**Individual Assignment questions and Answers (Social Network)**

**Student Name: Ioannis Konstantakopoulos**

**St. Number : 0960047**

**Question 1:** Read the research paper “Students under lockdown: Comparisons of students’ social networks and mental health before and during the COVID-19 crisis in Switzerland”, available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0236337

Answer following quetions within 600 words:

1. The five types of social relations the authors are taking into account to build the social network and how these networks might differ in function, intimacy and stability

They study the students in two undergraduate programs at a Swiss university during COVID. In the article, multiple dimensions of social relations has been taken into consideration. They assess change in five self-reported social networks within the student community: pleasant interactions, friendships, emotional support, informational support, and co-studying. Those five has been chossen as they differ in function, intimacy, and stability. They assume that stronger relationships will be characterized by an overlap in multiple dimensions .

We hypothesize that students at the time of the COVID-19 crisis nominate fewer fellow students than before in multiple dimensions of their social networks. It is expected that social networks between students are less connected, and thus will show a higher rate of students being socially isolated. They hypothesize that stronger ties are more likely to survive at the time of crisis, thus when they are characterized by a previous overlap in multiple relational dimensions.

1. What types of metrics do the authors use to measure the social network during and prior lockdown? And how do these metrics change?

Firstly they measured the five self-reported networks which are pleasant interaction, friendship, costudying, informational support, emotional support and also in mental health they where measuring these indicators : Depression, Anxiety, Stress and Loneliness.

For any of these items two key measures has been created to test hypothesis 1a and 1b.

The results where students reported less pleasant interactions and less styding partners during lockdown than before. Friendship networks remain stable while while informational support and emotional support nominations slightly increased. Changes in mental health, the students became more depressed, slightly more anxious more stressed and lonelier.

1. How do the authors analyze the face-to-face interaction network and digitial interaction network during COVID-19 crisis? What are their conclusions?

As a consequence of lockdown and social distancing only few face-to-face interactions were reported by students in general. They reported to interact through messaging, through calling either voice or video and using extensively the social media. The interaction network were overlapping, but not identical. Most of the calls and messaging were reported as pleasant interactions. Seventy-five percent of the physical interactions were overlapping with any of the digital interactions. Even though there was a lockdown and physical interaction were almost impossible students replace them by with digital interactions. Although the digital dawn the overall interactions were less during the COVID pandemic.

1. What do the author mean by ‘survived ties’? And which type of social ties survive better during COVID time?

Survived ties is the number of overlapping dimensions in the social network. The baseline survival rate which represents the proportion of interactions ties is been compared with the remaining survived ties between two individuals inside the network during the examination period of April 2019- April 2020. This way they can assess if having overlapping network dimensions at the start of the research in September 2019 contribute to the survival rate of ties until April 2020.

The type of ties that survived the wave was friendship and support networks as it stayed the same but the density of functional networks dropped in density.

5) What will be the policy implications that you will draw based upon their results?

The results on mental health of the students were clearly decreased such us social isolation, while other stressors like student competition are reduced during the COVID period. Overall although the mental health has been clearly decreased during the pandemic period. The results indicate the importance of social networks on mental health and wellbeing in general for instance students that are socially isolated and have less support they will probably develop more health problems. It is important to study how the pandemic affects the social networks and the mental health of the students. Even though measuring social isolation is important governments and universities should take into consideration the increasing in mental problems. Maybe a better policy of restricting the students but not having a full lockdown is needed as the result indicates the importance interaction to reduce mental problems.

**Question 2:** Rumors are the basis for viral marketing and, therefore, rumors diffusion is a topic widely studied. Zanette 2001 developed a rumor spreading model based on the epidemiological SIR model. Read the paper of Zanetee 2001 and answer 1) how did they reframe the popluations in epidemiology (i.e., S, I, R) into a rumor spreading process? Mazzoli et al. (2018) twisted the model into an agent-based model (ABM) to simulate rumor spreading. Read the paper of Mazzoli et al. (2018) and summarize 2) how did Mazzoli et al. set up an ABM to simulate the real-like diffusion of information and misinformation (e.g., what are the agents in their ABM, assumptions on the micro behaviours of agents, and key paramters)? And 3) how well did Mazzoli’s model reproduce the diffusion of information during the announcement of the discovery of the Higgs Boson on Twitter? (Provide your answer within 500 words, Question 2, 8 points)

Paper:

Dami´an H. Zanette, 2001. Critical behavior of propagation on small-world networks, avaliable from <https://arxiv.org/abs/cond-mat/0105596>.

Mattia Mazzoli, Tullio Re, Roberto Bertilone, et al, 2018. Agent Based Rumor Spreading in a scale-free network, available from: <https://arxiv.org/abs/1805.05999>.

1. They created a model with N elements where at each step, each element adopts one of the three possible states. As in SIR epidemiological models, susceptible (S), infected (I), and refractory (R). On each step a randomly selected infected element i contacts an element j. Then firstly, if j is in susceptible(S) state, it becomes infected, secondly, if j is infected or refractory, i becomes refractory. These rules are better explained and understanded in the frame of rumor spreading process, where S-elements they have not hear the rumor yet, I-elements have heard the rumor and they are willing to spread it and R-elements have no interest of transmitting the rumor anymore. At the start only one element is infected and the other(N-1) are in S-state. On first stage of the evolution the I-elements increases, this also indicates a growth on R-population and the contracts between themselves and R-elements happens more frequent. Later the I-elements starts to drop, until they vanish completely, and the evolution stops. There are NR elements that have been infected during any stage of the process.
2. On ABM there are debunkers agents that try to make spreaders change their mind on the rumors they spread. Two parameters has been used first, γ that represents the transition rate to spreader. Secondly, α represents the rate of transition to stifler. They created two node species for the model to be flexible for more research. They represent the news with a global variable ranges between 0-1 and these news are available to every agent with a certain probability to simulate the information overflow. Each agent has the opportunity to either spread or not the news based on personal preparation which is an individual threshold randomly assigned at the start of the model. In the model they represented three types of diffusion. Firstly, spontaneous spreading of the information after direct visualization happens when news surpasses the threshold, and the information is reliable. Secondly, collective influence when more than 30% of the agents friends are spreaders or a very influence hub agent between the friends of mine share the information then the threshold decreases and makes him more naïve. Lastly communication persuasion, happens when undeployed agents are friends of spreader agents, these send them messages to inform them about the validity of the news. If the interlocutors have a similar preparation on the topic, the undeployed agent’s threshold decreases to raise his probability of spreading due to augmented faith in his friend.
3. The conclusion was that the results where somewhat similar and the activation dynamics of the graphs are the same at lest from a qualitatively perspective. There where some similar trends of the density but the main difference between the real data and the simulated ones is the time they reached the peak. In the first the peak has been reached around t=80-85 as we see from the figure in the paper, and from the later the time is around t=25-40 roughly.

**Question 3:** Describe the principles of independent cascade, threshold models, and epidemiology models such as SIS and SEIR models. Can you interfere the assumptions that each of them make on the dismission of signal within the network? Under each model, provide one real-world example that might fit their assumptions and explain why. (Question 3, 8 points)

1. **Independent cascade Models:** Instead of a single probability infection, there is a probability of infection associated with each edge. The probability Pu, v is the probability of u infecting v. This probability can be assigned based on frequency of interactions, geographic proximity, or historical infection traces. Each node, once infected, has the ability to infect its neighbor in the next time step based on the probability associated with that edge.

An example: Lets say we have a product and we want to sell it to our friends. I am the one that sells it so I am the activated green node (once a node is activated (in our case has the product will continue telling that the product is worth buying it) can not be deactivated you can not for instance unbuy a product) that’s the t=0 moment. At the next step I have the chance to sell my product to my friends Lena, Hans and Jimmy, I have only achieved to activate hans, at the next step I will be inactive and only hans will have the ability to activate others nodes(friends) and they will continue to do the same.

1. **Epidemiology Models:** In this models there are 3 stages Susceptible: can catch be infected with probability of α within contract with an already infected node. Infected node which is capable of spread the virus with 1/γ time steps. The third one in the Recovered either healed or dead. The virus strength is the division between α and γ which is α/γ. In SIS model there is non status as inactive or immune and the node can be infected again. In SEIR the difference is there is another state, of expose (E) which mean the node is infected but not yet infectious (can not transmit the virus)

**An Example:** I will provide an example with a virus as the model called epidemiology. Starting all the people are in susceptible and everyone has a probability to get sick with the virus. All the infected people can spread the virous to their neighbors as long as the virus is in their system(some virus have longer time remaining in our system) After becoming infectious the person goes to the recovery status which is either dead or healed in either position the individual can be infected again only is SIS model which the virus do not cover long lasting immunity. Always is better to prevent being infected rather than getting sick and then stop the spread. So either the α has to be reduced with social distancing for instance or hand washing.

1. **Threshold Models:** In this model nodes can have 2 stages active(1)-inactive(0) and once a node is active will remain active all the time. The influence comes only from direct neighbor. Each node has an acceptance threshold ranging from 0-1. In each step each inactivated node reviews the status of all his/her direct neighbors N(i) and is activated if the weighted fraction of active nodes exceeds threshold. (To activate a node the sum of all influences from neighbors > threshold ) the model stops when all the nodes are activated of the activated ones are saturated.

**An example:** In our group assignment Me,hans,lena,jimmy if one saw a movie and want to say to all of us to watch it. In this model every node has a different influence to other nodes for instance my node with jimmy is bigger (I influence him more than hans) At t=0 I will influence all my friends and if my persuasiveness > threshold of each one of them lets say lena then I was able to activate only lena. At the next step t=1 me and lena will try to activate jimmy and hans. As the steps move on either all of my friends will see the movie or there might be one or two with big enough threshold that we cant activate them to watch the movie. Here Hans will not be active as his threshold is higher to the sum of our influence.

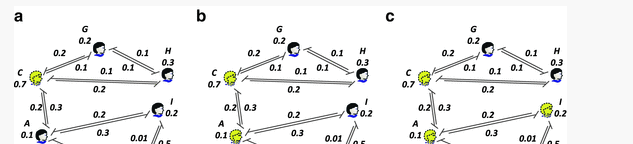


Figure 1: C:Me, A:Lena, I:Jimmy , B:Hans t=0 only me is active,t=1 I activate Lena and t=2 we activated Jimmy

**Question 4:** In Ex 3 of group assignment, the threshold model assumes a deterministic threshold, e.g., one will adopt only if the fraction of one’s neighbour exceed a certain amount (threshold *qi*):

F(xi)=1 | xi>=qi

F(xi)=0 | xi<qi

where F(xi) is the probability function of node i to adopt, subject to the xi, which is the fraction of adopters in node *i*’s neighbours; and qi is the threshold of node *i*.

Such a clear-cut adoption function, however, might violate the complexity in human being’s decision making process. Empirical adoption rates with k and k-1 neighboring adopters ofter have similar magnitudes. Rather than positing determinism, analyses of discrete choice problems typically hypothesize that individuals are random utility maximizers leading to positive choice probabilities. Embracing such complexities in human decision-making, efforts are made to fine tune the probability function of threshold model.

One way is to allow a small probability of below-threshold adoption.

F(xi)=1 | xi>=q1,i

F(xi)=pi | q2,i <xi<q1,i

F(xi)=0 | xi<q2,i

where F(xi) is the probability function of node i to adopt, subject to the xi, which is the fraction of adopters in node i’s neighbors; q1,i is slightly smaller to the original qi, and exceed which will lead to an adoption for sure; q2,i is somewhere between q1,i and 0, which will lead to a small chance of adoption (pi) if exceeds.

The other way is to change the deterministic adoption function to an ‘S-Shape’ logit function.

Where F(xi) is the probability function of node i to adopt, subject to the xi, which is the fraction of adopters in node *i*’s neighbours; and qi is the threshold of node *i*; b is a new parameter that is introduced to control the shape of the probability funciton.

For a given seed size, change the deterministic probability function to the above two more realistic assumptions (i.e., one allows below-threshold adoption, the other with S-shape logit function) and play around the parameters such as pi, q1,i, q2,i and b. Check how does such more realistic assumptions affect 1) final diffusion size, 2) the shape of diffusion curve, and 3) the target nodes of seeding. (Notes: to make results comparable, you should make sure the expected values of thresholds are the same cross different assumptions; you can work on the network and the threshold model for the vegetarian reciept in the group assignment or use your own data and threshold distribution; you should test the result for different seed sizes as it might change your conclusion) (Question 4, 14 points)

Using the Vegetarian to create the threshold model.

1) The diffusion size as we expect is not going to be the same as before. Simple contagion model is easier to reach everyone as there is no threshold. So its going to be slower, the sigmoid functions gives each node the addition that can be activated throughout the whole procedure not only at the start. As the node will be active throughout the experiment. The experiment will only stop if every node will be activated or the activated nodes are saturated.

2) The shape of the diffusion curve its a sigmoid S-curve.

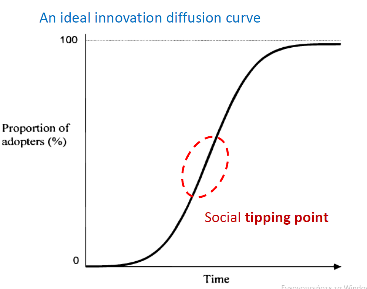
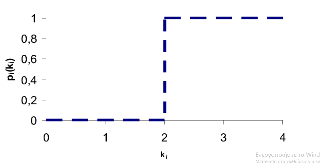
So that the curve will change from the sample plot to the left showing a binary curve it changes into a curved function. The code and the plots I created will not provide a plot as clear in it’s shape and we will not see it as in the figure below. However, one can notice by looking at the individual points within the plot that the values first increase more steeply

Figure: The change from simple to sigmoid shape curve.

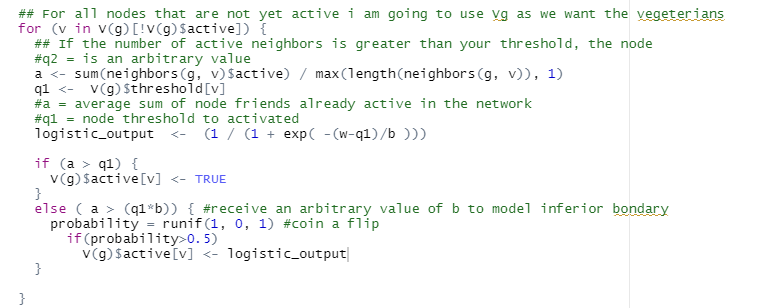
and then slowly as the s-curve indicates in the figure. After the social tipping point the curve looks as it works on its own and the vegetarians as they want the word to spread done correct moves to persuade more people to follow their trend.

3) The sigmoid function we introduced in this example it is still a function of the same parameters which is why I was unsure whether new seeds would actually improve the model (Almost same as question 36 of the group assignment).The seed is going to change with the sigmoid model as we will see in the code and the results different nodes will be selected from the model.

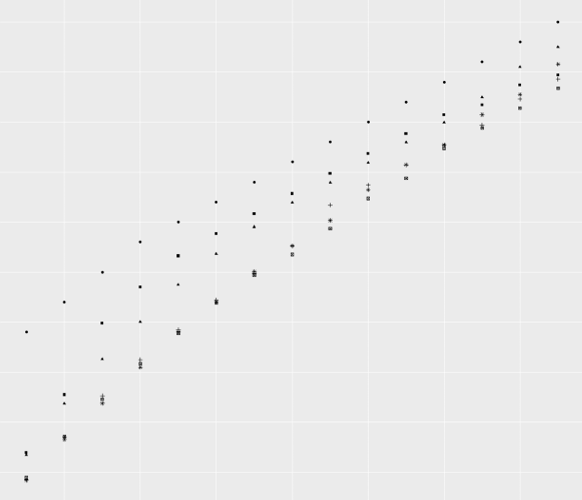
As we saw for question 36 of the group assignment which is similar to our question here with the simple plot the model will probably reach the full potential in short steps around 15(based on our group assignment) because the results are either 1 or 0. Now we introduce the b value to make the curve more smooth and close to our s-curve goal. Now we don’t know if the model will stop when it reaches the full network or before (according to the saturated activated nodes).

I will provide the code its almost the same as the group assignment but I made changes as this questions needs. First I will create the 2 thresholds we need q1 and q2 the q1 is the first and the q2 is the second. Our probability function F(xi)

That’s the code I used:



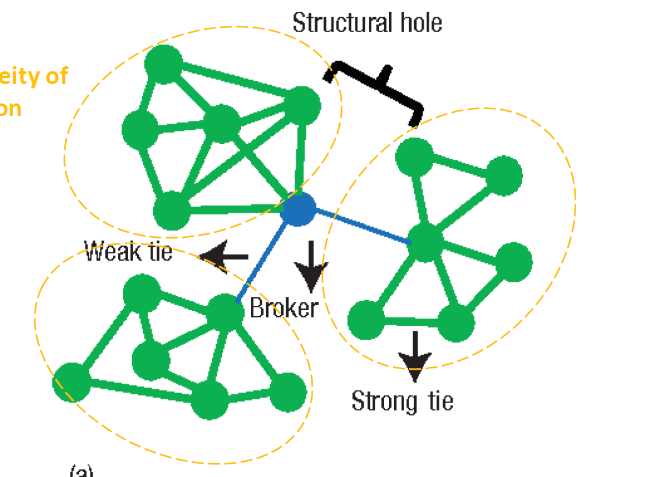
That’s the plot I got:



For different values of b as I said earlier, I have different curves of the plot on the Y-axis I have the spread on the network and on the X-axis I have the number of nodes. The curve might not look like an s-shape but I think that’s the perspective of solving this problem.

**Question 5(Bonus):** After the introductions of more realistic assumptions of threshold, do we need more weak ties to accelerate diffusion, or the opposite (i.e., we need more strong ties to accelerate diffusion)? Can you test your conjecture by rewiring the network strucutre (e.g., adding long ties or triadic closure) and monitoring the change of diffusion scale and speed? (Question 5, 5 points).

There is not one answer to this question in my opinion, because it depends if we are looking for a simple or complex contagion. In simple contagion we need more weak ties and in complex more strong ties.

I will go for complex contagion as I think is more important in real life problems. The sum of the neighbors influence need to influence be enough in order to surpass the threshold and activate the node. For instance in political conversations we need to amplify the strong nodes(close nodes) rather than the weak nodes, which are the ones connecting 2 communities together as we see in the figure below

In order to amplify the diffusion we need just enough nodes to reduce redundancy, using much more edges the connection breaks and there will be no diffusion as we can see in page 37 of lecture 7 and in complex contagion using the model with high clustering and no weak ties we have the faster results of spreading the information or the virus to the whole network.