Project Report R3

Ishan Mistry (ihmistry@ncsu.edu)
Andrew Bassett (apbasset@ncsu.edu)
Chirag Hegde (chegde@ncsu.edu)

Relevant Literature

First and foremost, the inspiration for our research is the work of Lonnberg et al. (2020), who created a series of Stochastic Differential Equations (SDEs) that generate a SIR model – frequently used in epidemiology – and mathematically approximates online meme engagement over time. We found this work as fascinating as it is important: we quickly realized that collecting real-world data in the context of meme propagation poses numerous challenges. Thus, creating models and simulations as Lonnberg et al. have done, is an important task. We like to think of our simulation framework as an extension of the "Lonnberg model" – a deep and detailed look inside of it with some novel advancements.

More research that is important to our work is that of Russo et al. (2023). We find the spread of online disinformation to be a significant societal issue, and they drafted a framework to combat disinformation through the idea of *emotional responses*. We incorporated this idea into our simulation as a configurable parameter, and analyzed how it affects meme propagation in various circumstances. There is also significant opportunity for future work in the combination of Russo et al. (2023) and our simulation, as their dataset is not yet public.

Another paper is Pellet-Rostaing et al. (2023). This is focused on the identification or classification of various modalities of memes: Textual memes, Image-based memes, and animated memes (gifs). This will play an important role in structuring our meme characteristics and network topologies.

What problem are you addressing?

We intend to analyze the growing prevalence of "meme culture" and its impact on discourse under various conditions. With so many variables at play, we need a simpler playground for rudimentary experimentation. This is exactly what we hope to achieve with our project; more specifically, we intend to accurately model the spread of propagation of memes online in a highly configurable and detailed simulation. Significant research has already occurred on how online anonymity promotes certain behavior, but memes specifically as an ultra-anonymous, decentralized, and rapidly changing landscape present interesting research scenarios about decentralized online social relationships.

Why is this problem important?

The problem we are addressing holds paramount significance in the contemporary landscape of digital communication and information dissemination. As society becomes increasingly interconnected through online platforms, understanding the mechanisms and dynamics of information propagation is essential for comprehending the broader implications on individual behaviors, societal structures, and the integrity of shared narratives.

In an era dominated by social media and digital interactions, the spread of information, often in the form of memes, exhibits a profound influence on public opinion, attitudes, and cultural trends. The virality of memes, characterized by their rapid dissemination and adoption, underscores the need for a comprehensive understanding of the factors that drive and impede their propagation.

Our project delves into the interplay between agent characteristics, network structures, and information dissemination patterns. By observing the role of polarization in networks, we aim to see the impact of ideological clustering on the spread of memes and the formation of echo chambers. Understanding how individuals with similar viewpoints cluster together, and the extent to which this clustering affects the diversity and reach of information, is crucial for fostering informed public discourse and mitigating the risks of misinformation.

Some of the ideas that we try to discover are :

1. Do different types of memes about the same topic impact engagement differently (Segev et al. (2015))? News articles show reduced engagement when there are many similar news items. In contrast, gossip spreads quickly but thinly. Memes combine elements of both.
2. Are structural properties of the networks a good identifier of the extent of the proliferation of memetic speech?
3. Memes that disseminate negative sentiment towards protected classes regularly receive widespread online engagement.
4. The sharing of hateful meme content correlates positively with the degree of anonymity provided by each social media platform.
5. Memes are positively correlated with social reputation[Liu et al. (2021), Kamvar et al. (2003)], especially trust.

How will you address this problem?

We try out an agent-based modeling approach to simulate and analyze the intricate interplay between agents, memes, and diverse network structures. At the core of our simulation framework are agents, representing individual entities in an online ecosystem, each characterized by a unique set of attributes, including susceptibility to information, humor preferences, and engagement tendencies. Memes, the vehicles of information, embody content with varying qualities, influencing agents' behaviors upon reception.

Networks

The network topology serves as the underlying scaffold for agent interactions, encompassing diverse structures such as Random Networks, Polarized Crowds, and Community Clusters. Each network type encapsulates distinct social configurations, reflecting real-world scenarios, from random connections to polarized groups and clustered communities.

In contrast, Polarized Crowds reflect social polarization, fostering strong internal connections while limiting interactions across ideological lines. Community Clusters emulate cohesive social groups, where meme diffusion is influenced by tight-knit community dynamics.

Random Networks

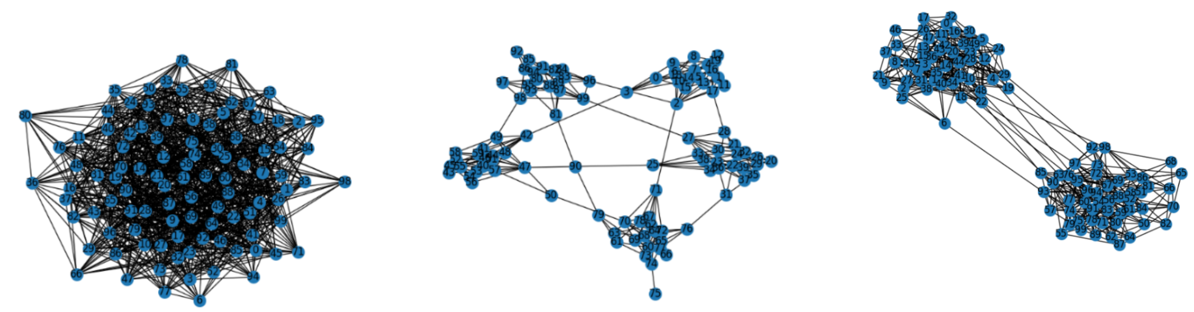
A Random Network is characterized by the absence of specific patterns or structures in the formation of connections among agents. In this network model, each agent has an equal probability of forming connections with any other agent, resulting in a decentralized and unpredictable social landscape. The Random Network model mirrors scenarios where connections lack inherent patterns, simulating a broad, decentralized landscape. The randomness in network connections introduces an element of unpredictability, enabling us to explore how memes navigate through the online landscape when social ties lack predefined structures.

Community Clusters

Community Clusters act as a nuanced and realistic depiction of social structures, where agents form cohesive groups based on shared interests, affiliations, or thematic commonalities. In this network model, individuals are organized into distinct clusters or communities, each characterized by tight-knit connections among its members. The aim here is to model scenarios where people tend to gravitate towards comments that align with their preferences. This network model enables us to investigate questions surrounding the resilience of information within specific interest groups, the formation of echo chambers, and the role of community dynamics in shaping the virality of content.

Polarized Networks

Polarized Networks serve as a representation of social structures that exhibit distinct clusters or groups with polarized viewpoints or preferences. In a Polarized Crowd network, we observe the creation of two separate groups, each forming its own set of connections, with limited interactions between the groups. These networks encapsulate the dynamic interplay between agents who tend to form connections predominantly within their own clusters while displaying limited interactions across group boundaries.



(a) Random Networks (b) Community Clusters (c) Polarized Networks

Figure 1: Generated Sample Networks(100 Nodes)

Agents

In our simulation project, agents serve as the dynamic entities that drive the dissemination and reception of information within complex social networks. These agents possess a diverse set of characteristics, including susceptibility to external stimuli, humor preferences(incorporated in the form of racism, anonymity etc.), and engagement propensities(possible actions - forward, consume, do nothing). Each node in the network represents an agent. At each time step the agents can take an action, determined by some probability

Simulation

For each simulation run, a network type and meme classification are chosen, and agents are initialized as either "Infected" or "susceptible." Since our simulations are based on the SIR model (frequently used in epidemiology and the basis of related meme researchLonnberg et al. (2020)), we define the three SIR metrics as having the following meanings:

1. **Susceptible**: An agent who has not yet engaged with the meme, but is in the same network as others who have
2. **Infected**: An agent who is engaging with the meme at the current time step
3. **Recovered**: An agent who has engaged with the meme at a previous time step, but could engage again.

Additionally, Agents can have the following characteristics, which are randomly initiated on a scale of 0-1:

1. **Ideology**: A rating of the agent's political leaning
2. **Racism**: A rating of the agent's racial biases
3. **Susceptibility**: A rating of how susceptible an agent is to disinformation
4. **Anonymity**: A rating of how anonymous the agent is in the network

These characteristics determine the likelihood of an agent interacting with a meme or replying to another agent who sent them the meme, depending on its attributes. Meme attributes are similarly represented by floats of values between 0-1 and are defined as follows:

1. **Ideology**: A rating of the meme's political messaging
2. **Racism**: A rating of the meme's racial biases
3. **Truthfulness**: A rating of the meme's level of disinformation

At each time step, agents will take any of the following actions:

1. **Send meme**: The agent sends the meme to one or more neighbors. Their state transitions to "recovered" and the neighbors' states transition to "infected."
2. **Consume meme**: The agent solitarily engages with the meme and updates its ε and φ values (more on this below). Its state transitions to "recovered."

3. **Reply to sender**: The agent makes a reply (an emotional reply [CITE PAPER], or otherwise) to the agent who sent them the meme. That agent updates its ϵ and φ values.

4. **Do nothing:** The agent does nothing, and its state is unchanged.

The foremost measure of meme engagement at a particular time step is the number of "infected" agents at that moment. In our simulation, there are two main attributes that determine whether an agent engages with a meme: engagement probability (ε) and reaction score (φ). Engagement probability ε is exactly as it sounds: the ultimate probability of whether an agent engages with a meme (actions 1-3), or does nothing. The reaction score φ is a related, but separate, measurement that determines things such as: "what kind of reply should Agent 2 send to Agent 1?", "How should Agent 1's scores be affected by consuming a meme of a particular content class?" and so on. As with all other attributes, φ is a float value within the range [0, 1], where 0 represents very strong engagement due to very strong agreement with the meme content, and 1 represents very strong engagement due to very strong disagreement with the meme content. A middling score (say, 0.4-0.6) would indicate low engagement and indifference towards the content. These values are computed as follows:

$$\varphi = mean(\Sigma(diff(attr_n)))) \tag{1}$$

$$\varepsilon = |1 - \varphi| \tag{2}$$

$$\varepsilon' = \varepsilon \pm \phi\varepsilon|\alpha - 1/2| \tag{3}$$

$$\varepsilon'' = \varepsilon' + \varepsilon\Omega\beta \tag{4}$$

Notes:

1. The reaction score φ is read as "the mean of the differences of all shared ideological characteristics between the agent and the meme." In the above images, these characteristics are (political) ideology and racism.

2. The value of ε" should be used in place of ε if the truthfulness of the meme is greater than 0.5.

3. The value of ε' should always be used in place of ε unless you do not intend to factor anonymity into the simulation.

Additional variables/functions:

1. diff(): The difference between a given ideological attribute of the agent and the meme

2. ϕ: The anonymity modifier, e.g., a value between [0, 1] that is a percentage amount representing the maximum increase or decrease in an agent's engagement level scaled to their level of anonymity

3. α : The anonymity level of an agent in the range [0, 1]

4. Ω: The susceptibility modifier, e.g., a value between [0, 1] that is a percentage amount representing the maximum increase in an agent's engagement level scaled to their susceptibility to disinformation

5. β: The susceptibility level of an agent in the range [0,1]

Defaults used in the above simulation runs: $\phi, = \{0.15, 0.05\}$

Please also view config.py in our codebase for additional configuration options.

Results

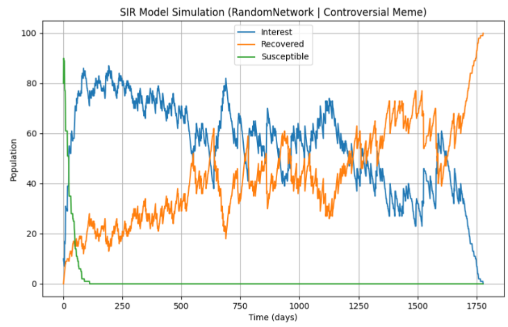


Figure 2: **Simulation 1:** Random Network, Controversial Meme Content, Mixed Anonymity

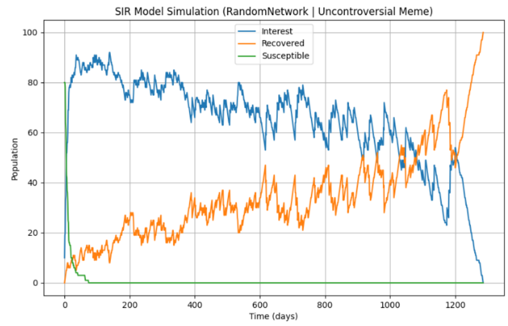


Figure 3: **Simulation 2:** Random Network, Uncontroversial Meme Content, Mixed Anonymity

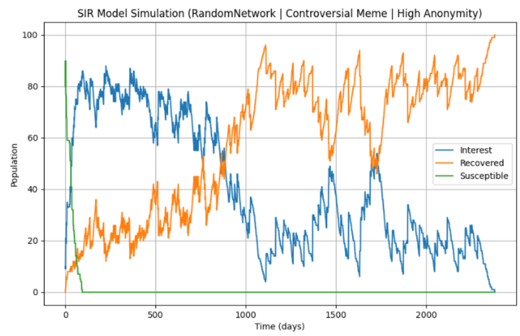


Figure 4: **Simulation 3:** Random Network, Controversial Meme Content, High Anonymity

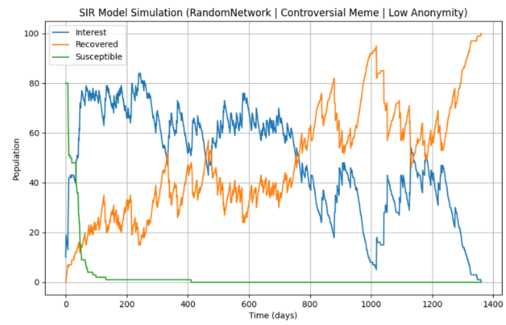


Figure 5: **Simulation 4:** Random Network, Controversial Meme Content, Low Anonymity

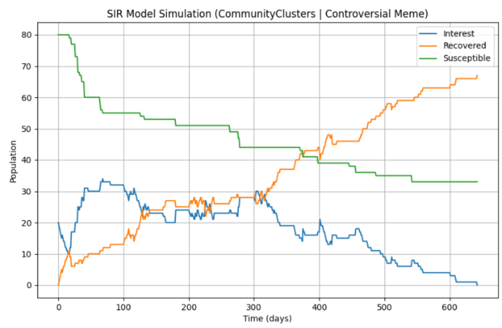


Figure 6: **Simulation 5:** Community Clusters, Controversial Meme Content, Mixed Anonymity

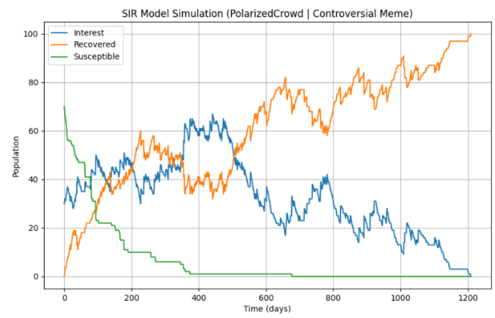


Figure 7: **Simulation 6:** Polarized Crowd, Controversial Meme Content, Mixed Anonymity

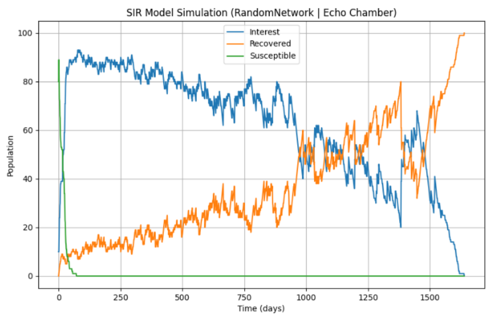


Figure 8: **Simulation 7: Echo Chamber -** All agents are highly ideologically aligned with the meme. High anonymity. Consistently observed heightened levels of engagement in the first quarter of the simulation time frame.

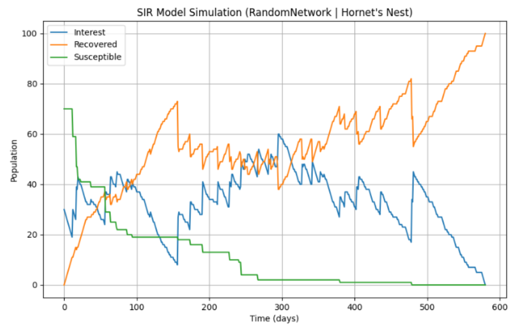


Figure 9: **Simulation 8: Hornet's Nest -** All agents are highly ideologically opposed to the meme. Very sporadic engagement at much lower levels than all other Random Network simulation runs.

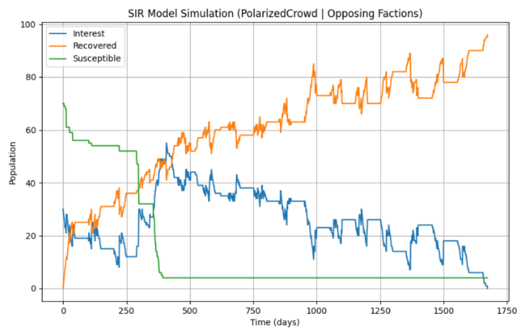


Figure 10: **Simulation 9: Opposing Factions -** Each polarized crowd in the simulation (50 agents each) is set up with opposing ideologies, where one faction is highly aligned with the meme's attributes and the other is highly opposed. Observed step-like engagement levels with moderate engagement.
a

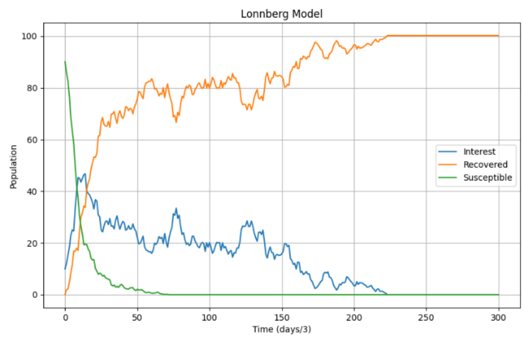


Figure 11: **Simulation 10 (Baseline):** The Lonnberg Model

Observations

1. Controversial meme content (whether racist or highly political) tends to have a longer propagation period than uncontroversial/neutral meme content
2. High anonymity networks tend to engage with memes for longer time periods
3. While both polarized networks and community clusters fail to reach the engagement levels that random, highly connected networks do, community clusters in particular do not facilitate high meme spread/engagement

What are some alternatives and how do you justify your approach?

Evaluating this problem is particularly difficult because we need to make decisions about how to categorize and quantify things that are quite abstract; for example, how do we model abstract notions such as anonymity and the various kinds of memes? As our approach is further developed, we will need to generate a clear framework that can guide our experiments. Fortunately, this is a problem also faced by existing literature which deals with many of the same topics, such as anonymity, hate speech, and disinformation online. Although it has largely been text-based (rather than image-based, in our case), researchers have certainly had to solve some of the "gray area" problems before with regard to these topics. We will be sure to determine best practices based on the "reusable knowledge" that they have already generated. Ideally, we can draw inspiration from some of their methods – with proper accreditation – and extend them to look at memes as a medium/model specifically.

How will you evaluate your approach?

For the deterministic approaches, we would try to replicate the results from the papers Lonnberg et al. (2020) on our curated data to see if we can validate the results from these models. The ground truth when it comes to engagement would be the Lonnberg curve as they claim to approximate Google Trends's data more accurately. The platform engagement metrics would vary across the platforms. Let's say we use the epidemiological model, we would try to simulate the proposed viral features using the equations that the SIR model uses and validate it on the basis of our version of the model from each of the platforms.

What are your main findings of the project?

The societal implications of our research can be extended to influencing fields such as digital marketing, political communication, and the design of online platforms. By deciphering the dynamics of meme propagation, we empower stakeholders with the knowledge to shape communication strategies, design interventions that foster diverse perspectives, and cultivate digital environments that promote constructive engagement.

Industry applications include :

1. Targeted Advertising Strategies: Digital marketers can leverage insights from the simulation to refine targeted advertising strategies
2. Product Managers: Product Managers who already possess a rich set of user personas can make use of such a framework as a test bed.
3. Consumer Behavior Analysis: Market researchers can apply the framework to analyze consumer behavior and preferences.
4. Crisis Management : The reactions and decision strategies in such critical situations can also be modeled using this framework

Term Paper

The persistent and overarching limitation of our problem formulation was the broadness of its scope. When forming our hypotheses, questions like "How do various social network structures and degrees of anonymity affect the propagation of memes online?" seemed reasonably focused but proved to be difficult (though, admittedly, we were warned about this in previous submissions!). In the real world, there are so many factors at play that make it difficult to measure or simulate these things with conclusive results. Defining and measuring meme "engagement" quickly expanded the scope of our work, and programmatically determining the content and sentiment of memes, real-world network structures, levels of user anonymity, and so many other things proved to be nearly impossible challenges given our time constraints. One of those limitations deserves a special mention: meme data is simply very challenging to get ahold of. There aren't really any notorious "meme datasets," and we quickly found that other, related research had to come up with creative ways to model meme engagement. For example, the work of Lonnberg et al. (2020), which served both as validation and inspiration for our simulation framework, simply resorted to validating their own SIR model by overlaying it with Google Trends data.

In consuming published work related to this topic, we did find that significant research has been accomplished regarding image-to-text translation for memes, sentiment analysis, and strategies for countering disinformation. However, it was less easy to find open-source code that spawned from this research, and we did not feel like we could write a system to do all these things in real-world tests in the allotted time. Thus, we settled on the main idea of expanding upon the work of Lonnberg et al. (colloquially known in this report as the "Lonnberg paper" or the "Lonnberg model"), and developing a simulation that would reproduce their work in a deep way, i.e., not just the output of a series of equations that resemble real-life data, but a dynamic network of agents that interact with each other in various ways for carefully calculated reasons that, when measured cumulatively, produce something that does look like their model, and thus, the real world.

There were, of course, many challenges with this, too. First, we had to have a deep understanding of the Lonnberg paper, which involved delving into epidemiology more than I expected. Their work is an augmentation of the SIR model, which is very cool, but this was a whole new paradigm of mathematics – specifically stochastic differential equations – that we had to familiarize ourselves with. It proved difficult to translate their raw formulas into Python code that we could ideally use as a baseline "simulation" to validate our results against, or, at worst, use to sample if our simulation failed to produce acceptable results. The main challenge then was, of course, building out the framework of our simulation and tuning its parameters to produce something that resembles the Lonnberg model. In a stroke of good luck, this came relatively quickly, and we then were able to start running numerous experiments with interesting combinations of parameters that could actually provide insights about our hypotheses.

In terms of future work, we feel that we have set the stage for numerous other experiments, both new and extensions of past research that we considered when building our simulation framework. First, we have a mechanism now in our simulation that utilizes the theory of emotional responses b Russo et al. (2023)

a mechanism to combat disinformation (which is an attribute of the Meme class in our program). The work of Russo et al. formulated the idea of responding to disinformation-spreaders with personalized, highly emotional replies that try to evoke empathy, and assembled a database of over 12,000 disinformative claims and emotional responses to them. We were interested in sampling this database as a part of our simulation, but the paper was published only days before we began our work, and their data was not yet public. However, this could add another layer of depth to the simulation in the future.

Additionally, another layer to the simulation could be "conversations" modeled by independent GPT-controlled agents. Our simulation already implements the ability for agents to reply to one another and adjust their influence weights accordingly. Pastor-Galindo et al. (2023) ran this experiment in a different framework with promising results. This could even be combined with the emotional response mechanism to produce a very detailed and lifelike simulation at the agent level, which would be, in our opinion, a very significant and novel contribution. A limitation, however, would be that using the ChatGTP API for agent-level responses would add enormous runtime penalties to the simulation, and could incur large fees due to high API usage. We would recommend starting with shorter, scaled-down simulations.

In the end, we think of the simulation framework we built as a sort of playground; we started with the Lonnberg model, expanded the depth and intricacy, and translated its configuration to something very human-readable. Although we ran numerous experiments in preparation for this paper, there are so many more interesting scenarios to model. We recommend anyone interested in our framework to focus on the idea of being creative and looking forward to any interesting results they might uncover. These results can and should be used to make insights about real life, but carefully: due to our time constraints, we only tested our simulation in limited circumstances, and so future work will likely require some additional tuning. That said, we found the results to be promising, and we are excited to leave the simulation open source on GitHub for the foreseeable future.

References

S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina. The eigentrust algorithm for reputation management in p2p networks. In *Proceedings of the 12th International Conference on World Wide Web*, WWW '03, page 640–651, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136803. doi: 10.1145/775152.775242. URL `https://doi.org/10.1145/775152.775242`.

H. Liu, X. Wang, L. Liu, and Z. Li. Co-evolutionary game dynamics of competitive cognitions and public opinion environment. *Frontiers in Physics*, 9, 2021. ISSN 2296-424X. doi: 10.3389/fphy.2021.658130. URL `https://www.frontiersin.org/articles/10.3389/fphy.2021.658130`.

A. Lonnberg, P. Xiao, and K. Wolfinger. The growth, spread, and mutation of internet phenomena: A study of memes. *Results in Applied Mathematics*, 6:100092, 2020. ISSN 2590-0374. doi: https://doi.org/10.1016/j.rinam.2020.100092. URL `https://www.sciencedirect.com/science/article/pii/S2590037420300029`.

J. Pastor-Galindo, P. Nespoli, and J. A. Ruipérez-Valiente. Generative agent-based social networks for disinformation: Research opportunities and open challenges. 2023. URL https://arxiv.org/abs/2310.07545.

A. Pellet-Rostaing, R. Bertrand, A. Boudin, S. Rauzy, and P. Blache. A multimodal approach for modeling engagement in conversation. *Frontiers in Computer Science*, 5, 2023. ISSN 2624-9898. doi: 10.3389/fcomp.2023.1062342. URL https://www.frontiersin.org/articles/10.3389/fcomp.2023.1062342.

D. Russo, S. P. Kaszefski-Yaschuk, J. Staiano, and M. Guerini. Countering misinformation via emotional response generation. 2023. URL https://arxiv.org/abs/2311.10587.

E. Segev, A. Nissenbaum, N. Stolero, and L. Shifman. Families and networks of internet memes: The relationship between cohesiveness, uniqueness, and quiddity concreteness. *Journal of Computer-Mediated Communication*, 20(4):417–433, 2015. doi: https://doi.org/10.1111/jcc4.12120. URL https://onlinelibrary.wiley.com/doi/abs/10.1111/jcc4.12120.