Investigating social phenomena prevalent in online Covid-19 discourse through network-agent based modelling

Anthony Reidy 18369643

Abstract

Vaccination programs for COVID-19 in high-income countries have benefited from their respective country's ability to produce high quality and thought-provoking public health campaigns. In parrel, there is growing amounts of Covid-19 misinformation online and thus, vaccine hesitancy. Countries such as Austria, have contemplated with introducing public health mandates to combat rises in the virus. In this paper, we propose a network-agent based model to simulate the spread of opinions on vaccine uptake. We explore how social compliance and social influence at the micro-level, can lead to poor information flow between contrasting opinions. Specifically, we utilise small world and scale-free networks, commonly used to model social networks. First, we demonstrate that our model mimics information flow in social networks by capturing a number of social network phenomena: Majority opinion, and opinion polarisation. Secondly, the rules of our model are validated by observing phenomena such a the spiral of silence and echo chambers that we did not explicitly code for. Thirdly, we show the influence of malicious agents and multiple opinions on the forming of these phenomena. A correlation between the number of malicious agents and the number of echo chambers is observed. Furthermore, we find malicious agents may be divisive, causing the network to become more one sided. Finally, we observe that if more opinions on vaccine uptake exist, the stronger the effect of the number of malicious agents on both the *formation* of echo chambers and the *ability* to influence the majority opinion.

Code: `https://gitlab.computing.dcu.ie/reidya3/ca4023-cap-reform`

1 Purpose

The rapid adoption of social media has left little time for a discussion of its sociol, political and cultural ramifications. Communication systems designed predominantly to engage the user's attention rather than foster discussion, may exacerbate people's self-insulation, create social fragmentation and promote ideological extremism (Sunstein 2017). Social media, while providing an unprecedented capacity for the public to communicate, has also been a major factor in the rise of fringe opinions damaging to public health. Reconciling principles of free speech with the policing of social media for damaging falsehoods remains a conundrum for many high-income democracies (Wilson & Wiysonge 2020). Modelling opinion spreading can give insight into the dynamics of public opinion formation, but also into the survival of minority opinions and the rise of extremism. One commonly use-case of agent-based model is for the simulation of opinion dynamics, in which each agent has an opinion that can be affected by other agents. This assignment presents an agent-based model that simulates the spreading of opinions in order to explore how an individual tendency for homophily at the micro-level can lead to poor information flow between contrasting opinions at the macro-level, in scale free and small world networks. We seek only to find the global effects of and validate our model against well understood phenomena, such as confirmation bias (Oswald & Grosjean 2004), the spiral of silence (Noelle-Neumann 1974), and echo chambers (Terren & Borge-Bravo 2021). In addition, we simplify the nature of social interaction, in order to isolate the systems dynamics associated with the inertia towards the average opinion. While in most models, a connecting link is sufficient to enable communication, we also base the probability of talking between agents to be a function of their history of interactions. Mimicking our tendency to reinforce bonds with those who support our own perceptions. By allowing agents to take two to six opinions on vaccinations, we replicate the conditions for nuanced discussion. To ascertain the effect of targeted opinion manipulation, we run several experiments that set a number of the most central nodes to act as malicious agents. By testing this effect with different networks, we may find which network characteristics afford the most resilience to targeted attacks on hubs. To the best of our knowledge, this model contains combines social theory like silent spirals and mechanics such as reputation-based trust, which has not been combined before.

2 Model Toplogy

In our model, agents are analogous to people. The model is based on the probable events that agents talk with each other and exchange on opinions. The principles of agent interaction, opinion exchange and maintenance of a trust system form the basis of this model. The following aspects were considered to characterize our agent-based model, (i) opinion domain, (ii) interacting agents neighbourhood and direction, (iii) updating function (iv) type of network

2.1 Opinion Domain

The first choice that has to be made is the numeric value chosen to represent the opinion on vaccine uptake. Though an opinion is intrinsically a qualitative and potentially multi-faceted feature, its study through an agent-based model requires it to be described by a numeric variable. Researchers in this domain generally adopt a binary opinion variable (Galam 2002, Afrasiabi et al. 2018, Shang et al. 2021), non-binary discrete variable (Martins 2008, Gambaro & Crokidakis 2017) or a continues variable over a bounded interval (Dong et al. 2016, Douven & Wenmackers 2017, Grabisch et al. 2018).

We seek to unify these different approaches. Firstly, we introduce the *opinions* hyperparameter which is a discrete variable, that can range from 2 to 6. This allows us to map each opinion value to a qualitative feature. For example, in the case of a binary opinion variable, we use the two values to represent a positive versus a negative opinion towards a COVID-19 vaccine. For 3 opinions, negative, neutral and positive are the qualitative features and so on. Secondly, the concept of preference is introduced (which is an initial uniformly distributed continuous variable from 0 to 1). This describes the strength or confidence of a person's belief in their chosen opinion. We do this (i) to represent people with radical opinions (opinion polarization, defined as preference greater >0.8) (ii) to model social influence. Social influence is the process by which individuals adapt their opinion, revise their beliefs, or change their behavior, as a result of social interactions with other people (Moussaïd et al. 2013). This behaviour is particularly prevalent in social networks due to their connective nature. To be more exact, we adopt this preference parameter to model two major attractors of opinion in COVID-19 discourse: the 'expert effect' (Das et al. 2018), induced by the presence of a highly confident individual in the group, and the 'bandwagon effect' (Bindra et al. 2022), caused by the presence of a critical mass of laypeople sharing similar opinions.

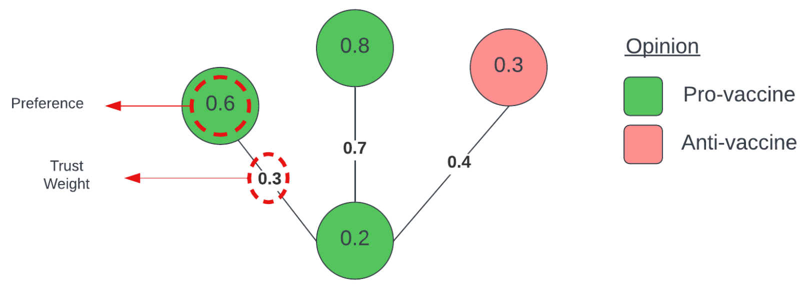

Figure 1: Network Topology

2.2 Network used

Both the Barabasi-Albert (scale-free) and Watts–Strogatz (small world) models are offered though the network_type hyperparameter (1 for Watts–Strogatz and 2 for Barabasi-Albert). A small-world network is a model in which most nodes are not neighbors of one another, but the neighbors of any given node are likely to be neighbors of each other. However, this model does not taken into account people with lots of friends. A scale-free network attempts to resolve this problem by capturing the presence of large hubs (a few nodes that are highly connected to other nodes in the network). Many networks have been reported to be scale-free, although statistical analysis has refuted many of these claims and seriously questioned others (Broido & Clauset 2019). Thus, we provide the user of utilising either or.

In the network, each agent does not leave its node position, nor form new links. A node, is an agent, capable of talking, listening, forming, shifting, changing or keeping an opinion. Each edge in the network has a weight. This weight is associated with the concept of "trust". Psychologists have known for decades that an important heuristic that humans employ in deciding whether to trust someone is the norm of reciprocity (Zonca et al. 2021). The norm of reciprocity refer to social strategies that individuals learn which prompt them to react to the positive actions of others with positives responses and the negative actions of others with negative response (Santos et al. 2021). To quantify this, imagine agent A_i's is considering A_j trustworthiness. The level of trust $(W_{i,j})$ between those agents is calculated as:

$$W_{i,j} = \frac{OT_{i,j}}{E_{i.j}}$$

where $OT_{i,j}$ is the number of times that A_i and A_j had the same opinion in the past and $E_{i,j}$ refers to the total number of encounters between two agents. Our intuition behind this is that agents are more eager to trust the opinion of another agent if their opinions have matched in the past. Generally, people would consider the possibility of taking a vaccine more seriously if it came from a friend who had a similar viewpoint. Therefore, an agent's probability to accept another's opinion as their own, depends on the proportion of past agreements. Agents are assigned edges once, at the start of a simulation but update their connection strengths through time. We initialise each 'trust' weight value as 0.5 to reflect an initial unbiased trust between agents.

2.3 Neighbourhood

Agents only ever consider interacting with their adjacent neighbours in the network. This is quite natural as the agents are actually part of a social network as it simulates the 'follower' or 'friends' relationship. Each node is connected to k (set by the N_neighbours hyper-parameter) nearest neighbors in the topology. As described in section 2.2, people who have have similar opinions are more likely to discuss the subject and therefore their opinions spread more easily between them. However, the spreading of opinion is not only dependent on the similarity of agents, but also on the similarity of the opinion that is being spread **at the moment**. Thus, agent cooperation is dependent on both the trust and the similarity of the opinion between to agents. Agents will decided whether to talk to their neighbours, via the following equation:

$$\begin{cases} |P_i - P_j| < \eta & \text{if } O_i = 0_j \\ trust > uniform[0,1] & \text{else} \end{cases}$$

where O_i and 0_j refer to the opinions of A_i and A_j, P_i and P_j refer to the preferences of A_i and A_j, and η is a user settable hyper parameter (similarity threshold). Agents only consider 'talking' to neighbours that fit the above threshold. This is done to replicate the phenomena of confirmation bias. Confirmation bias is the tendency of an individual to acquire and interpret information in such a way that one's existing beliefs and opinions are confirmed (Ling 2020). In addition, the order in which the agents talk is randomized in each step, as we feel that this resembles the real world.

However, the potential communicators of each agent is kept constant (i.e. the agents never move to meet other agents) in this particular model, Therefore, the above mechanism could result in unnaturally isolated agents. Our else statement rectifies this. We assumed that, if no agents has any similar opinion to other agents, it would be natural to interact with some agents. Therefore, if the agent meets neighbours that had very similar opinions in the past (trust), there will be a probability that this agent will interchange its opinion with someone of the same mind.

2.4 Malicious Agents:

A malicious agent is one that seeks to effect the overall system, by spreading a bias preference. Malicious agents are added to help us understand the conditions for which targeted propaganda or misinformation are able to influence peoples opinions on vaccines. A malicious agent is one that seeks to effect the overall system, by spreading a bias preference in order to effect the whole system. Certain networks may be more resilient to this perturbation. In addition, these agents may have an effect on the topology and information flow of the system. Thus, malicious agents are simulated with a random opinion, a preference of 1.0 (the maximum preference) and are the most connected agents (nodes with the highest degree). In addition, each malicious agents edge's weights have a trust value of 1.

2.5 Swingers:

To model changes in opinion due to external events, we introduce the concept of swingers. Swingers are randomly select agent(s) that are reassigned a random opinion and preference at each time step. The hyper parameter 'swingers' controls the number of agents assigned as swingers.

3 Model Setup

3.1 Initialisation

The model initializes by generating a random network of nodes. Each node is then populated by an agent, who receives a random opinion and a random preference. If malicious agents are simulated, they are too generated with a random opinion and a preference of 1.0 (the maximum preference). The swingers are only generated for time steps greater than 1.

3.2 Update Step

The model consists almost solely out of the actions of the agents, whose behaviour is all defined within the 'talk' function of the agent class. Using the OOP paradigm, we were able to encapsulate the communicative and operative closure of an automatic system, as society has been described in previous literature (Sevänen 2001).

The first submodule ('choose neighbours') selects which neighbours to talk with. This process has already been defined in section 2.3. In addition, this function updates the weights of the network. The second submodule ('form opinion') regulates the internal evaluation of newly encountered opinions. The way of adopting one's opinion to the environment is based on earlier literature (Starnini et al. 2016), where the preference $P_i(t)$ towards an opinion of agent i on time t is updated by the following rule:

$$P_i(t+1) = P_i(t+1) + K(\sum_{j \in N_i(t)} p_j - s_i(t))$$

where $N_i(t)$ is the set of selected neighbours at time-step t and K is the 'social influence' hyper-parameter.

Finally, there is two external functions, perturb_network and update_malicious_agents which regulates swingers and malicious agents. The perturb_network submodule makes sure that a given number of agents 'swings' their opinion and preference to new, random values. Note that this means that the agents do not necessarily adopt another opinion (the newly random-generated values could be the same as the original values). The update_malicious_agents submodule resets the value of trust of malicious agents, as well as their opinion and preference.

Hyper-Parameter	Description
Num agents	Number of nodes/agents
N neighbours	Number of edges that spawn from each node
Network type	Use a Watts-Strogatz or a Barabasi-Albert network
Beta Component	The β-component when using a Watts-Strogatz network
Similarity threshold	This controls the probability that agents will only talk to like-minded agent
Social influence	The proportion of an agents opinion that will be determined by its peers
Malicious	Number of malicious agents
Swingers	Number of agents that are reset to a random opinion and preference each step.
Echo limit	The threshold in trust value between two or more agents for echo chamber classification.
All majority	If true, the system is initialized with all agents preferring the same opinion.
Opinions	The number of opinions considered.

Table 1: User settable hyper-parameters of our model

4 Intended outputs

Output	Description
Radical opinions	The proportion of all agents that have a preference of 0.8 or higher
Percentage majority opinion	The proportion of all agents that adhere the majority opinion
Echo chambers	The number of Echo chambers that have formed
Silent Spirals	The silent spiral score
Community no	The number of communities of uniform opinion that have formed
Average Trust	The average of each weight in the network.
Transitivity	The degree to which agents indirectly influence other agents.

Table 2: Output variables of the model

Radicalisation A agent or group of agents that each have a preference value above 0.8 is considered to hold a radical opinion. Through experimentation, our goal to find what might cause increased radicalisation. Although, one should be careful to not presume too much from this crude measure.

Echo chamber Echo chambers are groups of agents where like-minded people communicate and reinforce pre-existing beliefs. This is measured to investigate if confirmation bias occurs.

Percentage Majority opinion The percentage of nodes that hold the majority opinion is tracked to determine under what conditions does the network become one-sided in their outlook.

Silent spiral The spiral of silence theory (Noelle-Neumann 1974) explains changes in people's willingness to express their opinion as the result of a fear of being socially isolated. People sense the opinions on controversial topics of those around them and modify their public behaviour accordingly. Over time, this results in the formation of a consensus, the establishment of a social norm. To find if those with minority views tend to communicate less, the least trusted 5% of nodes are quantified (i.e. the nodes with the lowest trust values to all neighbours). This silent spiral output details the proportion of these potential silent spirals that are part of the majority opinion. Values near 1.0 would suggests that the low connectivity happens in nodes that follow the majority opinion, which are not silent spirals. A Low value, near 0.0, indicate that nodes with low connectivity are mostly part of minority opinions, This could be example of a silent spiral(s). The silent spiral causes minority opinions to be left outside the conversation and are therefore less perceived and esteemed by their neighbours. Thus, the minority opinion does not go away but rather is suppressed. Some have linked this to increased radicalization.

Community no We detect communities of opinion utilising the *weighted* Louvian clustering algorithm (De Meo et al. 2011).It maximizes a modularity score for each community, where the modularity quantifies the quality of an assignment of nodes to communities. We calculate this metric to evaluate how much more densely connected the nodes within a community are (based on trust values) under different scenarios, compared to how connected they would be in a random network.

Average Trust The average trust throughout the network is measured to verify if the agents who generally agree with each other tend to take to each others opinion faster.

Transitivity Weighted transitivity is calculated as average clustering coefficient, where

$$transitivity = \frac{T^c}{T}$$

T is the number of all triplets and T^c is the number of closed triplets. Transitivity is dependent on both the number of possible opinions and the number of introduced malicious agents. It provides us a way to measure the degree to which agents indirectly influence other agents

5 Validation

To validate our model, we found several phenomena in the literature such as the silent spiral and the echo chambers, that are associated with opinion formation. Having not explicitly modeled these effects, yet finding them in our model, we have validated our model with the corresponding social theory. The following results enlighten this fact in detail.

6 Results

6.1 Sensitivity Analysis

To perform sensitivity analysis, we considered both the the Sobol (Nossent et al. 2011) and the one-factor-at-a-time (OFAT) (Frey & Sudarsanam 2008) technique. The Sobol method (which is a global sensitivity analysis) decomposes the variance of the model output into fractions belonging to different input parameters or different combinations of input parameters. In contrast, the one-factor-at-a-time analysis is a form of local sensitivity analysis, where one input parameter is changed while the others remain constant. Given page constraints, we focus on the Sobol method.

6.2 Sobol

Silent Spiral: The results of the first and total order sensitivity analysis for the silent spiral are represented in Figure 2 A,B. The first order analysis only considers the linear and nonlinear contributions of the individual input parameters. This form of analysis shows that all hyper-parameters except for the opinions parameter, have a quite similar influence on the output variability, namely around zero. However, the confidence interval indicates that these parameters still might play a role. The parameter opinions seems to have the most impact on the silent spiral output in the first order sensitivity analysis. When also including the interaction effects (see Figure 2B), the silent spiral output is approximately equally sensitive to changes in most input parameters. The number of neighbours deviates the most, and has the least influence

Community no: The first order Sobol analysis illustrates that changes in the opinions, echo limit, social influence and similarity threshold hyper-parameters have little influence on the number of communities, whereas the number of swingers have a relatively large impact (Figure 2C). Figure 2D illustrates that all parameters explain some variance in the output in the total order Sobol analysis. Similarly to the first order analysis, the number of swingers has considerably the most influence

Echo Chamber: Figure 2 E,F shows that all hyper-parameters contribute to variance in the number of echo chambers when including interaction effects, although the echo limit does not contribute individually. The number of swingers have the largest influence. The second largest effect on the variance is caused by the parameter opinions in the first order analysis and by the similarity threshold in the total order analysis

Majority percentage: The first order Sobol analysis, which is depicted in Figure 2G, indicates that variance in majority percentage is most influenced by the opinions and N neighbours hyper-parameters. Based on the total order Sobol analysis (see Figure 2H), we found that all factors have an influence. Although, the similarity threshold and the number of neighbours have the most impact.

Radical Opinions: Figure 2 I,J depicts the first and total order variance for the radical opinions result. The fraction of radical opinions in the network is considerably most sensitive to changes in similarity threshold in both orders. However, the confidence interval for the similarity threshold is relatively large, meaning that the accuracy is quite small.

Transitivity: According to the results of the first order sensitivity analysis (Figure 2k), the input parameter opinions has the most impact on the network. The number of neighbours is the second largest contributor. The total order sensitivity analysis shows that, the parameters opinions, echo limit, swingers, social influence and number of neighbours have a similar influence on the transitivity, namely around 0.8. It must be noted that the similarity threshold has the most impact

6.3 Individual Scenarios

To further measure the effects of social compliance and social influence on vaccine uptake, two different scenarios are performed. Firstly, in scenario 1 the influence of malicious agents in the network is measured. In other words, we analyzed data of a hundred time steps for an ascending number of malicious agents. Secondly, in scenario 2 the influence of multiple opinions and multiple malicious agents on various output parameters is explored by collecting data after a hundred time steps. Note, for each scenario we run the simulation 20 times and averaged the results. In addition, for the uninvestigated input parameters, we used the defaults, as specified in the Gitlab repository.

6.4 Scenario 1

No influence was found, when varying the number of malicious agents on the proportion of population holding the majority opinion when two possible opinions were present (Figure 3A). Next, we calculate the number of chambers. Echo chamber formation is measured as the proportion of cliques (subsets of vertices, all adjacent to each other) that supported only one opinion. From figure 3A, it is clear that echo chamber formation are dependent on both the number of possible opinions and the number of introduced malicious agents.

6.5 Scenario 2

Figure 4 illustrates heatmaps that describes the results for scenario 2. In scenario 2, the number of echo chambers shows a negative correlation with the number of possible opinions. This seems intuitive, as the population is divided over more opinions. This leaves less agents per opinion. Thus, it is harder to form a chamber where everybody has the same opinion on vaccine uptake. There also seems to be a strong positive correlation between the number of malicious agents and the number of echo chambers, for higher numbers of opinions. As stated earlier, radical opinions are measured as the number of agents with a preference of 80 percent or higher. In scenario 1, radical opinions are dependent on both the number of possible opinions and the number of introduced malicious agents. However, in scenario 2, The number of radical opinions show no correlation with the number of possible opinions. The number of malicious agents seems to correlate positively with the number of radical opinions. The case of two opinions and no malicious agents seems to behave differently. It is unexpected for this case to have a relatively high number of radical opinions.

We are also interested in transitivity. Transitivity is dependent on both the number of possible opinions and the number of introduced malicious agents. The transitivity of the network shows a negative correlation with the number of possible opinions. The number of malicious agents seems to correlate positively with transitivity.

7 Discussion & Conclusion

In conclusion, a correlation was found between the number of malicious agents and the number of echo chambers. Malicious agents may be divisive, causing local social networks to become more one sided, reducing information integration and therefore, vaccine uptake. The dividing of the network structure, into the number of plausible opinions was a common result (see Figure 3 and 4). Although, it must be noted that the chamber structure is more localized. Perhaps, the most interesting result is the scaling of macro-level impact that comes from increased similar opinion receptivity. In other words, although one agent's preference is not dramatic, the cumulative effect of multiple agent is. The formation of echo chambers provides an intermediary step between individual filtering to poor global information flow. Agents tend to form sub-communities where they reinforce the opinion of the sub-community. It seems that an an agent's desire to both conform and have its own opinion reinforced, causes the agent to prefer to communicate with those of a similar opinion. Over time, this results in more stable opinion spreading dynamics at the cost of functional connectivity.

There are several possibilities to extend this work in future. One of them is to check if agents become less trustful if malicious agents are introduced in the network. In addition, we would like to track the average preference of both the minority and majority opinion in a binary ”opinions” input variable simulation. In this setting, we would investigate different scenarios such as initialising a distrustful network work (all trust weights set to a value less than 0.3), and varying the number of swingers and malicious agents.

References

Afrasiabi, M. H., Guérin, R. & Venkatesh, S. (2018), ‘Opinion formation in ising networks’, *Online Social Networks and Media* **5**, 1–22.

Bindra, S., Sharma, D., Parameswar, N., Dhir, S. & Paul, J. (2022), ‘Bandwagon effect revisited: A systematic review to develop future research agenda’, *Journal of Business Research* **143**, 305–317.

Broido, A. D. & Clauset, A. (2019), ‘Scale-free networks are rare’, *Nature communications* **10**(1), 1–10.

Das, R., Kamruzzaman, J. & Karmakar, G. (2018), ‘Modelling majority and expert influences on opinion formation in online social networks’, *World Wide Web* **21**(3), 663–685.

De Meo, P., Ferrara, E., Fiumara, G. & Provetti, A. (2011), Generalized louvain method for community detection in large networks, *in* ‘2011 11th international conference on intelligent systems design and applications’, IEEE, pp. 88–93.

Dong, Y., Chen, X., Liang, H. & Li, C.-C. (2016), ‘Dynamics of linguistic opinion formation in bounded confidence model’, *Information Fusion* **32**, 52–61.

Douven, I. & Wenmackers, S. (2017), 'Inference to the best explanation versus bayes's rule in a social setting', *The British Journal for the Philosophy of Science* **68**(2), 535–570.

Frey, D. D. & Sudarsanam, N. (2008), 'An adaptive one-factor-at-a-time method for robust parameter design: Comparison with crossed arrays via case studies'.

Galam, S. (2002), 'Minority opinion spreading in random geometry', *The European Physical Journal B-Condensed Matter and Complex Systems* **25**(4), 403–406.

Gambaro, J. P. & Crokidakis, N. (2017), 'The influence of contrarians in the dynamics of opinion formation', *Physica A: Statistical Mechanics and its Applications* **486**, 465–472.

Grabisch, M., Mandel, A., Rusinowska, A. & Tanimura, E. (2018), 'Strategic influence in social networks', *Mathematics of Operations Research* **43**(1), 29–50.

Ling, R. (2020), 'Confirmation bias in the era of mobile news consumption: the social and psychological dimensions', *Digital Journalism* **8**(5), 596–604.

Martins, A. C. (2008), 'Mobility and social network effects on extremist opinions', *Physical Review E* **78**(3), 036104.

Moussaïd, M., Kämmer, J. E., Analytis, P. P. & Neth, H. (2013), 'Social influence and the collective dynamics of opinion formation', *PloS one* **8**(11), e78433.

Noelle-Neumann, E. (1974), 'The spiral of silence a theory of public opinion', *Journal of communication* **24**(2), 43–51.

Nossent, J., Elsen, P. & Bauwens, W. (2011), 'Sobol'sensitivity analysis of a complex environmental model', *Environmental Modelling & Software* **26**(12), 1515–1525.

Oswald, M. E. & Grosjean, S. (2004), 'Confirmation bias', *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory* **79**, 83.

Santos, F. P., Santos, F. C., Pacheco, J. M. & Levin, S. A. (2021), Social network interventions to prevent reciprocity-driven polarization, *in* 'Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems', pp. 1643–1645.

Sevänen, E. (2001), 'Art as an autopoietic sub-system of modern society: A critical analysis of the concepts of art and autopoietic systems in luhmann's late production', *Theory, Culture & Society* **18**(1), 75–103.

Shang, L., Zhao, M., Ai, J. & Su, Z. (2021), 'Opinion evolution in the sznajd model on interdependent chains', *Physica A: Statistical Mechanics and its Applications* **565**, 125558.

Starnini, M., Frasca, M. & Baronchelli, A. (2016), 'Emergence of metapopulations and echo chambers in mobile agents', *Scientific reports* **6**(1), 1–8.

Sunstein, C. R. (2017), *# republic*, Princeton University Press.

Terren, L. & Borge-Bravo, R. (2021), 'Echo chambers on social media: a systematic review of the literature', *Review of Communication Research* **9**, 99–118.

Wilson, S. L. & Wiysonge, C. (2020), 'Social media and vaccine hesitancy', *BMJ Global Health* **5**(10), e004206.

Zonca, J., Folsø, A. & Sciutti, A. (2021), 'Dynamic modulation of social influence by indirect reciprocity', *Scientific reports* **11**(1), 1–14.

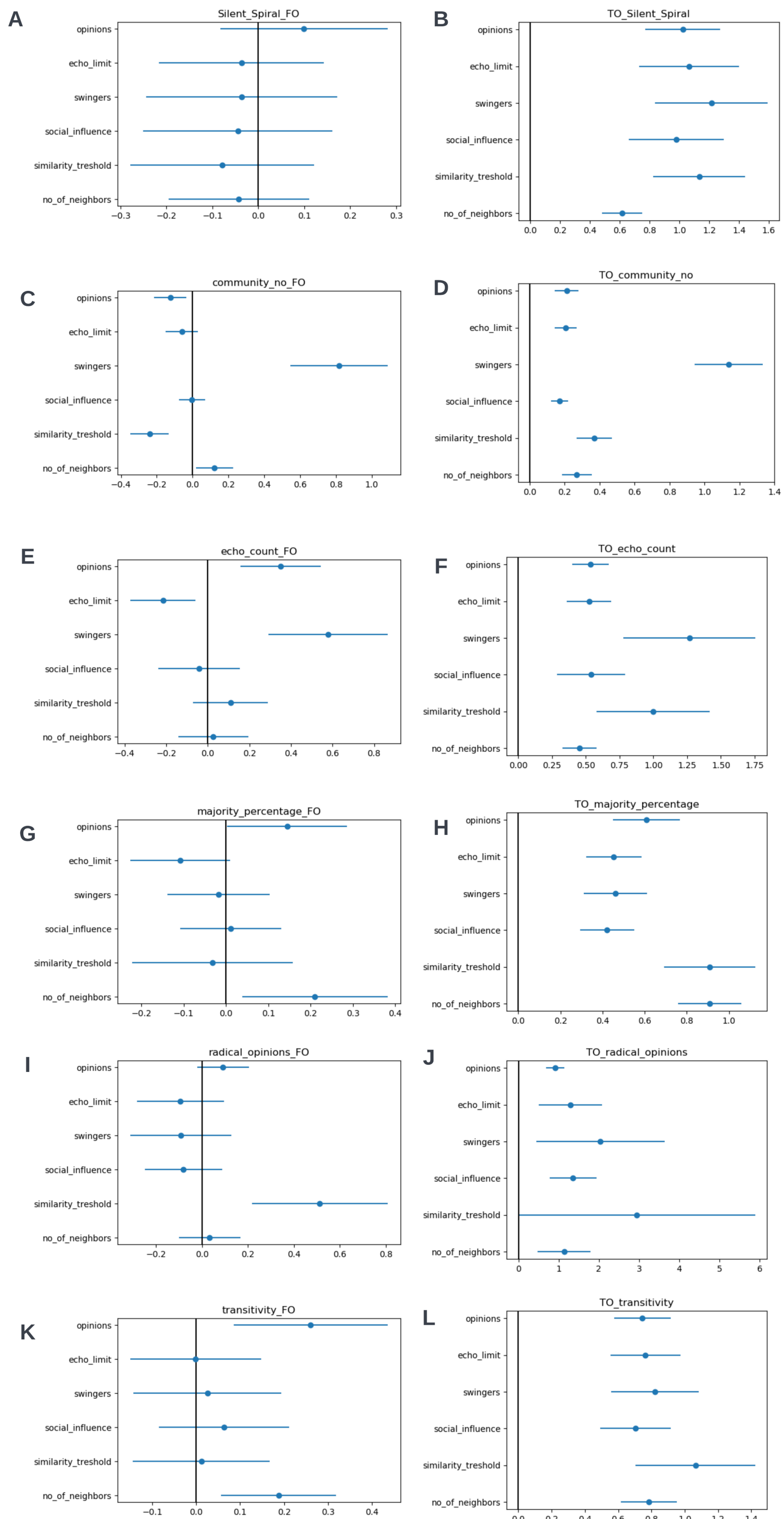

Figure 2: Results from the Sobodol method. This illustration displays graphs detailing the first-order and total-order Sobol indexes, varying the silent spiral (**A,B**), community no (**C,D**), echo chamber count (**E,F**), majority percentage (**G,H**), radical opinions (**I,J**) and transitivity (**K, L**) output parameters.

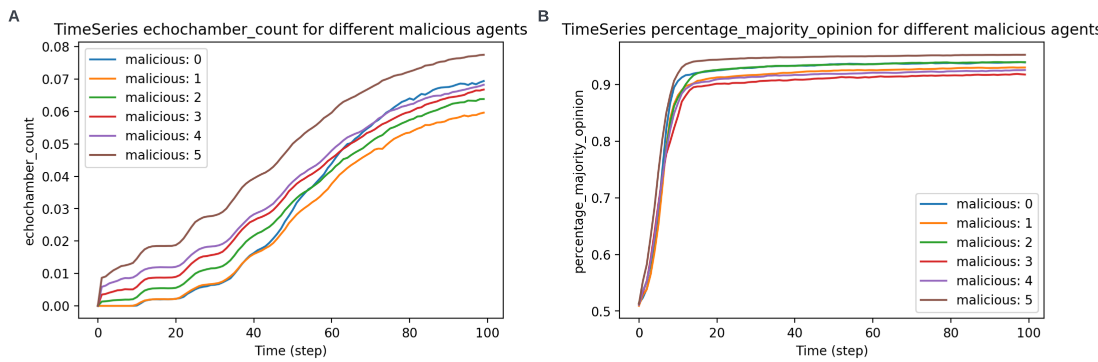

Figure 3: The two graphs displays results from scenario 1. **A:** A line graph detailing the count of echo chambers at different T, varying the number of malicious agents. **B:** A line graph detailing the percentage of agents who held the majority opinion at different T, varying the number of malicious agents.

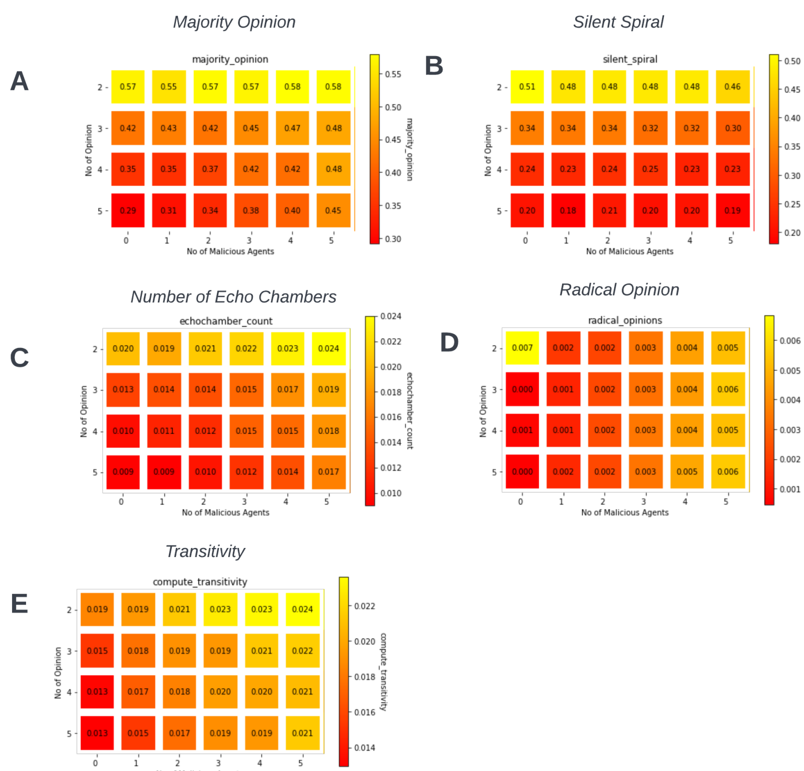

Figure 4: These 5 heat-maps display results for scenario 2. We vary both the number of opinions (columns) and malicious agents (row) to compute the following output parameters; percentage of agents who hold the majority opinion (**A**), silent spiral count (**b**) number of echo chambers (**c**), percentage of agents who hold radical opinions (**d**) and transitivity (**e**)