Surveying Efforts to Model the Propagation of Misinformation on Social Media

A Technical Report submitted to the Department of Computer Science

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

Shomik Ghose

Spring, 2020.

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

Anil Vullikanti, Department of Computer Science

David Wu, Department of Computer Science

Surveying Efforts to Model the Propagation of Misinformation on Social Media

Shomik Ghose
sg3tq@virginia.edu
University of Virginia
Charlottesville, Virginia, USA

ABSTRACT

Misinformation and disingenuous journalism are by no means a new phenomenon; however, the advent of digital information systems like social media have revolutionized the methods of content consumption and production alike, with broader implications for how misinformation develops and spreads. To enact effective strategies for mitigating the repercussions of misinformative content, there must first be a robust understanding of how misinformation propagates. Towards this end, researchers have leveraged advances in machine learning and artificial intelligence to conceptually model the spread of misinformation. In this literature review, we taxonomize recent models of misinformation propagation as "pathologic, sociogrammatic, or evolutionary" to take stock of current evidence and inform future directions. Future modelling efforts should focus on integrating interdisciplinary insights to better understand how sociodemographic factors nuance the spread of misinformation.

1 INTRODUCTION

Since its introduction in the early 2000's, social media has served as an effective tool to connect with like-minded individuals, spread information, and share ideas. More than ever, social media users feel better informed about current events, more intimately connected with friends and family, and uniquely empowered to effect sociopolitical change [25]. Unfortunately, due to its popularity and accessibility, the social media landscape has become increasingly polluted with a cacophony of opinion, dubious news reporting, and false information. Experts at Pew Research Center suggest that up to two of every five news articles shared on Facebook contain significant misinformative elements [2]. Simultaneously, Simultaneously, social media has become influential than ever: in 2018, 36% of US young adults ages 18-29 reported receiving their news primarily from social media outlets like Twitter and Facebook, a drastic increase compared to older age groups and previous years [26]. The juxtaposition of these two effects underscores the importance of challenging the flood of misinformation which plagues social media platforms.

1.1 Defining Misinformation

As there are many terms which relate to misinformation, we first define an operational definition of for misinformation. Wu et al. defines misinformation as *inaccurate* information that is *unintentionally* propagated by users [16]. Though often used synonymously, misinformation is distinct from disinformation which is fake information intentionally developed and propagated to mislead people [27]. While both involve the spread of false information, there is a key difference in the underlying intent. In practice, the ease of posting content to social media makes it difficult to ascertain if misinformative content is spread intentionally or not. For instance, someone posting material opposing vaccination may due so because of real safety concern or in an attempt to undermine trust in medical systems.

The term "fake news" experienced notable media attention in the wake of the 2016 US presidential election and overlaps with misinformation and disinformation. Fake news is distinct in that it predominantly arises from major media sources, but lacks a clear agreed upon definition [27, 31]. As fake news can encompass both satire, disinformation, and misinformation it is frequently employed without precision and we accordingly refrain from using it throughout this paper. Thus, excepting cases with clear malicious intent, we apply "misinformation" as an umbrella term to broadly encompass all forms of false information.

1.2 The Repercussions of Misinformation

Social media misinformation has already had tangible effects throughout society. Users can use social media immediately post opinions and content while events occur, making it difficult to distinguish sharing of official information from personal opinion, thereby reducing the overall integrity of the news [15]. For instance, during the 2016 US Presidential election, Americans were exposed to a wide variety of misinformative content including claims of voter fraud [1, 10]. Many experts believe that this false information had a significant effect on the electoral outcome of the election [1, 5]. Some of these unmitigated misinformative messages have even grown to become movements like Q-anon which boasts thousands of believers.

The repercussions of misinformation are also highly visible in healthcare: though connections between the MMR vaccine and autism have been thoroughly debunked, 21% of US parents still express substantial doubts about vaccine safety [7, 22]. In 2017, these misinformation-based choices drove a 31% increase in vaccine-preventable disease, resulting in thousands of lives lost and billions of dollars in expenditures [11, 22]. Even during the world-wide COVID19 pandemic, Facebook groups contesting the legitimacy of the virus and protesting the wearing of masks posed significant challenges to efforts to mitigate the epidemic. Estimates suggest that about 10-15% of adults in the US and Canada refused to wear masks, leading to thousands of additional infections and animosity towards healthcare professionals [29]. Recognizing the escalating role and consequences of misinformation in recent events, including the UK Brexit referendum, the 2016 U.S. presidential election, and the COVID19 pandemic, the World Economic Forum ranks the spread of digital misinformation as one of the foremost threats to global development [12].

Table 1: Summary of articles under consideration

Title	Year	Synopsis
Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks	2015	Translating traditional compartmental epidemiology models to understand the virus-like spread of misinformation
Using an Epidemiological Model to Study the Spread of Misinformation during the Black Lives Matter Movement	2021	Translating a newer epidemiological model to study misinformation during the 2020 DC Riots
Limiting the Spread of Misinformation in Social Networks	2011	Formalizing the *eventual influence limitation* problem and developing algorithmic solutions
The Spreading of Misinformation online: 3D Simulation	2018	Using sociograms to study the potential impact of a novel information verification technique
A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis	2018	Applying phylogeny to misinformative tweets to understand how they evolve and propagate

2 RELATED WORK

The literature confirms that social media is one of the most prolific sites for the spread of misinformation. Social media has been documented as a site of spread for misinformation regarding medicine, science, politics, and current events. Despite the significant risk that misinformation poses, many of our current interventions remain ineffective. For instance, the American Medical Association has put significant resources towards disseminating of peer-reviewed research, publishing expert fact-checked reports, and organizing publicity campaigns, but the number of individuals who believe in medical misinformation continues to steadily grow [30]. The inefficacy of these measures can be attributed to gaps in our understanding of the mechanisms which cause modern misinformation to propagate.

Conceptualizing the processes by which misinformation spreads can reveal important mechanical nuance while revealing gaps in our understanding. Reviews across different fields form the basis of our current conception of misinformation. Psychology indicates that the common culprits of partisanship and political motivations are surprisingly disconnected from the spread of misinformation [9, 21]. Instead, belief heuristics play a large role in determining how individuals judge their information. Sociology indicates that demographic factors like wealth, age and culture [4, 8] play significant roles in the propagation of misinformation through communities.

With the advent of COVID-19, there has been a significant push in the computer science community to understand the torrent of misinformative content being produced. Recent studies have elucidated characteristics of misinformative text [14, 32], studied user attention patterns [19] and exploited recurring motifs to identify false content [14, 18, 20]. However, there has not yet been a comprehensive review of these contributions to understand commonalities and patterns in these efforts. Towards this end, we seek to produce, to our knowledge, the first comprehensive review of computational modelling techniques for the propagation of misinformation on social media.

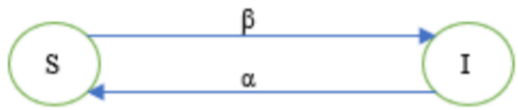
3 RESULTS

We seek to understand the findings and commonalities between the set of articles presented in 1. These articles were selected for significance and recency of the development. This therefore allows us us to understand where modelling began and how it has evolved. We taxonomize these different models as "pathologic, sociogrammatic, or evolutionary" and consider them within that framework.

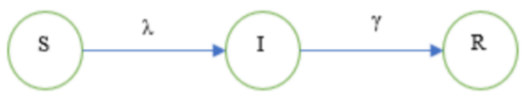
3.1 Pathologic Models

One of the first techniques used to model the spread of misinformation was to apply techniques used to predict the spread of real infectious diseases. This class of models, which we will term “pathological,” consider misinformation to be a communicable virus and focuses on modelling how users respond and adapt to exposure. The first attempts at developing pathological models drew inspiration directly from epidemiological mathematical models. In classical disease systems, most individuals unexposed and thus definitionally *susceptible* (S) until they are exposed to the contagion and become *infected* (I). Individuals afflicted by the virus may then recover (R) through proper treatment. These basic states form the basis for the two foundational compartmental epidemic models: *susceptible-infected-recovered* (SIR) and *susceptible-infected-susceptible* (SIS) models where the dynamics of the system are governed by ordinary differential equations (Fig. 1) representing the interconversions between these states [3]. While these models are effective at modelling infectious diseases, they cannot fully to capture the nuances of misinformation. One major limitation is that when someone in the *Susceptible* component is exposed to the disease, they can only transition to *Infected* state. This assumption translates poorly to the spread of misinformation, as individuals do not unilaterally nor instantaneously accept the information they are presented with and can require further exposure or even express skepticism.

To address this insufficiency, several new pathological models have been developed to specifically address the unique characteristics of misinformation spread on social media platforms. For instance, Tambuscio et al. propose a stochastic epidemic model derivative of SIS which seeks to model the unique social media phenomenon wherein the dissemination of misinformation is simultaneously accompanied by active debunking and fact-checking [28]. Unlike a real virus, where exposure invariably results in infection, per the authors’ new model, upon encountering misinformation users can either become ‘infected’ with the hoax or they can become non-believers and fight the propagation of falsehoods. After some



(a) SIS Model



(b) SIR Model

Figure 1: Basic compartmental models

period of time, both the believers and nonbelievers can return to a susceptible state. As such, the Tambuscio model can be interpreted as an SIS model wherein the infected compartment is partitioned into the misinformed *believers* (B) and the skeptic *fact-checkers* (F). These modifications to the traditional SIS approach are depicted below in Fig. 2:

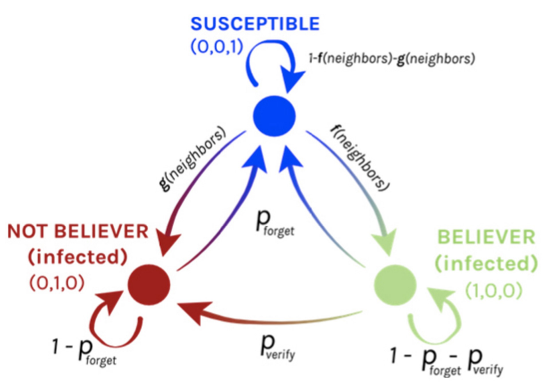


Figure 2: Tambuscio's Modified SIS Model

When considering applying this model to some arbitrary graph $G = (V, E)$ we observe three phenomena which occur in a given node $i \in V$:

(1) **Spreading** $[S \rightarrow B, S \rightarrow F]$: The way a susceptible individual becomes a believer or nonbeliever. The debunking process, g_i and infection process, f_i, are dependent on the believer and nonbeliever status of their neighbors and their own gullibility. Notably, $g_i(t) + f_i(t) = \beta$, where β represents the misinformation spreading rate of the classic SIS model. The balance between these two process is governed by the user's gullibility α_i.
(2) **Forgetting** $[B- > S, F- > S]$: Irrespective of a user's belief, the fast-paced nature of social media naturally leads some users to discard their belief status. Represented here with p_{forget}, if we jointly consider the B and F compartments to be the singular unique *Infected* state, we see that we recover the exact SIS model with p_{forget} equivalent to recovery probability α (Fig 1a).
(3) **Verifying** $[B- > F]$: A process unique to the Tambuscio model, wherein a believer fact-checks their information and accordingly corrects their misinformed beliefs, becoming a *fact-checker*. This process is represented by p_{verify} and is pivotal to better understanding how to mitigate the spread of misinformation.

Thus, we identify four key parameters: spread rate, β; gullibility, α_i; the likelihood of forgetting, p_{forget}; and likelihood of verifying, p_{verify}. The likelihood of information verification is particularly interesting, as the authors postulate that with sufficient fact-checking, misinformation can be purged from a network, irrespective of user gullibility. To validate their model, they apply it to heterogeneous (scale-free), homogeneous (random), and real Facebook networks. These practical demonstrations reveal that even small amounts of fact-checking behavior can entirely eradicate misinformation from networks irrespective of their topology in spite of high user gullibility. Finally, deriving mean-field equations for the model, the authors analytically identified a threshold for p_{verify} to ensure the eradication of misinformative content.

There have also been several alternative approaches to reconcile the differences between real infectious diseases and misinformation. The *susceptible-exposed-infected-skeptic* (SEIZ) model is one such popular attempts which looks to expand the number of compartments rather than subdividing the infected compartment as the Tambuscio model does. When a *susceptible* (S) individual encounters misinformation, they can become *infected* (I) and immediately propagate the hoax as in the SIS model, or they could instead be *exposed* (E) and require more time and information before deciding to spread the misinformation. Finally, a susceptible individual can also become a *skeptic* (Z) and not spread the misinformation; note that status as a skeptic does not imply verification as in the Tambuscio model, but instead reflects the choice to not share the encountered content. As with the SIS and SIR models, the SEIZ model is represented by a set of ordinary differential equations. The key advantage of the SEIZ model over the other compartmental SIS and SIR models is that a user can come in to contact with misinformation and either choose not to execute a reaction, or deliberate before taking action. This is particularly useful for modelling social media interactions on a platform like Twitter, where users are exposed to infinite streams of content with varying rates of engagement. Maleki et al. were the first to recognize this potential and translate the SEIZ model for the study of misinformation, accordingly developing the representation depicted in Fig 3 [17].

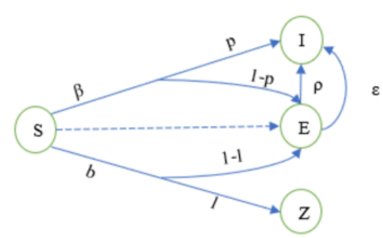


Figure 3: Maleki et al.'s SEIZ model

When considering applying this model there are three phenomena of importance

(1) **Initial Exposures** $[S->I, S->E]$: Users who have never heard of the hoax encountering *infected* users. The exposure rate β is analogous to the misinformation spreading rate of the classic SIS model. These users either immediately become *infected* by passing on the information with probability p or they can send time deliberating and become *exposed* with probability $1-p$.
(2) **Repeat Exposures** $[S->Z, S->E]$: Users who are already aware of the hoax encountering *skeptic* users. Individuals are recruited from the susceptible population with rate b and resolve not to share the misinformation and become *skeptics* with probability l or spend time deliberating and becoming exposed with probability $1-l$.
(3) **Exposure Resolution** $[E->I]$: Exposed individuals have encountered the hoax, wait until they resolve their beliefs to spread the hoax. This can occur in two different ways: exposed users encounter additional infected users with rate ρ and become infected, or users privately deliberate and decided to independently adopt the belief and become infected with rate ϵ.

Thus, we obtain six key parameters: contact rates between susceptible and infected users, β; between susceptible and skeptic users, b; and between exposed and infected users ρ; probabilities of expressing immediate skepticism l, immediately becoming infected p, and self-adoption of misinformation ϵ. These parameters outline three different methods for users to become infected, untangling the significance of these different routes can help to better understand specifically how people become infected.

To demonstrate the efficacy of the SEIZ model, Maleki et al. apply it to dissemination of misinformation regarding blackouts during the June 1st Washington, DC riots. In this context, infection constitutes spreading the hoax that the riots resulted blackouts using the hashtag #DCBlackOut. Their work indicated that SEIZ models were a better representation of misinformative processes than traditional SIS models. Furthermore, an analysis of the derived parameters revealed that immediate infection and contact between exposed and infected users was rare – the majority of users who became infected were first exposed and through deliberating the presented information, eventually self-adopted (ϵ). These results underscore the idea that infected users are not primarily created through heavy contact with infected peers, and instead perform a notable amount of private consideration during exposure. This highlights the importance of the verification action advocated for by Tambuscio et al and calls for more research to understand the factors leading to self-adoption [28].

3.2 Sociogrammatic Models

Another approach to modelling the spread of misinformation is to directly consider networks of interconnected actors or *sociograms* and apply concepts from networks theory to understand the mechanics of how misinformative campaigns propagate. These types of models generally adopt a more abstract approach to the problem of misinformation, and many contributions draw on algorithms concepts and seek to formalize and explore interesting subproblems within these networks. These models are popular as they are highly intuitive and their findings can be logically translated to actionable policies.

In 2011, Budak, Agrawal, and Abbadi helped to lay the groundwork for this subfield by presenting the first attempts to formalize the problem of limiting misinformation spread in social networks [6]. With the sheer volume content on social media, disseminating accurate information in response to unmitigated misinformation is extremely challenging. With the overarching goal of making social media a more reliable source of information, the authors seek to algorithmically determine an optimal method of disseminating accurate verification information which can mitigate the efficacy of misinformation spread throughout a network. To achieve this, they conceptualize a social network as a directed graph $G=(V,E)$. When a node v attempts to inform or misinform neighbor w along edge $e_{v,w} \in E$ it succeeds with probability $p_{v,w}$. To better understand their model and findings, we will use an illustrative sample network:

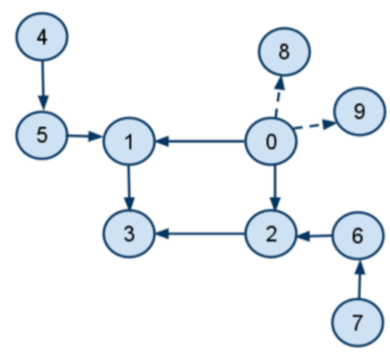


Figure 4: A sample network for understanding the *EIL* problem

In this example network, we may consider node 0 as an origin for a misinformation campaign C and expect this individual to misinform its neighbors (1, 2, 8, 9). We model the efficacy of these outreach attempts with a realization of $p_{v,w}$, indicating successes with solid lines (live edges) and failures with dotted lines (blocked edges). Thus, we expect node 0 to successfully misinform nodes 1 and 2 which in turn influence node 3. With this framework established, the authors pose the question: how can one best design a competing verification campaign L intended to limit the spread of misinformative campaign C. Campaign L shares the same outreach mechanism $p_{v,w}$ as misinformative piece C and once a node is claimed by either campaign, they cannot be reclaimed. They specifically look to minimize the number of nodes in the graph which become infected by campaign C at the conclusion of the information cascades and formalize this problem as the *eventual influence limitation problem (EIL)*.

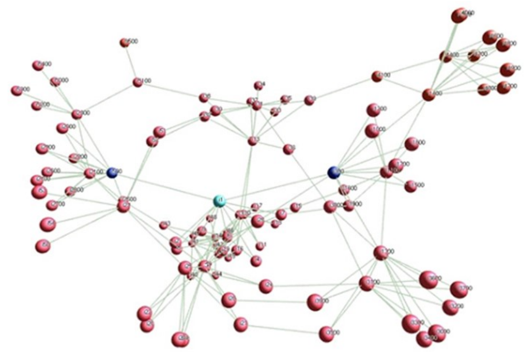
The authors show that *EIL* is NP-hard using an instance of the set-cover problem; however, given the simplifying assumption of $p_{v,w}=1$, they prove *EIL* to be submodular and monotone. This enables the use of a hill-climbing approach to develop a polynomial time greedy algorithm capable of selecting a set of nodes which would be most effective starting points for verification campaign L.

Noting the infeasibility of this expensive approach for real life social media networks, they develop several heuristics to test on real data sets of Facebook networks, finding that simple degree centrality heuristic exhibits performance comparable to the greedy solution. Furthermore, they find that increasing the number of L starting nodes failed to significantly change the permeance of C throughout real networks, rather starting the campaign as soon as possible was far more effective. Taken together, these results suggest that the most intuitive way to approach the *EIL* problem is by identifying few influential individuals (per high degree centrality) and having them intervene as early as possible.

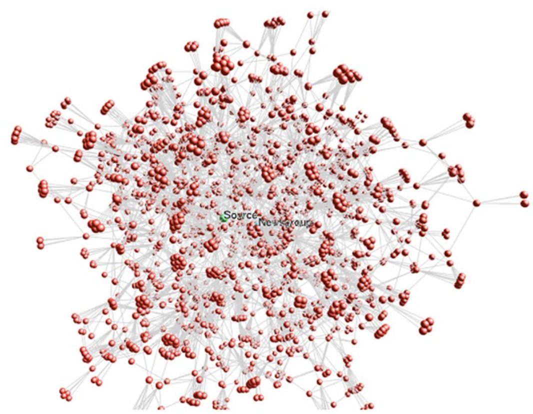
The *EIL* was the first of many questions asked of such conceptual social media networks which help to explore the properties of misinformation propagation. In 2018, Pourghomi, Dordevic, and Safieddine looked to similarly use an abstraction of a social network to investigate the impact of a particular verification method called *"click to authenticate"* [23]. Inspired in part by the work of Budak et al., the authors sought to understand how an easily accessible decentralized method of information verification may impact misinformation spread throughout a network. *Click to authenticate* is a theoretical verification intervention which enables users to right-click on a piece of media to access a real-time check for concurring reports, source metadata, and crowd-sourced feedback. A 2D modelling effort which looked at discrete time steps during the spread of misinformation revealed that such a tool could potentially combat misinformation [24]; in this study, the authors look to take the next step by demonstrating efficacy in a realistic spatial network.

Towards this end, the Pourghomi et al. used Biolayout, a popular biological network simulation tool to simulate several realistic scenarios. To understand the information verification action, they add several parameters to the existing directed graph $G = (V, E)$. In addition to the preexisting edge-related spreading probability parameter $p_{v,w}$, the model additionally includes parameter A representing the probability of a given node to immediately use "Click to authenticate" upon encountering new information, a parameter Cw representing the probability that a given node will authenticate given conflicting information, and a final parameter Rv indicating the chance that these authentication actions can correct a misinformed belief. Several simulations are then run with on randomly generated networks with various parameter values.

The authors notably confirm a key finding of the *EIL* problem, noting that if random users take an active interest in validating information, we can correct the beliefs approximately 30% of misinformed users, as represented by the red nodes (Fig. 5a). However, if we apply the greedy algorithm derived from the *EIL* problem to determine optimal authenticators we can save up to 54% of misinformed users (Fig. 5b). This finding helps to powerfully reinforce the idea that selecting correct influential users can greatly improve mitigation of misinformation. Additionally, the results suggest that for this specific network formulation, the population size becomes irrelevant as the extremities of the spreading trees are almost unilaterally saved through individual authentication and sharing, and thus misinformation is primarily confined to local subsections of the graph. This model is still in the process of being formalized and requires additional research to demonstrate practical relevance.



(a) Network without Influencers



(b) Network with Influencers

Figure 5: Networks subject to 30% verification. Red nodes are saved and blue are infected

3.3 Evolutionary Models

Unlike pathological models which focus on user reactions to encountering misinformation and sociogrammatic models which look at broader community interactions, evolutionary models seek to understand how misinformation itself changes during its spread to understand the associated propagation patterns. These kinds of models are a more recent development compared to pathological and network models and not yet as widely adopted. By offering insight into the nature of the misinformative content itself, we can develop a better intuition for understanding what kind of content will spread.

One of the first evolutionary studies was conducted by Jang et al., wherein the team developed a new framework for understanding the propagation of misinformation by borrowing ideas from traditional evolutionary biology [13]. Phylogenetic trees are tools used to do illustrate evolutionary relationships between organisms based on characteristic similarities and differences; in a more formal context, we can define then as connected unidirectional graphs without cycles and an explicit most common recent ancestor. The authors seek to translate these methods to the problem of misinformation to reveal how misinformative content spread and evolves during the spreading process. The authors use the Q-grams distance metric to reflect similarity between pairs of tweets, generating a table of pairwise values to construct an according evolution tree. To achieve this, the authors insert the various tweets into a priority queue based on chronology. The queue is then repeatedly popped to populate an empty tree, with each tweet being grafted to node with the minimum string distance, thereby producing a minimum spanning tree.

To test this novel framework, they use content aggregator Crimson Hexagon to choose sixty of the most popular news stories from various blogs, online publishers and news organizations, thirty being truthful and and thirty were real. The truthfulness and falsehood of these stories was confirmed with fact checking service *snopes.com* and tweets pertaining to each of these sixty stories were then collected from the Crimson Hexagon database. Using the algorithm outlined above, a phylogenetic tree is created for each of the stories revealing that there are significantly different evolution patterns generated during the propagation of factual versus fake content.

To gain a better understanding of their findings, we consider two particular phylogenies illustrated in Fig. 6. Fig. 6a depicts the propagation of misinformative rumor claiming that Donald Trump was born in Pakistan. Fig 6b depicts the propagation of the factual information that Barack Obama contacted Hilary Clinton through her private email server. Analyses indicate that, on average, misinformative phylogenies have far greater depth, whereas trees for true information display greater breadth. Tree depth is directly reflective of changes to tweet content, while tree breadth is indicative of wider content dispersion. Thus, these findings suggest that real news is generally shared in its original form, while misinformation is instead modified during the propagation process, reflecting individual opinion and distortion of content. Furthermore, the authors note differing chronologies between real and misinformative news: true stories are usually shared immediately upon publishing across many reputable news sources whereas misinformative stories tends to linger and spread more slowly. These contrasting patterns encourage misinformative articles to accumulate more comments and opinions from those users engaging with these topics, often contributing new misinformative content to an existing hoax. Overall, we find that users who engage with misinformation either share the content with opinionated modifications or not at all.

We also garner insights regarding the sources and seeding of misinformation. Normal users – not celebrities or politicians – are sharing 87% of misinformative stories—that is twenty-six out of thirty fake-news items. Of the remaining 13]% of stories, half of them were flagged by Twitter's misinformation systems and blocked while the remaining half published by influential individuals remain active. In addition to being generated predominantly by ordinary individuals, 43% of misinformative stories cite dubious news sites frequently flagged as misinformative. In contrast, 43% for truthful stories are generated by reliable news media sources, with 66% of them citing verified mainstream media sources like The New York Times. Evolutionary models are a burgeoning field that can better elucidate the mechanisms through which misinformation spreads.

4 CONCLUSION

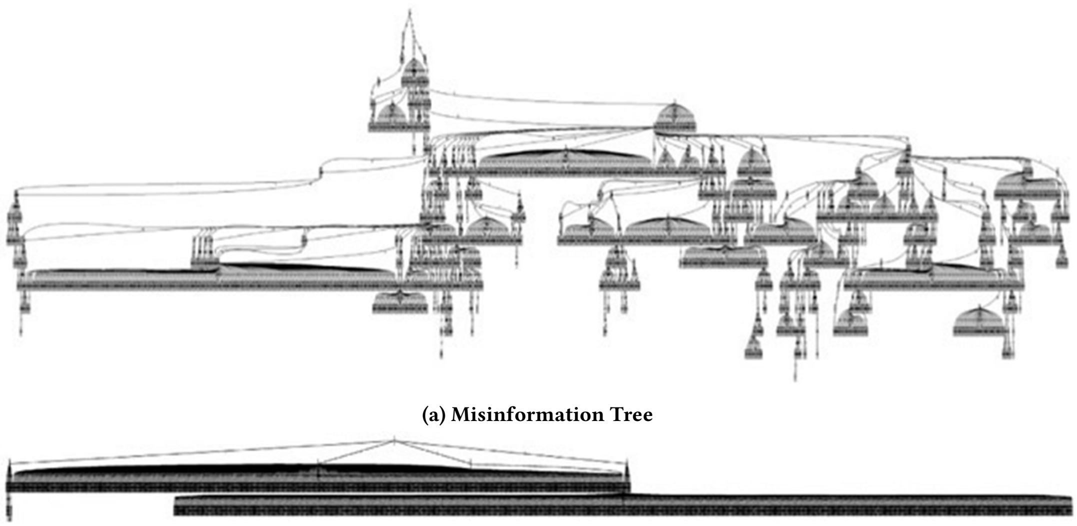
In this paper, we survey the current approaches used to model the propagation of misinformative content through social media spaces, paying specific attention to the conceptual underpinnings of each study. Methodology and conclusions are summarized to uncover complimentary in findings and understand the specific niche of the parallel research approaches. Based on these analyses, we taxonomize current modelling efforts to reveal potential opportunities.

Pathological models conceptualize misinformation as a disease and leverage techniques from epidemiology to understand it's spread. Compartmental mathematical models are used to represent the status of an individual and reflect the way that their interactions change their beliefs and thereby status. Pathological approaches like those explored by Tambuscio et al. and Maleki et al. offer a unique insight into the user-level actions and psychology behind the viral spread of misinformation, highlighting the significance of the surprisingly similar mechanisms behind skeptical verification and self-adoption of misinformation.

Sociogrammatic models look at social media as a web of interconnected actors, using concepts from algorithms and network theory to understand how misinformation can propagate through them. Work by Budak et al. and Pourghomi et al. help to abstract the problem of misinformation propagation, demonstrating the importance of understanding the underlying network structure to identify meaningful actors within the networks. Thus, these types of models are excellent at providing guidance for designing macro-level misinformation mitigation interventions.

Evolutionary models consider misinformation as something iteratively changing and growing throughout the spreading process. One such model by Jang et al. combines basic natural language processing with phylogenetic analysis from biology to study how the content of tweets change as they travel through social media. They uncover key differences in user input and information verifiablity. This class is still developing and future work can certainly look into translating other concepts from evolutionary biology to the misinformation domain.

These different classes of models all broadly serve to explicate the spread of misinformation, but provide different perspectives and deepen our understanding. Combining the learning from the studies surveyed, we learn that misinformation is often generated on a "grassroots" basis. Ordinary users encounter misinformation with misleading sources, deliberate and form individual opinions on the topics, then share this content with their own personal tinting, further obfuscating the evidentiary path and preventing reliable verification of information. We further learn that one of the most effective ways to limit the spread of misinformation is to identify influential actor within networks use them as conduits to verify information and dispel misinformation. These lessons help to additionally inform our future directions: of particular interest is the deliberation process - few people post misinformation immediately or due to peer pressure - what processes lead people to become believers and how can we correct this? Additionally, future studies may seek to integrate the current sets of models; for instance, sociogrammatic analyses can improve by modelling each node's state in terms of an SEIZ model rather than a black and white infection



(a) Misinformation Tree

(b) Real News Tree

Figure 6: Phylogenetic Trees for one false and one real story

status. We hope that further research can provide practical guidelines to inform policy and mitigate the spread of misinformation through social media.

One notable gap in the modelling efforts is a clear lack of consideration for sociodemographic factors affecting misinformation spread. Current modelling efforts treat all users uniformly, however, research in psychology and sociology strongly indicate that factors like education and wealth are the strongest predictors of individuals becoming believers of misinformative content [4, 8, 9, 21]. Future work should attempt to incorporate these interdisciplinary learnings, studying how the propagation of misinformation differs for these vulnerable groups and thereby developing more effective interventions.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives* 31, 2 (May 2017), 211–36. https://doi.org/10.1257/jep.31.2.211

[2] Jenna Anderson and Lee Rainie. 2017. *The Future of Truth and Misinformation Online*. https://www.pewresearch.org/internet/2017/10/19/the-future-of-truth-and-misinformation-online/

[3] Norman T. J. Bailey. [n.d.]. .

[4] Jay Amol Bapaye and Harsh Amol Bapaye. 2021. Demographic Factors Influencing the Impact of Coronavirus-Related Misinformation on WhatsApp: Cross-sectional Questionnaire Study. *JMIR Public Health Surveill* 7, 1 (30 Jan 2021), e19858. https://doi.org/10.2196/19858

[5] Ceren Budak. 2019. What Happened? The Spread of Fake News Publisher Content During the 2016 U.S. Presidential Election. In *The World Wide Web Conference* (San Francisco, CA, USA) *(WWW '19)*. Association for Computing Machinery, New York, NY, USA, 139–150. https://doi.org/10.1145/3308558.3313721

[6] Ceren Budak, Divyakant Agrawal, and Amr El Abbadi. 2011. Limiting the Spread of Misinformation in Social Networks. In *Proceedings of the 20th International Conference on World Wide Web* (Hyderabad, India) *(WWW '11)*. Association for Computing Machinery, New York, NY, USA, 665–674. https://doi.org/10.1145/1963405.1963499

[7] Hilary D. Marston Catharine I. Paules and Anthony S. Fauci. 2019. Measles in 2019 — Going Backward. *New England Journal of Medicine* 380 (June 2019), 2185–2187. https://doi.org/DOI:10.1056/NEJMp1905099

[8] Xinran Chen and Sei-Ching Joanna Sin. 2013. 'Misinformation? What of it?' Motivations and individual differences in misinformation sharing on social media. *Proceedings of the American Society for Information Science and Technology* 50, 1 (2013), 1–4. https://doi.org/10.1002/meet.14505001102 arXiv:https://asistdl.onlinelibrary.wiley.com/doi/pdf/10.1002/meet.14505001102

[9] John Cook, Ullrich Ecker, and Stephan Lewandowsky. 2015. *Misinformation and How to Correct It*. American Cancer Society, 1–17. https://doi.org/10.1002/9781118900772.etrds0222 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118900772.etrds0222

[10] Andrew Guess, Brendan Nyhan, and Jason Reifler. 2018. Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements. http://www.ask-force.org/web/Fundamentalists/Guess-Selective-Exposure-to-Misinformation-Evidence-Presidential-Campaign-2018.pdf

[11] Juhani Eskola Samuel L. Katz Scott Ratzan Heidi J. Larson, Louis Z. Cooper. 2011. Addressing the vaccine confidence gap. *The Lancet* 38, 9790 (2011), 526–535. https://doi.org/10.1016/S0140-6736(11)60678-8

[12] Jennifer Jerit David Schwieder James H. Kuklinski, Paul J. Quirk and Robert F. Rich. 2000. Addressing the vaccine confidence gap. *Journal of Politics* 62, 3 (2000), 790–816. https://doi.org/10.1111/0022-3816.00033

[13] S. Mo Jang, Tieming Geng, Jo-Yun Queenie Li, Ruofan Xia, Chin-Tser Huang, Hwalbin Kim, and Jijun Tang. 2018. A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis. *Computers in Human Behavior* 84 (2018), 103–113. https://doi.org/10.1016/j.chb.2018.02.032

[14] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent Features of Rumor Propagation in Online Social Media. In *2013 IEEE 13th International Conference on Data Mining*. 1103–1108. https://doi.org/10.1109/ICDM.2013.61

[15] David M. J. Lazer. 2018. The Science of Fake News. *Science* 359 (2018), 1094–1096. https://pubmed.ncbi.nlm.nih.gov/29590025/

[16] Kathleen M. Carley Liang Wu, Fred Morstatter and Huan Liu. 2019. Misinformation in Social Media: Definition, Manipulation, and Detection. *ACM SIGKDD Explorations* 21, 2 (Nov. 2019), 36–44. https://doi.org/10.1145/3373464.3373475

[17] Maryam Maleki, Esther Mead, Mohammad Arani, and Nitin Agarwal. 2021. Using an Epidemiological Model to Study the Spread of Misinformation during the Black Lives Matter Movement. arXiv:2103.12191 [cs.SI]

[18] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, and Christos Faloutsos. 2012. Rise and Fall Patterns of Information Diffusion: Model and Implications. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (Beijing, China) *(KDD '12)*. Association for Computing Machinery, New York, NY, USA, 6–14. https://doi.org/10.1145/2339530.2339537

[19] Delia Mocanu, Luca Rossi, Qian Zhang, Marton Karsai, and Walter Quattrociocchi. 2015. Collective attention in the age of (mis)information. *Computers in Human Behavior* 51 (2015), 1198–1204. https://doi.org/10.1016/j.chb.2015.01.024 Computing for Human Learning, Behaviour and Collaboration in the Social and Mobile Networks Era.

[20] Nam P. Nguyen, Guanhua Yan, My T. Thai, and Stephan Eidenbenz. 2012. Containment of Misinformation Spread in Online Social Networks. In *Proceedings of the 4th Annual ACM Web Science Conference* (Evanston, Illinois) *(WebSci '12)*. Association for Computing Machinery, New York, NY, USA, 213–222. https://doi.org/10.1145/2380718.2380746

[21] Gordon Pennycook and David G. Rand. 2021. The Psychology of Fake News. *Trends in Cognitive Sciences* 25, 5 (2021), 388–402. https://doi.org/10.1016/j.tics.2021.02.007

[22] Gregory A. Poland and Ray Spier. 2010. Fear, misinformation, and innumerates: How the Wakefield paper, the press, and advocacy groups damaged the public health. *Vaccine* 28, 12 (2010), 2361–2362. https://doi.org/10.1016/j.vaccine.2010.02.052

[23] Pardis Pourghomi, Milan Dordevic, and Fadi Safieddine. 2018. The Spreading of Misinformation online: 3D Simulation. In *2018 5th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*. 299–304. https://doi.org/10.1109/ICITACEE.2018.8576937

[24] Pardis Pourghomi, Fadi Safieddine, Wassim Masri, and Milan Dordevic. 2017. How to stop spread of misinformation on social media: Facebook plans vs. right-click authenticate approach. In *2017 International Conference on Engineering MIS (ICEMIS)*. 1–8. https://doi.org/10.1109/ICEMIS.2017.8272957

[25] Kristen Purcell and Lee Rainie. 2014. *Americans Feel Better Informed Thanks to the Internet.* https://www.pewresearch.org/internet/2014/12/08/better-informed/

[26] Elisa Shearer. 2020. *Social media outpaces print newspapers in the U.S. as a news source.* https://www.pewresearch.org/fact-tank/2018/12/10/social-media-outpaces-print-newspapers-in-the-u-s-as-a-news-source/

[27] Kai Shu, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements. arXiv:2001.00623 [cs.SI]

[28] Marcella Tambuscio, Giancarlo Ruffo, Alessandro Flammini, and Filippo Menczer. 2015. Fact-Checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15 Companion)*. Association for Computing Machinery, New York, NY, USA, 977–982. https://doi.org/10.1145/2740908.2742572

[29] Steven Taylor and Gordon J. G. Asmundson. 2020. Negative attitudes about facemasks during the COVID-19 pandemic: The dual importance of perceived ineffectiveness and psychological reactance. *medRxiv* (2020). https://doi.org/10.1101/2020.11.17.20233585 arXiv:https://www.medrxiv.org/content/early/2020/11/18/2020.11.17.20233585.full.pdf

[30] Samuel P Trethewey. 2020. Strategies to combat medical misinformation on social media. *Postgraduate Medical Journal* 96, 1131 (2020), 4–6. https://doi.org/10.1136/postgradmedj-2019-137201 arXiv:https://pmj.bmj.com/content/96/1131/4.full.pdf

[31] Yuxi Wang, Martin McKee, Aleksandra Torbica, and David Stuckler. 2019. Systematic Literature Review on the Spread of Health-related Misinformation on Social Media. *Social Science Medicine* 240 (2019), 112552. https://doi.org/10.1016/j.socscimed.2019.112552

[32] Zhixuan Zhou, Huankang Guan, Meghana Bhat, and Justin Hsu. 2019. Fake News Detection via NLP is Vulnerable to Adversarial Attacks. *Proceedings of the 11th International Conference on Agents and Artificial Intelligence* (2019). https://doi.org/10.5220/0007566307940800