A Bayesian Model for False Information Belief Impact, Optimal Design, and Fake News Containment

Amin Khajehnejad  
3Red Trading Group LLC., University of Michigan Ann Arbor  
Email: aminkh@umich.edu, amin.khajehnejad@3redgroup.com

Shima Hajimirza  
Texas A&M University  
Email: shima.hm@tamu.edu

Abstract—This work is a technical approach to modeling false information nature, design, belief impact and containment in multi-agent networks. We present a Bayesian mathematical model for source information and viewer’s belief, and how the former impacts the latter in a media (network) of broadcasters and viewers. Given the proposed model, we study how a particular information (true or false) can be optimally designed into a report, so that on average it conveys the most amount of the original intended information to the viewers of the network. Consequently, the model allows us to study susceptibility of a particular group of viewers to false information, as a function of statistical metrics of the their prior beliefs (e.g. bias, hesitation, open-mindedness, credibility assessment etc.). In addition, based on the same model we can study false information “containment” strategies imposed by network administrators. Specifically, we study a credibility assessment strategy, where every disseminated report must be within a certain distance of the truth. We study the trade-off between false and true information-belief convergence using this scheme which leads to ways for optimally deciding how truth sensitive an information dissemination network should operate.

I. INTRODUCTION

Adversarial spread of false information based on exploitation of viewer’s naivety, gullibility and bias has been remarked numerously over social media and news networks, most notably in the midst of the most recent US presidential election and Brexit [1]–[5]. The notion is familiarly referred to as fake news. However, manipulatory demonstration and spread of false information is not limited to news, and indisputably exists in other domains such as advertising [6]–[9], financial markets (electronic trading) [10], [11], cyber-security [12], adversarial machine learning [13], [14] and more, often referred to by terms such as misinformation or disinformation. When it comes to studying false information spread, the goals are multifaceted and significantly vague. This is due to various reasons. Firstly, The nature of false information is highly context-dependent and debatable. The notion of fake news for instance is not fully resolved and well-defined for many [15]–[17]. The practice of fake news fabrication and broadcasting can take various forms, from an individual creating and promoting fake social media posts, news websites etc. [18], to manipulative journalism of well-established news organizations doing so with the intention of appealing to a certain group of audience [19]. False news creation and spread is also not necessarily bound to malicious intention, and can be a mere reflection of erroneous, exaggerative or neglectful quoting of absorbed facts, rumors, panic epidemics and similar phenomena. Furthermore, false information spread can affect different systems in different ways, and consequently, administrators decide to handle it in various forms: identification, detection, prediction, regulation, education or even ignorance are common mechanism that are usually adopted to combat false information [20]. Nevertheless, one needs to identify and study the following aspects of false information in their corresponding contexts:

1. Nature. What is false information, how is it originated/designed, and what distinguishes it from the truth?
2. Impact. How does false information or fake news impact beliefs of the users/viewers of a system/network, and how does that in turn affect the dynamics of information flow and system operation?
3. Detection. What methodologies can be used to identify false information accurately and efficiently?
4. Spread (epidemic). In an interactive network where information is propagated in a distributed way by users and outside feeds, what are the chances of epidemic false news spread as a function of network parameters?
5. Containment and mitigation. What mechanisms can prevent false information dissemination, or offset it with truth? At what costs can those strategies be implemented and, what are the expected outcomes and damages?

Mathematical modeling of false information is far more challenging than semantic and social analysis, due to lack of basic fundamentals. Yet there are a handful of recent works that have modeled various aspects of false information nature and spread using technical methods. Most notably, the majority of works have investigated fraudulent accounts and credibility of news in online social networks such as twitter [21]–[26]. A few other works have proposed models for disseminating good information in social networks to mitigate the effect of misinformation, thereby improving reliability [27]–[31]. Misinformation and manipulation in financial limit order markets are probably studied far more in-depth, especially using game-theoretic approaches (see e.g. [32]–[35]), due to much simpler and readily available
numerical representations of information and beliefs. There are recent works that have tried to adopt those approaches in the context of social networks. For example, the study of [30] offers a game-theoretic approach to truth and fake news dissemination based on a mixed-strategy formulation of volunteer’s dilemma in an adhoc social network. However, those studies are still preliminary and immature compared to areas where numerical representation of information is available.

The current work is a technical (yet different) approach to false information design, impact and containment in multi-agent networks. We present a Bayesian mathematical model for information, viewer’s belief, and how the former impacts the latter in a media (network) of broadcasters and viewers. Given the proposed model, we study how a particular information (true or false) can be optimally designed into a report, so that on average it conveys the most amount of the original intended information to the viewers of the network. The model then allows one to study susceptibility of a particular group of viewers to false information, as a function of statistical metrics of the their prior beliefs (e.g., bias, hesitation, openness-mindedness, etc.). In addition, based on the same model we can study false information “containment” strategies imposed by network administrators. Specifically, we study a credibility assessment strategy, where every disseminated report must be within a certain distance of the truth. We study the trade-off between false and true information containment using this scheme which leads to ways for optimally deciding how truth sensitive an information dissemination network should operate.

A. Proposed Model

We consider an information dissemination network (medium) with an information source, a transmitter (reporter), a network administrator and multiple viewers (audience). The schematics of such a network are illustrated in Figure 1.

![Fig. 1: Schematics of an information dissemination network with source, reporter, administrator and viewer.](image)

The transmitter intends to disseminate in the network a piece of information \( x \) from the source. However, he can modify the information content and design a report \( y \), in order for the viewers to accept as much of the source information as possible. There is an authenticity filter in the network that verifies every input report by comparing it with the truth. The report is not permitted to be broadcast if it is too distant from the truth \( x \). Therefore, the reporter’s optimal design must fall in an acceptable neighborhood of the truth which we denote by \( C(x) \). We assume that \( x_s, x_t, \) and \( y \) are all real vectors of length \( n \). For the present, we use the simple euclidean distance as the measure of deviation between facts: 
\[
d(x_1, x_2) = \|x_1 - x_2\|.
\]

Therefore, the authenticity filter has the following mathematical form:
\[
y \in C(x_t) = \{ x : \|x - x_t\| \leq \epsilon \}
\]
where \( \epsilon \) is a positive constant. There are \( N >> 0 \) viewers, and every viewer \( i \) has a prior distribution \( p_i(x) \) about the truth. When a viewer receives a report \( y \), he/she forms a posterior distribution given the prior belief and the observation, and consequently adopts a new belief in a Bayesian manner, which we denote by \( \zeta_i(y) \). The details of the viewers belief impact is described in the next subsection. We make the assumption of the well-informed reporter. This means that the reporter is fully aware of the truth \( x_s \), and also of the network administration policy \( \epsilon \). However, the reporter is not aware of every individual’s prior distribution. Instead, he only knows average statistics such as mean of first and seconds moments of the viewers’ priors, as will be explained the next section. The reporter’s goal is to convey the source’s information to as many viewers as possible. For a single viewer \( i \), information conveyance can be measured by the distance of the original information \( x_s \) and the adopted belief \( \zeta_i(y) \):
\[
\tau_{i,s}(y) = \|x_s - \zeta_i(y)\|
\]
For simplicity of future reference, we call this quantity the “reporter-viewer information convergence”, or in short “convergence”. The smaller \( \tau_{i,s}(y) \), the more successful the reporter has been for impacting that particular viewer. However, the statistical notion of information throughput is obtained by averaging \( \tau_{i,s}(y) \) over all viewers:
\[
\tau_s(y) = \mathbb{E}[\|x_s - \zeta_i(y)\|]
\]
The reporter’s goal is therefore to minimize average convergence \( \tau_s(y) \) for a valid report \( y \) which is permitted to pass through the network, i.e.
\[
\arg\min_y \tau_s(y) \\
\text{s.t. } y \in C(x_t)
\]
B. Bayesian Viewer

We model the viewer of the news (audience) as an agent \( i \) that has a prior belief about the truth at a particular point in time. We denote the probability distribution of this prior with \( p_i(\cdot) \). Upon receiving a piece of report \( y \), the viewer’ instantaneous belief of the truth (his point of view) is the Maximum A Posteriori fact:
\[
\zeta_i(y) = \arg\max_x p_i(x|y),
\]
where the posterior distribution of the true news is obtained from the Bayes rule:

\[ p_i(x|y) \propto p_i(x)p_i(y|x) \quad (6) \]

The likelihood distribution \( p_i(y|x) \) is a function of the uncertainty and credibility of the source of news in the viewer’s mind. This is essentially an estimation of the statistics of truth manipulation (in the mind of the viewer) that an information source performs on the facts before disseminating them. If in the viewer’s mind, the source of the news is very noisy, exaggerative or inaccurate, then the likelihood distribution \( p_i(y|x) \) is very wide and has a large variance. If there is a reporting bias to the journalism which is known to the viewer, then the distribution is nonzero mean. In summary, the viewer blends his prior opinion with what he/she observes, taking into account the presumed credibility of the source (which could in all likelihood be false).

In a perfectly Gaussian world, we can assume that the prior and likelihood are both \( n \)-dimensional Gaussian distributions:

\[ \begin{align*}
    p_i(x) &\sim \mathcal{N}(\mu_i, \Sigma_i), \\
    p_i(x|y) &\sim \mathcal{N}(y, \Sigma_{y,i}),
\end{align*} \quad (7) \]

where \( \Sigma_i \) and \( \Sigma_{y,i} \) are covariance matrices of the viewer’s prior and the conditional likelihood distribution, respectively. Intuitively, \( \Sigma_i \) is a multi-dimensional measure of viewer’s intrinsic hesitation about the state of truth at the time being, while \( \Sigma_{y,i} \) measures suspicion of noisy/false reporting against the news source. The well-informed reporter assumption means that the reporter is not aware of individual viewer characteristics \( \mu_i, \Sigma_i, \Sigma_{y,i} \), but has perfect knowledge of the joint distribution of these quantities. Combining equations (5), (6), (7), we get:

\[ \log p_i(x|y) \propto -(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i) - (x-y) \Sigma_{y,i}^{-1} (x-y) \]

and by maximizing this, we get that the viewers’ post-report belief becomes:

\[ \zeta_i(y) = A_i y + B_i \mu_i \quad (9) \]

where

\[ \begin{align*}
    A_i &= (\Sigma_i^{-1} + \Sigma_{y,i}^{-1})^{-1} \Sigma_{y,i}^{-1}, \\
    B_i &= (\Sigma_i^{-1} + \Sigma_{y,i}^{-1})^{-1} \Sigma_{y,i}^{-1} \mu_i
\end{align*} \quad (10) \]

Therefore, in the world of Gaussian news and beliefs, a viewer adopts a linear combination of his prior belief and the viewed report as his/her new belief.

### C. Optimal Reporter

Combining equations (3), (4) and (9) together, the optimal report \( y^* \) is revealed to be the solution of the following optimization:

\[ \begin{align*}
    \min_{y} &\quad E\|A_i y + B_i \mu_i - x_s\|, \\
    \text{s.t.} &\quad \|y - x_i\| \leq \epsilon.
\end{align*} \quad (11) \]

(11) is a convex optimization. Therefore, using the method of Lagrange multipliers (see e.g., [35]), \( y^* \) is the minimizer of the following utility function for some nonnegative \( \lambda \geq 0 \):

\[ L(y, \lambda) = E\|A_i y + B_i \mu_i - x_s\|^2 + \lambda \|y - x_i\|^2 \quad (12) \]

Taking derivatives with respect to \( y \), we get:

\[ \frac{\partial L(y, \lambda)}{\partial y} = (A_i^2 + \lambda I) y + A_i B_i \mu_i - A_i x_s - \lambda x_i \quad (13) \]

where:

\[ A_i^2 = EA_i^2 A_i, \quad A_i B_i = EA_i^T B_i \mu_i, \quad A_i = EA_i. \quad (14) \]

Finally, by setting (13) equal to zero, the optimal report is obtained as:

\[ y^* = (A_i^2 + \lambda I)^{-1} (\lambda x_i + A_i x_s - A_i B_i \mu_i) \quad (15) \]

where \( \lambda \) is the smallest nonnegative constant that makes \( \|y^* - x_i\| \leq \epsilon \). The \( \lambda \) in (15) is a regularization factor that forces the report to be closer to reality. Note that the well-informed reporter assumption means that \( A_i \), \( A_i B_i \) are computable for the reporter, hence so is the optimal report \( y^* \).

### D. Ergodic Covariances

For further simplicity of the analysis, we consider the case of ergodic covariances, i.e., \( \Sigma_i = \Sigma \) and \( \Sigma_{y,i} = \Sigma_s \) for all viewers, and only mean belief \( \mu_i \) varies. Under this assumption, it is easy to show that

\[ y^* = (A_i^2 + \lambda I)^{-1} (\lambda x_i + A_i x_s - A_i B_i \mu_i) \quad (16) \]

When \( \lambda = 0 \), there is no authenticity requirement for the report, thus the optimal report can be designed solely based on the input source information and the viewers’ belief statistics. Specifically:

\[ y^* = A_i^{-1} x_s - A_i^{-1} B \mu = x_s + \Sigma_s^{-1} (x_s - \mu) \quad (17) \]

In other words, the optimal report is the linear sum of the news source and a skewed version of the average news surprise \( x_s - \mu \), which can also be interpreted as exaggeration. The optimal magnitude and rotation of such exaggeration is determined by the relative ratio of the covariance matrices \( \Sigma \) (hesitation) and \( \Sigma_s \) (source reliability). The more reliable a source (smaller \( \|\Sigma_s\|_s \)) or the more open minded/less certain the viewers (larger \( \|\Sigma_s\|_s \)), more exaggeration is allowed. \( \|X\|_s \) denotes the nuclear norm of a matrix \( X \) which can be a measure of the magnitude of a covariance matrix. However, this also tells us that elements of the report (i.e. different coordinates of \( y \)) can be exaggerated differently, according to the certainty distribution. It should be noted that even in the case of a perfectly truthful source \( x_s = x_i \), the optimal report is exaggerative and not completely honest. A credibility assessment strategy or filtration policy should take that fact into consideration. In other words, enforcing truthful reporting will compensate the performance of an optimal reporter completely based on the truth if \( \|\Sigma_s \Sigma_s^{-1} (x_s - \mu)\| >> \epsilon \). The optimal convergence metric \( \tau_s(y^*) \) for \( \lambda = 0 \) in this case becomes:
\[ \tau_s(y^*) = \mathbb{E}(\mu_t - \mu)^TB^2(\mu_t - \mu) \] (18)

which is independent of the particular piece of news \( x_s \), and only depends on the source credibility and viewers’ belief statistics.

E. Authentication Policy

Assuming that every reporter does the optimal action knowing the specified filter parameter \( \epsilon \), it is up to the network administrator to set a value that promotes reporting based on truthful sources and limits the unauthentic ones. This requires definition of an appropriate truth dissemination utility function. For a given source \( x_s \), policy parameter \( \epsilon \) and optimal report \( y^* \), we denote the optimal convergence with \( c(x_s, \epsilon) \), which can be obtained from the derivations of the previous section as:

\[ c(x_s, \epsilon) = \mathbb{E}_i [Ay^*_i + B\mu_t - x_s] \] (19)

The criteria for \( \epsilon \) in filter design is intuitive: average convergence for a source of false information must be farther than that of the authentic source. In addition, \( \epsilon \) should not be too small to limit the truthful optimal reporter. These criteria can be represented by:

\[
\begin{align*}
U_1(\epsilon) &= \mathbb{E}_{x_s, x_t} \left( c\left(x_s, \epsilon\right) - c\left(x_t, \epsilon\right) \mid \|x_t - x_s\| \geq d_{min}\right) \geq \delta, \\
U_2(\epsilon) &= \frac{\mathbb{E}c(x_t, \infty)}{\mathbb{E}c(x_t, \epsilon)} \geq \alpha,
\end{align*}
\] (20, 21)

where \( \delta \) and \( \alpha \) are positive constants. The first equation states that the average convergence distance for false news information sources must be farther than that of the true information source by a minimum margin \( \delta \). Note that the expectation is over the distribution of all information sources, and an information vector is considered false if it is at a distance \( d_{min} \) from the truth. The second equation on the other hand guarantees that the network is not too restrictive for the optimal reporter of truth compared with a free network with no authentication policy. Combining these metrics, we can define a unified utility function, and choose the maximizer \( \epsilon^* \):

\[
\epsilon^* = \arg\max_{\epsilon} U(\epsilon)
\]

\[
= \mathbb{E}_{x_s, x_t} \left( c\left(x_s, \epsilon\right) - c\left(x_t, \epsilon\right) \mid \|x_t - x_s\| \geq d_{min}\right) + \beta \frac{\mathbb{E}c(x_t, \infty)}{\mathbb{E}c(x_t, \epsilon)}
\] (22)

We set up a basic numerical simulation to calculate and illustrate the convergence metrics and utility functions of true and unauthentic information sources based on the derivations of the previous section. We consider a simple case of \( n = 2 \), with ergodic covariances \( \Sigma, \Sigma_s \) given below in (23). We assume that the true information \( x_t \) and source information \( x_s \) are both independently generated from a uniform distribution of vectors on the unit 2-sphere, \( S_2 \). The average viewers’ belief \( \bar{\mu} \) is also drawn from the same distribution.

\[
\Sigma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \Sigma_s = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}, \quad \mu_t \sim \mathcal{N}\left(\mu_t, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix}\right)
\]

\[
\bar{\mu}, x_t, x_s \sim S^2,
\] (23)

In Figure 2, the average convergence for true source \( \mathbb{E}\|\zeta(y^*_t) - x_t\| \) is compared with the convergence of a random source \( \mathbb{E}\|\zeta(y^*_s) - x_s\| \), as a function of network filter parameter \( \epsilon \). The standard deviations of the same quantities are displayed as error areas. Since the mean belief is completely independent of the source and truth, this figure is a representative of a convergence trade-off in a network of indifferent audience.

Figure 3 shows convergence trade-off conditioned on the fact that average belief is closer to the source (false) information than the truth, i.e., \( \|\bar{\mu} - x_s\| \leq \|\bar{\mu} - x_t\| \). We can think of this scenario as “uneducated audience”. Note that for larger \( \epsilon \), the false source convergence error area is below the average convergence of truth, meaning that the false information reporter has a significant chance of conveying more information that the truth.
Fig. 3: Statistics of belief convergence of optimal report for true information (blue) and random unauthentic information (red), for uneducated audience \( \|\hat{\mu} - x_s\| \leq \|\hat{\mu} - x_t\| \).

Figure 4 shows convergence trade-off conditioned on the fact that average belief is farther to the source (false) information than the truth \( \|\hat{\mu} - x_s\| \geq \|\hat{\mu} - x_t\| \).

Fig. 4: Statistics of belief convergence of optimal report for true information (blue) and random unauthentic information (red), for educated audience \( \|\hat{\mu} - x_s\| \geq \|\hat{\mu} - x_t\| \).

Finally, an example of network utility function for \( \beta = 1.6, d_{min} = 1.1 \) as a function of \( \epsilon \) is shown in Figure 5.

Fig. 5: Example of network utilities for \( \beta = 1.6, d_{min} = 1.1 \).

III. Conclusion and Future Work

A Bayesian model for information, viewer’s belief and the impact of the former on the latter was proposed in this work. The model was studied from the point of an optimal reporter intending to convey the message of a particular source to a statistically large group of viewers, through a network that implements a basic authentication policy. The simplifying assumption of numerical information vectors and Gaussian priors, posteriors and belief impact was made. The reporter was also assumed to be well-informed, which means he has the knowledge of truth, network policy and joint distribution of viewers’ belief statistics. Consequently, we derived the formulation of the optimal reporter, and showed that it is a linear combination of a portion of truth, source information and exaggeration. In addition, knowing this optimal strategy, we studied the optimal choice of the network authentication policy parameter. If the policy is too restrictive, it hinders the performance of a well-intended reporter that only uses exaggerative techniques to convey the truth to an uneducated or biased audience. However, too loose of a policy grants a large capacity for false information dissemination to false or malicious sources. We proposed a computable utility function that takes these factors into account, and can be optimized for the best choice of authentication policy.

Given that this study is the first of its own kind, the proposed model is expected to be significantly more progressed in the future work, in order to account for more practicalities. First and foremost, our primary goal for a follow-up work would be to validate the model with real-world data in various contexts (social media, news websites and electronic trading platforms). Secondly, we intend to incorporate less perfect assumptions such partially informed reporters, and selective or random authentication policies, and provide more analytical and concrete derivations than those of the current work.

Another limitation of the current model is that it is very static. In reality the source credibility parameters \( \Sigma_{s,i} \) is a function of the long term reporting of a source, and varies through the time. Furthermore, viewers have ways of cross-validating and comparing information coming from different sources. Other model parameters are also expected to be dynamic.

The present model is also passive, which means the viewers do not spread the information, and are only receivers of it. A very interesting study would be to adopt the current model and study dynamics of belief evolution and propagation in a network where viewers act as two-way agents passing and receiving information.

Finally, the role of education can be explained and studied with the current model to some extent. Education can be considered as a meta-function that skews viewers beliefs towards the truth over time. Given the importance of education in the age of social media and abundance of information, we believe that it is essential to provide modeling tools for studying its role in prevention of misinformation.
REFERENCES

[1] A. Benham, E. Edwards, B. Fractenberg, L. Gordon-Murmane, C. Het- henington, D. A. Liptak, M. Smith, C. Thompson, and A. P. Mintz, “Web of deceit: Misinformation and manipulation in the age of social media,” Information Today, Inc., 2012.

[2] H. Allcott and M. Gentzkow, “Social media and fake news in the 2016 election,” Journal of Economic Perspectives, vol. 31, no. 2, pp. 211–36, 2017.

[3] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, “Truth of varying shades: Analyzing language in fake news and political fact-checking,” Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing, pp. 2931–2937, 2017.

[4] A. Kucharski, “Post-truth: Study epidemiology of fake news,” Nature, vol. 540, no. 7634, p. 525, 2016.

[5] H. Rainie, J. Q. Anderson, and J. Albright, “The future of free speech, trolls, anonymity and fake news online,” Washington, DC: Pew Research Center, 2017.

[6] J. C. Olson and P. A. Dover, “Cognitive effects of deceptive advertising,” Journal of Marketing Research, pp. 29–38, 1978.

[7] G. M. Armstrong, M. N. Gurol, and F. A. Russ, “Detecting and correcting deceptive advertising,” Journal of Consumer Research, vol. 6, no. 3, pp. 237–246, 1979.

[8] S. Romani, “Price misleading advertising: effects on trustworthiness toward the source of information and willingness to buy,” Product and Brand Management, vol. 15, no. 2, pp. 130–138, 2006.

[9] J. Richards, Deceptive advertising: Behavioral study of a legal concept. Routledge, 2013.

[10] E. J. Lee, K. S. Eom, and K. S. Park, “Microstructure-based manipulation: Strategic behavior and performance of spoofing traders,” Journal of Financial Markets, vol. 16, no. 2, pp. 227–252, 2013.

[11] P. Kollock, “The production of trust in online markets,” Advances in group processes, vol. 16, no. 1, pp. 99–123, 1999.

[12] P. W. Singer and A. Friedman, Cybersecurity: What everyone needs to know. Oxford University Press, 2014.

[13] L. Huang, A. D. Joseph, B. Nelson, B. I. Rubinstein, and J. D. Tygar, “Adversarial machine learning,” in Advances in Financial Markets, vol. 16, no. 1, pp. 99–123, 1999.

[14] A. Chakraborty and B. Yolmaz, “Manipulation in market order models,” Journal of Financial Markets, vol. 7, no. 2, pp. 130–138, 2006.

[15] N. P. Nguyen, G. Yan, M. T. Thai, and S. Eidenbenz, “Containment of misinformation spread in online social networks,” in Proceedings of the 4th Annual ACM Web Science Conference. ACM, 2012, pp. 213–222.

[16] H. Zhang, M. A. Alim, X. Li, M. T. Thai, and H. T. Nguyen, “Misinformation in online social networks: Detect them all with a limited budget,” ACM Transactions on Information Systems (TOIS), vol. 34, no. 3, p. 18, 2016.

[17] A. El Abbadi, “Limiting the spread of misinformation in social networks,” in Proceedings of the 20th international conference on World wide web. ACM, 2011, pp. 665–674.

[18] M. Barreno, B. Nelson, R. Sears, A. D. Joseph, and J. D. Tygar., “Limiting the spread of misinformation in social networks,” in Proceedings of the 21st international conference on World Wide Web. ACM, 2012, pp. 71–80.

[19] C. Budak, D. Agrawal, and A. El Abbadi, “Limiting the spread of misinformation in social networks,” in Proceedings of the 20th international conference on World wide web. ACM, 2011, pp. 665–674.

[20] 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. ACM, 2012, p. 2.

[21] M. Gupta, P. Zhao, and J. Han, “Evaluating event credibility on twitter,” in Proceedings of the 2012 SIAM International Conference on Data Mining. SIAM, 2012, pp. 153–164.

[22] S. M. Shariff, X. Zhang, and M. Sanderson, “User perception of information credibility on news tweets,” in ECIR. Springer, 2014, pp. 513–518.

[23] K. Thomas, D. McCoy, C. Grier, A. Kolcz, and V. Paxson, “Trafficficking fraudulent accounts: The role of the underground market in twitter spam and abuse,” in USENIX Security Symposium, 2013, pp. 195–210.

[24] C. Yang, R. Harkreader, J. Zhang, S. Shin, and G. Gu, “Analyzing spammers’ social networks for fun and profit: a case study of cyber crime,” in Proceedings of the 21st international conference on World wide web. ACM, 2012, pp. 71–80.