JASSS
THE JOURNAL OF ARTIFICIAL SOCIETIES AND SOCIAL SIMULATION

Using Agent-Based Modeling to Analyze Intentional Diffusion of Ideas

Stefano Di Filippo[1]

[1]*Università Commerciale Luigi Bocconi*
Correspondence should be addressed to *stefano.difilippo@studbocconi.it*

Journal of Artificial Societies and Social Simulation xx(x) x, 20xx
Doi: 10.18564/jasss.xxxx Url: http://jasss.soc.surrey.ac.uk/xx/x/x.html

Received: dd-mmm-yyyy Accepted: dd-mmm-yyyy Published: dd-mmm-yyyy

Abstract: This paper presents a COCA-derived approach to study information diffusion in a network under the action of an external influence. Our main interest is in determining under what conditions an influencing actor, merely capable to induce public attention and not directly persuading, can obtain a shift in the population's distribution of opinions. We analyze the results under a wide range of starting conditions and derive implications both for prediction of social contagion and for normative statements directed to "influencing agents" or to those who wish to oppose them.

Keywords: Agent-Based Models, Information Diffusion, CODA Models

Introduction

1.1 The phenomenon of information and ideas spreading in a social network has been already extensively studied in the literature, often by means of adapting epidemiological models (Castiello et al. (2023), Bettencourt et al. (2006)). Though many of these studies were motivated by offering insight into the dynamics of public opinion and disinformation (Ndii et al. (2018), Castiello et al. (2023)), to our knowledge there seems to be little research considering the presence and motivations of an agent influencing the idea spread. In this work, we try to answer the questions of how and based on what factors information provided by an external agent shapes public opinion and what are the possible approaches the agent could take to optimize its behaviour. The agent in question is just seen, in the context of our study, as interested in having some idea or information be accepted by the greater possible fraction of the population. This implies that we are mainly interested in ideas which are mostly opposed by the society under investigation, but not necessarily. The agent may be easily identified with a dictator, media mogul, or terrorist group, but can actually represent any actor with some capability of influencing public discussion.

1.2 These questions are explored by making use of simple agent-based models with a network topology, and the results of simulations are analyzed and discussed, also in terms of induced policy implications. We believe that the line of research of the present study may shed significant light on both positive and normative questions, by complementing the tools already used to investigate issues in the area of information diffusion and management. Our hope is to provide a framework with the ability to model social contagion under external influences, inform policy decisions, and potentially predict information pollution due to destabilizing influences.

Background

2.1 Modeling news and rumors contagion has been frequently done through adaptation of models first developed to analyze the spread of diseases, due to similarity of these two phenomena. Examples of such models are the ISR (ignorants-spreaders-recovered) model and the IESZ (ignorants-exposed-spreaders-skeptics) model, directly derived from the SIR family (Castiello et al. (2023), Bettencourt et al. (2006)). Approaches further divide on whether they make use of a network topology or not (Ndii et al. (2018)); since we consider the structure of the social relationships to be of interest and importance in our context, our model is gonna be of the network type. Alternative ways to study information diffusion make use of game-theory based models, which have a focus on modeling the individuals' decisions, payoffs and actions to analyze behaviour strategies in equilibrium (Jiang et al. (2014), Razaque et al. (2019)).

2.2 Among the most up-to-date information diffusion models found in the literature, the CODA (Continuous Opinions and Discrete Actions) models represent agents' opinions as continuous variables, which individuals pass through a threshold to choose their action from a discrete set of possibilities (Razaque et al. (2019), Chowdhury et al. (2016)); in turn it's the actions, and not the opinions themselves, which then influence and update the individuals' opinion through their contacts. The CODA model provides great flexibility, and numerous extensions and restrictions have been analysed or applied to specific purposes, for instance with the additional modeling of trust effects or with the restriction to multi-level discrete opinions (Martins (2013), Varma & Morărescu (2017)). In the particular case in which agents have access to their neighbors' opinions, we talk about COCA models (Chowdhury et al. (2016)). Since the models developed in this paper keep the signal for updating at the level of opinions, they could be seen as part of the COCA family. However, the model here presented is found to be different from the CODA specifications mostly found in literature, both for its updating rule and for the context in which it is investigated, namely under the presence of an influence external to the network.

Models

3.1 We consider a network of n nodes denoting individuals of a population, and m edges denoting close relationships between the individuals. Outside the population, an actor, which will be referred to as "influencer", is interested in having an idea I be accepted by as many agents as possible, and has some means to influence the spread of such idea in the network. Each agent i is characterized by an "opinion" attribute p_i, representing the personal position with respect to the idea. $p_i \in [0,1]$, where 0 represents "full disagreement", 0.5 "neutrality", and 1 "full agreement". We model the starting distribution of p_i across the nodes with a Beta distribution: $p_i(0) \sim \text{Beta}(\alpha, \beta)$, where we choose $\alpha < \beta$ for ideas which are in contrast with the population's culture overall. The parameters of the distributions are set as $\alpha = s \cdot \mu$, $\beta = s \cdot (1-\mu)$, where $\mu = E[p_i(0)] \in [0,1]$, $s \in (0, +\infty)$; thus, $1/\mu$ is a measure of the disagreement the idea encounters at the start, and so how unorthodox and clashing I is for the population. Noticing the relation of s with the variance

$$V(p_i(0)) = \frac{\mu(1-\mu)}{s+1}$$

and with the absolute skewness of the distribution

$$|\mathsf{sk}(p_i(0))| = \frac{|2-4\mu|\sqrt{s+1}}{\sqrt{\mu(1-\mu)}(s+2)}$$

implying that $\frac{\mathsf{d}|\mathsf{sk}(p_i(0))|}{\mathsf{d}s} < 0$, we take $1/s$ as a measure of the polarizing nature of I. By any means, in our context we can then represent the idea as a vector of these two parameters: $I = (1/\mu, 1/s)$.

3.2 After instantiating the network, generated either from the Albert-Barabási model or as an Erdős–Rényi random graph, we assign the opinion attribute to the nodes by sampling from our chosen distribution. Once the p_i have been assigned, we can compute for each node the value of an "exposure" attribute, defined as the average opinion of neighbors, weighted by their opinion similarity: $e_i = \frac{1}{k_i - \sum_j (|p_j - p_i|)} \sum_{j \in N(i)} (1 - |p_j - p_i|) p_j$.

3.3 The influencer controls a model variable denoted as $M(t) = [\, M_1(t) \quad \dots \quad M_n(t)]$ representing the degree of public discussion/media coverage on the idea; we thus assume that the external actor is able to influence public awareness and attention on I, but not to directly convince agents, whose opinions only directly change due to their interactions. M goes from 0 (no awareness or discussion), to 1 (bombarding coverage or propaganda). Initially, $M(0) = 0$, and the network is inactive: the opinions attributes represent private or even unconscious belief about I, but have no way to induce any interaction until people start discussing the topic. Once the influencer increases M, the network becomes active. We also incorporate in the model a mechanism by which, when the network is active, each node has a probability of removing the link with the neighbor with the most distant opinion, provided that one of them disagrees and one of them agrees, that is, $\text{sign}(0.5 - p_j) \neq \text{sign}(0.5 - p_i)$.

3.4 At each time step t, the following happens:

1. $M(t)$ is set, through some strategy.

2. For each node i, if $M_i(t) > 0$, get its neighbor j such that $p_j = \arg\max_{p_j} |p_j - p_i|$. If $\text{sign}(0.5 - p_j) \neq \text{sign}(0.5 - p_i)$, then remove the i, j link with probability $p_{\text{rem}} = M_i(t)|p_j - p_i|$.
3. The exposure attributes for each node $e_i(t)$ are computed on the basis of $p_j(t-1)$.
4. The opinions attributes are updated according to:

$$p_i(t) = (1 - M_i(t))p_i(t-1) + M_i(t)e_i(t) \tag{1}$$

3.5 We implement two different versions of the model based on the type of strategy implemented by the influencer. In our first implementation, we consider a completely uniform influence: $M(t) = \kappa$ for all time steps t, and for all nodes i. We run the model with different values of κ also to account for the greater or smaller ability of the influencer to steer public discussion. In our second implementation, we consider a strategy of the type $M_i(t) = \kappa \cdot 1_q(i,t)$, where κ is again the intensity of the influence, while the indicator function expresses either the following condition:

$$1_q(i,t) = \begin{cases} 1 & \text{if } \; p_i(t) \geq F_p^{-1}(q) \\ 0 & \text{if } \; p_i(t) < F_p^{-1}(q) \end{cases} \tag{2}$$

or the opposite one:

$$1_q(i,t) = \begin{cases} 1 & \text{if } \; p_i(t) \leq F_p^{-1}(q) \\ 0 & \text{if } \; p_i(t) > F_p^{-1}(q) \end{cases} \tag{3}$$

where $F_p^{-1}(q)$ is the quantile function of the distribution of $p(t)$ and q a model parameter. This alternative approach allows the influencer to generate public discussion only in a subset of the population that satisfies a certain condition. This condition is that the individual opinion is higher (or lower) than the q-quantile of the opinion distribution at time t. Thus, this amounts to assuming that the influencer can target the individuals with opinion above (or below) a certain quantile q, and can continue doing this as the opinion at quantile q, $F_p^{-1}(q)$, changes over time.

Simulation and Results

4.1 In analyzing the behaviour of the model, we have chosen sets of parameter values to cover the entire spectrum of possible realistic conditions, and performed a model simulation for each combination. In particular, we considered:

- $\mu \in \{0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9\}$, coding for extremely rejected ideas up to widespread accepted ones;
- $s \in \{0.01, 0.1, 0.5, 1, 5, 10, 100\}$, going from high variance and polarization up to low variation around the mean opinion. This set of values considers a range of different Beta distribution density regimes: U-shaped (high polarization, lower value of s), sloping (medium polarization, medium values of s), skewed and symmetrical bell-shaped (low polarization, high values of s);
- $(i,t) \to M_i(t) \in \{0.2, 0.5, 0.8\}$, where we are now using the constant and uniform strategy type, ranging from low to high intensity;

4.2 We applied the model with each parameter combination to two different networks: a more realistic scale-free network generated from the Albert-Barabási model and a Poisson random network. Both graphs have been generated to have 1000 nodes and an average degree of around 50[1]. The simulations have been run for 50 steps each. In Fig. 1, we observe the change produced in the average opinion by the end of the simulation, for the scale-free network. About μ, we notice three main effects:

- The net change tends to be positive for $\mu > 0.5$, and tends to be negative for $\mu < 0.5$.
- The net change can be either positive or negative for $\mu \approx 0.5$.
- The greater the value of $|0.5 - \mu|$, the greater the absolute net change in average opinion.

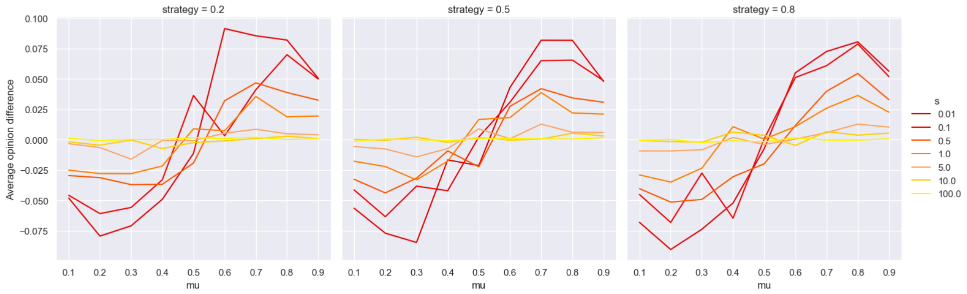

Figure 1: Change in average opinion across model parameters for the AB network

4.3 As the value of s increases, i.e. as we have less polarization, the less change in average opinion we observe. The values of s corresponding to bell-shaped beta distributions (100, 10) lead to an almost zero change at all values of μ; the values corresponding to slope-shaped beta distributions (5, 1, 0.5) exhibit changes in average opinion; finally, values of s corresponding to U-shaped beta distributions (0.1, 0.01), coding for a highly polarizing idea, exhibit differences that are more relevant and more sensible to changes in μ.

4.4 We also notice that the strategy choice, here just the intensity of the influence, does not seem to affect the results[2]. We deduce that, in the context modelled by our system, it would be optimal for the influencer to choose smaller values of intensity of diffusion, just enough to activate the network[3].

4.5 In Fig. 2 we plot the change in the standard deviation of opinions by the end of the simulation, across μ and s. We notice that the variance of opinions is decreased in each model specification, but that it becomes more relevant the higher is the polarization $1/s$. Also μ has an effect, with more extreme values leading to bigger changes in variance, thereby forming a U-shape whose convexity is directly proportional to $1/s$.

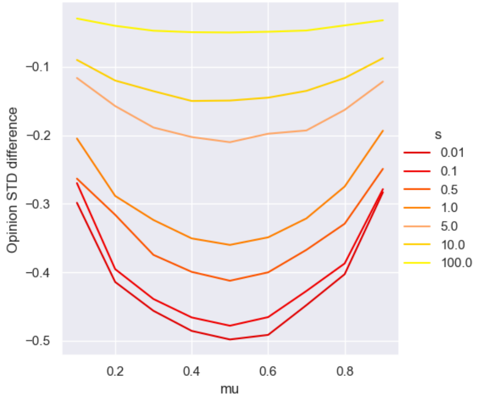

Figure 2: Change in standard deviation of opinions across model parameters for the AB network

4.6 By the previous inspections, we can state that uniform and constant influence will only be able to make an already accepted idea even more accepted (i.e. $\mu > 0.5 \rightarrow \bar{p}'(t) > 0)$, and an already rejected one even more rejected by the population (i.e. $\mu < 0.5 \rightarrow \bar{p}'(t) < 0)$. Through the same process, the influencer would also reduce the variability of opinions whatever the starting conditions. Indeed, this can be seen in Fig. 3, where each panel shows the density of opinions at the start and at the end of the simulation, for the combination of parameters given by $\mu \in \{0.2, 0.8\}, s \in \{0.01, 10\}$. We see that the beta distributions converge to point masses, but not at the initial mean: Beta$(s\mu, s(1-\mu)) \xrightarrow{t} \delta(E(p(\infty)))$, where $E(p(\infty)) - E(p(0))$ is a function of μ and s. This means that the influencer has the potential to change the average opinion of the population just resorting to stimulating public discussion, but we have also shown that she cannot regularly induce "corrective effects", that is produce a positive net change for a mostly rejected opinion, and vice versa. With our current model, we only have glimpses of corrective effects when agreement to an idea is approximately evenly split

[1] Also other (n, k) combinations have been tested, leading to similar results.

[2] Thus, unless specified otherwise, lack of the "strategy" variable in a plot implies it has been set to 0.5.

[3] Note that the number of steps for reaching stationarity has also been compared, and it shows little difference across the range of strategies.

in the population. In general, our analysis detects that the model works by converging fast to a point mass distribution of opinions centered at a value in $[\mu - 0.1, \mu + 0.1]$, with the average opinion reaching equilibrium in a seemingly exponential fashion. Though reaching up to a 10% absolute increase or decrease in agreement is not negligible, especially with modest influencing tools, this potential quickly vanishes, as the new constant opinion distribution does not allow for further exploitation. This is consistent with the evidence that, for lower polarization, the effects the influencer can obtain become less and less significant, which implies that, in the context under study, only polarizing ideas can be subject of strategic diffusion. It is thus evident how the system exploits variation in opinions to slightly increase or decrease average agreement, though the mechanism is still not entirely clear.

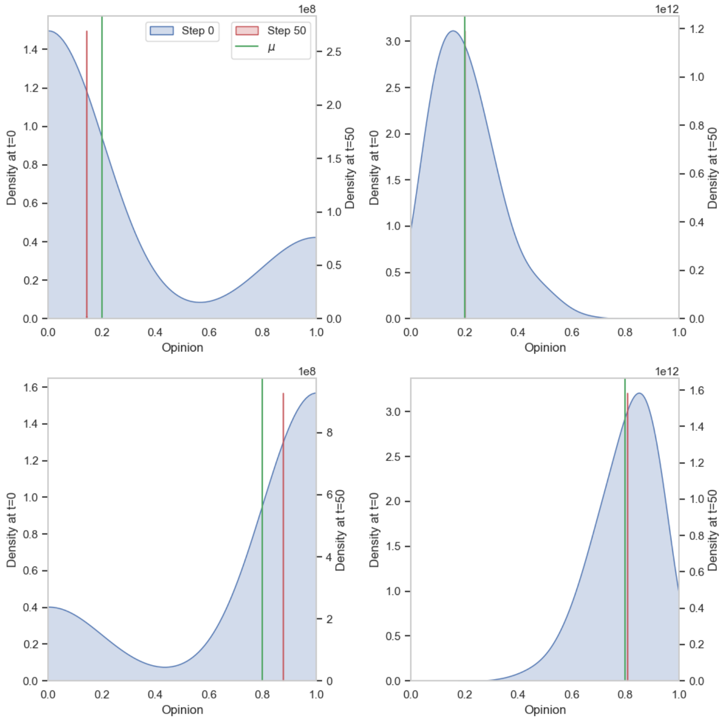

Figure 3: Densities for 4 model specifications

4.7 The effect of network topology is shown in Fig. 4 and Fig. 5, where the previous plots for the AB model are plotted together with results from simulation of Poisson random graphs with 1000 nodes and 50 average degree. We observe similar results for each parameter combination. The only relevant difference is that for the Poisson graph the change in average opinion goes uniformly to 0 for all s as μ approaches 0.5, without the variability shown for the scale-free network.

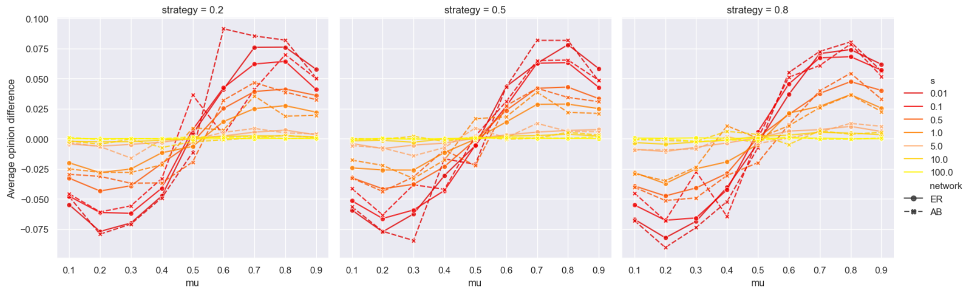

Figure 4: Change in average opinion

4.8 On the bases of the highlighted weak spots of our model, we implemented a second version with the two aims of allowing for corrective effects and for larger and sustainable changes in average opinion. The influencer is now assumed to be able to involve only a fraction of the population in public discussion, meaning $M_i(t)$ is now multiplied by an indicator function among eq. 2 and eq. 3, parameterized by a quantile q. We run the system again on a scale-free network and on a Poisson random graph, both generated to have 400 nodes and an average

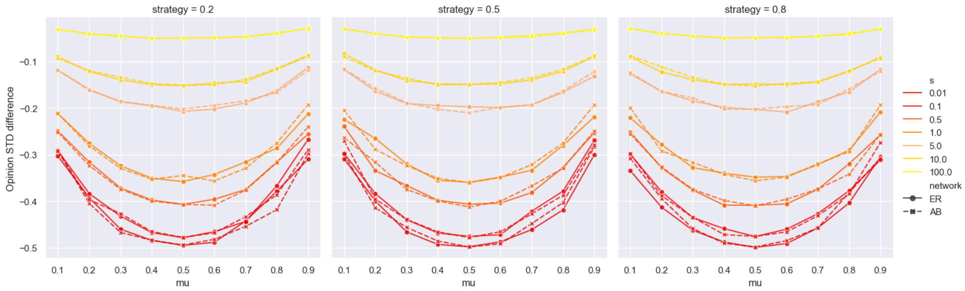

Figure 5: Change in standard deviation of opinions

degree of around 20. The model is simulated for each set of parameters combining the same previous choices for μ and s, with $q \in \{0.3, 0.5, 0.7\}$, and a boolean parameter *contrast*. Each specification is run for 100 steps. Depending on the choice of μ, *contrast* determines the form of the influence function $M_i(t)$ as follows:

- If (contrast $= 1$ and $\mu < 0.5$) or (contrast $= 0$ and $\mu >= 0.5$), then we select eq. 2.
- If (contrast $= 1$ and $\mu >= 0.5$) or (contrast $= 0$ and $\mu < 0.5$), then we select eq. 3.

4.9 In other terms, this parameter codes for the following:

- contrast $= 1$ → influence is restricted to the individuals with opinion no more negative (resp. positive) then a relative certain threshold, if the opinion is generally considered negative (resp. positive). Thus discussion and influence will mainly involve those that tend to disagree more with the general opinion.
- contrast $= 0$ → influence is restricted to the individuals with opinion no more positive (resp. negative) then a relative certain threshold, if the opinion is generally considered negative (resp. positive). Thus discussion and influence will mainly involve those that tend to agree more with the general opinion.

4.10 κ is kept constant at 0.5. Fig. 6 plots the change in average opinion by the end of the simulation, for the scale-free network. In the second row, we show the simulations with *contrast* = 1; we see that they exhibit reinforcing effects like for the simple model, and again polarization is directly related to the size of the difference. However, the shape of the opinion change as a function of μ is different from the one seen in Fig. 1: there, there was a smooth increase of the average opinion difference, going from its lowest for μ near 0, approaching zero as μ approaches 0.5, and then going to its highest for μ near 1. In the second row of Fig. 6 we see that the absolute value of the difference is greatest exactly as μ is far from the extremes and near 0.5, but it falls fast to 0 when reaching the middle point. This may flag the presence of a threshold or nonlinear phenomena for this kind of model. Still focusing on the second series of plots, we also notice that the absolute opinion difference is much more sizable at basically all level of μ and s with respect to the previous model.

4.11 Looking at the first row of plots in Fig. 6, we investigate what happens when *contrast* is set to 0. The analysis shows that indeed this setting, meaning selection of eq. 3 for $\mu < 0.5$ and eq. 2 for $\mu >= 0.5$, leads to corrective effects: we observe relevant changes in average opinion in the direction opposite to the sign of general agreement. Furthermore, for all plots we observe little change due to the choice of quantile, which prompts us to derive the following policy implication:

- If the influencer aims to generate (strong) reinforcing effects (*contrast* = 1), then it is equivalent whether she opts for: (a) exclusion or limitation in joining the public discussion of those individuals that have very strong opinions in the opposite direction w.r.t. the influencer's objective; (b) targeting public discussion only to those individuals that have very strong opinions in the same direction w.r.t. the influencer's objective; (c) any point in between.
- If the influencer aims to generate corrective effects (*contrast* = 0), then it is equivalent whether she opts for: (a) exclusion or limitation in joining the public discussion of those individuals that have very strong opinions in the same direction w.r.t. the influencer's objective; (b) targeting public discussion only to those individuals that have very strong opinions in the opposite direction w.r.t. the influencer's objective; (c) any point in between.

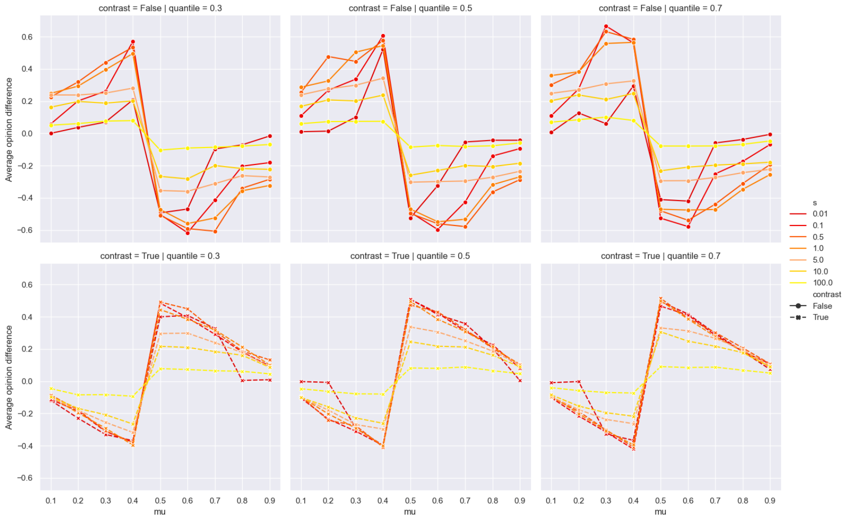

Figure 6: Change in average opinion

4.12 In contrast with the previous model, the obtained effects are also sustainable in time, although this is likely a consequence of the less weak assumption of the influencer being able to continuously track and target a certain opinion-based population fraction.

4.13 In Fig. 7, each panel shows the density of opinions at the start and at the end of the simulation, for the combination of parameters given by $\mu \in \{0.2, 0.8\}, s \in \{0.01, 10\}$. We set *contrast* = 0 to each model, $q = 0.7$ if $\mu = 0.2$ and $q = 0.3$ if $\mu = 0.8$. We observe the presence of corrective effects in all plots, with the densities moving in the opposite direction of $\mu - 0.5$. Also, we see that after 100 steps none of the beta distributions converged to just around a single point. In particular, for higher values of polarization, after 100 steps we still observe a similar variance in opinions compared to the starting distribution.

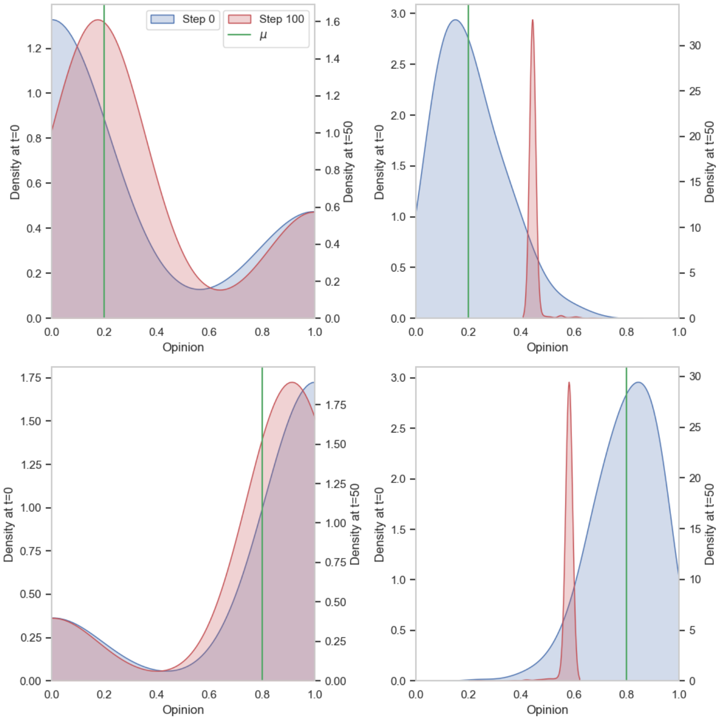

Figure 7: Densities for 4 model specifications

4.14 We conclude the presentation of results by showing how network topology affects this new system, in Fig. 8 and Fig. 9. No change in the behaviour of the model based on whether the graph has a power law or binomial distribution can be detected.

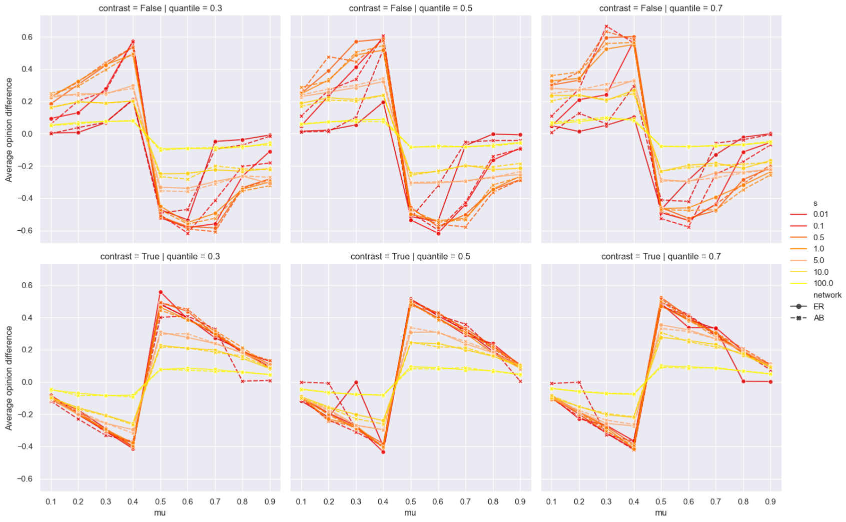

Figure 8: Change in average opinion

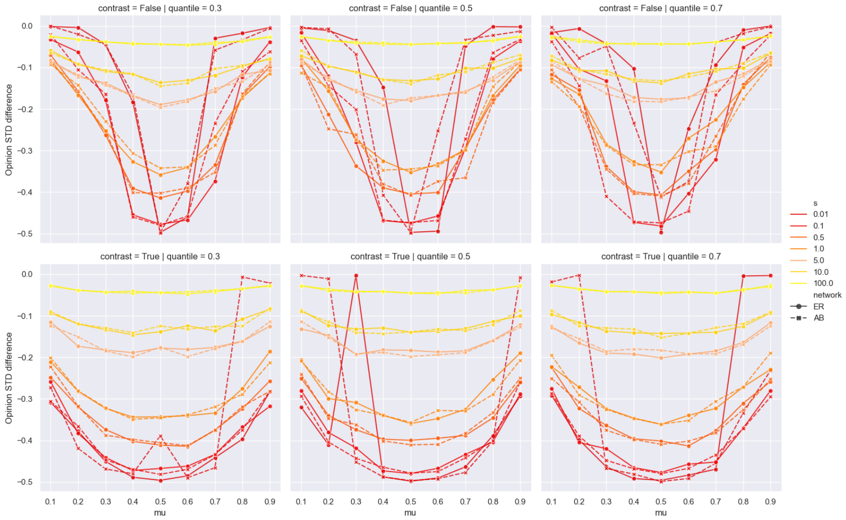

Figure 9: Change in standard deviation of opinions

Conclusions

5.1 In this paper we have first shown that in a context where the influencer is only capable to stimulate public discussion, and can do so just through simple constant strategies, it can still produce changes in the average opinion of the population. However, this potential is quickly exploited, and the changes are moderate and always reinforcing the status quo. We have also observed that only polarizing ideas, with sufficient variability of opinions across individuals, can be chosen by the influencer for such information diffusion process. There is also evidence that the influencer has no benefit in choosing a higher strategy intensity nor in trying to reduce the scale-free nature of the network, for example by targeting hubs, as random networks show similar behaviours. From this simple model, we could claim that someone who wishes to decrease the presence or acceptance of a fringe idea should opt for increasing public discussion about it, and the same is true for increasing the spread of a mainstream position. Indeed, it also seems that this is the case in many autocratic regimes where the public discussion on certain topics of propaganda can be incessant, although the possibility to more directly persuade and spread misinformation is also a factor.

5.2 In the second part of our analysis, we assumed that the influencer can stimulate public awareness/discussion mainly in just a subset of individuals, based on their position in the distribution of opinions of the population. We have shown that this type of strategy can lead to opinion changes that are much bigger and more sustained in time, though this is also due to other little-investigated assumptions. The main result, however, is that through this strategy the influencer can obtain both reinforcing and corrective effects. In general, correcting effects are obtained by targeting a fraction of population which is somewhat skewed towards agreeing with the general opinion. The likely reason behind this phenomenon is that, by inducing stronger opinion updates in these nodes, the "more conforming" opinions are diluted, and the distribution moves from agreement to disagreement, or vice versa. The exact mechanism is however still not clear, and it shows evidence of nonlinear effects still to be investigated.

5.3 We believe these models still hold potential for providing further interesting results, especially as the mechanisms behind the change in opinions both in the first and in the second model are not clear. Further research should also focus on analyzing rigorously the evolution of the opinion distribution through time, and, in this respect, the model seems quite appropriate to be cast into a Bayesian procedure. Finally, more sophisticated and general strategies by the influencer could be studied, so that the optimal actions on the basis of the main model parameters may be derived.

References

Bettencourt, L. M., Cintrón-Arias, A., Kaiser, D. I. & Castillo-Chávez, C. (2006). The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and its Applications*, *364*, 513–536. doi:https://doi.org/10.1016/j.physa.2005.08.083

Castiello, M., Conte, D. & Iscaro, S. (2023). Using epidemiological models to predict the spread of information on twitter. *Algorithms*

Chowdhury, N. R., Morarescu, I.-C., Martin, S. & Srikant, S. (2016). Continuous opinions and discrete actions in social networks: a multi-agent system approach. *https://arxiv.org/pdf/1602.02098*

Jiang, C., Chen, Y. & Liu, K. (2014). Graphical evolutionary game for information diffusion over social networks. *Selected Topics in Signal Processing, IEEE Journal of*, *8*, 524–536. doi:10.1109/JSTSP.2014.2313024

Martins, A. C. (2013). Trust in the coda model: Opinion dynamics and the reliability of other agents. *Physics Letters A*, *377*(37), 2333–2339. doi:https://doi.org/10.1016/j.physleta.2013.07.007

Ndii, M. Z., Carnia, E. & Supriatna, A. K. (2018). *Mathematical Models for the Spread of Rumors: A Review*. CRC Press

Razaque, A., Rizvi, S., khan, M. J., Almiani, M. & Rahayfeh, A. A. (2019). State-of-art review of information diffusion models and their impact on social network vulnerabilities. *Journal of King Saud University – Computer and Information Sciences*

Varma, V. S. & Morărescu, I.-C. (2017). Modeling stochastic dynamics of agents with multi-leveled opinions and binary actions. In *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*, (pp. 1064–1069). doi: 10.1109/CDC.2017.8263798