MISINFORMATION IN GOVERNMENT AND SOCIETY

A Research Paper submitted to the Department of Engineering and Society

Presented to the Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree

Bachelor of Science, School of Engineering

Samer Nahed Kadih

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Advisor

S. Travis Elliott, Department of Engineering and Society

Disinformation is false information deliberately meant to mislead. The subject of "fake news," and social media's role in exacerbating its effects, has been center-stage in political discussions since the 2016 U.S. presidential election (Tam, 2017), whose integrity is questioned as a consequence. More recently, according to the Congressional Research Service, social media plays an active role in hampering efforts to mitigate the spread of the COVID-19 virus (Gallo & Cho, 2021, pp. 14-18). A variety of reasons may contribute to the general public's unwillingness to screen the news presented to them for bogus statements, such as not making time in their busy lives or not feeling confident about their technical literacy (Horrigan, 2017). Why does this information spread in social media? What channels do politically motivated influencers have for propagating that information? Given the complex political relations and motivations involved in either spreading or suppressing misinformation, there is a plethora of relevant groups involved. In order to better understand the motivations and processes involved in peddling misinformation in the national political network, this STS research paper will employ Actor-Network Theory to model the relationship between these actors. (Callon, 1984). From a technical perspective, in order to keep up with the volume of data disseminating over the internet, machine learning models for natural language processing are typically employed for classifying the credibility of information (Asr & Taboada, 2019, p. 5). The success of these models, however, is contingent on access to large volumes of high-quality data for training and testing. A technical discussion, with advisement from Professor Daniel G. Graham, explores the development of a commercial pipeline process over a summer internship, meant to commodify big data quality and cleanliness. When coupled, these discussions aim to shed light on the importance of big data for algorithmically modeling networks, with a focus on how they can inform social media engineers' platform decisions in a way that is politically responsible.

IDENTIFYING THE POLITICAL NETWORK

Social media platforms are a business, and are justifiably programmed with a focus on their profit margins. Some of the bigger players–Facebook, Twitter, YouTube, and Wikipedia–employ an advertising business model, and so their motivation is in maximizing the amount of time a user spends on their platform (Bovet & Makse, 2019). These platforms structure their user experience so that it is a force for compelling their interest, but this invariably makes it a catalyst for provocative or sensational content. This is not a novel concept; even in 18th-century London, academics noted that "stories or gossip made it into newspapers which had just began to circulate among a broad public" (Asr & Taboada, 2019, p. 2). Today, the development of the internet amplifies the effects of this malpractice to a global scale. The human tendency to surround oneself with the familiar is the impetus for echo chambers: a phenomenon driven by human bias to consume content relevant to their platform, and avoiding any ideas that challenge their pre-existing beliefs (pp. 3-4).

Some of the psychological phenomenon that echo chambers appeal to are the influence of repetition, pronounceability, familiarity, and imagery (Greifeneder et al., 2020). It is observed that repeated claims increase acceptance among an audience. This effect is most pronounced with claims that people feel uncertain about, but it also tends to increase agreement among people who know that those claims are false. By virtue of a repeated claim feeling familiar, people are more apt to nod along in agreement. Though attenuated, this effect persists even after people are introduced to diagnostic evidence (p. 78). Pronounceability refers to the idea that claims whose sources are easier to pronounce endow them with higher credibility and trustworthiness (p. 76). An echo chamber's familiarity is the main driver of its users' confirmation bias (Bovet & Makse, 2019). Frequent exposure to information not only increases

its apparent truth, but also the belief that it came from a trustworthy source. These errors in attributing the claim to a credible source increases the likelihood that people convey that information to others in their network, who themselves are more likely to propagate it, and so on. Finally, social media posts have the freedom of linking strong imagery to their statements (p. 96). Even if these images are non-probative, such that they provide no evidence to a claim, they promoted a sense of truthiness nonetheless.

An experiment run by the Stanford History Education Group found that Americans of all ages, regardless of their digital or academic affinity, fail to ask important questions about the content they consume (Steinmetz, 2018). On average, people are inclined to believe false news 20% of the time. This motivates the idea that forces against truth, such as economically or politically motivated local and foreign entities, or just simply conspiracists, can appeal to social media platforms for distributing disinformation with greater frequency, and, as a result, greater effectivity. On the other end of the spectrum, forces for truth can appeal to instituting better local and national policies for educating the general public on classifying and filtering out disinformation (Asr & Taboada, 2019, p. 2). Though technology can help to dampen the influence of these phenomenon in propagating fake news, this is a problem that requires an interdisciplinary approach.

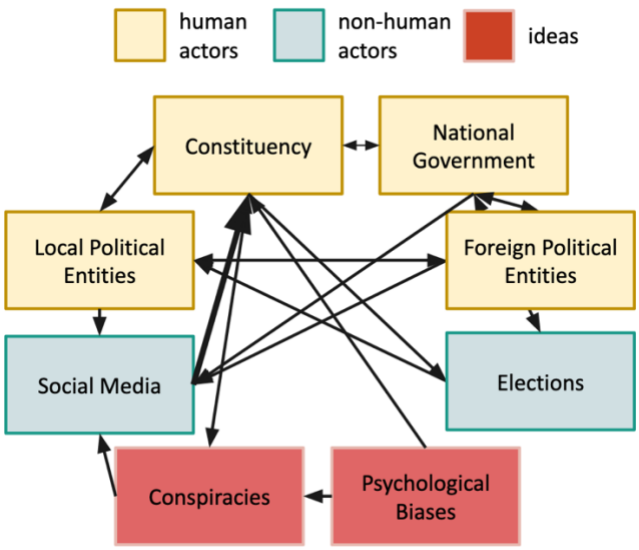

Figure 1. Misinformation Actor Network. Misinformation in government and society. Kadih, S. (2022).

Though technology can help to dampen the influence of these phenomena in propagating fake news, this is a problem that requires an interdisciplinary approach. Figure 1 outlines the actors discussed in this actor-network and how they influence one another.

MACHINE LEARNING FOR CLASSIFYING RELIABLE NEWS

Corrupt actors in social networks, typically motivated by economic or political means, exploit echo chambers by planting sensational posts on social media and having them organically spread throughout the platforms (Asr & Taboada, 2019, p. 4). The speed at which a non-sensical idea can reach the masses is frightening, but even more frightening are the lasting consequences: “people tend to remember facts and events that have been repeatedly mentioned, even when the repeated mention is in the context of a retraction or myth debunking” (p. 4). Detecting misinformation in news sources is an integral part of helping regular consumers of social media responsibly navigate the political environment. The logic behind how the content of participants in a social network is distributed is programmatic by nature. Appealing to the moral framework described by duty ethics, humans ought to have greater autonomy in their reactions to

provocative segments of news, and to avoid falling prey to the psychological tactics used by peddlers of fake news (Vallor, 2021). The idea of one's identity is at odds with its interpretation in social media platforms, where users are algorithmically profiled and classified to general categories. Social media is an attractive space for constructing one's identity. It is not so clear, however, whether a user's identity is more-so formed by the expression of ideas genuinely and proportionally belonging to their network, or by the influences of a corrupt few. To appeal to the ethical demands of identity, social media engineers have a responsibility to programmatically detect fake news before it is able to spread throughout a social network, so to protect and promote the interests of their users in autonomously managing their identities.

There is a plethora of well-documented, rudimentary models for detecting fake news (Asr & Taboada, 2019, p. 5). A popular method is adaptive graph convolutional networks, in which various kinds of information structured arbitrarily, like visual and text data from a newsfeed, can be leveraged to better classify that information (Qian, Hu, Fang, & Xu, 2021, pp. 7-11). A sophisticated model proposed by Ruchansky et al. splits the learning into different modules, as depicted by Figure 2 on page 6. The "capture" module uses a recurrent neural network (RNN) to characterize the temporal patterns present in the distribution of an article, with the hope of being able to discriminate between verifiable and fake news sources based on how they propagate over a network. The "score" module attempts to characterize user behavior when they engage with an article, specifically with respect to the identified credibility of that article. Finally, the "integrate" module attempts to construct a relationship between the "capture" and "score" modules to make a conclusion on the veracity of that article. This poses a silent ultimatum for users of a social media platform, whose influence is linked to how responsible they are in sharing disinformation.

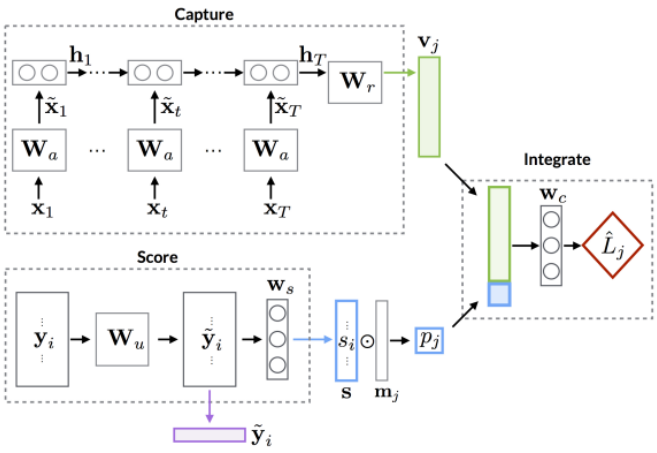

Figure 2. CSI model specification. Kadih, S. (2022).

A critical element to the success of social media engineers in dampening the effects of fake news is in gathering a large pool of quality data regarding misinformation in social media, so that the machine learning models are trained with a non-trivial set of data for practical applications. Then, different flavors of models, such as predicative modeling or multi-modal representations, are tested and tuned against readily available benchmarks. Finally, it is pivotal that the interface for this project can access a pool of news sources which can be catered to the end-user, such as by a third-party news API, after the news that ought to be catered to the user is properly identified. A diverse palette of news will likely be suggested to minimize bias. A proper model will classify user-typed or viewed social media posts as either true or false. If news is classified as the latter, then, either by the same or a different model, derive which articles from verified news sources in recent history are related to the context of the accused news. News from verified sources is considered better suited for conveying information to unsuspecting consumers of social media with regards to the present political landscape.

NETWORK PROPOGATION ALGORITHMS FOR RELIABLE NEWS

It is important to explore the propagation of news in social media settings. Specifically, analyzing different models for *how* and *why* misinformation propagates over media channels from a technological perspective and its impact on government and society. One of the U.S. government's chief subjects of interest involve mitigating media channels' role in exacerbating misinformation (Gallo & Cho, 2021, pp. 14-18). With respect to the COVID-19 virus, whose subject is most relevant to the present political environment, channels of misinformation regarding the dangers of the vaccine are a detriment to alleviating the issue (Gradoń et al., 2021, pp. 1-4). How this information propagates, resulting in polarizing ideas in social media, can be modeled by immediate constituent analysis: a profound method for analyzing complex networks (Prasetya & Murata, 2020, p. 7). To anchor this model to a problem that is still actively a detriment to society, the Biden administration's efforts to increase vaccination rates were met with hesitation by the public, and a surge in hospitalized patients was the consequence (Ivory et al., 2021). This is likely a cause of targeted peddling by corrupt actors in the network. If a suitable model can be determined for how the dynamics of our present political network behave, then a method for minimizing the effects of fake news propagation can be employed.

Actor-network theory (ANT) is invoked on the particular case of misinformation in the political network. Different actors involved in the network have different influences in how misinformation is spread, each with their own psychological trends and oddities that are outlined in present-day research. The general trends of a network can be described by the probabilistic laws governing the network. Although ANT cannot feasibly predict the actions of any one actor, over a large sample these populations can be understood to act in a certain way with some level of confidence, and thus can be modeled by stochastic processes (Venezuela et al., 2019, pp. 808-

810). Figure 3 demonstrates how Bayesian statistics can be used to fit parameters to how a network operates. Combined with an interpretation of the different actors involved in a network, and their links to one another, such as the actor-network outlined in Figure 1, these parameters give us insight into how strong the relationships are between actors. When compared to other models proposed by Venezuela et al., it appears that politically active citizens in a network tend to distribute misinformation more often than those who are less engaged. By jointly using ANT to study the motivations of different actors, as well as breaking down the actors into measurable stochastic processes, there are then two different avenues of approaching the problem of minimizing the dissemination of fake news. This can either be done between the relationships of different actors, or on a more granular level, as with an actor's relationship to their motivations.

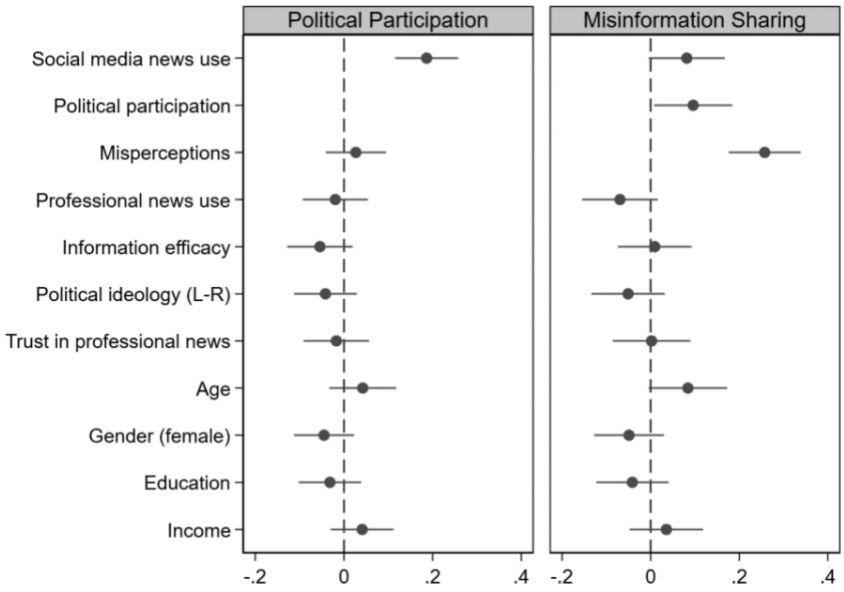

Figure 3. Standardized coefficients. Kadih, S. (2022).

Figure 4 highlights a suggested model for information dissemination. This simple model makes a few assumptions, or hypotheses, about our political environment to explain its dynamics accordingly (Valenzuela et al., 2019, pp. 806-807). Hypothesis 1 (H1) is that social media news use is positively correlated with political participation. Hypothesis 2 (H2) is that political

participation is positively correlated with the spread of misinformation. Hypothesis 3 (H3) is that misperceptions about the news cycle will moderate that relationship between political participation and the sharing of misinformation in a network so that the relationship is stronger for misinformed users and weaker for informed users.

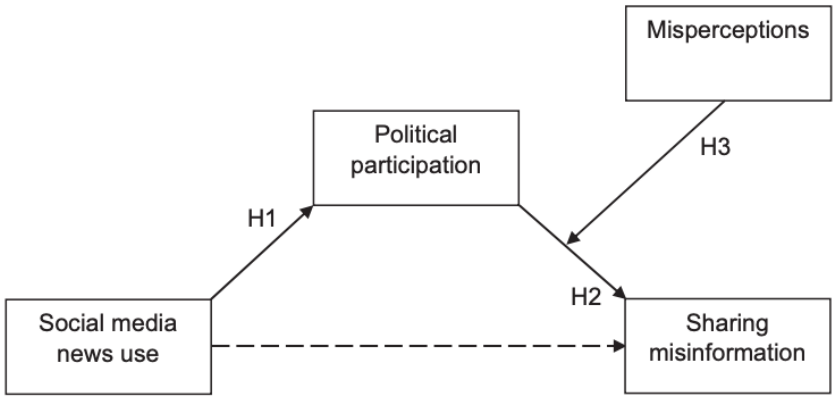

Figure 4. Model of disinformation dissemination. Kadih, S. (2022).

Algorithms for how information propagate over a network can have drastic implications on the consequences of that information. It is pivotal to understand the mechanisms behind information diffusion within a network. Those mechanisms can be constructed or modified to distribute information in a decidedly fairer way, so that demographic biases are minimized in the spread of that information in a social network (Stoica & Chaintreau, 2019, p. 1). The starting seed, or initial outreach, of information in a network is particularly indicative of how fair it will go on to become in terms of diversity of information diffusion.

ANALYSIS AND REMEDY OF MISINFORMATION NETWORK

Social media platforms, provided they maintain an advertising revenue model that motivates sensational posts and claims, are a detriment to preserving a user's autonomy and personal identity. There are large volumes of data available to social media engineers whose commodification can be crucial for the development of better machine learning and network propagation models to better identify the different actors involved in the dissemination of misinformation. Though an interdisciplinary approach is required to educate the different actors in a network to combat fake news, ultimately, we must admit that the fundamental nature of our network needs to change. In order to minimize the influence of bad actors, it is important for misinformation to be dealt with at the source. Machine learning models can make use of quality news data by quickly detecting patterns among the behavior of credible news in contrast to fake news. It is of equal importance that misinformation that leaks through a systematic filter to have limited outreach. Coincidentally, algorithms that maximize the diversity of information diffusion have this feature built-in. For future work by academic researchers, and a better understanding of the different actors involved in social networks, it is pivotal for there to be a movement for transparency among the different social media platforms, so that there is public access to clean, reliable, high-quality data.

REFERENCES

Asr, F. T., & Taboada, M. (2019). Big data and quality data for fake news and misinformation detection. *Big Data & Society*, 6(1). https://doi.org/10.1177/2053951719843310

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *In Nature Communications*, 10(1). https://doi.org/10.1038/s41467-018-07761-2

Callon, M. (1984). Some elements of a sociology of translation: domestication of the scallops and the fishermen of St Brieuc Bay. *The Sociological Review*, 32(1), 196–233. https://doi.org/10.1111/j.1467-954X.1984.tb00113.x

Chan, M. S., Jones, C. R., Jamieson, K. H., & Albarracín, D. (2017). Debunking: A meta-analysis of the psychological efficacy of messages countering misinformation. *Psychological Science*, 28(11), 1531-1546. https://doi.org/10.1177/0956797617714579

Gallo, J. A., & Cho, C. Y. (2021). Social media: Misinformation and content moderation issues for Congress. *Congressional Research Service*. https://crsreports.congress.gov/product/pdf/R/R46662

Gradoń, K. T., Hołyst, J. A., Moy, W. R., Sienkiewicz, J., & Suchecki, K. (2021). Countering misinformation: A multidisciplinary approach. *Big Data & Society*, 8(1). https://doi.org/10.1177/20539517211013848

Greifeneder, R., Jaffé, M. E., Newman, E. J., & Schwarz, N. (Eds.). (2020). The Psychology of Fake News. *Routledge*. https://doi.org/10.4324/9780429295379

Horrigan, J. B. (2017). How people approach facts and information. *Pew Research Center*. https://www.pewresearch.org/internet/2017/09/11/how-people-approach-facts-and-information/

Ivory, D., Leatherby, L., & Gebeloff, R. (2021). Daily report: Least vaccinated U.S. counties have something in common: Trump voters. *The New York Times*. https://www.nytimes.com/interactive/2021/04/17/us/vaccine-hesitancy-politics.html

Kadih, S. (2022). *Misinformation Actor Network*. [Figure 1]. *STS Research Paper*: Misinformation in government and society (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Kadih, S. (2021). *CSI model specification*. [Figure 2]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Kadih, S. (2021). *Standardized coefficients*. [Figure 3]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Kadih, S. (2021). *Model of information dissemination*. [Figure 4]. *Prospectus* (Unpublished undergraduate thesis). School of Engineering and Applied Science, University of Virginia. Charlottesville, VA.

Prasetya, H. A., & Murata, T. (2020). A model of opinion and propagation structure polarization in social media. *Computational Social Networks, 7(1).* https://doi.org/10.1186/s40649-019-0076-z

Qian, S., Hu, J., Fang, Q., & Xu, C. (2021). Knowledge-aware multi-modal adaptive graph convolutional networks for fake news detection. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 17(3), 1–23. https://doi.org/10.1145/3451215

Ruchansky, N., Seo, S., Liu, Y. (2017). CSI: A hybrid deep model for fake news detection. *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 797-806. https://doi.org/10.1145/3132847.3132877

Steinmetz, K. (2018, August 9). How your brain tricks you into believing fake news. Time Magazine, 192(7).

Stoica, A.-A., & Chaintreau, A. (2019, May 13). Fairness in social influence maximization. companion proceedings of the 2019 World Wide Web Conference. *WWW '19: The Web Conference.* https://doi.org/10.1145/3308560.3317588

Tam, P. (2017). Daily report: The continued creation and dissemination of fake news. *The New York Times*. https://www.nytimes.com/2017/01/19/technology/daily-report-the-continued-creation-and-dissemination-of-fake-news.html

Valenzuela, S., Halpern, D., Katz, J. E., & Miranda, J. P. (2019). The paradox of participation versus misinformation: Social media, political engagement, and the spread of misinformation. *Digital Journalism*, 7(6), 802-823. https://doi.org/10.1080/21670811.2019.1623701

Vallor, S. (2021). Social networking and ethics. *The Stanford Encyclopedia of Philosophy*, (Fall 2021 Ed.). https://plato.stanford.edu/archives/fall2021/entries/ethics-social-networking