Epidemiology Inspired Framework for Fake News Mitigation in Social Networks

Bhavtosh Rath, Jaideep Srivastava

University of Minnesota, USA

Abstract

Research in fake news detection and prevention has gained a lot of attention over the past decade, with most models using features generated from content and propagation paths. Complementary to these approaches, in this position paper we outline a framework inspired from the domain of epidemiology that proposes to identify people who are likely to become fake news spreaders. The proposed framework can serve as motivation to build fake news mitigation models, even for the scenario when fake news has not yet originated. Some models based on the framework have been successfully evaluated on real world Twitter datasets and can provide motivation for new research directions.

Keywords

Fake news spreaders, Social networks, Epidemiology.

1. Introduction

The wide adoption of social media platforms like Facebook, Twitter and WhatsApp has resulted in the creation of behavioral big data, thus motivating researchers to propose various computational models for combating fake news. So far the focus of most research has been on determining veracity of the information using features extracted manually or automatically through techniques such as deep learning. We propose a novel fake news prevention and control framework that incorporates people's behavioral data along with their network structure. Like in epidemiology, models proposed within the framework cover the entire life cycle of spreading: i.e. before the fake news originates, after the fake news starts spreading and containment of its further spreading. The framework is not to be confused with popular information diffusion based models [1] because they a) usually categorize certain nodes and cannot be generalized to all nodes, b) consider only the propagation paths but not the underlying graph structure and c) can be generalized to information diffusion and need not be particular to fake news spreading.

Related Work: Literature of research in fake news detection and prevention strategies is vast, and can be divided broadly into three categories: Content-based, Propagation-based and User-based.

In *content-based* approach the problem is formulated as identifying whether content of a spreading information is fake or not. Most proposed models rely on using linguistic or visual based features. While earlier work relied mostly on hand engineering relevant features, more recently deep learning based models have gained popularity as they can automatically generate relevant features. *Propagation based* approaches consider propagation paths of fake news and are mostly inspired from information diffusion and cascade models. They are used to understand how information spreading patterns can help distinguish fake news from true news. These models are usually integrated with content-based features to improve prediction performance. Majority of computational models for fake news detection from these two categories are summarized in [2]. *User-based* approaches focus more on peoples' psychology. While user-specific features can be included as part of content-based models, there has also been some research exploring behavior patterns of individuals who spread fake news. Behavioral principles like naive realism and confirmation bias (at individual level) have been found to make fake news perceived as true, as stated in [3]. A phenomenon called echo chamber effect (at group level) has also been found to reinforce people's pre-existing biases, making them averse to accepting opposing opinions [4]. The role of bots in fake news spreading has also been studied. More recently work has been done to identify fake news spreaders [5] which focus on modelling linguistic features but they do not integrate underlying network structure. *Not many computational models have been proposed exploring psychological concepts from historical behavioral data that make people vulnerable to spreading fake news, which our proposed framework can be used to address.*

Proceedings of the CIKM 2020 Workshops. October 19-20, Galway, Ireland.
Editors of the Proceedings: Stefan Conrad, Ilaria Tiddi.
EMAIL: rathx082@umn.edu (B. Rath); srivasta@umn.edu (J. Srivastava)
ORCID:

Table 1
Mapping Epidemiological concept to fake news spreading.

	Epidemiology context	Fake news spreading context
Infection	Infection	Fake news
Population	People and communities	Nodes and modular sub-graphs
Vulnerable	Likely to become infection carriers	Likely to become fake news spreaders
Exposed	Neighbors are infected	Neighbor nodes are fake news spreaders
Spreaders	Infected people	Fake news spreaders
Prevention	Medication	Refutation news
Control	Immunization	Refutation news
Recovered	Infection cured	Retract fake news and/or spread refutation news

A major limitation with existing models is that they rely on the presence of fake news to generate meaningful features, thus making it difficult to model fake news mitigation strategies. Our framework proposes models using two important components that do not rely on the presence of fake news: underlying network structure and people's historical behavioral data.

The rest of the paper is divided as follows: We explain how epidemiological concepts can be mapped directly to the problem of fake news spreading and mitigation. We then explain proposed models for detecting fake news spreader using the Community Health Assessment model and also summarize current and future research based on the ideas. Finally we give our concluding remarks.

2. Epidemiology Inspired Framework

Epidemiology is the field of medicine which deals with the incidence, distribution and control of infection among populations. In the proposed framework fake news is analogous to infection, social network is analogous to population and the likelihood of people believing a news endorser in the immediate neighborhood is analogous to their vulnerability to getting infected when exposed. We consider fake news as a pathogen that intends to infect as many people as possible. An important assumption we make is that fake news of all kinds is generalized as a single infection, unlike in epidemiology where people have different levels of immunity against different kinds of infections (i.e. the framework is information agnostic). Also we do not distinguish bots in the network population.

The likelihood of a person getting infected (i.e. believing and spreading the fake news) is dependent on two important factors: a) the likelihood of trusting a news endorser (a person is more likely to spread a news without verifying its claim if it is endorsed by a neighbor they trust); and b) the density of its neighborhood, similar to how high population density increases the likelihood of infection spreading, a modular network structure is more prone to fake news spreading. After the infection spreading is identified there is a need to de-contaminate the population. A medicinal cure is used to treat the infected population and thus prevent further spreading of infection. In the context of fake news, a refutation news can serve this purpose. Refutation news can be defined as true news that fact-checks a fake news. Contents from popular fact-checking websites[1] are examples of refutation news. In epidemiology the medicine can have two purposes: As control mechanism (i.e. medication), with the intention to cure infected people (i.e. explicitly inform the fake news spreaders about the refutation news) and as prevention mechanism (i.e. immunization), with the intention to prevent uninfected population from becoming infection carriers in future (i.e. prevent unexposed population from becoming fake news spreaders). An infected person is said to have recovered if he either decides to retract from sharing the fake news or decides to share the refutation news, or both. Mapping of epidemiological concepts to the context of fake news spreading is summarized in Table 1.

3. Contributions

In this section we show how the framework has been applied so far and how it is used to propose relevant models.

3.1. Community Health Assessment model

A social network has the characteristic property to exhibit community structures that are formed based on

[1] https://www.snopes.com/, https://www.politifact.com/

Table 2
Neighbor, boundary and core nodes for communities in Figure 1.

com	$\mathcal{N}_{com}$	$\mathcal{B}_{com}$	$\mathcal{C}_{com}$
1	D_2	C_1	A_1,B_1,E_1,D_1,F_1,G_1
2	A_6,E_6	C_2,D_2	A_2,B_2,E_2,F_2
3	D_1,D_5,E_6	A_3,C_3	B_3,D_3,E_3,F_3
4	D_3	C_4	A_4,B_4,D_4,E_4,F_4
5	D_4,D_8,E_8	D_5,A_5,C_5	E_5,B_5
6	A_5	D_6	A_6,B_6,C_6,E_6
7	B_6	A_7	B_7,C_7,D_7,E_7,F_7, G_7
8	$F7$	A_8	B_8,C_8,D_8,E_8,F_8

inter-node interactions. Communities tend to be modular groups where within-group members are highly connected, and across-group members are loosely connected. Thus members within a community would tend to have a higher degree of trust among each other than between members across different communities. If such communities are exposed to fake news propagating in its vicinity, the likelihood of all community members getting infected would be high. Thus it is important to identify vulnerable individuals that lie in the path of fake news spread to limit the overall spreading of fake news in the network. The idea is illustrated in Figure 1. In the context of Twitter, directed edge $B_1 \longrightarrow A_1$ represents B_1 follows A_1. Thus information flows from A_1 to B_1 when B_1 decided to retweet an information endorsed by A_1. The goal would be to identify nodes that are likely to believe and spread the fake news. Subscript of the nodes denote the community they belongs to. Motivated by the idea of ease of spreading within a community we proposed the Community Health Assessment model. The model identifies three types of nodes with respect to a community: neighbor, boundary and core nodes, which are explained below:

1. Neighbor nodes: These nodes are directly connected to at least one node of the community. The set of neighbor nodes is denoted by $\mathcal{N}_{com}$. They are not a part of the community.

2. Boundary nodes: These are community nodes that are directly connected to at least one neighbor node. The set of boundary nodes is denoted by $\mathcal{B}_{com}$. It is important to note that only community nodes that have an outgoing edge towards a neighbor nodes are in $\mathcal{B}_{com}$.

3. Core nodes: These are community nodes that are only connected to members within the community. The set of core nodes is denoted by $\mathcal{C}_{com}$.

The neighbor, boundary and core nodes for communities in Figure 1 are listed in Table 2.

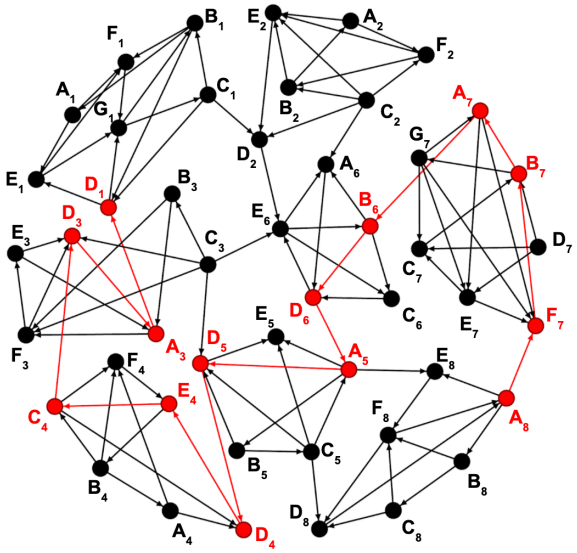


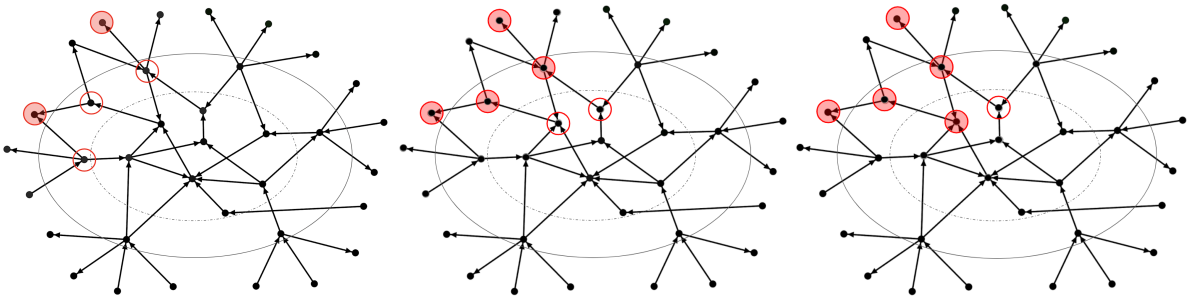
Figure 1: Motivating example. Red nodes denote fake news spreaders.

3.2. Assessment, identification and prevention

To model a person's likelihood to endorse a fake news based on their belief in the endorser, we applied the Trust in Social Media (TSM) algorithm. It assigns a pair of complementary trust scores, called *Trustingness* and *Trustworthiness* to every node in a social network. While trustingness quantifies the propensity of a node to trust its neighbors, trustworthiness quantifies the willingness of the neighbors to trust the node. Implementation details for the algorithm can be found in [6]. Below we propose three phases for the framework and summarize models implemented so far with future directions.

1. Vulnerability assessment of population: In epidemiology, it is important to identify individuals and groups that are vulnerable to fake news before the spreading begins. Borrowing ideas from the community health assessment model, we proposed metrics that quantify the vulnerability of nodes and communities in a network. Through experiments on real world information spreading networks on Twitter, we showed that our proposed metrics are more effective in identifying fake news spreaders compared to true news spreaders, confirming our hypothesis that fake news relies strongly on inter-personal trust to propagate while true news does not. Details regarding the model implementation can be found in [7].

2. Identification of fake news spreaders: While determining the veracity of information has been widely



(a) Fake news reaches $\mathcal{N}_{com}$ (b) Fake news reaches $\mathcal{B}_{com}$ (c) Fake news reaches $\mathcal{C}_{com}$

Figure 2: Community health assessment model perspective for fake news prevention and control.

researched, it is equally important to determine the authenticity of the people who are spreading information. A model for automatic identification of people spreading fake news by leveraging the concept of *Believability* (i.e. the extent to which the propagated information is likely to be perceived as truthful) is proposed. With the retweet network edge-weighted by believability scores, network representation learning is used to generate node embeddings, which is leveraged to classify users as fake news spreaders or not using a recurrent neural network classifier. Based on experiments on a very large real-world rumor dataset collected from Twitter, we could effectively identify false information spreaders. Further details can be found in [8].

3. Prevention and control of infection spreading: Motivation for this problem can be explained through Figure 1. D_1, a neighbor node for community 3 is a fake news spreader. Node A_3, a boundary node is exposed and likely to start fake news spreading in community 3. To prevent such a scenario it is important to predict boundary nodes of all communities in a network that are likely to become fake news spreaders when the infection has reached neighbor nodes. Similarly, consider the scenario where A_3 is a fake news spreader. Members of the community B_3, D_3 and E_3 which are immediate followers of A_3 are now exposed to the fake news, and the remaining community members are two steps away. Due to their close proximity they too are vulnerable to believing A_3 and causing infection to spread throughout the community. Thus it is important to identify core nodes that would become likely spreaders when the infection has reached boundary nodes. The scenarios are explained in Figure 2 applying the community health assessment model. Nodes inside the dotted oval denote core nodes, between dotted and solid oval denote boundary nodes and outside the solid oval denote neighbor nodes. (a) shows the scenario where fake news has reached the two neighbor nodes (highlighted in red). Three boundary nodes (circled in red) are exposed to the fake news. In (b) two out of three exposed boundary nodes become spreaders, and marks the beginning of fake news spreading within the community. And in (c), one of the two exposed core nodes become spreader.

Thus using community health assessment model we can build models that predict both exposed (i.e. boundary nodes) and unexposed (i.e. core nodes) nodes that would likely become fake news spreaders after infection spreading has begun (i.e. fake news has reached neighbor nodes). Effective mitigation strategies could then be deployed against predicted spreaders.

4. Conclusion

In this position paper we proposed a novel epidemiology inspired framework and showed how the community health assessment model can be used to build models for fake news mitigation, a problem less explored compared to fake news detection. What makes it different from most existing research is that a) it proposes a more spreader-centric modelling approach instead of content-centric approach, and b) it does not rely on features extracted from fake news thus serving as motivation to build fake news mitigation strategies, even for the scenario when fake news has not yet originated. Recent work that apply few of the ideas have shown encouraging results, thus serving as motivation to pursue the idea further. A limitation of our model is that it does not not incorporate the dynamic nature of social network structure. As part of future work we would like to incorporate eliminating the presence of bots as we are focusing on modeling psychological and sociological properties based on behavioral data.

References

[1] F. Jin, E. Dougherty, P. Saraf, Y. Cao, N. Ramakrishnan, Epidemiological modeling of news and rumors on twitter, in: Proceedings of the 7th workshop on social network mining and analysis, 2013, pp. 1–9.

[2] K. Sharma, F. Qian, H. Jiang, N. Ruchansky, M. Zhang, Y. Liu, Combating fake news: A survey on identification and mitigation techniques, ACM Transactions on Intelligent Systems and Technology (TIST) 10 (2019) 1–42.

[3] K. Shu, A. Sliva, S. Wang, J. Tang, H. Liu, Fake news detection on social media: A data mining perspective, ACM SIGKDD explorations newsletter 19 (2017) 22–36.

[4] M. Del Vicario, G. Vivaldo, A. Bessi, F. Zollo, A. Scala, G. Caldarelli, W. Quattrociocchi, Echo chambers: Emotional contagion and group polarization on facebook, Scientific reports 6 (2016) 37825.

[5] J. Bevendorff, B. Ghanem, A. Giachanou, M. Kestemont, E. Manjavacas, M. Potthast, F. Rangel, P. Rosso, G. Specht, E. Stamatatos, et al., Shared tasks on authorship analysis at pan 2020, in: European Conference on Information Retrieval, Springer, 2020, pp. 508–516.

[6] A. Roy, C. Sarkar, J. Srivastava, J. Huh, Trustingness & trustworthiness: A pair of complementary trust measures in a social network, in: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2016, pp. 549–554.

[7] B. Rath, W. Gao, J. Srivastava, Evaluating vulnerability to fake news in social networks: A community health assessment model, in: 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2019, pp. 432–435.

[8] B. Rath, W. Gao, J. Ma, J. Srivastava, Utilizing computational trust to identify rumor spreaders on twitter, Social Network Analysis and Mining 8 (2018) 64.