Development of a Mutable Meme Model in the Context of a Contagion Spread Simulation

Koby Lewis, Joseph Mattson, Caelen Miller, Michael DeBuse, and Sean Warnick[†]

Abstract—The prevalence of misinformation in social networks is one of modern society's greatest challenges. One of the causes of this is the mutation of memes (e.g. ideas or facts) into topically similar yet semantically different memes. In many cases, individuals in social networks who mutate memes in such a way can cause a spread of misinformation that leads to a large group of people rejecting the initial meme. The goal of this study is to create a model to represent meme mutation in contagion simulations. In order to accomplish this, we develop a tree structure that represents memes in a meaningful and mutable way. We simulate the resulting model dynamics on a social network using a complex contagion model, and illustrate how memes described by this tree structure reflect many phenomena observed in real-world social networks, such as the formation of echo chambers, the impact of influencers, and the importance of simplicity in meme spread.

I. INTRODUCTION

In today's world, information spreads rapidly through social networks and can have a significant impact on society [1], [2]. It is therefore crucial to understand the mechanisms behind the spread of ideas and information, particularly when it comes to misinformation. Through the rest of this paper, we use the term "meme" to refer to ideas and facts in a social context.

As memes spread through social networks, they are commonly misinterpreted by actors in the network [3], [4]. In essence, these actors "mutate" the meme into something topically similar, yet semantically different. Let us consider a meme indicating that international applications are down at 40% of universities in the United States [5]. This meme could be spread via a social network from a source to various actors. Each actor has the potential to misinterpret the meme in some way. For example, one actor could interpret the meme to imply that international applications are down by 40% on average at universities instead of down at 40% of universities. This actor could then spread this modified meme to others in the network. Consequently, these mutated memes can be further propagated through the network to other actors. The focus of this study is to observe the spread of misinformation through this mechanism.

This study aims to contribute to the understanding of this mechanism by using an agent-based model to demonstrate the mutation of ideas in a social network. There are many

†Koby Lewis, Joseph Mattson, Caelen Miller, Michael DeBuse, and Sean Warnick are with the Information and Decision Algorithms Laboratories (IDeA Labs), Department of Computer Science, Brigham Young University, Provo, UT 84602, USA kobyjlewis@gmail.com, joseph.d.mattson@gmail.com, caelendmiller@gmail.com, mdebuse3@gmail.com, and sean@cs.byu.edu

nuances in memes that make them difficult to model while appropriately capturing their meaning [6], [7]. One of these is the presence of qualitative concepts. Looking at our previous example, the concept of international admissions is important to the meaning of the meme. A mutation to this part of the meme, such as changing it to refer to *general* admissions, is difficult to model. In order to render a tractable model of the mutation of memes, we limit the model to only mutate quantitative and structural components of the meme. Despite this limitation, we can capture many of the complexities of how ideas are mutated in a social network, illustrating how changes in the structure of a meme affects the way it spreads through a network [8].

In order to examine the behavior of the mutation in this model, we implement a stochastic complex contagion model to represent the social network. Similar models are commonly used to represent the discrete spread of information within a social network [8], [9], [10], [11]. This contagion model generally functions by setting up a graph of nodes and connections and initializing one node with the "contagion." In this case, the contagion is the original meme. In each step of the simulation, uninfected nodes have a chance to be "exposed" to a meme with a probability based on the number of connections they share with nodes that are "infected" with that meme. In our version of this model, actors also have a chance to mutate the meme to which they are "exposed." We include other parameters that change the nature of the propagation of memes in this model.

II. RELATED WORKS

Cinelli *et. al.* [2] collected evidence surrounding the spread of misinformation about COVID-19 on social networks, including Twitter, Instagram, and Reddit. They fit the spread data to epidemic models. Where our study is a proof of concept, this one by Cinelli *et. al.* contains data from the spread of real-world misinformation. In a future study, we could compare how our model holds up against actual misinformation data like these.

Yan *et. al.* [3] discussed the effects of actors spreading modified information on social media. Real data from Twitter posts was used for this study. They found that there is a positive correlation between mutating the ideas presented in original sources to make a Tweet and a faster spread of those Tweets. While they studied the specific changes made to the original memes, they did not propose a computational model to represent those memes as we do here.

West and Bergstrom [5] collected a list of science misconceptions that are used for some of our sample memes.

Rabb *et. al.* [10] developed a cognitive contagion model describing the spread of misinformation. They used a modified complex contagion model to run their simulations. This research parallels ours in that they developed a model for contagion spread in an attempt to mirror reality. However, our research focuses on a logical representation of memes, while the research done by Rabb *et. al.* focuses exclusively on the cognitive state of the actors in the network.

Emery et. al. [11] worked with complex contagion models to attempt to understand how to mitigate the spread of misinformation. By eliminating certain problematic users from the network, misinformation spread dropped dramatically. The methods used in this study are similar to ours, however, our focus is more on the model of the meme itself rather than the dynamics of misinformation in the network.

Törnberg *et. al.* [8] studied the formation of echo chambers in social networks. They employed contagion models to represent the spread of misinformation in these networks. Echo chambers appear in our data and play a significant role in the spread dynamics of our system. While we use a similar contagion model to Törnberg *et. al.*, they do not focus on the actual memetic mutation model for misinformation as we do here.

Raponi *et. al.* [12] performed an overarching study of fakenews propagation models supported by real-world datasets. This included epidemiological, forest fire, energy, and information cascade models. Unlike the model that we created, the models discussed focused on modifying the nature of the individuals and their possible states in the network rather then the type of information being spread.

Lavorgna *et. al.* [13] studied the ability of social media influencers to propel or limit the spread of misinformation using an online experiment. The power of influencers is reflected in our model, though not to the degree of nuance presented in their research. In future research we may include some of this nuance as a means to study influencers.

Harff *et. al.* [14] studied the distribution of information by influencers within the context of a specific medical social media platform. They found that when properly moderated by experts in a field, the platform naturally self corrects, preventing individuals who spread fake news from becoming influencers.

Gabora [15] demonstrated the viability of meme spread and evolution by describing a computer model in which an entity holds a meme represented as a small neural network that controls a set of parameters. These parameters were fed into a fitness function, and agents in the simulation did not adopt a new meme unless it produced fitter parameters than the old meme. While these parameters were given a concrete meaning (that of positions of body parts), the memes themselves were quite abstract and only capture quantitative aspects of memes.

III. MEME MUTATION & CONTAGION MODELS

The model we develop consists of three major components, which are described in detail below. These components are:

- 1) A model for representing an individual meme
- 2) Transformation rules for mutating a meme
- 3) A model for the spread and contagion of memes in a social network

The transformation rules we define are based on a small collection of misconceptions, fallacies and miscommunications, and is by no means complete. Instead, we intend to provide a proof of concept and a sample for how the mutation of memes in a social network can be represented. The goal is to demonstrate a system that can represent a breadth of the sort of ideas that might spread in a real social network.

A. Structure of a Meme

We model memes as a set of Boolean expressions and conditional probabilities that can be represented in a recursive tree structure. As a simple example, the meme "There is a chance of catching a cold this winter" could be represented as $P(\text{``catching a cold''} \mid \text{``winter''}) = s\%$ where s is some small, non-zero percentage. This can be structurally divided into the probability expression, the equality operator and the percentage s. The probability expression includes two propositions. A full grammar that can be used to generate memes under this model is described in Figure 1.

The notation $\langle Meme \rangle$ (see Figure 1) refers to the set of all possible expressions of the given structure. Memes, in this model, are thus represented by a list of expressions. If an agent is infected with a particular meme, we interpret this to mean that they believe each of the expressions composing that meme. An expression in this model can either be a proposition, a comparison statement, a combination of other expressions (either a conjunction or disjunction), or a negation of an expression. Numerical values can be a numerical variable (e.g. "Months since COVID-19 recovery"), a number literal (e.g. "5" or "40%"), a conditional probability statement, or a mathematical expression combining other numerical values using the "+" or "-" operators—basically any expression that evaluates to a number. A conditional probability statement is made up of the probability function P, an expression as the argument, and optionally another expression as the condition.

B. Mutations

Once a set of memes is defined, agents can mutate them from their original form into a different version. To do this, we collect a set of common misconceptions that act on the *structure* of a meme, transforming it. This is done by recursively searching a meme's nodes for a node that matches the structure of one of the transformation rules listed in Table I. For example, as shown in Figure 2, one misconception could be, "There is a 40% chance applications are down", where the original meme was, "Applications are down at 40% of schools". This change could be caused by the "Loss of Condition" transformation rule.

The "Exaggeration" rule is based on the idea that numbers are often communicated imprecisely, which can lead to the recipient of a meme interpreting an exaggerated version of the probability of an event occurring. For instance, given the

Name	Transformation	Example
Exaggeration	$p \implies \frac{1}{1+\frac{p}{1-p}-2}$	(P("Getting a computer virus" "Is a Linux
	$1+\frac{1}{1-p}$	OS'') = 0.1) \Longrightarrow (P("Getting a computer
		virus" "Is a Linux OS") = 0.0122)[16]
Loss of Context	$\langle Meme \rangle \implies \langle Meme \rangle \setminus \langle BinExp \rangle_i$	("The Earth has experienced higher CO ₂ levels
		before","The Earth has never experienced such
		a sudden rise in CO_2 levels") \Longrightarrow ("The Earth
		has experienced higher CO ₂ levels before")
Reverse Implication	$P(a b) \Longrightarrow P(b a)$	P("Having low cholesterol" "Having cancer") = $n \Longrightarrow$
		P("Having cancer" "Having low cholesterol") = n
Faulty Negation	$P(a \neg(b)) \Longrightarrow P(\neg(a) b)$	P("Being reinfected with COVID-19" ¬("Months since
		recovery" > 3)) = $n \implies P(\neg("Being reinfected with$
		COVID-19") ("Months since recovery" > 3)) = n [5]
Common Cause	$P(a c) \langle CompOp \rangle_1 x, P(b c)$	P("Ice cream sales are up" "It is summer") = n ,
	$\langle CompOp \rangle_2 \ y \Longrightarrow P(a b) \langle CompOp \rangle_1 \ x$	P("Drowning rates are up" "It is summer") = $n \Longrightarrow$
		P("Ice cream sales are up" "Drowning rates are up") = n
Loss of Condition	$P(a b) \implies P(a)$	P("The amount applications are down" > 0
		"We are analyzing one school") \Longrightarrow P("The
		amount applications are down" > 0) [5]
Probability Collapse	$P(a > 0) = n \implies a = n$	P("The amount applications are down" > 0) = 40%
		\implies "The amount applications are down" = 40% [5]

TABLE I

MEME TRANSFORMATION RULES

meme "There is a small probability of a computer running Linux getting a virus", the meme might mutate to "There is a *very* small probability of a computer running Linux getting a virus".

The "Loss of Context" rule simply states that often not all of the information is communicated properly. While the remaining information might still be true, without surrounding context it could lead to incorrect action. This is demonstrated by the meme, "The Earth has experienced these levels of CO₂ before, but never this fast of a rise in CO₂," becoming, "The Earth has experienced these levels of CO₂ before," from which one might conclude that we need not be concerned about climate change.

The "Reverse Implication" rule is based on confusing which direction a condition points for two events that occur together. For instance, the meme, "People who have cancer also tend to have low cholesterol," might be misunderstood as "People who have low cholesterol also tend to have cancer".

"Faulty Negation" occurs when an unwarranted conclusion is made that if a condition is negated, then so must the main argument of a probability expression. This can be seen in the meme, "You are unlikely to be reinfected with COVID-19 if it has been less than three months since recovering from a previous infection," becoming, "You are likely to be reinfected with COVID-19 if it has been three or more months since recovering from a previous infection."

"Common Cause" is a fallacy that occurs when two events share a common condition, but one is assumed to be the cause of the other. A classic example of this is that both ice cream sales and drowning rates are up during summer months, which might be interpreted to mean that ice cream sales *cause* drowning.

"Loss of Condition" was demonstrated above with the example of admissions, but that misconception can be taken further with "Probability Collapse", which transforms a probability that some value a is greater than 0 is n, into

simply that the value of a is n. For instance, we can take the meme that, "The chance that applications are down is 40%," and transform it further into, "Applications are down by 40%".

C. Contagion Model and Variants

To simulate the spread of the meme, an undirected graph with fixed topology is used. In particular, we used the Infect Dublin data set [17]. More information on the properties of Infect Dublin can be found at https://networkrepository.com/ ia-infect-dublin.php. Each node of the graph can either hold no meme, or some variant of a meme. An original meme is given to a starting node and the simulation moves on to the next time step. At each time step, each node N_i that does not hold a meme will check its predecessors for memes, and will select the most common meme variant $M_{most-common}$ among them (ties are broken arbitrarily). The node N_i will then have some probability p_{adopt} of adopting the meme $M_{most-common}$, which it will hold from then on. However, if $M_{most-common}$ is chosen to be adopted, there is a probability p_{mutate} that an attempt to mutate $M_{most-common}$ will be made. If the attempt is successful, N_i will instead adopt the mutated version. Any number of mutated variants can be present in the network at any given time, provided at least one node holds that meme. Each node can only hold up to one variant at a time. The hyper parameters p_{adopt} and p_{mutate} are the same for every node. In our simulations, we give them the values 1.0 and 0.2, respectively.

The process for attempting a mutation is to shuffle all of the available transformations described in Table I, and sequentially test to see if the transformation matches the structure of the meme. The first transformation that matches is used and the mutation succeeds. If no transformation matches the meme, the mutation fails.

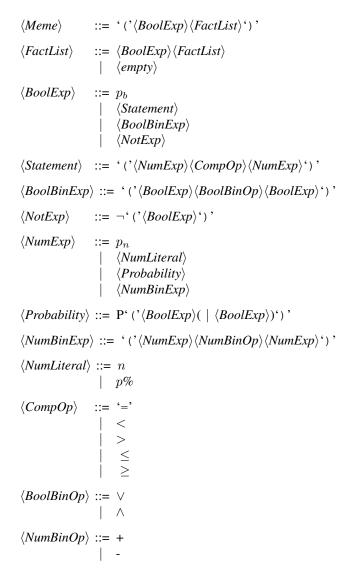


Fig. 1. Grammar for Memes. This context-free grammar lays out the rules for the structure of a Meme in our model. '::=' can be read as 'is constructed from' or 'has the constituents'. '|' can be read as 'or'. A < NumExp>, as shown in the grammar, is either a numeric proposition (such as "the number of applications"), a < NumLiteral>, a < Probability>, or a < NumBinExp>. A < Meme> is made up of a < BoolExp> and a < FactList> surrounded by parentheses. A meme can be generated by starting with a < Meme> node and replacing all bracketed items with one of their options for constituents and repeating until there are no bracketed items remaining.

We make several assumptions about the agents in the network — represented by the graph nodes — in order to run this simulation. Each agent acts exactly the same as all other agents in the way they receive and propagate memes. Each agent only has access to memes held by other agents they are connected to. Each agent can only hold one meme, and differing memes have no interactions other than competition to spread through the network.

There are a number of parametric variations on the base contagion model that were also tested. These parameters are listed below.

A COMPLEXITY AFFECTS SPREAD: If set, this parameter makes more complex memes less likely to spread to

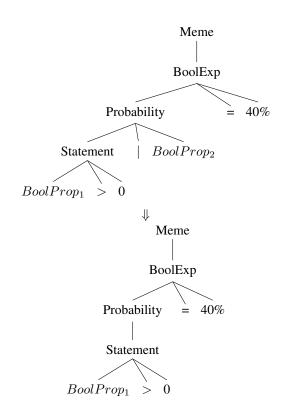


Fig. 2. Representation of Meme Structure. This example shows the representation of an initial meme in a tree structure and a potential mutation into a different meme. In this case, $BoolProp_1$ is the proposition describing "the amount international applications are down" and $BoolProp_2$ is the proposition that "only one school is being examined." The mutation removes this second proposition, implying that applications are down at *all* schools.

their neighbors. A meme's complexity is calculated by totalling up the number of literals and propositions in the meme. The probability of the meme being adopted p_{adopt} is divided by the complexity of the meme.

B CAN CHANGE MIND: When set, nodes can be reinfected. In other words, nodes that already hold a meme still check their predecessors for memes. The node has a probability $p_{change-mind}$ to adopt the new most common meme among its predecessors. All nodes have the same value for $p_{change-mind}$, and in our simulations we set this value to 0.1.

IV. RESULTS

The purpose of our model is to simulate real world meme mutation and spread, thereby providing a framework to conduct research on the spread of misinformation. Our novel approach to meme representation creates opportunities to study the mutation of information and its natural propagation through a social network. Results thus far have been promising, proving a model that is both intuitive and accurate to previously studied network dynamics. Through manipulating the parameters, various real-world phenomena can be simulated. These include the rapid spread of simple-to-understand but false information, the formation of echo chambers, and an increased rate of spread due to influential individuals. As our model manipulates both the information

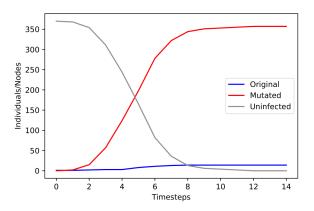


Fig. 3. Domination by Simple Mutated Memes. This graph shows the spread of memes through a network over time. This simulation was generated by allowing the complexity of the meme affect its spread. The simulation ran until it reached a steady state, which in this case took 14 time steps. In this and all following figures, red is used to represent the combined set of all mutated variants, while blue represents only the original variant. The mutated memes dominate the network, as they are simpler and thus have a higher infection probability than the original meme. This propagation closely mimics generic epidemic spread, thus showing the efficacy of our model.

being propagated and the nature of the propagation, there are still many unexplored possible applications.

A. Rapid Spread of Simplified Memes

With the "Complexity Affects Spread" parameter on, less complex memes tend to spread much faster. This is due to the probability of adoption at each time step being inversely proportional to the complexity of the meme. As seen in Figure 3, this leads to the mutated, simplified memes completely taking over, regardless of whether nodes can be reinfected. This is closely mirrored in reality, where nuanced, yet true, memes are often distilled into simple, though distorted, memes.

Our model allows for results like this to be generated organically based on the natural mutations and spread of memes. Without a model like ours, researchers would need to use high level abstractions to reproduce similar phenomena. Because of the hierarchical structure of our meme model, it is simple to calculate the complexity of a meme, which can then be used to determine the rate of spread of that meme.

B. Formation of Echo Chambers

When memes are able to infect individuals that were previously infected by a different meme, echo chambers (groups of individuals that all hold the same beliefs) often form. This may manifest in a variety of ways, with the most common being the formation of isolated but well insulated groups, as shown in Figure 4. Although these groups are small compared to their opposition, they self reinforce to the extent that they become stable. Another manifestation of echo chambers is the complete polarization of the network. In these cases, two distinct groups form in the network, one for the original meme and one for the mutated memes. Any

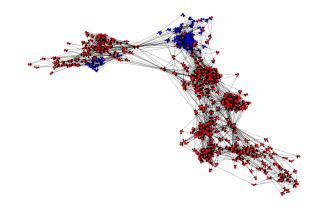


Fig. 4. Small, Stable Echo Chambers. This network was produced while allowing individuals in the network to be infected by a new meme after having already been infected. An echo chamber formed in which a group continued to hold to the original meme (blue) despite being surrounded by actors with mutated memes (red). This stayed relatively stable until it reached 40 time steps, at which point the simulation ended.

individual that has beliefs contrary to the majority in either group is rapidly converted, resulting in each group purely consisting of one meme.

C. Influential Actors

Within social networks, some actors have significantly more connections than most others (high degree vertex). These influential actors, or influencers, have a tremendous impact on the final result of the network. The meme that infects an influencer first has a large advantage over its competition, typically taking over the majority of the network unless the other meme also infects an influencer within a short time frame. A meme that is unable to reach a major influencer in time is generally either eliminated or develops into small echo chambers. As shown in Figure 5, allowing true memes to reach an influencer early in the simulation minimizes the spread of mutated variants in the network.

D. Effect of Allowing Opinion Changes

When opinions are not allowed to change after initial infection by a meme, the model rapidly reaches a stable point. However, when using an option to allow individuals to change their minds as shown in Figure 6, there is a more dynamic competition between the different memes present in the network. An equilibrium is reached where the network is divided between the different memes and individuals swap between them. This is related to the echo chambers seen in Figure 4 in that the groups with the same meme are relatively stable over a period of time, but have members at the fringes that will occasionally swap.

V. CONCLUSION

In this paper, we demonstrate the efficacy of our mutable meme model. Using our model, we can observe the spread of memes in a social network and how mutations to memes can occur and spread in such an environment. By allowing simpler memes to have a higher likelihood of

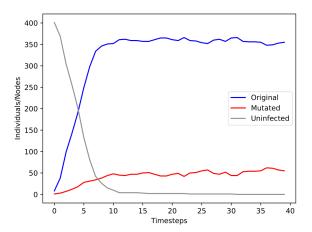


Fig. 5. Influencers. This graph shows the result of allowing the original meme to start on the most influential node in the network. Although the mutated memes spread more quickly than the original meme due to their increased simplicity, the advantageous starting position of the original meme allows it to dominate. This simulation ran for 40 time steps.

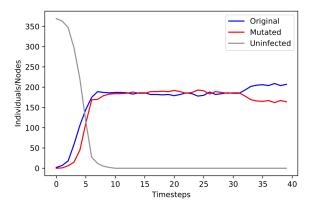


Fig. 6. Opinion Change. This graph was produced by allowing individuals in the network to be infected by memes even if they have already been infected. The simulation ran for 40 time steps. The potential for reinfection in this experiment causes an equilibrium where individuals swap between different memes.

spreading, mutated memes come to dominate social networks in simulation. In addition, we observe the formation of echo chambers centered around memes in these simulations. These phenomena and others discussed in this paper give us insight into the spread and behavior of misinformation in social networks.

While this model is effective in demonstrating meaningful mutations, it has limitations. Because of the model's focus on logical structure, many potential semantic mutations are infeasible. For example, in the meme "There is a chance of catching a cold this winter," mutating the season "winter" to "spring" is impossible with our current model. We also have no way to model the effect such a mutation would have on a meme's potential propagation.

In addition to the lack of semantic representation in our meme model, our process is limited by the homogeneity of the population in the network. Due to the lack of detail in each actor in the network, our simulations ignore the potential effect each actor's internal state may have on the spread of information.

At the moment, our model is a proof of concept. In future research, we plan to address these limitations. Specifically, we are working to represent an internal "opinion" state of each actor. This opinion will affect an individual's likelihood to adopt a meme based on the similarity of the meme and the opinion.

REFERENCES

- [1] A. Korsunska, "The spread and mutation of science misinformation," in *Information in Contemporary Society: 14th International Conference, iConference 2019, Washington, DC, USA, March 31–April 3, 2019, Proceedings 14.* Springer, 2019, pp. 162–169.
- [2] M. Cinelli, W. Quattrociocchi, A. Galeazzi, C. M. Valensise, E. Brugnoli, A. L. Schmidt, P. Zola, F. Zollo, and A. Scala, "The covid-19 social media infodemic," *Scientific reports*, vol. 10, no. 1, pp. 1–10, 2020.
- [3] M. Yan, Y.-R. Lin, and W.-T. Chung, "Are mutated misinformation more contagious? a case study of covid-19 misinformation on twitter," in 14th ACM Web Science Conference 2022, 2022, pp. 336–347.
- [4] J. Akaishi and T. Arita, "Misperception, communication and diversity," 01 2003.
- [5] J. West and C. T. Bergstrom, "Misinformation in and about science," Proceedings of the National Academy of Sciences, vol. 118, no. 15, 2021
- [6] K. Atkinson, T. Bench-Capon, and P. McBurney, "Computational representation of practical argument," *Synthese*, vol. 152, pp. 157–206, 2006
- [7] H. J. Levesque, "Knowledge representation and reasoning," *Annual review of computer science*, vol. 1, no. 1, pp. 255–287, 1986.
- [8] P. Törnberg, "Echo chambers and viral misinformation: Modeling fake news as complex contagion," *PLoS one*, vol. 13, no. 9, p. e0203958, 2018
- [9] D. A. Sprague and T. House, "Evidence for complex contagion models of social contagion from observational data," *PloS one*, vol. 12, no. 7, p. e0180802, 2017.
- [10] N. Rabb, L. Cowen, J. P. de Ruiter, and M. Scheutz, "Cognitive cascades: How to model (and potentially counter) the spread of fake news," *Plos one*, vol. 17, no. 1, p. e0261811, 2022.
- [11] B. Emery, C. Ting, N. Johnson, and J. Tucker, "A projected network model of online disinformation cascades." Sandia National Lab.(SNL-NM), Albuquerque, NM (United States), Tech. Rep., 2021.
- [12] S. Raponi, Z. Khalifa, G. Oligeri, and R. Di Pietro, "Fake news propagation: A review of epidemic models, datasets, and insights," vol. 16, no. 3, 2022. [Online]. Available: https://doi.org/10.1145/3522756
- [13] "Fake news, influencers and health-related professional participation on the web: A pilot study on a social-network of people with multiple sclerosis," *Multiple Sclerosis and Related Disorders*, vol. 25, pp. 175– 178, 2018.
- [14] D. Harff, C. Bollen, and D. Schmuck, "Responses to social media influencers' misinformation about covid-19: A pre-registered multipleexposure experiment," *Media Psychology*, vol. 25, no. 6, pp. 831–850, 2022.
- [15] L. Gabora, "Meme and variations: A computer model of cultural evolution," in 1993 Lectures in complex systems, D. S. L. Nadel, Ed. Mass Reading, 07 1995, ch. 10, pp. 471–486.
- [16] K. K. Mookhey, "Linux: Security, audit and control features," p. 128.
- [17] R. A. Rossi and N. K. Ahmed, "The network data repository with interactive graph analytics and visualization," in AAAI, 2015. [Online]. Available: https://networkrepository.com