

Individual Assignment Social Network

Mirthe Hendriks 6866999 m.m.j.hendriks@students.uu.nl

Question 1

Students under lockdown: Comparisons of students' social networks and mental health before and during the COVID-19 crisis in Switzerland",

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

Elmer, Mephram and Stadtfeld (2020) identify five types of social relations: pleasant interaction, friendships, emotional support, informational support, and co-studying. Logically these differ in function, intimacy and stability.

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

The metrics used to measure the social network are the outdegree of each respondent and the proportion of out-isolates. The outdegree represents the number of ties that the participant reports in each social network. Out-isolates is the fraction of the participants who answered that they have no ties in a network. Participants in the current-year and previous-year cohort filled out a survey with social network items. The authors used a name generator approach, allowing participants to list the names of other students. For each mentioned name, a tie was drawn between this participant and the individual within the particular social network. Mutual nomination refers to the relative frequency of two participants reporting one another in the same social network.

The authors hypothesized that students at the time of COVID-19 will report fewer fellow students in various networks (H1a), and that social networks will be less connected with more socially isolated students (H1b). These hypotheses were partially confirmed; the within-person comparison showed that overall outdegree decreased for social interaction and co-study networks and the proportion of out-isolates increased. The metrics did not change significantly for friendship and social support networks. Similarly, the between-person comparison showed decreased outdegree for pleasant interaction and co-study networks, while the proportion of out-isolates did not change significantly.

3) How do the authors analyze the face-to-face interaction network and digital interaction network during COVID-19 crisis? What are their conclusions?

Authors analyzed the face-to-face interaction network and digital interaction network during COVID-19 using multiplex network visualizations. The participants reported on communication with fellow students in the network, with specific items on face-to-face interaction (physical communication) and digital interaction with messaging applications, calls and social media. The relative density and overlap between social interaction networks in COVID-19 times is visualized with a Venn diagram. The authors concluded that the overall number of interactions between students during COVID-19 decreased. Students had fewer face-to-face interactions (less pink ties), and instead interacted in the digital realm. The Venn diagram shows dense interaction networks for digital contact.

4) What do the author mean by 'survived ties'? And which type of social ties survive better during COVID time?

The authors label a tie as a 'survived tie' when a tie was reported in April 2020 after also being reported in September 2019. Practically, this would imply that in September 2019 a participant filled in a name as a connection in a specific network and then did this again in April 2020. The ties were maintained during the social distancing phase and thus survived the COVID-19 crisis.

The second hypothesis in this paper was that stronger ties are more likely to survive the COVID-19 crisis. The stronger social ties are those that overlap in multiple relational dimensions. Measuring this entailed a count of the overlap of the number of ties between two individuals in the social network dimensions (pleasant interaction, friendship, co-studying, informational support, emotional support, where pleasant interaction is the excluded reference). The results indicate that the weaker, one-dimensional ties were less likely to survive compared to the strong ties with overlapping dimensions. Only 45.4% of the one-dimensional interaction ties survived, while between 84.9% and 87.2% ties survived when there was overlap in two to four networks. Similarly, looking at co-study ties, the survival rate was 5.9% for one-dimensional ties and 57.1% for four-dimensional ties. As such, hypothesis two was supported.

5) What will be the policy implications that you will draw based upon their results? (Question 1, 8 points)

The paper addresses the necessity to contemplate potential negative effects on individuals' social networks and mental health resulting from the implementation of COVID-19 measures. Universities should make it their policy to advance hybrid teaching strategies and promote opportunities for students to socialize and interact in an informal social context. These opportunities are especially relevant for recently enrolled students who have not yet been able to form social ties with fellow students. The COVID-19 measures might make face-to-face interaction unattainable, but university staff should consider facilitating online events. This could entail digital study groups, mentoring and peer group work. Besides, additional support for at-risk students should be considered, for example by providing psychological counseling.

Question 2

Read the paper of Zanette 2001 and answer 1) how did they reframe the populations in epidemiology (i.e., S, I, R) into a rumor spreading process?

Zanette reframed the SIR model into a rumor spreading process. The epidemiological SIR model assumes three states for a N -element population; susceptible(S), infected(I), and refractory(R). The diffusion process entails that at time t an infected element (i.e. person) comes in contact with other elements, and if these elements are in the susceptible state they will be infected. After a certain time period, the infected element recovers and reaches the refractory state. In the rumor spreading process, elements in state S have not yet heard the rumor, but they are susceptible to hearing it. The I -elements have heard the rumor and are willing to spread it and the R -elements have heard the rumor but they have lost interest so they will not spread the rumor anymore. In the rumor spreading process, it is assumed that the elements are nodes of a small-world network, where only connected elements can contact each other. Over time, the rumor spreads; there is an increase in I -elements. But at a certain point, R -elements become more common, the I -population decreases, and the rumor is spread less and less.

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 parameters)?

Mazzoli et al. argue that the real-life diffusion of (mis)information is more complex social behaviour than the SIR rumor spreading model assumes. They adopt an agent-based model (ABM), which assumes that agents are autonomous, social and perceptive beings capable of making (proactive and reactive) decisions out of self-interest. These agents can decide to be non-spreaders (undeployed) or spreaders (deployed) of the rumor. In contrast to the epidemiologic models, the ABM's rumor spreading process contains 'emergent debunking behavior agents', i.e. agents who tend to fact-check the information they encounter. These debunker agents attempt to convince spreaders that they are promoting misinformation. ABM is built on certain assumptions on the micro behaviors of agents. First, agents have a threshold of skepticism; if information is reliable enough to pass this threshold, the agent will spread the information. Secondly, neighbors in the network influence the agents' threshold. For example, if 30% of the neighbors or a hub neighbor spreads the information, this information might be considered reliable enough to exceed the agent's threshold. Additionally, communication persuasion between agents influences this process. Once undeployed agents come in contact with spreaders, the spreader agent will try to convince the undeployed agent of the reliability of the information.

The ABM is generated based on the Barabasi-Albert model with scale free networks and is regulated by two key parameters: " α representing the rate of transition to stifer (R) and λ representing the transition rate to spreader (S)".

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? (500 words, Question 2, 8 points).

The ABM model reproduced the diffusion of information on the Higgs Boson discovery on Twitter fairly well. Mazzoli et al. reproduced the diffusion by measuring and modeling the density of activated Twitter users over time in response to the discovery. In turn, they compared the log-scaled plot on the density of active users versus time of the real data with the same plot made for the diffusion process reproduced with the ABM. The plots show similar activation curves and the authors conclude that the activation dynamics are the same from a qualitative perspective. The model thus seems to perform well on this real social network diffusion process.

Question 3

Describe the principles of independent cascade, threshold models, and epidemiology models such as SIS and SEIR models. Can you infer the assumptions that each of them make on the dissemination 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)

The independent cascade (IC), threshold, SIR and SEIR models are used to model diffusion processes in human social networks. IC, SIR and SEIR are simple contagion models, characterized by the assumption that one single successful contact in a network is sufficient for diffusion. First, the IC model. A defining principle of the IC model is that it assumes people in the network can have two states: active ($S=1$) and inactive ($S=0$). The model also

assumes that once a node has been activated, it cannot become inactive again. The activation of nodes is measured at subsequent points in time; at time $t=0$, k nodes are activated (i.e. seeded). The model assumes that any activated node only has a single chance to activate its neighbor nodes. Whether this activation is successful is determined by the probability for the edge connecting the nodes. The diffusion process will stop when all nodes are activated or when the number of activated nodes are saturated (no more nodes can be active), the IC model assumes that only when this happens the signal will be dismissed.

A principle in which the SIR model varies from the IC model, is that SIR establishes that nodes can have three states: susceptible(S), infected(I), and recovered(R). SIR is an epidemiological model, which is recognizable in the definitions of the three states. Susceptible refers to the people in a network who are not yet infected, but have a certain probability of being infected upon contact with an infected element. An infected person can potentially infect susceptible people in the network for a certain period of time ($1/\lambda$). After time period t , infected individuals will have recovered (or died). The SEIR model is similar to the SIR model, except that it assumes an additional state between S and I: exposed(E), i.e. infected but not yet infectious. These epidemiological models assume for the diffusion process that once an individual is recovered, they can no longer be infected. The signal diminishes and the diffusion process comes to an end when no one in the network is exposed or infectious anymore.

In contrast to the previous models, the threshold model is related to the complex contagion. Complex contagion is based on the assumption that human beings need more than one contact in their social network to convince them of a certain behavior. Similar to the IC model, according to the threshold model people can have two states: active ($S=1$) and inactive ($S=0$), and activated nodes won't become inactive. The main principle of the threshold model is that a person's behavior is influenced by its direct neighbors in the network. Each person in the network has an acceptance threshold (between 0 and 1), and if for an inactivated person the weighted fraction of active neighbors exceeds this threshold, this person will be activated. For example, if an inactivated person hears from one active neighbor, this person might be convinced (the acceptance threshold is not exceeded). However, if this person has three active neighbors trying to convince him/her of something, the acceptance threshold might be passed and the person would be activated. The signal will be dismissed when all people in the network are activated or the number of activated people is saturated.

Examples

A real-world example that would fit the assumptions of the IC model is Kevin buying an extra fluffy pillow and communicating this to his social network. He talks to his friends about the pillow, and some of his friends might also buy the pillow. In accordance with the model's assumptions, a person is either active (bought the pillow) or not. For a step in time, with a certain probability, Kevin might activate his friend Jiamin. Now Jiamin has bought the pillow, for a next step in time, she might activate people in her social network. The assumption of simple contagion holds for this real-life example, as one contact is sufficient for a friend to become active.

An exemplary real-life case of the epidemiologic SEIR model is the spread of the COVID-19 virus. As the model assumes, with the spread of the virus, a susceptible individual can be exposed and become infected once they come in contact with another infected individual. After a certain time period, an infected individual recovers (or dies). One assumption of the

SEIR model that does not hold for the real-life case is that unfortunately an recovered individual can be infected again with COVID-19.

A real-world example of the threshold model would be the process were a social group attempt to convince Mirthe to drink beers instead of study for the exam. This example meets the threshold model's assumption of complex contagion; more than one contact in Mirthe's social network needs to convince Mirthe to drink in order for her to become 'active' and drink beer. Mirthe has a certain acceptance threshold (e.g. 0.7) which needs to be exceeded by the weighted fraction of the 'active' friends in the social network. Every friend has an influence Eva has 0.4, Rosanne has 0.3 and Robin had 0.1. $0.8 > 0.7$ so Mirthe is activated, she has been convinced by her friends to have a beer.

Question 4

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 p_i , $q_{1,i}$, $q_{2,i}$ 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. (Question 4, 14 points)

The probability function with the deterministic threshold entails that a person will only be activated once the weighted fraction of the neighbors exceeds this set threshold (with notation q_i). The probability for a person to be activated is either 1 (activated) or 0 (not activated). If the fraction of activated neighbors in this persons social network x_i is larger than this person acceptance threshold q_i , the person is activated. In other words, the person has been convinced, influenced by his/her neighbors.

Nevertheless, the clear-cut adoption function of the deterministic threshold model might not be in line with the complexities of the human decision making process. A below-threshold adoption function was coded to allow for a small probability of below-threshold activation. In the LT function (see below), for all nodes that are not yet active, it is specified that (1) if for a node i the weighted fraction of active neighbor nodes x_i is larger than the acceptance threshold $q_{1,i}$, the probability of node i to adopt is 1, and (2) if the weighted fraction of active neighbor nodes x_i is smaller than threshold $q_{2,i}$, the probability of node i to adopt is 0. Now additionally, (3) if the faction of adopters x_i is larger than $q_{2,i}$ (the lower threshold) but smaller than $q_{1,i}$ (the upper threshold), there is a small probability that node i will adopt nonetheless (notation $F(x_i)=p_i$).

A first finding is that the final diffusion size (i.e. the proportion of adopters) is dependent on the various acceptance thresholds of the nodes in the social network. The below-threshold adoption function has two threshold values. The upper threshold value q_1 has the most impact on the diffusion size and speed. If all nodes in the network have a higher q_1 the diffusion process will take more time and the final diffusion size will be lower. In other words, after a number of time steps, the proportion of adopts will be considerably lower than if the people in the network overall would have had lower q_1 thresholds. The diffusion curves in figure 1 visualize this. We also see that if throughout the network the q_1 thresholds increase, the diffusion curve becomes less steep.

Nevertheless, when comparing the diffusion curves of the deterministic threshold function and the below-threshold adoption function (keeping other parameter stable), we might find that the below-threshold adoption function has a steeper S-shape. With below-threshold adoption the proportion of adopters might become larger. In turn, less time would

be needed to reach the 'critical mass', i.e. the size of the population to reach the social tipping point. Passing this social tipping point would imply that diffusion would really take off. Logically this makes sense; if more people in a social network have become activated, other inactive individuals will have more active neighbors who might convince them to become adopt as well. We might witness an exponential growth in diffusion. As a result, the diffusion curve would have a steeper S-Shape (figure 2 shows what this would look like). The final diffusion size would also be larger if the diffusion curve has a steeper S-Shape and the social tipping point is reached.

Additionally, the 'S-Shape' logit function was used in the linear threshold model to shape the probability of node i to adopt ($F(x_i)$) in a less deterministic way. Again the probability of node i to become active is no longer either 1 or 0, instead the probability is determined by the S-Shape logit function $F(x_i)=1/(1+\exp(-(x_i+q_i)/b))$. Figure 3 shows the diffusion curves for the vegetarian recipe with the deterministic and S-Shape linear threshold models. In the function, parameter b can be specified to control the shape of the probability function. With a larger b , the probability to adopt becomes smaller and vice versa. If for the nodes in the network the probability to adopt is smaller, the diffusion curve will most likely be less steep (flatter) and the final diffusion size would be lower.

Finally, how would the target nodes of seeding differ for the deterministic threshold model and the more realistic threshold models. The greedy linear threshold function identified the nodes to be seeded for the largest diffusion. The greedy threshold function bases the output on the centrality of the node (a large number of connections) and the highness of the threshold. Central nodes with a high threshold are ideally seeded for a large diffusion. In the non-deterministic threshold functions there is a certain randomness in the probability with which nodes is activated, so there would be more randomness in the diffusion process. On one iteration the seeding of a certain node i might randomly results in a larger diffusion, then in another iteration. The LT_Greedy function with the below-threshold activation and S-Shape logit in the LT function, outputs different nodes with the highest spread over multiple iterations. Thus, it would be more difficult to determine what node should be seeded for the largest diffusion and influence maximization.

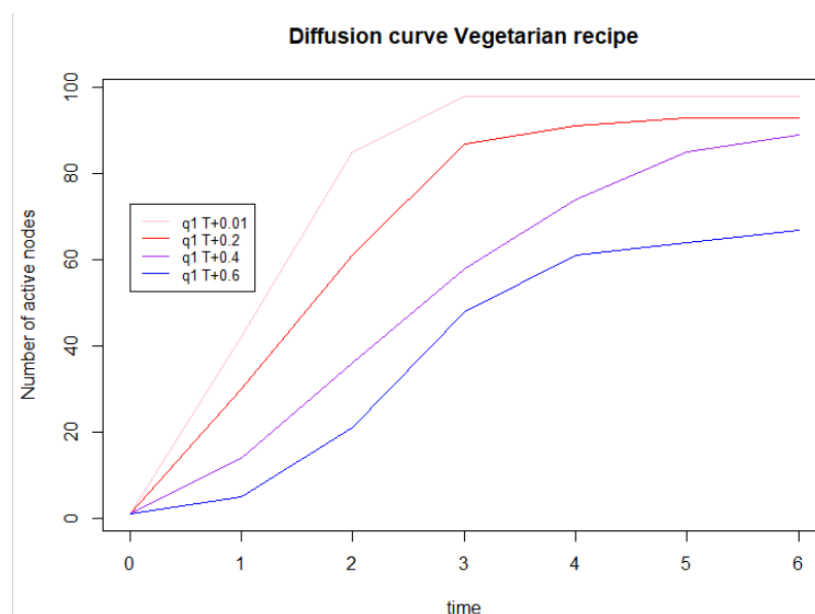


Figure 1. Different q1 thresholds

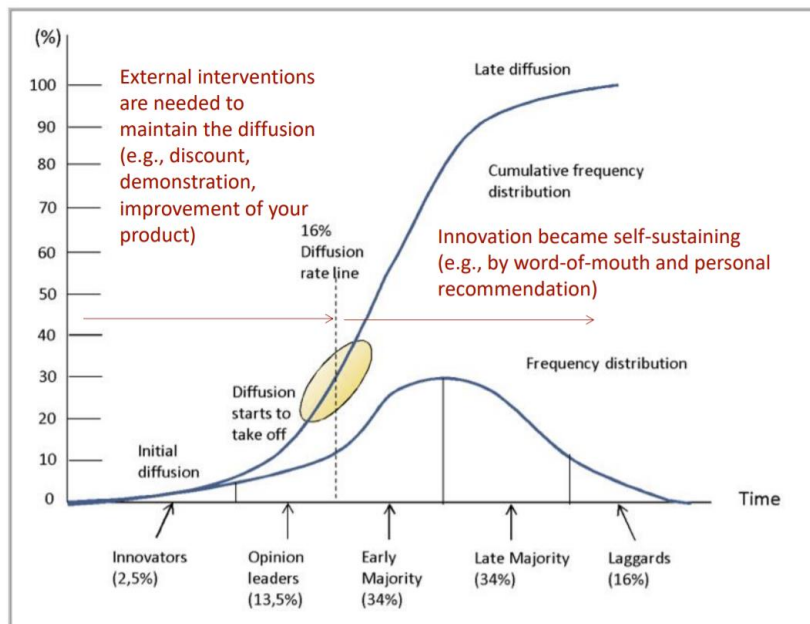


Figure 2. Diffusion curve shapes

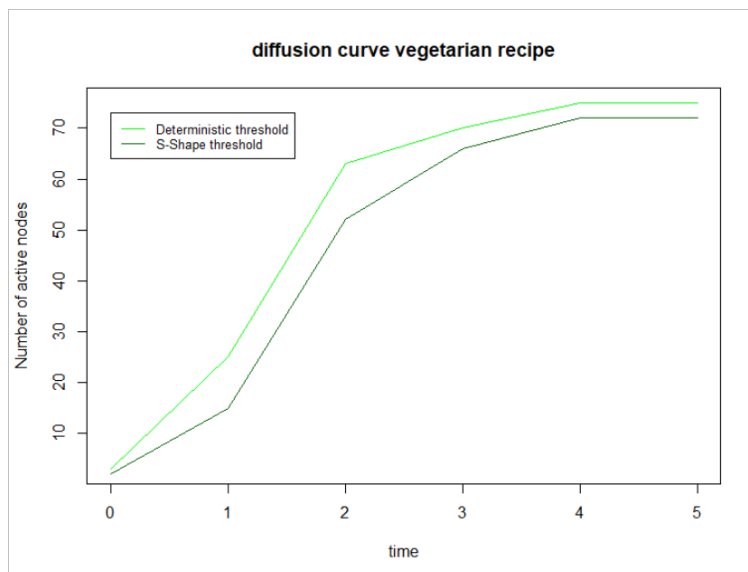


Figure 3. Diffusion curve vegetarian recipe deterministic and S-Shape threshold

```

LT <- function(g, activated, plot = TRUE) {
  ## Transform the graph to an undirected graph if it is undirected
  if (is_directed(g)) {
    g <- as.undirected(g)
    cat(paste0("The inserted (directed) graph is for simplicity transformed to a undirected graph.\n"))
  }
  if (is.null(E(g)$weight)) {
    stop("Specify the individual activation thresholds as V(g)$threshold")
  }
  if (plot) {
    l <- layout.fruchterman.reingold(g)
  }
  out <- list() ## Empty list for the output
  ## Set all nodes to inactive
  V(g)$active <- FALSE
  ## Except the activated nodes
  V(g)$active[activated] <- TRUE
  ## If changes occur, continue the while loop (set to true initially)
  changes <- TRUE

  n_active <- sum(V(g)$active) ## Number of active nodes
  ## Timepoint
  t <- 1
  while(changes) {
    changes <- FALSE ## Set changes to false
    ## And set the time parameter
    t <- t + 1
    if (plot) {
      V(g)$color <- ifelse(V(g)$active, "red", "lightblue")
      plot(g, layout = l)
    }

    ## For all nodes that are not yet active
    for (v in V(g)[!V(g)$active]) {
      xi <- (sum(neighbors(g, v)$active) / max(length(neighbors(g, v)), 1))
      q1 <- V(g)$threshold[v]
      q2 <- V(g)$threshold2[v]
      ## If the proportion of active neighbors is greater than your threshold, the node
      ## gets activated, also allow for a small probability of below threshold adoption
      if (xi > q1) { #xi>q1 --> 1
        V(g)$active[v] <- TRUE
      }
      if (q2 < xi & xi < q1) { # q2 < xi < q1 --> pi
        if (q2+((q1-q2)/2) > xi) {
          V(g)$active[v] <- FALSE
        }
        else {
          V(g)$active[v] <- TRUE
        }
      }
      else { ## xi<q2 --> 0
        V(g)$active[v] <- FALSE
      }
    }
    ## The number of active nodes at timepoint t
    n_active[t] <- sum(V(g)$active)
    ## If the number of active nodes is still increasing, continue the while loop
    if (n_active[t] > n_active[t-1]) {
      changes <- TRUE
    }
  }
  if (plot) {
    V(g)$color <- ifelse(V(g)$active, "red", "lightblue")
    plot(g, layout = l)
  }
  c(active = n_active)# Return the output
}

```

Code: LT function below-threshold adoption