Topological effects in the spread of information and misinformation in social networks

Gonçalo Simões*
Instituto Superior Técnico, Universidade de Lisboa, Portugal
(Dated: October 5, 2021)

Fake news and misinformation are an increasing problem at local and global scale. This work focuses on the analysis of opinion dynamics and interventions strategies able of reducing the spread of fake news. To this end, we develop a novel simulator of spreading processes where social interactions are described as a network of connected nodes. We resort to different classes of scale-free networks to assess two types of interventions and rumour spreading dynamics: 1) vaccination-like method to inform nodes, making them resistant to fake information; 2) competition between true and fake rumours. We show that in low clustering structures, the spread of both true and fake news is facilitated. In networks portraying hierarchical topologies, the flow of information is largely dominated by top nodes, which suggests the use of targeted intervention policies. Moreover, if social networks are shaped by sparsely connected communities, the low inter-community connections lead to a slower spread but still reaches similar final values as in low clustering structures. Finally, community structure leads to high levels of polarisation, as a result of the emergence of "belief bubbles". A similar effect may also be observed in hierarchical topologies with high clustering coefficient.

I. INTRODUCTION

Fake news or misinformation are pieces of false information that are (often maliciously) created or spread without fact-checking. They can also be seen as rumours that threaten public opinion in current matters by creating unnecessary discord [1, 2], potentially undermining the credibility of news markets [3].

One aspect that characterises fake news is how they spread in the network. People tend to share fake news more often, reaching farther, faster, deeper and more broadly in the network [4]. Similar findings also seem to suggest that humans tend to share fake news more than bots. Fake news also tend to have a lower heterogeneity than accurate news, due to having less dominant broadcasters [5]. Typically, misinformation has fewer initial sharers and grows over time, based on branch spreading processes, eventually reaching many individuals. Current literature mostly assumes the spread of misinformation as an analogue process to a virus outbreak. Contrary to trustworthy news, which tends to stay mainly in the vicinity of the main broadcasters, fake news curls its way around the network and its individuals much more deeply.

In this context, we may wonder how we can fight the growing threat of fake news in social media? Hartley and Vu [6] propose insights into intervention policies fighting fake news through an equilibrium model. They conclude that an intervention policy that shifts a digital citizen's behaviour to one of a "high effort" mindset can make the critical evaluation of fake news easier. Another intervention policy is to reduce the utility one gets from engaging with fake news. The former empowers the individual's awareness and detection of fake news, while the latter tries to minimise the dissemination of fake information on social networks and social media (also discussed by Lazer et al. [2]).

Here we aim to take advantage of the multidisciplinary nature of Network Science, with a focus on the spread of rumours in structured populations. To this end, we develop a novel modelling platform of spreading processes where social interactions are described as a network of connected nodes. We resort to different classes of scale-free networks to assess two types of interventions and rumour spreading dynamics and assess their impact on the spread of truthful and false rumours.

Our focus here will be on intervention types where we try to raise individual awareness and resistance to fake news. This can be done via training, promotion of fact-checking, etc. Let us consider that misinformation starts from a single individual. Then, several approaches can be considered to describe how to prepare the population against it. Here we consider two.

First, we shall consider external (top-down) interventions on the population to increase individuals' awareness on false information, creating resistance to it. For simplicity, and as an analogy to classical vaccination, aware individuals do not share their awareness to others around them. Secondly, we evaluate self organised intervention mechanisms, where we consider the direct competition between fake and truthful rumours. The fake news and the awareness of it each start in their own individual. These then spread as viruses to their neighbours and so on, competing in an arms-race. The spread of the awareness can be seen as a viral vaccination.

Both of these approaches can help us see how three different scale-free networks change how misinformation and awareness spread. For simplicity and as an analogy to epidemiology, the individual who is aware of the truth will be referred to as vaccinated, and awareness or awareness spreading referred to as vaccination (a process which

* goncalo.a.simoes@tecnico.ulisboa.pt

does not occur in the realm of disease outbreaks). Their structures, low clustering, highly clustered hierarchical, and community, give three different points of view for information spreading and how each network property affect it. Does having a hierarchical society make it more susceptible to attacks of misinformation, mainly targeted attacks? Can we combat it through more robust and faster targeted information procedures? Do communities prevent misinformation from spreading, or do they facilitate it? What is the polarisation of opinions emerging from these structures?

To answer these questions, we extend a rumour spreading model to study various situations regarding how misinformation and awareness start and spread.

II. PROPOSED APPROACH

A. Rumour Model

The model used is one with two additions to the Nekovee et al. extension of the Maki-Thompson model [7]. One is to account for vaccinated/educated individuals with a new state V. The other is a stifling interaction between vaccinated individuals and spreaders. In the resulting model, we get four states: ignorants (S), spreaders (I), stiflers (R) and vaccinated (V). S represent the neutral individual who doesn't know either the truthful or false rumour. States I and V are both "spreader" states, and try to convert S's to their side. They cannot convert members of the other side. R are individuals who believe or know of the fake rumour but do not spread it because they have been stifled. The vaccinated/spreader stifling interaction means V's will actively try to convince others around them to stop spreading false information. If they're convinced, they become stiflers. Without this stifling, V and S do not interact with each other.

We call this model SIRV, seen in Figure 1. The transitions between states and their parameters are as follows:

$\boldsymbol{\beta}$: S→I. When an I individual interacts with an S, the ignorant can become "infected" with probability β.

$\boldsymbol{\theta}$: S→V. When a V individual interacts with an S, the ignorant can become vaccinated with probability θ. The vaccinated state works like a virus. The idea is that this vaccination is the truth opposite to the misinformation spread by I individuals.

$\boldsymbol{\gamma}$: I→R. When an I individual interacts with another I or an R, the spreader can be stifled with probability γ.

$\boldsymbol{\delta}$: I→R. An I individual can forget the misinformation with probability δ, becoming an R.

$\boldsymbol{\zeta}$: I→R. When an I individual interacts with a V, it can become an R with probability ζ.

For $\zeta = 0$ we can study vaccination spreading without vaccination stifling. With $\zeta = \theta = 0$, we can study classical vaccination. On Table I we present the various model parameters for reference.

Parameter	**Symbol**	**Range**
Number of nodes	N	$\mathbb{N}$
Infection rate	β	[0,1]
Stifling rate	γ	[0,1]
Forgetting rate	δ	[0,1]
Vaccination rate	θ	[0,1]
Vaccination stifling rate	ζ	[0,1]

TABLE I. Model parameters and their value ranges.

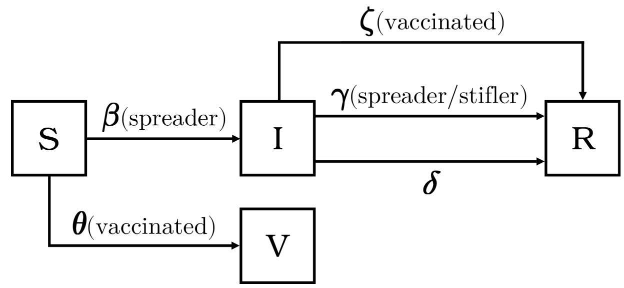

FIG. 1. Diagram of the SIRV extension model of the Nekovee et al model [7]. The parameters are as follows: infection rate β, stifling rate γ, forgetting rate δ, vaccination rate θ, and vaccination stifling rate ζ. The four states represent: ignorants (S), spreaders (I), stiflers (R) and vaccinated (V).

B. Networks

Three scale-free network types where chosen for study, each with different properties and topologies. Barabási–Albert (BA) [8] serves as a baseline and gives us networks with very low clustering coefficient. Dorogovtsev–Mendes–Samukhin (DMS) [9] on the other hand gives us a highly clustered hierarchical topology. Lancichinetti-Fortunato-Radicchi (LFR) [10] provides a multi community structure with a low clustering coefficient.

1. *Barabási–Albert (BA)*

BA are scale-free networks with very low clustering (see Table II). Since there are little to no triangles in the network, nodes will not be connected to the neighbours of their neighbours. This will lead to nodes being connected to other nodes outside of their immediate vicinity, which results in a lower Average Path length (APL). This is expected to facilitate both the spread of the misinformation and vaccination.

2. *Dorogovtsev–Mendes–Samukhin (DMS)*

DMS on other hand provides us with a highly clustered scale-free network. Its network algorithm starts

Metrics	BA	DMS	LFR
Degree Distribution Exponent	3.05	2.92	2.69
Average Degree	3.99	3.99	4.33
Clustering Coefficient	low (0.03)	high (0.74)	low (0.14)
Average Path Length	4.08	4.91	6.80

TABLE II. Metrics for the various networks used. BA and DMS used 1,000 randomly generated networks, while LFR used the 100 generated networks that are used throughout this paper. Degree distribution does not take into account the low and high cutoff points, which would bring the results closer to the expected value of 3.

with 3 fully connected nodes (a triangle), then randomly selects an edge and adds a node connected to the nodes of this edge (adds another triangle). It repeats these last 2 steps until we reach the number of desired nodes. This creates a highly clustered network which has an hierarchical structure. This topology has interesting properties for information propagation since it resembles a tree like structure. Given the starting triangle (the core nodes), every node that branches off an edge of this triangle will never share a direct edge or non-direct path (not passing through the core) between them. This forces a funnel of information to pass through at least one of the core nodes. If we block any core node, information can still flow freely because of the core triangle. If we block two core nodes, then all of the branches for that core edge will become disconnected between themselves and also to the rest of the network. If we block the three core nodes, every branch becomes disconnected. This makes the propagation of information inefficient. Regardless of where it starts, it will always be limited to that network branch.

3. Lancichinetti-Fortunato-Radicchi (LFR)

LFR, similarly to DMS, gives us a different topology from BA and allows us to study how communities can influence the spread of information, yet keeping a power law degree distribution.

The LFR algorithm does not necessarily create scale free networks, but if we supply the degree distribution exponent parameter with value 3, it will have the same expected degree distribution for BA and DMS. From this, the other parameters can be picked more freely according to what we want [11]. We chose values that produced networks with communities sufficiently large but not too connected to each other. This allows us to study what effects this community isolation produces, like the polarisation found in "echo chambers" (social groups that repeat the opinion they believe and ignore information that would challenge it). Values chosen besides the defaults of the networkx algorithm were: community size distribution exponent with 2; fraction of inter-community edges at 0.07 (intra-community edges $(1 - mu)\deg(u)$); average degree with 5; and minimum community size at 50 nodes. With these values, the algorithm has a propensity to generate networks that are not connected (communities or nodes not connected to the rest of the network). In order to combat this we generated a large number of networks and verified that they were indeed connected. We did this until we had 100 networks saved (which can be found here [12]) which were reused throughout the rest of this paper to obtain the results.

The APL (see Table II) on the other hand increases by almost 3 steps, compared to BA. In our LFR we don't have a core like DMS or an abundance of paths between nodes. Instead, we have multiple dense communities, connected by few paths, making the distance to travel between two nodes, on average, that much higher, and the nodes become much more disconnected to others in other communities.

With regards to hubs, the chance of a hub being the connection between communities is higher since having a higher degree means a higher chance to be the one with the inter-community edges, given that we use a very low value for this fraction.

C. Software

For obtaining results via simulations, an open source software for simulation of propagation models on networks with a focus on modularity was done. The software repository can be found here [13]. One important aspect of the simulations is that they were done via batching. Instead of simulating one event per time step, we simulate every event possible in that time step, so long as they don't interfere with a previous event's result. If an event causes a node to change from state A to B, then we won't simulate another event that might've changed it from A to C. All possible events are randomly shuffled to avoid biased behaviour.

III. RESULTS: EXTERNAL INTERVENTIONS

If we were to fight misinformation on a population that does not share truthful information and can only learn it, how could we approach this? By learning truthful information an individual would be considered vaccinated. In the classical sense of a vaccination protocol, individuals learn truthful information via a top-down approach. How can the structure influence the outcomes and our options to fight misinformation with a top down approach?

For the following results we used networks with 1,000 nodes, and for SIRV, we used stifling rate γ and forgetting rate δ of 0.1, and vaccination rate θ of 0 to simulate classical vaccination. The other parameters, infection rate β and vaccination stifling rate ζ, varied from 0 to 0.9, with a step of 0.1. For γ and δ, the value was not changed because they're not the focus of this study.

Misinformation always starts in a single node, which is chosen randomly, unless otherwise specified. For each combination of parameter values, 1,000 simulations were executed for statistical purposes, and the results were obtained as a mean over these simulations. For both BA and DMS, we generated 100 random networks at the start of every simulation batch (which usually contains enough simulations to produce a figure) and use those for that batch specifically. As said in Section II B 3, LFR uses the same 100 networks for all results, except when for comparison of results by network size was done.

The size of 1,000 nodes was chosen by doing a comparison between it and 10,000. This comparison, done for all networks and tested for various vaccination fractions, did not produce statistically significant differences, which led to choice of 1,000 nodes for simulation ease.

A. Vaccination Fraction

For fighting misinformation in a classical vaccination scenario, we have to employ strategies on what nodes to vaccinate and how many. We'll be focusing on Random, Acquaintance and Hub strategies. Random works as a baseline for comparison. By varying the infection rate β, we obtained results for Total Misinformed (TM), the number of stiflers (R) at the end of a simulation, regarding all 3 strategies and all 3 networks as seen in Figure 2. The data points have intervals of 0.1 for both axis, which means intervals with large changes may not be indicative of the actual tendencies, as seen in the Hub data. This will be addressed further on.

We can see that, as β increases, the gains from increasing vaccination become minimal past a certain cutoff point. This point depends on the network type and the strategy employed. Above 0.50 β it appears that increasing the vaccination fraction yields minimal gains, independent of the network or strategy.

Random Strategy. As expected, this strategy works rather poorly on the various networks, since it doesn't try to take advantage of the underlying structure like Acquaintance or Hub. Nevertheless it behaves better in DMS and LFR than in BA. In DMS, despite the cutoff value being similar to LFR, the overall TM has a slower increase as the vaccination fraction decreases, which as discussed below, seems to be due to its hierarchical structure.

Acquaintance Strategy. With this strategy, since it has a higher chance of vaccinating hubs, the cutoff points drop considerably for all network types. While the chance of vaccinating hubs is relatively the same across all of them, since they have similar degree distributions, DMS has a significant drop when compared to the others. This difference can again be attributed to the network's topology.

Hub Strategy. For this strategy, where vaccination blocks the main highway nodes for information, we get another considerable drop. All networks now generally require less than 10% of the population to be vaccinated to effectively curb the misinformation spread. Since it falls between the interval of data points, it becomes impossible to see the actual data tendency, but we can redo the results for this interval at smaller steps. What we see is a difference in orders of magnitude between DMS and BA\LFR. By looking at the middle cutoff point of the contour results (around where TM passes the 50% threshold at high β's), we see a cutoff point of 0.2 to 0.3% (2 to 3 nodes, respectively) for DMS, while LFR has one of 5% and BA has one of 8%.

B. Structure

How does the structure influence these results? With DMS, we know the extreme importance of the core nodes in propagating information, so any strategy that can vaccinate hubs with a higher chance can take advantage of this. The core nodes will most likely be the largest hubs, and it explains the striking difference in results between Random and Acquaintance\Hub strategies. As said in Section II B 2, by vaccinating 2 or 3 core nodes, we can cutoff a great number of branches from the rest of the network, or even every other branch, respectively. By doing this, the misinformation is contained inside those branches and is effectively stopped from spreading through the network. As discussed above in Hub Strategy, the middle cutoff point of 0.2% to 0.3%, 2 to 3 nodes respectively, match what we expect to cutoff the network considerably, since these are likely to be in the core. Random still behaves better compared to BA/LFR because random nodes are likely to be near the core, thus blocking branches/sub-branches.

In BA on the other hand, given the higher abundance of paths between nodes in the network, randomly vaccinating nodes is not very effective, since even if we do manage to block a hub node, the misinformation can still spread around that blockade. Thus, the vaccination fraction has to be higher to compensate for this in order to block more paths. Strategies that are likelier to vaccinate hubs will block information highways, leading to higher drops in TM.

With LFR we discussed before that hubs have a higher chance to be the connections between communities. Given its community topology, with few paths connecting communities, the higher the chance of vaccinating a hub the lower the TM, like with BA and DMS. Due to having less paths than BA, in general we will need to vaccinate less nodes, even with Random, since a high enough fraction will tend to block the inter-community paths that exist. When we vaccinate hubs, the probability of paths being blocked is higher while the middle cutoff point is still around 50 nodes. This means we vaccinate both hubs inside the community and ones with inter-community edges. The inside hubs are only good

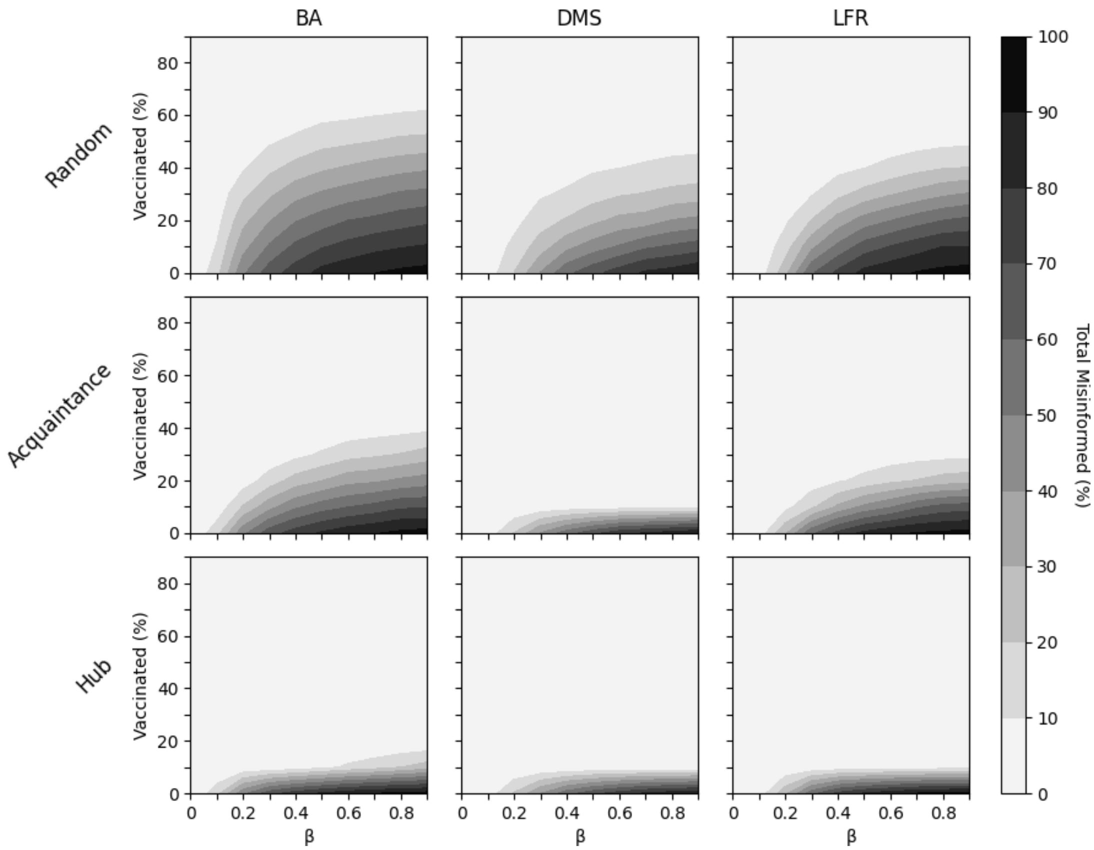

FIG. 2. Contour graph depicting how different strategies of vaccination (rows: Random, Acquaintance, and Hub) affect different networks (columns: BA, DMS, and LFR) considering vaccination fraction and infection rate β. The colours shown represent values for Total Misinformed, which is the percentage of stiflers at the end of a simulation. 1,000 repeats for each point of data were done. Vaccination fraction and β vary from 0 to 0.9 with a step of 0.1 giving us 100 data points.

for curbing the propagation inside their community once misinformation reaches it. Thus, we requires more hubs to be vaccinated.

C. Clustering Coefficient

It's difficult to separate clustering from structure. As clustering increases so does the number of nodes with similar states that interact with each other. The increase in clustering comes with a higher polarisation, which we will discuss in Section IV. In this case, since vaccination does not spread, misinformation tends to takeover more clustered areas, forming a polarised zone. This will, in turn, increase the stifling that happens, because spreaders will stifle each other, while also forgetting and starting to stifle those around them. More clustered zones or communities, which are dominated by the spreaders, will then begin to die out from the inside, due to stifling, with a "virus" like behaviour. This should theoretically force misinformation to disappear sooner.

D. Concluding Remarks

From these results, we can see why the structure is very important. Hierarchical structures have an inherent strong resistance against misinformation spread, while networks with low clustering show a very weak one. The difference between these "extremes" is the degree of order that they have. We assume the correlation is that: the more order that there is in the structure of a network, the more resistant it becomes to the spread of misinformation, and the easier it becomes to vaccinate against it.

IV. RESULTS: SELF-ORGANIZED INTERVENTIONS

What if, instead, individuals in the population tried to spread the truthful information in order to try to reduce the effect of misinformation? How does the structure affect this competition between two spreading informations? This is what we seek to study in this section with spreading vaccination.

For this section, for the results obtained, we kept the same size of networks and values for stifling rate γ and forgetting rate δ in SIRV, as in Section III. Both misinformation and vaccination start in a single node, since we're simulating the bottom up approach of vaccination. The networks used also remain the same, with BA and DMS having 100 networks generated randomly per simulation batch and using the same 100 LFR networks. A size comparison of results between 1,000 and 10,000 nodes was also done and found no statistically significant differences.

A. Timescales

In order to vary the speed of vaccination against misinformation, we can implement timescales through the use of random number generation within an interval. The interval we use is $[0, 1+w]$, with w being the timescale. If the number is in $[0, 1]$ then we do a batch of misinformation events (transitions). If, instead, it's in $]1, 1+w]$ then we do a batch of vaccination events. If w has value 1, each information spreads at the same speed (on average). With value 0.5, misinformation spreads two times faster. With 2, the reverse happens. With this, we can try to understand how individuals, less or more active on the side of vaccination, can affect how misinformation spreads.

Considering different timescales for the spread of vaccination through the use of the timescale factor, how does it affect misinformation spread? By varying timescale for values 0.5, 1, 2 and 4, and vaccination rate θ and infection rate β from 0 to 0.9, with a step of 0.1, we obtained the results which can be seen in Figure 3. Overall, independent of the network, by increasing timescale, we increase the effect that θ has. Thus, for lower values of θ we can reach the same TM values by increasing timescale. On lower timescales and lower β, increasing θ has little effect on TM. As β increases, so does the effect that θ has on lowering TM. Nonetheless, increasing timescale provides the largest decreases in TM, regardless of θ or β, since individuals will "work" faster to spread the truthful information.

Similar to what was seen in Section III, DMS has lower values overall compared to BA and LFR, but as timescale increases, the increased effect to vaccination is lower. At timescale 2, the three networks behave similarly (with the exception of very low θ), but at timescale 4, DMS behaves worse for higher β. This can be explained by the structure and will be discussed below. Another clear result is that BA and LFR behave very similar with regards to TM, but given their different topologies, the way the misinformation/vaccination spread on the networks must be fundamentally different. This will also be discussed in the following section.

1. Structure

Starting with DMS, it gives lower TM values overall when compared to BA and LFR for low timescales, since the hierarchical topology makes the spread of information more difficult when nodes can be blocked (vaccinated). Given the structure and that both informations start with 1 node, whichever information reaches and controls the core first will tend to control the rest of the network, since it will cutoff the opposite information from spreading outside of its branch or sub-branch. As timescale increases, what helped DMS to lower the spread of the misinformation, now hinders the spread of the vaccination, leading to the higher values (compared to BA and LFR) on timescale 4, and slightly for timescale 2. Despite spreading, on average, at 4 times the speed of the misinformation and most likely reaching the core first, the vaccination cannot stop the misinformation from spreading on the branch or sub-branches that misinformation has already blocked, since there are no paths around that blockage, like in BA and LFR. On higher β, the misinformation can spread more effectively despite being slower and gain more ground before vaccination can stop its advance.

While BA and LFR give similar results, the propagation of information is fundamentally different, because of their structures. In BA the information can spread more freely and faster while in LFR it is forced into the few inter-community paths in order to spread between communities. This is what we see if we look at the progression over time of nodes per state, specifically I and R, as seen in Figure 4. The peak of spreaders is reached much sooner in BA and, naturally, a higher value because of this. The propagation in LFR is much slower, because the spreading between communities takes longer, but also because spreading inside of a community as a random starting node (most likely a node with low degree due to the distribution) takes more time. Even despite this difference in propagation, the TM (number of stiflers at the end of the simulation) remains similar, albeit slightly lower for LFR. Most of this is maintained even with misinformation starting in a hub, which will be discussed further down. It is also worth mentioning that as the number of spreaders grows, so does the number of successful stifling and forgetting events. As more stiflers appear, stifling events will keep increasing and misinformation will die faster.

2. Other Simulations

Other simulations like misinformation or vaccination starting in the largest hub instead were also done. We found that BA and LFR still have similar results in these cases. Whatever information starts in a hub can quickly spread around the network in BA, but in LFR it still means starting in a community which means inter-community bottlenecks. Despite this, it helps in spread-

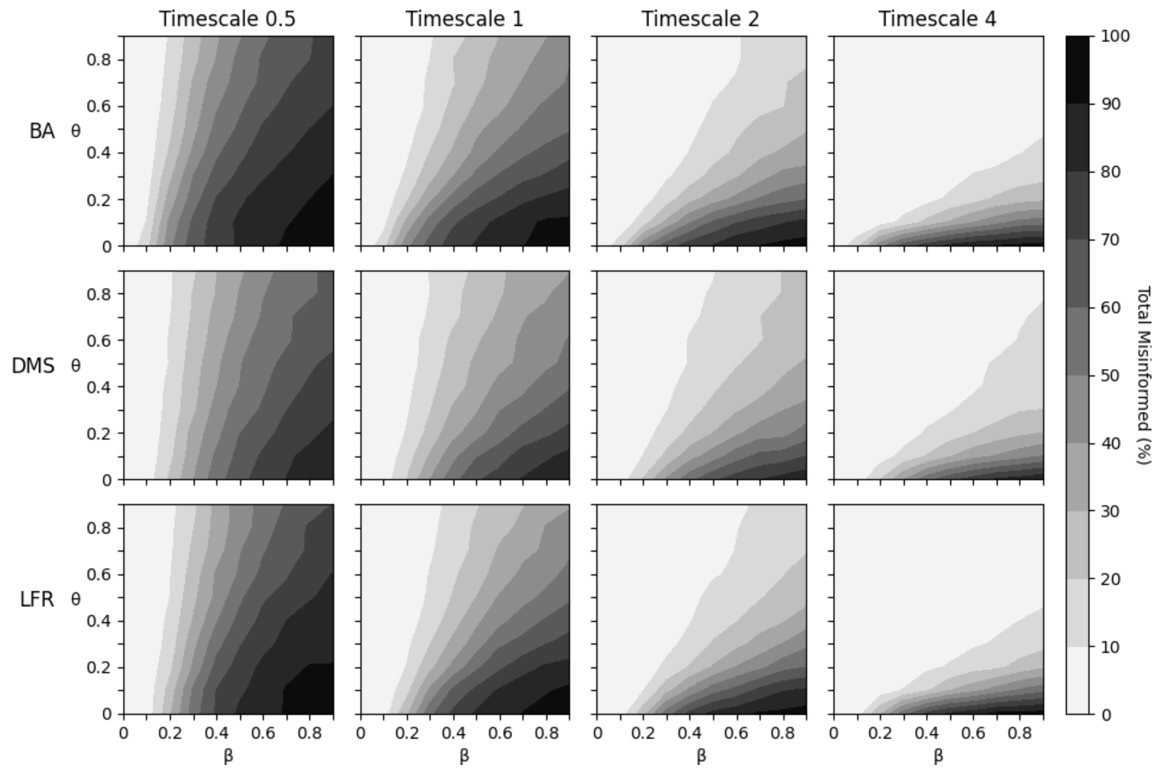

FIG. 3. Contour graphs depicting Total Misinformed values for various networks (rows: BA, DMS, and LFR) and various timescales (columns: 0.5, 1, 2, and 4) given infection rate β and vaccination rate θ, with both misinformation and vaccination starting in a random node. Both β and θ vary from 0 to 0.9 with step 0.1.

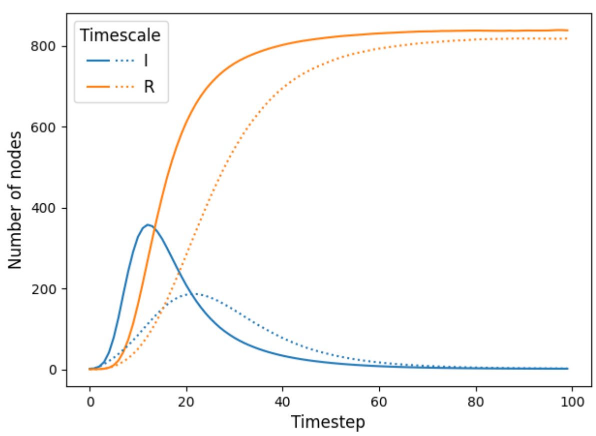

FIG. 4. Progression over time of the number of nodes per state (spreaders and stiflers) with random vaccination and misinformation, where both start with a single node. The solid lines represents BA states and the dotted lines represent LFR. The data was obtained with infection rate $\beta = 0.5$ and vaccination rate $\theta = 0.2$, averaged over 10,000 repeats.

ing quickly within the starting community, moving the peak of spreaders slightly earlier. On DMS, it dictates the supremacy of whichever information started in the hub, leading to either high resistance in diminishing TM or domination of vaccination. As expected, hub misinformation leads to large increases in Total Misinformed (TM) overall or large decreases in the case of hub vaccination. The effect of vaccination rate θ and timescales is significantly reduced in hub misinformation, but they are much more effective at reducing TM in BA and LFR, contrary to DMS due to its restrictive structure. In hub vaccination, with timescales of 4, TM is reduced to negligible levels ($\geq$ 10%) with at least $\theta = 0.1$, with the exception of some cases with very high β.

B. Polarisation

While the Total Misinformed can show us how effective increasing timescale or β is, it does not paint the full picture of what happens in the networks. One thing we can look at, to help our insight into how the spread of both informations characterise the networks, is polarisation. By this we mean to measure how polarised (or segregated) individuals and communities are if they have a tendency to be neighbours with individuals who share the same opinion (information in our case). The higher that tendency is, the more polarised a network is, the more segregated individuals are. In order to study this,

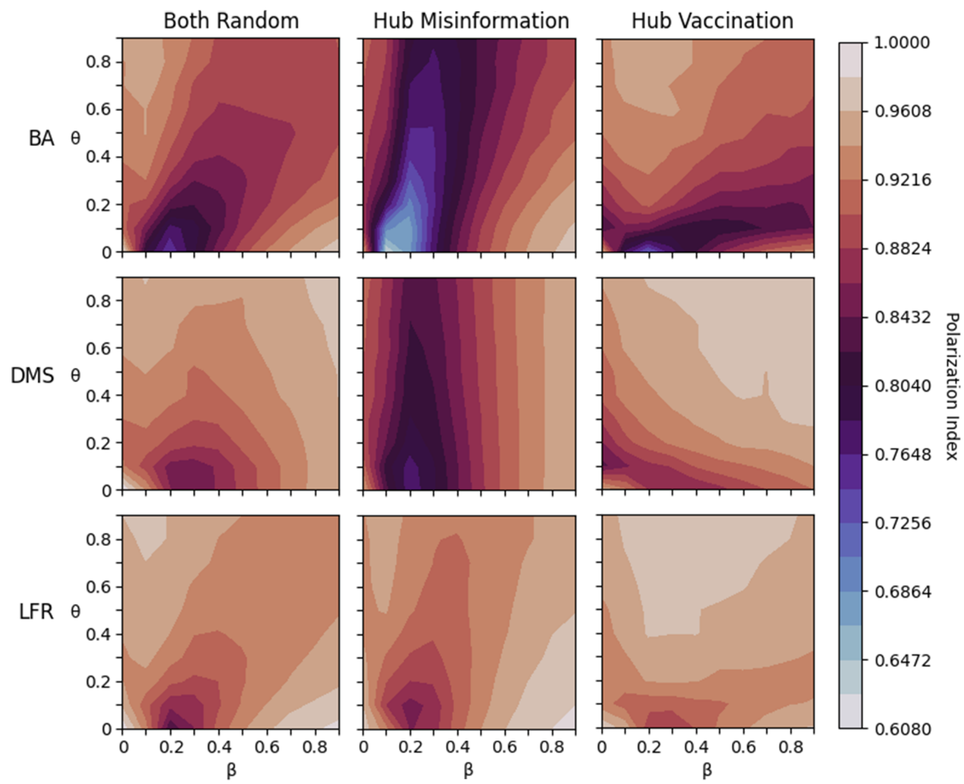

FIG. 5. Contour graphs for the polarisation index with various networks (rows: BA, LFR, and DMS) and information origin (columns: Misinformation and Vaccination start in a random node, Misinformation starts in the largest hub, and Vaccination starts in the largest hub). Timescale is fixed at 1, while both infection rate β and vaccination rate θ vary from 0 to 0.9 with step 0.1.

we devised a simple polarisation index P.

$$P = 1 - mean(\frac{\text{cross edges}}{\text{total edges}}, \text{per community}) \quad (1)$$

where cross edges are edges between nodes of different opinions (considering ignorants, stiflers, and vaccinated). This index is focused on communities, but since BA and DMS don't have communities per se, we consider them as a single community, where as in LFR we use the designated communities by the generation algorithm. What this index tells us is the polarisation average per community. The closer the value is to 1, the more homogeneous are the communities (or community).

With this index applied at the end of each simulation, we can rerun the simulation batches done for studying timescales, and obtain polarisation results for each networks, timescale, and origin of information. The condensed results, focusing on timescale 1, can be seen in Figure 5. We can see generally that BA has the lowest index values. Its structure causes the most cross interactions due to being almost clusterless, leading to a more heterogeneous community. LFR's structure helps to keep the three opinions mostly separated within their communities. Since the spread between communities takes longer, spreaders and vaccinated will more easily spread inside of a community than outside it. It is also possible to have entire communities of susceptibles if the paths becomes stifled or the simulation ends before vaccination can reach it. For DMS, the hierarchical structure forces both misinformation and vaccination to meet only at the edges of the areas that they control, leading to a lower number of cross edges between them. The susceptible that remain on the network will depend on the parameters. Given this, we can see a clear difference in the shape of change that BA and LFR have compared to DMS, which like the results for timescales, implies there is some relation between BA and our subset of LFR.

We can define their shapes roughly as canyons, where polarisation index starts high (more homogeneous) then lowers before rising high again. This canyon separation usually shows us the separation between two zones that represent a different information dominating the network. Using BA with both random starter nodes as example, the left margin of the canyon shows us an area where vaccinated (and susceptibles) dominate the network due to low β's. The right margin instead has misinformation dominating the network due to high β's and low θ's. The canyon then represents the zone where both informations compete the most, and cause more chaos in opinion relations. The depth of the canyon depends on the speed of spreading but also on how consistently each information spreads, since consistency will leave less gaps

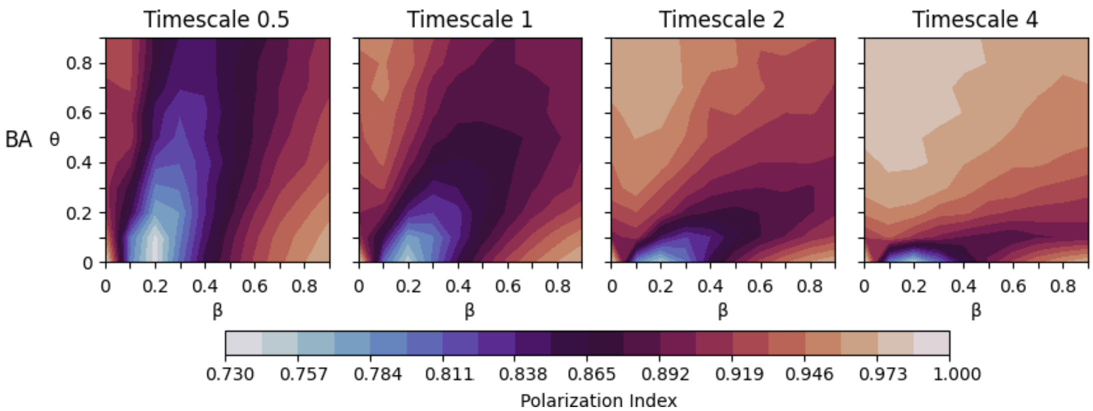

FIG. 6. Contour graphs for the polarisation index with BA networks and various timescales (columns: 0.5, 1, 2, 4). Both informations start in a random node. Both infection rate β and vaccination rate θ vary from 0 to 0.9 with step 0.1.

which increase the area of contact between opinions. This is more important in DMS, and LFR, due to clustering. The shape of the canyon depends on the structure since it's what mainly affects spreading.

Another factor for the shape of the canyon is the timescale. As we increase it or decrease it, the vaccination margin will gain or lose size, respectively. We can see how BA with random starter nodes changes as timescale varies in Figure 6. The left margin, which is mostly composed of vaccinated and susceptibles grows as timescale increases, and shrinks the canyon and right margin due to its added strength. The canyon shrinks due to less competition, and for the right margin it becomes increasingly more difficult for misinformation to dominate.

C. Concluding Remarks

Timescales give us the opportunity to study how two informations with similar or different speeds spread and affect each other.

We showed that BA and LFR (our subset of possible LFRs) have extremely similar results for Total Misinformed (TM), independent of where either information starts. Despite leading to similar TM at the end of our simulations, where LFR always has the lower value, LFR has a different spreading pattern due to its restrictive structure. This raises the implication that both networks share some implicit similarity in their structure despite their differences.

With DMS, we saw that hierarchical networks can generally resist the spread of misinformation but also the spread of vaccination, so it's a double edged sword. However, when either start at the top of the hierarchy, their spreading potential is amplified tremendously, regardless of other variables. In the case of misinformation, even with very high timescales, its extremely difficult to reduce misinformation spread significantly, due to its restrictive structure. Otherwise, increasing the timescale provides great reduction of misinformation, but past a certain point the structure forces diminishing returns.

When we look at polarisation across all three starting scenarios, the same order of overall polarisation is maintained. BA is the least polarised (more heterogeneity) and LFR is the most. Despite their similarity in TM on timescale results, LFR networks leads to much more polarised networks due to the community structure which helps to maintain the information "trapped" (information bubbles) inside the community and keep out new information. BA's chaotic structure enforces more heterogeneity since information can freely propagate and create a higher area of cross contact. DMS's hierarchical structure also causes information bubbles due to the easily blocked paths for propagation, but the polarisation is lower than LFR, because the structure leads to a higher area of cross contact. In LFR this area is minimised because of the clustered communities and low inter-community edges.

From these differences in polarisation, we can make a generalised assumption: increase in order (i.e. increase in average clustering coefficient) does not correlate with an increase in polarisation. A better metric for a simple correlation with polarisation might be Average Path Length. Remembering the metrics from Table II, BA had on average ~4 steps between every node, DMS had ~4.9 and LFR ~6.8. These match the order we see for polarisation increase, and also the amount of difference between them (except with both random starting positions where DMS and LFR are close, but LFR has more polarisation overall). As nodes become more distant, on average, from each other, the longer information will take to reach the various nodes in the network. By taking more time, information will spread more in the vicinity of where it started and less in the faraway nodes of the network. This will create clusters of the same information, leading to more polarisation. More distance also means more time to spread consistently where it already is (nearer to the start). In LFR with a high enough infection rate β or vaccination rate θ, the starting community will likely be almost fully infected before the information even reaches the furthest community.

There is no generalised conclusion on whether increasing θ or β, separately, leads to increase or decrease in polarisation, since it depends on the network, the value of the other variable and the timescale.

V. CONCLUSIONS

This research allowed us to understand better how structure can influence the spread of misinformation when the truth is introduced to fight it. Regardless of whether we are dealing with top-down (classical) or bottom-up (dynamic) vaccination, structure plays the most important role.

Our results showed that hierarchical networks can negatively impact both sides' spread by increasing the resistance to change when dealing with bottom-up interventions. On a top-down approach, the structure provides us with a sturdy effect in lowering misinformation. In this case, both Acquaintance and Hub strategies provide low values of misinformation with tiny vaccination fractions. However, on community and low clustering structures, we do not see this similarity between strategies. These structures give rise to similar results, albeit marginally lower with communities, but with a significant difference between Acquaintance and Hub. Thus, the Hub strategy is better if we can get global information on the network. Otherwise, we require a much more significant fraction of vaccinated to get similar results. On dynamic vaccination, both of these structures behave similarly, except for more fringe cases with vaccination stifling. That being so, community structures that are sparsely connected slow down the spread of misinformation but ultimately give us the same results as low clustering networks. This said, sparsely connected community structures lead to substantially more polarisation, meaning that individuals inside the same community mostly share the same opinion.

We also saw how these structures behave under targeted attacks (misinformation starting in the largest hub) and targeted defences (vaccination of the largest hub). While targeted attacks produce more misinformation in the networks, we can still reduce it if the structure allows it. Whereas low clustering and community structures allow us to increase the timescale and vaccination efficacy to reduce misinformation spread, since hierarchical structures hinder the spread of information, it is increasingly difficult to combat misinformation as the misinformation becomes more viral, even with the increase of timescale and vaccination efficacy. On targeted defences, though, these structures produce the lowest misinformation due to the same reasons, as misinformation cannot compete against the advantageous spread of truthful information.

From this research, several questions emerged that could be the focus of future works. Firstly, with our three networks, we can test various types of structures, but real social networks tend to be a mix of the qualities shared by all these. Therefore, real social networks could be used with our model to see how our predictions with synthetic networks compare to real ones. Finally, given the similar results of BA and LFR, we can study if this behaviour maintains as we change the generation parameters for LFR networks.

[1] H. Allcott and M. Gentzkow, Social media and fake news in the 2016 election, Journal of economic perspectives **31**, 211 (2017).

[2] D. M. Lazer, M. A. Baum, Y. Benkler, A. J. Berinsky, K. M. Greenhill, F. Menczer, M. J. Metzger, B. Nyhan, G. Pennycook, D. Rothschild, *et al.*, The science of fake news, Science **359**, 1094 (2018).

[3] B. Martens, L. Aguiar, E. Gómez-Herrera, and F. Mueller-Langer, The digital transformation of news media and the rise of disinformation and fake news, (2018).

[4] S. Vosoughi, D. Roy, and S. Aral, The spread of true and false news online, Science **359**, 1146 (2018).

[5] Z. Zhao, J. Zhao, Y. Sano, O. Levy, H. Takayasu, M. Takayasu, D. Li, J. Wu, and S. Havlin, Fake news propagates differently from real news even at early stages of spreading, EPJ Data Science **9**, 7 (2020).

[6] K. Hartley and M. K. Vu, Fighting fake news in the covid-19 era: policy insights from an equilibrium model, Policy Sciences **53**, 735 (2020).

[7] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, Theory of rumour spreading in complex social networks, Physica A: Statistical Mechanics and its Applications **374**, 457 (2007).

[8] A.-L. Barabási and R. Albert, Emergence of scaling in random networks, science **286**, 509 (1999).

[9] S. N. Dorogovtsev, J. F. Mendes, and A. N. Samukhin, Size-dependent degree distribution of a scale-free growing network, Physical Review E **63**, 062101 (2001).

[10] A. Lancichinetti, S. Fortunato, and F. Radicchi, Benchmark graphs for testing community detection algorithms, Physical review E **78**, 046110 (2008).

[11] Python library Networkx's LFR algorithm documentation can be found here.

[12] The 100 used LFR networks can be found here.

[13] Software repository found here.