

Boğaziçi University
Department of Industrial Engineering

MS THESIS PROPOSAL

Dynamic Analysis of False Information Spread Over Social Media

Orkun İrsoy

Supervisor
Prof. Dr. Yaman Barlas

Istanbul, February 2021

INTRODUCTION

Means of communication in today's world have undergone a significant shift towards digital communication. Social media is being used by a vast majority of people including people from every age group or socio-economic status. Consequently, more and more people are using social media to obtain and spread information on various subjects. Reuters report that nearly two-thirds of adult people in US use social media as a news source (Moon, 2017).

This new way of communication provides many advantages: it promotes engagement, reduces barriers among people all around the world by providing an alternative to face-to-face socialization. The information on these social networks propagates faster and to a wider audience. However, since the content is created by the users without any review process unlike traditional media, validity of the content is not verified. As a result, people, either willingly or not, might spread the false information. This is such a prevalent phenomenon that an article trying to quantify the exposure to fake news of US population before the presidential election on 2016 stated that average adult has seen at least one fake stories among a quite narrow dataset of 156 fake stories (Allcott, H and Gentzkow, 2017). Another problem is existence of fake accounts also called bots which are controlled by a person, a group of people or an artificial intelligence to manipulate people. The number of bot accounts in Twitter is estimated between 9 to 15% (Varol et al., 2017) and for Facebook the number is claimed as 60 million (Lazer et al., 2018).

Spread of false information may cause a wide range of problems such as political manipulation of huge masses (Varol et al., 2017), blocking the relief operations during a disaster or may even cause a terror attack (Kumar and Shah, 2018). One such case is Pizza Gate conspiracy which ended up a person firing a gunshot at a local store due to rumors about child-trafficking (Morstatter, Carley and Liu, 2019). Another example is the accusation of Facebook due to manipulation of citizens on the Presidential Election in 2016 (Lazer et al., 2018). Given the severity of such consequences, false information spread in digital platforms is classified as one of the current threats to society according to the World Economic Forum (Lee Howell et al., 2013).

Policy makers try to alleviate the problem by endorsing research on detecting the false information on digital platforms thus there is a vast literature about the subject at hand. Blackbox models are used to detect the false news by exercising various data mining methods such as natural language processing or image processing. Another way is to track the spread pattern and try to detect fake information by the propagation which is entertained by network scientists. Finally, simulation is used to make an inference about the system and evaluate different interference scenarios. A detailed review of such models is presented by Kumar and Shah (2018).

The problem of false information spread over digital social networks has many drivers. At individual level many psychological factors are effective. Firstly, ability of people distinguishing true and false information is surprisingly close to random with accuracies between 53 - 78% (Kumar and Shah, 2018). Moreover, other biases such as confirmation bias (People's tendency to trust their existing information), desirability bias (People's tendency to accept information that pleases them). Some of these biases stem from social interactions within the network such as availability cascade (People's tendency to accept information as the popularity of such opinion increases) or validity effect (People's tendency to accept information as the number of exposures to that information increases) (Zhou and Zafarani, 2018). At an aggregate level, many social, political or economic drivers are effective depending on the classification of information.

Apart from the psychological and social factors, network structures are also effective on how false information spreads among people. Since the social media networks are usually preferential, they are prone to effects such as echo chambers (repetitive exposure of a specific information due to homogenous social clusters) (Vicario et al., 2016). This requires an interdisciplinary perspective to solve the problem (Zhou and Zafarani, 2018) at hand which encourages to utilize systems thinking.

In their paper Ammara, Bukhari, and Qadir (2020) introduce system dynamics methodology to the subject and build a small model to show that causal and holistic perspective is advantageous in coming up with permanent solutions. Using different archetypes, they claim that efforts to eliminate such a problem cannot go beyond naïve interventions which sometimes are backfired through feedback loops and they encourage further utilization of this methodology in this field.

PROBLEM DEFINITION

To analyze how's and why's of false information spread over social media, we must first define what is false information. Many definitions and classifications of false information exist in the literature. Intention of the agent is one dimension where "misinformation" refer to unintentionally spreading information whereas "disinformation" is intentional. Another categorization is the knowledge-based differentiation i.e., whether the information is purely factual or opinion-based (Kumar and Shah, 2018).

Drivers, propagation dynamics and mitigation strategies may differ for each aforementioned classification for the false information. As an example, in the case of misinformation, financial motive is one of the key motivations and consequently decreasing the financial gain from click-based web activity is one of the trivial solutions. However, same mitigation strategy would not benefit much for the opinion-based political news in which the preferential network structure of the social media is the main effective factor. On the other hand, some of the proposed mitigation strategies such as increasing news literacy might be common for both classifications.

Parallel to the diversity of classifications, research on how to combat false information spread usually focused on specific type of information or concentrated on a single stage of the process such as how to detect the false information from content, from propagation, what should be the action after detection and so on. This compartmental approach, however, often fails to build a holistic solution to problem of dis/misinformation over social media. An example to further solidify the need for a systemic perspective might be the increasing use of data mining methods to classify whether an information is false or not. Those methods try to determine the authenticity of the information using content-based (the text or video that the information is composed of) or context-based (the time the information is shared, how many shares it received during the first hour etc.) data but these results also pave the way for building more effective AI's to generate such fake information (Ammara, Bukhari and Qadir, 2020).

Therefore the main objective is to build a model that will utilize the current knowledge on each focal point but still captures the interaction between these would create a huge value in that it will allow us to make inference about the influential drivers,

reveal the endogenous feedback loops, test the effectiveness of current mitigation strategies in a holistic perspective and come up with more structural solutions rather than the naïve interventions. System dynamics methodology is especially suitable for such a task since it allows us to combine the accumulated qualitative and quantitative knowledge in a variety of disciplines.

METHODOLOGY & POSSIBLE RESEARCH DIRECTIONS

The first step is the extensive literature survey on false information spread over social media. Particular emphasis should be made on the common and different components of each information classification. At the core, an information spread over social media module will be common for each type of information. The core model can be expanded if there are structural differences that can affect the behavior or the recommended mitigation strategy for each type. In other words, the idea is to start with a broader definition of false information and build the model and then add different modules to specify different types to see if the resulting behavior changes or not.

Another branch of the literature survey should focus on finding the validation scenarios for the model at hand. Although the qualitative validation might be a difficult case for this model, extreme conditions, possible mind experiments and etc. should be extracted from qualitative literature. A research on simulation in social sciences would add great value to this model as a literature approved methodology will built a confidence to the model which is essential for a simulation-based approach.

Finally, literature survey should include the proposed mitigation strategies both for the core and extended models. This will construct a base for the scenario analysis in which we should try to answer questions such as: “Which of these strategies provide long lasting solutions/ which are naïve attempts?” “Do strategies work for all type of false information? Are there any conflicting mitigation strategies that backfires? If so, how should we implement them?” and so on.

RESEARCH PLAN

Literature Survey	6 Weeks
Review of Models and Model Construction	6 Weeks

Parameter Estimation & Calibration	2 Weeks
Validation of the Model	3 Weeks
Testing & Experimentation	3 Weeks
Reporting	4 Weeks

REFERENCES

1. Moon, Angela. “Two-Thirds of American Adults Get News from Social Media: Survey.” *Reuters*, Thomson Reuters, 8 Sept. 2017, www.reuters.com/article/us-usa-internet-socialmedia/two-thirds-of-american-adults-get-news-from-social-media-survey-idUSKCN1BJ2A8.
2. Wu, L., Morstatter, F., Carley, K. M., & Liu, H. (2019). Misinformation in social media: Definition, manipulation, and detection. *ACM SIGKDD Explorations Newsletter*, 21(2), 80–90.
https://www.public.asu.edu/~huanliu/papers/Misinformation_LiangWu2019.pdf
3. O. Varol et al., in Proceedings of the 11th AAAI Conference on Web and Social Media (Association for the Advancement of Artificial Intelligence, Montreal, 2017), pp. 280–289.
4. Allcott, H., & Gentzkow, M. (2017). *Social Media and Fake News in the 2016 Election*. 31(2), 211–236.
5. Lazer, D. M. J., Baum, M. A., Benkler, Y., Berinsky, A. J., Greenhill, K. M., Menczer, F., Metzger, M. J., Nyhan, B., Pennycook, G., Rothschild, D., Schudson, M., Sloman, S. A., Sunstein, C. R., Thorson, E. A., Watts, D. J., & Zittrain, J. L. (2018). The science of fake news. *Science*, 359(6380), 1094–1096.
<https://doi.org/10.1126/science.aao2998>
6. Kumar, S., & Shah, N. (2018). False Information on Web and Social Media: A Survey. *ArXiv*, April.
7. Lee Howell et al. Digital wildfires in a hyperconnected world. WEF Report, 3:15–94, 2013.
8. Zhou, X., & Zafarani, R. (2018). *A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities*. <https://doi.org/10.1145/3395046>

9. Ammara, U., Bukhari, H., & Qadir, J. (2020). *Analyzing Misinformation Through The Lens of Systems Thinking*. *Tto*, 55–63.
10. Sterman, J.D. *Business Dynamics: Systems Thinking and Modeling in a Complex World*. McGraw-Hill, Boston, 2000.
11. Vicario, M. Del, Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., Stanley, H. E., & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences of the United States of America*, 113(3), 554–559. <https://doi.org/10.1073/pnas.1517441113>