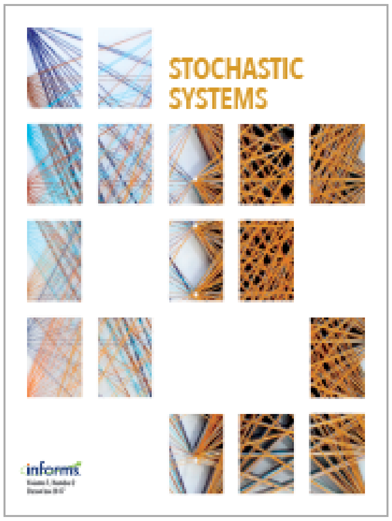


Stochastic Systems

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Simple Markovian Spreading Process with Mobile Agents

Refael Hassin

To cite this article:
Refael Hassin (2021) A Simple Markovian Spreading Process with Mobile Agents. Stochastic Systems 11(1):19-33.
https://doi.org/10.1287/stsy.2019.0061

informs
http://pubsonline.informs.org/journal/stsy

STOCHASTIC SYSTEMS
Vol. 11, No. 1, March 2021, pp. 19–33
ISSN 1946-5238 (online)

A Simple Markovian Spreading Process with Mobile Agents

Refael Hassin[a]

[a] Department of Statistics and Operations Research, Tel Aviv University, 69978 Tel Aviv, Israel
Contact: hassin@post.tau.ac.il, https://orcid.org/0000-0001-9631-6162 (RH)

Received: February 27, 2020
Accepted: June 8, 2020
Published Online in Articles in Advance: October 9, 2020

https://doi.org/10.1287/stsy.2019.0061

Abstract. We investigate a spreading process where each agent is represented by a continuous-time Markov chain with two states, L and M. State L refers to "home," whereas state M refers to a "meeting place." When two agents stay together at M, they "meet" and form a contact. This means, according to the application, that they can exchange information, infect each other, perform an act of trade, and so on. We assume that initially all are at state L, and exactly one of the agents possesses a piece of information (or is infected by a contagious disease, etc.) The process can generally be classified as a spreading process with mobile agents, and its simplicity allows us to demonstrate several interesting properties. We provide an efficient way for computing the propagation time and investigate the dependence of the spreading process on parameters such as the number of agents, the number of uninformed agents at the end of the process, and the contact intensity.

Funding: This research was supported by the Israel Science Foundation [Grant 355/15].

Keywords: spreading process • propagation time

1. Introduction

In a spreading process, an entity, such as information, opinion, or infectious disease, is transferred among agents through a contact mechanism. The literature on spreading processes is extensive and can be studied through several excellent recent surveys, for example, Vespignani (2012), Pastor-Satorras et al. (2015), Nowzari et al. (2016), and Jin et al. (2017).

The common approach to modeling spreading processes considers a network where individuals are represented by nodes, and neighboring nodes communicate and transmit a piece of information (we will apply this terminology throughout). At any point of time, a node that possesses the information may play the role of a *spreader* that acts to spread the information to its neighbors. Other nodes may represent ignorant agents (susceptible-infectious model) and informed agents who are not active in further transmitting the information (susceptible-infectious-recovered model). There are many variations of this model that include randomness and status changes, as surveyed by de Arruda et al. (2018). Central questions investigated include the expected time it takes to deliver the message to all nodes (e.g., Chierichetti et al. 2011) or characterizing the steady state of the system.

Specifically, we consider a finite set of N mobile individuals (agents), each moving between their individual home locations $L_1, \ldots, L_N$ and a common meeting location denoted M.[1] Movements are controlled by a continuous-time Markov chain. The time spent in state L_i is exponentially distributed with parameter λ, after which the individual moves to M. The time duration at M is exponential with parameter μ, and then the individual returns to his or her origin L_i. Initially, all of these N individuals are at their origins; one of them is informed about a piece of information, often referred to as a *rumor*. An informed person at the meeting location M shares the information with the others who are at M. We are interested in the way the information propagates in the population and specifically the time until everybody is informed, that is, the *propagation time*.

The underlying graph of our network is a *star graph* with a single *hub node* M representing the meeting place and leaves $L_1, \ldots, L_N$ representing origins of the individuals in the population. We will simplify the exposition by merging the origins $L_1, \ldots, L_N$ into a single location L in which no contact is created and information is not shared. Thus, instead of considering a star graph, we simply have a network with two nodes.

A similar model is considered by Gilboa-Freedman and Hassin (2016), in which one of the individuals in the population plays the role of a *leader*. The authors provide a formula for the expected time until the leader has met any given number of nonleaders.

Epidemic models often use the unitless *reproductive number* $R = \lambda(k)/\mu$, where $\lambda(k)$ is the rate by which an individual with k contacts spreads the disease, and μ is the individual recovery rate. Analogously, we use a unitless *contact intensity* measure $\rho = \lambda/\mu$ giving the ratio of individual transition rate from L to M to the rate of reverse movement. A large contact intensity means that the time spent at M is significantly longer than the time spent at L, and the intensity of mutual contacts is high.

Common models of spreading processes are considerably different from ours. These processes either reach a steady state or end up at a single absorbing state (such as when all individuals are cured). In our model, there is a single absorbing state when all are informed.

Our model belongs to the family of spreading processes with mobility. The relevant literature is surveyed in section IX of Pastor-Satorras et al. (2015). Most of these models assume that each node of the network hosts a single agent, whereas the paper by Colizza and Vespignani (2008) investigates *metapopulation systems*, where each node of a network hosts multiple agents who move along the arcs of the network. Other papers also assume Markovian movements. For example, our model can be considered as a very special case of the general model by Sattenspiel and Dietz (1995). These authors combine network mobility and disease spreading as follows: Each individual has an origin location (node in the network). Individuals move in the network according to a transition (migration) matrix that has a special component for returning to the origin node. The infection process is also general and depends on both the number of contacts with others in the node and its neighboring nodes and also on the node itself. In our model, individuals always return from M to their origins L_i, and we assume that contagion is certain at M and impossible at L_i.

Most of the previous literature focuses on imitating real-life spreading processes. This results in complex networks and types of interactions that are impossible to solve, even for simple special cases. In contrast, our goal is to suggest a new approach and exemplify the way it operates by investigating a simplified case. Because of this simplicity, it is possible to provide definite answers to central questions concerning the behavior of the process.

Perhaps closest to our approach is the paper by Groenevelt et al. (2005) and, in particular, their unrestricted multicopy protocol, where a source agent initially possesses information that needs to be delivered to a destination agent. Each pair of agents within a group, which includes these two, meets for a very short time in intervals of exponential length and exchange information. The main question is characterizing the distribution of the source-to-destination time. Zhang et al. (2007) derive closed-form formulas for a fluid approximation of this model.

A main feature of our model, which differs from conventional models of spreading processes, is that we allow *simultaneous interaction* (transmission of information, contagion), whereas the common convention assumes single interactions. Consequently, our model does not define a birth–death process. A spreading process on a star graph is also considered by Milling et al. (2015).

We now provide an outline of this paper together with our main results and insights.

An increase in the number of agents N has two opposite effects on the time of propagation—more agents must be informed but also more serve to spread the information. We demonstrate an interesting dependence of the propagation time on the number of agents in two extreme cases of contact intensity (Section 2.1). When the contact intensity is very high, the propagation time is proportional to the harmonic sum $H(N)$ and increases with N. In contrast, when the contact intensity is very low, the propagation time is proportional to $H(N)/N$ and decreases with N. We numerically investigate the dependence for intermediate contact intensities, showing that depending on the contact intensity, the function can be monotone increasing or first decreasing and then increasing, except for from $N = 2$ to $N = 3$, where it always increases.

We present an efficient way for solving for the expected propagation time that emerges from an interesting presentation of the state transition diagram (Section 2.2). Rather than solving a system of $O(N^2)$ linear equations, it is sufficient to solve $N - 1$ systems of N equations, which is clearly an improvement as long as there is no linear-complexity algorithm for solving the equations.

With a given population size N, the expected time it takes to inform the next agent increases with the number of already informed agents. Allowing a small fraction of agents to remain uninformed considerably shortens the propagation time (Section 2.3).

We find that the probability that all agents are informed by time t as a function of t is S shaped (Section 3.1). Moreover, the expected number of customers informed by a given time t has an S shape (Section 3.2). This finding conforms with the theory of *diffusion of innovations*, as detailed in Rogers (1995). This theory emerged from observing that in numerous studies in various fields the rate of adoption of an innovation has an S shape. This finding has since been verified in many other instances and supported by theoretical models. For a random sample of recent examples in diverse areas that demonstrates empirical and theoretical ubiquitous existence of the S-shaped curve of diffusion, see Centola (2010), Isella et al. (2011), Feola and Butt (2017), Simpson and Clifton (2017), Manzo et al. (2018), Mascia and Mills (2018), and Xiong et al. (2018).

We also investigate possible controls of the process, such as the dependence of the propagation time on the contact intensity (Section 4.1); the contact intensity required to achieve a given target expected propagation time (Section 4.2); and the effect of increasing the seed size, that is, the number of agents who are initially informed (Section 4.3). This setup is similar to the influence-maximization problem considered by Sun et al. (2018). We observe that the seed size has a strong effect when the seed is either very small or very large, and the effect is quite moderate in between. This observation may have consequences when deciding on a promotion campaign that intends to inform the market on a new product.

2. Expected Propagation Time

Recall that at any point of time, either all or none of those at M are informed. The state of the system can be described by three variables: the number of informed agents at L, denoted i; the number of uninformed agents at L, denoted j; and the state, informed or uninformed, of the other $N-i-j$ agents who are at M. Figure 1 illustrates the transitions between states when $N=3$.

Define $T_{i,u}$ as the expected time to the termination of the spreading process when there are i informed individuals and u uninformed individuals at L, and the rest (who are at M) are informed. We define $\hat{T}_{i,u}$ similarly but assuming that those at M are uninformed.

There is no need to define $T_{N,0}$ or $T_{N-1,0}$ because the first time all are informed is when the process terminates, and there must be at least two agents at M. For $i=0,\ldots,N-2$,

$$T_{i,0}=0. \tag{1}$$

For $u\geq 1$ and $i+u<N$,

$$\left[(i+u)\lambda+(N-i-u)\mu\right]T_{i,u}=1+i\lambda T_{i-1,u}+u\lambda T_{i,u-1}+(N-i-u)\mu T_{i+1,u} \quad i\geq 0 \tag{2}$$

(there is no need to define $T_{-1,u}$ because its coefficient is zero) and

$$\left[(i+u)\lambda+(N-i-u)\mu\right]\hat{T}_{i,u}=1+i\lambda T_{i-1,u}+u\lambda \hat{T}_{i,u-1}+(N-i-u)\mu \hat{T}_{i,u+1} \quad i\geq 1 \tag{3}$$

($\hat{T}_{0,u}$ is not defined because there must be at least one informed agent).

For $i=1,\ldots,N$ and $i+u=N$,

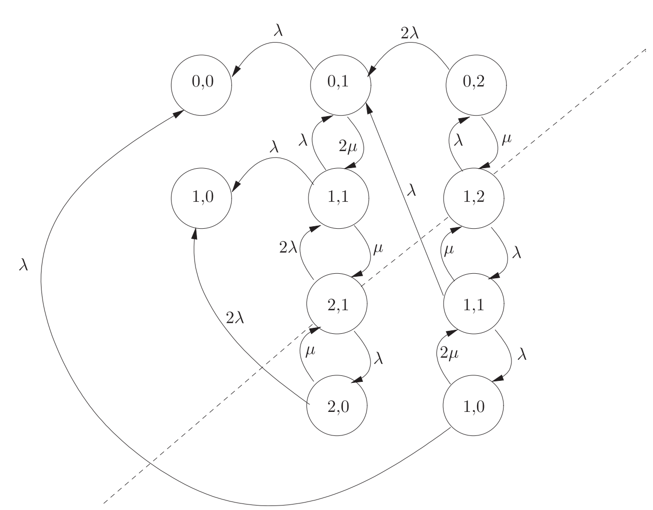
$$N\lambda T_{i,N-i}\left(\equiv N\lambda\hat{T}_{i,N-i}\right)=1+i\lambda T_{i-1,N-i}+(N-i)\lambda\hat{T}_{i,N-i-1}. \tag{4}$$

(State $(0,N)$ is not possible because there must be at least one informed agent, and state $(N,0)$ is not possible, as explained earlier.)

Note that Equation (4) is a special case of (3) when $i+u=N$. We could therefore delete (4) and change the range for (3) to $i+u\leq N$ while noting that $T_{i,N-i}\equiv\hat{T}_{i,N-i}$.

Our main interest is in investigating the properties of $T_{1,N-1}$ ($\equiv\hat{T}_{1,N-1}$).

Figure 1. Transition Diagram $N=3$



Note. States above the dashed diagonal are those where agents at M are informed; in those below the diagonal, they are uninformed; and on the diagonal, there are no agents at M.

2.1. Extreme Cases

2.1.1. $N = 2$**.** The solution when $N = 2$ is easy and can be used to illustrate the system's sensitivity to its parameters. The transition equations are

$$T_{1,1} = \frac{1}{2\lambda} + \frac{1}{2}T_{0,1} + \frac{1}{2}\hat{T}_{1,0},$$
$$\hat{T}_{1,0} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}T_{1,1},$$
$$T_{0,1} = \frac{1}{\lambda+\mu} + \frac{\mu}{\lambda+\mu}T_{1,1}.$$

The solution is

$$\frac{T_{1,1}}{1/\lambda} = \frac{3}{2} + \frac{1}{2\rho}, \quad \frac{\hat{T}_{1,0}}{1/\lambda} = \frac{T_{0,1}}{1/\lambda} = 1 + \frac{1}{2\rho}, \tag{5}$$

where the unitless variable $\frac{T_{i,u}}{1/\lambda}$ measures the propagation time in units of expected sojourn time at L. Thus, the normalized propagation time linearly increases in the inverse contact intensity $1/\rho$, as can also be seen from Figure 3 and the upper figure in Figure 9.

2.1.2. $\rho \to \infty$**.** Consider $\rho \to \infty$. This means that with high probability all agents will move to M before any of them moves back to L. In this case, the expected time until the first agent moves to M is $\frac{1}{(N-1)\lambda}$, and then it takes expected time $\frac{1}{(N-2)\lambda}$ until the second move, and so on. Therefore,

$$\frac{T_{1,N-1}}{1/\lambda} = \left(\frac{1}{N} + \frac{1}{N-1} + \cdots + 1\right) = H(N),$$

where H is the harmonic function that asymptotically behaves like $\ln N$. Note that this expression is a lower bound for $T_{1,N-1}$ for any $\mu > 0$.

2.1.3. $N \to \infty$**.** Suppose that ρ is fixed and $N \to \infty$. In this case, once the informed agent moves to M, which takes, on average, $1/\lambda$ time, a large group of informed agents is formed because of the high flow of uniformed agents from L to M. Then, with high probability, there is always at least one informed agent at M, and the situation is as in the case of $\mu \to 0$; namely, the expected time to termination of the process is $\frac{H(N)}{\lambda}$. Therefore, we conclude that

$$1 < \frac{T_{1,N-1}}{1/\lambda} < H(N) + 1.$$

The lower bound is approached when μ is very small, so those who move to M stay there with high probability, and the upper bound is approached when μ is large, so counting actually starts only after the initially informed agent moves to M. This behavior is illustrated in Figure 3.

2.1.4. $\rho \to 0$**.** When $\rho \to 0$, the time spent at M is relatively very short. Consider a typical state where all N agents are at L, and i of them are informed. There are $(N-i)i$ pairs of an informed agent and an uninformed agent, and we are interested in the time until the first pair meets. When $\rho \to 0$, these meetings can be considered as independent events. Consider a given pair. It is easy to see that the steady-state probability that one of these individuals is at L and the other is at M is $2\rho/(1+2\rho+2\rho^2) \to 2\rho$. The average number of transitions into the state where both are at M is $2\lambda\rho$. Given that there are $(N-i)i$ pairs, and using the independence assumption, the rate by which meeting events occur is $2\lambda\rho(N-i)i$. The expected time until the first meeting occurs is therefore $\frac{1}{2\lambda\rho(N-i)i}$, and then we have $i+1$ informed agents. Therefore,

$$\frac{T_{1,N-1}}{1/\lambda} = \sum_{i=1}^{N-1}\frac{1}{2\rho(N-i)i} = \frac{1}{2\rho}\sum_{i=1}^{N-1}\frac{1}{N}\left(\frac{1}{N-i} + \frac{1}{i}\right) = \frac{H(N-1)}{\rho N}. \tag{6}$$

2.2. Computation

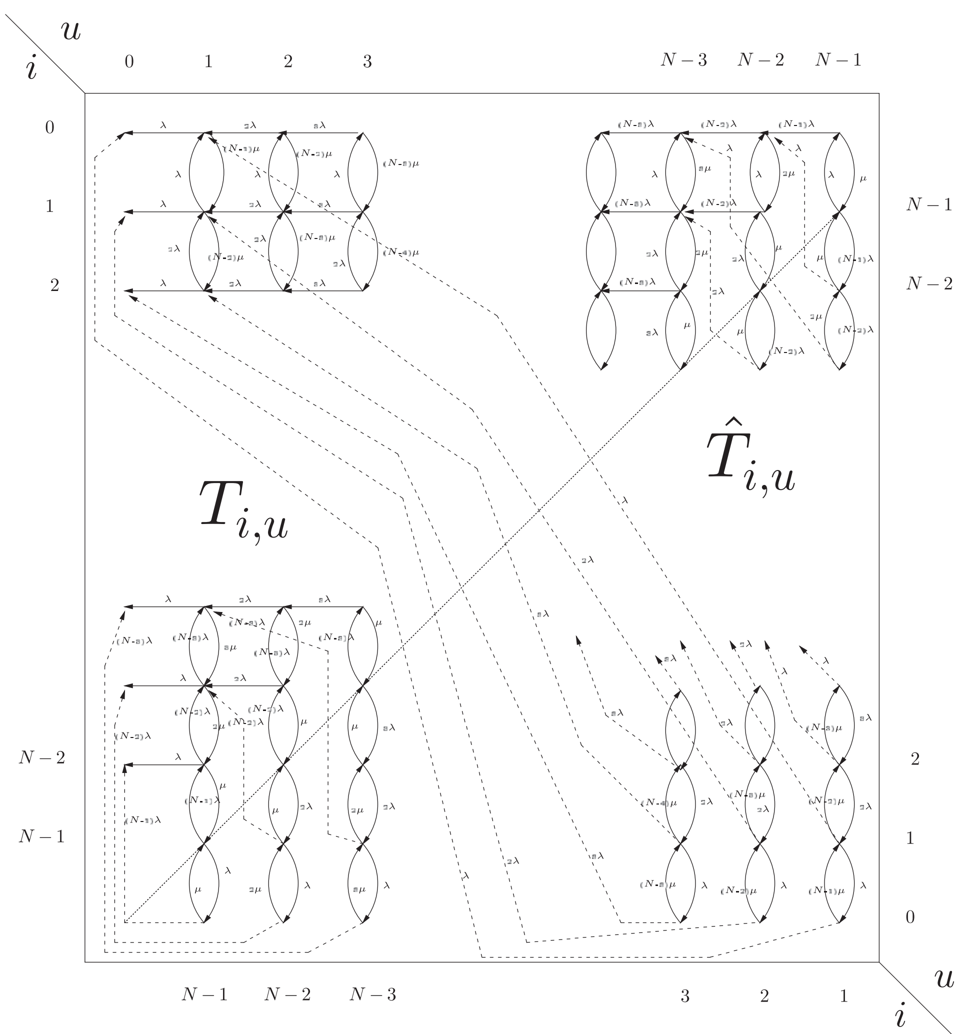
Figure 2 shows the transition diagram for the process. Above the diagonal we have the states where the agents at M are informed, in the states below the diagonal they are uninformed, and on the diagonal we have the states with no agents at M. Note that indexing above the diagonal is different from the indexing below the diagonal: the i and u coordinates are swapped and obtain complementary values that sum up to N. The long dashed arrows indicate transitions from states where those at M are uninformed, and an informed agent moves from L to M simultaneously informing all agents already there.

Remark 1. Column j of the transition diagram, starting the count from $j = 0$, corresponds to the set of states with a total number of j uninformed agents and $N - j$ informed agents. To see this property, observe that the states above the diagonal in this column have j uninformed agents (the other $N - j$ agents are informed because these states correspond to $T_{1,N-1}$), whereas states under the diagonal have $N - j$ informed agents (the other j agents are uninformed because these states correspond to $\hat{T}_{1,N-1}$).

We observe that *there are no transitions from left to right*; therefore, the computation can be done one column at a time from left to right. In our computations, we use a common indexing system for all entries. Thus, entry (p, q) refers to the state in row p and column q of the transition diagram, as given in Figure 2. The range of indices is $p = 0, 1, \ldots, N$ and $q = 0, 1, \ldots, N - 1$. Thus, for $p + q < N$, the state is above diagonal and corresponds to a T value; if $p + q > N$, then the state is below diagonal and corresponds to a $\hat{T}$ value; when $p + q = N$, the T and $\hat{T}$ values coincide.

The initial conditions are $T_{i,0} = 0$ for the left-most column. Suppose that the T and $\hat{T}$ values corresponding to columns $j < q$ have already been computed. We now compute the values corresponding to column q as follows. Let X_p be the expected time to termination of the process when the system is at the state

Figure 2. Transition Diagram



Note. States above the diagonal are those where agents at M are informed; in those below the diagonal, they are uninformed; and on the diagonal, there are no agents at M.

corresponding to row p in column q (we eliminate q from this notation because we assume that it is fixed). Thus, for $p \le q$, we have $X_p = T_{p,q}$; and for $p \ge q$, we have $X_p = \hat{T}_{N-q,N-p}$.

We now write Equations (2)–(4) in terms of the unknowns X_p and the already known variables $T_{i,u}$ corresponding to columns $j < q$. The transformation of Equations (2) and (4) is straightforward, and we simply replace i by p. For $p = 0, \ldots, N-q-1$, Equation (2) becomes

$$-p\lambda X_{p-1} + \big[(p+q)\lambda + (N-p-q)\mu\big]X_p - (N-p-q)\mu X_{p+1} = 1 + q\lambda T_{p,q-1}. \tag{7}$$

For $p = N-q$, Equation (4) becomes

$$-p\lambda X_{p-1} + N\lambda X_p - q\lambda X_{p+1} = 1.$$

Care must be taken when writing the equation for $p = N-q+1, \ldots, N$. In this case, when dealing with below-diagonal entries corresponding to $\hat{T}$ values, we use the relations $q = N-i$ and $p = N-u$. Here $\hat{T}_{i,u}$ corresponds to the unknown X_p, $\hat{T}_{i,u-1}$ corresponds to X_{p+1}, $\hat{T}_{i,u+1}$ corresponds to X_{p-1}, and $T_{i-1,u}$ is the already computed $T_{N-q-1,N-p}$. Equation (3) then becomes

$$-(p+q-N)\mu X_{p-1} + \big[(2N-q-p)\lambda + (p+q-N)\mu\big]X_p - (N-p)\lambda X_{p+1} = 1 + (N-q)\lambda T_{N-q-1,N-p}. \tag{8}$$

Note that X_{-1} and X_{N+1} need not be defined because they appear with zero coefficient.

Thus, we repeatedly solve a system of $N+1$ linear equations in $N+1$ unknowns and then increase u by one until we obtain the complete solution.

Figure 3 shows the expected propagation time for various values of N and μ. We normalize the time in our numerical examples so that $\lambda = 1$. The lower graph corresponds to $\mu = 0.1$; as explained in Section 2.1.2, $T_{1,N-1}$ is approximately the harmonic sequence.

Increasing N has two contradicting effects: more agents need to be informed, but also more serve to spread the information. Which effect dominates depends on the number of agents and on ρ. We observe, however, that for any value of ρ (or μ in the figures where λ was normalized to one), the curves always increase when N goes from two to three (the increase is less significant as μ increases). We can prove the correctness of this observation as follows.

Proposition 1. $T_{1,2}^{\{N=3\}} > T_{1,1}^{\{N=2\}}$.

Proof. The proof applies path analysis. Consider a two-agent process G2 and a three-agent process G3. The agents in G2 are i (the informed agent) and u_1 (the uninformed agent). Initially, both are at L. In G3, we initially have an additional uninformed agent u_2 at L. Consider the first transition of i to M. There are four possible cases for the agents that agent i meets at M:

1. Only u_1
2. Both u_1 and u_2
3. Only u_2
4. Nobody

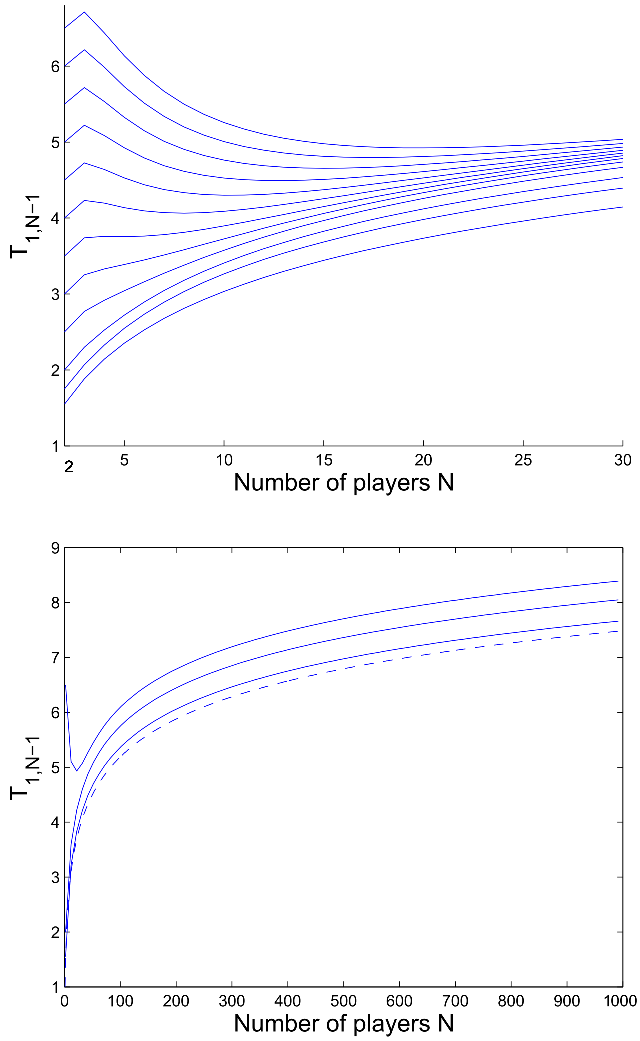
In the second case, both processes terminate. Cases 1 and 3 have equal probabilities. In case 1, G2 terminates, whereas in case, 3 G3 has not yet terminated. Thus, so far, G2 has a shorter expected residual propagation time. Consider now case 4. If the next event is the return of i to L, then we are back at the starting point. Otherwise, there is an equal probability that the next event is the transition of u_1 or u_2 to M. Again, in the former case, G2 terminates, whereas, in the latter case, G3 is still ongoing. The conclusion is again that G2 has a smaller expected propagation time. □

Remark 2. We observed that the propagation time increases when the number of agents increases from two to three. The effect of a further increase in N depends on μ. For small values of μ, $T_{1,N-1}$ is monotone increasing in N. For larger values of μ, the function decreases first and then increases. For even larger values of μ, the function is monotone decreasing. The μ value where $T_{1,2} = T_{1,3}$ is approximately 4.33.

2.3. Expected Time to Inform a Given Fraction

Suppose that we are interested in the expected times $T_{i,u}^{(\alpha)}$ and $\hat{T}_{i,u}^{(\alpha)}$ it takes to inform a fraction α of the population. Then $T_{i,u}^{(\alpha)} = 0$ if the number of informed agents $N-u$ exceeds $N\alpha$ or $u \le N(1-\alpha)$. Similarly, $\hat{T}_{i,u}^{(\alpha)} = 0$ if $i \ge \alpha N$. Let $N_0 = \lfloor (1-\alpha)N \rfloor$. Then, according to Remark 1, these states are exactly those in the first $N_0 + 1$ columns of the transition diagram of Figure 2. Thus, the only change in the balance equations is that the initial

Figure 3. $T_{1,N-1}$, $\lambda = 1$



Notes. In the upper figure, $N \le 30$ and $\mu = 0.1, 0.5, 1, 2, \ldots, 10$. In the lower figure, $N \le 1{,}000$, $\mu = 0.1, 1, 10$; the dashed line shows $H(N)$. Increasing μ values correspond to the graphs from the bottom up.

conditions given by Equation (1) are extended so that all variables corresponding to the first N_0 columns in Figure 2 are zero. For the other columns, Equations (7) and (8) do not change.

Remark 3. Figure 4 shows typical examples. We see that the expected time it takes to inform an additional agent is initially short and then increases. Thus, any relaxation in the requirement to inform the entire population is very meaningful in terms of propagation time. For example, when $\mu = 1$, allowing 5 of 100 agents to be left uninformed reduces the expected propagation time from 5.76 to 3.46.

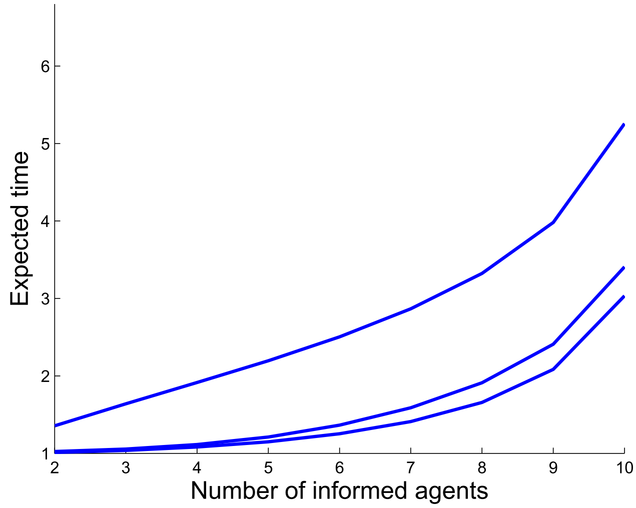
Figure 5 presents similar graphs but in terms of the fraction of targeted agents versus the fraction of time it takes to reach this target, in expectation, relative to the time it takes to inform the whole population.

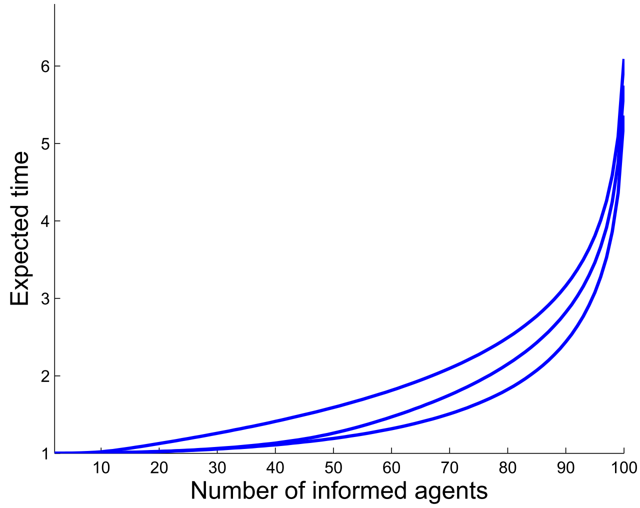
Remark 4. We observe that the curves associated with smaller values of μ are lower at zero but overtake those related to larger values of μ and thus take higher relative times to reach high fractions of the population. This reflects the property that when μ is small, the process is relatively faster at the beginning because there is a higher probability of meeting an informed agent at M.

3. Propagation Within a Time Window

We now consider information that is relevant only within a given time window $[0, t]$, and after this interval, agents stop spreading the information. We are interested in the number of agents exposed to the information within the time window and the probability that the process terminates before t after all agents are informed.

Figure 4. The Expected Time Until at Least $x = 2, \ldots, N$ Agents Are Informed





Notes. Here $\lambda = 1$. In each figure, the lower graph corresponds to $\mu = 0.1$, the middle graph corresponds to $\mu = 1$, and the upper graph corresponds to $\mu = 10$. In the upper figure, $N = 10$, and in the lower figure, $N = 100$.

3.1. The Cumulative Distribution Function of $T_{1,N-1}$

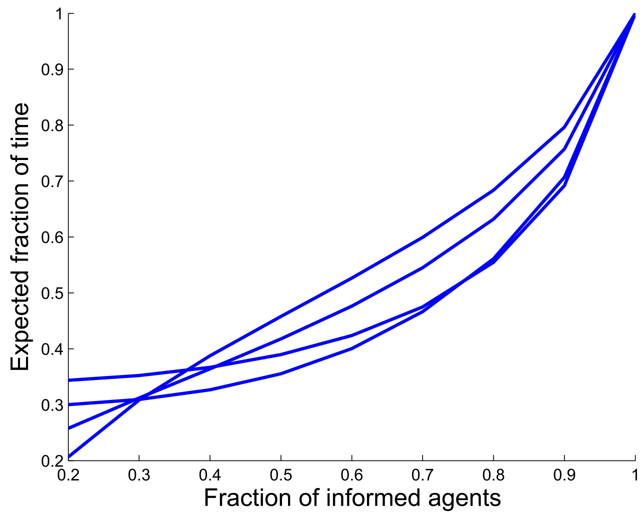
We now consider the cumulative distribution function (CDF) of the propagation time. For $t > 0$, we compute the probability of success; that is, the process terminates (all are informed) no later than t. Let $P(i, u, t)$ and $\hat{P}(i, u, t)$ denote the probabilities that the process terminates no later than t time units from now, given the current state (i, u), and the agents at M are informed and uninformed, respectively. These probabilities change with t as follows.

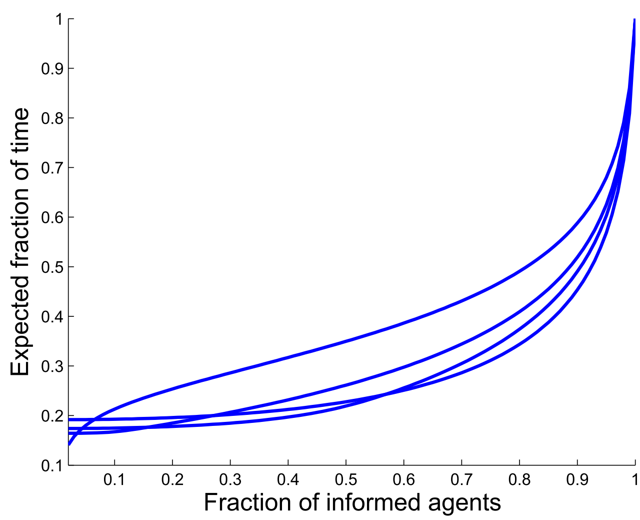
Initially, $P(i, u, 0) = 1$ if $u = 0$ and $P(i, u, 0) = 0$ if $u > 0$. Similarly, $\hat{P}(i, u, 0) = 0$ if $i < N$ and $\hat{P}(N, 0, 0) = P(N, 0, 0) = 1$. Also, for all $t > 0$ and all i, $P(i, 0, t) = 1$.

For $t > 0$ and $i + u < N$,

$$\begin{aligned} P(i,u,t+dt) = {} & \left[1 - (i+u)\lambda dt - (N-i-u)\mu dt\right] P(i,u,t) \\ & + i\lambda dt P(i-1,u,t) \qquad \text{if } i > 0 \\ & + u\lambda dt P(i,u-1,t) \qquad \text{if } u > 0 \\ & + (N-i-u)\mu dt P(i+1,u,t) + o(dt), \end{aligned}$$

Figure 5. The Expected Fraction of Time Until a Given Fraction of the Population Is Informed





Notes. The upper figure has $N = 10$, and the lower figure has $N = 100$. In each figure, the graphs correspond to $\mu = 0.01, 1, 10, 100$ from top down near zero and from bottom up near one.

$$\begin{aligned}\hat{P}(i,u,t+dt) = {} & \big[1-(i+u)\lambda dt-(N-i-u)\mu dt\big]\hat{P}(i,u,t)\\ & + i\lambda dt P(i-1,u,t) \qquad \text{if } i>0\\ & + u\lambda dt\hat{P}(i,u-1,t) \qquad \text{if } u>0\\ & + (N-i-u)\mu dt\hat{P}(i,u+1,t)+o(dt).\end{aligned}$$

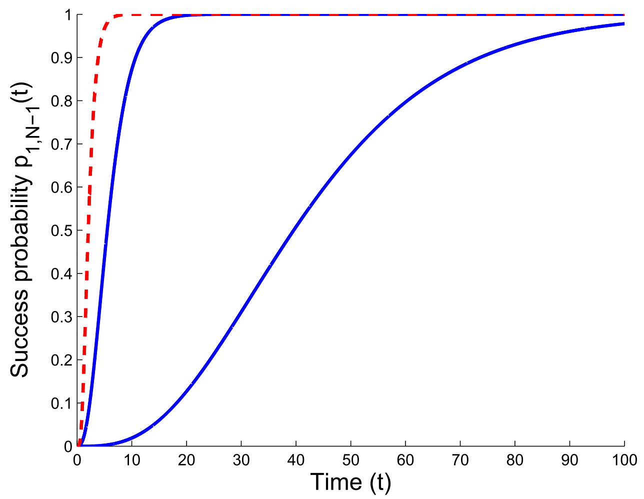
For $t > 0$ $i > 0$, $u > 0$, and $i + u = N$, $P(N,0,t) = 1$ for $t > 0$, and when $i < N$,

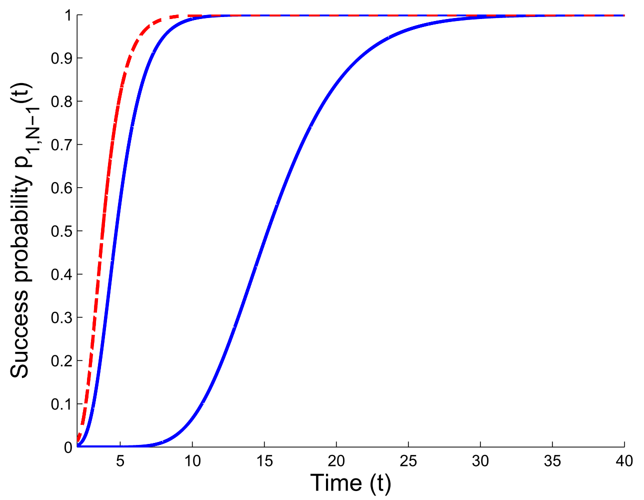
$$\begin{aligned}P(i,u,t+dt) = \hat{P}(i,u,t+dt) = {} & (1-N\lambda dt)P(i,u,t)\\ & + i\lambda dt P(i-1,u,t)\\ & + u\lambda dt\hat{P}(i,u-1,t)+o(dt).\end{aligned}$$

Figure 6 shows the typical graph of the probability $P(1,N-1,t)$ that all are informed by time t (the success probability). Note that when $\rho \to \infty$, the process terminates when all agents move to M. The probability that this happens before time t is simply the probability that a given agent moves to M before t raised to the power of N, that is, $(1-e^{-\lambda t})^N$. This probability is illustrated by the dashed curves in Figure 6 together with the computed CDF for $\mu = 10$ and $\mu = 50$ when $N \in \{5, 30\}$ and $\lambda = 1$.

Stochastic Systems, 2021, vol. 11, no. 1, pp. 19–33, © 2020 The Author(s)

Figure 6. Cumulative Distribution Functions for $T_{1,N-1}$ When $\mu \to 0$ (Upper Dashed Graph), $\mu = 10$ (Middle Graph), and $\mu = 100$ (Lower Graph)





Notes. Here $\lambda = 1$. In the upper figure, $N = 5$, and in the lower figure, $N = 30$.

Remark 5. The CDF has an S shape, similar to the normal distribution. Indeed, $T_{1,N-1}$ is the sum of random variables corresponding to the time it takes to increase the number of informed customers by one. However, these random variables are not independent; for example, several agents are often simultaneously informed. We can also see that the distribution does not converge to the Normal distribution from the formula $T_{1,N-1} \longrightarrow (1 - e^{-\lambda t})^N$ when $\rho \longrightarrow \infty$.

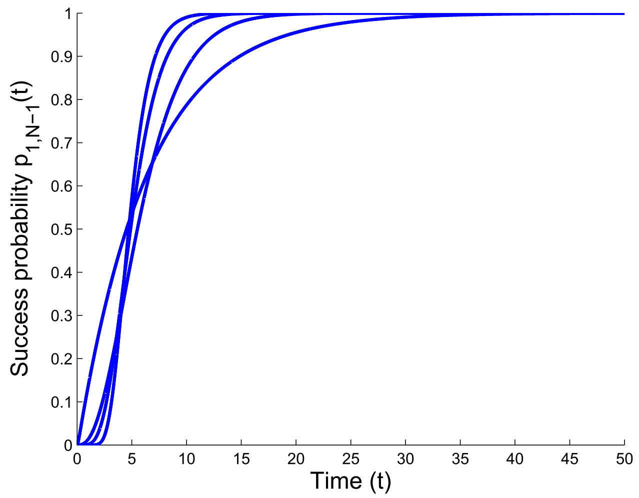
Remark 6. Figure 7 shows the sensitivity of the CDF to changes in N. We see that as N increases, both probabilities for a very short and a very long propagation time decrease. This can be explained by noting that, on the one hand, when N increases, more agents need to be informed, and the probability that this will be accomplished within a short time period decreases. On the other hand, when N increases, a seed of informed agents forms quickly, and from this point, the propagation is fast.

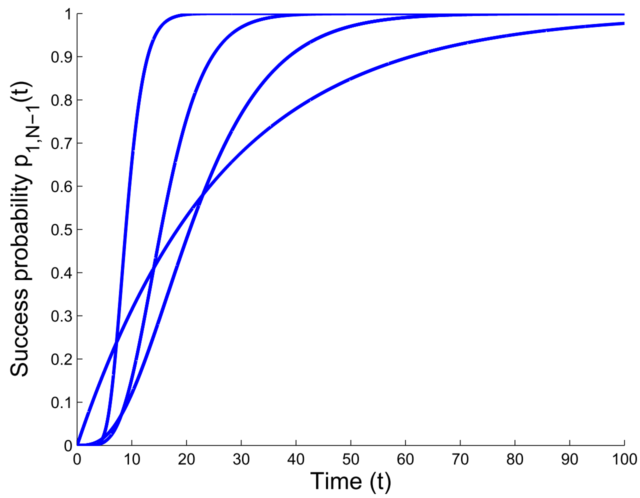
3.2. Number of Informed Agents

How many agents will be informed t time units after the beginning of the process? We now compute the state probabilities at time t and in particular the expected number of informed agents.

It is easy to compute this number when $\rho \longrightarrow \infty$. In this case, a given individual is informed at time t if either this is the initially informed agent or both this agent and the initially informed have moved from L to M before time t. Therefore, the expected number of informed agents at t is $1 + (N-1)(1 - e^{-\lambda t})^2$.

Figure 7. Cumulative Distribution Functions for $T_{1,N-1}$ When $\mu = 10$ (Upper Figure) and $\mu = 50$ (Lower Figure)





Notes. Here $\lambda = 1$. Also, $N = 2, 5, 10, 30$, where N increases from top down near $t = 0$ and from bottom up for large values of t.

We now describe the general solution. Let $Q(i,u,t)$ and $\hat{Q}(i,u,t)$ denote the probabilities of the corresponding states at time t. Initially at time $t = 0$, $Q(1, N-1, 0) = 1$ and, equivalently, $\hat{Q}(1, N-1, 0) = 1$, whereas all other states have zero probability. Then we have the following differential equations (up to an $o(dt)$ tolerance).

For $i + u < N$,

$$\begin{aligned} Q(i,u,t+dt) = {} & \big[1 - (i+u)\lambda dt - (N-i-u)\mu dt\big] Q(i,u,t) \\ & + (N-i-u+1)\mu dt Q(i-1,u,t) \qquad \text{if } i > 0 \\ & + (i+1)\lambda dt Q(i+1,u,t) \qquad\qquad\qquad\qquad (*) \\ & + (u+1)\lambda dt Q(i,u+1,t) \qquad \text{if } i+u+1 < N \\ & + (i+1)\lambda dt \hat{Q}(i+1,u,t) \qquad \text{if } i+u+1 < N. \end{aligned}$$

(The strict inequality in the last line prevents duplication with $(*)$ when $i + u + 1 = N$.) Thus,

$$\begin{aligned} \hat{Q}(i,u,t+dt) = {} & \big[1 - (i+u)\lambda dt - (N-i-u)\mu dt\big] \hat{Q}(i,u,t) \\ & + (N-i-u+1)\mu dt \hat{Q}(i,u-1,t) \qquad \text{if } u > 0 \\ & + (u+1)\lambda dt \hat{Q}(i,u+1,t). \end{aligned}$$

For $i + u = N$,

$$\begin{aligned} Q(i,u,t+dt) \equiv \hat{Q}(i,u,t+dt) = [1 - n\lambda dt]Q(i,u,t) & \\ + \mu dt Q(i-1,u,t) & \qquad \text{if } i > 0 \\ + \mu dt \hat{Q}(i,u-1,t) & \qquad \text{if } u > 0. \end{aligned}$$

The probability that at time t there are least k informed agents is

$$\sum_{\substack{u \le N-k \\ i+u \le N}} Q(i,u,t) + \sum_{\substack{i \ge k \\ i+u<N}} \hat{Q}(i,u,t).$$

Note that the probabilities $P(i,u,t)$ and $\hat{P}(i,u,t)$ are related to the probabilities $Q(i,u,t)$ and $\hat{Q}(i,u,t)$. For example, $P(1,N-1,t)$ can be computed as $\sum_{i=0}^{N} Q(i,0,t)$.

Figure 8 shows the behavior of the expected number of informed agents as a function of time. The curves are S shaped, similar to common empirical data and theoretical models.

4. Process Control

4.1. Dependence of the Propagation Time on Contact Intensity

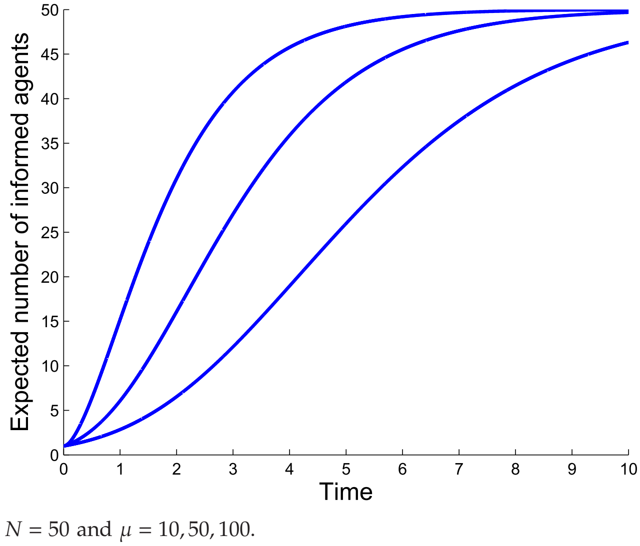
If we deal with information spreading, we may be interested in accelerating the propagation process. In contrast, if it is about disease contagion, we will be interested in slowing it. A natural control variable is the contact intensity ρ, which can be controlled by changing λ or μ. Figure 9 shows the dependence of the expected propagation time on μ for various sizes of the population and $\lambda = 1$.

Remark 7. Interestingly, the curves intersect. It can be seen from the formulas in Sections 2.1.2 and 2.1.4 that for a small μ (high ρ) propagation time, $H(N)/\lambda$ increases with N, so the process is faster when N is small. The opposite holds when μ is large (ρ is small), where propagation time $(\mu/\lambda^2)H(N-1)/N$ decreases with N (for $N > 2$). (Compare this observation with Remark 2.)

4.2. Achieving a Target Propagation Time

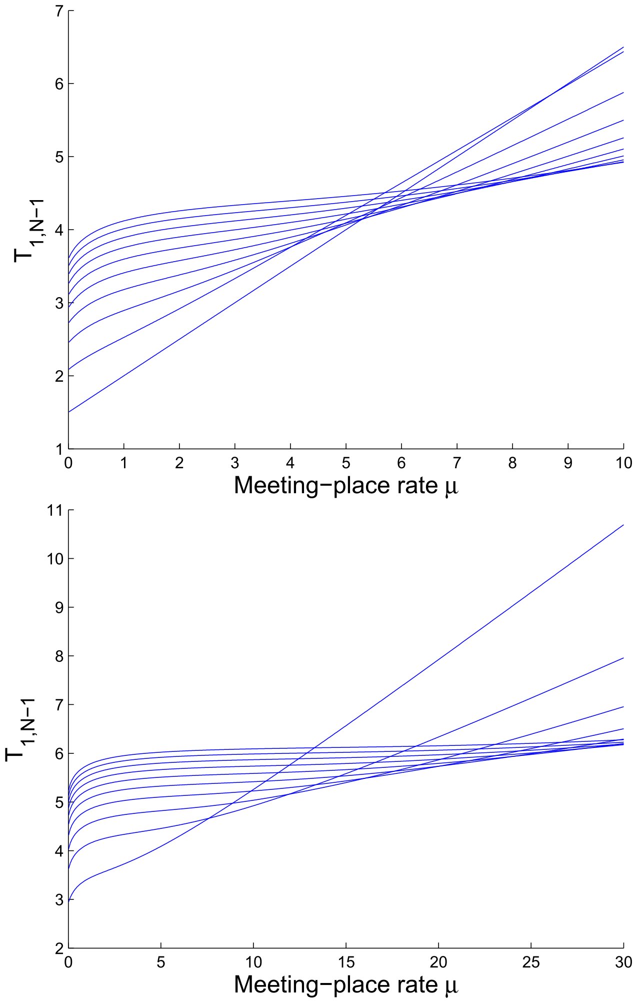
Figure 10 shows the meeting-place rates necessary to achieve a given $T_{1,N-1}$ as a function of N when $\lambda = 1$. We can explicitly compute the intersections of the curves with the axes. The intersection with the μ-axis has $N = 2$, and with $\lambda = 1$, $T_{1,1} = 1.5 + \mu/2$, giving $\mu = 2T_{1,1} - 3$. The intersection with the N-axis has $\mu = 0$, and then the target propagation time equals the harmonic sum $H(N)$. For example, as can be seen in the figure, $H(30) \approx 4$; therefore, the lower curve, which corresponds to $T_{1,N-1} = 4$, hits $\mu = 0$ near $N = 30$. Similarly, the curve above it, which corresponds to $T_{1,N-1} = 4.5$, will hit zero at $H(50) \approx 4.5$.

Figure 8. Expected Number of Informed Agents



Notes. Here $\lambda = 1$. From top down, $N = 50$ and $\mu = 10, 50, 100$.

Figure 9. $T_{1,N-1}$ as a Function of the Meeting-Place Rate μ



Notes. Here $\lambda = 1$. Also, $N = 2, 4, \ldots, 20$ (upper figure) and $N = 10, 20, \ldots, 100$ (lower figure). Increasing N values are attached to graphs from bottom up near zero and from top down for large values of μ.

Remark 8. The curves first decrease from two to three and then increase, eventually decreasing until they hit zero where $H(N) \approx T_{1,N-1}$. Thus, when N is large, we need a small μ to achieve a given target propagation time. But also when N is small, we need a relatively small μ.

4.3. Selecting the Size of the Seed

Section 2.3 considered a process that terminates when a given number of informed agents is reached. In contrast, in this section, we assume that the process starts with a given number of informed agents.

Suppose that we start with a *seed* of k informed agents. We now investigate the effect of the size of the seed on the propagation time.

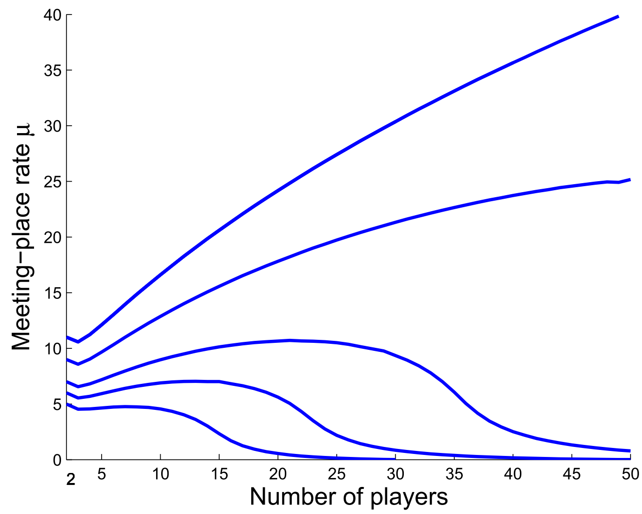
When $\rho \to 0$, it is easy to extend the derivation given in Section 2.1.4. Given a seed of size k, the expected time until all agents are informed is (similar to Equation (6))

$$\sum_{i=k}^{N-1} \frac{1}{N}\left(\frac{1}{N-i} + \frac{1}{i}\right) = \frac{1}{N}\left(\sum_{j=1}^{N-k} \frac{1}{j} + \sum_{i=1}^{N-1} \frac{1}{i}\right) = \frac{1}{N}(H(N-k) + H(N-1)).$$

Figure 11 shows the expected propagation time for $N = 30$ and $N = 50$ and all possible seed sizes $k = 1, \ldots, N$.

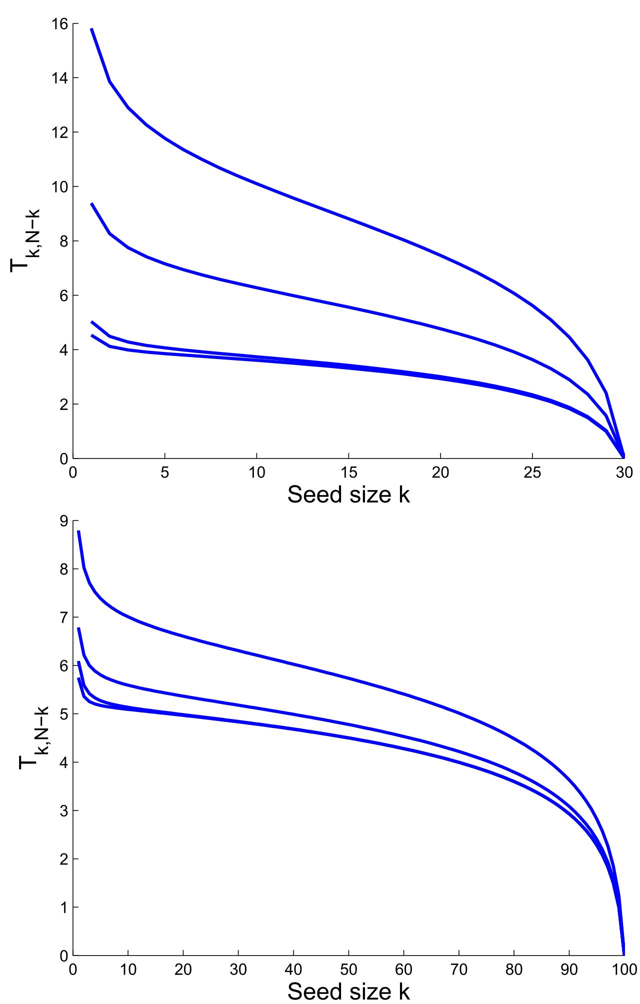
Remark 9. We see from Figure 11 that an increase in the seed size has a strong effect on reducing the propagation time when the seed is either very small or very large, and the effect is quite moderate in between.

Figure 10. Meeting-Place Rate μ Necessary to Achieve a Given Expected Propagation Time $T_{1,N-1}$ As a Function of the Number of Agents N, for $T_{1,N-1} = 4, 4.5, 5, 6, 7$ and $\lambda = 1$



Note. Increasing T values correspond to the graphs from bottom up.

Figure 11. Expected Propagation Time $T_{k,N-k}$ As a Function of Seed Size k



Notes. In the upper figure, $N = 30$, and in the lower figure, $N = 100$. Also, $\mu = 1, 10, 50, 100$ with increasing values from bottom up, and $\lambda = 1$.

5. Concluding Remarks

This paper considers a simple spreading model with mobile agents on a star network that can be reduced to a two-node network. Yet the analysis is far from being straightforward, and the results are not always easy to predict. A compact representation of the transition diagram enables an efficient computation of the expected propagation time. We provide analytical bounds to extreme cases that are complemented by numerical results and sensitivity analyses. For example, increasing the size of the population has two contradicting effects that, as we show, cause the time of propagation to initially increase, then decrease, and eventually increase, with asymptotic behavior similar to the harmonic sum. Related to this, the expected time it takes to inform an additional agent is initially short and then increases, and the probability of informing all agents within a given deadline has an S shape. We also provide insights on the way the process can be affected by changing modeling assumptions such as the initial and final number of informed agents and the contact intensity.

Although the simplicity of our model leads to useful insights, it is of interest to consider extensions to more complicated structures and types of interactions that, in particular, more closely model real-world situations.

Endnote

[1] A natural extension to our model has multiple meeting places and a transition probability matrix for the movements among them. However, we aim at avoiding such complications in order to define a simple model for which we can obtain definite conclusions.

References

Centola D (2010) The spread of behavior in an online social network. *Science* 329(5996):1194–1197.

Colizza V, Vespignani A (2008) Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *J. Theoret. Biol.* 251(3):450–467.

De Arruda GF, Rodrigues FA, Rodríguez P, Cozzo E, Moreno Y (2018) A general Markov chain approach for disease and rumour spreading in complex networks. *J. Complex Networks* 6(2):215–242.

Feola G, Butt A (2017) The diffusion of grassroots innovations for sustainability in Italy and Great Britain: An exploratory spatial data analysis. *Geographical J.* 183(1):16–33.

Chierichetti F, Lattanzi S, Panconesi A (2011) Rumor spreading in social networks. *Theoret. Comput. Sci.* 412(24):2602–2610.

Gilboa-Freedman G, Hassin R (2016) When Markov chains meet: A continuous-time model of network evolution. *Statist. Probability Lett.* 116:131–138.

Groenevelt R, Nain P, Koole G (2005) The message delay in mobile ad hoc networks. *Performance Evaluation* 62:210–228.

Isella L, Stehlé J, Barrat A, Cattuto C, Pinton J-F (2011) What's in a crowd? Analysis of face-to-face behavioral networks. *J. Theoret. Biol.* 271(1):166–180.

Jin M, Liu F, Zhou C (2017) Rumor spreading: A survey. *2nd Internat. Conf. Artificial Intelligence Engrg. Appl.* (AIEA 2017), 263–269.

Manzo G, Gabbriellin S, Roux V, M'Mbogori FN (2018) Complex contagions and the diffusion of innovations: evidence from a small-N study. *J. Archaeological Method Theory* 25(4):1109–1154.

Mascia MB, Mills M (2018) *When conservation goes viral: The diffusion of innovative biodiversity conservation policies and practices*. *Conservation Lett.* 11(3):e12442, https://doi.org/10.1111/conl.12442.

Milling C, Caramanis C, Mannor S, Shakkottai S (2015) Distinguishing infections on different graph topologies. *IEEE Trans. Inform. Theory* 61(6):3100–3120.

Nowzari C, Preciado VM, Pappas GJ (2016) Analysis and control of epidemics. *IEEE Control Systems Magazine* 36(1):26–46.

Pastor-Satorras R, Castellano C, Van Mieghem P, Vespignani A (2015) Epidemic processes in complex networks. *Rev. Modern Phys.* 87(3):925–979.

Rogers EM (1995) *Diffusion of Innovations*, 4th ed. (Free Press, New York).

Sattenspiel L, Dietz K (1995) A structured epidemic model incorporating geographic mobility among regions. *Math. Biosciences* 128(1–2):71–91.

Simpson G, Clifton J (2017) Testing diffusion of innovations theory with data: Financial incentives, early adopters, and distributed solar energy in Australia. *Energy Res. Soc. Sci.* 29:12–22.

Sun H, Cheng R, Xiao X, Yan J, Zheng Y, Qian Y (2018) Maximizing social influence for the awareness threshold model. Pei J, Manolopoulos Y, Sadiq S, Li J, eds. *Database Systems for Advanced Applications*, Lecture Notes in Computer Science, vol. 10827 (Springer, Cham), 491–510.

Vespignani A (2012) Modelling dynamical processes in complex socio-technical systems. *Nature Phys.* 8:32–39.

Xiong H, Wang P, Bobashev G (2018) *Multiple peer effects in the diffusion of innovations on social networks: a simulation study*. *J. Innovation Entrepreneurship* 7(1):1–18.

Zhang X, Neglia G, Kurose J, Towsley D (2007) Performance modeling of epidemic routing. *Comput. Networks* 51(10):2867–2891.