectors into a high dimensional space. When compared to the simpler models, it is clear that the gap in performance on unlinked data can be attributed to features that cannot be linearly separated.

With the application of implicit hashtag link discovery, classification accuracy saw an improvement of 13.5% and 9.9% for the Castillo and Snopes datasets respectively. Classification results using pruned hashtag linking is shown in

Table 3: Classification with Hashtag Linkage

Castillo Dataset 231 Convs 45% Rumors	Accuracy	Precision	Recall	F1
DT	0.7785	0.7823	0.7427	0.7607
DSTS	0.7923	0.7994	0.7536	0.7722
Hybrid SVM	0.7967	0.7927	0.7809	0.7814
RBF	0.7876	0.7820	0.7800	0.7765
Snopes Dataset 3,323 Convs 41% Rumors				
DT	0.9416	0.9250	0.9325	0.9284
DSTS	0.9224	0.8993	0.9118	0.9052
Hybrid SVM	0.9311	0.9111	0.9206	0.9156
RBF	0.9260	0.8938	0.9280	0.9103

Table 4: Classification with Web Linkage

Snopes Dataset 5,947 Convs 93% Rumors	Accuracy	Precision	Recall	F1
DT	0.9825	0.9889	0.9924	0.9907
DSTS	0.9650	0.9728	0.9903	0.9814
Hybrid SVM	0.9724	0.9801	0.9906	0.9853
RBF	0.9815	0.9877	0.9926	0.9901

Table 3. At the same time, the volume of classifiable conversations was improved for both datasets. The largest gain was within the significantly pruned Snopes dataset - an improvement of over three times as many classifiable instances. Additionally, the gap in classification performance that was caused by difficulty in separability as seen on the unlinked data vanishes.

In both datasets, the discriminative quality of the features was improved to the point where a high level of accuracy was obtainable with simpler, less time consuming classifiers that do not rely on projecting the features into a higher dimensional space. Web linking was performed on the Snopes dataset, however, a low number of classifiable instances and highly redundant data of the Castillo dataset did not allow for the discovery of a suitable number of links. Classification results for web linking can be seen in Table 4. Web linking improved classification accuracy for several of the underperforming methods by 2-3% while also having a positive impact on recall.

5. RELATED WORK

The study of rumor spread and detection has been approached from several angles. The impact, spread, and perception of rumors within social media was examined in order to propose strategies for increasing credibility within social search [10, 6, 13]. These works provide insights into the spread of rumor ideas which can be leveraged by automatic rumor detection models. Such automated methods can be classified into three general types: Epidemic Models, User and Message-based Classification, and Propagation-enhanced Classification.

5.1 Epidemic Models

The spread of information and rumors within social media share many characteristics in common with epidemics. Several studies have employed epidemic models on network and information flow [1, 11]. Jin et al. proposed an extension which specifically targets rumors [7]. Their model, the SEIZ (susceptible, exposed, infected, skeptic) enhanced epidemic model, accurately captures the spread of both news and rumors through the inclusion of an additional state which helps to model exposure delay caused by preheld views as well

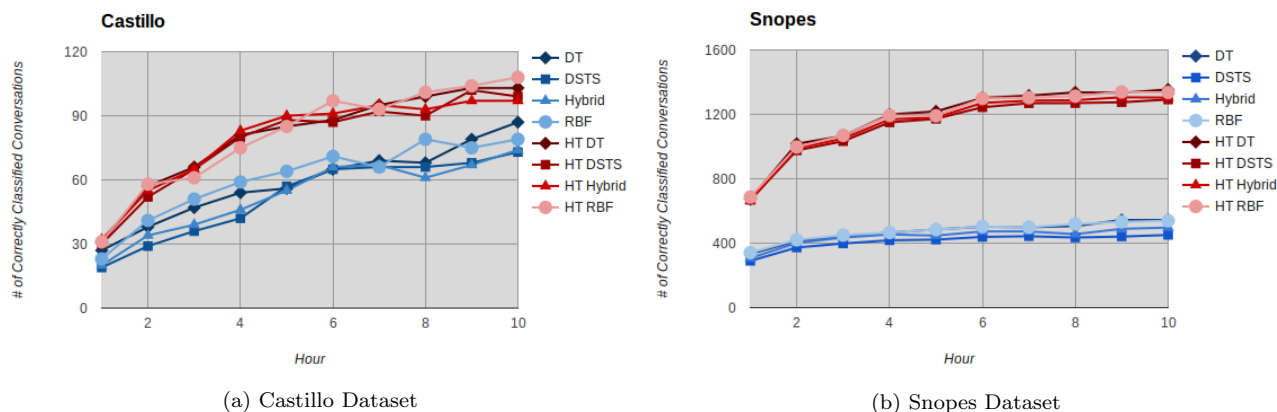


Figure 3: The number of correctly classified rumors as time progresses. With the addition of implicit links, the number of correctly classifiable instances for both rumor and non-rumor conversations see a large increase. This is a combined result of the reduction of noise due to additional tweet statistics as well as an improvement in classification accuracy for conversations that were previously too small for classification.

as other potential factors [7]. Additionally, Takayasu et al. leveraged the SIR model to study the impact of corrective tweets upon rumors during crisis events [15]. While these methods are useful for understanding and generalizing the underlying rumor spreading mechanics, they do not easily allow for prediction or classification which are cornerstone tasks that enable response and preemption.

5.2 User and Message-Based Classification

In order to allow for rapid response, several studies have employed machine learning techniques with the intent of creating classifiers that can, with a high degree of accuracy, decide whether a conversation is credible. Castillo et al. proposed a classifier capable of differentiating news from chatter while also providing initial credibility assessment of conversations [3]. A large body of work has proposed classifier improvements and alternative features to improve the performance of the information credibility task [2, 4, 5, 12, 14, 17]. Other diverging works employ rumor detection through discovery of enquiry content [20]. This has lead to several systems that attempt to automatically detect and verify or debunk rumor conversations [8, 19]. These works provide a strong basis for rumor detection but are severely hampered by datasets that often contain only a few hundred labeled rumor instances. Additionally, all of these previous works perform some form of data pruning method to reduce noise within their data and provide an improvement in classification accuracy. While this usually a reasonable tradeoff when early detection of rumors is not a factor, the practice has become prevalent within works specifically designed to provide early results. This leads to misleading results without real applicability. Fortunately, our work provides a major stepping stone for rumor detection methods which will allow most recent work to function without the need to prune data by leveraging related data.

5.3 Propagation-Enhanced Classification

More recently, conversation propagation statistics have been shown to be an effective source of information in order to detect rumors. Wu et al. [16] employed a random walk kernel in order to measure the similarity in the underlying conversation networks. This form of structural information

was shown to be nearly as effective as the combination of all other classification features. In conjunction with additional features this form of detection achieved a 6% gain over other methods. While this reliance on network structure can achieve particularly impressive results, the highest level of accuracy is only seen after the conversation structure has been allowed to develop for more than 24 hours. Other work [18, 9] proposes a timeseries-based method that captures the change of standard statistics as a conversation progresses. While our contribution focuses primarily on early detection, we do propose a modification to these methods which allow rumors to be accurately classified regardless of initial volume of statistics, overcoming one of the few significant problems inherent to timeseries modeling.

6. CONCLUSION AND FUTURE WORK

In this paper, we analyzed how the implicit structure within social media conversations can be leveraged to allow for the detection of emergent rumors. We find that implicit linkage can significantly improve the ability to classify rumors within their formative stages by leveraging information within related conversations. Through the use of implicit links, it was shown that rumor classification can be augmented by historical data in order to achieve state-of-the-art results with computationally cheap classification models. This discovery paves the way for rumor detection methods to progress beyond pruning methods while maintaining a high degree of accuracy on as few as a single tweet.

We discussed the implicit linkage capabilities of hashtags, web links, and content similarity. We discovered that with the addition of implicit structure, not only can rumor detection on a single tweet be significantly improved, but the gain in performance achieved by addition estimated data does not erode performance as the conversation grows. We find that hashtag-based linkage provided the strongest improvement with a gain of up to 30% over unlinked classification on an unpruned dataset. Additionally, we discovered that under-performing classification methods saw improvement from the remaining implicit link types. Further analysis showed that the use of implicit links must be tempered with caution as an over-abundance of external conversation

statistics may harm detection accuracy. Our experimental results suggest that consistent accuracy gains through linking can be obtained by limiting the number of links formed to between five and ten.

This work has many possible extensions. While hashtags were greatly beneficial for the discovery of similar conversations, the hashtag graph structure between conversations revealed usage patterns that may provide additional clues for early rumor detection. Likewise, the affinity of certain types of implicit links to be found within each type of conversation could be harnessed as a detection feature. Finally, there are many other forms of potential implied links such as content similarity through document embedding which could provide new insights into the information veracity problem.

Though not explicitly explored in this work, the creation of a multi-label dataset allows for many exciting avenues of exploration. For example, implicit link distribution may vary by type of rumor as it does for non-rumors. This may allow for the discovery of additional characteristics related to the veracity of a rumor that can be harnessed for implicit link creation as well as improved veracity inference.

7. ACKNOWLEDGEMENTS

This work is sponsored, in part, by Office of Naval Research grants N00014-16-1-2257 and N00014131083.

8. REFERENCES

- [1] L. M. Bettencourt, A. Cintrón-Arias, D. I. Kaiser, and C. Castillo-Chávez. The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, 364:513–536, 2006.
- [2] C. Boididou, S. Papadopoulos, Y. Kompatsiaris, S. Schiffrer, and N. Newman. Challenges of computational verification in social multimedia. In *Proceedings of the companion publication of the 23rd international conference on World wide web companion*, pages 743–748. International World Wide Web Conferences Steering Committee, 2014.
- [3] C. Castillo, M. Mendoza, and B. Poblete. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684. ACM, 2011.
- [4] C. Castillo, M. Mendoza, and B. Poblete. Predicting information credibility in time-sensitive social media. *Internet Research*, 23(5):560–588, 2013.
- [5] A. Gupta and P. Kumaraguru. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st Workshop on Privacy and Security in Online Social Media*, page 2. ACM, 2012.
- [6] M. Indrawan-Santiago, M. Miyabe, A. Nadamoto, and E. Aramaki. How do rumors spread during a crisis? analysis of rumor expansion and disaffirmation on twitter after 3.11 in japan. *International Journal of Web Information Systems*, 10(4):394–412, 2014.
- [7] F. Jin, E. Dougherty, P. Saraf, Y. Cao, and N. Ramakrishnan. Epidemiological modeling of news and rumors on twitter. In *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, page 8. ACM, 2013.
- [8] X. Liu, A. Nourbakhsh, Q. Li, R. Fang, and S. Shah. Real-time rumor debunking on twitter. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pages 1867–1870. ACM, 2015.
- [9] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong. Detect rumors using time series of social context information on microblogging websites. In *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*, pages 1751–1754. ACM, 2015.
- [10] M. R. Morris, S. Counts, A. Roseway, A. Hoff, and J. Schwarz. Tweeting is believing?: understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*, pages 441–450. ACM, 2012.
- [11] M. E. Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.
- [12] V. Qazvinian, E. Rosengren, D. R. Radev, and Q. Mei. Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pages 1589–1599. Association for Computational Linguistics, 2011.
- [13] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, and F. Menczer. Truthy: mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th international conference companion on World wide web*, pages 249–252. ACM, 2011.
- [14] T. Takahashi and N. Igata. Rumor detection on twitter. In *Soft Computing and Intelligent Systems (SCIS) and 13th International Symposium on Advanced Intelligent Systems (ISIS), 2012 Joint 6th International Conference on*, pages 452–457. IEEE, 2012.
- [15] M. Takayasu, K. Sato, Y. Sano, K. Yamada, W. Miura, and H. Takayasu. Rumor diffusion and convergence during the 3.11 earthquake: a twitter case study. *PLoS one*, 10:e0121443–e0121443, 2015.
- [16] K. Wu, S. Yang, and K. Q. Zhu. False rumors detection on sina weibo by propagation structures. In *IEEE International Conference on Data Engineering, ICDE*, 2015.
- [17] F. Yang, Y. Liu, X. Yu, and M. Yang. Automatic detection of rumor on sina weibo. In *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, page 13. ACM, 2012.
- [18] Y. Yang, K. Niu, and Z. He. Exploiting the topology property of social network for rumor detection. In *Computer Science and Software Engineering (JCSSE), 2015 12th International Joint Conference on*, pages 41–46. IEEE, 2015.
- [19] Q. Zhang, S. Zhang, J. Dong, J. Xiong, and X. Cheng. Automatic detection of rumor on social network. In *Natural Language Processing and Chinese Computing*, pages 113–122. Springer, 2015.
- [20] Z. Zhao, P. Resnick, and Q. Mei. Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1395–1405, 2015.