Christopher Elchik (cwelchik), Zack Haugan (zkhaugan), Ethan Telep (ejtelep)

Modeling the Spread of Trends Within Various Social Network Distributions

1. What problem are you addressing?

We are addressing the problem of understanding the diffusion of trends and information in a social network. Given the rise of the internet, the number of social interactions that each person encounters in their daily life has skyrocketed, leading to an increased exposure to new ideas and trends. For example, Large Language Models have seen a meteoric rise in popularity and usage in the last two years, and their adoption into multiple aspects of society took place in record time. Additionally, 'overnight' linguistic trends such as the rapid rise and fall of the interest in the term 'demure' due to a popular internet video are becoming more prevalent. As trends and information spread through social networks, the speed and extent of their diffusion is influenced by a complex set of factors. The problem arises in attempting to understand the dynamics of these properties, as certain trends take hold or fail even against educated predictions, as evidenced by popular and unpopular marketing campaigns. The spread of trends is a difficult and substantial concept to measure in existing social groups due to the fact that it is not easy to identify when someone first heard of an idea or to operationalize how engaged an individual is in a trend.

Now more than ever, trends can disseminate at a rapid pace, and with a limited understanding of this process, we seek to better understand this process through modeling and simulation. We hypothesize that a more connected social network will tend to spread trends quickly, but also result in a faster 'dropping' of said trend. However, we also hypothesize the existence of anomalies, where certain interconnected groups within a social network like cliques may be so strongly interconnected that they never become disinterested in an idea once introduced. We also predict that different distributions for the values of apathy and susceptibility will lead to different behaviors in the overall social network. Lastly, we predict that higher initial engagement values for the "trend-setter" will lead to a higher peak engagement for the entire network.

(See section 3 for an outline of how we computationally define apathy, susceptibility, engagement, and a trend).

2. Why is this problem important?

Understanding the spread of trends through social networks is important due to the severity of the possible applications when attempting to spread (or block) information as fast as possible, especially since the large-scale usage of social media has rapidly accelerated the spread of trends. Entities can therefore take advantage of understanding the flow of information, whether they wish to apply it for public safety (i.e., amber alerts), political gain, advertising, etc. By using such a model of the spread of trends, data from small studies such as interest groups could help companies estimate how much engagement a new idea would have on the general population.

3. **How will you address this problem?**

We plan to study the spread of information through social networks by developing varying agent-based models of social networks and simulating the flow of trends through these models, iteratively recording the agents' engagement with a trend at all steps. Additionally, we recorded the derivative of the society's overall engagement with respect to time, so we can analyze the speed at which trends spread (and die out) in various social networks.

We plan to acquire existing data of social networks as a starting point to construct our agent-based network. As such, we opted to not create custom, special-case social network scenarios (i.e., "celebrity" agent scenarios), because we were strictly interested in the behaviors of naturally-occurring social networks that are publicly available. Understanding the behaviors of natural social networks allows for us to more effectively generalize our simulated results to the real world, as opposed to attempting to generalize the results of randomly generated networks.

To best simulate a natural social network, we used Stanford's SNAP Facebook Social Circles Dataset. This dataset includes real-world friend lists from Facebook's social network, collected via a survey, where the social network graphs are stored in an edge list format. To store graphs computationally, we utilized the NetworkX python package, using its DiGraph feature for directed graphs. We intertwined NetworkX's graph abilities with Mesa's agent-based simulation framework to simulate the flow of trends throughout agents in the network. All visualizations are created via MatPlotLib.

We used the following four networks for simulations of real-world social networks (ordered by file size):

Network	**Nodes**	**Edges**	**Average Degree**	**Clustering Coefficient**	**Transitivity**
3980.edges	52	292	11.23	0.4617	0.4500
686.edges	168	3312	39.43	0.5338	0.4536
3437.edges	534	9626	36.05	0.5437	0.4489
1912.edges	747	60050	160.78	0.6354	0.7000

Figure 1: Characteristics of Social Networks from SNAP Social Circles Dataset

Our model consisted of a NetworkX DiGraph, where agents are nodes and relationships are edges. The edges were unidirectional, to reflect the "follower"/"following" relationships commonly seen in social media. An agent's outgoing edges represent who they follow, and an agent's incoming edges represent their followers. Because we are modeling a snapshot of a virtual space, these relationships between agents will remain static. That is, agents will keep their followers/following throughout the study and not interact with others who they don't follow. The dynamic part of the study will be observed in the values held by each edge/node, as outlined below.

We will add variables to the agents (nodes) and relationships (edges) to impact how they propagate information to other agents. These internal variables will include:

- Susceptibility to new trends (decimal between 0 and 1)
  - Property of each edge
  - This property will be modeled unidirectionally, meaning that an agent A may be highly susceptible to trends coming from an agent B, but not vice versa.
  - This value would influence how much engagement can be transferred from a social interaction between two agents, with a higher value meaning that more engagement is transferred.
- Apathy rate (decimal between 0 and 0.5)
  - Property of each individual agent
  - Refers to how quickly the agent gets "bored" of the trend and drops it, with a higher value meaning that the agent's decrease in engagement is greater
- Engagement value (decimal between 0 and 1)
  - Property for each agent
  - Refers to how "engaged" an agent is with a particular trend
- Trend (decimal between 0 and 1)
  - Computationally, a trend will be modeled as the average of all agent engagement values

We decided to vary the simulations based on the following initial conditions:

Graph Structure: Which social network was used as the structure to perform the simulation.
- One of [3980.edges, 686.edges, 3437.edges, 1912.edges]

Initial Engagement: For the simulations, a single agent was chosen as the trendsetter, meaning that they alone have a non-zero initial engagement, while all other agents in the network have 0 initial engagement. This value can be thought of as the initial trendsetter's enthusiasm about the trend they hope to spread, with 1 representing maximum enthusiasm.
- One of [0.25, 0.5, 0.75, 1]

Susceptibility Distribution: These values represent the lower and upper bounds for the uniformly randomly generated susceptibility values for each edge in the network. The three values for this variable can be thought of as describing a barely susceptible, mildly susceptible, and highly susceptible society in relation to a trend.
- One of [(0,0.5), (0.25,0.75), (0.5,1)]

Apathy Rate: This value is related to both how much an agent's engagement decreases and how much each susceptibility value for each incoming edge for the agent decreases per step. The three values for this variable can be thought of as describing a barely apathetic, mildly apathetic, and highly apathetic society in relation to a trend.
- One of [0.1, 0.25, 0.5]

To perform proper hypothesis testing, we ran every permutation of initial conditions as stated above, choosing to vary one of the three non-graph structure variables while keeping the other three variables constant per graph. This allowed for easy analysis for independent variables, as we could analyze data that kept three variables constant while varying one of the conditions.

With each iteration of the simulation, each agent is updated according to the following code to update each metric. To determine how agent engagements are updated, we implemented the model using probability, where the probability that an agent increases their engagement with a trend is based on how susceptible they are with the neighbor introducing the trend to them. Through experimentation, we found that this method for updating an agent's internal variables yielded the most natural results.

Python

```python
# Iterate over each neighbor with an incoming edge
for neighbor, susceptibility in self.neighborsSusceptibility.items():
    # get neighboring agent
    nbr_agent = self.model.agentDict[neighbor]

    # probability of increasing engagement = the susceptibility of the
incoming edge
    if random.uniform(0,1) < susceptibility:
    # Increase engagement by the neighboring agent's engagement scaled by
the suceptibility of the edge
        self.updatedEngagement += susceptibility * \
nbr_agent.getEngagement()
    else:
        # Decrease engagement by amount proportional to apathy
        self.updatedEngagement -= (self.apathy_rate) * 0.1

    # Keep engagement between 0 and 1
    self.updatedEngagement = min(1, max(0, self.updatedEngagement))

    # scale edge susceptibility by apathy every step
    self.updatedNeighborsSusceptibility[neighbor] = \
self.neighborsSusceptibility[neighbor] * (1 - self.apathy_rate)
```

Figure 2: Relevant Simulation Code used for Updating an Agent's Internal Variables

For each time step of the model, each agent's updated values are calculated, but not changed immediately. This is so that for any arbitrary order of agent updates, during one 'step' of the model, an agent is only affected by its immediate neighbors. After each agent's values are updated, the update is put into effect by the model, using these updated values for the next step.

We collected data over 200 iterations for each model, as 200 steps gave plenty of time to capture all interesting aspects of the simulation, showing trend patterns before the metrics converge.

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

An alternative to this approach would be analyzing existing trend data in certain communities, as there are plenty of studies that analyze the spread of information in isolated communities. However, these studies are one dimensional, and represent a singular specific scenario. Our approach allows us to tune parameters to see the resulting macro changes in information distribution, allowing for a more subtle approach to the problem. More specifically, our approach allowed for us to tune apathy rates, initial engagements, and susceptibility, all in various social networks of different sizes and connectivities.

Another alternative would be to model this information using existing deterministic mathematical models, such as the diffusion of innovation or even disease models (Kudryashov et. al.). However, existing models limit our ability to explore how grouped social behaviors (i.e., the susceptibility/apathy of a society) affect trend propagation. The introduction of engagement value, apathy rate, and susceptibility are used to further explore these highly personal agents and relationships. Additionally, our probabilistic model includes an aspect of randomness that the mathematical models lack, allowing for our model to align more closely with the randomness inherent in human interactions. By using a natural model of individuals in regards to their particular tendencies of trend acquisition, dissemination, and loss of interest, we may be able to gain greater insight into what properties most influence the spread of trends in a social network.

One challenging aspect of this project was determining a system to modify an agent's engagement based on the various variables in the system. After much experimentation, we settled on the current system as described in Figure 2. When updating the values of each agent in the simulation, we decided to increase engagement in a manner that takes into account both the engagement level of the neighbor introducing the trend and the susceptibility of the incoming edge (representing the one-way connection from the neighbor to the agent). We believe that this approach is intuitive, as real-world trends are often spread via highly-engaged people. Additionally, the decrease in engagement is directly proportional to the agent's apathy rate, which we believe is also intuitive.

We could have alternatively generated our own social networks, using the Watts-Strogatz Small World network generation method. However, we opted to remain with the real-world dataset for more interesting insights that are applicable to real social networks. Modeling each edge, or connection between two agents, in the graph as having its own susceptibility value allows for a more realistic model of human interaction in the simulation. We believe that this approach provides a more accurate, controllable, and generalizable model for investigating how social behaviors influence the spread of trends.

5. **How did you evaluate your approach?**

We analyzed each variable of the system experimentally, changing one variable at a time while keeping all others constant. These variables consisted of Graph Structure, Initial Engagement, Susceptibility Distribution, and Apathy Rate (as outlined in Section 3). For each trial, we recorded the average engagement of the model for each step, and the derivative of this engagement value for each step. We created plots for the average engagement over time, and included the maximum and minimum rate of change for the total engagement on the graph

We decided to confirm the validity of our results through comparison with the patterns seen in third-party trend analysis from Google Trends. We decided upon Google Trends because it is currently the leading and most trusted service for examining the popularity of trends in society. For example, below shows the google search interest over time for the Fidget Spinners trend which took the internet by storm in 2017.

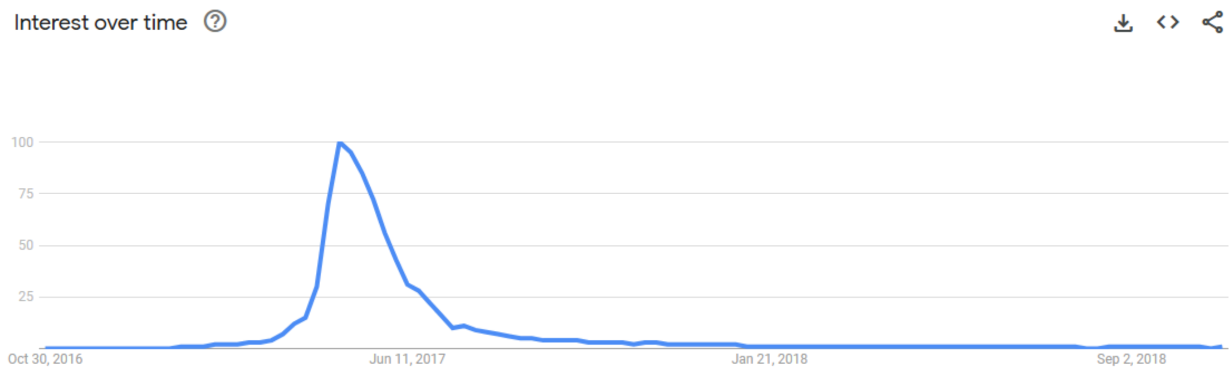

Figure 3: United States Google search interest in the term "Fidget Spinner" over time

Additionally, the TikTok linguistic trend "Demure" shows a similar pattern of a sharp rise in overall interest, followed by a peak and a slow decline.

Figure 4: United States Google search interest in the term "Demure" over time

Both of these examples highlight a general pattern in trend interest or engagement over time, and is something that we sought to capture in our simulation. Additionally, we decided to use these "overnight" trends as examples, with a quick appearance and gradual decline, as this best reflects the usual human tendency to acquire trends quickly, while losing interest tends to be a longer process.

For certain values of our initial variables, we found that these patterns can emerge from our simulation, highlighting the ability of our simulation to emulate social interest in real-world trends. Then, through analysis, we can deduce the variable(s) that play the largest role in creating this pattern (i.e., apathy, susceptibility, social network composition, initial engagement, etc.).

6. What are your main findings of the project?

One finding we noticed was the impact of edge susceptibility on the maximum popularity of a trend, as well as the speed of a trend's growth. In Figure 5 below, we see that as the susceptibility increases, the maximum engagement of the trend (height of the peak) increases. Additionally, the maximum derivative of the total engagement (the maximum speed at which the trend spread), is higher (represented as the first number in the last tuple in the legend) for more susceptible populations. We believe that this result reflects how a more susceptible society can more easily be influenced by a trend.

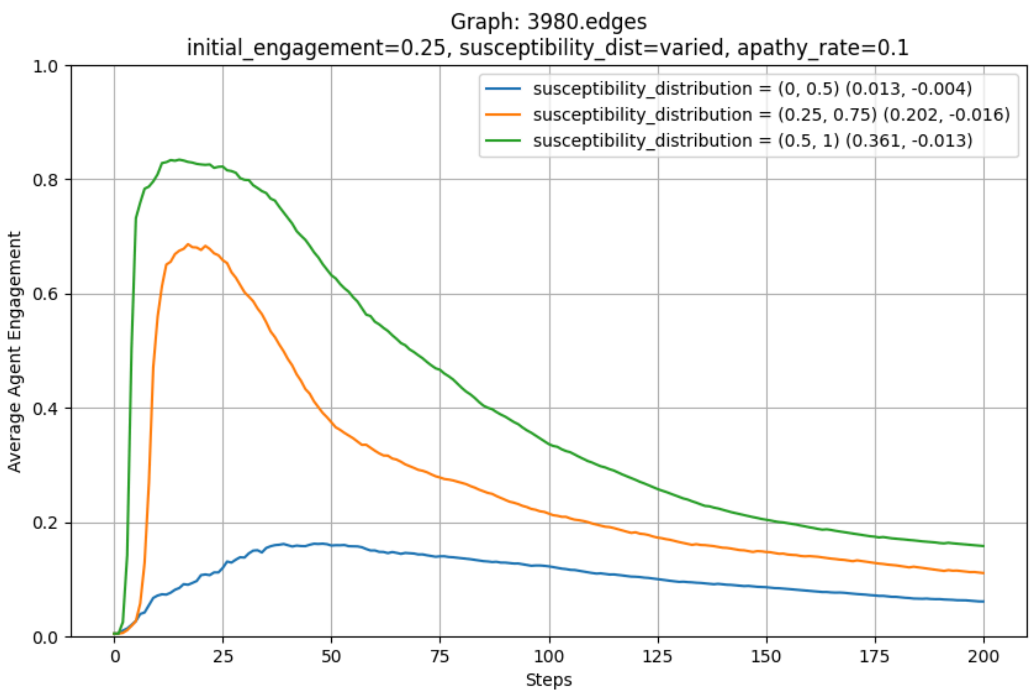

Figure 5: Graph of Total Engagement Over Time with Varying Levels of Society Susceptibility

Overall, we found that modifying the sole trendsetter's initial engagement did not greatly impact the distribution of the final graph. However, under certain circumstances like those in Figure 6 below, we found that low values of initial engagement could result in the trend not taking off at all, as evidenced by the blue line associated with the lowest initial engagement. We interpret this finding as the idea that there is a minimum enthusiasm that the initial trendsetter must have in order for their idea to take off in a society. Additionally, we found a slight correlation between initial engagement and maximum total engagement, with higher initial engagement leading to a higher peak engagement for the simulation. However, this correlation is not very significant.

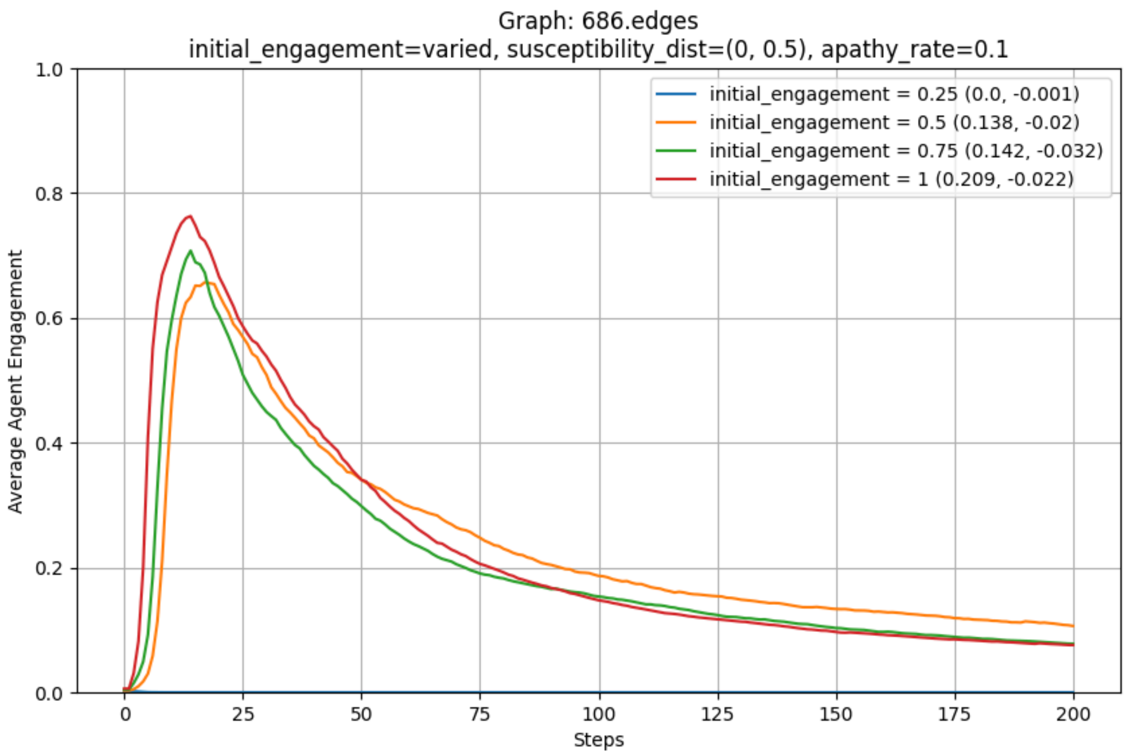

Figure 6: Graph of Total Engagement Over Time with Varying Levels of Initial Engagement

We also found in Figure 7 that when there is a high initial engagement of the "trend-setter," and low susceptibility across a society, the trend was ultimately unable to take off unless there was a very low apathy rate. Even then, with the low apathy rate, the peak societal engagement was still very low, capping at a little over 0.2. In society, this could suggest the importance of susceptibility in spreading a trend, as societies with the lowest apathy rates were barely able to spread a trend in this scenario. However, when comparing Figure 7 and Figure 8 (both below), notice that the susceptibility is slightly increased (while all other variables are the same), and trends were able to spread much more rapidly.

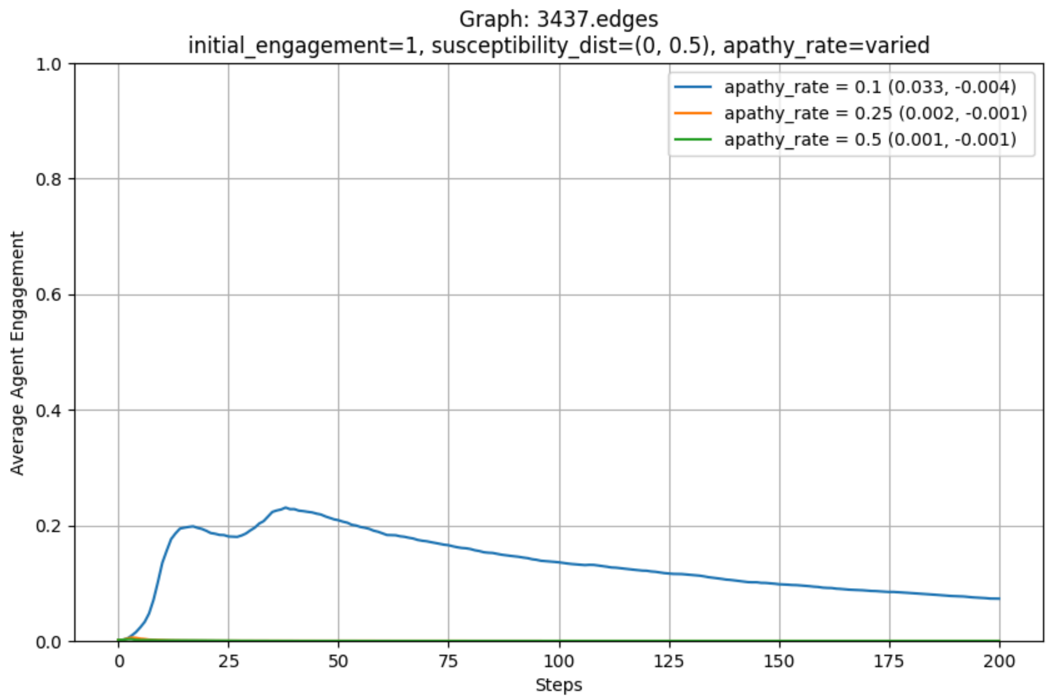

Figure 7: Graph of Total Engagement Over Time with Low Society Susceptibility and Varying Levels of Apathy Rate

When varying the apathy rate in a mildly susceptible society, we noticed that the societies with a lower apathy rate tended to have higher levels of maximum engagement with the trend, as well as a faster adoption of said trend. We also noticed that the final value of total engagement for the simulation (described as an indicator of trend retention) is higher for lower apathy rates. This makes intuitive sense, as a society with lower apathy will retain engagement with a trend for longer.

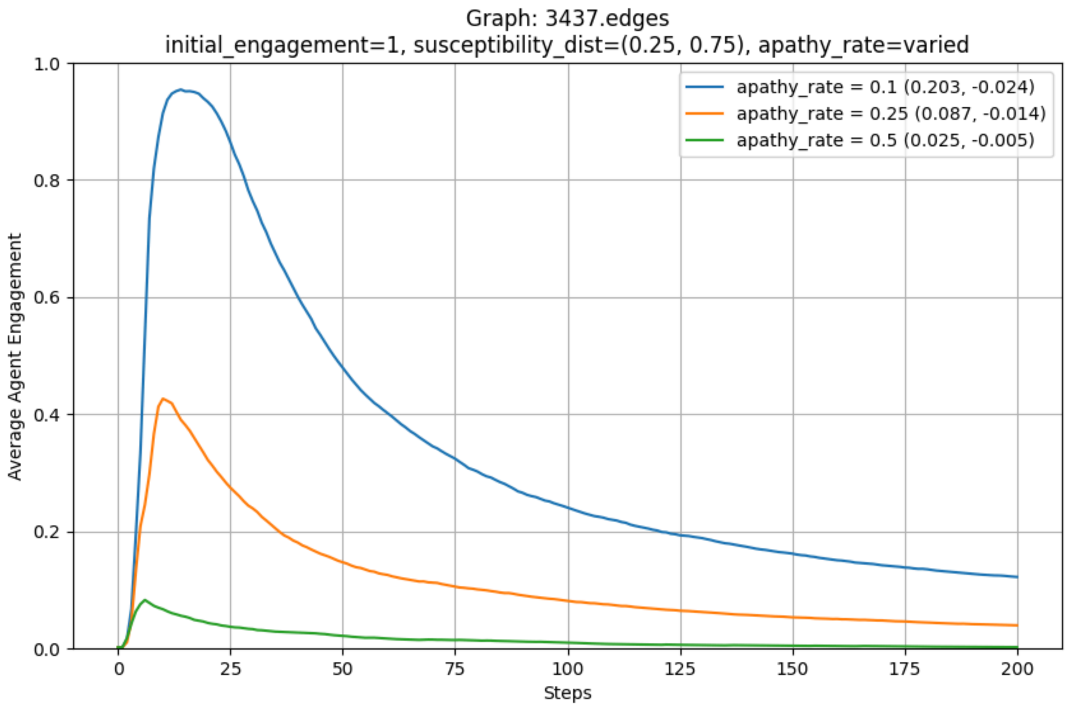

Figure 8: Graph of Total Engagement Over Time with Medium Society Susceptibility and Varying Levels of Apathy Rate

Upon analyzing the spread of trends across the different types of social networks, we found no meaningful relationship between the connectivity of a network and the spread of a trend. Also, throughout all simulations, we were unable to capture an instance of a trend re-emerging, although there were many simulations in which the rate of loss of engagement was very slow.

7. Practical applications and potential misuse of our approach

The techniques developed in this project have the potential to impact a range of stakeholders, including public health agencies, social media platforms, marketing firms, and policymakers, all of which may need to spread information as fast as possible. Public health agencies could leverage these methods to improve the dissemination of critical information, such as during health crises, while social media platforms might enhance user engagement by refining algorithms to prioritize meaningful and positive trends. Marketing firms could benefit by identifying key influencers within networks, allowing for more effective targeting, and policymakers might use these insights to shape public opinion or distribute essential information more efficiently.

However, alongside these benefits, there is a risk of misuse, such as manipulating public perception, spreading misinformation, or reinforcing societal divides. Studies show that on Twitter, “true stories [take] about six times as long to reach 1,500 people as it does for false stories to reach the same number of people,” (Dizikes) highlighting the higher susceptibility of social media users to fake news. Given the greater ability for false information to be spread, an adversary could use this model to uncover methods to more effectively spread misinformation in social networks.

The approach offers significant advantages, including amplifying positive messages, fostering awareness, and uniting communities around shared goals. Yet, there are inherent risks of harm. For instance, the methods could be exploited to spread discriminatory content, manipulate public sentiment, or target vulnerable populations with harmful campaigns. Such misuse could deepen societal inequalities, perpetuate abuse, and damage public trust. The most effective way to mitigate any harmful use of our results is to make the public aware of where they may be most susceptible, as the best way to combat the spread of false information is to be aware of one’s vulnerabilities and think critically.

References

Dizikes, P. (2018, March 8). Study: On Twitter, false news travels faster than true stories. MIT News; Massachusetts Institute of Technology. https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308

Kudryashov, N. A., Chmykhov, M. A., & Vigdorowitsch, M. (2021). Analytical features of the SIR model and their applications to COVID-19. In *Applied Mathematical Modelling* (Vol. 90, pp. 466–473). essay, ScienceDirect. Retrieved October 4, 2024, from https://doi.org/10.1016/j.apm.2020.08`.057.

Google. (n.d.). Google Trends. Retrieved November 22, 2024, from https://trends.google.com/.

Leskovec, J., & Krevl, A. (2014, June). SNAP Datasets: Stanford Large Network Dataset Collection. Retrieved November 22, 2024, from http://snap.stanford.edu/data