Stockholms universitet

Some rumor spreading models on complete graphs and Erdős–Rényi graphs

Jan Alexandersson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2019:13
Matematisk statistik
Juni 2019

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2019:13**
http://www.math.su.se

Some rumor spreading models on complete graphs and Erdős–Rényi graphs

Jan Alexandersson*

June 2019

Abstract

In this thesis we study three rumor spreading models, the Daley–Kendall model, Maki–Thompson model and Daley–Kendall model with memory. The main area of use for rumor spreading models is naturally to describe how a rumor is spread in a population but can be used in other areas as well, for example the spread of viruses on the Internet. Rumor spreading is closely related to epidemics as the rumor can be seen as an infection. Due to the complexity of real world networks some simplified models for the networks are required and we choose to consider both homogeneously mixed populations as well as Erdős–Rényi networks for the rumor to be spread in. The rumor spreading models we study all divide the population into three subgroups; ignorants, spreaders and stiflers. An ignorant is an individual who is unaware of the rumor, a spreader is an individual who actively tries to spread the rumor and a stifler is an individual who is aware of the rumor but no longer tries to further spread the rumor.

The main problem to be answered in this thesis is the number of individuals who know about the rumor when there is no one left who is interested in further spreading the rumor. This number depends on the size of the population, the initial number of individuals who knows about the rumor, as well as the average number of friends the individuals have. We also examine when people first stop being interested in spreading the rumor further. We make use of deterministic models and stochastic models as well as simulations to study the rumor spreading models.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: janalexandersson97@gmail.com. Supervisor: Pieter Trapman and Abhishek Pal Majumder.

Acknowledgements

I would like to thank my supervisors Pieter Trapman and Abhishek Pal Majumder for their entusiastic encouragement and for their valuable advise and guidance through this thesis. I would also like to thank my brother Per Alexandersson for helpful assistance with the implementation of the Mathematica code used in this thesis.

Contents

1 Introduction

A rumor is information which is spread around by some interaction between individuals and the expression *information transmission* is often used to refer to the spreading of news, rumors or viruses on the Internet. Information transmission is also closely related to epidemics, where the rumor can be associated to an infection. In reality the spreading of a rumor is complex and every individual is unique, behaves in its own way and the structure of the social network for different populations can differ significantly. One can for example imagine that a rumor spreads differently in an office workplace population compared to a school population. To construct a model which takes every aspect of human behavior into consideration is virtually impossible.

In this thesis we study some of the most well-known models of rumor spreading, namely the Daley–Kendall model [DK65] and the Maki–Thompson model [Mak73]. We also introduce a third model which is a variant of the Daley–Kendall model where the individuals remember who they first heard the rumor from. This model will therefore go under the name Daley–Kendall model with memory. We will onwards abbreviate these models as the DK (Daley–Kendall), MK (Maki–Thompson) and the DKM (Daley–Kendall model with memory) models. The widely used way to abbreviate the Maki–Thompson model is MK even though MT might be more intiutive. Before specifying the models we need to introduce some definitions and assumptions which applies to all three models.

First of all we shall only consider closed populations; this means that there are no births, death or migrations occurring. Secondly, the population is divided into three subgroups; ignorants, spreaders and stiflers. An *ignorant* is an individual in the population who is unaware of the rumor, a *spreader* is an individual who actively spread the rumor and finally, a *stifler* is an individual who knows about the rumor but no longer is interested in spreading it further.

Furthermore, we assume that the population is homogeneously mixed, meaning that it is equally likely for all individuals to interact with each other. In other words you can say that everyone in the population are friends with everyone else and their level of friendships are equal. However, this assumption is relaxed when we in a later section apply our models on types of social networks represented by random graphs, so called Erdős–Rényi graphs. Rumor transmission only proceed by pair-wise contact between individuals meaning that simultaneous transmission in groups larger than two is not allowed.

1.1 Glossary

- *Informed* – If an individual is aware of the rumor then that individual is informed. In our case an individual is informed if belonging to state S or R.
- *Source* – If person A is the first to inform person B about the rumor then

A is the source to B.

- *Final size of a rumor* – The proportion of informed individuals in the population when there is no longer any spreading going on.
- *Stifling* – The event of an individual transitioning from state S to state R.

2 Models on rumor spreading

Since we only consider closed populations, the total number of spreader, stiflers and ignorants will remain constant throughout time. We denote ignorants with I, spreaders as S and stiflers as R. This notation should not be confused with the notations from the very well know Susceptible–Infectious–Removed (SIR) model used in epidemics. Furthermore we let the total size of the population be $n + a$ where n denotes the initial number of ignorants and a the initial number of spreaders. Consequently there are no initial stiflers at the beginning of the process.

2.1 Homogenously mixed networks as complete graphs

Before we describe how our models works we want to introduce graphs, which can be used to represent social networks. A graph $G = (V, E)$ consists of a set of *nodes*, V, representing the individuals in the social network and a set of *edges*, E, with subsets of node pairs, representing relationships between the nodes. A homogeneously mixed population is represented by what is called a *complete graph*, where there exist an edge between any pair of nodes in the graph. A property of the complete graph is that the total number of edges in a graph with N nodes is $N(N-1)/2$. The degree of a node v in the graph is the number of neighbors to v, or in other words the number of edges that is connected to v.

2.2 The Daley–Kendall model

As previously stated, the rumor is spreading through pair-wise contact between individuals. The outcome of these interactions is directed by a set of rules. In the case of a spreader interacting with an ignorant, the ignorant transitions into a spreader. In the case of a spreader interacting with a stifler the spreader also becomes a stifler. Lastly, if two spreaders interact then both will become stiflers. Any other interaction leaves the situation unchanged.

These rules can also be expressed briefly as

$$S + I \to 2S, \tag{1}$$

$$S + R \to 2R, \tag{2}$$

$$S + S \to 2R. \tag{3}$$

These interactions will occur continuously over time, however a different use of time will be utilized. In the interval between time point t and $t+1$ exactly one interaction which leads to a state transition, (1), (2) or (3), will occur. Any other interactions also happens over time but as they do not change the state of any individual then are not interesting to us, which justifies the unusual use of time. We call the elapsed time between t and $t+1$ an *epoch* following [Sud85].

An example of rumor spreading in the DK model is shown in Figure 1 below. In the example the initial conditions $n = 4$ and $a = 1$ are used.

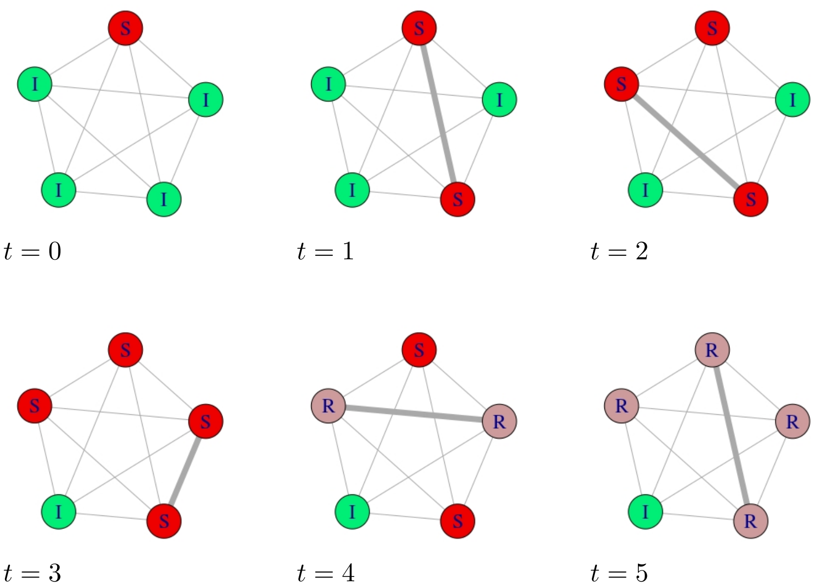


Figure 1: Example of a rumor progression using the Daley–Kendall model with a population of size 5. Green nodes represents ignorants, red represents spreaders and grey represents stiflers. The thick edge indicates interaction.

2.3 The Maki–Thompson model

The Maki–Thompson model (MK model) is a variant of the DK model, except that the rumor is spread via *directed* interaction. Consequently, if a spreader interact with another spreader or a stifler, only the spreader who initiate the interaction becomes a stifler. This is a reasonable modification to the DK model since if an individual attempt to spread to an already informed individual, the initiating individual will regard the rumor as well-known and therefore not continue to spread. However, the second individual will have even more reason to further spread the rumor as the rumor appears to still be interesting due to the spreading attempt from the initiating individual.

The rules for interaction for the MK model can now be expressed as

$$
\begin{aligned}
S + I &\to 2S, \\
S + R &\to 2R, \\
S + S &\to R + S \text{ (previously } R + R \text{ in the DK model).}
\end{aligned}
$$

2.4 Daley–Kendall model with memory

The model we are now going to consider takes into account a serious flaw present in the previous models. Both the DK model and the MK model do not take into account that the individuals can remember from who they first heard the rumor. A consequence of this is that a transitioned spreader can attempt to spread the rumor right back to it's source, which is unreasonable in human behavior. In this model an individual who gets informed will not try to spread the rumor back to the spreader who informed the individual.

The specific situation avoided is illustrated below in Figure 2.

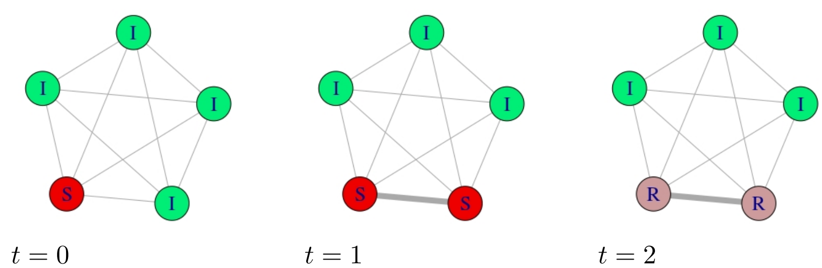


Figure 2: Example of a rumor progression using the Daley–Kendall model with a population of size 5. Green nodes represents ignorants, red represents spreaders and grey represents stiflers. The thick edge indicates interaction. This example illustrates a rumor ending after only two transitions.

The set of rules for the interactions for the DKM model is the same as for the DK model with one exception; an individual can no longer interact with their source. In terms of graph theory, we can say that after an ignorant becomes informed the edge between that individual and its source is removed forever.

An example of a rumor spreading using the Daley–Kendall model with memory is found in Figure 3.

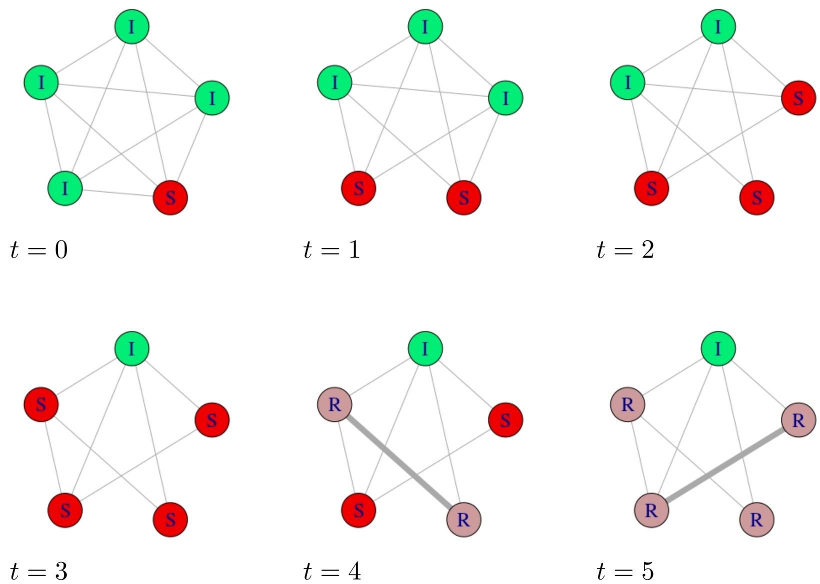


Figure 3: Example of a rumor progression using the Daley–Kendall model with memory with a population of size 5. Green nodes represent ignorants, red represent spreaders and grey represent stiflers. The thick or removed edge indicates interaction.

This model is however very similar to the DK model in case of large populations for complete graphs. The DKM model becomes more interesting for graphs with lower average degree, because the probability of a node trying to spread the rumor back to the source will be larger. We will later see larger differences between the two models when applying the models on Erdős–Rényi graphs, with low average degree.

3 Stochastic processes and deterministic approximation of the models on complete graphs

Stochastic models are often preferable over deterministic models, particularly when studying smaller communities [Bri10], yet deterministic models can still be sufficient for addressing the types of questions which the primary focus in this thesis, namely the final size of the rumor.

A natural question that comes up when talking about a rumor is how many people that in the end will hear the rumor. Consider a high school student who did something embarrassing at a party in the weekend, and assuming that high school students enjoy to gossip, everyone present at that party will try to spread what happened at school on Monday. That one student will naturally wonder how many people at the school will hear about it in the end. It is also of interest if the amount of people at the party influence the final size of the rumor, assuming everyone at the party is informed and interested in spreading the rumor on Monday. To translate this example back to our case a is the number of people at the party and n is the number of students at the school who did not go to the party.

3.1 Rumor spreading as stochastic processes

Our models can be described as stochastic processes. Let $I(t)$, $S(t)$ and $R(t)$ denote the number of ignorants, spreaders and stiflers at time t. Recall that $I(t) + S(t) + R(t) = n + a$, for all $t \geq 0$, and that we initially have $I(0) = n$, $S(0) = a$ and $R(0) = 0$.

Now we can introduce the three-dimensional process

$$\Big\{S(t), I(t), R(t)\Big\}_{t\geq 0}.$$

This process will be a Markov process for the DK model and the MK model as the process as future transitions in the process is independent of the past given that we know the present state of the process. In other words, the next state of the process is only dependent of the number of ignorants, spreaders and stiflers in the present and independent of the past.

3.1.1 Stochastic process for the Daley–Kendall model

For the DK model there are three possible interactions that changes the state of the process. These interactions are the pair-wise contacts which includes at least one spreader, which were presented in the description of the model, (1)-(3). Therefore, the interactions which not include any spreader are uninteresting, in the sense that they do not change the state of the process. Furthermore, we can view these interactions as non-occuring. As a result of our defined use of time there will always be exactly one process-changing event in each time epoch, t to $t + 1$. We now want to derive the probabilities for these interactions to occur, in a time epoch starting at time t. We do this by randomly selecting an edge from the edge set which includes at least one spreader and compute the probabilities that the edge representing one of the three possible interactions.

Considering a homogeneously mixed population, the number of ways we can choose an unique pair of nodes consisting of a spreader and an ignorant is $S(t) \cdot I(t)$, since every spreader has an edge connected with every ignorant. The same applies for the number of ways to choose an edge between a spreader and a stifler and we get $S(t) \cdot R(t)$ ways. The number of ways to choose a unique pair of two spreaders is the same as the number of edges in a complete graph with $S(t)$ nodes, which is $S(t)(S(t)-1)/2$. Using this knowledge we can now write down the transition probabilities for the DK model,

$$
\begin{aligned}
P\left[\begin{pmatrix} I(t+1) \\ S(t+1) \end{pmatrix} = \begin{pmatrix} I(t) \\ S(t) \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle| \begin{pmatrix} I(t) \\ S(t) \end{pmatrix} = \begin{pmatrix} i \\ s \end{pmatrix}\right] &= \frac{si}{si + sr + \frac{s(s-1)}{2}} \\
P\left[\begin{pmatrix} S(t+1) \\ R(t+1) \end{pmatrix} = \begin{pmatrix} S(t) \\ R(t) \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle| \begin{pmatrix} S(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} s \\ r \end{pmatrix}\right] &= \frac{sr}{si + sr + \frac{s(s-1)}{2}} \\
P\left[\begin{pmatrix} S(t+1) \\ R(t+1) \end{pmatrix} = \begin{pmatrix} S(t) \\ R(t) \end{pmatrix} + \begin{pmatrix} -2 \\ 2 \end{pmatrix} \middle| \begin{pmatrix} S(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} s \\ r \end{pmatrix}\right] &= \frac{s(s-1)/2}{si + sr + \frac{s(s-1)}{2}}.
\end{aligned}
$$

3.1.2 Stochastic process for the Maki–Thompson model

The MK model is described by $(S(t), I(t), R(t))_{t \in N}$, which is a $S := \{(x, y, z) : x + y + z = n + a\}$ valued discrete time stochastic process, with transition probabilities given by

$$
\begin{aligned}
P\left[\begin{pmatrix} I(t+1) \\ S(t+1) \end{pmatrix} = \begin{pmatrix} I(t) \\ S(t) \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle| \begin{pmatrix} I(t) \\ S(t) \end{pmatrix} = \begin{pmatrix} i \\ s \end{pmatrix}\right] &= \frac{i}{n + a - 1} \\
P\left[\begin{pmatrix} S(t+1) \\ R(t+1) \end{pmatrix} = \begin{pmatrix} S(t) \\ R(t) \end{pmatrix} + \begin{pmatrix} -1 \\ 1 \end{pmatrix} \middle| \begin{pmatrix} S(t) \\ R(t) \end{pmatrix} = \begin{pmatrix} s \\ r \end{pmatrix}\right] &= 1 - \frac{i}{n + a - 1}.
\end{aligned}
$$

Finding the transition probabilities for the MK-model is easier than for the DK model. This is a consequence of the interaction between a spreader and a stifler, $S + R$, and the interaction between two spreaders, $S + S$, will result in the same transition in the process. The probability of the transition corresponding to the event $S + I$, is $I(t)/(n + a - 1)$ at time $t + 1$. This is realized by first choosing a spreader which will initiate a contact with an ignorant and then choosing the ignorant, which gives the probability $I(t)/(n + a - 1)$, the fraction of ignorants excluding the first chosen spreader as individuals can not contact themselves. The probability for the second possible transition in the process (from interaction $S + S$ or $S + R$) follows from the fact that it is the complement to the first.

3.1.3 Daley–Kendall with memory process

Deriving any general transition probabilities for this model is hard as they can not be expressed solely by the number of ignorants, spreaders and stiflers as the transition probabilities depend on how many edges that have been removed and also the states of the previously connected nodes to the removed edges. Therefore, any expressions for the transition probabilities is expected to be complicated. This process is not a Markov process when considering only the number of ignorants, spreaders and stiflers. In Figure 21 and Figure 22 in Appendix, we can see two different examples of a rumor progression using the DKM model where both examples end up with the same number of ignorants, spreader and stiflers after the same number of transitions. Even the number of edges and the number of each type of interaction is the same for both examples, but we end up having different transition probabilities for the next interaction. These transition probabilities are shown in the figures. Recall that when deriving the transition probabilities we do not consider interactions among ignorants and stiflers as the do not change any states of the individuals.

3.2 Deterministic models on rumor spreading

3.2.1 Deterministic approximation of the DK model

Recall that the initial number of individual of each state are $S(0) = a$, $I(0) = n$ and $R(0) = 0$. We also know that the end state of the process, when $t \to \infty$, is $\{n - \zeta, 0, a + \zeta\}$, where ζ is the number of the initial ignorants who eventually becomes informed. Furthermore, $Z = \zeta/n$ is the proportion of initial ignorants that eventually becomes informed. We can then write the end state of the process as $\{n(1 - Z), 0, nZ + a\}$. Naturally we also get

$$\tilde{Z} = \frac{nZ + a}{n + a},$$

where $\tilde{Z}$ is defined as the final size of the rumor, the proportion of informed in the population when the process has reached its end state.

By using the transition probabilities for the DK model we can get the deterministic approximation for the corresponding process. Using the same method as in [Wat87] we are able to derive the final size of the rumor as a function of n and a. The deterministic approximation is represented by what is called the mean-field equations, which is a system of differential equations presented below

$$\begin{aligned}
\frac{di}{dt} &= -m^{-1} is \\
\frac{ds}{dt} &= m^{-1}\big[is - (\overbrace{n + a - i - s}^{r})s - 2 \cdot \frac{1}{2}s(s-1)\big] &&= m^{-1}(2i - n - a + 1)s \\
\frac{dr}{dt} &= m^{-1}\big[(n + a - i - s)s + 2 \cdot \frac{1}{2}s(s-1)\big] &&= m^{-1}(n + a - 1 - i)s.
\end{aligned}$$

For ease of notation we suppressed the parameter t and defined m as $m(t) = i(t)s(t) + r(t)s(t) + s(t)(s(t)-1)/2$, which is the total sample space of edges representing a process-changing transition.

Now consider the function $k(t) = -i(t)+r(t)-(n+a-1)\ln[i(t)]$. The significant property of this function is that the derivative is zero.

$$\begin{aligned}
k' &= -i' + r' - (n+a-1)\frac{i'}{i} \\
&= m^{-1}is + m^{-1}(n+a-1-i)s - (n+a-1)\frac{m^{-1}is}{i} \\
&= m^{-1}(n+a-1)s - (n+a-1)m^{-1}s = \\
&= 0.
\end{aligned}$$

Note that the function is constant through time as a consequence of the derivative being 0 for all t, and is called a constant of motion. To derive the value of the constant we use our known initial values $i(0) = n$, $s(0) = a$ and $r(0) = 0$, and get $k(0) = -n - (n+a-1)\ln(n)$, which in turn gives that

$$-i(t) + r(t) - (n+a-1)\ln[i(t)] = -n - (n+a-1)\ln(n),$$

which holds for all $t \geq 0$, particularly also when $t \to \infty$. Combining this with our knowledge about the end state of the process, $\big(n(1-z), 0, nz+a\big)$, we get that

$$\ln(1-z) + \frac{2nz+a}{n+a-1} = 0. \tag{4}$$

Now define z^*, a function of n and a, as the positive solution to equation 4, with regards to z. We can not explicitly write z^* in terms of n and a, so the solution has to be derived numerically. Now recall that z^* is the deterministic approximation of the proportion of the initial number of ignorants who eventually get informed. Furthermore, it follows that the final size of the rumor, denoted $\tilde{z}$, is

$$\tilde{z} = \frac{nz^* + a}{n+a}. \tag{5}$$

3.2.2 Deterministic approximation of the MK model

Close to analougous we can derive the same result for the MK model. We get the following mean-field equations, using the transition probabilities corresponding to the MK model,

$$
\begin{aligned}
\frac{di(t)}{dt} &= -\frac{i}{n+a-1} \\
\frac{ds(t)}{dt} &= 2\frac{i}{n+a-1} - 1 \\
\frac{dr(t)}{dt} &= 1 - \frac{i}{n+a-1}.
\end{aligned}
$$

Similar for the DK model we can express a constant of motion $k(t) = -i(t) + r(t) + (n + a - 1)\ln[i(t)]$, with only the sign for the last term changed compared to the function for the DK model. We show that it is a constant of motion by computing its derivative

$$
k'(t) = -\big(-\frac{i}{n+a-1}\big) + \big(1 - \frac{i}{n+a-1}\big) + (n+a-1)\frac{-\frac{i}{n+a-1}}{i} = 0.
$$

The rest of the calculations follows in like manner as for the DK model and the resultturns out to be identical as we derive the same equation, $\ln(1-z)+\frac{2nz+a}{n+a-1} = 0$ also for the MK model.

3.2.3 Deterministic approximation of the DK model with memory

As we do not know the transition probabilities for this model the deterministic approximation for this model can not be derived using the same method as for the DK and MK model. However, if we let $n \to \infty$ this model will behave as the regular DK model. When we let $n \to \infty$ the probability of a spreader spreading the rumor back to its own source or spreading the rumor twice to the same person, converge to 0 in the regular DK model. As $n \to \infty$ the degree of a node in the graph also tends to infinity and thus the possible interactions for that node tend to infinity. As a consequence the probability for a forbidden edge to be the next interaction converge to 0, where a forbidden edge is an edge which would be removed using the DK model. The conclusion is that for large n the DK model and the DKM model behaves approximately the same.

3.2.4 Impact of the initial number of spreaders

We are now able to answer the question about the impact of the party size in our example. By numerically solving equation 4 for different values of a we surprisingly notice in Figure 4 that z^* is a *decreasing* function by a. The interpretation is that a bigger party will result in a smaller fraction of ignorants will be informed when the spreading begin. This is a consequence of initial spreaders from the party will talk about the rumor with each other and become bored of spreading it further and transition into stiflers. The rate of stifling will be high at the start of the spreading.

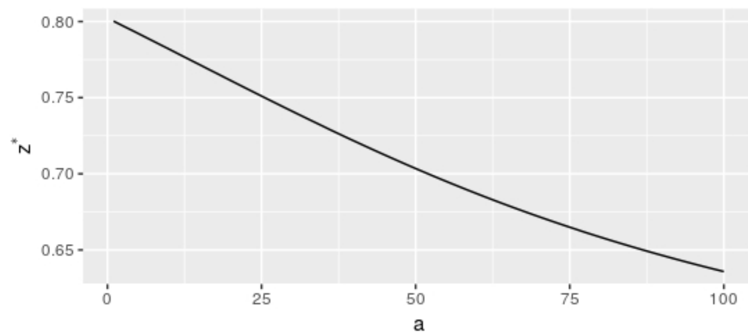


Figure 4: The deterministic approximation of the proportion of *ignorants* that eventually becomes informed, for a population of size 100 and a initial spreaders.

Notice that by increasing a we also decrease n as we let the population size stay constant.

Now if we instead consider equation 5 we find that $\tilde{z}$ instead is *increasing* with a, as we see in Figure 5.

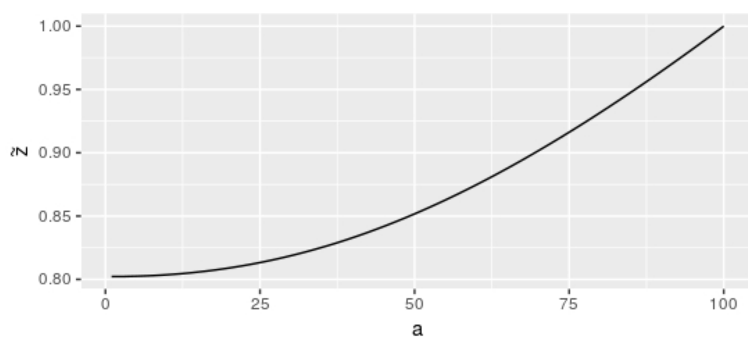


Figure 5: The deterministic approximation of the final size of a rumor, $\tilde{z}$, for a population of size 100 and a initial spreaders.

The interpretation is now that the total proportion of the school that will be informed in the end will increase with a and moreover that the one student probably wishes that the party was as small as possible.

3.2.5 Deterministic results for large populations

Now if we instead let a be fixed and increase n we see in Figure 6 that $\tilde{z}$ is a decreasing function which also seems to converge. It is reasonable for $\tilde{z}$ to be decreasing due to there being more ignorants for the rumor to reach. A higher amount of initial ignorants will thus lead to a smaller proportion that will be informed in the end with the whole population taken into account.

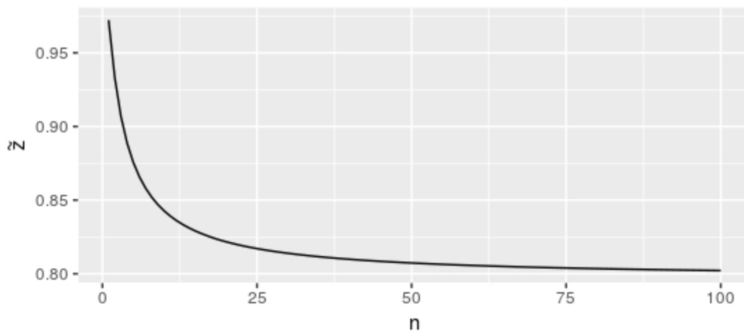


Figure 6: The deterministic approximation of the final size of a rumor, $\tilde{z}$, for a population with n initial ignorants and $a = 1$ initial spreaders.

Furthermore we can examine our intuition of convergence by letting $n \to \infty$. If we apply this to equation 4 we get

$$\begin{aligned}
&\lim_{n\to\infty}\left[\ln(1-z) + \frac{2nz + a}{n+a-1}\right] \\
&= \ln(1-z) + 2z \lim_{n\to\infty} \frac{n}{n+a-1} + \lim_{n\to\infty} \frac{a}{n+a-1} \\
&= \ln(1-z) + 2z = 0.
\end{aligned}$$

By numerically solving this equation we get the positive root $z^* \approx 0.797$, assuming $a \geq 1$. If we would allow a to be 0 then we naturally get the solution $z^* = 0$ because no one in the whole population initially knows about the rumor. Furthermore, we get that also $\tilde{z} \to z^*$ when $n \to \infty$ from

$$\lim_{n\to\infty} \frac{nz^* + a}{n+a} = \lim_{n\to\infty} z^* \cdot \lim_{n\to\infty} \frac{n}{n+a} + \lim_{n\to\infty} \frac{a}{n+a} = z^* \cdot 1 \approx 0.797,$$

using the property of limits which says that the limit of two functions is the product of their limits, if their limits exists [Rud76, p.85]. We observe when $n \to \infty$ that $\tilde{z}$ is independent of a, hence the initial number of spreaders is irrelevant when the number of initial ignorants grows to infinity.

3.3 Comparison of the models

Previous section show that there is no difference in the deterministic approximation regarding the DK and MK model although there clearly is a difference in the how the two models work. Simulations suggest that the difference lies in the distribution of $\tilde{Z}$.

Figure 7 suggest the variance of $\tilde{Z}$ is slightly larger for the DK model compared to the MK model and we can also observe in Figure 8 that the amount of events required to reach the end state also differ between the two models. The explanation for this may be due to the fact that in the MK model there is no possibility to increase the number of stiflers with more than one in one time epoch while in the DK model there is a possibility that the number of stiflers can increase with two in one time epoch. Therefore the rumor is more likely to die out in fever time epochs as the rate of stifling is allowed to be higher for the DK model.

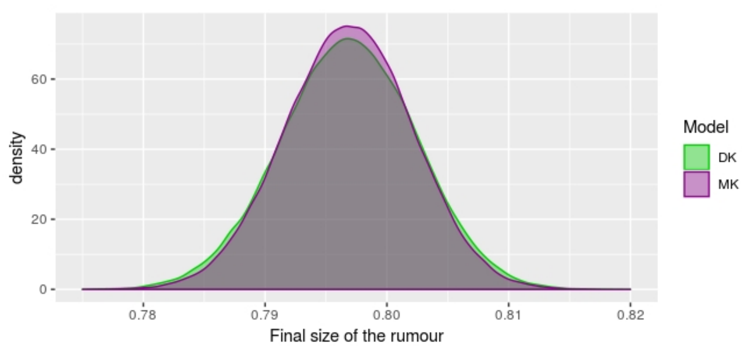


Figure 7: Simulated distribution of the final size of the rumor considering the Daley–Kendall model and the Maki–Thompson model for populations of size 10000 using 100000 simulations.

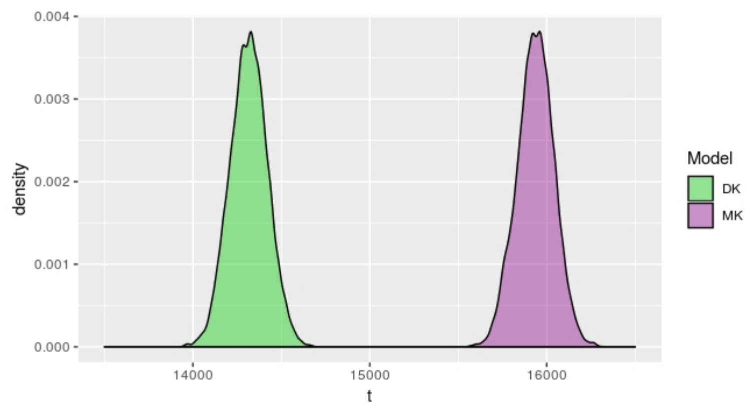


Figure 8: Simulated distribution of the number of time epochs considering the Daley–Kendall model and the Maki–Thompson model for populations of size 10000 using 100000 simulations.

Observing Figure 7 again we notice the that the distributions for the final size of the rumor, considering the DK and MK model, look normally distributed and it was showed in [Wat87] that the distributions is asymptotically normal with mean approximately 0.7968. The variance for the MK model is approximately $0.2728n^{-1}$ compared to $0.3017n^{-1}$ for the DK model. Thus we can confirm that our observation from the simulations were correct, that the variance is in fact larger for the DK-model.

The difference in the stifling mechanic between the model can also be seen in Figure 9 where we notice that the amount of spreaders decrease more rapid after the peak for the DK model compared to the MK model.

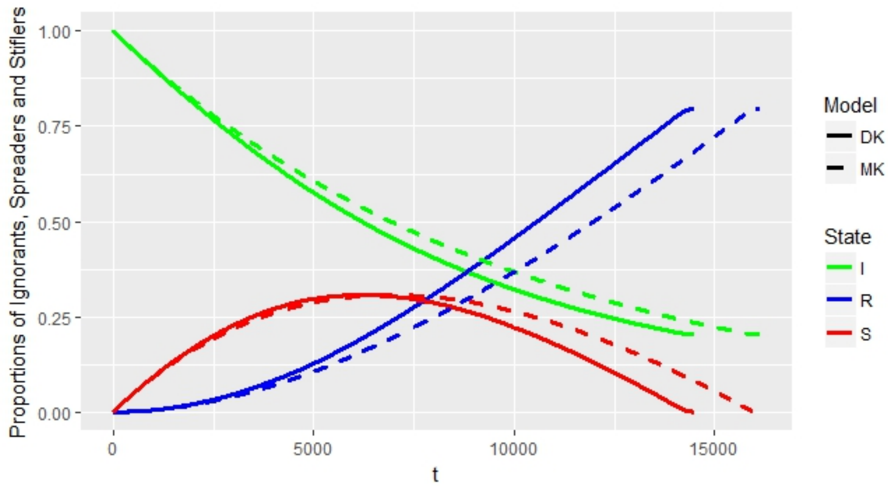


Figure 9: Simulated proportion of ignorants, spreaders and stiflers over time by taking the mean of 100 simulations considering a population of size 10000 with one initial spreader. The different line types represent the models Daley–Kendall (solid) and Maki–Thompson (dashed).

This is also a consequence of the interaction of two spreaders transitions both of them to stiflers for the DK model compared to only one for the MK model.

As previously said the DKM model is expected to behave the same as the DK model for large n, and we can already see that for $n = 100$ in Figure 10 that the distributions are very similar.

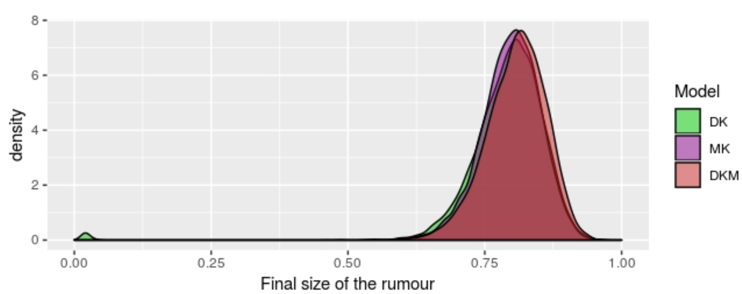


Figure 10: Simulated distribution of the final size of the rumor for the Daley–Kendall model, the Maki–Thompson model and the Daley–Kendall model with memory for a population of size 100 with one initial spreader. Each of the three models have been simulated 10000 times.

We can however expect the mean to be slightly larger than for the DK model as the removal of the edge lowers the possibility of stifling compared to the DK model and we should therefore expect a larger proportion of the population to be informed. In Figure 10 we can also see that the small density right after 0 for the DK model is not present for the DKM model. This is because that small density is outcomes where the rumor dies after only two interactions, as described by Figure 2. The probability of such an outcome using the DK model is fairly large at

$$\frac{1}{\underbrace{2(n+1-2)}_{\#\text{ of S-I edges at } t=1}+1},$$

assuming initial value $a = 1$, and is therefore expected once in every 199 simulations for $n = 100$.

However, for smaller n there are larger differences between the models. This is expected as the probability for an interaction between an individual and its source is larger for smaller populations and the rate of stifling will be significantly lower for the DKM model.

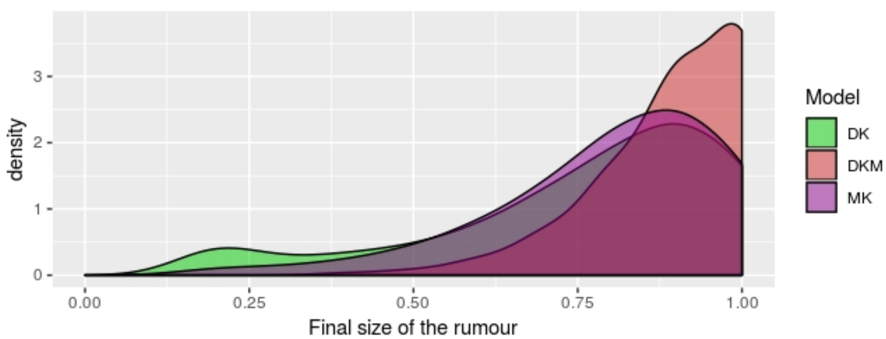


Figure 11: Simulated distribution of the final size of the rumor for the Daley–Kendall model, the Maki–Thompson model and the Daley–Kendall model with memory for a population of size 10 with one initial spreader. Each of the three models have been simulated 10000 times.

This is observable in Figure 11 where the population consist of only 10 individuals and we can see the final size of the rumor being closer to 1 using the DKM model.

4 Applying the models on Erdős–Rényi networks

While a homogeneously mixed network is easy to deal with it is not really realistic that everyone is friends with everyone in a population. Therefore we want to consider some other type of network. While Erdős–Rényi networks may not representative of real world human networks they are however more realistic than homogeneously mixed networks as we allow for individuals to have different number of friends and add some randomness.

4.1 Erdős–Rényi graphs

The Erdős–Rényi-graph [Dur06], denoted $G(n,p)$, is a random graph with n nodes. For each pair of nodes there is a probability p, independent from every other edge, that there is an edge connecting them. An edge is therefore present according to a Bernoulli variable with probability of success p. The degree distribution for a chosen node is therefore a sum of independent Bernoulli variables which gives the binomial distribution

$$P(k) = \binom{n-1}{k} p^k (1-p)^{n-1-k},$$

where $P(k)$ is the probability for the chosen node to have degree k. We have $n-1$ as a consequence of nodes not being allowed to connect to themselves.

Furthermore, if we let $n \to \infty$ and set $p = \frac{\lambda}{n}$ then $P(k)$ will converge to the Poisson distribution

$$P(k) \to e^{-\lambda} \frac{\lambda^k}{k!}.$$

The expected average degree for a node in the graph where $p = \frac{\lambda}{n}$ therefore always λ for all n.

4.2 Early stages of the rumor

During the early stages of the rumor spread the large majority of the individuals will be ignorants. Therefore in the early stages there will be almost only transitions where ignorants become spreaders due to low number of spreaders making the probability low for a stifling event to occur. It is then natural to think about when the first stifling event will occur and how many spreaders present at that point.

We do this by examining the discrete time process when we do not allow for stifling events to take place, meaning that in every time step the will be a contact between a spreader and an ignorant. However, we can still consider the probability for a stifling event to occur at time point t if we were to allow it. Now instead of keeping track of the numbers of individuals in each of the three states we now want to keep track of the numbers of the different edges. Let $X(t)$ denote the number of spreader-ignorant(SI) edges, $Y(t)$ the spreader-spreader(SS) edges and for completion $Z(t)$ the ignorant-ignorant edges, at time point t. The last type of edge is however not of interest as we are not interested in contacts among ignorants. Furthermore, we do not have a known graph at the start of the process but instead think of exploring the graph as the spreading goes on. So at the start we only know the initial spreader and it's neighbors and after spreading the rumor to one of the neighbors we reveal the neighbors of that node, and so on. We can think of it as building the graph as the spreading goes on.

Using the DK or the MK model where we do not remove edges after spreading there will be at least t SS-edges at time point t because an edges of that type will always be created in the event of a contact between a spreader and an ignorant, which we say occur every time step. Moreover, there will also be a random number of additional SS-edges for edges not primarily created due to a spreader-ignorant contact. Using the property that edges is present independently with probability $\frac{\lambda}{n+1}$ we get that the number of additional SS-edges, to the certain number of SS-edges t, is $\mathrm{Bin}\big(\frac{(t+1)t}{2} - t, \frac{\lambda}{n+1}\big)$ distributed. At time point t we have $t+1$ spreaders and there is $\frac{(t+1)t}{2}$ possible edges between them, but we also know that t of those edges already exists. Similarly we also know that the number of SI-edges is $\mathrm{Bin}\big((t+1)(n-t), \frac{\lambda}{n+1}\big)$ distributed, as we have $t+1$ spreaders and $n-t$ ignorants, at time point t, which in turn gives $(t+1)(n-t)$ possible SI-edges, each present with the same probability. To sum it up we have that

$$
\begin{aligned}
X(t) &\sim \mathrm{Bin}\left((1+t)(n-t), \frac{\lambda}{n+1}\right)\\
Y(t) &\sim \mathrm{Bin}\left(\frac{(t+1)t}{2} - t, \frac{\lambda}{n+1}\right) + t\\
Z(t) &\sim \mathrm{Bin}\left(\frac{(n-t)(n-t-1)}{2}, \frac{\lambda}{n+1}\right).
\end{aligned}
$$

Furthermore, the probability for a stifling event to occur at time point $t+1$, given that no previous stifling event has occurred, is $P(t) = \frac{Y(t)}{X(t)+Y(t)}$. These random variables are also independent at a given time point as the random variables are sums of different sets of independent Bernoulli variables. For the DKM model, where we do not allow individuals to spread the rumor back to their source, we can just remove the certain SS-edges which is created when spreaders contact ignorants and instead get $Y(t) \sim \mathrm{Bin}\big(\frac{(t+1)t}{2} - t, \frac{\lambda}{n+1}\big)$.

While $E[P(t)]$ is not necessarily equal to $\frac{E[Y(t)]}{E[X(t)]+E[Y(t)]}$, simulations show that it can be used as a good approximation for Erdős–Rényi networks with high average degree. The average of 100 simulations compared with $\frac{E[Y(t)]}{E[X(t)]+E[Y(t)]}$ can be seen in Figure 12, and is the basis for why our approximation is appropriate.

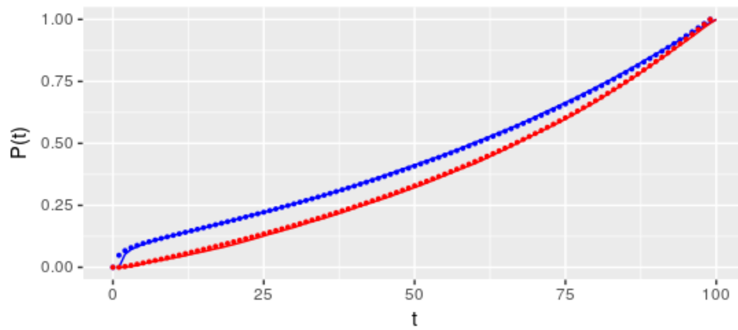


Figure 12: Average $P(t)$ of 100 simulations (solid line) and $\frac{E[Y(t)]}{E[X(t)]+E[Y(t)]}$ (dotted line) for Erdős–Rényi networks with average degree 10, considering a population of size 100 with one initial spreader. Blue color represents spreading using DK and MK model and red represents the DKM model.

For the DK and MK model we have that

$$\frac{E[Y(t)]}{E[X(t)]+E[Y(t)]} = \frac{\left(\frac{(t+1)t}{2} - t\right)\frac{\lambda}{n+1} + t}{\left(\frac{(t+1)t}{2} - t\right)\frac{\lambda}{n+1} + t + (1+t)(n-t)\frac{\lambda}{n+1}}$$

and for the DKM model we get

$$\frac{E[Y(t)]}{E[X(t)]+E[Y(t)]} = \frac{\frac{(t+1)t}{2} - t}{\frac{(t+1)t}{2} - t + (1+t)(n-t)}.$$

Interestingly we can see that for the DKM model that the approximation of $E[P(t)]$ does not depend on the average degree λ and we will in a later section see that the final size of the rumor is nearly constant with regards to the average degree (see Figure 20). This also means that the expected time until the first stifling event is the same for networks with different average degree.

Furthermore, the probability of no stifling event occurring up until time t is $\tilde{P}(t) = \prod_{j=0}^{t}(1-P_j)$. This probability of significant interest as we want to know how many spreaders are present before the first stifling event and $\tilde{P}_{t-1} \cdot P(t)$ tells us the probability for the first stifling event to occur at time t, in which time point we have $t+1$ spreaders. In Figure 13 and Figure 14 we can see simulated $P(t)$ and $\tilde{P}_t$ for networks with average degree 3 and 10.

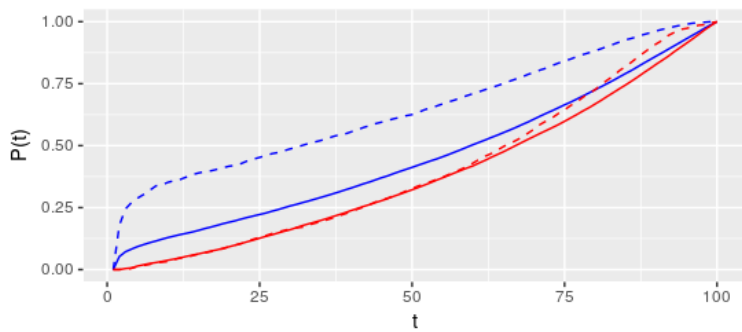


Figure 13: Average $P(t)$ of 100 simulations for Erdős–Rényi networks with average degree 3 (dashed line) and 10 (solid line), considering a population of size 100 with one initial spreader. Blue color represents spreading using DK and MK model and red represents the DKM model.

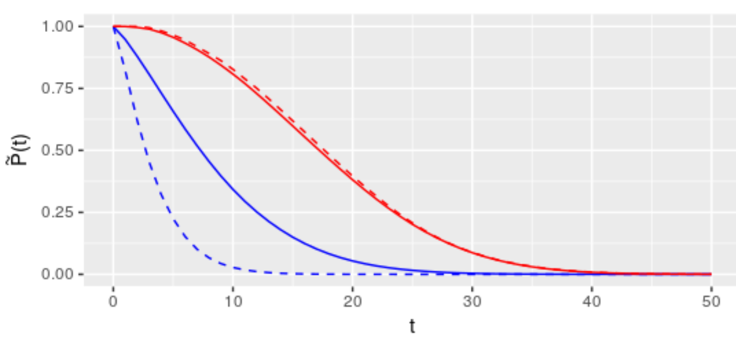


Figure 14: Average $\tilde{P}(t)$ of 100 simulations for Erdős–Rényi networks with average degree 3 (dashed line) and 10 (solid line), considering a population of size 100 with one initial spreader. Blue color represents spreading using DK and MK model and red represents the DKM model.

For the DKM model we can also derive an approximated distribution for $P(t)$. First we rewrite $P(t)$ as

$$P(t) = \frac{Y(t)}{X(t)+Y(t)} = \frac{1}{\frac{X(t)}{Y(t)}+1} = \frac{1}{\frac{X(t)/n_{x,t}}{Y(t)/n_{y,t}} \cdot \frac{n_{y,t}}{n_{x,t}}+1} = \frac{1}{T(t) \cdot \frac{n_{y,t}}{n_{x,t}}+1},$$

where $n_{x,t} = (t+1)(n-t)$ and $n_{y,t} = \frac{(t+1)t}{2} - t$. Using Lemma 1 in Appendix have that $\log(T(t))$ is approximately normally distributed with mean 0, as $\log(\frac{p_x}{p_y}) = 0$ because $p_x = p_y = \frac{\lambda}{n+1}$ in our case, and variance $(\frac{1}{n_{x,t}} + \frac{1}{n_{y,t}})(\frac{n+1}{\lambda} - 1)$. Lemma 1 requires that $X(t)$ and $Y(t)$ are independent of each other, which in our case is satisfied. Furthermore, $\widetilde{T}(t) = \frac{n_{y,t}}{n_{x,t}} T(t)$ is approximately log-normal with $\mu = \log(\frac{n_{y,t}}{n_{x,t}})$ and $\sigma^2 = (\frac{1}{n_{x,t}} + \frac{1}{n_{y,t}})(\frac{n+1}{\lambda} - 1)$. We now have that $P(t) = \frac{1}{\widetilde{T}(t)+1}$. Using Lemma 2 in Appendix we get that $P(t)$ is approximately logit-normal with parameters

$\mu = -\log\left(\frac{\frac{(t+1)t}{2} - t}{(t+1)(n-t)}\right)$ and $\sigma^2 = \left(\frac{1}{(t+1)(n-t)} + \frac{1}{\frac{(t+1)t}{2} - t}\right) \cdot \left(\frac{n+1}{\lambda} - 1\right)$.

In Figure 15 we compare $P(t)$ from simulations with the derived logit-normal distribution and observe that it seems to be an appropriate approximation. As it does not exists any analytical expression for the expected value of a logit-normal distribution we instead simulate a large number of values from the distribution and take the mean, for each t.

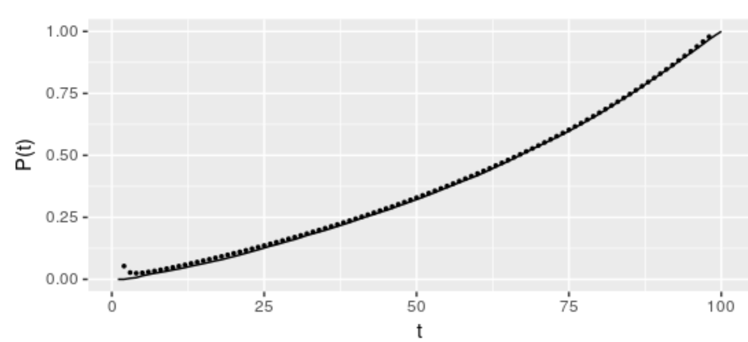


Figure 15: Average $P(t)$ of 100 simulations for Erdős–Rényi networks with average degree 10 (solid line) and the average of 100000 values from the logit-normal distribution (dotted line) for each t, considering a population of size 100 with one initial spreader using the DKM model.

4.3 Mean-field equations for the MK model

Rumor spreading on Erdős–Rényi networks has already been studied in the past [NMBM07], and mean-field equations were derived for the MK model.

Now let $I_k(t)$, $S_k(t)$ and $R_k(t)$ be the fraction of ignorants, spreaders and stifler nodes in the network with degree k at time t. The condition $I_k(t)+S_k(t)+R_k(t) = 1$ is satisfied. The mean-field equation can now be presented as

$$\begin{aligned}
\frac{dI_k(t)}{dt} &= -k\gamma I_k(t)\sum_{l=0}^{n} P(l|k)S_l(t),\\
\frac{dS_k(t)}{dt} &= k\gamma I_k(t)\sum_{l=0}^{n} P(l|k)S_l(t) - k\sigma S_k(t)\sum_{l=0}^{n}\Big((S_l(t)+R_l(t))P(l|k)\Big),\\
\frac{dR_k(t)}{dt} &= k\sigma S_k(t)\sum_{l=0}^{n}\Big((S_l(t)+R_l(t))P(l|k)\Big),
\end{aligned}$$

where γ is the rate of which an ignorant becomes a spreader when interacting with a spreader. Additionally, σ is the rate of which a spreader becomes a stifler when interacting with a stifler or another spreader. Remember that for the MK model only the initiating spreader transitions to a stifler in a contact consisting of two spreaders. Now we are not using our previously defined use of time and instead use a more standard use of time where t is continuous and the spreading proceed using rates. Previously we did not define any rates of which transitions occur which equivalent of setting $\gamma = 1$ and $\sigma = 1$ in this case. This is also the case we are going to study further so we can make comparisons to previous analysis.

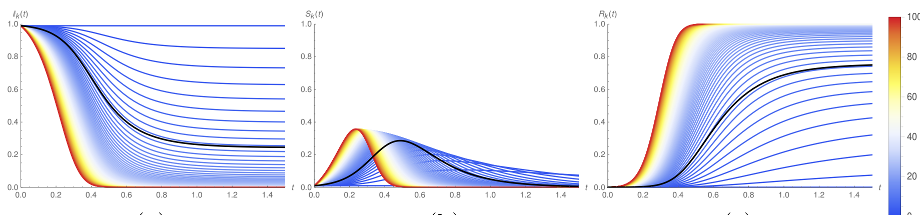
In these equations $P(l|k)$ is the degree-degree correlation function, the conditional probability that a random chosen node with l edges is connected to a node with degree k. Therefore, this function can be written $P(l|k) = \frac{lP(l)}{\lambda}$, where $P(l)$ is the degree distribution of the network and $\lambda = \sum_k kP(k)$ is the average degree of nodes in the network. Notably, the conditional probability $P(l|k)$ is proportional to $lP(l)$ as a consequence that edges will be biased to be connected to nodes of high degree. In words, the degree-degree correlations tell us whether the neighbors of an individual with, for example, high degree tend to have high or low degrees.

However, when deriving these mean-field equations some approximations were made. Consider the probability that a node with l links is in the spreader state given that it is connected to an ignorant node with degree k. This probability is approximated by the density of spreader nodes in connectivity class k (nodes which have the degree k). Using this approximation we ignore any correlations between the states of neighboring nodes. Although we do not take these correlations into account these equations may still be useful if the impact of the absence of these correlations is small.

Notably, these equations only take the network into account via the degree-degree correlation function $P(l|k)$ which implies that we do not actually have to generate any network for our analysis. Consequently this makes it easy to study rumor spreading on other networks than the Erdős–Rényi if we have the quantity $P(l|k)$. For the Erdős–Rényi we have the degree-degree correlation function

$$P(l|k) = \frac{l\binom{n-1}{l}p^l(1-p)^{n-1-l}}{\lambda}.$$

These non-linear differential equations can be solved numerically using a computer. Solutions to the equations are illustrated in Figure 16 for a population of size 100 individuals with the initial conditions $I_k(0) = 99/100$, $S_k(0) = 1/100$ and $R_k(0) = 0$. Note here that k cannot be 100 as a node can not connect to itself with an edge. The figure is based on a Erdős–Rényi network with $\lambda = 10$.



(a) (b) (c)

Figure 16: Time evolution for the proportions of ignorants (a), spreaders (b) and stiflers (c) considering the MK model for Erdős–Rényi network with a population of 100 individuals and $\lambda = 10$. Each curve represents the probability that a node of degree k is in one of the three states, where the color represents the degree k. The initial conditions are $I(0) = 0.99$, $S(0) = 0.01$ and $R(0) = 0$.

The black line in the figure represents the weighted average functions $\tilde{I}(t) = \sum_k P(k)I_k(t)$, $\tilde{S}(t) = \sum_k P(k)S_k(t)$ and $\tilde{R}(t) = \sum_k P(k)R_k(t)$ in (a), (b) and (c) respectively. This is the line which best represents the overall state of the proportions of ignorants, spreaders and stiflers at time t. We can see that ignorant nodes with low degree tends to stay ignorant when t grows and converge to a higher value in proportion than nodes with a relative high degree.

Remarkably we can see that the final size of the rumor, which can be approximated by the black line in (c), seems to converge to a value nearby the deterministic approximation we got for a complete graph, which is approximately 0.802 using equation 5, already for $\lambda = 10$. The final size of the rumor for a network with for example $\lambda = 2$ will however show to be significantly lower. For a network with $\lambda = 2$ we get the final size of the rumor $\tilde{R}(t) \approx 0.592$ when $t \to \infty$.

In Figure 17 we can see that the converged value of $\tilde{R}(t)$ for large t is increasing for networks with higher average degree which in other words means that more edges in the network implies a higher value for the final size of the rumor.

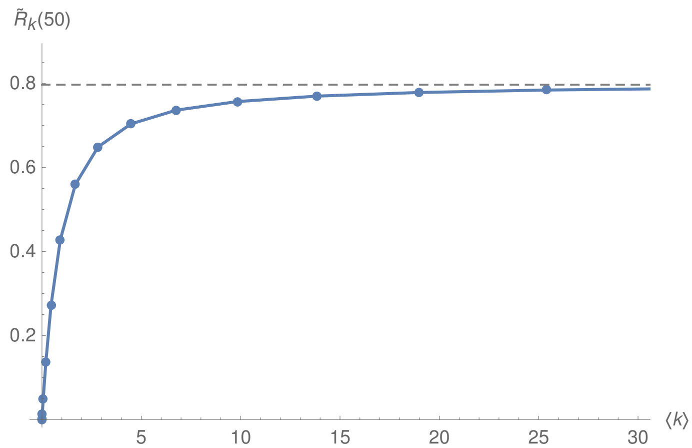


Figure 17: Converged weighted average function $\tilde{R}_k$ considering a population of size 100 on Erdős–Rényi networks with average degree λ, at time 50. The function $\tilde{R}_k(t)$ can be regarded as converged for $t = 50$. The gray line represent the deterministic final size of a rumor for a complete graph.

To determine if the approximation made in the derivation of the mean-field equations severely impacts the results we can compare our result in Figure 17 with simulations in Figure 18.

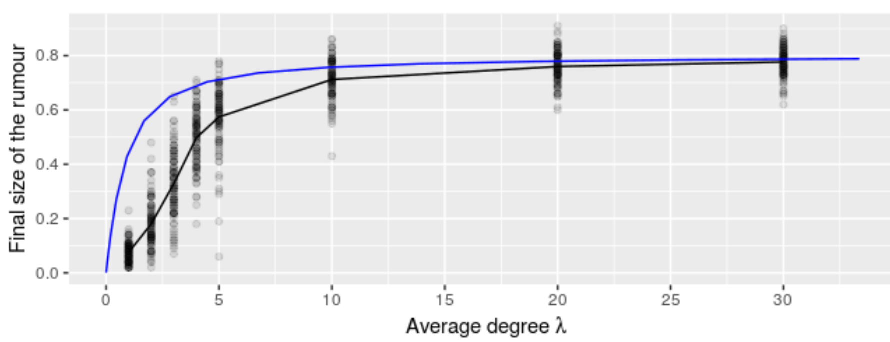


Figure 18: Simulated final size of the rumor considering the Maki–Thompson model. Each simulation generates a Erdős–Rényi graph with 100 nodes with one randomly chosen initial spreader. The final size of the rumor is simulated 100 times for each average degree λ in the figure. The black line represents the mean of the simulations for each λ and the blue line is the same as in Figure 17.

Unfortunately, we can see that the mean-field equations differ for smaller λ compared to what our simulations show. We can then conclude that the correlations between the states of neighboring nodes have a large impact for networks with low average degree.

However, a deterministic model consider the population sizes on a continuous scale, which may be acceptable for large populations [CCBvdD$^+$02, p.85], but can cause serious problems for smaller populations. In our case, a population of size 100 might not be sufficiently large for our deterministic model to be appropriate. In our model we have only one initial spreader and it is possible that the rumor spreading never takes off and stops with only spreading to a low fraction of the population, which is something that the deterministic model does not take into account. This can also be a reason why our deterministic model show a higher average in the final size of the rumor than what simulations show.

We can also see in Figure 18 how the values are distributed, where points with high opacity is overlapping points which in turn means higher density. We can see that the variance of the distribution of the final size of the rumor is higher for small λ than for larger λ.

4.4 Comparison between models on Erdős–Rényi networks

In previous sections we came to the conclusion that the final size of the rumor was in average the same using the DK model as for the MK model when spreading the rumor on a complete graph. We now want to determine if this also holds for Erdős–Rényi graphs. From simulations we can conclude that this is not the case. We see in Figure 19 that for Erdős–Rényi graphs with small average degree the two models differ considerably, as the final size of the rumor is larger for the MK model. However, both models converge to the same value when increasing λ, namely the value for the final size of the rumor for a complete graph.

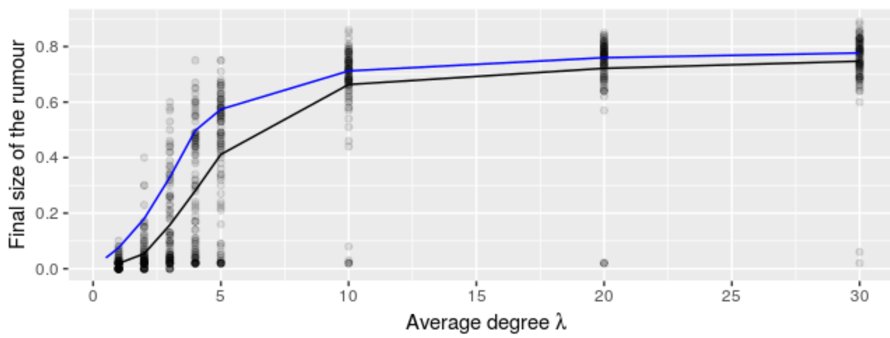


Figure 19: Simulated final size of the rumor considering the Daley–Kendall model. Each simulation generates a Erdős–Rényi graph with 100 nodes with one randomly chosen initial spreader. The final size of the rumor is simulated 100 times for each average degree λ in the figure. The black line represents the mean of the simulations for each λ and the blue line is the respective line considering the MK model.

We can also see that the variance for the final size of the rumor is large for small λ, meaning that the probability of the rumor spreading to a high fraction of individuals is high, but the probability for the rumor to only reach a low fraction is also high. For larger λ however, we can see that the rumor almost always reach a high fraction of individuals.

Moreover, recall that the difference between the DK model and the DKM model was not very significant in cases of large populations when applying the models on complete graphs due to the low probability of individuals spreading back to its source. The low probability for that event is a consequence of the nodes having large degrees. Now if we instead consider applying these two models on a Erdős–Rényi network we can expect larger differences between the models for smaller λ because nodes with lower degree have a higher chance of interacting with its source, after becoming a spreader than nodes with higher degrees have. Therefore we can expect the final size of the rumor to be significantly higher considering the DKM model for small λ and similar to the DK model for large λ. We confirm this by simulations and we can see the comparison between the two models in Figure 20.

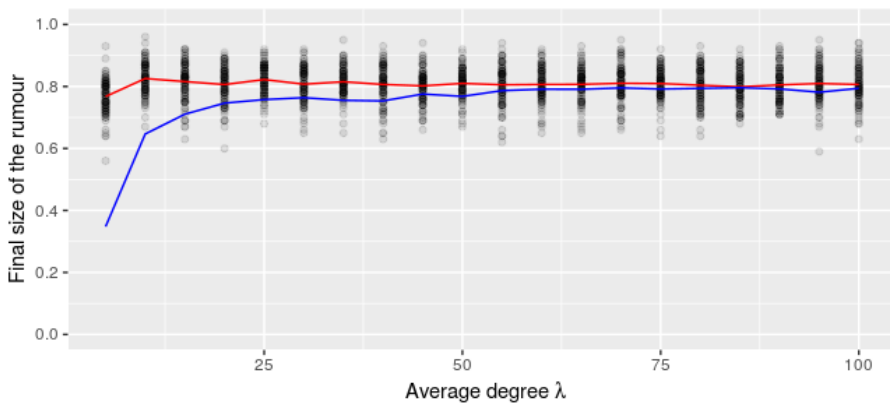


Figure 20: Simulated final size of the rumor considering the Daley–Kendall model with memory. Each simulation generates a Erdős–Rényi graph with 100 nodes with one randomly chosen initial spreader. The final size of the rumor is simulated 100 times for each average degree λ in the figure. The red line represents the mean of the simulations for each λ and the blue line represents the mean of simulations considering the Daley–Kendall model in Figure 19.

We see that by removing the possibility for individuals spread the rumor back to their source we also lower the variance considering the final size of the rumor and we always reach a high fraction of informed individuals. The final size of the rumor appears to be constant with regards to λ which indicates that it does not depend on λ. However, for really small λ we could probably see a change as the number of nodes with no neighbors increases.

5 Conclusion and further questions

In this thesis we have studied different rumor spreading models, the two well-known Daley–Kendall model and Maki–Thompson model and also the modified Daley–Kendall model with memory. We considered the spreading in homogeneously mixed networks and as well as in Erdős–Rényi networks. Using deterministic approximation we came to the conclusion that the DK and MK model share the same final size of the rumor considering homogeneously mixed networks but simulations showed that the DK model have a slightly higher variance. We also saw that a large number of initial spreaders will actually spread the rumor to a lower fraction of the remaining ignorants than a low number of initial spreaders. However, the final size of the rumor will increase with a higher initial number of spreaders. The final size of the rumor is a decreasing function in n, the population size, and converges to approximately 80% of the population knowing about the rumor in the end. For very small populations there were some notably differences between the models but for large population the end result is nearly identical for the models.

We then moved on to apply the models on Erdős–Rényi networks where we expected to see larger differences between the models because the difference in the stifling mechanic. This is a result of a lower average degree in the network. We could see that by preventing a spreader to try to spread back to it's own source the rumor reaches a significantly higher fraction of individuals compared to when it is allowed, for networks with a low average degree.

However, while the models we have studied are interesting they are not fit for prediction of a rumor spreading in a real human network as the models are not realistic. A more realistic model could include, for example, different strength of friendship between the individuals, a forgetting mechanic, tolerance before transforming into spreader or stifler meaning that an individual does not become a spreader after just hearing the rumor once. Also, different rumors can be of different interest which obviously will be a large factor on how large the spread will be.

6 References

[Bri10] T. Britton, *Stochastic epidemic models: A survey*, Mathematical Biosciences **225** (2010), no. 1, 24–35.

[CCBvdD+02] C. Castillo-Chávez, S. Blower, P. van den Driessche, D. Kirschner, and A. Yakubu, *Mathematical approaches for emerging and reemerging infectious diseases: an introduction*, Springer, New York, 2002.

[DK65] D.J. Daley and D.G. Kendall, *Stochastic rumours*, IMA Journal of Applied Mathematics (Institute of Mathematics and Its Applications) **1** (1965).

[Dur06] R. Durrett, *Random graph dynamics*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2006.

[Gut09] A. Gut, *An intermediate course in probability*, Springer New York, 2009.

[KBAP78] D. Katz, J. Baptista, S. P. Azen, and M. C. Pike, *Obtaining confidence intervals for the risk ratio in cohort studies*, Biometrics **34** (1978), no. 3, 469–474.

[Mak73] D. Maki, *Mathematical models and applications : with emphasis on the social, life, and management sciences*, Prentice-Hall, Englewood Cliffs, N.J, 1973.

[NMBM07] M. Nekovee, Y. Moreno, G. Bianconi, and M. Marsili, *Theory of rumour spreading in complex social networks*, Physica A: Statistical Mechanics and its Applications **374** (2007), no. 1, 457–470.

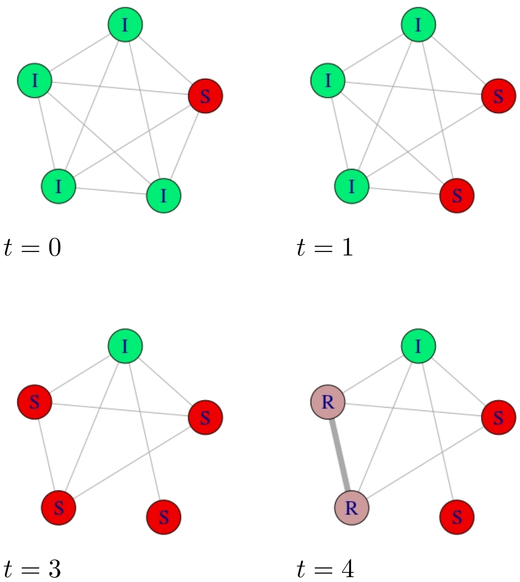
[Rud76] W. Rudin, *Principles of mathematical analysis*, International series in pure and applied mathematics, McGraw-Hill, 1976.

[Sud85] A. Sudbury, *The proportion of the population never hearing a rumour*, Journal of Applied Probability **22** (1985), no. 02, 443–446.

[Wat87] R. Watson, *On the size of a rumour*, Stochastic Processes and their Applications **27** (1987), 141–149.

7 Appendix

Figure 21 and Figure 22 below is used as an example why the Daley–Kendall model with memory is not a Markov process, in Section 3.1.3.

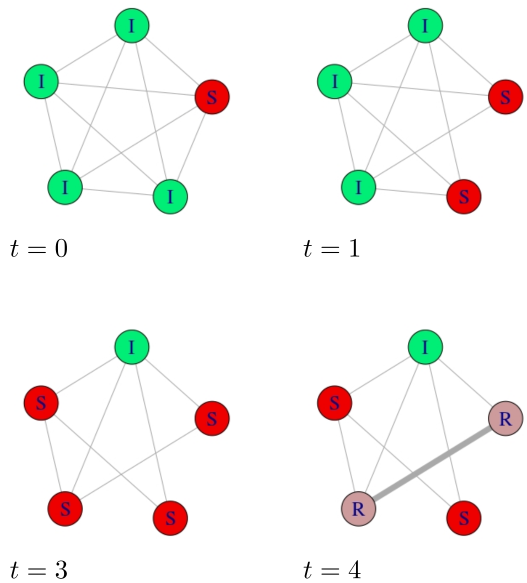


$$P\begin{pmatrix} \mathrm{I(5)=I(4)–1,} \\ \mathrm{S(5)=S(4)+1} \end{pmatrix} = \frac{2}{4}$$

$$P\begin{pmatrix} \mathrm{S(5)=S(4)–1,} \\ \mathrm{R(5)=R(4)+1} \end{pmatrix} = \frac{2}{4}$$

$$P\begin{pmatrix} \mathrm{S(5)=S(4)–2,} \\ \mathrm{R(5)=R(4)+2} \end{pmatrix} = 0$$

Figure 21: Example of a rumor progression using the Daley–Kendall model with memory with a population of size 5. Green nodes represents ignorants, red represents spreaders and grey represents stiflers. The thick or removed edge indicates interaction.



$$P\begin{pmatrix} \text{I(5)=I(4)–1,} \\ \text{S(5)=S(4)+1} \end{pmatrix} = \frac{2}{4}$$

$$P\begin{pmatrix} \text{S(5)=S(4)–1,} \\ \text{R(5)=R(4)+1} \end{pmatrix} = \frac{1}{4}$$

$$P\begin{pmatrix} \text{S(5)=S(4)–2,} \\ \text{R(5)=R(4)+2} \end{pmatrix} = \frac{1}{4}$$

Figure 22: Example of a rumor progression using the Daley–Kendall model with memory with a population of size 5. Green nodes represents ignorants, red represents spreaders and grey represents stiflers. The thick or removed edge indicates interaction.

Lemma 1. *If $X \sim Bin\big(n_x, p_x\big)$ and $Y \sim Bin\big(n_y, p_y\big)$ is independent of each other and $T = \frac{X/n_x}{Y/n_y}$, then $\log(T)$ is approximately normally distributed with mean $\log(\frac{p_x}{p_y})$ and variance $(\frac{1}{p_x} - 1)/n_x + (\frac{1}{p_y} - 1)/n_y$.*

Proof. See proof in [KBAP78].

□

Lemma 2. *If $Z = \frac{1}{L+1}$, where L is log-normal with mean μ and variance σ^2, then Z is logit-normally distributed with parameters $-\mu$ and σ^2.*

Proof. We can derive the probability density function for Z by using the change of variable formula [Gut09, 20-21]. We have that

$$Z = g(L) = \frac{1}{L+1} \implies g^{-1}(Z) = \frac{1}{Z} - 1 \implies \frac{dg^{-1}(z)}{z} = -\frac{1}{z^2}.$$

Now applying the change of variable formula we get

$$\begin{aligned} f_Z(z) &= \left|\frac{dg^{-1}(z)}{z}\right| \cdot f_L(g^{-1}(z)) \\ &= \frac{1}{z^2} \cdot \frac{1}{\frac{1-z}{z}\sqrt{2\pi}\sigma} \exp\left\{ - \frac{\Big(\log(\frac{1-z}{z}) - \mu\Big)^2}{2\sigma^2} \right\} \\ &= \frac{1}{(1-z)z\sqrt{2\pi}\sigma} \exp\left\{ - \frac{\Big(\log(\frac{z}{1-z}) - (-\mu)\Big)^2}{2\sigma^2} \right\}, \end{aligned}$$

which is the density of a logit-normal distribution with parameters $-\mu$ and σ^2. □