Adv. Appl. Prob. **46,** 560–584 (2014)
Printed in Northern Ireland

CONVERGENCE AND MONOTONICITY FOR A MODEL OF SPONTANEOUS INFECTION AND TRANSMISSION

ERIC FOXALL,* *University of Victoria*

Abstract

A version of the contact process (effectively an SIS model) on a finite set of sites is considered in which there is the possibility of spontaneous infection. A companion process is also considered in which spontaneous infection does not occur from the disease-free state. Monotonicity with respect to parameters and initial data is established, and conditions for irreducibility and exponential convergence of the processes are given. For the spontaneous process, a set of approximating equations is derived, and its properties investigated.

Keywords: Continuous-time Markov process; interacting particle system; SIS model

2010 Mathematics Subject Classification: Primary 60J25
Secondary 92B99

1. Introduction

The contact process is a continuous-time Markov process that was first introduced by Harris [3] in 1974. In Harris's original paper, the setting for the process is the d-dimensional lattice $\mathbb{Z}^d$, and each site is either in state 0 or 1, so that the state space is the set of configurations $\{0, 1\}^{\mathbb{Z}^d}$. Each site x has a set of neighbours $\mathcal{N}(x)$: for example, we might have $\mathcal{N}(x) = \{y \in \mathbb{Z}^d : \|y - x\| = 1\}$ for some norm $\|\cdot\|$. At site x, transition from state 0 to state 1 occurs at rate $\lambda n_1(x)$, where $n_1(x)$ is the number of neighbours of x in state 1 and $\lambda > 0$ is a parameter, and transition from state 1 to state 0 occurs at rate 1. The state of the process at time t is typically denoted by ξ_t, which is a configuration. For each $t \geq 0$, $\xi_t \in \{0, 1\}^{\mathbb{Z}^d}$, and $\xi_t(x)$ specifies the state of site x at time t. Then $(\xi_t)_{t\geq 0}$ specifies a realization of the process. The contact process can be thought of as a model of the spread of an infection: sites in state 1 are infected, and sites in state 0 are healthy.

In this paper we consider a version of the contact process in which the set of sites S is finite, transition rates are allowed to depend on sites, and there is the additional possibility of spontaneous infection. In agreement with the epidemiology literature, the following notation is used. For a site x, α_x is the rate of spontaneous infection (spontaneous transition from state 0 to state 1) at x and γ_x is the rate of recovery (transition from state 1 to state 0). For a pair of sites $x \neq y$, β_{xy} is the rate of transmission from x to y. For this process, the notation ξ_t is used to specify the state of the process at time t. In this case the set of neighbours $\mathcal{N}(y)$ of a site $y \in S$ is the set $\{x \in S : \beta_{xy} > 0\}$. Therefore, at site y at time t, transition from state 0 to state 1 occurs at rate

$$\alpha_y + \sum_{\{x\in\mathcal{N}(y):\, \xi_t(x)=1\}} \beta_{xy}$$

Received 19 December 2012; revision received 23 April 2013.
* Postal address: Department of Mathematics and Statistics, University of Victoria, PO Box 3060 STN CSC, Victoria BC, V8W 3R4, Canada. Email address: e.t.foxall@gmail.com

and transition from state 1 to state 0 occurs at rate γ_y. The basic contact process on the infinite lattice $S = \mathbb{Z}^d$ is obtained by taking $\alpha_x = 0$, $\gamma_x = 1$, and $\beta_{xy} = \beta \mathbf{1}\{y \in \mathcal{N}(x)\}$; this ensures that the process looks the same from any site in S (i.e. is translation invariant), and reflects the fact that the recovery rate can be made equal to 1 by rescaling time, provided the recovery rate is not equal to 0 (if the recovery rate is equal to 0 then we have an essentially different process, in which infected sites remain infected for all time). The parameter values α_x, β_{xy}, and γ_x are nonnegative, but are not in general required to be positive, and it is not required that $\beta_{xy} = \beta_{yx}$. We call a Markov process of this type a *spontaneous process*—terminology also used by Harris. By disallowing infection when all sites are healthy, we obtain what we call an *absorbing process*.

One of the main results of this paper is Theorem 4, whose statement we include here. Suppose that we have an irreducible spontaneous process (see Section 1.2 of [8] for the definition of an irreducible process), and let π denote its unique stationary distribution. Define

$$\lambda = \min_{x \in S} \left(\alpha_x + \gamma_x - \sum_{y \neq x} \beta_{xy} \right),$$

and suppose that $\lambda > 0$. Let P^t denote the transition semigroup for the spontaneous process. Then, for $0 \leq s \leq t$,

$$\rho(\mu P^t, \pi) \leq \mathrm{e}^{-\lambda(t-s)} \, \rho(\mu P^s, \pi).$$

Therefore, λ gives a lower bound on the exponential rate of convergence of the process to its unique stationary distribution. Corollary 2 gives an analogous result for the absorbing process, and indeed much of Section 5 (namely, Theorem 2, Proposition 3, and Theorem 3) is devoted to adapting previously known convergence results to the setting of a conditional distribution.

A summary of the paper is now provided. In Section 2 we describe the graphical construction of any spontaneous (and absorbing) process, and use this construction to prove monotonicity with respect to parameters and initial data. In Section 3 we describe the forward equation for the spontaneous process, and re-express it at the level of individual sites, anticipating the derivation of the reduced equations. In Section 4 we provide a necessary and sufficient condition for irreducibility (or 'conditional irreducibility', defined in that section) in terms of the associated rates, since an irreducibility assumption is required for the results that follow. In Section 5 we use a notion called the transportation distance in order to obtain conditions for exponential convergence of an irreducible (or conditionally irreducible) process to its unique stationary (or quasistationary) distribution. In Section 6 we apply these results to the spontaneous process and absorbing process, to obtain conditions for exponential convergence. In Section 7 we derive the reduced equations as an approximation to the time evolution of the spontaneous process, and we give two proofs of monotonicity, one for monotonicity of the unique steady state with respect to parameters in the case of small interactions, and one for monotonicity of solutions with respect to parameters and initial data. The first proof relies on the implicit function theorem. The second proof involves a characterization of the coefficients in the Taylor series expansion of the reduced equations, in terms of infection paths on the directed graph consisting of the set of sites together with the set of directed transmission links between sites. In Section 8 we consider an example of a spontaneous process on a star graph, and we compute some relevant statistics of the reduced equations, from the perspective of the central site.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

2. Graphical construction and monotonicity

Fix a set of sites S and a set of parameters

$$\Theta = ((\alpha_x)_{x\in S}, (\beta_{xy})_{xy\in\Lambda}, (-\gamma_x)_{x\in S}),$$

where $\Lambda = \{xy \in S \times S : x \neq y\}$; the use of $(-\gamma_x)$ rather than (γ_x) should become clear in Proposition 1. The state space for the process is the set of configurations $\Omega = \{0, 1\}^S$. A configuration is a function that assigns a 0 to a healthy site and a 1 to an infected site. Note that each configuration can be specified by its set of infected sites, and this gives a natural bijection between Ω and the powerset of S. To obtain the spontaneous process, we proceed as follows. To each site x attach a Poisson random variable $(U_t(x))_{t\geq 0}$ of rate α_x and a Poisson random variable $(W_t(x))_{t\geq 0}$ of rate γ_x, and to each ordered pair of distinct sites xy attach a Poisson random variable $(V_t(xy))_{t\geq 0}$ of rate β_{xy}, and let the variables all be independent. Now construct a random graph living on $S \times \mathbb{R}^+$ as follows (an example is given in Figure 1). If $U_t(x)$ has a jump at time t then draw a circle at (x, t) to denote infection. If $W_t(x)$ has a jump at time t then draw a cross to denote recovery. If $V_t(xy)$ has a jump at time t then draw an arrow from (x, t) to (y, t) to denote transmission. Given the configuration at time 0, draw bold lines moving upwards from all initially infected sites, and from all locations where there is a circle, terminating the line if a cross is encountered. If an arrow juts out from a bold line, continue the bold line along the arrow and then vertically upwards. Continue in this way, all the way up the graph; note that not necessarily all transmission, spontaneous infection, and recovery events are 'used'. Then site x is infected at time t if and only if there is a bold line passing through the point (x, t). This defines, on the same probability space, a family $(\xi_t^A)_{t\geq 0}$ of Ω-valued Markov processes, one for each subset A of initially infected sites. To obtain the absorbing process, we make the following additional modification: if T is any time such that all sites are healthy

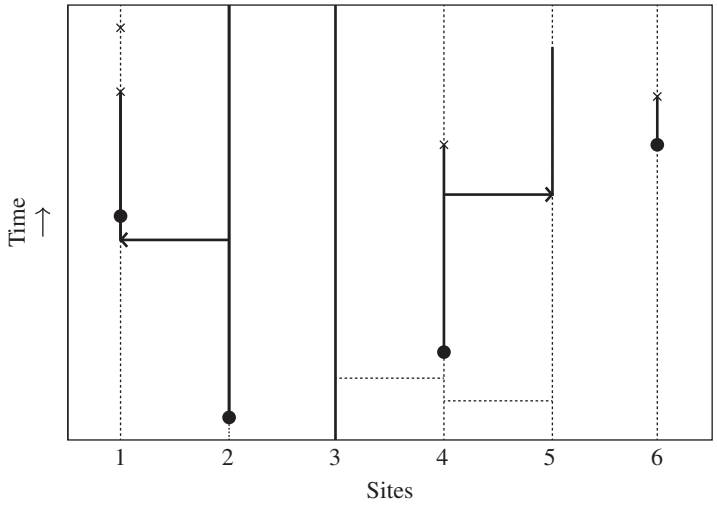

FIGURE 1: An illustration of the graphical construction for the spontaneous process. There are six sites, and time evolves in the upward direction; here, only site 3 is initially infected. Circles denote opportunities for spontaneous infection, crosses denote opportunities for recovery, and arrows denote opportunities for transmission. Infected sites and transmission events are denoted by the bold lines and bold arrows, respectively.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

then, for $t > T$, all the bold lines that have been drawn are erased, since the state remains the same after T.

This construction is called the graphical construction, and is due to Harris [4]; see also [5] or [7] for examples of the graphical construction for other variants of the contact process.

One property is immediate from the graphical construction. For configurations ξ and ξ', use $\xi \vee \xi'$ to denote the configuration for which $(\xi \vee \xi')(x) = \max(\xi(x), \xi'(x))$ for each $x \in S$. If $A = \{x : \xi(x) = 1\}$, and $A' = \{x : \xi'(x) = 1\}$, then $A \cup A' = \{x : (\xi \vee \xi')(x) = 1\}$. Then the process is additive in the sense that, for any sets A and A' of sites, for the above family of processes, we have $\xi_t^A \vee \xi_t^{A'} = \xi_t^{A \cup A'}$ for each $t \geq 0$.

There is another useful property, called monotonicity, which is not hard to prove using the graphical construction. For configurations $\xi^{(1)}$ and $\xi^{(2)}$ and parameters Θ_1 and Θ_2, say that $\xi^{(1)} \leq \xi^{(2)}$ and that $\Theta_1 \leq \Theta_2$ if an inequality holds entrywise. Intuitively, $\xi^{(1)} \leq \xi^{(2)}$ if all sites that are infected in configuration $\xi^{(1)}$ are infected in configuration $\xi^{(2)}$, and $\Theta_1 \leq \Theta_2$ if the parameters in Θ_2 lead to a process in which infection is more rapid, and recovery slower, than the process obtained using parameters in Θ_1.

Proposition 1. *Let $(\xi_t^{(1)})_{t \geq 0}$ and $(\xi_t^{(2)})_{t \geq 0}$ be two realizations of either the spontaneous process or of the absorbing process, with respective parameters $\Theta_1 \leq \Theta_2$. If $\xi_0^{(1)} \leq \xi_0^{(2)}$, there is a common probability space for $(\xi_t^{(1)})_{t \geq 0}$ and $(\xi_t^{(2)})_{t \geq 0}$ on which $\xi_t^{(1)} \leq \xi_t^{(2)}$ for all $t \geq 0$.*

Proof. Parameters are labelled so that $\Theta_i = ((\alpha_{x,i})_{x \in S}, (\beta_{xy,i})_{xy \in \Lambda}, (-\gamma_{x,i})_{x \in S})$ for $i = 1, 2$. Define independent Poisson processes $(U_t(x))$ of rate $\alpha_{x,1}$, $(W_t(x))$ of rate $\gamma_{x,2}$, $(V_t(xy))$ of rate $\beta_{xy,1}$, $(U_t^{(2)}(x))$ of rate $\alpha_{x,2} - \alpha_{x,1} \geq 0$, $(W_t^{(1)}(x))$ of rate $\gamma_{x,1} - \gamma_{x,2} \geq 0$, and $(V_t^{(2)}(xy))$ of rate $\beta_{xy,2} - \beta_{xy,1} \geq 0$. Construct a random graph as before, now using black circles for $(U_t(x))$, red circles for $U_t^{(2)}(x)$, black crosses for $(W_t(x))$, red crosses for $W_t^{(1)}(x)$, black arrows for $(V_t(xy))$, and red arrows for $(V_t^{(2)}(xy))$ (the colour change is just to be able to tell them apart). Run the first process using the initial configuration $\xi_0^{(1)}$, and using the black circles for spontaneous infection, the black and red crosses for recovery, and the black arrows for transmission. Run the second process using the initial configuration $\xi_0^{(2)}$, and using the black and red circles for spontaneous infection, the black crosses for recovery, and the black and red arrows for transmission. Note that if X_t and Y_t are independent Poisson random variables of rates λ_1 and λ_2, then $X_t + Y_t$ is a Poisson random variable of rate $\lambda_1 + \lambda_2$ (see Section 2.4 of [8]), which guarantees that the processes have the correct rates of infection, transmission, and recovery. Now suppose that $\xi_0^{(1)} \leq \xi_0^{(2)}$. Drawing in the bold lines for each process, we observe that every bold line drawn from $t = 0$ in the first process is also a bold line from $t = 0$ in the second process. Also, every spontaneous infection event for the first process is a spontaneous infection event for the second process, every transmission event for the first process is a transmission event for the second process, and every recovery event for the second process is a recovery event for the first process. From this, it follows easily that every bold line for the first process is a bold line for the second process. In other words, $\xi_t^{(1)} \leq \xi_t^{(2)}$ for all $t \geq 0$.

3. Forward equation

There is an equivalent description of the spontaneous process, directly in terms of transition rates on the state space Ω, that gives a differential equation for the distribution over time. First, associate to each set $A \subset S$ the unique element of Ω that is equal to 1 on A and equal to 0 on A^c; this gives a bijection between Ω and the powerset of S. In other words, we have the

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

correspondence

$$\xi \leftrightarrow \{x : \xi(x) = 1\}$$

between configurations ξ and subsets of sites S. For each $y \in A$, there is a transition from A to $A \setminus \{y\}$ at rate γ_y. For each $y \in A^c$, there is a transition from A to $A \cup \{y\}$ at rate $\alpha_y + \sum_{x \in A} \beta_{xy}$. A continuous-time Markov process on a finite state space can be specified by its Q-matrix, whose entries are given by the rate of transition from state to state (see [8, Chapter 2] for a description). In this case the Q-matrix has entries

$$q_{AB} = \begin{cases} \alpha_y + \displaystyle\sum_{x \in A} \beta_{xy} & \text{if } B = A \cup \{y\}, \\ \gamma_y & \text{if } A = B \cup \{y\}, \\ 0 & \text{otherwise,} \end{cases}$$

for $A \neq B$ and

$$q_{AA} = - \sum_{\{B \subset S :\, B \neq A\}} q_{AB}.$$

The absorbing process can be obtained by setting the transition rates $q_{\varnothing\{x\}}$ equal to 0 for each $x \in S$; note that $\varnothing$ denotes the disease-free state in which there are no infected sites. Incidentally, as shown in [8], any continuous-time Markov process that can be described in terms of a Q-matrix can be constructed from the transition rates in the Q-matrix, which is another way of constructing the spontaneous and absorbing processes.

For a process $(\xi_t)_{t \geq 0}$ with a given Q-matrix Q on a finite state space, the transition semigroup $(P^t)_{t \geq 0}$ for the process with entries

$$P^t_{AB} = \mathbb{P}(\xi_t = B \mid \xi_0 = A)$$

is given by $P^t = \mathrm{e}^{Qt}$, a fact that is proved in [8, Chapter 2] and which will be useful later.

If $(\xi_t)_{t \geq 0}$ is a realization of the spontaneous process, the distribution $p(t)$ with entries $p_A(t) = \mathbb{P}(\xi_t = A)$ satisfies the vector differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} p = pQ, \tag{1}$$

which in [8] is called the *forward equation*. Note that $0 \leq p_A(t) \leq 1$ for $A \subset S$ and $t \geq 0$, and that, for each $t \geq 0$, $\sum_{A \subset S} p_A(t) = 1$. For a site x, the probability $p_x(t) = \mathbb{P}(\xi_t(x) = 1)$ of infection at x at time t is related to the values $p_A(t)$ through the equation

$$p_x(t) = \sum_{\{A \subset S :\, x \in A\}} p_A(t)$$

(note that $p_x(t) \neq p_{\{x\}}(t)$, the probability that x is the only infected site). Note that $0 \leq p_x(t) \leq 1$ for $x \in S$ and $t \geq 0$, but that the vector $(p_x(t))_{x \in S}$ is not a distribution, so that $\sum_{x \in S} p_x(t)$ is not in general equal to 1. It can be verified directly from the forward equation and from the transition rates in the Q-matrix that, for the spontaneous process, for each site $y \in S$, $p_y(t)$ satisfies the differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t} p_y(t) = \alpha_y(1 - p_y(t)) + \sum_x \mathbb{P}(\xi_t(x) = 1,\, \xi_t(y) = 0)\beta_{xy} - \gamma_y p_y(t). \tag{2}$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

Intuitively, if we consider configurations in which y is not infected, there is the spontaneous infection rate α_y, and these configurations have probability $1 - p_y(t)$. For each $x \neq y$, considering configurations in which y is not infected but x is infected, there is the additional infection rate β_{xy}, and these configurations have probability $\mathbb{P}(\xi_t(x) = 1,\ \xi_t(y) = 0)$. Considering configurations in which y is infected, y recovers at rate γ_y, and these configurations have probability $p_y(t)$. A formal proof of the validity of (2) uses the following fact.

Lemma 1. *Let $\Gamma \subset \Omega$ be a subset of the state space. Then*

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_{A \in \Gamma} p_A(t) = \sum_{A \in \Gamma,\ B \in \Gamma^{\mathrm{c}}} (p_B(t) q_{BA} - p_A(t) q_{AB}).$$

Proof. From the forward equation, we have, for each $A \subset S$,

$$\frac{\mathrm{d}}{\mathrm{d}t} p_A(t) = \sum_{B \in \Omega} p_B(t) q_{BA} = \sum_{B \neq A} (p_B(t) q_{BA} - p_A(t) q_{AB}),$$

using the fact that $q_{AA} = -\sum_{B \neq A} q_{AB}$. Then, summing over $A \in \Gamma$, note that, for $A, A' \in \Gamma,\ A' \neq A$, the term $p_{A'}(t) q_{A'A} - p_A(t) q_{AA'}$ in $\mathrm{d}p_A(t)/\mathrm{d}t$ is cancelled by the corresponding term $p_A(t) q_{AA'} - p_{A'}(t) q_{A'A}$ in $\mathrm{d}p_{A'}(t)/\mathrm{d}t$, so that it suffices to sum $p_B(t) q_{BA} - p_A(t) q_{AB}$ over $A \in \Gamma$ and $B \in \Gamma^{\mathrm{c}}$, whence the formula follows.

Let $\Gamma_y = \{A \subset S\colon y \in A\}$. Then $B \in \Gamma_y^{\mathrm{c}}$ if and only if $y \notin B$. There is a nonzero rate of transition from $A \in \Gamma_y$ to $B \in \Gamma_y^{\mathrm{c}}$, and from $B \in \Gamma_y^{\mathrm{c}}$ to $A \in \Gamma_y$, exactly when $B = A \setminus \{y\}$. In this case the transition rate from A to B is γ_y, and the transition rate from B to A is $\alpha_y + \sum_{x \in B} \beta_{xy}$. Thus, by Lemma 1,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} p_y(t) &= \sum_{A \in \Gamma_y} p_A(t) \\
&= \sum_{A \in \Gamma_y,\ B \in \Gamma_y^{\mathrm{c}}} (p_B(t) q_{BA} - p_A(t) q_{AB}) \\
&= \alpha_y (1 - p_y(t)) + \sum_x \mathbb{P}(\xi_t(x) = 1,\ \xi_t(y) = 0) \beta_{xy} - \gamma_y p_y(t),
\end{aligned}$$

since $\sum_{B \in \Gamma_y^{\mathrm{c}}} p_B(t) = \mathbb{P}(\xi_t(y) = 0) = (1 - p_y(t))$, $\sum_{\{B \in \Gamma_y^{\mathrm{c}}\colon\, x \in B\}} p_B(t) = \mathbb{P}(\xi_t(x) = 1,\ \xi_t(y) = 0)$ and $\sum_{A \in \Gamma_y} p_A(t) = p_y(t)$, giving (2).

4. Irreducibility

Let Ω^* denote the state space minus the disease-free state. Say that the absorbing process is *conditionally irreducible* if Ω^* is a communicating class for the absorbing process. We obtain necessary and sufficient conditions both for the spontaneous process to be irreducible and for the absorbing process to be conditionally irreducible (see Section 1.2 of [8] for the definitions of irreducibility and a communicating class). We begin with a definition.

Definition 1. For sites x and y, a *spontaneous infection path* is a list $x = x_0 x_1 \cdots x_k = y$ such that $\alpha_{x_0} \beta_{x_0 x_1} \cdots \beta_{x_{k-1} x_k} > 0$. A site y has a spontaneous infection path if there is some x such that there is a spontaneous infection path from x to y; this is in particular the case if $\alpha_y > 0$. For sites x and y, a *transmission path* is a list $x = x_0 x_1 \cdots x_k = y$ such that $\beta_{x_0 x_1} \cdots \beta_{x_{k-1} x_k} > 0$.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

The following lemma gives necessary and sufficient conditions both for irreducibility of the spontaneous process, and conditional irreducibility of the absorbing process.

Lemma 2. *A necessary and sufficient condition for the spontaneous process to be irreducible is that, for each site y, $\gamma_y > 0$ and y has a spontaneous infection path. If the absorbing process is defined on a set of at least two sites then a necessary and sufficient condition for the absorbing process to be conditionally irreducible is that, for each site y, $\gamma_y > 0$ and either y has a spontaneous infection path or, for each $x \neq y$, there is a transmission path from x to y.*

If the spontaneous process is irreducible then it has a unique stationary distribution π. If the absorbing process is conditionally irreducible then it has a unique quasistationary distribution ζ supported on Ω^*.

Proof. Consider first the spontaneous process. If there exists $y \in S$ such that $\gamma_y = 0$ for some $y \in S$ then state $\varnothing$ is inaccessible from state $\{y\}$. If there exists $y \in S$ that has no spontaneous infection path then state $\{y\}$ is inaccessible from state $\varnothing$. If both conditions are satisfied then, for any subset of sites A and any initial state, there is a positive probability that first all sites become infected, and then the sites in $S \setminus A$ recover; in other words, every state is accessible from every other state, which is what it means to be irreducible.

For a finite state Markov Chain, irreducibility implies that there exists a unique stationary distribution π given by the equation

$$\pi Q = 0.$$

See [8, Section 3.5] for the proof of existence and uniqueness of the stationary distribution.

Consider now the absorbing process. If there exists $y \in S$ such that $\gamma_y = 0$ then, for $x \neq y$, state $\{x\}$ is inaccessible from state $\{y\}$. If there exists $y \in S$ that has no spontaneous infection path, and there is $x \neq y$ such that there is no transmission path from x to y, then state $\{y\}$ is inaccessible from state $\{x\}$. If both conditions are satisfied then, for any subset of sites A, from any state with at least one infected site there is a positive probability that first all sites become infected, and then the sites in $S \setminus A$ recover; in other words, every state in Ω^* is accessible from every other state in Ω^*. Since the disease-free state does not have access to any state in Ω^*, it follows that Ω^* is a communicating class; in other words, the absorbing process is conditionally irreducible.

Since, for any $t \geq 0$, the restriction of P^t to Ω^* is a primitive matrix, the Perron–Frobenius theory [1] guarantees a unique positive eigenvector ζ of the restricted transition matrix, with a positive eigenvalue, such that $\sum_{x \in \Omega^*} \zeta(x) = 1$, and this is what defines a quasistationary distribution.

5. Transportation distance

In the case where the processes are irreducible or conditionally irreducible, the material covered in this section will help to provide sufficient conditions for the exponentially fast convergence of the spontaneous process and the absorbing process. The convergence itself is discussed in the next section.

Let $G = (V, E)$ be a finite connected graph with undirected edges; here connected means that, for any two vertices, there is a path running along the edges from one vertex to the other vertex. For the spontaneous and absorbing processes, G corresponds to the graph with vertex set $V = \Omega = \{0, 1\}^S$ and edge set corresponding to the set of pairs of configurations that differ in the state of exactly one site. To each path assign a length that corresponds to the number of edges on the path. For vertices u and v, let $\ell(u, v)$ be the length of any shortest path between

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

u and v. Then $\ell(u, v)$ is a metric on G, which is called the path metric. Let $\mathrm{Diam}(G)$ denote the graph diameter of G, defined by

$$\mathrm{Diam}(G) = \max_{(x,y)\in V\times V} \ell(x, y).$$

For a finite or countably infinite set A, $\mathcal{P}(A)$ is used to denote the set of probability measures on A, that is, functions $\mu\colon A \to \mathbb{R}^+$ such that $\mu(x) \geq 0$ for each $x \in A$ and $\sum_{x\in A} \mu(x) = 1$. The words distribution and probability measure are used here interchangeably. For a pair of distributions μ and ν on V, a *coupling* is a distribution η on $V \times V$ with marginals μ and ν, that is, $\mu(x) = \sum_{y\in V} \eta(x, y)$ and $\nu(y) = \sum_{x\in V} \eta(x, y)$.

If X and Y are a pair of random variables defined on a common probability space and having respective distributions μ and ν, then the joint distribution is a coupling of μ and ν. Conversely, any coupling of μ and ν defines a pair of random variables X and Y, with respective distributions μ and ν, on a common probability space. Therefore, it is possible to think of couplings either in terms of distributions, or in terms of random variables.

A coupling η of μ and ν gives a rule that redistributes probability mass in such a way that μ is sent to ν, the quantity $\eta(x, y)$ being the amount of probability mass at x that is moved to y. The condition $\sum_{y\in V} \eta(x, y) = \mu(x)$ says that η sends the amount $\mu(x)$ from the vertex x to the vertices of the graph in some way, and the condition $\sum_{x\in V} \eta(x, y) = \nu(y)$ says that η sends the amount $\nu(y)$ from the vertices of the graph to the vertex y in some way.

The *expected distance* $\mathbb{E}_\eta \ell$ for η, defined as

$$\mathbb{E}_\eta \ell = \sum_{(x,y)\in V\times V} \eta(x, y)\ell(x, y),$$

corresponds to the average distance travelled by the probability mass in sending μ to ν, assuming that mass travels along a shortest path. A natural question to ask is what is the most efficient way of sending μ to ν, in terms of expected distance. This is called the *transportation distance* $\rho(\mu, \nu)$ between μ and ν and is given by

$$\rho(\mu, \nu) = \inf_{\eta\in\mathcal{C}} \mathbb{E}_\eta \ell,$$

where $\mathcal{C}(\mu, \nu)$ is the set of couplings of μ and ν. As it turns out, ρ is a metric on the set of distributions on V. Nonnegativity of ρ follows from the fact that the path metric is nonnegative and from the fact that couplings are by definition nonnegative functions. Symmetry of ρ follows from the fact that the path metric is symmetric and from the fact that θ is a coupling of μ and ν if and only if the function η defined by $\eta(x, y) = \theta(y, x)$ for $(x, y) \in V \times V$ is a coupling of ν and μ. For a proof of the triangle inequality for ρ, see [6, Lemma 14.3].

In the topology defined by the total variation metric (see [6] for a definition of the total variation metric), $\mathcal{C}(\mu, \nu)$ is a compact subset of the set of distributions on $V \times V$. Therefore, there exists $\theta \in \mathcal{C}(\mu, \nu)$ such that $\mathbb{E}_\theta \ell = \rho(\mu, \nu)$. Any such θ is called an *optimal coupling*.

As shown in the following lemma, the transportation distance is Lipschitz continuous; continuity of the transportation distance is used in Proposition 3 below, but Lipschitz continuity is easier to prove.

Lemma 3. *For the transportation distance ρ and distributions μ, ν, μ_1, and ν_1,*

$$|\rho(\mu, \nu) - \rho(\mu_1, \nu_1)| \leq C \max(\|\mu - \mu_1\|_\infty, \|\nu - \nu_1\|_\infty),$$

where $C = |V|^2 \,\mathrm{Diam}(G)$ and $\|\cdot\|_\infty$ is defined on $\mathbb{R}^V$ by

$$\|u\|_\infty = \max_{x\in V} |u(x)|.$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

Proof. Let

$$\delta = \max(\|\mu - \mu_1\|_\infty, \|\nu - \nu_1\|_\infty),$$

and let η be an optimal coupling of μ and ν. Define $\underline{\eta}$ on $V \times V$ by

$$\underline{\eta}(x, y) = \max(\eta(x, y) - \delta, 0),$$

and observe that $0 \leq \eta(x, y) - \underline{\eta}(x, y) \leq \delta$ for each $(x, y) \in V \times V$. Let $\varepsilon = 1 - \sum_{(x,y)\in V\times V} \underline{\eta}(x, y)$. Then, summing the last inequality over $(x, y) \in V \times V$, it follows that $0 \leq \varepsilon \leq |V|^2\delta$. Define $\underline{\mu}$ and $\underline{\nu}$ by

$$\underline{\mu}(x) = \sum_{y\in V} \underline{\eta}(x, y), \qquad \underline{\nu}(y) = \sum_{x\in V} \underline{\eta}(x, y).$$

Let $x \in V$. Then, if $\eta(x, y) \geq \delta$ for some $y \in V$ then $\underline{\mu}(x) \leq \mu(x) - \delta$, and if $\eta(x, y) < \delta$ for each $y \in V$ then $\underline{\mu}(x) = 0$. Since $|\mu_1(x) - \mu(x)| \leq \delta$ and $\mu_1(x) \geq 0$, it follows that $\mu_1(x) \geq \underline{\mu}(x)$. In the same way $\nu_1(x) \geq \underline{\nu}(x)$.

Define θ on $V \times V$ by

$$\theta(x, y) = \underline{\eta}(x, y) + \frac{1}{\varepsilon}(\mu_1(x) - \underline{\mu}(x))(\nu_1(y) - \underline{\nu}(y)).$$

Then θ is nonnegative, and a calculation shows that θ is a coupling of μ_1 and ν_1. Note that

$$\sum_{x\in V}(\mu_1(x) - \underline{\mu}(x)) = \sum_{y\in V}(\nu_1(y) - \underline{\nu}(y)) = 1 - \sum_{(x,y)\in V\times V} \underline{\eta}(x, y) = \varepsilon,$$

so that

$$\sum_{(x,y)\in V\times V} (\mu_1(x) - \underline{\mu}(x))(\nu_1(y) - \underline{\nu}(y)) = \varepsilon^2.$$

Taking $\sum_{(x,y)\in V\times V} \theta(x, y)\ell(x, y)$ and using the fact that $\ell(x, y) \leq \mathrm{Diam}(G)$,

$$\mathbb{E}_\theta \ell \leq \mathbb{E}_{\underline{\eta}} \ell + \varepsilon\,\mathrm{Diam}(G) \leq \mathbb{E}_\eta \ell + |V|^2\delta\,\mathrm{Diam}(G),$$

recalling that $\varepsilon \leq |V|^2\delta$ and observing that $\underline{\eta} \leq \eta$. Since η is optimal,

$$\rho(\mu_1, \nu_1) \leq \mathbb{E}_\theta \ell \leq \rho(\mu, \nu) + |V|^2\delta\,\mathrm{Diam}(G).$$

Reversing the roles of μ_1, ν_1 and μ, ν completes the proof.

For $x \in V$, let $m(x) = \min(\mu(x), \nu(x))$. For $\eta \in \mathcal{C}(\mu, \nu)$, note that $\eta(x, x) \leq m(x)$. As shown in the following lemma, there is an optimal coupling for which equality is achieved, simultaneously for all $x \in V$.

Lemma 4. *Let μ and ν be distributions on V. There is an optimal coupling θ so that, for each $x \in V$, $\theta(x, x) = m(x)$.*

Proof. Let η be an optimal coupling, and let $x \in V$. Set $\delta = m(x) - \eta(x, x)$, and suppose that $\delta > 0$. Set $a = \sum_{u\neq x} \eta(u, x)$ and, $b = \sum_{v\neq x} \eta(x, v)$, and define θ as follows:

- $\theta(x, x) = m(x) = \eta(x, x) + \delta$;
- $\theta(u, x) = \eta(u, x) - \delta a^{-1}\eta(u, x),\ u \neq x$;

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

- $\theta(x, v) = \eta(x, v) - \delta b^{-1}\eta(x, v),\ v \neq x;$
- $\theta(u, v) = \eta(u, v) + \delta(ab)^{-1}\eta(u, x)\eta(x, v),\ u \neq x$ and $v \neq x.$

Note that

$$\sum_{v\neq x}\theta(x, v) = \sum_{v\neq x}\eta(x, v) - \delta.$$

Then

$$\sum_{v\in V}\theta(x, v) = \sum_{v\in V}\eta(x, v) = \mu(x)$$

and

$$\sum_{v\neq x}\theta(u, v) = \sum_{v\neq x}\eta(u, v) + \delta a^{-1}\eta(u, x),$$

so

$$\sum_{v\in V}\theta(u, v) = \sum_{v\in V}\eta(u, v) = \mu(u)$$

when $u \neq x$. Analogous computations reveal that $\sum_{u\in V}\theta(u, x) = \nu(x)$ and that $\sum_{u\in V}\theta(u, v) = \nu(v)$ for $v \neq x$, which proves that θ is a coupling of μ and ν.

To show that θ is an optimal coupling, first note that $\ell(x, x) = 0$, so that $\theta(x, x)\ell(x, x) = \eta(x, x)\ell(x, x) = 0$. We consider the remaining terms of the sum in three parts. For the first two parts, we have

$$\sum_{u\neq x}\theta(u, x)\ell(u, x) = \sum_{u\neq x}\eta(u, x)\ell(u, x) - \delta a^{-1}\sum_{u\neq x}\eta(u, x)\ell(u, x)$$

and

$$\sum_{v\neq x}\theta(x, v)\ell(x, v) = \sum_{v\neq x}\eta(x, v)\ell(x, v) - \delta b^{-1}\sum_{v\neq x}\eta(x, v)\ell(x, v).$$

For the remaining part, let $U = V \setminus \{x\}$. Then $\sum_{(u,v)\in U\times U}\theta(u, v)\ell(u, v)$ is equal to

$$\sum_{(u,v)\in U\times U}\eta(u, v)\ell(u, v) + \delta(ab)^{-1}\sum_{(u,v)\in U\times U}\eta(u, x)\eta(x, v)\ell(u, v).$$

Since $\ell(u, v) \leq \ell(u, x) + \ell(x, v)$, it follows that

$$\sum_{(u,v)\in U\times U}\eta(u, x)\eta(x, v)\ell(u, v) \leq \sum_{(u,v)\in U\times U}\eta(u, x)\eta(x, v)(\ell(u, x) + \ell(x, v)).$$

The right-hand side is equal to

$$\sum_{v\neq x}\eta(x, v)\sum_{u\neq x}\eta(u, x)\ell(u, x) + \sum_{u\neq x}\eta(u, x)\sum_{v\neq x}\eta(x, v)\ell(x, v),$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

which is just $b\sum_{u\neq x}\eta(u,x)\ell(u,x)+a\sum_{v\neq x}\eta(x,v)\ell(x,v)$, so

$$\begin{aligned}\sum_{(u,v)\in U\times U}\theta(u,v)\ell(u,v) &\leq \sum_{(u,v)\in U\times U}\eta(u,v)\ell(u,v)\\ &\quad+\delta(ab)^{-1}\left(b\sum_{u\neq x}\eta(u,x)\ell(u,x)+a\sum_{v\neq x}\eta(x,v)\ell(x,v)\right)\\ &= \sum_{(u,v)\in U\times U}\eta(u,v)\ell(u,v)\\ &\quad+\delta a^{-1}\sum_{u\neq x}\eta(u,x)\ell(u,x)+\delta b^{-1}\sum_{v\neq x}\eta(x,v)\ell(x,v).\end{aligned}$$

Combining the three parts, from the resulting cancellation we have

$$\mathbb{E}_\theta\ell=\sum_{(u,v)\in V\times V}\theta(u,v)\ell(u,v)\leq\sum_{(u,v)\in V\times V}\eta(u,v)\ell(u,v)=\rho(\mu,\nu),$$

so θ is an optimal coupling. Note that $\theta(y,y)\geq\eta(y,y)$ for all $y\in V$. Therefore, repeating the procedure for each $x\in V$ establishes the result.

In what follows, all optimal couplings are assumed to satisfy the property stated in Lemma 4.

Let P be the transition matrix of a discrete-time Markov chain with state space Ω that corresponds to the vertices of G. The following theorem gives a sufficient condition for P to contact the transportation distance.

Theorem 1. (Bubley and Dyer [2].) *Suppose that there exists a $\lambda>0$ such that*

$$\rho(P(x,\cdot),P(y,\cdot))\leq\mathrm{e}^{-\lambda}$$

whenever $\ell(x,y)=1$. Then, for any pair of distributions μ and ν on Ω,

$$\rho(\mu P,\nu P)\leq\mathrm{e}^{-\lambda}\rho(\mu,\nu).$$

A proof of this theorem is given in [6, Proof of Theorem 14.6]. A quick examination of the proof reveals that the following slightly stronger result also holds.

Proposition 2. *Let $\Gamma\subset\Omega$ be a subset of the state space such that the subgraph of G induced by the states belonging to Γ is connected, and which is such that, for each $x,y\in\Gamma$, there is a shortest path from x to y that is contained in Γ. Suppose that there exists a $\lambda>0$ such that, for all $x,y\in\Gamma$,*

$$\rho(P(x,\cdot),P(y,\cdot))\leq\mathrm{e}^{-\lambda}$$

whenever $\ell(x,y)=1$. Then, for any pair of distributions μ and ν supported on Γ,

$$\rho(\mu P,\nu P)\leq\mathrm{e}^{-\lambda}\rho(\mu,\nu).$$

For a function $f:\mathbb{R}^+\to\mathbb{R}$, define the right-hand upper derivative Df as

$$Df(t)=\lim_{h\to0^+}\frac{1}{h}(f(t+h)-f(t)).$$

Suppose now that $(P^t)_{t\geq0}$ is the transition semigroup of a continuous-time Markov chain with state space Ω. The following corollary is a continuous-time version of Proposition 2.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

Corollary 1. *Let Γ be as in Proposition 2. Suppose that there exists a $\lambda > 0$ such that, for all $x, y \in \Gamma$,*

$$\rho(P^h(x,\cdot), P^h(y,\cdot)) \le 1 - \lambda h + o(h)$$

for small $h > 0$, whenever $\ell(x, y) = 1$. Then, for any pair of distributions μ and ν supported on Γ and any $t \ge 0$,

$$D\,\rho(\mu P^t, \nu P^t) \le -\lambda\,\rho(\mu P^t, \nu P^t).$$

Proof. Let $\varepsilon > 0$. There exists $\tau > 0$ such that $o(h) \le \varepsilon h$ for $h \in [0, \tau]$. Apply Proposition 2 to the distributions μP^t and νP^t on Γ, using the transition matrix P^h, to find that

$$\rho(\mu P^{t+h}, \nu P^{t+h}) \le (1 - (\lambda - \varepsilon)h)\,\rho(\mu P^t, \nu P^t).$$

Subtracting $\rho(\mu P^t, \nu P^t)$ from both sides, diving by h, and taking the lim sup as $h \to 0^+$,

$$D\,\rho(\mu P^t, \nu P^t) \le -(\lambda - \varepsilon)\,\rho(\mu P^t, \nu P^t).$$

Since $\varepsilon > 0$ is arbitrary, the result follows.

Let $x^* \in \Omega$ be a distinguished state, and let $\Omega^* = \Omega \setminus \{x^*\}$. For a distribution μ on Ω with $\mu(x^*) < 1$, let μ^* be the corresponding conditional distribution on Ω^*, that is, the distribution on Ω^* defined by

$$\mu^*(x) = \mu(x)(1 - \mu(x^*))^{-1}$$

for $x \in \Omega^*$. The following estimate relates the transportation distance of distributions and their corresponding conditional distributions.

Theorem 2. *Let μ and ν be a pair of distributions on Ω with $\mu(x^*) < 1$ and $\nu(x^*) < 1$. Then*

$$\rho(\mu^*, \nu^*) \le \frac{1}{1 - m(x^*)}[\rho(\mu, \nu) + |\mu(x^*) - \nu(x^*)|\,\mathrm{Diam}(G)].$$

If $\mu(x^*) = \nu(x^*)$ then

$$\rho(\mu^*, \nu^*) = \frac{1}{1 - m(x^*)}\,\rho(\mu, \nu).$$

Proof. Set $a = \mu(x^*)$ and $b = \nu(x^*)$. Since ρ is symmetric and the map $\eta \mapsto \theta$ given by $\theta(x, y) = \eta(y, x)$ is a bijection from the set of couplings of μ and ν to the set of couplings of ν and μ, there is no loss of generality in supposing that $m(x^*) = b$. Let η be an optimal coupling. By Lemma 4 we may suppose that $\eta(x^*, x^*) = m(x^*) = b$. Since $b = \nu(x^*) = \sum_{x \in S}\eta(x, x^*)$ and η is nonnegative with $\eta(x^*, x^*) = b$, it follows that $\eta(x, x^*) = 0$ for $x \in \Omega^*$. Define θ on $\Omega^* \times \Omega^*$ as follows. For $(x, y) \in \Omega^* \times \Omega^*$, let

$$\theta(x, y) = (1 - b)^{-1}\eta(x, y) + \frac{1}{(1 - a)(1 - b)}\mu(x)\eta(x^*, y).$$

A straightforward calculation shows that $\theta(x, y)$ is a coupling of μ^* and ν^*, and that $\sum_{(x,y)\in\Omega^*\times\Omega^*}\theta(x, y)\ell(x, y)$ is not larger than the right-hand side of the inequality in the statement of the theorem, noting that

$$\sum_{y \in \Omega^*} \eta(x^*, y) = \mu(x^*) - \eta(x^*, x^*) = a - b$$

and that $\ell(x, y) \le \mathrm{Diam}(G)$.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

Suppose that $\mu(x^*) = \nu(x^*) = m(x^*)$. Let $\mathcal{C}$ denote the set of couplings θ of μ and ν that satisfy $\theta(x^*, x^*) = m(x^*)$, and let $\mathcal{C}^*$ denote the set of couplings of μ^* and ν^*. Since $\sum_{y\in\Omega}\theta(x^*, y) = \mu(x^*) = m(x^*)$, $\sum_{x\in\Omega}\theta(x, x^*) = \nu(x^*) = m(x^*)$ and $\theta \geq 0$ for any coupling θ of μ and ν, if $\theta \in \mathcal{C}$ then $\theta(x^*, x) = \theta(x, x^*) = 0$ for $x \in \Omega^*$. Therefore, if $\theta \in \mathcal{C}$ then the function θ^* defined by

$$\theta^*(x, y) = (1 - m(x^*))^{-1}\theta(x, y), \qquad (x, y) \in \Omega^* \times \Omega^*,$$

is a coupling of μ^* and ν^*. Since the map from $\mathcal{C}$ to $\mathcal{C}^*$ given by $\theta \mapsto \theta^*$ is $1:1$ and onto, it is a bijection of $\mathcal{C}$ with $\mathcal{C}^*$. Moreover, for $\theta \in \mathcal{C}$,

$$\mathbb{E}_{\theta^*}\ell = c\mathbb{E}_\theta\ell,$$

where $c = (1 - m(x^*))^{-1}$ is a constant, which implies that θ^* is an optimal coupling of μ^* and ν^* if and only if θ is an optimal coupling of μ and ν. Since η is an optimal coupling that satisfies $\eta(x^*, x^*) = m(x^*)$, the result follows.

Let $(P^t)_{t\geq 0}$ be the transition semigroup of a continuous-time Markov chain on Ω. Suppose that x^* is an absorbing state for $(P^t)_{t\geq 0}$, that is, the distribution δ_{x^*} supported on x^* satisfies $\delta_{x^*}P^t = \delta_{x^*}$. Let ζ be a *quasistationary distribution* supported on Ω^* with *decay rate* $v \geq 0$, that is, $\mathrm{d}\log(1 - (\zeta P^t)(x^*))/\mathrm{d}t = -v$ and $(\zeta P^t)^* = \zeta$ for $t \geq 0$. Note that

$$v = -\frac{1}{(\zeta P^t)(\Omega^*)}\frac{\mathrm{d}}{\mathrm{d}t}(\zeta P^t)(\Omega^*),$$

so v measures the rate at which mass leaks out from Ω^*, starting from ζ. Also, note that the condition $(\zeta P^t)^* = \zeta$ for $t \geq 0$ is equivalent to having $(\zeta P^t)(x) = (1 - (\zeta P^t)(x^*))\zeta(x)$ for $x \in \Omega^*$.

For a distribution μ on Ω, set $\mu_t = \mu P^t$ and $a_t = \mu_t(x^*)$. Define the distribution ν_0 on Ω by $\nu_0(x^*) = a_0$ and $\nu_0(x) = (1 - a_0)\zeta(x)$ for $x \neq x^*$, and, for $t \geq 0$, set $\nu_t = \nu_0 P^t$ and $b_t = \nu_t(x^*)$. Let s be such that $(\zeta P^s)(x^*) = a_0$. Then $\nu_0(x) = (\zeta P^s)(x)$ for all x, which implies that $\nu_t = \zeta P^{t+s}$ for $t \geq 0$, and, in particular, that $\nu_t^* = \zeta$ and $\mathrm{d}\log(1 - b_t)/\mathrm{d}t = -v$. Set $u_t = -\mathrm{d}\log(1 - a_t)/\mathrm{d}t$; note that

$$u_t = -\frac{1}{\mu_t(\Omega^*)}\frac{\mathrm{d}}{\mathrm{d}t}\mu_t(\Omega^*),$$

so u_t measures the rate at which mass leaks out from Ω^*, starting from μ. The following estimate is used in the proof of Theorem 3.

Proposition 3. *Let $f(t) = \rho(\mu_t, \nu_t)$, let $g(t) = \rho(\mu_t^*, \nu_t^*) = \rho(\mu_t^*, \zeta)$, and let u_t and v be as above. Suppose that the function $t \mapsto (P^t)_{t\geq 0}$ is continuous. Then*

$$Dg(0) \leq \frac{1}{1 - a_0}[Df(0) + \min(u_0, v)f(0)] + |u_0 - v|\,\mathrm{Diam}(G).$$

Proof. From Theorem 2, since $a_0 = b_0$,

$$\rho(\mu_0^*, \nu_0^*) = \frac{1}{1 - a_0}\rho(\mu_0, \nu_0), \tag{3}$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

and, for $h > 0$,

$$\rho(\mu_h^*, \nu_h^*) \le \frac{1}{1 - \min(a_h, b_h)}[\rho(\mu_h, \nu_h) + |a_h - b_h| \operatorname{Diam}(G)]. \tag{4}$$

Setting

$$m_h = \frac{1}{1 - \min(a_h, b_h)} = \min\left(\frac{1}{1 - a_h}, \frac{1}{1 - b_h}\right),$$

the right-hand side of (4) can be rewritten as

$$\frac{1}{1 - a_0}\rho(\mu_h, \nu_h) + \left(m_h - \frac{1}{1 - a_0}\right)\rho(\mu_h, \nu_h) + m_h|a_h - b_h| \operatorname{Diam}(G). \tag{5}$$

Since a_t and b_t are nondecreasing in t, $a_0 = b_0$, and the function $1/(1 - x)$ is monotone increasing on $(0, 1)$,

$$\lim_{h \to 0^+} \frac{1}{h}\left(m_h - \frac{1}{1 - a_0}\right) = \min\left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{1 - a_t}\bigg|_{t=0}, \frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{1 - b_t}\bigg|_{t=0}\right) = \frac{1}{1 - a_0}\min(u_0, v),$$

and, since $a_0 = b_0$,

$$\lim_{h \to 0^+} \frac{|a_h - b_h|}{h} = \left|\frac{\mathrm{d}}{\mathrm{d}t}(a_t - b_t)\right|\bigg|_{t=0} = (1 - a_0)|u_0 - v|.$$

Recall that $\nu_t^* = \zeta$ for $t \ge 0$. Subtract (3) from (4) (with the right-hand side as in (5)), divide by h, and take $\limsup_{h \to 0^+}$ to obtain the desired result. Note that continuity of ρ and of P^t are assumed; continuity of P^t is a hypothesis, and continuity of ρ follows from Lemma 3.

Combining Corollary 1 and Proposition 3, the following estimate is obtained.

Theorem 3. *Suppose that, for $\lambda > 0$, the transition semigroup $(P^t)_{t \ge 0}$ satisfies the conditions of Corollary 1, and let μ_t, $f(t)$, $g(t)$, etc. be as in Proposition 3. Let*

$$r(t) = \lambda - \min(u_t, v), \qquad c(t) = |u_t - v| \operatorname{Diam}(G),$$

and let

$$\phi(s, t) = \exp\left(-\int_s^t r(\tau)\,\mathrm{d}\tau\right).$$

Then, for $t \ge 0$,

$$Dg(t) \le -r(t)g(t) + c(t),$$

and, for $0 \le s \le t$,

$$g(t) \le \phi(s, t)g(s) + \int_s^t \phi(\tau, t)c(\tau)\,\mathrm{d}\tau.$$

Proof. Applying Corollary 1, $Df(0) \le -\lambda f(0)$, and applying the second equations in Theorem 2, $f(0) = (1 - a_0)g(0)$; so, from Proposition 3,

$$Dg(0) \le [\min(u_0, v) - \lambda]g(0) + |u_0 - v| \operatorname{Diam}(G).$$

Replacing μ with μ_t in the discussion leading up to Proposition 3 reveals that

$$Dg(t) \le [\min(u_t, v) - \lambda]g(t) + |u_t - v| \operatorname{Diam}(G) \quad \text{for } t \ge 0,$$

that is,

$$Dg(t) \le -r(t)g(t) + c(t) \quad \text{for } t \ge 0.$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

Therefore, for $0 \le s \le t$, $g(t)$ is bounded above by the solution to the differential equation

$$y' = -ry + c, \qquad y(s) = g(s),$$

which proves the second inequality.

In words, the result of Theorem 3 says that we have exponential convergence of $g(t)$ at rate $\lambda - \min(u_t, v)$, up to the constant penalty $|u_t - v| \operatorname{Diam}(G)$. Therefore, if the full system converges at rate at least $\lambda > v$ and mass is leaking out of Ω^* at a rate that is equal to the quasistationary rate, then we have exponential convergence at rate at least $\lambda - v$. If a_t and b_t are identically 0, that is, if $\mu_t^* = \mu_t$ and $\nu_t^* = \nu_t$ for $t \ge 0$, the result simplifies to the statement $g(t) \le \mathrm{e}^{-\lambda(t-s)} g(s)$.

6. Convergence

Recall that, for the spontaneous and absorbing processes, there is a natural connected (undirected) graph $G = (V, E)$ with vertex set $V = \Omega = \{0, 1\}^S$ and edge set E given by the set of pairs of states that differ in the label of exactly one site. For the spontaneous process, any sets of sites $A \subset S$ and $B \subset S$, and any $t \ge 0$, the graphical construction defines a coupling of the distributions of ξ_t^A and ξ_t^B. Let ℓ denote the path metric. We have the following estimate for the expected distance between adjacent pairs of states after a short time, in the coupling given by the graphical construction.

Proposition 4. *Let $A \subset B \subset S$ be such that $B = A \cup \{x\}$ for some $x \in S$, and consider the spontaneous process. With respect to the coupling θ_t given by the graphical construction, for small $t > 0$, the expected distance $\mathbb{E}_{\theta_t} \ell$ between ξ_t^A and ξ_t^B is given by*

$$\mathbb{E}_{\theta_t} \ell = 1 + \Bigg(\sum_{y \in A^{\mathrm{c}}} \beta_{xy} \Bigg) t - (\alpha_x + \gamma_x) t + O(t^2),$$

where A^{c} denotes the complement of A.

Proof. The graphical construction defines $(\xi_t^A)_{t \ge 0}$ and $(\xi_t^B)_{t \ge 0}$ on the same probability space. Let τ denote the time of the first event. If site x becomes infectious in (ξ_τ^A), or recovers in (ξ_τ^B), then $\xi_\tau^A = \xi_\tau^B$; this event occurs in the interval $[0, t]$ with probability $\alpha_x t + \gamma_x t + O(t^2)$. If site x causes a subsequent infection then ξ_τ^A and ξ_τ^B differ at two sites, and this event occurs in the interval $[0, t]$ with probability $\sum_{y \in A^{\mathrm{c}}} \beta_{xy} t + O(t^2)$. For any other event, ξ_τ^A and ξ_τ^B differ at a single site. The probability of two or more events occurring in the interval $[0, t]$ is of $O(t^2)$. Let θ_t denote the coupling of ξ_t^A and ξ_t^B given by the joint distribution. Then

$$\begin{aligned}
\mathbb{E}_{\theta_t} \ell &= \sum_{n=0}^{\infty} n \mathbb{P}(\ell = n) \\
&= \sum_{n=0}^{2} n \mathbb{P}(\ell = n) + O(t^2) \\
&= 1[1 - (\mathbb{P}(\ell = 0) + \mathbb{P}(\ell = 2))] + 2\mathbb{P}(\ell = 2) + O(t^2) \\
&= 1\Bigg[1 - \Bigg(\sum_{y \in A^{\mathrm{c}}} \beta_{xy} + \alpha_x + \gamma_x \Bigg) t \Bigg] + 2 \Bigg(\sum_{y \in A^{\mathrm{c}}} \beta_{xy} \Bigg) t + O(t^2) \\
&= 1 + \Bigg(\sum_{y \in A^{\mathrm{c}}} \beta_{xy} \Bigg) t - (\alpha_x + \gamma_x) t + O(t^2).
\end{aligned}$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

A consequence of this result is that if two copies of the process are started from states $\varnothing$ and $\{x\}$, then, if

$$\sum_{y \neq x} \beta_{xy} < \alpha_x + \gamma_x,$$

the distributions of the two processes over a short period of time converge, and if

$$\sum_{y \neq x} \beta_{xy} > \alpha_x + \gamma_x,$$

the distributions of the two processes over a short period of time diverge. Using the results of Section 5, the following sufficient condition is obtained for the spontaneous process to converge exponentially quickly to its stationary distribution.

Theorem 4. *Suppose that the spontaneous process is irreducible, and let π denote its unique stationary distribution. Let*

$$\lambda = \min_{x \in S} \left(\alpha_x + \gamma_x - \sum_{y \neq x} \beta_{xy} \right),$$

and suppose that $\lambda > 0$. Let P^t denote the transition semigroup for the spontaneous process. Then, for any pair of distributions μ and ν on Ω and any $t \geq 0$,

$$D\,\rho(\mu P^t, \nu P^t) \leq -\lambda\,\rho(\mu P^t, \nu P^t),$$

and, in particular, for $0 \leq s \leq t$,

$$\rho(\mu P^t, \pi) \leq \mathrm{e}^{-\lambda(t-s)}\,\rho(\mu P^s, \pi).$$

Proof. Vertices in V have distance 1 when the corresponding states differ at one site, which is exactly the case considered in Proposition 4. The result follows using Corollary 1, and integrating.

It is not hard to show that the quantity $\alpha_x + \gamma_x$ is the rate of convergence of the state of site x to its stationary distribution, when the influence of other vertices is disregarded. The quantity $\sum_{y \neq x} \beta_{xy}$ measures the effect of the state of site x on the state of other sites. Therefore, Theorem 4 says that if at each site the rate of convergence to its stationary value of the state of that site is larger than the total rate at which the state of that site affects the state of neighbouring sites, then the state of the system as a whole converges exponentially to its stationary value.

Theorem 4 shows that in order to increase the rate of convergence to the stationary distribution, it suffices to decrease transmission rates, or to increase the rate of recovery. This reflects the intuition that, when a disturbance is introduced into a system, the system returns to its steady state more quickly when transmission rates are reduced, or when the rate of recovery is increased. Increasing the rate of spontaneous infection would also improve convergence, but this is of course undesirable.

Since the absorbing process is identical to the spontaneous process except for its behaviour at the disease-free state, Theorem 3 has the following corollary.

Corollary 2. *Suppose that the absorbing process is conditionally irreducible, and let ζ denote its unique quasistationary distribution. Let P^t denote its transition semigroup, and let μ be any distribution on Ω. Let λ be as in Theorem 4. Let $g(t)$ denote $\rho(\mu P^t, \zeta)$, and let*

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

$\phi(s,t) = \exp(-\int_s^t r(\tau)\,\mathrm{d}\tau)$ and $c(t) = |u_t - v|\,\mathrm{Diam}(G)$ be as in Theorem 3. Then, for $0 \le s \le t$,

$$g(t) \le \phi(s,t)g(s) + \int_s^t \phi(\tau,t)c(\tau)\,\mathrm{d}\tau.$$

Let μ_t denote μP^t. It follows, from the forward equation (1) for the process and from the fact that $q_{\eta\eta} = -\sum_{\xi\neq\eta} q_{\eta\xi}$ holds for each configuration η, that, for any subset $\Gamma \subset \Omega$ of the state space,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t(\Gamma) = \sum_{\xi\in\Gamma,\,\eta\in\Gamma^{\mathrm{c}}} \mu_t(\eta)q_{\eta\xi} - \mu_t(\xi)q_{\xi\eta}.$$

Letting $\Gamma = \Omega^*$, $(\Omega^*)^{\mathrm{c}}$ is the disease-free state, which is absorbing, so we have $q_{\eta\xi} = 0$ for $\eta \in (\Omega^*)^{\mathrm{c}}$ and $\xi \in \Omega$, and $q_{\xi\eta} \neq 0$ if and only if $\xi = \{x\}$ for some site x, in which case the transition rate $q_{\{x\}\varnothing} = \gamma_x$. Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mu_t(\Omega^*) = -\sum_{x\in S} \mu_t(\{x\})\gamma_x.$$

The terms u_t and v are described in Section 5 and are related to the additional convergence penalty $c(t) = |u_t - v|\,\mathrm{Diam}(G)$ for the absorbing process. They are given by

$$u_t = \frac{-1}{\mu_t(\Omega^*)} \sum_{x\in S} \mu_t(\{x\})\gamma_x, \qquad v = -\sum_{x\in S} \zeta_t(\{x\})\gamma_x.$$

To assess convergence, it is therefore enough to know λ and the mass at the states $\{x\}$ for $x \in S$.

7. Reduced equations

Recall that $p_x(t) = \mathbb{P}(\xi_t(x) = 1)$. We can argue as follows to obtain differential equations whose solutions approximate the values $p_x(t)$. For convenience, the notation $p_x(t)$ is also used for the quantities in these equations. The following assumption is made. If site y is not infectious at time t then it is infectious at time $t+\Delta t$ with probability $(\alpha_y + \sum_x p_x(t)\beta_{xy})\Delta t + O((\Delta t)^2)$. If site y is infectious at time t then it is infectious at time $t + \Delta t$ with probability $1 - \gamma_y\Delta t + O((\Delta t)^2)$. Therefore,

$$p_y(t+\Delta t) = \left(\alpha_y + \sum_x p_x(t)\beta_{xy}\right)\Delta t(1 - p_y(t)) + (1 - \gamma_y\Delta t)p_y(t) + O((\Delta t)^2),$$

which means that

$$\frac{p_y(t+\Delta t) - p_y(t)}{\Delta t} = \left(\alpha_y + \sum_x p_x(t)\beta_{xy}\right)(1 - p_y(t)) - \gamma_y p_y(t) + O(\Delta t),$$

and, taking the limit as $\Delta t \to 0^+$,

$$\frac{\mathrm{d}}{\mathrm{d}t}p_y(t) = \left(\alpha_y + \sum_x p_x(t)\beta_{xy}\right)(1 - p_y(t)) - \gamma_y p_y(t). \tag{6}$$

These are called the *reduced equations* and have been studied in [9]—in that paper the equations are referred to as 'the n-intertwined model'. Comparing to (2), we find that the assumption is equivalent to the assumption that

$$\mathbb{P}(\xi_t(x) = 1,\ \xi_t(y) = 0) = p_x(1 - p_y) = \mathbb{P}(\xi_t(x) = 1)\mathbb{P}(\xi_t(y) = 0),$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

that is, coordinates are pairwise independent. We can construct a stochastic process from the reduced equations using the values $p_x(t)$ obtained from the reduced equations, and then letting site y have the rate of infection $\alpha_y + \sum_x p_x(t)\beta_{xy}$ and the rate of recovery γ_y.

We can solve for the steady states $(\pi(y))_{y\in S}$ of the reduced equations by setting $\mathrm{d}p_y(t)/\mathrm{d}t = 0$ in (6), obtaining

$$\pi(y) = \frac{U(y)}{U(y) + \gamma_y}, \tag{7}$$

where

$$U(y) = \alpha_y + \sum_x \pi(x)\beta_{xy} \geq \alpha_y$$

and can be interpreted as the rate of infection at y. Under this interpretation, and interpreting $\pi(y)$ as the steady state probability of infection at site y, (7) in words says that the steady state probability of infection is given by the ratio of the infection rate at y to the sum of the infection and recovery rates. Note that the infection rate at y depends on the neighbouring values of $\pi(x)$, so (7) indeed gives a consistency condition for the steady state values. If $\beta_{xy} = 0$ for every distinct ordered pair of sites xy, the above equations decouple to give

$$\pi(y) = \frac{\alpha_y}{\alpha_y + \gamma_y}, \tag{8}$$

which is exactly the steady-state probability of infection of a vertex for the spontaneous process, in the absence of interaction.

It might be expected that the reduced equations have similar properties to the spontaneous process which they approximate. For example, we might suppose that the reduced equations are monotonic with respect to parameters and initial data, and that, when the spontaneous process is irreducible, the corresponding reduced equations have a unique steady state which is monotonic with respect to the parameters. These two suppositions are explored, in reverse order, in the following two subsections.

7.1. Monotonicity of π for small interactions

Recall that Θ denotes the set of parameters

$$\Theta = ((\alpha_x)_{x\in S}, (\beta_{xy})_{xy\in\Lambda}, (-\gamma_x)_{x\in S}),$$

where $\Lambda = \{xy \in S \times S : x \neq y\}$. Suppose that in some region of parameter values there is a locally unique branch of steady states $\pi(\Theta)$, one for each value of the parameter set Θ. Rewriting the steady state equations (7) as a function of parameters,

$$F(\Theta, \pi(\Theta)) = \pi(\Theta),$$

where $F(\Theta, \pi(\Theta))$ is the vector function with entries $U(y)/(U(y) + \gamma_y)$. Letting

$$G(\Theta, \pi) = F(\Theta, \pi) - \pi,$$

the steady state equations become

$$G(\Theta, \pi(\Theta)) = 0, \tag{9}$$

where the 0 on the right-hand side denotes the vector $(0, 0, \ldots, 0) \in \mathbb{R}^S$. Differentiating (9) with respect to Θ and suppressing arguments,

$$\frac{\mathrm{d}G}{\mathrm{d}\Theta} = \partial_\Theta G + \partial_\pi G \frac{\mathrm{d}\pi}{\mathrm{d}\Theta} = 0$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

or

$$\partial_\Theta G = -\partial_\pi G \frac{\mathrm{d}\pi}{\mathrm{d}\Theta},$$

which, if $-\partial_\pi G$ is invertible gives

$$\frac{\mathrm{d}\pi}{\mathrm{d}\Theta} = (-\partial_\pi G)^{-1}\partial_\Theta G.$$

Now,

$$\partial_\pi G = \partial_\pi F - I,$$

where I is the identity, so $\partial_\pi G$ is singular if and only if 1 is an eigenvalue of $\partial_\pi F$. The matrix $\partial_\pi F$ has entries

$$(\partial_\pi F)_{xy} = \begin{cases} \partial_{\pi(x)} \dfrac{U(y)}{U(y)+\gamma_y}, & \beta_{xy} \neq 0, \\ 0, & \text{otherwise}, \end{cases}$$

where, using the quotient rule and the fact that $\partial_{\pi(x)} U(y) = \beta_{xy}$,

$$\partial_{\pi(x)} \frac{U(y)}{U(y)+\gamma_y} = \frac{\beta_{xy}\gamma_y}{(U(y)+\gamma_y)^2}.$$

The nonzero entries of $\partial_\pi F$ satisfy

$$0 \leq (\partial_\pi F)_{xy} \leq \frac{\beta_{xy}}{\alpha_y + \gamma_y},$$

which implies in particular that $\partial_\pi F$ is a nonnegative matrix. Moreover, the column sums of $\partial_\pi F$ satisfy the inequality

$$0 \leq \sum_x (\partial_\pi F)_{xy} \leq \frac{\sum_x \beta_{xy}}{\alpha_y + \gamma_y}.$$

If the column sums of $\partial_\pi F$ are less than 1, that is,

$$\sum_x \beta_{xy} < \alpha_y + \gamma_y \tag{10}$$

for each $y \in S$, then the matrix $-\partial_\pi G$ is positive on the diagonal, nonpositive on the off-diagonal, and strictly diagonally dominant, and so it is a nonsingular M matrix (see Chapter 6 of [1] for several equivalent definitions of an M-matrix), and, in particular, $(-\partial_\pi G)^{-1}$ exists and is a nonnegative matrix.

When there is no interaction ($\beta_{xy} = 0$ for all distinct ordered pairs of sites xy), we have the unique steady state (8). Condition (10) is satisfied for all parameter values in a neighbourhood of the 'interaction-free' values. By the implicit function theorem, we therefore have a unique branch of steady states $\pi(\Theta)$, as hoped for, which can be continued so long as condition (10) is satisfied. Using the fact that $\partial_\Theta G$ has entries

$$(\partial_{\alpha_x} G)_y = \partial_{\alpha_x} \frac{U(y)}{U(y)+\gamma_y} = \frac{\gamma_y}{(U(y)+\gamma_y)^2} \geq 0,$$

$$(\partial_{\beta_{xy}} G)_y = \partial_{\beta_{xy}} \frac{U(y)}{U(y)+\gamma_y} = \frac{\pi(x)\gamma_y}{(U(y)+\gamma_y)^2} \geq 0,$$

$$(\partial_{-\gamma_y} G)_y = \partial_{-\gamma_y} \frac{U(y)}{U(y)+\gamma_y} = \frac{U(y)}{(U(y)+\gamma_y)^2} \geq 0$$

implies that $\partial_\Theta \pi \geq 0$; in other words, $\Theta_1 \geq \Theta_2$ implies that $\pi(\Theta_1) \geq \pi(\Theta_2)$.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

7.2. Monotonicity of the reduced equations

Consider the reduced equations

$$\frac{\mathrm{d}}{\mathrm{d}t}p_y(t) = \left(\alpha_y + \sum_{\{x\in S:\, xy\in E\}} p_x(t)\beta_{xy}\right)(1 - p_y(t)) - \gamma_y p_y(t). \tag{11}$$

Let $p(t, p_0, \Theta)$ denote the probability of infection of sites as a function of time for a set of parameters Θ and initial data $p_0 = p(0, p_0, \Theta)$; in other words, $p(t, p_0, \Theta)$ is the flow corresponding to the vector field defined by (11).

For fixed p_0, Θ, and t, let p be the vector $p(t, p_0, \Theta)$ with entries p_x. Our first goal is to define a notion of access that depends on p. For x and y vertices, $x \neq y$, and a positive integer k, define

$$C_k(x, y) = \sum_{x=x_0x_1\cdots x_k=y} \prod_{i=1}^{k} \beta_{x_{i-1}x_i}(1 - p_{x_i}).$$

Say that $x \to y$ if $x = y$ or if there is a positive integer k such that $p_x C_k(x, y) > 0$. Since

$$C_{j+k}(x, y) = \sum_z C_j(x, z)C_k(z, y), \tag{12}$$

it follows that '$\to$' is transitive. On the set $\{(x, y) \in S \times S:\, x \to y\}$, define $\mathrm{d}(x, y)$ as follows. If $x = y$, let $\mathrm{d}(x, y) = 0$. If $x \neq y$, let $\mathrm{d}(x, y)$ be equal to k such that $p_x C_k(x, y) > 0$ and $p_x C_j(x, y) = 0$ for $j < k$. Then d satisfies the triangle inequality. To see this, note that if $x \to y \to z$ and either $x = y$, $x = z$, or $y = z$, then it holds trivially that $\mathrm{d}(x, y) \leq \mathrm{d}(x, z)+\mathrm{d}(z, y)$. If x, y, and z are all distinct, argue as follows. Since, for any k, x, y, $C_k(x, y)$ is nonnegative, then, using (12), if $C_j(x, z) > 0$ and $C_k(z, y) > 0$ then $C_{j+k}(x, y) > 0$. From the definition of $\mathrm{d}(x, y)$, it then follows that $\mathrm{d}(x, y) \leq j + k \leq \mathrm{d}(x, z) + \mathrm{d}(z, y)$.

The following lemma uses the access notion just defined to describe the dependence of derivatives of $p(t, p_0, \Theta)$ on entries and parameter values.

Lemma 5. *For fixed p_0 and Θ, let $p(t)$ be the function $p(t, p_0, \Theta)$ with entries $p_x(t)$, $x \in S$. Let $x, y \in S$, and let k be a positive integer. If $\mathrm{d}(x, y) = k$ then, for $t \geq 0$,*

$$\partial_{p_x(t)} p_y^{(k)}(t) > 0.$$

If $\mathrm{d}(x, y) = k - 1$ then

$$\partial_{-\gamma_x} p_y^{(k)}(t) > 0.$$

If $\mathrm{d}(x, y) = k - 1$ and, for some positive integer j, $\partial_{\alpha_x} p^{(j)} \neq 0$, then

$$\partial_{\alpha_x} p_y^{(k)}(t) > 0.$$

If $\mathrm{d}(x, y) = k - 1$ then, for any z, if, for some positive integer j, $\partial_{\beta_{zx}} p_y^{(j)}(t) \neq 0$, then

$$\partial_{\beta_{zx}} p_y^{(k)}(t) > 0.$$

Proof. Since t is fixed, for convenience, denote the entries $p_x(t)$ and their derivatives $p_x^{(j)}(t)$ by p_x and $p_x^{(j)}$. First, observe that $\mathrm{d}p_y/\mathrm{d}t$ depends on p_x such that $\mathrm{d}(x, y) \leq 1$ and on α_x, $\beta_{\cdot x}$, and γ_x such that $\mathrm{d}(x, y) = 0$. If $\mathrm{d}(x, y) = k$ then, by the triangle inequality, $\mathrm{d}p_x/\mathrm{d}t$ depends on p_z such that $\mathrm{d}(z, y) \leq k + 1$, and depends on α_z, $\beta_{\cdot z}$, and γ_z such that $\mathrm{d}(z, y) \leq k$.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

Let $F(p, \Theta; k, k')$ be a blanket notation for a polynomial depending on p_x such that $\mathrm{d}(x, y) \leq k$ (on entries at a distance less than or equal to k) and on α_x, β_x, and γ_x such that $\mathrm{d}(x, y) \leq k'$ (on parameters at a distance less than or equal to k'). Using the product rule and the above fact about the derivatives of the p_x, it follows that $\mathrm{d}F(p, \Theta; k, k)/\mathrm{d}t = F(p, \Theta; k+1, k)$.

For a positive integer k, define

$$A_k(y) = \sum_{\{x:\ \mathrm{d}(x,y)=k\}} p_x C_k(x, y).$$

Observe that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} p_y &= \left(\alpha_y + \sum_x p_x \beta_{xy}\right)(1 - p_y) - \gamma_y p_y \\ &= A_1(y) + \alpha_y(1 - p_y) - \gamma_y p_y \\ &= A_1(y) + F(p, \Theta; 0, 0).\end{aligned}$$

Assume inductively that

$$p_y^{(k-1)} = A_{k-1}(y) + F(p, \Theta; k-2, k-2).$$

For each k, $C_k(x, y)$ is a polynomial in p and Θ that depends on entries and on parameters at a distance less than or equal to $k-1$. Therefore, differentiating $A_{k-1}(y)$,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} A_{k-1}(y) &= \sum_{\{x:\ \mathrm{d}(x,y)=k-1\}} \frac{\mathrm{d}}{\mathrm{d}t} p_x C_{k-1}(x, y) + p_x \dot{C}_{k-1}(x, y) \\ &= \sum_{\{x:\ \mathrm{d}(x,y)=k-1\}} \left[\left(\alpha_x + \sum_{\{z:\ \mathrm{d}(z,y)=k\}} p_z \beta_{zx}\right)(1 - p_x) - \gamma_x p_x\right] C_{k-1}(x, y) \\ &\quad + F(p, \Theta; k-1, k-2). \qquad (13)\end{aligned}$$

Pushing a few more terms into the polynomial F,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} A_{k-1}(y) &= \sum_{\{x:\ \mathrm{d}(x,y)=k-1\}} \sum_{\{z:\ \mathrm{d}(z,y)=k\}} [p_z \beta_{zx}(1 - p_x)] C_{k-1}(x, y) + F(p, \Theta; k-1, k-1) \\ &= \sum_{\{z:\ \mathrm{d}(z,y)=k\}} \sum_{\{x:\ \mathrm{d}(x,y)=k-1\}} [p_z \beta_{zx}(1 - p_x)] C_{k-1}(x, y) + F(p, \Theta; k-1, k-1) \\ &= \sum_{\{z:\ \mathrm{d}(z,y)=k\}} p_z C_k(z, y) + F(p, \Theta; k-1, k-1) \\ &= A_k(y) + F(p, \Theta).\end{aligned}$$

Then, differentiating $p^{(k-1)}$,

$$p_y^{(k)} = A_k(y) + F(p, \Theta; k-1, k-1),$$

which proves the induction. If $\mathrm{d}(x, y) = k$ then, by definition, $C_k(x, y) > 0$ and so

$$\partial_{p_x} p_y^{(k)} = C_k(x, y) > 0.$$

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

To assess the dependence of $p_y^{(k)}$ on parameters, differentiate $p_y^{(k-1)}$ and use (13) to write $p_y^{(k)}$ as

$$p_y^{(k)} = \sum_{\{x:\ \mathrm{d}(x,y)=k-1\}} \left[\left(\alpha_x + \sum_{\{z:\ \mathrm{d}(z,y)=k\}} p_z \beta_{zx} \right)(1-p_x) - \gamma_x p_x \right] C_{k-1}(x,y) + F(p, \Theta; k-1, k-2).$$

If $\mathrm{d}(x, y) = k - 1$ then, by definition, $C_{k-1}(x, y) > 0$ and so

$$\partial_{\alpha_x} p_y^{(k)} = (1-p_x) C_{k-1}(x,y) \geq 0,$$
$$\partial_{\beta_{zx}} p_y^{(k)} = p_z (1-p_x) C_{k-1}(x,y) \geq 0,$$
$$\partial_{-\gamma_x} p_y^{(k)} = p_x C_{k-1}(x,y) > 0,$$

as claimed. If $\partial_{\alpha_x} p_y^{(k)} = 0$ then $(1 - p_x) = 0$ and $\mathrm{d}p_x/\mathrm{d}t$ does not depend on α_x. Since, for any positive integer j, α_x only appears in the expression for $p_y^{(j)}$ as a result of differentiating p_x, it follows that $\partial_{\alpha_x} p_y^{(j)} = 0$. Similarly, if $\partial_{\beta_{zx}} p^{(k)} = 0$ then $p_z(1 - p_x) = 0$ and $\mathrm{d}p_x/\mathrm{d}t$ does not depend on β_{zx}. Since, for any positive integer j, β_{zx} only appears in the expression for $p_y^{(j)}$ as a result of differentiating p_x, it follows that $\partial_{\beta_{zx}} p_y^{(j)} = 0$.

Theorem 5 below establishes monotonicity of solutions to the reduced equations with respect to p_0 and Θ, assuming that the flow is locally analytic. The following lemma establishes this fact.

Lemma 6. *For fixed values of p_0, Θ, and a, there is a neighbourhood U of a such that the function $p(t, p_0, \Theta)$ is analytic for $t \in U$.*

Proof. Fix Θ and p_0, and let $p(t) = p(t, p_0, \Theta)$. Let $L = \|\Theta\|_\infty$; L is the least upper bound of absolute values of parameters. Note first that, for any $y \in S$,

$$\frac{\mathrm{d}}{\mathrm{d}t} p_y(t) = \alpha_y - (\alpha_y + \gamma_y) p_y + \sum_x p_x \beta_{xy} - \sum_x p_x \beta_{xy} p_y,$$

so

$$\left| \frac{\mathrm{d}}{\mathrm{d}t} p_y(t) \right| \leq L + 2L p_y + L \sum_x p_x + L \sum_x p_x p_y.$$

Let $d = \max_{y \in S} \#\{x:\ \beta_{xy} \neq 0\}$, where # denotes cardinality. Then $|\mathrm{d}p_y(t)/\mathrm{d}t|$ is bounded by a polynomial in $(p_x(t))_{x \in S}$ of degree at most 2, containing at most $1+1+d+d = 2(1+d) := C$ monomial terms, each of absolute value at most $2L$. Examine the monomial $p_{x_1} p_{x_2} \cdots p_{x_k}$, where the x_i are not necessarily distinct. Differentiating,

$$\frac{\mathrm{d}}{\mathrm{d}t} p_{x_1} p_{x_2} \cdots p_{x_k} = \sum_{i=1}^{k} \frac{\mathrm{d}}{\mathrm{d}t} p_{x_i} \prod_{j \neq i} p_{x_i},$$

so $|\mathrm{d}p_{x_1} p_{x_2} \cdots p_{x_k}/\mathrm{d}t|$ is bounded above by a polynomial in $(p_x(t))_{x \in S}$ of degree at most $k + 1$, containing at most kC terms, each of absolute value at most $2L$. Assume inductively that $|\mathrm{d}^k p_y(t)/\mathrm{d}t^k|$ is bounded above by at most $k! C^k$ monomial terms, each of absolute value at most $(2L)^k$. Then it follows that $\mathrm{d}^{k+1} p_y(t)/\mathrm{d}t^{k+1}$ is bounded in absolute value by at most $(k + 1)C(k! C^k) = (k + 1)! C^{k+1}$ monomial terms, each of absolute value at most

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

$2L(2L)^k = (2L)^{k+1}$. The induction is proved. In particular, for each positive integer k and each $t \geq 0$,

$$\left| \frac{\mathrm{d}^k}{\mathrm{d}t^k} p_y(t) \right| \leq k!\,(2CL)^k,$$

which implies that, for each $a \geq 0$, there is a neighbourhood of a on which $p_y(t)$ can be expressed as the convergent power series

$$\sum_{k=0}^{\infty} \frac{1}{k!} \frac{\mathrm{d}^k}{\mathrm{d}t^k} p_y(a)(t-a)^k.$$

Theorem 5. *If $\Theta_1 \geq \Theta_2$ and $p_{0,1} \geq p_{0,2}$ then, for $t \geq 0$,*

$$p(t, p_{0,1}, \Theta_1) \geq p(t, p_{0,2}, \Theta_2).$$

Proof. Denote $p(t, p_{0,1}, \Theta_1)$ and $p(t, p_{0,2}, \Theta_2)$ by $p_1(t)$ and $p_2(t)$ respectively. For a proof by contradiction, let $\mathcal{T}$ denote the set of times $t \geq 0$ such that there exists y such that $p_{1,y}(t) - p_{2,y}(t) < 0$. Then $\mathcal{T}$ is the union over y of the inverse image of the negative reals under $p_{1,y} - p_{2,y}$, and is therefore open. If $\mathcal{T}$ is empty, the assertion holds. If $\mathcal{T}$ is not empty, let $a = \inf \mathcal{T}$. Since $p_{0,1} \geq p_{0,2}$, it follows that $\mathcal{T}$ does not contain 0; therefore, $a \geq 0$. Then, since $\mathcal{T}$ is open, $a \notin \mathcal{T}$. There exists an $\varepsilon > 0$ and a y such that $p_{1,y}(t) - p_{2,y}(t) < 0$ for $a < t < a + \varepsilon$. Using Lemma 6, we expand $p_{1,y}(t) - p_{2,y}(t)$ as a power series centred at a:

$$p_{1,y}(t) - p_{2,y}(t) = \sum_{k \geq 0} \frac{1}{k!} [p_{1,y}^{(k)}(a) - p_{2,y}^{(k)}(a)](t-a)^k.$$

Since $a \notin \mathcal{T}$, $p_{1,y}(a) \geq p_{2,y}(a)$. Thus, for some $m > 0$, $p_{1,y}^{(k)}(a) = p_{2,y}^{(k)}(a)$ for $k < m$ and $p_{1,y}^{(m)}(a) < p_{2,y}^{(m)}(a)$. Denote $p_1(a)$ and $p_2(a)$ by p_1 and p_2, and define the relation $x \to y$ and the function $\mathrm{d}(x, y)$ according to p_1 (works equally well using p_2). Ignore parameters α_x and β_{zx} for which $\partial_{\alpha_x} p_{1,y}^{(j)} = 0$ and $\partial_{\beta_{zx}} p_{1,y}^{(j)} = 0$ for all j. Suppose that, for $k < m$, the entries of p_1 and p_2 agree up to a distance $k-1$ and that the parameters Θ_1 and Θ_2 agree up to a distance $k-2$, a fact which is vacuously true for $k = 1$. Now, $p_{1,y}^{(k)}$ and $p_{2,y}^{(k)}$ are equal, and each depends on its entries up to a distance less than or equal to k and on its parameters up to a distance less than or equal to $k-1$. By assumption, only the entries at distance k and the parameters at distance $k-1$ may differ. However, by Lemma 5, if any of these values differed, then $p_{1,y}^{(k)}$ and $p_{2,y}^{(k)}$ would be different, which is not the case. Therefore, entries at distance k and parameters at distance $k-1$ must agree. By induction, this forces agreement of entries up to a distance less than or equal to $m-1$ and parameters up to a distance less than or equal to $m-2$. Since $p_1 \geq p_2$ and $\Theta_1 \geq \Theta_2$, and since $p_{1,y}^{(m)}$ and $p_{2,y}^{(m)}$ depend on entries at a distance less than or equal to m and on parameters at a distance less than or equal to $m-1$, again applying Lemma 5 now establishes that $p_{1,y}^{(m)} \geq p_{2,y}^{(m)}$, which contradicts $p_{1,y}^{(m)} < p_{2,y}^{(m)}$. It follows that $\mathcal{T}$ is empty; in other words, $p_1(t) \geq p_2(t)$ for $t \geq 0$.

8. Sample paths for a star network

We conclude by examining the sample paths for a star network. It is interesting to observe the dynamics of the spontaneous process from the perspective of a single site, when there is interaction. As a simple example, consider N sites having identical rates $\alpha > 0$ and $\gamma > 0$ of spontaneous infection and recovery. Endow the sites with a star graph topology by selecting a central site x, and setting $\beta_{xy} = \beta_{yx} = \beta > 0$ if $y \neq x$ and $\beta_{yz} = 0$ if $y \neq x$ and $z \neq x$.

https://doi.org/10.1239/aap/1401369707 Published online by Cambridge University Press

For this model, the stationary probability of infection of a site (according to the reduced equations) can be computed explicitly. Since the dynamics are invariant under permutation of the noncentral sites, the stationary probability of infection must be constant over noncentral sites, and so it can take on at most two distinct values. Let $\pi(x)$ and $\pi(y)$ denote the corresponding values, where x is the central site and y is any noncentral site. Then (7) gives the system of equations

$$\pi(x)[\alpha + (N-1)\pi(y)\beta + \gamma] = \alpha + (N-1)\pi(y)\beta,$$
$$\pi(y)[\alpha + \pi(x)\beta + \gamma] = \alpha + \pi(x)\beta.$$

Substituting the second equation into the first equation, gives the quadratic equation $AX^2 + BX + C = 0$ for $\pi(x)$, with

$$A = (\alpha+\gamma)\beta + (N-1)\beta^2,$$
$$B = (\alpha+\gamma)^2 + (N-2)\alpha\beta - (N-1)\beta^2,$$
$$C = -[\alpha(\alpha+\gamma) + (N-1)\alpha\beta],$$

which, since $A > 0$ and $C < 0$, means that the roots $[-B \pm \sqrt{B^2 - 4AC}]/2A$ are real and distinct, with one positive and one negative root. Therefore, the positive root is equal to $\pi(x)$, the stationary probability of infection at x; $\pi(y)$ can then also be obtained.

We could try to simulate the dynamics at x just by defining the rate of infection $\alpha + (N-1)\pi(y)\beta$ and the rate of recovery γ. This would give the correct stationary probability of infection at x. However, the dynamics at x are more colourful, as shown below.

Let $(\xi_t)_{t\geq 0}$ be a realization of the process such that $\mathbb{P}(\xi_t(y) = 1)$ is the same for all $y \neq x$. Since the dynamics are invariant under any permutation of the noncentral sites, for this property to hold, it is sufficient that it hold for $t = 0$. If $y \neq x$ then in terms of the path $(\xi_t(x))_{t\geq 0}$ at x,

$$\frac{\mathrm{d}}{\mathrm{d}t} p_y(t) = (\alpha + \mathbf{1}_{\{\xi_t(x)=1\}}\,\beta)(1 - p_y(t)) - \gamma p_y(t),$$

which can be integrated to give $p_y(t)$ in terms of $p_y(0)$ and the sample path $\{\xi_s(x)\colon 0 \leq s \leq t\}$. Defining $m(t) = \alpha + \mathbf{1}_{\{\xi_t(x)=1\}}\,\beta + \gamma$, $M(s,t) = \exp(-\int_s^t m(\tau)\,\mathrm{d}\tau)$, and $i(t) = \alpha + \mathbf{1}_{\{\xi_t(x)=1\}}\,\beta$,

$$p_y(t) = M(0,t)\,p_y(0) + \int_0^t i(s)M(s,t)\,\mathrm{d}s.$$

The quantity $m(t)$ can be understood as the instantaneous rate of mixing at y, since it is the rate at which the influence of the state at previous times decays. The quantity $i(t)$ can be understood as the instantaneous rate of infection. The instantaneous rate of infection at x at time t is then given by $\alpha + (N-1)p_y(t)\beta$, which depends on the history at x. From x's perspective, the dynamics are non-Markovian.

Acknowledgements

The author thanks Rod Edwards and Pauline van den Driessche for helpful discussions, and the anonymous referee for useful comments.

References

[1] Berman, A. and Plemmons, R. J. (1994). *Nonnegative Matrices in the Mathematical Sciences.* Society for Industrial and Applied Mathematics, Philadelphia, PA.

[2] Bubley, R. and Dyer, M. (1997). Path coupling: a technique for proving rapid mixing in Markov chains. In *Proc. 38th Annual Symp. Foundations of Computer Science*, pp. 223–231.

[3] Harris, T. E. (1974). Contact interactions on a lattice. *Ann. Prob.* **2,** 969–988.
[4] Harris, T. E. (1978). Additive set-valued Markov processes and graphical methods. *Ann. Prob.* **6,** 355–378.
[5] Krone, S. M. (1999). The two-stage contact process. *Ann. Appl. Prob.* **9,** 331–351.
[6] Levin, D., Peres, Y. and Witmer, E. L. (2006). *Markov Chains and Mixing Times.* American Mathematical Society.
[7] Neuhauser, C. (1992). Ergodic theorems for the multitype contact process. *Prob. Theory Relat. Fields* **91,** 467–506.
[8] Norris, J. R. (1997). *Markov Chains.* Cambridge University Press.
[9] Van Mieghem, P., Omic, J. and Kooij, R. (2009). Virus spread in networks. *IEEE/ACM Trans. Networking* **17,** 1–14.