Università degli Studi di Milano

Dipartimento di Informatica

Course “Simulation” - Exam Project, A.Y. 2022/2023

Student: Luca Annese

Project Title: A novel approach to the rumor spreading model.

1. Introduction

A rumor is defined as “an unofficial interesting story or piece of news that might be true or invented, and quickly spreads from person to person”[1], and thus both mathematical and sociological view, as well as economics and other various social sciences, have been studying the behavior of this phenomenon. A common analogy is between epidemics spreading and rumor propagation: in both scenarios we consider infected individuals (people who have already been infected by the rumor) whose function is to spread their condition to other who have not been contacted yet by the disease. In this kind of setting, it is easy to develop and consider new features to embody in it such as: introduce the population of people who have been cured via a vaccine, people who avoided the infection, what kind of infection it is and many more, which can be easily translated into a rumor spreading framework. In my project I decided to study two different simulation models: a system dynamics approach which is more common for the nature of the problem whose aspect will be highlighted after and an agent-based model in which I consider a new view on the problem, based on the common assumption that “the like seeks the like”, from the latin phrase “*similia similibus curentur*”. In this approach I consider that each individual tends to believe only those similar to him, with respect to their educational level, which can be seen as a summa of one’s beliefs, cultural level as well as academic achievements. Moreover, despite the classical SIR (Susceptible Infected Recovered) model on rumor spreading, I consider another possible actor: the Counters, whose job is to restrain the spreading of the starting rumor, by spreading a counter-rumor in turn.

The focus of the project is to study and analyze the duration of a rumor as a function of the average educational level.

2. Model

The project consists of two different approaches: the first one is the classic view on the topic of rumor diffusion via a system dynamics approach modeled like the SIR (Susceptible-Infected-Recovered) model defined by Kermack and McKendrik, while the other one is a take on the agent-based model, having in mind the same structure of the first one, yet exploiting the possibility of managing each individual singularly.

2.1 System dynamics

The rumor spreading model is based on the infection diffusion model SIR, in which an agent can assume one of three possible states: susceptible are those who have not been affected by the infection

[1] https://dictionary.cambridge.org/dictionary/english-italian/rumour?q=rumor

but can be, infectious are those who have been infected and can infect others and recovered those who have been infected but are no longer infectious. The parallelism with a rumor spreading model is clear, and thus, these classes become:

- Ignorants: individuals who have never heard the rumor.
- Spreaders: individuals who have heard the rumor and are actively spreading it
- Stiflers: individuals who have heard the rumor and don't spread it or individuals who are not interested in the rumor.

Finally, my model considers two other classes: the Counters and the Heard. The latter consists of the individuals who have heard the rumor, while the utter class consists of those individuals who have heard the rumor and actively spread a counter-rumor.

The rumor initializes in only a small fraction of the population and starts spreading among the Ignorants via the Spreaders, who actively contact some individuals. After an Ignorant hears the rumor, it joins the Heard state, where it will decide whether to become: a Stifler, if the rumor is not interesting enough, the Spreaders if the rumor is interesting enough and its educational level not high enough, or the Counters. The Spreaders contact the Ignorants trying to convince them to join their side, while the Counters contact the Ignorant trying to make them become Stifler, thus, to avoid the Spreaders increasing their numbers.

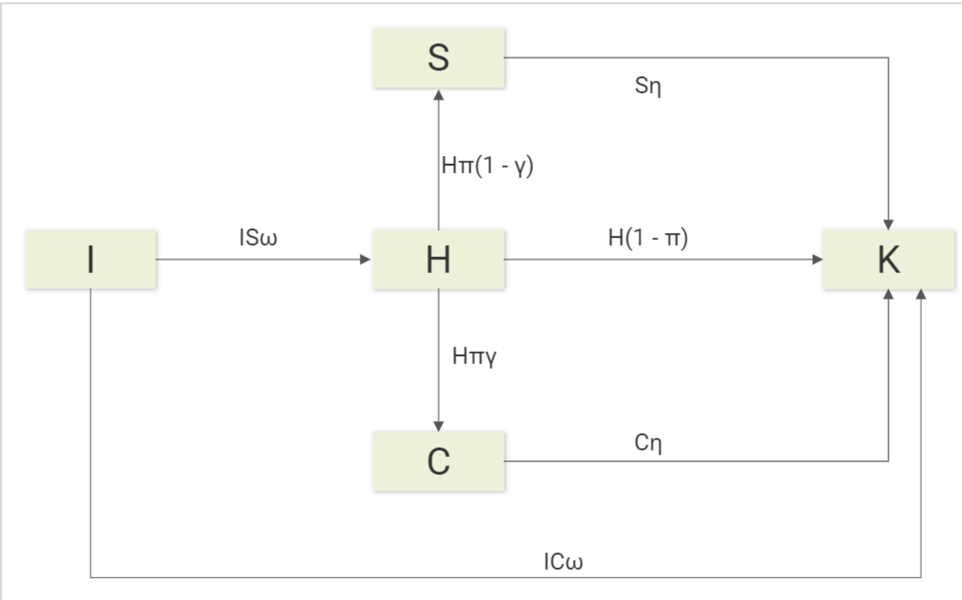


Figure 1: Flowchart for the IHSCK model

I, H, S, C and K, represents respectively the Ignorants, Heard, Spreaders, Counters and Stiflers. Also, consider that the size of the initial population, N, stays constant over time, thus: I + H + S + C + K = N, at every time step.

The model considers some common parameters between the states: I represent how interesting the rumor is with π, as an approximated probability that a person would find the rumor interesting. So, if one finds it interesting one will join the Spreader/Counter side, otherwise I consider the probability that it is not interesting as (1 - π). To represent the chance of interaction between two people ω ,I assume as a possible value *6/person per day*[2]. η is the forgetting rate and represents the probability that an individual spontaneously forgets the rumor. The forgetting rate is set as constant value, although it can be implemented as variable parameter, yet its behavior as such can be easily sensed. The chosen value, taken from literature, is 0.2[2]. At last, γ is the average educational level of the whole population. In this scenario is considered as the probability that an individual would *not* believe the rumor.

The differential equations for each flow are as follows:

$$\frac{dI}{dt} = -\frac{IS\omega}{|pop|} - \frac{IC\omega}{|\text{pop}|} \tag{1}$$

In equation (1) the first term describes the interaction between Spreaders and Ignorant through the parameter ω, as well as the second term. Both terms depend on the size of Spreaders/Counters and Ignorant and are proportional to the number of contacts per day. The term |pop|, which is the size of the initial population, helps define a flow, thus what percentage of the population follows the transition.

$$\frac{dH}{dt} = \frac{IS\omega}{|pop|} - H\pi(1-\gamma) - H\gamma\pi - H(1-\pi) \tag{2}$$

Equation (2) has two different types of flow: the first term refers to incoming population from the Ignorants, equal to the outgoing population in (1); the second part of the equation refers to the outgoing flow. In particular, the first one describes the people that have heard the rumor and are interested in it, but due to their low educational level, become Spreader; the second term accounts for the complement, which is joining the Counter party. The last one indicates the individuals who find the rumor not attractive enough, so they will flow into the Stiflers.

$$\frac{dS}{dt} = H\pi(1-\gamma) - S\eta \tag{3}$$

$$\frac{dC}{dt} = H\pi\gamma - C\eta \tag{4}$$

Both (3) and (4) have similar behavior so it makes sense describing them together. Spreaders and Counters both lose their population only due to the forgetting rate, and both are proportional to it. In both equations the first term describes the incoming flows, which are one the complement of the other.

$$\frac{dK}{dt} = \frac{IC\omega}{|pop|} + C\eta + S\eta + H(1-\pi) \tag{5}$$

[2] https://www.nature.com/articles/s41598-020-62585-9/tables/3

Equation (5) describe the incoming flow for the Stiflers, which collects the above seen outgoing flows, in order: the individuals convinced by the Counters to avoid the rumor, the Spreaders/Counters forgetting the rumor independently and the people not interested in the rumor after having heard of it.

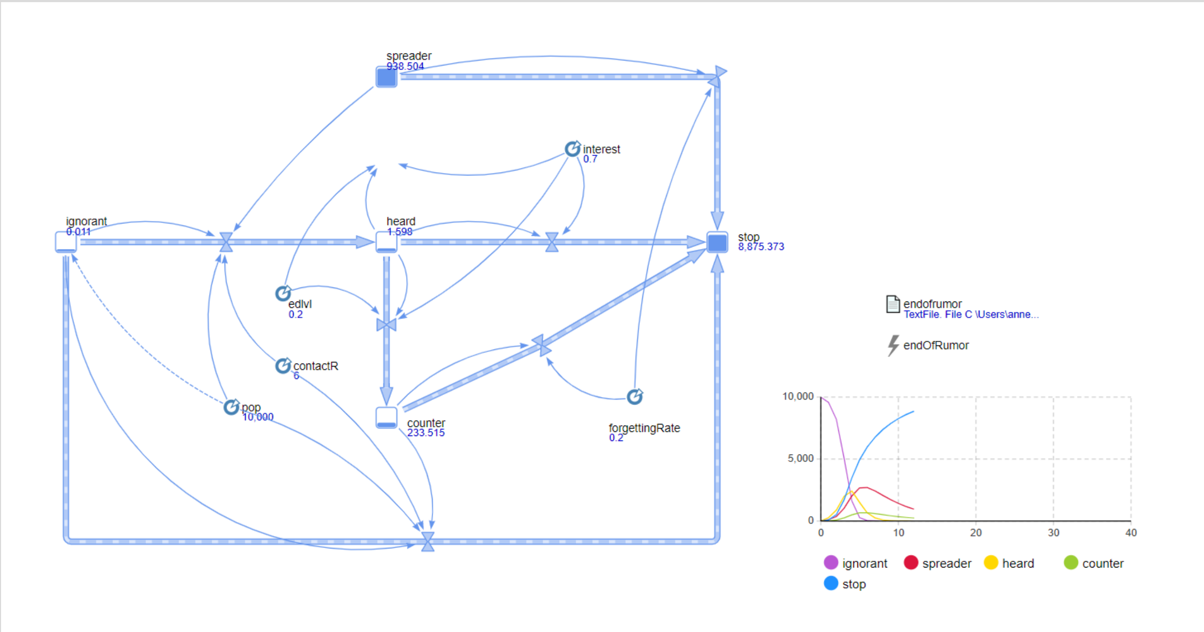


Figure 2: Anylogic view of a simulation run in the systems dynamic model. On the left the flowchart with the flows in-between each state and the parameters; on the right a graph explaining the population distribution at the moment

2.2 Agent-based

The agent-based model describes a similar yet different view on the topic. The model considers the same fundamental bases of the system dynamics approach (without the interest factor, which I consider not that useful for my investigation) so, it includes the same main classes as before: Susceptibles (Ignorants), Spreaders, Counters and Stiflers. As before the Spreaders will advocate for the rumor and the Counters will try to avoid its spreading by contacting the Susceptible to make them join the Stiflers. Even though the premises are the same, the agent-based model helps fashion the topic under a different light: every agent is crafted alone and takes part in the model as an independent agent with its own educational level and interval of believing, which will be discussed after, the model enables the use of a different approach than the being-contacted-join-a-side system more focused on the pairwise interaction of agents and lastly allows to mold the overall structure on a more realistic framework, such as the social network environment.

The classes are shaped differently: the two main agents are the Susceptibles and the Spreaders. Spreaders, at the beginning, are intended as *every media a person has access to*, like social networks, TV shows or newspapers, that start the rumor. Susceptibles are instead *the people that use social media*. Each Susceptible, after making contact with the rumor, can be in two main states: either Counter or Spreader, whose function is like in the previous case; convince the remaining population to join the Stiflers and avoid the spreading of the rumor or join the Spreader. In these two states the agent will repeatedly *post*, as intended in a social network fashion, about the rumor/counter rumor they heard. Each agent is set inside a network, to simulate the structure of a social network setting.

Let’s see in more detail the inner structure of these agents.

The Spreaders (agent type) have a very intuitive and simple structure:

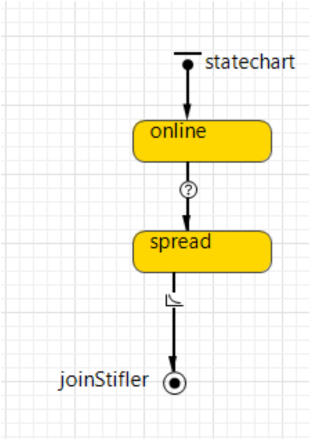


Figure 3: Agent-type Spreader flow chart

Where their only function is starting the rumor among the population and with a rate equal to the forgetting rate above seen, they will gradually and independently join the Stiflers and stop spreading the rumors. The Stiflers (agent type) are just a toy agent used to contain those that are no more involved in the general spreading model.

The Susceptible, instead, have a more complicated structure:

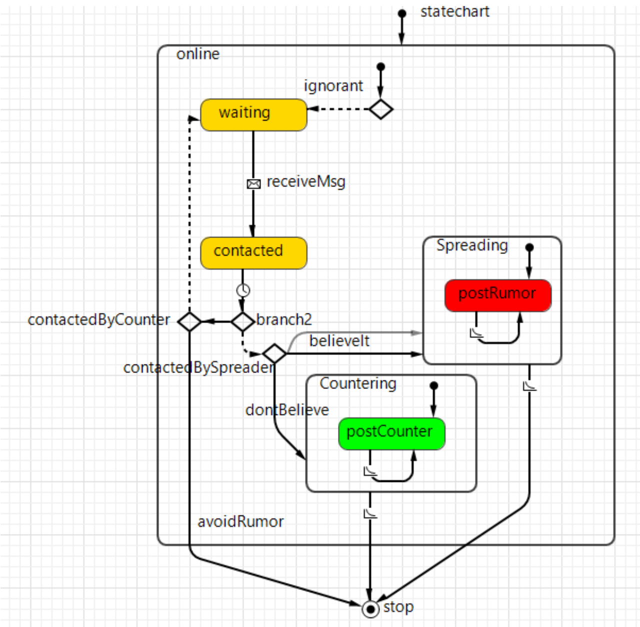


Figure 4: Agent-type Susceptible flow chart

Every agent can be in one of different states:

- *Waiting*: the agent has not yet been convinced by either a Spreader or a Counter.
- *Contacted*: the agent knows the rumor yet is still deciding what side to join.
- *PostCounter*: the agent starts contacting its contact to try convincing them to join the Stiflers and avoid the Spreaders.
- *PostRumor*: the agent repeatedly contacts its connections to spread the rumor.

When an agent is contacted by a Spreader it will get into the *contacted* state, then it will decide to either believe the Spreader and join them or refuse and join become a Counter. When the one contacting is a Counter, instead, the contacted agent will pass anyway through the contacted state and then decide whether to believe the Counter case and avoid the rumor joining the Stiflers or do not believe it and returning to the waiting state, expecting to be contacted by a more similar Counter or convincing Spreader. In both *post* states an agent can only contact those individuals who have not yet heard about the rumor, who are marked by an inner flag: *rumorHeard*. As before, a Susceptible will become a Stifler with the same rate, the forgetting rate, seen in the previous analysis.

The intuition under the "believing a person" is based on the common fact that a person would believe one who is *enough similar to him*. In particular, I supposed that a person with high educational level, in the same sense we considered above, tends to believe one who is *very similar to him,* while a person with lower education level tends to believe a broader audience, yet, not too dissimilar to him whatsoever. To model this *interval of credibility* I propose the following formulation, in which the sender is the one contacting and the receiver is the one being contacted:

1. Compute the absolute distance between the two agents' educational level, in other words how much they are dissimilar.
2. Compute the interval in which the receiver will believe a person.
3. Compare the two measures: if the distance is contained in the interval, then it will believe it and join the Spreaders side, otherwise the receiver will join the Counters side.

In this formulation I consider the fact that the receiver knows the sender, so, the agent knows which side the sender is on.

The interval of believing is computed in function of the receiver educational level as follow:

$$\varepsilon_{\gamma_{rec}} = \frac{100 - \gamma_{rec}}{1 + e^{\alpha(\gamma_{rec} - \beta)}} \tag{6}$$

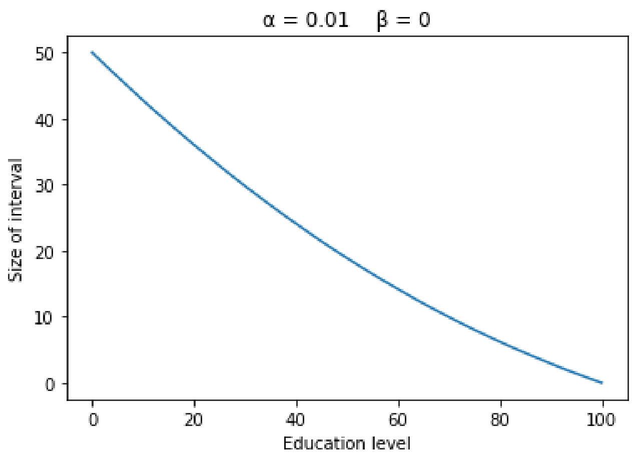


Figure 5: Interval of believing function.

The shape of (6) derives from the sigmoid function which is firstly being flipped along the y-axis to achieve a decreasing function in the interval [0, 100] and then I added some parameters on the exponent of the denominator to get the desire shape: α controls the slope, while β monitor the y intercept in the origin. The employed parameters, the ones displayed above the graph, are those that performed better and fit well with the representation of the topic I intend to describe. They were obtained through a trial-and-error approach.

The main parameter is the educational level γ. Each agent has an educational level randomly extracted from a Normal distribution, in particular a truncated Normal distribution with a variable mean μ_γ and a fixed standard deviation of 25, via the Anylogic built-in function:

normal(0, 100, μ_γ *, 25)*

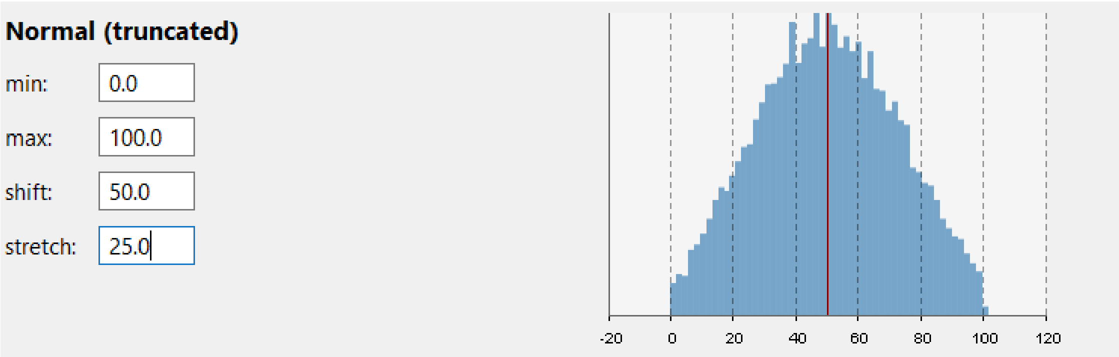


Figure 6: Anylogic view of the distribution tool generator of the Normal (truncated) distribution

The value of the average educational level is set accordingly before each experiment run.

To simulate the environment of a social network, the agents have been put inside a pre-built scale-free network. This feature is only used to shape the connections between agents and show them directly, despite being a characteristic that can lead to some different considerations.

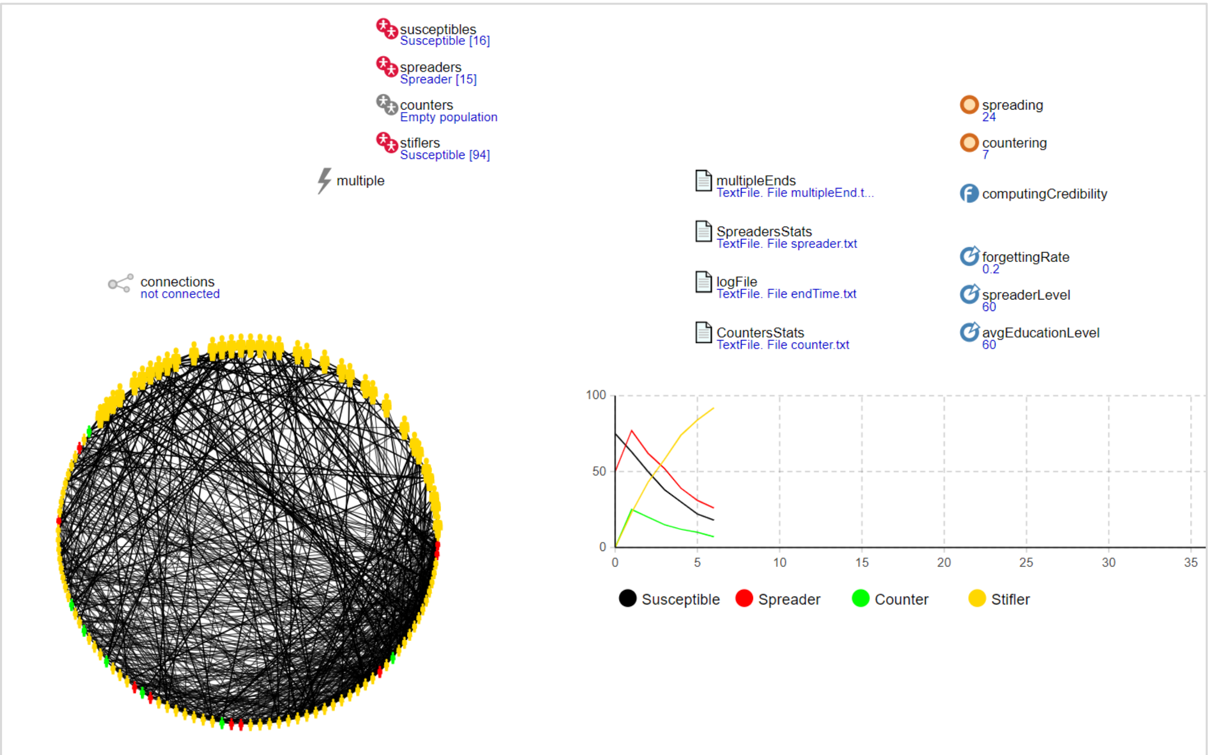


Figure 7: Anylogic view of a simulation run in the agent-base model. On top the agent-type put in the environment; on the left, the environment with the scale-free network; on the right, the employed files and events, the model parameters and the population graph.

3. Experimental results and analysis

The expectation for this scenario is that the duration, in days, of a rumor should be lower with an increase of the average educational level, as know should expect. Firstly, consider the system dynamic approach, then the agent based and lastly a comparison between the two.

In both cases the experimental setting is the same: multiple simulations have been carried out with the Experiment tool of Anylogic, in which I measure the duration of a rumor as the number of days elapsed between the start, which coincides with the beginning of the simulation run and the day the last Spreader becomes a Stifler. The choice is because a rumor is gone when no one is spreading it actively, the Counters can still be doing their job, yet when the Spreaders are zero, the rumor cannot be brought back up. In both approaches the day of end is flagged by an event which is triggered when the population of Spreaders reaches 0. The experimental runs are performed by increasing the value of the average educational level by 10 in the agent-based case and by 0.1 in the other case.

In the system dynamics picture the graph looks like what I expected:

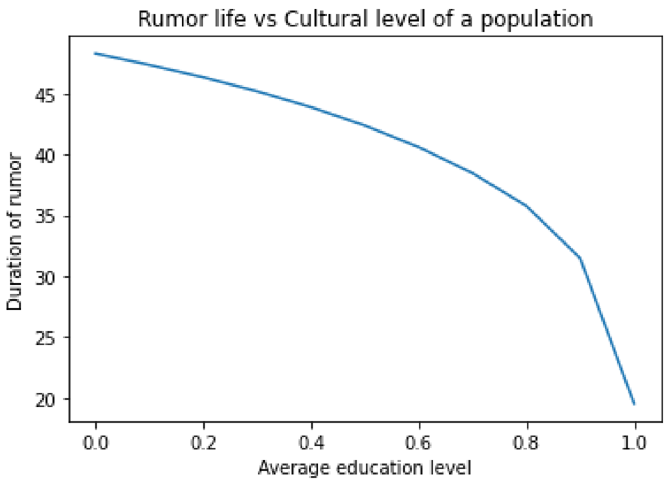


Figure 8: Duration of a rumor in function of the average educational level in a system dynamics model

Where the number of lasting days for a rumor decreases as the average educational level increases. Observe that in the rightest part, where the level approaches 1, the days quickly decrease due to the fact that the average level is intended as a probability, so, basically, each agent will become a Counter.

In the agent-based case the same relationship can be spotted with some respect:

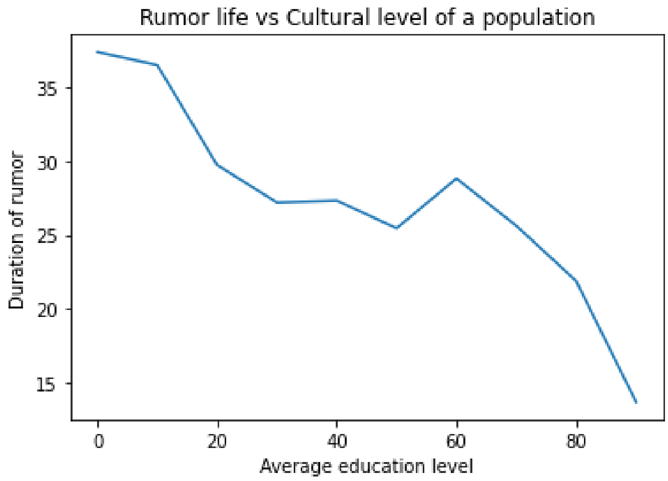


Figure 9: Duration of a rumor in function of the average educational level in an agent-base model

If we consider the mean of multiple runs, each of one having different random seed, instead of a single run which is subjected to the randomness of the simulation:

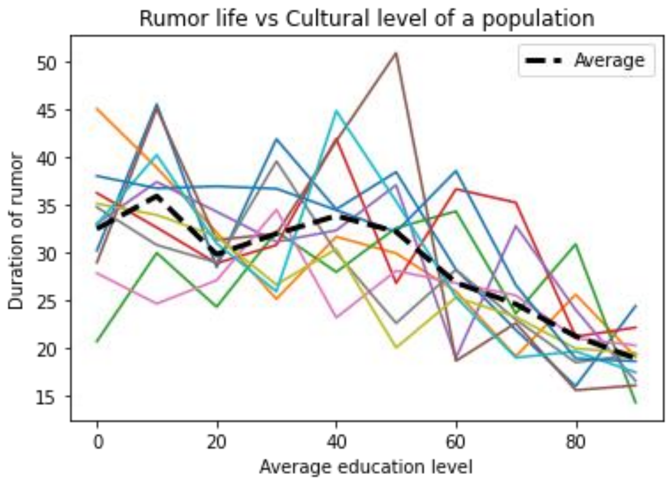


Figure 10: Multiple graphs plotted together to show the relationship between duration of a rumor and the average educational level, avoiding stochastic influences.

Note that some measures have very high peaks that can be seen as outliners due to the randomness of the simulation. Moreover, if we plot the two results together:

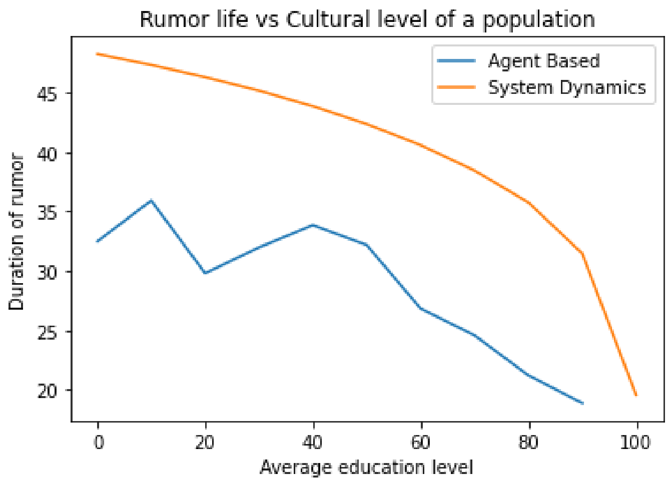
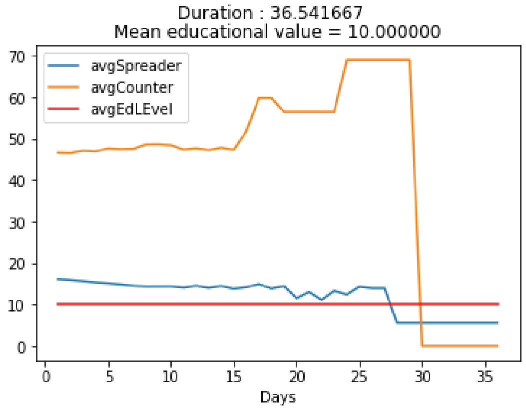
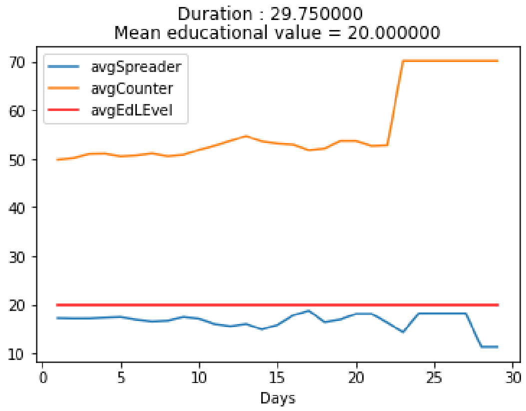


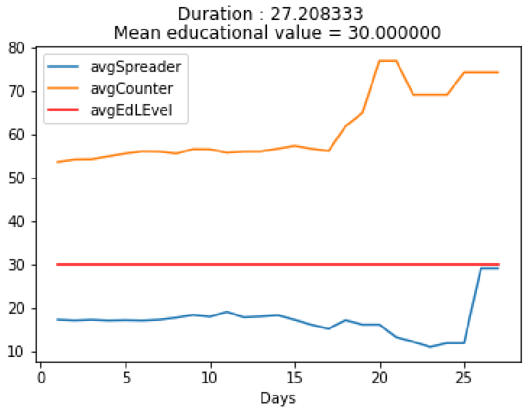
Figure 11: Comparison between the results of the two employed approach

The agent-based simulation seems to produce a rumor that lasts less, this is because of the initial population employed: in the first case I used 10000 Ignorants while in the second case 1000 Susceptibles, due to the limitations of the imposed by the machine. As expected the two variables are highly negatively correlated: with a -0.91 correlation coefficient for systems dynamics and -0.89 correlation coefficient for the agent-based model; results that one more confirms my intuition.

In the agent-based scenario more analysis has been carried out about the average level, per day, of the Spreaders and Counters. It is notable that the average level of education of the Spreaders and the Counters remain still between the runs, denoting that the average level may influence the duration of the rumor, but not the individuals involved in the spreading-countering of it.







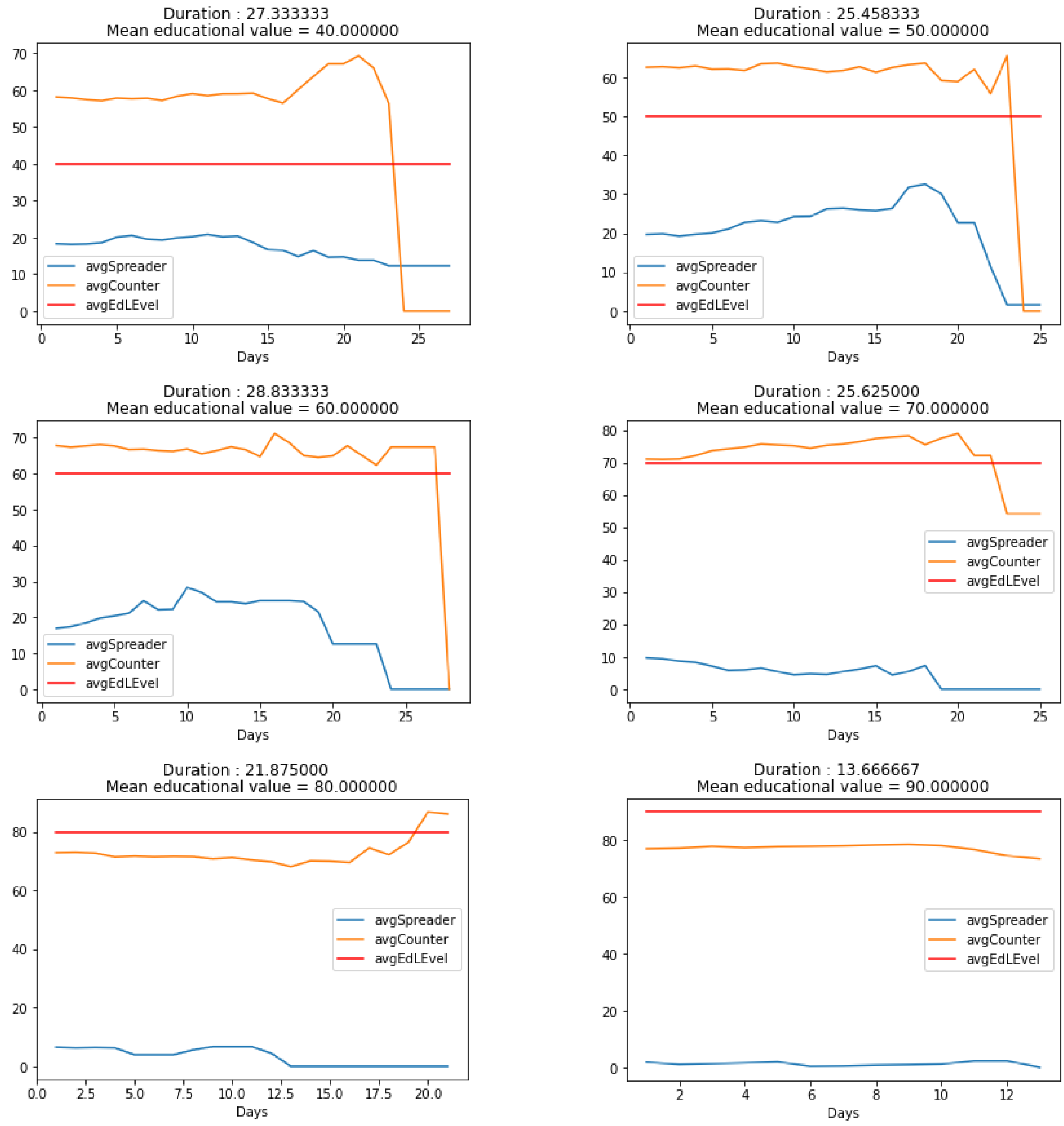


Figure 12: Average educational level per day in each simulation run for the agent-based model.

4. Conclusions

The project is based on estimating and studying the duration of a rumor among a population modeled with an educational level that wants to represent an individual as a value alone. Two different simulation approaches have been used to discover the underlying relationship, which at the end results in what one could imagine. The models contain some approximations and some parameters that have not been exploited very much, which leave room for a more focused study. For example, instead of employing a single value to represent each agent, one can define a vector of values, each of which models a different aspect like: age, interests, academic achievements, status and so on. My approach has been focused on the paradigm "*similia similibus curentur*" that has brought up the "interval of believing" function to determine whether a person will or will not believe another opinion.

References

Chen, X., Wang, N. Rumor spreading model considering rumor credibility, correlation and crowd classification based on personality. Sci Rep 10, 5887 (2020). https://doi.org/10.1038/s41598-020-62585-9

Dabke, D., & Arroyo, E. (2016). Rumors with personality: A differential and agent-based model of information spread through networks. Society for Industrial and Applied Mathematics Undergraduate Research Online, 9, 453-467.

Haixin Ding, Li Xie, Simulating rumor spreading and rebuttal strategy with rebuttal forgetting: An agent-based modeling approach, Physica A: Statistical Mechanics and its Applications, Volume 612, 2023, 128488, ISSN 0378-4371, https://doi.org/10.1016/j.physa.2023.128488.

Kermack, W.O. and Mckendrick, A.G. (1991) Contributions to the Mathematical Theory of Epidemics-I. Bulletin of Mathematical Biology, 53, 89-118.

Martine Collard, Philippe Collard, Laurent Brisson, Erick Stattner. Rumor Spreading Modeling: Profusion versus Scarcity. ASONAM 2015: IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Aug 2015, Paris, France. pp.1547 - 1554, ff10.1145/2808797.2809299ff.ffhal-01252936

Tseng, S.-H.; Son Nguyen, T. Agent-Based Modeling of Rumor Propagation Using Expected Integrated Mean Squared Error Optimal Design. Appl. Syst. Innov. 2020, 3, 48. https://doi.org/10.3390/asi3040048

Wang Y, L W, Qing F. Rumor dynamic model considering intentional spreaders in social network. Discrete Dyn Nat Soc (2022) 2022:1–10. doi:10.1155/2022/6044099