Combating Misinformation Online: Identification of Variables and Proof-of-Concept Study

Milan Dordevic, Fadi Safieddine, Wassim Masri, Pardis Pourghomi

Combating Misinformation Online: Identification of Variables and Proof-of-Concept Study

Milan Dordevic[1], Fadi Safieddine[1], Wassim Masri[1], and Pardis Pourghomi[1],

[1] The American University of the Middle East,
P.O.Box: 220, Dasman, 15453, Kuwait
{Milan.Dordevic, Fadi.Safiedinne, Wassim.Masri,
Pardis.Pourghomi}@aum.edu.kw

Abstract. The spread of misinformation online is specifically amplified by use of social media, yet the tools for allowing online users to authenticate text and images are available though not easily accessible. The authors challenge this view suggesting that corporations' responsible for the development of browsers and social media websites need to incorporate such tools to combat the spread of misinformation. As a step stone towards developing a formula for simulating spread of misinformation, the authors ran theoretical simulations which demonstrate the unchallenged spread of misinformation which users are left to authenticate on their own, as opposed to providing the users means to authenticate such material. The team simulates five scenarios that gradually get complicated as variables are identified and added to the model. The results demonstrate a simulation of the process as proof-of-concept as well as identification of the key variables that influence the spread and combat of misinformation online.

Keywords: Misinformation, Information, Simulation, Social Media, Authentication.

1 Introduction

The process by which information and misinformation travels online and specifically by social media users has been the subject of several publications [1,2,3,4,5]. The challenges in combating misinformation on social media could be greatly enhanced should researchers be able to simulate the different scenarios of information and misinformation cascades. Specifically here, researchers need to consider the factors involved in the travel of misinformation and the factors involved in combating the spread of misinformation. In this paper, the authors identify the factors that influence the travel of information and misinformation as both theoretically start from one single node and travel across a network of nodes and points. Variables are identified for which the authors develop a fuller picture of what influences the process, speed, and success pace of fighting misinformation online. In the process of identifying these variables, the team simulates five scenarios as new variables are identified with each simulation and added to the model.

2 Literature Review

Oxford Internet Survey of 2013 results show online social networks as becoming one of the key sources of information and news especially among younger generations [6]. Thus, the spread of misinformation has increased as a result of the increase in the number of people using social networks [7]. Due to the lack of accountability of social media users spreading information and not having appropriate filtering techniques similar to reviewing and editing information in traditional publishing, social media have become a significant media for the spread of misinformation [4]. Thus, the spread of diverse forms of information, misinformation, and propaganda involves the distribution of false information through an information diffusion process involving users of social networks where the majority of users may not be attentive to the untruth story. In one study, researchers state that the acceptance of misleading information by the people greatly depends on their prior beliefs and opinions [8]. In another study [9], researchers state that the spread of misinformation in online social networks is context specific with topics such as health, politics, finances, theology, and technology trends are key sources of misinformation. People believe things which support their past thoughts without questioning them [10]. We have used the term misinformation to denote any type of false information spreading in social networks.

Considering the dark side of social networks, the environment facilitates the arrangement of groups and campaigns with the intention of undertaking unethical activities as well as mimicking widespread information diffusion behavior [4],[10]. Consequently, this facilitation of potential misconducts in online environments has encouraged some users to spread misinformation that results in greater support to cult-like views in a variety of topics [5]. What is more, those views are sometimes contagious and the individuals behind them make great efforts to spread them to others. The persistence of misinformation in the society is dangerous and requires analysis for its prevention and early detection [10,11,12]. The lack of accountability and verifiability however afford the users an excellent opportunity to spread specific ideas through the network while not discouraging freedom of expression and freedom of ideas [4].

In online social networks, the enormous distribution of data has resulted in persistent pockets of misinformation. Thus finding reliable information requires sifting-out the different types of misinformation in online social networks which has become a computationally puzzling task [10].

2.1 Related Work

In a research conducted by Lee et al. [1], the authors aimed at identifying and engaging "information propagators" which refers to people willing to help propagate information on social media. By modelling their characteristics and using that model to predict their willingness to propagate information in the future, the authors have been able to identify three characteristics of people willing to propagate information and misinformation online. These characteristics are: (1) personal traits of users such as personality and readiness to share or pass on that information; (2) the wait-time of a user based on the previous time lapses between passing on the information to predict the next time they share that information again; and (3) a recommender system based on the two previous

components to select the right set of users with a high likelihood of Re-sharing of information. While the paper focuses on Twitter as an example, parallels could be drawn to other social media applications.

In a research conducted by Hoang and Lim [2], the authors aimed to identify and model factors that contribute to viral diffusion based on the interrelationships among items, users, and the user-user network. This time the team categorized these factors into two sets. The first set includes external factors such as advertising, while the second set includes internal factors such as: a) Item virality which is the ability of an item to spread the adoptions by users through the follow links; b) the virality of the users diffusing the item which is their ability to spread the adoptions to other users through the follow links; and c) the susceptibility of the user adopting the item, which is the ability of a user to adopt items easily as other neighbouring users diffuse the items to others. The authors then proposed a Mutual Dependency Model that measures all three factors above simultaneously based on a set of principles that help to distinguish each property from others in viral diffusion.

In a research conducted by Jin et al. [3] the authors applied epidemiological modelling techniques to understand information diffusion on Twitter, in relation to the spread of both news and rumours. Epidemiological models are usually used to better understand how information diffuses by dividing users into several groups that reflect their statuses. The possible groups in which a user has been classified are: susceptible (S), exposed (E), infected (I), and recovered (R). Users could move from one group to another with a certain probability that could be estimated from data. Several models were introduced such as the SI model in which a susceptible (S) user can get infected (I) by one of his neighbours and will stay permanently in this state; SIS model where users can move back and forth between being (S) susceptible and (I) infected; the SIR model where users could move to a recover (R) state which is not really used to in news cascades models; and a model called SEIZ model (susceptible, exposed, infected, sceptic) proposed originally by Bettencourt et al. (2006) [13] which added a new state: exposed (E). Jin et al. (2013) suggested instead to represent the case where a user may take some time while in the exposed (E) state before believing a rumour (i.e. move to an infected (I) state).

For simulating spread of misinformation online, Budak et al. [4] presented a network algorithm that could be tested in case of two competing campaigns using two variations of the Independent Cascade Model (ICM) termed: (1) Multi-Campaign Independent Cascade Model (MCICM) and (2) Campaign-Oblivious Independent Cascade (COICM) to consider how information and misinformation spreads online. The paper, theoretically, relies on the design of the system itself and the input of 'influential' people to counter 'bad' campaign and limit misinformation. This could potentially be useful during time-sensitive political campaigns or breaking news events. Budak et al. acknowledge a limitation shared by other publications in this area when it comes to lab modelling of information diffusion acknowledging that lab models may not reflect the full extent of influences in real life. Thus, lab simulations will still need to be tested in the real world of social media [4] p. 667.

While previous work has provided important literature into the behaviour and challenges of spreading and combating misinformation online, there does not seem to be

one uniform method to how the spread of information is modelled. Nor does there seem to be uniformed agreed method for modelling the spread of misinformation. In addition, despite repeated review of the literature, the team could not find any viable or applied proposal on how to combat the spread of misinformation online.

2.2 The 'Right-click Authenticate' in combating misinformation online

In a prior publication [5], the authors suggested an approach to combating misinformation on social media. The team proposed an automated approach, dubbed as 'Right-click Authenticate' option that could review, rank, and identify misinformation using tools already found online. However, these tools have not been combined together in a setup that helps online users in their pursuit of authentication of the information they view. Three categories of authentication have been identified: textual, imagery, and video misinformation, however the paper focused on the first two: Textual and imagery authentication. In that process, users who are unsure about the content could right-click and select authenticate as conceptualized in figure 1.

Using reverse image search [14], a search that requires the user to upload an image or copy the image's web address to search for matches to that actual image online, users are able to identify the sources and dates of the first appearances of that image online as well as the context in which the image is presented. Some of the highly refined reverse image searches are able to detect even modifications of the image including color tones changes, photo editing, cropping and writing made onto the original image. Second layer is to validate any meta-data linked with the questioned image including the camera used, date it was created, and what photo-editing tools have been used. Meta-data may also help detect if any image editing tools have been used [15,16].

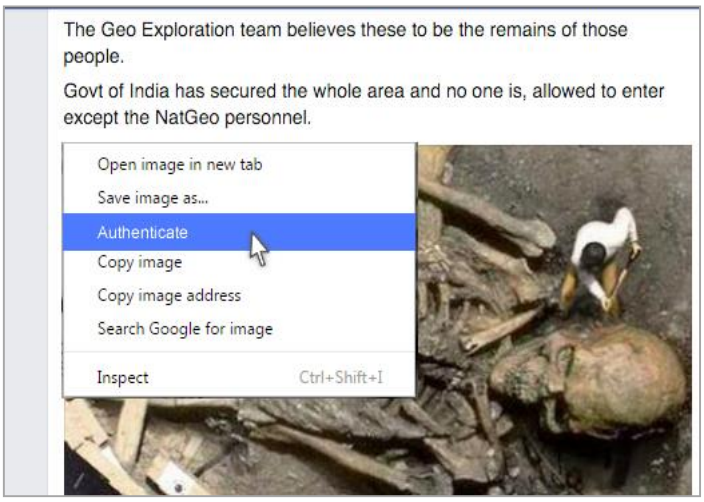


Fig. 1. Conceptualizing a right-click 'Authenticate' option [5]

The third part is an editorial feedback written in the same format and style Wikipedia operates authentication of information [17] with regards to the authenticity of that image. Image editorial feedback is combine with explanations based on the origin, date, meta-

data, where the image appears online, or article that dismisses or confirms that image. Finally, a crowdsourcing of feedback is represent the final indication on what the majority of users judging this information. These four sections are combined as: Image Match, Image Metadata, Editorial, and Feedback respectively. The solution is the bundling of these four sections into one single right-click option as conceptualized in figure 2. To ensure the successful results, the same algorithm used for online search engines to be used here. Thus, images that get frequently selected as a match to get higher ranking than those images that do not get selected as a match.

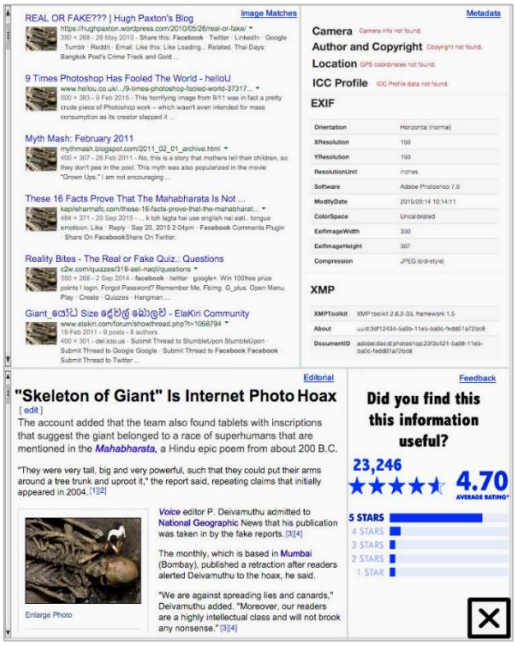


Fig. 2. Conceptualization of the 'Authenticate' outcome [5]

The right-click search authenticate option can also be used to authenticate a selection of news by title or text since the option to select and search text is already a well-established right-click search option on variety of browsers [5]. Another benefit for this right-click authenticate for images and text is that copyright infringements on intellectual rights are become more easily detected. In the paper, the authors acknowledge that new images and breaking news to require longer time to be authenticated. This proposal just based on theory and has not been simulated, implemented, or tested.

3 Research Questions and Methodology

The team acknowledges that the 'right-click authenticate' method for tackling the spread of misinformation needs to be demonstrated and proven as proof-of-concept. This paper attempts to further develop this theory to answer some key questions:

1. What variables are influencing the spread of information and misinformation on social media?

2. By means of simulation, can the process by which information and misinformation be modeled as proof-of-concept?

The team used graph theory computational simulation with observational research method [18]. In the process of identifying variables, the team used reflective analysis [19] to review progressively different scenarios in the spread of information and misinformation on social media. This approach is comparable to other approaches identified in the literature [1,2,3,4].

In lab conditions, the team observed the different two-dimensional simulations of information as it travelled from the source to a theoretical maximum reach. The two-dimensional simulation represents a slice of what a real-world multi-dimensional simulation of information would likely resemble. Successively analyzing and observing simulations of scenarios, the team subsequently evolved their model of simulation to identify and introduce new variables. With the introduction of new variables, a reflective analysis considered the logical impact of the new variable. Changes to the simulation and the justifications are then considered. While conducting the simulation, the team suggested values for such variables that are not based on any specific scenario or research, but solely for the purpose of facilitating a reflective analysis to re-evaluate the simulation and considering missing factors.

One of the main assumptions agreed at the start of the simulation is that the phenomena by which information and misinformation travels can be simulated despite unpredictability generally dominating human behavior online. This assumption is consistent with other academic publishers in this area of research. Without a preset of simulation scenario or the number of variables, the team developed five simulations and identified a total of ten variables. The demonstration, simulation, and identification of variables presented in this paper will be extended in further research aiming to design a formula by which success rate of combating misinformation online could be used for computational simulation.

4 Variables and Graph Modelling

Spread of misinformation in social networks can be modeled by using graph theory. The team considered a weighted directed graph $G = (V, E)$ consisting of V vertices and edges E. V can be viewed as the users of the social network. Among the vertices in V, the team distinguishes two types of vertices:

(1) Vertices which belong to set S, the set of sharing vertices, which represents users that send and receive information;

(2) Vertices which belong to set R, the set of reading vertices, which represents users who only receive information- accordingly $R \subseteq S$. A vertice r is a neighbor of a vertice s if and only if there is $e_{s,r} \in E$, an edge from r to s in G. Furthermore, all vertices from V can be divided into subsets (layers) depending on length l. Where $V = V_1 \cup V_2 \cup ... \cup V_l$ from figure 3.

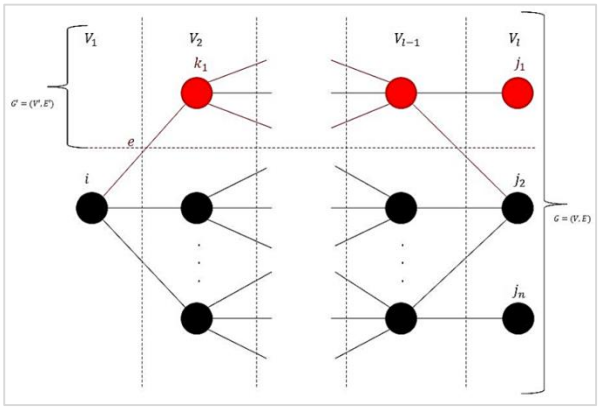


Fig. 3. Graph of Misinformation Modeling

Assuming i and j are any vertices of the given graph, the vertices i and j_n are connected by certain chains of edges going through different layers. The main goal of the team's approach is to see the effect of cascade labeling in models that they created. Note that cascade labeling symbolizes pressing the Right-click 'Authenticate'.

As illustrated in figure 3, by selecting edge $e, where\ e \in E$, the entire sub graph G', from $k_1 \in V_2$ to $j_1 \in V_l$, where l stands for length, is colored in red. This step is known as cascade labeling. Subsequently, this cascade labeling results in coloring some of the vertices from G into red. Coloring in red symbolizes the node authentication of the information to be untrue and the exclusion of sharing misinformation. The authors assume that pressing the 'Right-click Authenticate' can happen more than once in a demonstrated model.

Given graph G by sequentially repeating the cascade labeling process i.e. pressing the 'Right-click Authenticate' button, the number of vertices colored in red increases while the number of vertices colored in black decreases.

Since selecting an edge e results in coloring some vertices of sub graph G' into red, repeating the same process on any other edge from E in a graph G will result in coloring some more vertices into red.

Eventually, after n repetitions of this process in graph G, all vertices from subset V_l can be colored in red. Therefore, by implying cascade labeling procedure, some of the destination vertices $j_1, j_2, \dots, j_n$ will be preserved of receiving misinformation.

The model in figure 3 is assuming that only one vertex authenticates the information and passes that information on. The first vertex to authenticate and turn from black to red is modeled as red with black line and labeled as vertex $s_2 \in S_2$, where $S_2 \subseteq V_2$. The extended version of that model is shown in figure 4.

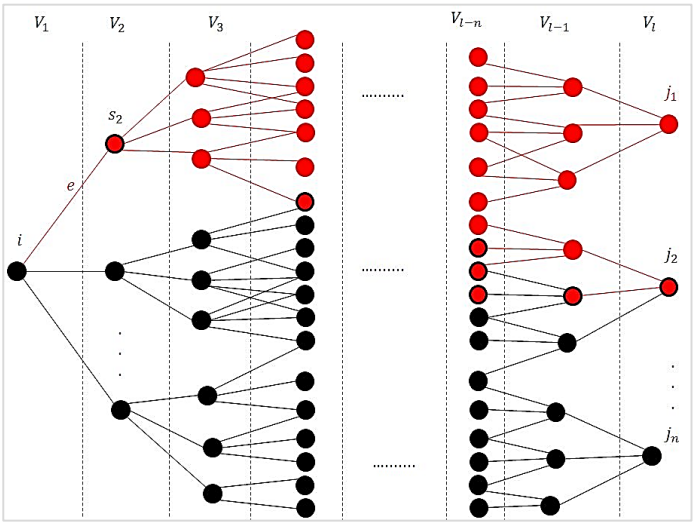


Fig. 4. Extended Graph of Misinformation Modeling

In the next scenario, the team studied next three variables that need to be considered in combating misinformation online: rate of authentication, rate of sharing, and rate of cross-wire.

The rate of Authentication (A) is a variable that represents the rate of users willing to authenticate the information. This usually occurs when online users are not sure of an information or when they get conflicting information. Thus, these users might decide to authenticate such information to start a correction cascade or at least stop the cascade of misinformation from their part.

The rate of Authentication (A) could be anything between 0 and 100%; although the team acknowledges it is unlikely to be either extreme. For the simulation in figure 5, the team predicts that the percentage of users who will authenticate to be around 30%. Thus for the simulation purposes, the team have assumed the probability of authentication as $A = 0.3$.

The passing on Rate (P) is a variable that represents the ratio of users who read the information and then perform an action of actively disseminate it further. Thus, the ratio shows the probability that vertex which authenticate will pass that correct information to anyone else as well as the ratio of those vertex that pass on misinformation.

The synonyms used for passing on rate are average of forwarding, liking, and sharing rate. We assume that the rate of willingness to share is probably the same for those who believe the misinformation. To demonstrate this scenario, the team assumed the probability of sharing information by online users regardless they believe it or not to be $P = 0.5$. Although if the research determines differences in sharing between those who believe and those who do not believe the information, variations of this variable could be created as P_1 for those who believe the information to be true and P_o for those who do not believe it.

The Cross-Wire (Cw) is a variable that represents the probability that user who received different information from different sources will react to validate. In such a case, online users exposed to misinformation are sufficiently skeptical to question it and use the 'Right-click Authenticate' to validate it. In figure 5 vertex c_1 received different information from sources a and b and accept the information received from a while discard the misinformation received from b.

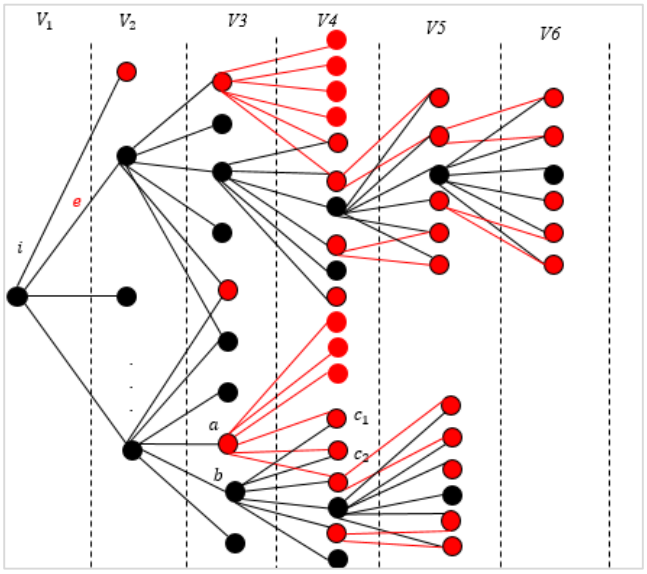


Fig. 5. The Authenticate, Passing on rate, and Cross-Wire rate simulation

For the purpose of simulation, the team assumed $Cw = 0.3$. The simulation in figure 5 shows the usage of variables A, P and Cw. Moreover, the figure shows how the speed of misinformation spread is slowed down compared to the scenario in figure 4 and again for those who authenticate the information.

As demonstrated in figure 5, providing means of authentication can have important impact on the spread of misinformation online. Red nodes are shown to be playing a role in limiting the spread of misinformation.

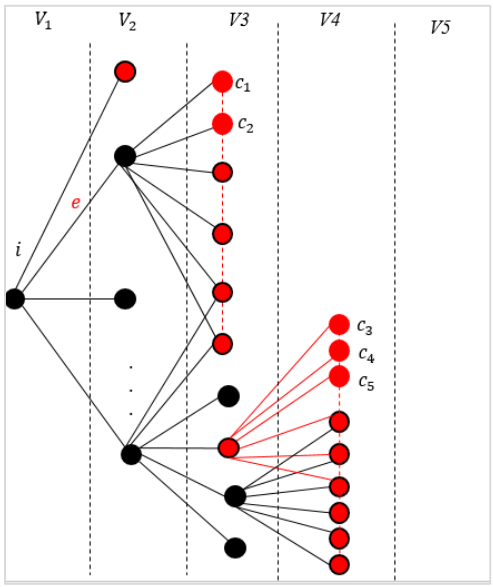
For the next simulation, the team considers Same Level Communication (Sl) as a variable that represents the probability that users who authenticate information and leave feedback encourages other users from the same level also to authenticate. That includes passing on vertices on the same level thus turning several of these vertices from black to red. For example, in figure 6, vertices c_1 and c_2 have validated the misinformation. Provided c_1 or c_2 left a feedback, this turns the remaining vertices from black to red. The same happens to $c_3, c_4, and\ c_5$.

The team assumes that all other variables and assumptions are kept in place.

A vertex that is red or just turned red with probability 0.5 will alert other online users that the information is not true. In such scenario, the team assumes that other vertices will, in turn, discard the misinformation and turn into red. The reason behind this assumption is that the first online user to authenticate has a greater impact on subsequence

online users. So the team assumed that probability of Same Level Communication is $Sl = 1$.

Considering the rate of authentication A, passing on information rate P, average cross-wire rate Cw, success rate of Same Level communication rate Sl, where $A = 0.3$, $P = 0.5$, $Cw = 0.3$, and $Sl = 1$ respectively, excluding misinformation does not extend beyond V_4. The simulation of this scenario specific lab scenario, demonstrated in figure 6, where at level V_4 all vertices are red.



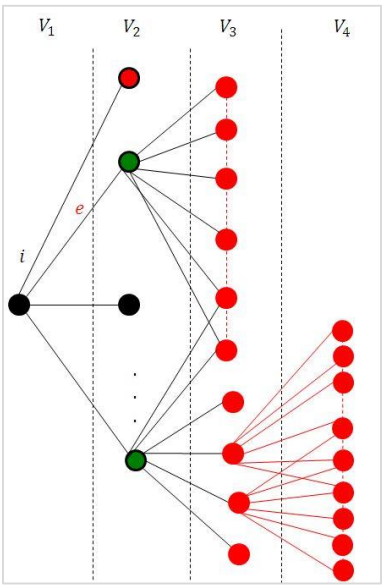


Fig. 6. Reverse Validation (Rv). **Fig 7**. Same Level Communication (Sl)

For the final simulation, the authors have considered Reverse Validation (Rv) Variable. Rv represents a probability that the user who initially believed the misinformation, while being informed by other users through their feedback that the information is not true, either removes the post or rectify the post, thus turning red node themselves. However, to differentiate them from other red node, the team decided to label such node green. This is a backflow to a previous or source vertex. The output of applying the Rv, is shown in figure 7 as green vertices. The team considered this final variable at probability that the source vertex will take action to rectify the misinformation as $Rv = 0.5$.

5 Results and Limitations

The combinations of all these variables and the assumptions that the team made to understand how combating misinformation works has resulted in identifying some key variables where i is the first vertex and j_n is the last vertex of the given simulation. V_1 represents the first phase of spread of misinformation and l represents the maximum possible reach of information through the network. The authors conclude that combating misinformation online is also be influenced by the following variables: rate of authentication A, passing on information rate P, average cross-wire rate Cw, success rate

of Same Level communication rate Sl, and Reverse Validation rate Rv. Thus the paper demonstrates by means of simulation how misinformation travels online. The paper also shows how 'right-click authenticate' process can reduce the spread of misinformation online. Thus suggesting a viable solution for combating misinformation online by identifying and demonstrating key variables and factors.

The proof-of-concept has been constrained with assumptions that are based mostly on observations of computer simulation and reflective analysis subjective to individual experiences of the team. However, the approach has been backed by similar observations done in other academic publications [1, 2, 3]. The team acknowledges that the proposed variables may not be exclusive, and that further research may reveal additional factors influencing the travel of information and the means of combating misinformation online. Furthermore, the identification of the variables is lab based and further proof should be drawn from examples from existent event observations once the formula is developed. This is a limitation acknowledged in the literature when it comes to lab modeling as opposed to real life simulation [4]. The 'right-click authenticate' process has two key limitations in application and implementation [5]. In application, the authors acknowledge that the 'authentication' option has little or no real impact at authenticating breaking news. For the process to work, time is needed for the information or image to be authenticated and a review written. For the implementation limitation, the building of the 'right-click authenticate' option requires authorization and collaboration from a reverse image search engine, which may not be forth coming.

6 Conclusion

The team set out to demonstrate a proof-of-concept and identified the variables involved in the travel of information and the 'Right-click Authenticate' idea suggested in a previous publication [5]. The team believes that some headway has been achieved but that still work to be done to develop the formula and conduct simulations to further validate the concept. Two parallel lines of further research are expected to follow. First, the team will be working towards developing the formula and run computational simulations of the formula using MATLAB and BioLayout Express for three dimensional simulation. Second and equally important, the team intend to develop a prototype browser based on an existing open source applications that allows demonstration of the concept and the running of actual simulations thus allowing lab and field simulations.

7 References

1. Lee, K., Mahmud, J., Chen, J., Zhou, M., Nichols, J.: Who will retweet this?: Automatically identifying and engaging strangers on twitter to spread information. In: The 19th international conference on Intelligent User Interfaces, ACM, pp. 247-256 (2014)
2. Hoang, T. A., Lim, E. P.:Virality and Susceptibility in Information Diffusions. In: ICWSM (2012)
3. Jin, F., Dougherty, E., Saraf, P., Cao, Y., Ramakrishnan, N.: Epidemiological modeling of news and rumors on twitter. In: Proceedings of the 7th Workshop on Social Network Mining and Analysis, ACM. p.8 (2013)
4. Budak, C., Agrawal, D., El Abbadi, A.: Limiting the spread of misinformation in social networks. In: Proceedings of the 20th international conference on World Wide Web, ACM, pp. 665-674 (2011)

5. Safieddine, F., Masri, W., Pourghomi, P.: Corporate Responsibility in Combating Online Misinformation. International Journal of Advanced Computer Science and Applications(IJACSA), 7(2),pp. 126-132 (2016)
6. Dutton, W.H., Blank, G., Gorseli, D.: Cultures of the Internet: The Internet in Britain. Oxford Internet Survey 2013 Report: University of Oxford (2013)
7. World Economic Forum Report.: Top 10 trends of 2014: The rapid spread of misinformation online (2014)
8. Libicki, MC.: Conquest in cyberspace: National security and information warfare. Cambridge University Press, New York, USA (2007)
9. Karlova, NA., Fisher, KE.,: Plz RT: A social diffusion model of misinformation and disinformation for understanding human information behaviour. Inform Res, 18(1), pp. 1–17 (2013)
10. Kumar, K.K., Geethakumari, G.,: Detecting misinformation in online social networks using cognitive psychology. Human-centric Computing and Information Sciences, 4(1), pp. 1-22, (2014)
11. Lewandowsky, S., Ecker, U.K., Seifert, C.M., Schwarz, N., Cook, J.: Misinformation and its correction continued influence and successful debiasing. Psychol Sci Public Interest, 13(3), pp.106–131 (2012)
12. De Neys, W., Cromheeke, S., Osman, M.: Biased but in doubt: Conflict and decision condence. PLoSONE, vol. 6, e15954 (2011)
13. Bettencourt, L.M., Cintrón-Arias, A., Kaiser, D.I., Castillo-Chávez, C.: The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. Physica A: Statistical Mechanics and its Applications, 364, pp. 513-536 (2006)
14. Martin, J.: How to do a reverse Google Image search on Android or iPhone. PC Advisor (online). (2016)
15. Buchholz, F.: On the role of file system metadata in digital forensics. Digital Investigation, 1(4), pp. 298–309 (2004)
16. Castiglione, A., Cattaneo, G., De Santis, A.: A Forensic Analysis of Images on Online Social Networks. IEEE Conference, pp. 679-684 (2011)
17. Wikipedia Contributors.: Wikipedia : Version 1.0 Editorial Team. Wikipedia, The Free Encyclopedia, https://en.wikipedia.org/wiki/Wikipedia:Version_1.0_Editorial_Team/Assessment
18. Osmond, J., Darlington, Y.: Reflective analysis: Techniques for facilitating reflection. Australian social work, 58, pp. 3-14 (2015)
19. Altmann, J.: Observational study of behavior: sampling methods.Behaviour, 49.3, pp. 227-266 (1974)