Rumours, consensus and epidemics on networks

A. J. Ganesh

May 31, 2012

1 Review

1.1 The Poisson process

The Poisson process is one of the most widely used models of a continuous time stochastic process. Along with Brownian motion, it can be considered as a fundamental example of such processes. The Poisson process is a counting process N_t, $t \geq 0$, (a non-decreasing integer-valued process which counts the number of 'points' or 'events' up to time t) indexed by the time parameter t, which will be positive real-valued in most of our examples.

We say that a counting process is a Poisson process of rate (or intensity) λ if the following hold:

- [1] The random variable $N_{t+s} - N_t$ is independent of $\{N_u, 0 \leq u \leq t\}$, for all $s, t \geq 0$.
- [2] The random variable $N_{t+s} - N_t$ has a Poisson distribution with mean λs, i.e.,

$$\mathbb{P}(N_{t+s} - N_t = k) = \frac{(\lambda s)^k}{k!} e^{-\lambda s}, \quad k = 0, 1, 2, \ldots \tag{1}$$

Property 2 above can be equivalently restated in either of the following two ways:

- [2a] For all $t \geq 0$,

$$\begin{aligned}\mathbb{P}(N_{t+h} - N_t = 1) &= \lambda h + o(h) \\ \mathbb{P}(N_{t+h} - N_t = 0) &= 1 - \lambda h + o(h) \\ \mathbb{P}(N_{t+h} - N_t \geq 2) &= o(h),\end{aligned}$$

  where, for a function f, we write $f(h) = o(h)$ if $f(h)/h \to 0$ as $h \to 0$. Loosely speaking, we say a function is $o(h)$ if it tends to zero faster than h.

- [2b] Let $T_1, T_2, T_3, \ldots$ be the increment times of the counting process $(N_t, t \geq 0)$, i.e.,

$$T_n = \inf\{t \geq 0 : N_t \geq n\}.$$

  The process $(N_t, t \geq 0)$ is a Poisson process of rate λ if the random variables $T_{n+1} - T_n$ are independent and identically distributed (iid) with the Exp(λ) distribution, i.e., $\mathbb{P}(T_{n+1} - T_n \geq t) = \exp(-\lambda t)$ for all $t \geq 0$.

Recall that a random variable X has the Exp(λ) distribution if $\mathbb{P}(X > t) = e^{-\lambda t}$. Now, by Bayes' theorem,

$$\mathbb{P}(X > t+s|X > t) = \frac{\mathbb{P}(X > t+s \cap X > t)}{\mathbb{P}(X > t)} = \frac{\exp(-\lambda(t+s))}{\exp(-\lambda t)} = \mathbb{P}(X > s).$$

If we think of X as the time to occurence of an event, then knowing that the event hasn't occured up to time t tells us nothing about how much longer we need to wait for it to occur; the distribution of the *residual time* until it occurs doesn't depend on how long we have already waited. This is referred to as the *memoryless property* of the exponential distribution.

We shall establish the equivalence of [2], [2a] and [2b] by showing that [2] $\Rightarrow$ [2*a*] $\Rightarrow$ [2*b*] $\Rightarrow$ [2]. The first implication is obvious by letting s tend to zero.

For the second implication, it suffices to show that T_1 has an $Exp(\lambda)$ distribution since, by [1], T_1, $T_2 - T_1$, $T_3 - T_2$, ... are iid. Let F denote the cdf of T_1. Observe that

$$\begin{aligned}\mathbb{P}(T_1 \in (t, t+h]) &= \mathbb{P}(N_t = 0, N_{t+h} - N_t \geq 1) \\ &= \mathbb{P}(N_t = 0)\mathbb{P}(N_{t+h} - N_t \geq 1 | N_u = 0, 0 \leq u \leq t) \\ &= \mathbb{P}(N_t = 0)\mathbb{P}(N_{t+h} - N_t \geq 1) \\ &= (1 - \lambda h + o(h))\mathbb{P}(T_1 > t),\end{aligned}$$

where the third equality follows from [1] and the last equality from [2a]. The above equation implies that

$$\mathbb{P}(T_1 \in (t, t+h] | T_1 > t) = 1 - \lambda h + o(h),$$

i.e., that $1 - F(t+h) = (1 - \lambda h + o(h))(1 - F(t))$. Letting h tend to zero, we obtain $F'(t) = -\lambda(1 - F(t))$. Solving this differential equation with the boundary condition $F(0) = 0$, we get $F(t) = 1 - \exp(-\lambda t)$, which is the cdf of an Exp(λ) random variable. Thus, we have shown that T_1 has an Exp(λ) distribution, as required.

To show the last implication, we show (1) by induction on k. Without loss of generality, let $t = 0$. For $k = 0$, (1) reads $\mathbb{P}(N_s = 0) = e^{-\lambda s}$. Since the event $N_s = 0$ is the same as $T_1 > s$, this follows from the exponential distribution of T_1. Next, suppose that (1) holds for all $j \leq k$. Let f denote the density function of T_1, the time to the first increment. By conditioning on the time to the first increment, we have

$$\begin{aligned}
\mathbb{P}(N_s = k+1) &= \int_0^s f(u)\mathbb{P}(N_s = k+1 | N_u = 1)du \\
&= \int_0^s f(u)\mathbb{P}(N_{s-u} = k)du \\
&= \int_0^s \lambda e^{-\lambda u} \frac{(\lambda(s-u))^k}{k!} e^{-\lambda(s-u)} du \\
&= e^{-\lambda s} \frac{(\lambda s)^{k+1}}{(k+1)!}.
\end{aligned}$$

The second equality follows from [1], and the third from [2b] and the induction hypothesis. Thus, we have shown that [2b] implies [2], completing the proof of the equivalence of [2], [2a] and [2b].

Remarks.

1. An implication of properties [1] and [2] above is that, if X_1 and X_2 are independent Poisson random variables with mean λ_1 and λ_2 respectively, then $X_1 + X_2$ is a Poisson random variable with mean $\lambda_1 + \lambda_2$. To see this, consider a Poisson process of unit rate, take X_1 to be N_{λ_1} and X_2 to be $N_{\lambda_1+\lambda_2} - N_{\lambda_1}$. You may want to verify, by direct calculation, that Poisson random variables do indeed have this property. (The simplest way to perform this calculation is using generating functions.)

2. The definition of a Poisson process can be extended to $t \in \mathbb{R}$ instead of just $t \geq 0$; to do so, simply use properties [1] and [2]. In this case, for $t < 0$, we let $-N_t$ denote the number of points of the Poisson process in $[-t, 0]$. With this definition, N_t is still a non-decreasing, integer-valued process.

3. The Poisson process we defined is 'homogeneous' in that the intensity parameter λ is a constant. This can be generalised to allow λ to be a (measurable) function of t, in which case the Poisson process is said to be inhomogeneous. Property [1] still holds, while property [2] is modified to say that the number of points in an interval $(s, t]$ is a Poisson random variable with mean $\int_s^t \lambda(x)dx$.

4. Finally, the definition of a Poisson process can also be extended to "spatial point processes". In this general setting, property [2] generalises to say that if A is a measurable subset of $\mathbb{R}^d$, then the number of points of the Poisson process in A is a Poisson random variable with mean λ (the intensity) times the Lebesgue measure (area or volume) of A. Property 1 generalises to say that if $A_1, A_2, \ldots$ are disjoint subsets of $\mathbb{R}^d$, then the number of points in these sets are mutually independent random variables.

We now state two key properties of Poisson processes that we shall make use of repeatedly throughout the course.

Lemma 1 *Suppose $N_t^1, t \geq 0$ and $N_t^2, t \geq 0$ are independent Poisson processes, of rate λ_1 and λ_2 respectively. Define $N_t = N_t^1 + N_t^2$ to be the superposition of the two Poisson processes (i.e., it counts points of both). Then, $N_t, t \geq 0$ is a Poisson process of rate $\lambda = \lambda_1 + \lambda_2$.*

Lemma 2 *Suppose $N_t, t \geq 0$ is a Poisson process of rate λ, and that $X_1, X_2, X_3, \ldots$ is a sequence of iid Bernoulli(p) random variables (i.e., $\mathbb{P}(X_i = 1) = p = 1 - \mathbb{P}(X_i = 0)$, $i = 1, 2, 3, \ldots$) independent of $N_t, t \geq 0$. Define*

$$N_t^1 = \sum_{k=0}^{N_t} X_k, \quad N_t^2 = N_t - N_t^1.$$

Then, $N_t^1, t \geq 0$ and $N_t^2, t \geq 0$ are independent Poisson processes, of rate λp and $\lambda(1-p)$ respectively.

The proof of these lemmas isn't difficult, so we shall only sketch an outline. For the proof of Lemma 1, observe that N_t inherits property [1] of a Poisson process from N_t^1 and N_t^2 and their independence, while property [2] holds because the sum of two independent Poisson random variables is Poisson, as noted above. For Lemma 2, property [1] follows from the fact that N_t is a Poisson process, and that the X_i are an iid sequence independent of the Poisson process N_t. The independence of N_t^1 and N_t^2 also follows from these facts. Property 2 is most easily checked in the form of [2a]; for the process N_t^1 to have an increment in the interval $[s, s+h]$, the process N_t must also have an increment, and the corresponding Bernoulli random variable should take the value 1. (In fact, we should consider the possibility of N_t having multiple increments during $[s, s+h]$, and exactly one of the corresponding Bernoulli random variables taking value 1, but the probability of two or more increments is $o(h)$, which is negligible.)

1.2 Continuous time Markov chains

Consider a continuous time stochastic process X_t, $t \geq 0$ on a discrete (finite or countable) state space S. The process is called a continuous time Markov chain (CTMC) or a Markov process on this state space if

$$\mathbb{P}(X_{t+s} = y | \{X_u, u \leq s\}) = \mathbb{P}(X_{t+s} = u | X_s), \tag{2}$$

for all states y and all $s, t \geq 0$. In words, this says that the future of the process X_t, $t \geq 0$ is *conditionally* independent of the past given the present.

If the above probability only depends on t and X_s but not on s, then we say that the Markov chain is time homogeneous. In that case, we can represent the conditional probabilities in (2) in the form of a matrix $P(t)$ with entries

$$P_{xy}(t) = \mathbb{P}(X_{t+s} = y | X(s) = x) = \mathbb{P}(X_t = y | X_0 = x).$$

The matrices $\{P(t), t \geq 0\}$ satisfy the following properties:

$$P(0) = I, \quad P(t+s) = P(t)P(s). \tag{3}$$

(Why?)

It should be clear that the Poisson process introduced in the last section has the above independence property, and hence is a Markov process. It turns out that the Poisson process already possesses many of the features of

general continuous-time Markov processes, as we shall see below, and hence serves as a template for them.

Let us return to such a process X_t, $t \geq 0$on a countable state space S, with transition probability matrices $\{P(t), t \geq 0\}$. We saw earlier that these matrices obeyed equation (3). Thus,

$$P(t+\delta) - P(t) = P(t)(P(\delta) - I) = (P(\delta) - I)P(t).$$

This suggests that, if $\frac{1}{\delta}(P(\delta) - I)$ converges to some matrix Q as δ decreases to zero, then

$$P'(t) = QP(t) = P(t)Q. \tag{4}$$

This is correct if the Markov process is finite-state, as will generally be the case in this course; in this case, the matrices $P(t)$ are finite-dimensional. For countable state chains, the first equality still holds but the second may not. The two equalities are known, respectively, as the **Chapman-Kolmogorov** backward and forward equations. The equations have the solution

$$P(t) = e^{Qt} = I + Qt + \frac{(Qt)^2}{2!} + \dots$$

What does the matrix Q look like? Since $P(\delta)$ and I both have all their row sums equal to 1, it follows that the row sums of $P(\delta) - I$, and hence of Q must all be zero. Morever, all off-diagonal terms in $P(\delta)$ are non-negative and those in I are zero, so Q must have non-negative off-diagonal terms. Thus, in general, we must have

$$q_{ij} \geq 0 \, \forall \, i \neq j; \quad q_{ii} = -\sum_{j \neq i} q_{ij},$$

where q_{ij} denotes the ij^{th} term of Q. Define $q_i = -q_{ii}$, so $q_i \geq 0$. In fact, $q_i > 0$ unless i is an absorbing state. (Why?)

The matrix Q is called the generator (or infinitesimal generator) of the Markov chain. The evolution of a Markov chain can be described in terms of its Q matrix as follows. Suppose the Markov chain is currently in state i. Then it remains in state i for a random time which is exponentially distributed with parameter q_i. In particular, knowing how long the Markov chain has been in this state gives us no information about how much longer it will do so (memoryless property of the exponential distribution). At the end of this time, the chain jumps to one of the other states; the probability

of jumping to state j is given by $q_{ij}/\sum_{k\neq i} q_{ik} = q_{ij}/q_i$, and is independent of how long the Markov chain spent in this state, and of everything else in the past. The Poisson process is a special case where the exponential random time spent in a state has the same mean for all states, and the only possible jump from a state $n \in \mathbb{N} \cup \{0\}$ is to state $n+1$.

If we observed a CTMC only at its jump times, we'd get a discrete time Markov chain (DTMC) on the same state space, with transition probabilities $p_{ij} = q_{ij}/q_i$. Conversely, we can think of a CTMC as a DTMC which spends random, exponentially distributed times (instead of unit time) in each state before transiting to another state.

If we know the state of a Markov process at time 0 and its infinitesimal generator, then we can work out its probability distribution at any future time. Specifically, if we let $\mu(t)$ denote the probability distribution on S at time t, then, by the discussion above,

$$\mu(t) = \mu(0)P(t) = \mu(0)e^{Qt}.$$

We shall be interested in particular in the long-term behaviour of the Markov process, i.e., in the behaviour of $\mu(t)$ as t tends to infinite.

In the following, we shall restrict our attention to finite state Markov processes as we shall only be dealing with these in this course, and it avoids certain technicalities that can arise in countable state chains.

Classification of states

Consider a Markov process $\{X_t, t \geq 0\}$ on a finite state space S. A state $x \in S$ is said to be recurrent if, starting from this state, the process returns to it with probability 1, i.e.,

$$\mathbb{P}(\exists t \geq 1 : X_t = x | X_0 = x) = 1.$$

It is called transient if this return probability is strictly smaller than 1.

State j is said to be accessible from state i if it is possible to go from i to j, i.e., $P_{ij}(t) > 0$ for some $t \geq 0$. In particular, each state is accessible from itself since $P_{ii}(0) = 1$. States i and j are said to communicate if i is accessible from j and j from i. Note that this defines an equivalence relation. (A relation R is said to be an equivalence relation if it is (a) symmetric: xRy implies yRx for all x and y, (b) reflexive: xRx for all x, and (c) transitive: xRy and yRz together imply xRz for all x, y and z.) Hence, it partitions

the state space into equivalence classes, which are called communicating classes. (Sets $A_1, A_2, \ldots$ are said to form a partition of A if $\cup_{i=1}^{\infty} A_i = A$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$.) Note that states in the same communicating class have to all be transient or all be recurrent.

A finite state Markov process eventually has to leave the set of transient states and end up in one or other recurrent communicating class. A Markov process is said to be irreducible it it consists of a single communicating class. In that case, for a finite-state process, that class has to be recurrent.

Invariant distributions

Suppose we observe a finite-state Markov chain over a long period of time. What can we say about its behaviour? If it starts in a transient state, it will spend some time in a transient communicating class before entering one or another recurrent class, where it will remain from then on. Thus, as far as long-term behaviour is concerned, it is enough to look at the recurrent classes. Moreover, each such communicating class can be studied in isolation since the Markov chain never leaves such a class upon entering it. Finally, the Markov chain within a single class is irreducible.

Hence, we restrict our attention in what follows to irreducible Markov chains. Consider such a chain on a finite state space S, and let Q denote its generator matrix. We have:

Theorem 1 *There is a unique probability distribution π on the state space S such that $\pi Q = \mathbf{0}$, where $\mathbf{0}$ denotes the all-zero vector.*

It is easy to see that Q has the all-1 vector as a right eigenvector corresponding to the eigenvalue 0, since all its row sums are zero. The above theorem says that the corresponding left eigenvector is also non-negative, and that there is only such eigenvector corresponding to an eigenvalue of 0 if the generator corresponds to an irreducible chain. The assumption of irreducibility is necessary for uniqueness. We won't give a proof of this theorem, but one way is to use the Perron-Frobenius theorem for non-negative matrices (applied to the stochastic matrix $P(t) = e^{Qt}$ for an arbitrary $t > 0$).

The probability distribution π solving $\pi Q = \mathbf{0}$ is called the *invariant distribution* or *stationary distribution* of the Markov chain. The reason is that, if the chain is started in this distribution, then it remains in this distribution forever. This is because the probability distribution $\mu(t)$ at any time t is

given by

$$\mu(t) = \mu(0)e^{Qt} = \pi\Big(I + Qt + \frac{Q^2t^2}{2!} + \ldots\Big) = \pi.$$

The invariant distribution describes the long-run behaviour of the Markov chain in the following sense.

Theorem 2 (Ergodic theorem for Markov chains) *If $\{X_t, t \geq 0\}$ is a Markov process on the state space S with unique invariant distribution π, then*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t 1(X(t) = j)dt = \pi_j,$$

irrespective of the initial condition. The convergence holds almost surely.

The above theorem holds for both finite and countable state spaces, assuming the invariant distribution exists (which it may fail to do in the countable case, even if the Markov chain is irreducible). The theorem says that π_j specifies the fraction of time that the Markov chain spends in state j in the long run.

1.3 Graphs and networks

We shall use the terms graph and network interchangeably to refer to a finite set of *nodes* or *vertices*, some of which may be connected by *edges* or *arcs*. We write $G = (V, E)$ to denote a graph G with vertex set V and edge set E. An edge is a pair of vertices. Thus, E is a subset of $V \times V$, the Cartesian product of the vertex set with itself. An edge (i, j), $i, j \in V$ is called undirected if the order of the vertices doesn't matter, i.e., if (i, j) and (j, i) are the same, and is called directed otherwise. If $(i, j) \in E$ is a directed edge, we say it is directed from i to j. A graph is called directed or undirected if all its edges are directed or undirected respectively. In this course, all graphs will be either one or the other. If $(i, j) \in E$, then j is called a neighbour of i. If the graph is undirected, the degree of i is defined as the number of neighbours it has. If it is directed, we use the terms in-degree and out-degree, with the obvious meanings (number of edges directed to i and from i respectively).

A path is a sequence $v_1, v_2, \ldots, v_k$ of nodes such that $(v_i, v_{i+1}) \in E$ or $(v_{i+1}, v_i) \in E$ for each i between 1 and $k-1$. If $v_1 = v_k$, the path is called a

cycle. The path or cycle is called directed if the underlying graph is directed and $(v_i, v_{i+1}) \in E$ for each i between 1 and $k-1$. We use the term simple path and simple cycle to mean that the path (cycle) does not contain a proper subset which is itself a cycle.

A graph which contains no cycles is called acyclic. A graph is said to be connected if, for any two nodes $u, v \in V$, there is a path between u and v (i.e., $\exists n$ and $v_1, v_2, \ldots, v_n$ such that $v_1 = u$, $v_n = v$ and $(v_i, v_{i+1}) \in E$ or $(v_{i+1}, v_i) \in E$ for all i between 1 and $n-1$). A connected, acylic graph is called a tree. Observe that if a tree has k nodes, it must have exactly $k-1$ edges. (Show this by induction on k.) An undirected graph $G = (V, E)$ is called a complete graph if $E = V \times V$, i.e., there is an edge between every pair of nodes.

Let $G = (V, E)$ be a graph and let $V' \subseteq V$. The subgraph induced by the nodes in V' is defined as the graph $G' = (V', E')$ where $E' = \{(i, j) \in E : i, j \in V'\}$. In other words, it is the graph consisting of the nodes in V' as well as those edge in the original graph which connect these nodes. If a subgraph is a complete graph (on its node set), then it is called a clique. A set of nodes $V' \subseteq V$ is called an independent set if, for any two nodes in V', the edge between them is absent (whereas in a clique, for any two nodes, the edge between them is present).

1.4 Probability Inequalities

What can we say about the probability of a random variable taking values in a certain set if we only know its moments, for instance, or its generating function? It turns out that they give us some bounds on the probability of the random variable taking values in certain specific sets. We now look at some examples.

Let X be a non-negative random variable with finite mean $\mathbb{E}X$. Then, for all $c > 0$, we have

$$\textbf{Markov's inequality:} \qquad \mathbb{P}(X \geq c) \leq \frac{\mathbb{E}X}{c}.$$

The proof is straightforward. Suppose X has a density, and denote it by f. Then

$$\mathbb{E}X = \int_0^\infty x f(x) dx \geq \int_c^\infty x f(x) dx \geq \int_c^\infty c f(x) dx = c\mathbb{P}(X \geq c).$$

Re-arranging this gives us Markov's inequality. (Why does X have to be non-negative?)

Next, let X be a random-variable, not necessarily non-negative, with finite mean $\mathbb{E}X$ and finite variance $\mathrm{Var}(X)$. Then, for all $c > 0$, we have

$$\textbf{Chebyshev's inequality:} \qquad \mathbb{P}(|X - \mathbb{E}X| \geq c) \leq \frac{\mathrm{Var}(X)}{c^2}.$$

The proof is an easy consequence of Markov's inequality. Note that the event $|X - \mathbb{E}X| \geq c$ is the same as the event $(X - \mathbb{E}X)^2 \geq c^2$, and apply Markov's inequality to the non-negative random variable $Y = (X - \mathbb{E}X)^2$. Note that $\mathbb{E}Y = \mathrm{Var}(X)$.

Finally, let X be a random-variable, not necessarily non-negative, and suppose that its moment-generating function $\mathbb{E}[e^{\theta X}]$ is finite for all θ. Then, for all $c \in \mathbb{R}$, we have

$$\textbf{Chernoff's inequality:} \qquad \mathbb{P}(X \geq c) \leq \inf_{\theta > 0} e^{-\theta c}\mathbb{E}[e^{\theta X}].$$

The proof follows by noting that the event $X \geq c$ is identical to the event $e^{\theta X} \geq e^{\theta c}$ for all $\theta > 0$ (the inequality gets reversed for $\theta < 0$), applying Markov's inequality to the non-negative random variable $Y = e^{\theta X}$, and taking the best bound over all possible θ.

2 Rumour spreading on the complete graph

Consider the following model of rumour spreading on a complete graph. There are n nodes, a single one of which initially knows the rumour. There are n independent unit rate Poisson processes, one associated with each node. At a time when there is a jump of the Poisson process $N_i(t)$ associated with node i, this node becomes active, and chooses another node j uniformly at random with which to communicate. If node i knows the rumour at this time and node j doesn't, then i informs j of the rumour; otherwise there is no change. This is called the "push" model as information is pushed from i to j. It should be obvious what is meant by the pull and push-pull models. Start time at 0 and let T denote the first time that all nodes know the rumour. Then T is a random time and we can ask about its expected value or its distribution, and how this depends on n, the size of the graph.

The model description above takes a node-centred perspective. But it is equivalent to a description where the clocks sit on the edges rather than

the nodes. Let's consider a node i and ask what the probability is that it chooses to communicate with node j during the time interval $([t, t + dt)$. For this to happen, node i should become active during this time interval, and it should choose node j to communicate with. The first of these events corresponds to $N_i(t + dt) - N_i(t) = 1$ (the Poisson process at node i has an increment during this time period), which has probability $1 \cdot dt$ (since the Poisson process has unit rate) independent of the past at all nodes. The second event has probability $\frac{1}{n-1}$, independent of everything else, since node i chooses another node to communicate with uniformly at random. Hence, the probability that the time period $[t, t + dt)$ sees a communication event from i to j is $\frac{1}{n-1}dt$, independent of the past. In other words, the model can be recast as follows: there are $n(n-1)$ independent Poisson processes of rate $1/(n-1)$, one associated with each *directed* edge (i, j) in the complete graph on n nodes. When there is a jump in the Poisson process on edge (i, j), the rumour is pushed from node i to node j if node i is informed at this time. Otherwise, nothing happens.

Remarks

1. Note that, in both models above, the Poisson processes on all nodes and edges start at time zero. Would it not be more natural to start the process at node i only when this node first learns of the rumour? Indeed it would, but it makes no difference. Can you see why?

2. A discrete-time version of the model was studied by Boris Pittel (On spreading a rumor, *SIAM J. Appl. Math.*, 1987, pp. 213–223). In this version, time proceeds in rounds. In each round, each informed node picks another node uniformly at random and pushes the rumour to it. The models are very similar. We have chosen to work with the continuous time model because it is somewhat easier to analyse.

We now want to analyse the model described above, and find out how long it takes for all nodes to learn the rumour. Observe from the verbal description that, if we let S_t denote the set of informed nodes at time t, then $S_t, t \geq 0$ evolves as a continuous time Markov chain. The state space of this Markov chain is the set of all subsets of the node set $\{1, 2, \ldots, n\}$, which is of size 2^n. This is a rather large state space. In fact, from the symmetry in the problem, we can see that it is enough to keep track of the number of informed nodes, rather than of exactly which nodes are informed. Even for this reduced state descriptor, the evolution is Markovian. This reduces the size of the state space to n, and makes the problem more tractable.

Let T_i denote the first time that exactly i nodes are informed, so that $T_1 = 0$ and $T_n = 1$. Then, $T_{i+1} - T_i$ is the random additional time it takes for the $(i+1)^{\text{th}}$ node to be informed, after the i^{th} node has been. Let S_{T_i} denote the set of nodes that are informed at time T_i. There are i nodes in this set, and $n-i$ nodes in its complement, so that there are $i(n-i)$ edges between S_{T_i} and $S_{T_i}^c$. There are independent Poisson processes of rate $1/(n-1)$ associated with each of these edges, according to which some node in S_{T_i} contacts some node in $S_{T_i}^c$ and informs it of the rumour. (Communications taking place on edges with S_{T_i} or $S_{T_i}^c$ have no effect.)

Now, using the fact that the superposition of independent Poisson processes is a Poisson process with the sum of their rates, we conclude that the time to inform a new node is the time to the first jump in a Poisson process of rate $i(n-i)/(n-1)$. In other words, $T_{i+1} - T_i$ is an $Exp(\frac{i(n-i)}{n-1})$ random variable, independent of T_i, and of the past of the rumour spreading process. Hence, recalling the formulas for the mean and variance of an exponential random variable, we have,

$$\mathbb{E}[T_{i+1} - T_i] = \frac{n-1}{i(n-i)}, \quad \text{Var}(T_{i+1} - T_i) = \Big(\frac{n-1}{i(n-i)}\Big)^2. \tag{5}$$

Using a partial fraction expansion, we can rewrite the above as

$$\begin{aligned} \mathbb{E}[T_{i+1} - T_i] &= \frac{n-1}{n}\Big(\frac{1}{i} + \frac{1}{n-i}\Big), \\ \text{Var}(T_{i+1} - T_i) &= \Big(\frac{n-1}{n}\Big)^2\Big(\frac{1}{i^2} + \frac{1}{(n-i)^2} + \frac{2}{n}\big(\frac{1}{i} + \frac{1}{n-i}\big)\Big). \end{aligned} \tag{6}$$

Next, we note that the time until all nodes know the rumour is given by $T_n = \sum_{i=1}^{n-1}(T_{i+1} - T_i)$ since $T_1 = 0$. Hence, by (6) and the linearity of expectation, we have

$$\begin{aligned} \mathbb{E}[T_n] &= \sum_{i=1}^{n-1} \mathbb{E}[T_{i+1} - T_i] = \frac{n-1}{n}\sum_{i=1}^{n-1}\Big(\frac{1}{i} + \frac{1}{n-i}\Big) \\ &= 2\frac{n-1}{n}\sum_{i=1}^{n-1}\frac{1}{i} \sim 2\log n. \end{aligned} \tag{7}$$

Notation: For two sequences f_n and g_n, we write $f_n \sim g_n$ (read f_n is asymptotically equivalent to g_n) to mean that $\lim_{n\to\infty} f_n/g_n = 1$.

To show that $\sum_{i=1}^{n} \frac{1}{i} \sim \log n$, note that the sum is bounded below by $\int_0^n \frac{1}{x+1} dx$ and above by $1 + \int_1^n \frac{1}{x} dx$.

Thus, we have shown that the mean time needed for the rumour to spread to all nodes in a population of size n scales as $2 \log n$. We shall show that, in fact, the random time concentrates closely around this value. In order to do so, we first need to compute its variance. Recall that the random variables $T_{i+1} - T_i$ for successive i are mutually independent. Hence, $\mathrm{Var}(T_n) = \sum_{i=1}^{n-1} \mathrm{Var}(T_{i+1} - T_i)$, and we obtain using (6) that

$$\mathrm{Var}(T_n) = \Big(\frac{n-1}{n}\Big)^2 \sum_{i=1}^{n-1} \Big(\frac{1}{i^2} + \frac{1}{(n-i)^2} + \frac{2}{n}\big(\frac{1}{i} + \frac{1}{n-i}\big)\Big) \sim \frac{\pi^2}{3}. \quad (8)$$

We have used the fact that $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$ to obtain the last equivalence.

Now, let's apply Chebyshev's inequality to the random variable T_n, the time for the rumour to reach all nodes. Using the estimates for the mean and variance of T_n in (7) and (8), we find that the statement

$$\mathbb{P}(|T_n - 2 \log n| \geq c) \leq \frac{\pi^2}{3c^2}$$

is approximately true for large n. More precisely, for every $\epsilon > 0$, it holds for all n sufficiently large that

$$\mathbb{P}(|T_n - 2 \log n| \geq c) \leq \frac{(1+\epsilon)\pi^2}{3c^2},$$

which tends to zero as c tends to infinity. In words, what this says is that the random variable T_n grows roughly like $2 \log n$. The fluctuation around this mean value does not grow unboundedly with n, but remains bounded.

3 Rumour spreading on general graphs

In the last section, we saw that the time it takes for a rumour to spread on the complete graph (for the specific model considered) grows logarithmically in the population size. What can we say more generally? How does the shape of a graph affect the spreading time?

Let $G = (V, E)$ be a directed graph on n nodes, and let P be an $n \times n$ stochastic matrix with the property that $p_{ij} = 0$ if $(i, j) \notin E$ (i.e., the

non-zero entries in P correspond to edges in the graph G). We consider the following rumour spreading model. There are n independent unit rate Poisson processes, one associated with each node. At an increment time of the Poisson process at node i, this node becomes active and picks a neighbour j with probability p_{ij}, the ij^{th} element of the matrix P. If i knows the rumour at this time and j doesn't, then i pushes the rumour to j. Otherwise, nothing happens. Note that if G is the complete graph, and P is the matrix with zero diagonal elements and all off-diagonal elements equal to $1/(n-1)$, then we recover the model of the previous section.

Again, we start with a single node s (called the source node) which is initially informed of the rumour, and are interested in the random time T until all nodes become informed. The dynamics can be modelled as a Markov process if we take the state to be the *set of informed nodes* at that time. It is not enough to keep track of the number of informed nodes, as the future dynamics depend on where in the network these nodes are located. Likewise, the distribution of the random variable T may well depend on which node we start with as the source of the rumour. As mentioned in the last section, the state space becomes very large (of size 2^n) if we have to use the subset of infected nodes as the state variable. This makes it a lot harder to obtain estimates of the mean and the variance that are sharp. What we shall do in this section instead is derive an *upper bound* on the rumour spreading time T based on some simple properties of the graph G and the matrix P that determines the choice of communication contacts.

As in the previous section, it is helpful to go from a node-centric to an edge-centric description. Pick a directed edge $(i,j) \in E$. What is the probability that i attempts to push the rumour to j in some infinitesimal time interval dt? Two things need to happen for this: node i has to be activated, and it has to choose j as its contact. The first event has probability $1 \cdot dt$ (as the Poisson process at node i is of unit rate), the second has probability p_{ij}, and they are independent of each other and of the past of all processes. In other words, the process of communications along the directed edge (i,j) is a Poisson process of rate p_{ij}, and communications along different edges are mutually independent. We shall make use of the following definition in our analysis of the rumour spreading time.

Definition. The *conductance* of the non-negative matrix P is defined as

$$\Phi(P) = \min_{S \subset V, S \neq \emptyset} \frac{\sum_{i \in S, j \in S^c} p_{ij}}{\frac{1}{n}|S| \cdot |S^c|}. \tag{9}$$

The minimum is taken over all non-empty proper subsets S of the vertex set V, S^c denotes the complement of S, and $|S|$ denotes the size of the set S.

As in the analysis for the complete graph, let T_k denote the first time that exactly k nodes are informed of the rumour. Thus, $T_1 = 0$ and T_n is the time that all nodes are informed. Let S_k denote the (random) subset of nodes that are informed at time T_k. What can we say about $T_{k+1} - T_k$? For each node $i \in S_k$ and $j \in S_k^c$, the events of i contacting j occur according to a Poisson process of rate p_{ij}. Moreover, these are mutually independent for distinct ordered pairs of nodes. Thus, using the fact that the superposition of independent Poisson processes is a Poisson process with the sum of their rates, we see that the total rate at which an informed node contacts an uninformed node and informs it of the rumour is given by $\sum_{i \in S_k, j \in S_k^c} p_{ij}$. (Contacts between nodes within S_k, or within S_k^c, result in no change of the system state, so we can ignore them.) Hence, conditional on S_k, the time $T_{k+1} - T_k$ until an additional node becomes informed is exponentially distributed with parameter $\sum_{i \in S_k, j \in S_k^c} p_{ij}$. Consequently,

$$\mathbb{E}[T_{k+1} - T_k | S_k] = \frac{1}{\sum_{i \in S_k, j \in S_k^c} p_{ij}}. \tag{10}$$

In order to compute $\mathbb{E}[T_n]$ exactly, we would have to consider every possible sequence of intermediate sets along which the system can go from just the source being informed to all nodes being informed, computing the probability of the sequence and using the conditional expectation estimate above. This is impractical for most large networks. Instead, we shall use the conductance to bound the conditional expectation of $T_{k+1} - T_k$. Observe from (9) and (10) that

$$\mathbb{E}[T_{k+1} - T_k | S_k] \leq \frac{1}{\Phi(P)} \frac{n}{k(n-k)} = \frac{1}{\Phi(P)} \Big(\frac{1}{k} + \frac{1}{n-k} \Big). \tag{11}$$

We have used the fact that $|S_k| = k$ by definition, and so $|S_k^c| = n - k$. Note that while the exact conditional expectation of $T_{k+1} - T_k$ depends on the actual set S_k of informed nodes at time T_k, the bound does not; it only depends on k, the number of informed nodes at this time.

We can use this bound to easily obtain a bound on the expected time to inform all nodes, following the same steps as in the analysis of the complete graph. First, write $T_n = \sum_{i=1}^{n-1} T_{i+1} - T_i$ since $T_1 = 0$. Next, use the linearity

of expectation, and the bound in (11), to get

$$\mathbb{E}[T_n] \leq \sum_{k=1}^{n-1} \frac{1}{\Phi(P)} \Big(\frac{1}{k} + \frac{1}{n-k} \Big) = \frac{2}{\Phi(P)} \sum_{k=1}^{n-1} \frac{1}{k} \sim \frac{2 \log n}{\Phi(P)}. \tag{12}$$

The above expression is a bound on the expected value of the random variable T_n. Can we also something about the distribution of the random variable. Using Markov's inequality, we obtain

$$\mathbb{P}\Big(T_n > \frac{c \log n}{\Phi(P)} \Big) \leq \frac{\Phi(P) \mathbb{E}[T_n]}{c \log n} \leq \frac{2}{c},$$

which tends to zero as c tends to infinity. In other words, the random variable T_n is of order $\log n / \Phi(P)$ in probability.

Example. Let $G = (V, E)$ be the complete graph, and suppose $p_{ij} = 1/(n-1)$ for every ordered pair (i, j), $i \neq j$. This is exactly the model of rumour spreading on a complete graph that we first analysed. What does the bound tell us in this case? To answer that, we need to compute the conductance $\Phi(P)$ for this example. Fix a subset S of the node set consisting of k nodes, where k is not equal to zero or n. For each node in this set, there are $n - k$ edges to nodes in S^c. The communication rate on each of these edges is $1/(n-1)$. Hence, we get

$$\sum_{i \in S, j \in S^c} p_{ij} = \frac{k(n-k)}{n-1},$$

and so

$$\frac{\sum_{i \in S, j \in S^c} p_{ij}}{\frac{1}{n} |S| \cdot |S^c|} = \frac{n}{n-1},$$

irrespective of the choice of S. Hence, taking the minimum over S gives us $\Phi(P) = \frac{n}{n-1}$. Substituting this in (12), we obtain that $\mathbb{E}[T_n]$ is bounded by a quantity that is asymptotic to $2 \log n$, which is precisely the same as the exact analysis gave us. Thus, the bound is tight in this example. In general, of course, it won't be tight, but in many examples, it may be good enough to be useful, and yield at least the right scaling in n of the rumour spreading time, though not the exact constants.

4 Consensus on complete graphs: the classical voter model

Consider a population of n individuals, each of whom has a preference for one of two political parties, which we shall denote by 0 and 1. They interact in some way, and change their opinion as a result of the interaction. In practice, their opinion would also be influenced by external factors, such as the policies or performance of the parties, but we don't consider that in our models. We want to know how the interaction alone influences the evolution of preferences over time.

There are many possible ways to model interactions. It may be that individuals change their opinion if some fraction of their friends or family have a different opinion and influence them. Individuals may differ in how big this fraction needs to be before they change their mind, and also in the relative weights they ascribe to the opinions of different people in their social circle. While it is possible to build models incorporating many of these features, we shall consider a much simpler model, described below.

We can think of the n individuals as the nodes of a complete graph, K_n. The nodes are started in an arbitrary initial state $\mathbf{X}(0) = \{X_v(0), v \in V\}$, where $X_v(0) \in \{0, 1\}$ specifies the initial preference of node v. There are n independent unit rate Poisson processes, one associated with each node. At every time that there is an increment of the Poisson process associated with node v, node v becomes active, chooses a node w uniformly at random from the set of all nodes (including itself), and copies the preference of node w at that time. Thus, individuals in this model are easily persuaded and show no resistance to changing their mind, which is obviously unrealistic. However, it leads to tractable models.

It should be clear from the description above that $\mathbf{X}(t), t \geq 0$ is a continuous-time Markov process as, given the state $\mathbf{X}(t)$ at time t, both the time to the next jump, and the state reached after that jump are independent of the past of the process before time t. Letting e_v denote the unit vector with a 1 corresponding to node v and zeros for all other elements, we can write the

transition rates for this Markov process as follows:

$$\begin{aligned} q(\mathbf{x}, \mathbf{x} + e_v) &= \frac{1}{n}(1 - x_v) \sum_{w \in V} x_w, \\ q(\mathbf{x}, \mathbf{x} - e_v) &= \frac{1}{n} x_v \sum_{w \in V} (1 - x_w). \end{aligned}$$

The $1 - x_v$ term in the first equation says that an increment in the v^{th} element of $\mathbf{x}$ is possible only if this element is 0. In that case, it changes to 1 at rate $1/n$, the contact rate between any two nodes, times the number of nodes which are in state 1, which is given by $\sum_{w \in V} x_w$. Likewise, the second equation gives the rate for node v to move from state 1 to state 0.

The state space of the Markov process $\mathbf{X}(t), t \geq 0$ is $\{0, 1\}^V$, the set of $0-1$ valued vectors indexed by the vertex set. Clearly, the all-0 and all-1 vectors, which we'll denote $\mathbf{0}$ and $\mathbf{1}$ are absorbing states; no vertex can change state if the system has reached either of these states. Moreover, all other states form a single communicating class as it is possible to move from any of them to any other. Hence, the Markov process eventually hits one of these two absorbing states, and becomes absorbed. We say that consensus is reached, either on the value 0 or the value 1, which is adopted by all nodes. The questions we want to address are:

- How likely are we to reach consensus on, say, the value 1, given the initial state?
- How long does it take to reach consensus?

These are the questions we shall address in the remainder of this section.

4.1 Hitting probabilities

We begin by observing that the process representation above yields a Markov chain on 2^n states, where $n = |V|$ is the number of nodes. This is unnecessarily large. Given the symmetry in the model, it suffices to keep track of the number of nodes in each state. If we let $Y(t) = \sum_{v \in V} X_v(t)$ denote the number of nodes in state 1 at time t, then $Y(t), t \geq 0$ is a Markov process on the much smaller state space $\{0, 1, 2, \ldots, n\}$. The states 0 and n are absorbing, and our questions above reduce to determining the htting probabilities

of each of these states, as well as the absorption time. First, note that the transition rates for the new Markov process $Y(t)$ are given by

$$q(k, k+1) = \frac{(n-k)k}{n}, \quad q(k, k-1) = \frac{k(n-k)}{n}. \tag{13}$$

To see this, note that a transition from k to $k+1$ happens if any of the $n-k$ nodes in state 0 becomes active (which happens at a total rate of $n-k$) and chooses one of the nodes in state 1 to contact (which has probability k/n). Likewise, the second term corresponds to one of the k nodes in state 1 becoming active and choosing a node in state 0 as its random contact. In order to compute the hitting probabilities of the two absorbing states 0 and n, starting from an arbitrary initial state, we shall make use of results from the theory of martingales, which we now introduce, first in discrete time and then in continuous time.

Definition A stochastic process $X_t, t \in \mathbb{N} \cup \{0\}$ is called a martingale if $\mathbb{E}[X_{t+1}|X_t, X_{t-1}, \ldots, X_0] = X_t$ for all t.

Note that this is different from the Markov property. It doesn't say that the probability distribution of X_{t+1} given the past depends only on X_t; it only says that the mean only depends on X_t. However, it is very restrictive about the form of this dependence, as it says that the mean has to be equal to X_t. Intuitively, we think of a martingale as representing one's fortune in a fair game of chance. The fortune after the next play (next roll of dice or spin of the rouletter wheel or whatever) is random, but is equal in expectation to the current fortune. An example would be a game of dice in which you win five times your bet (plus your stake) if the die comes up 6, and lose your stake otherwise. Note that this game would be a martingale whatever fraction of your wealth you decided to stake each time.

It is clear, by induction, that $\mathbb{E}[X_{t+s}|X_t, X_{t-1}, \ldots, X_0] = X_t$ for all $s \geq 1$. Taking $t = 0$, $\mathbb{E}[X_s] = \mathbb{E}[X_0]$ for all $s \geq 1$. In fact, it turns out this relationship not all holds for all fixed (deterministic) times s, but also for random times satisfying certain conditions.

Definition A random time T is called a stopping time for a stochastic process $X_t, t \in \mathbb{N} \cup \{0\}$ if the event $T = t$ is measurable with respect to $\{X_s, s \leq t\}$.

In less measure-theoretic language, the random T is a stopping time if you can decide whether $T = t$ just by observing $X_s, s \leq t$. In other words, T is a function of the past and present, not of the future.

Example. Let X_t be the fortune after t time steps in a game where you bet 1 pound repeatedly on rolls of a die where you get 6 pounds (including your stake) if the die comes up 6. Suppose your initial fortune X_0 is 10 pounds. Let T_1 be the first time that your fortune is either 20 pounds or zero. Let T_2 be the time one time step before your fortune hits zero. Then T_1 is a stopping time, whereas T_2 is not. (Both are perfectly well-defined random variables on the sample space of infinite sequences of outcomes of rolls of the die.)

Theorem 3 *(Optional Stopping Theorem) Let $X_t, t \in \mathbb{N} \cup \{0\}$ be a bounded martingale (i.e., there is a finite constant M such that $|X_t| \leq M$ for all $t \geq 0$), and let T be a stopping time for it. Then $\mathbb{E}[X_T] = \mathbb{E}[X_0]$.*

The requirement that the martingale be bounded can be relaxed, but needs to be replaced with conditions that are more complicated to state, and to check in applications. It will suffice for us to confine ourselves to the bounded case.

The definitions and theorem above extend to continuous time martingales. We won't restate the theorem, or the definition of a stopping time, which are identical, but just the definition of a martingale, which extends in the obvious way.

Definition A continuous-time stochastic process $X_t, t \geq 0$ is called a martingale if $\mathbb{E}[X_{t+s}|(X_u, u \leq t)] = X_t$ for all $s, t \geq 0$.

Let us now go back to the continuous-time Markov process $Y_t, t \geq 0$ representing the number of nodes in state 1. We saw in equation (13) that the rates for going from k to $k+1$ and $k-1$ are equal. Hence, it is equally likely that the first jump from state k is to either of these states, which implies the following lemma.

Lemma 3 *The stochastic process $Y_t, t \geq 0$ is a martingale.*

Proof. Fix $t \geq 0$. If $Y_t = 0$ or $Y_t = n$, then Y_t remains constants at all subsequent times, and hence $\mathbb{E}[Y_s|(Y_u, u \leq t)] = Y_t$ for all $s \geq t$. Hence, it only remains to verify this equality if $Y_t = k$ for some $k \in \{1, 2, \ldots, n-1\}$. We suppose from now that this is the case.

Clearly, Y_s is constant and equal to Y_t for all $s \in (t, \tau)$ where τ is the random time of the first jump after time t. Hence $\mathbb{E}[Y_s|(Y_u, u \leq t)] = Y_t$ for

all $s \in (t, \tau)$. Moreover,

$$\mathbb{E}[Y_\tau | (Y_u, u \leq t)] = \frac{1}{2}(Y_t + 1) + \frac{1}{2}(Y_t - 1) = Y_t.$$

Thus, the equality $\mathbb{E}[Y_s | (Y_u, u \leq t)] = Y_t$ holds for all s up to and including the first jump time after t. By induction on the sequence of jump times, it holds for all $s \geq t$.

This way of proving that Y_t is a martingale basically reduces the continuous-time process to a discrete-time process watched at the jump times. Another approach is to look at the change over infinitesimal time intervals. Observe from equation (13) that

$$\mathbb{E}[Y_{t+dt} - Y_t | (Y_u, u \leq t)] = (+1) \cdot \frac{(n-k)k}{n} dt + (-1) \cdot \frac{(n-k)k}{n} dt = 0,$$

since the possible values of $Y_{t+dt} - Y_t$ are $+1$ or -1, and we multiply them by the corresponding rates and take the average. (Jumps of 2 or more have probability $o(dt)$, which we ignore.) The fact that the change in conditional expectation is zero over infinitesimal time intervals implies that the conditional expectation is constant, and hence that Y_t is a martingale. □

It is now straighforward to compute the hitting probability of each of the absorbing states starting from any given initial state k. Let $T = \inf\{t \geq 0 : Y_t = 0 \text{ or } n\}$ denote the random time to absorption. Then, T is clearly a stopping time, and we have by the Optional Stopping Theorem that

$$\mathbb{E}[Y_T] = \mathbb{E}[Y_0] = k.$$

But

$$\mathbb{E}[Y_T] = n\mathbb{P}(Y_T = n) + 0\mathbb{P}(Y_T = 0),$$

and so

$$\mathbb{P}(Y_T = n) = \frac{\mathbb{E}[Y_T]}{n} = \frac{k}{n}.$$

This gives us the probability of hitting n before 0 as a function of the initial state k.

4.2 Hitting times

We shall derive bounds on the time to consensus by establishing a duality with coalescing random walks. Suppose we want to know whether, on a

specific sample path of the random process, consensus has been reached by a given time, τ. We can determine this by following the evolution of the process, given its initial condition, from time 0 to time τ. Alternatively, we can do so by following the evolution backwards from τ. Imagine that, at time τ, each node is occupied by a single particle.

Let $\tau - T_1$ denote the last time before τ (first time looking backwards) that there is a contact between any two nodes. Suppose that, at time $\tau - T_1$, a node denoted v_1 copies a node denoted u_1. As v_1 is involved in no further communications after time $\tau - T_1$, and neither is any other node, it is clear that the state of v_1 at time τ, and consequently the state of all nodes at time τ, is fully determined by the state of *all nodes other than* v_1 at time $\tau - T_1$. We represent this by saying that the particle at v_1 has moved to u_1 and coalesced with the particle there at time $\tau - T_1$.

As we continue following the process backwards from time $\tau - T_1$, there may be a time $\tau - T_2$ at which some node $v_2 \neq v_1$ copies some other node, denoted u_2. We are not interested in times before $\tau - T_1$ at which v_1 copies some other node, because that won't affect the final state of v_1. Again, we represent this by the particle at v_2 moving to u_2. If $u_2 \neq v_1$, then the moving particle coalesces with the one occupying u_2. If $u_2 = v_1$, then the particle at u_2 moves to v_1 but there is no particle there to coalesce with.

The above is a verbal description of the process, looking backwards in time from τ, but we would like a probabilistic model. What can we say about the random time T_1, which is the first time looking back from τ that a contact occured between two nodes? Nodes becomes active at the points of independent unit rate Poisson processes. It is worth noting that the time reversal of a Poisson process is also a Poisson process of the same rate. Hence, looking back from τ, node activation times are again independent unit rate Poisson processes. When a node v becomes active, it contacts a node u chosen uniformly at random (again, time reversal doesn't change this) and copies its state. In our description, this corresponds to the particle at v moving to u, and coalescing with the particle already there, if any. Thus, the process backwards from τ corresponds to particles performing independent continuous time random walks on the complete graph until they meet another particle and coalesce. Coalesced particles behave just like any other particle. We thus arrive at the following probabilistic model for the process looking back from τ.

Initially, there are n particles, one at each node of the complete graph.

Particles move independently of each other. Each particle waits for a random time exponentially distributed with unit mean, then moves to a node chosen uniformly at random (including itself). If there is a particle at that node, the two particles coalesce, and henceforth behave as a single particle obeying the same rules. It is clear from this description that the process can be modelled as a Markov chain $Z_t, t \geq 0$ on the state space $\mathcal{P}(V)$, the set of all subsets of the vertex set. The state at time t denotes the set of nodes occupied by a particle. The event that consensus has occured by time τ is then equivalent to the event that the set Z_τ of occupied nodes at time τ of the backwards-time process is a set of nodes which all have the same initial condition in the forward-time process. Thus, this event depends on the initial condition in general. We would like to obtain a bound on the consensus time that doesn't depend on the initial condition. The only way to ensure that all nodes in Z_τ have the same initial state, irrespective of the initial condition, is if Z_τ is a singleton set, i.e., consists of a single node. This is the case if all n particles have coalesced into a single particle. We will now analyse the distribution of the random time for this event to occur.

In a general graph, we would have to keep track of the locations of all extant particles, but as we are working on the complete graph, all nodes are identical. Hence, we only need to keep track of the number of particles at any time t, which we shall denote by W_t. Let T_k denote the first time that $W_t = k$. We have $T_n = 0$ as we start with n particles. We want to estimate T_1, the random time that all particles have coalesced to a single one. Now, what can we say about $T_{k-1} - T_k$? Each of the k particles alive at time T_k becomes active according to a unit-rate Poisson process, and moves to a node chosen uniformly at random from all n nodes (including its current location). Using the fact that Bernoulli splittings of Poisson processes are Poisson, we can model this by associating independent Poisson processes of rate $1/n$ with each directed edge of a complete directed graph on n nodes. If the Poisson clock on the directed edge (u, v) goes off, then the particle at u moves to v and coalesces with the particle there, if any. As we are only interested in coalescences, it is enough to look at those directed edges which link occupied nodes. As there are k occupied nodes, there are $k(k-1)$ such edges. Each of these has an independent rate $1/n$ Poisson clock associated with it. Using the fact that superpositions of independent Poisson processes are Poisson, the time until the clock on some edge between occupied nodes rings is given by an $Exp(k(k-1)/n$ random variable. When this happens,

two particles coalesce. Hence,

$$T_{k-1} - T_k \sim \mathrm{Exp}\Big(\frac{k(k-1)}{n}\Big), \quad \mathbb{E}[T_{k-1} - T_k] = \frac{n}{k(k-1)} = n\Big(\frac{1}{k-1} - 1k\Big).$$

Recalling that $T_n = 0$, we obtain that

$$\mathbb{E}[T_1] = \sum_{k=2}^{n} \mathbb{E}[T_{k-1} - T_k] = n.$$

As we argued above, the random time T_1 is an upper bound on the time to consensus. Thus, the above result tells us that the mean time to consensus is bounded above by n, the number of nodes. While we did not derive a lower bound, this is in fact the correct scaling relationship. A more detailed but tedious calculation shows that the mean time to consensus, starting from an initial condition in which a fraction α of nodes are in state 0, is given by $nh(\alpha)$, where $h(\alpha) = -\alpha \log \alpha - (1-\alpha)\log(1-\alpha)$ denotes the binary entropy function evaluated at α. Thus, the time to reach consensus on the complete graph scales linearly in the number of nodes. This is in contrast to the rumour spreading time, which is much smaller, scaling logarithmically in the number of nodes.

5 Consensus on general graphs

Let $G = (V, E)$ be a directed graph. Each node $v \in V$ can be in one of two states, 0 or 1. We denote by $X_v(t)$ the state of node v at time t, and by $\mathbf{X}(t)$ the vector, $(X_v(t), v \in V)$. The process by which nodes change their state is as follows. Associated with each directed edge $(v, w) \in E$ is a Poisson process of rate $q_{vw} > 0$, at the points of which node v copies the state of node w. The Poisson processes associated with distinct edges are mutually independent. It follows from this description that $\mathbf{X}(t)$ is a Markov process.

We assume that the graph G is *strongly connected*, i.e., that there is a *directed path* from v to w for every pair of nodes v, w. If there is such a directed path, then node w can influence node v. Clearly, the all-0 and all-1 states, which we denote by $\mathbf{0}$ and $\mathbf{1}$ respectively, are absorbing. The assumption says that every node can influence every other node, and hence that all states other than $\mathbf{0}$ and $\mathbf{1}$ form a single communicating class, from which both these states are accessible. Hence, the Markov chain eventually

reaches one of these absorbing states, and we want to know the probability of reaching **0** and **1** starting from an arbitrary initially state.

Note that the assumption of strong connectivity is essential for absorption in one of these two states to be guaranteed. To see this, we consider a counterexample. First, consider a graph consisting of 3 nodes: $V = \{u, v, w\}$ and $E = \{(vu), (v, w)\}$. Consider the initial state, $X_u(0) = 0$, $X_w(0) = 1$, and arbitrary $X_v(0)$. In this graph, both u and w can influence v but they can't influence each other and neither can influence the other. Hence, $X_u(t) = 0$ for all $t \geq 0$, $X_w(t) = 1$ for all $t \geq 0$, while $X_v(t)$ keeps oscillating between 0 and 1. Hence, there is no absorption in this example. Equally, if the graph is disconnected, there may be multiple absorbing states; for example, it is posssible that all nodes in one strongly connected component converge to the 0 state, while those in another converge to the 1 state.

Let Q be a rate matrix (infinitesimal generator matrix) with off-diagonal elements q_{vw} as specified above; the diagonal elements are then given by the requirement that the row sums should all be zero. Since Q is the rate matrix of a finite state Markov chain on the state space V, it has an invariant distribution π, namely a probability vector on the vertex set V which solves the global balance equations $\pi Q = \mathbf{0}$. Under the assumption that the graph is strongly connected, the invariant distribution is unique, and strictly positive (i.e., π_v is not zerofor any $v \in V$). Define $M(t) = \pi \mathbf{X}(t)$, the product of the row vector π and the column vector $\mathbf{X}(t)$. We now claim the following.

Lemma 4 *The stochastic process $M(t)$ is a martingale.*

Proof. Observe that

$$\mathbb{E}[M(t+dt) - M(t) | (\mathbf{X}(u), u \leq t)] = \sum_{v \in V : X_v(t) = 0} \pi_v \mathbb{P}(X_v(t+dt) = 1 | \mathbf{X}(t))$$
$$- \sum_{v \in V : X_v(t) = 1} \pi_v \mathbb{P}(X_v(t+dt) = 0 | \mathbf{X}(t)) + o(dt), \tag{14}$$

since the possible changes in the state $\mathbf{X}(t)$ over an infinitesimal time interval dt are the change in state of some single node from 0 to 1 or 1 to 0; by the definition of $M(t)$, a change in state of X_v from 0 to 1 increases $M(t)$ by π_v, while a change from 1 to 0 decreases it by π_v. Now, by the description

of the model,

$$\begin{aligned}\mathbb{P}(X_v(t+dt)=1|\mathbf{X}(t)) &= (1-X_v(t))\sum_{w:(v,w)\in E} q_{vw}X_w(t),\\ \mathbb{P}(X_v(t+dt)=0|\mathbf{X}(t)) &= X_v(t)\sum_{w:(v,w)\in E} q_{vw}(1-X_w(t));\end{aligned}$$

in each case, we add up the rates q_{vw} of contacting nodes w in the opposite state to node v. Substituting these transition probabilities in (14), we get

$$\begin{aligned}&\mathbb{E}[M(t+dt)-M(t)|(\mathbf{X}(u),u\le t)]\\ &=\sum_{v\in V:X_v(t)=0}\pi_v(1-X_v(t))\sum_{w:(v,w)\in E}q_{vw}X_w(t)\\ &-\sum_{v\in V:X_v(t)=1}\pi_v X_v(t)\sum_{w:(v,w)\in E}q_{vw}(1-X_w(t))+o(dt).\end{aligned}$$

We can extend both sums on the RHS above to all $v\in V$ since the restriction is effectively imposed by the terms $(1-X_v(t))$ and $X_v(t)$ inside the sums. Moreover, we can also extend the sum over w to all $w\in V$ if we define q_{vw} to be zero whenever $(v,w)\notin E$. Dropping the $o(dt)$ term as well for simplicity, we can rewrite the above in a less notationally cumbersome way as

$$\begin{aligned}&\mathbb{E}[M(t+dt)-M(t)|(\mathbf{X}(u),u\le t)]=\\ &\sum_{v,w\in V}\pi_v q_{vw}(1-X_v(t))X_w(t)dt-\sum_{v,w\in V}\pi_v q_{vw}X_v(t)(1-X_w(t))dt\\ &=\sum_{v,w\in V}\pi_v q_{vw}(X_w(t)-X_v(t))dt. \qquad (15)\end{aligned}$$

In fact, we should have been able to write down this expression directly. Changes in state may occur at points of the Poisson processes on the edges of the graph. These occur at rate q_{vw} on edge (v,w). If nodes v and w are in the same state, i.e., $X_w(t)-X_v(t)=0$, then there is no change. If $X_w(t)-X_v(t)=1$, i.e., node w is in state 1 and node v is in state 0, then $X_v(t)$ changes from 0 to 1, and M_t increases by π_v. Conversely, if $X_w(t)=0$ and $X_v(t)=1$, then $X_v(t)$ changes from 1 to 0, and $M(t)$ decreases by π_v. Now,

$$\sum_{v,w\in V}\pi_v q_{vw}X_v(t)=\sum_{v\in V}X_v(t)\pi_v\sum_{w\in V}q_{vw}=0,$$

since the row sums of the Q matrix are 0, whereas

$$\sum_{v,w\in V}\pi_v q_{vw}X_w(t)=\sum_{w\in V}X_w(t)\sum_{v\in V}\pi_v q_{vw}=0,$$

since π satisfies the global balance equations $\pi Q = \mathbf{0}$. Substituting in (15), we conclude that

$$\mathbb{E}[M(t+dt) - M(t)|(\mathbf{X}(u), u \leq t)] = 0,$$

which implies that $M(t)$ is a martingale, as claimed. □

It is now straightforward to compute the hitting probability of the all-zero and all-one states, starting from an arbitrary initial condition $\mathbf{X}(0)$. Let $M(0) = \pi\mathbf{X}(0)$ denote the value of the martingale corresponding to this initial condition. Let T denote the absorption time in one of the two absorbing states, and note that it is a stopping time. Hence, by the Optional Stopping Theorem, $\mathbb{E}[M(T)] = M(0) = \pi\mathbf{X}(0)$. But

$$\begin{aligned} \mathbb{E}[M_T] &= \mathbb{P}(\mathbf{X}(T) = \mathbf{1}) \cdot \pi\mathbf{1} + \mathbb{P}(\mathbf{X}(T) = \mathbf{0}) \cdot \pi\mathbf{0} \\ &= \mathbb{P}(\mathbf{X}(T) = \mathbf{1}) \cdot 1 + \mathbb{P}(\mathbf{X}(T) = \mathbf{0}) \cdot 0. \end{aligned}$$

It follows that

$$\mathbb{P}(\mathbf{X}(T) = \mathbf{1}) = \mathbb{E}[M(T)] = M(0) = \pi\mathbf{X}(0).$$

We do not consider the problem of estimating the consensus time on a general graph. While the duality with coalescing random walks holds just as on the complete graph, the resulting process is much more difficult to analyse as it requires keeping track of the locations of all the particles, and not just their number.

6 The contact process on a general graph

Let $G = (V, E)$ be an undirected graph with vertex set V and edge set E. The contact process, or SIS epidemic, on G, is described as follows. Each node can be in one of two states, infected (I) or healthy and susceptible to infection (S). Healthy nodes may become infected by infected neighbours in the graph, and infected nodes may spontaneously recover, but are then susceptible to re-infection. Let us denote by S_t the set of infected nodes at time t. The infection and recovery processes are described as follows. Each susceptible node j becomes infected at rate $\alpha|\{i \in S_t : (i,j) \in E\}|$. The parameter α is called the infection rate, and the set $\{i \in S_t : (i,j) \in E\}$ denotes the set of infected neighbours of the node j. Thus, each infected

neighbour can pass on infection at rate α. More formally, you can associate independent Poisson processes of rate α with each edge in the graph; if there is an increment of the Poisson process on edge (i, j) at time t, and one of these nodes is infected and the other susceptible at time t, then the other node also becomes infected. Next, every infected node recovers spontaneously at rate β, independent of everything else. Again, formally, associate independent Poisson processes of rate β with each node. If the Poisson process at node i has an increment at time t and this node is infected, then it becomes susceptible at time t. Otherwise, nothing happens.

It should be clear from the description that the set of infected nodes, S_t, evolves as a continuous time Markov chain. Its state space is $\mathcal{P}(V)$, the power set of V, namely the set of all subsets of the vertex set. The null set is absorbing, i.e., if $S_t = \emptyset$, then $S_u = \emptyset$ for all $u > t$; once all nodes are susceptible, no node can subsequently become infected. Since the graph G is connected, you should also be able to see that the Markov process $S_t, t \geq 0$ has to communicating classes, the null set $\emptyset$, and the set of all other subsets of V. The null set is recurrent, and the other communicating class is transient, so eventually the Markov chain hits the null set. In other words, eventually the infection dies out from the population. Starting with an arbitrary initial set S_0 of infected nodes, we would like to say something about the random time T until the infection dies out. Exact analysis doesn't seem feasible, but can we obtain bounds on the expectation of T or some other properties of its distribution?

In order to obtain such bounds, we shall first provide a different representation of the epidemic process, and then define an auxiliary Markov process that dominates it stochastically. We shall denote the state of the epidemic at time t by a vector $\mathbf{X}(t) = (X_v(t), v \in V)$ taking values in the state space $\{0, 1\}^V$. Here, $X_v(t) = 1$ denotes that node v is infected at time t, and $X_v(t) = 0$ denotes that it is healthy, and hence susceptible to infection. Then $\mathbf{X}(t), t \geq 0$ is a Markov process with the following transition rates:

$$\begin{aligned} q(\mathbf{x}, \mathbf{x} + e_v) &= \alpha(1 - x_v) \sum_{w:(w,v)\in E} x_w \\ q(\mathbf{x}, \mathbf{x} - e_v) &= \beta x_v. \end{aligned} \tag{16}$$

Here, e_v denotes the unit vector with a 1 in the v^{th} position and zeros elsewhere. The first equation above says that infection propagates at rate α from each infected neighbour of v, namely each neighbour w such that $x_w = 1$; however, node v can only become infected if it is susceptible, i.e.,

$x_v = 0$. The second equation says that node v can be cured at rate β, but only if it is infected, i.e., $x_v = 1$.

We shall now define an auxiliary Markov process $\mathbf{Y}(t), t \geq 0$, coupled with the process $\mathbf{X}(t)$, such that $\mathbf{Y}(t) \geq \mathbf{X}(t)$ for all $t \geq 0$, where the inequality holds coordinate-wise. The state space for this dominating process is $\{0, 1, 2, \ldots\}^V$. We first give a verbal description of this process, then write down its transition rates, and then describe the coupling of the two processes.

We can think of $\mathbf{Y}(t)$ as an epidemic process also, but imagine that there is an infinite susceptible population at each node. We think of $Y_v(t)$ as denoting the number of infected individuals at node v at time t. Each of these individuals transmits infection to each neighbouring node at rate α, i.e., at the points of a rate α Poisson process, with the Poisson processes corresponding to distinct individuals or edges being mutually independent. Recall that the superposition of independent Poisson processes is a Poisson process with the sum of their rates. Since each infected individual at a node w neighbouring v transmits infection to v at rate α, the total rate at which infection passes along the edge (w, v) at time t is given by $\alpha Y_w(t)$. Thus, we think of there being independent Poisson processes of rate $\alpha Y_w(t)$ on each directed edge (w, v) incident on node v. (The graph is undirected but we think of each undirected edge as corresponding to two directed edges, one in each direction, as the rates of transmission of infection may be different in the two direction.) Now, at the first time that there is an increment in any one of these Poisson processes, we see it as giving rise to an infection event at node v, resulting in one more individual becoming infected at this node. Conversely, each infected individual at node v becomes cured at rate β, independent of everything else, and so the number of infected individuals at node v decreases by one at rate βY_v. It follows from the description above that the process $\mathbf{Y}(t), t \geq 0$ is Markovian, with the following transition rates:

$$\begin{aligned} \tilde{q}(\mathbf{y}, \mathbf{y} + e_v) &= \alpha \sum_{w:(w,v)\in E} y_w \\ \tilde{q}(\mathbf{y}, \mathbf{y} - e_v) &= \beta y_v. \end{aligned} \tag{17}$$

The main difference to note from the transition rates $q(\cdot, \cdot)$ for the original epidemic process is that there is no $1 - y_v$ term; even if infection is already present at a node, new infections can pass to it, causing additional infections at that node. The effect of this change is that the transition rates $\tilde{q}(\cdot, \cdot)$ are now linear functions of the state. This makes the $\mathbf{Y}(t)$ process more

tractable analytically, as we shall see.

The $\mathbf{Y}(t)$ process is called a branching random walk; you can think of each infectious individual as giving birth to a random number of infectious individuals at neighbouring nodes before dying, each of which does the same. If you trace the descendants of an infected individual keeping track of where they are born, these look like a random walk on the graph.

We want to couple the $\mathbf{X}$ and $\mathbf{Y}$ processes in such a way that $\mathbf{X}(t) \leq \mathbf{Y}(t)$ (in the coordinate-wise sense) for all $t \geq 0$ provided this inequality holds for $t = 0$. Formally, we say that two or more random variables, or stochastic processes, are coupled if they are defined on the same probability space. If that is too abstract, an operational definition is that two random variables or stochastic processes are coupled if we simulate them using a computer program that draws upon the output of the same random number generator (i.e., uses a common sequence of random numbers). It doesn't have to use exactly the same random numbers to simulate both random variables - it may use overlapping subsets, in which case they will be dependent in some way, or it may use disjoint subsets, in which case they will be independent.

Example 1. Let X and Y be exponential random variables with parameters λ and μ respectively, and suppose $\lambda > \mu$. We could simulate X and Y independently, in which case there is non-zero probability for both the events $X \geq Y$ and $X \leq Y$. However, can we simulate them in such a way that the inequality $X \leq Y$ holds deterministically? We can achieve this by first simulating two independent random variables, X_1 from an $\text{Exp}(\mu)$ distribution and X_2 from an $\text{Exp}(\lambda - \mu)$ distribution, and then taking $Y = X_1$, $X = \min\{X_1, X_2\}$. Then X and Y have the desired distributions, but $X \leq Y$ deterministically.

Example 2. In the last example, we coupled two random variables. Now we consider a more complicated example, where we couple two random processes. Let $X_t, t \geq 0$ be a Poisson process of rate λ, and $Y_t, t \geq 0$ a Poisson process of rate $\mu < \lambda$. Can we construct the processes X_t and Y_t in such a way that $X_b - X_a \geq Y_b - Y_a$ for arbitrary intervals $[a, b]$? If we are to do this, then the set of points/events of the Y_t process (the times at which this process increases by 1) must deterministically be a subset of the set of points of the X_t process. We now show how this can be done.

Let T_1^X and T_1^Y denote the time of the first jump of the X_t and Y_t process respectively. Then, T_1^X and T_2^X are exponentially distributed with parameter λ and μ respectively. Since $\lambda > \mu$, we simulate two random variables

$Z \sim \text{Exp}(\lambda)$ and $W \sim \text{Exp}(\mu)$ as in Example 1, such that $Z \leq W$ deterministically. If $Z = W$, we set $T_1^X = T_1^Y = Z$, i.e., both the X_t and Y_t process have a point at the random time Z, which is their first point after time 0. We can then simulate the future of these two processes as if starting afresh at this time.

If $Z < W$, then we set $T_1^X = Z$, and defer the simulation of T_1^Y. Thus, the X_t process has its first jump at the random time Z, but this is not a jump time of the Y_t process. But, by the memoryless property of the exponential distribution, the residual time until the first jump of the Y_t process has an $\text{Exp}(\mu)$ distribution. So, we can again restart the simulation of the X_t and Y_t processes from the random time Z (where we put down a point for X, but not for Y in this case.)

By repeating this construction, we simulate a sequence of random times, some of which are increment times for both the X and Y process, while some are only increment times for the X process; there are no times which are only increment times for Y process. This gives a simultaneous (coupled) construction of the X_t and Y_t processes which has the property we wanted. This construction is similar in spirit to the way we shall construct the Markov processes $\mathbf{X}(t)$ and $\mathbf{Y}(t)$ (the epidemic process and the branching random walk) so that $\mathbf{X}(t)$ is dominated by $\mathbf{Y}(t)$ deterministically for all t.

Exercise. Explain a different way to simulate the Poisson processes X_t and Y_t of rates λ and μ simultaneously, so that $X_b - X_a \geq Y_b - Y_a$ for all intervals $[a, b]$ if $\lambda \geq \mu$. For your construction, you have available to you a random number generator that can generate a sequence of iid exponential random variables (of any specified parameter) and an independent sequence of iid Bernoulli random variables (also of any specified parameter).

We now describe briefly how to couple the epidemic process and the branching random walk in such a way that $\mathbf{X}(t) \leq \mathbf{Y}(t)$ for all $t \geq 0$. We assume that $\mathbf{X}(0) \leq \mathbf{Y}(0)$, which is a necessary condition for such a coupling to be possible. We shall describe the coupling by induction on the sequence of jump times. We begin by generating the first jump time of the $\mathbf{Y}(t)$ process. To do this, for each node v, we need to sample the time to the first infection at that node from an exponential distribution with parameter $\sum_{w:(w,v)\in E} \alpha Y_w(0)$. Independent of this, we sample an exponential with parameter $\beta Y_v(0)$ for the first cure time at this node (the first time at which one of the $Y_v(0)$ infected agents at this node becomes healthy). Similarly, we sample exponential infection and recovery times at each node, all mutually

independent of each other. The time to the first jump is then the minimum of all these exponential random variables, and it specifies what transition is going to take place at the first jump time. Based on this, we specify the corresponding transition in the $\mathbf{X}(t)$ process as described below.

Suppose that the first jump happens at the random time τ and corresponds to the transition $Y_v(\tau) = Y_v(0) + 1$, i.e., the number of infectious individuals at node v increases by 1. The rate of this transition was $\sum_{w:(w,v)\in E} \alpha Y_w(0)$ in the $\mathbf{Y}$ process. In the $\mathbf{X}$ process, the corresponding possible transition is from node v susceptible to node v infected. If node v was already infected at time 0, $X_v(0) = 1$, then nothing happens in the $\mathbf{X}$ process at time τ. If node v was healthy at time 0, then we set it to infected at time τ with probability

$$\frac{\sum_{w:(w,v)\in E} \alpha X_w(0)}{\sum_{w:(w,v)\in E} \alpha Y_w(0)},$$

and leave it susceptible with the residual probability. This construction ensures that this transition happens with the correct rate in the $\mathbf{X}$ process. It is feasible because the fraction above lies in the range $[0, 1]$ by the assumption that $\mathbf{X}(0) \leq \mathbf{Y}(0)$, and hence is indeed a probability. The point to note is that, Y_v increases by 1 at the random time τ, while X_v may or may not increase. Hence, $Y_v(\tau) \geq X_v(\tau)$ if $Y_v(0) \geq X_v(0)$. The inequality holds for all other nodes w as well, because both X_w and Y_w remain unchanged at time τ.

Next, what if the transition sampled (minimum of exponential random variables sampled) corresponded to a decrease from $Y_v(0)$ to $Y_v(0) - 1$ at the random time τ? For this to happen, we must have had $Y_v(0) \geq 1$, and the rate of this transition is $\beta Y_v(0)$. With probability $(\beta X_v(0))/(\beta Y_v(0))$, which is positive only if $X_v(0) = 1$ and corresponds to the ratios of the rates in the two processes, we force a transition from 1 to 0 in the $\mathbf{X}$ process at time τ. With the residual probability, the $\mathbf{X}$ process remains unchanged. Thus, it is possible for Y_v to decrease, while X_v remains unchanged. Could this result in $Y_v(\tau)$ being smaller than $X_v(\tau)$, even though $Y_v(0)$ was at least as big as $X_v(0)$? It turns out this is impossible by our construction above. Indeed, if Y_v decreases by 1 at time τ, then the probability that X_v also decreases is $X_v(\tau-)/Y_v(\tau-)$ (which denote the values just before time τ). Hence, either both had the same value just before time τ, in which case both decreased, or Y_v had a strictly bigger value, in which case the inequality $X_v(\tau) \leq Y_v(\tau)$ still holds.

Thus, we have coupled the construction of the first jump in the $\mathbf{X}$ and $\mathbf{Y}$ processes in such a way that the inequality $\mathbf{X}(t) \leq \mathbf{Y}(t)$ is preserved at all times up to and including the first jump, if it was true at time zero. By the Markov property, we can simulate both processes after the first jump time based only on the state at the first jump time. Hence, by induction, the inequality $\mathbf{X}(t) \leq \mathbf{Y}(t)$ holds for all $t \geq 0$ under the coupling described above. (*Warning:* Strictly speaking, the proof only works for all time if the number of jumps in any finite time remains finite. If infinitely many jumps happen within some finite, possibly random, time T^*, then the description of the process from one jump to the next doesn't tell us what happens after T^* - the process is undefined after T^*. The property that, with probability one, only finitely many jumps happen in finite time is called non-explosivity of the Markov process. Non-explosivity is automatic if all the diagonal terms in the rate matrix Q are uniformly bounded, as is the case in finite-state chains like the $\mathbf{X}$ process for instance, but is not the case for the $\mathbf{Y}$ process. Nevertheless, the $\mathbf{Y}$ process is also non-explosive.)

Starting from an arbitrary initial condition $\mathbf{X}(0)$ for the epidemic process, we want to obtain a bound on the random variable $T = \min\{t \geq 0 : \mathbf{X}(t) = \mathbf{0}$, which denotes the time for the Markov process $\mathbf{X}(t)$ to hit the unique absorbing state $\mathbf{0}$ in which all nodes are susceptible, i.e., for the epidemic to die out. We shall obtain this bound by considering the Markov process $\mathbf{Y}(t)$ started in the same initial state $\mathbf{X}(0)$, and using the fact that $\mathbf{X}(t)$ remains bounded by $\mathbf{Y}(t)$ at all subsequent times. We first state our bound as a theorem, and then give a proof of it. In order to state the result, we need to introduce some terminology.

The adjacency matrix of a graph $G = (V, E)$ on n nodes is an $n \times n$ matrix A with $\{0, 1\}$-valued entries; the element a_{ij} is 1 if (i, j) is an edge, and 0 if it is not. If the graph G is undirected, then its adjacency matrix A is symmetric. Hence, all its eigenvalues are real. Denote the eigenvalues by $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$. The spectral radius ρ of the matrix A is defined as the maximum modulus of all its eigenvalues. Therefore $\rho = \max\{|\lambda_1|, |\lambda_n|\}$ for an undirected graph. We now have the following upper bound on the epidemic survival time on an undirected graph.

Theorem 4 *Let $G = (V, E)$ be an undirected graph on n nodes, and let ρ denote the spectral radius of its adjacency matrix. Consider the SIS epidemic or contact process on G with pairwise infection rate α and cure rate β, and with arbitrary initial condition. If $\alpha\rho < \beta$, then the random time T for the*

epidemic to die out satisfies

$$\mathbb{P}(T > t) \leq ne^{-(\beta-\alpha\rho)t}$$

for all $t \geq 0$*, and consequently*

$$\mathbb{E}[T] \leq \frac{\log n + 1}{\beta - \alpha\rho}.$$

Proof. We shall use the equations (17) to derive a differential equation for $\mathbb{E}[\mathbf{Y}(t)]$, where $\mathbf{Y}(t)$ denotes the branching random walk bounding the epidemic process $\mathbf{X}(t)$ as described above. Observe from (17) that

$$\mathbb{E}[Y_v(t+dt) - Y_v(t)|\mathbf{Y}(t)] = \alpha dt \sum_{w:(w,v)\in E} Y_w(t) - \beta Y_v(t)dt + o(dt),$$

and so, taking expectations on both sides, dividing by dt and taking limits, we get

$$\frac{d}{dt}\mathbb{E}[\mathbf{Y}(t)] = (\alpha A - \beta I)\mathbb{E}[\mathbf{Y}(t)].$$

This is a linear differential equation, and has the solution

$$\mathbb{E}[\mathbf{Y}(t)] = e^{(\alpha A-\beta I)t}\mathbf{Y}(0) = e^{(\alpha A-\beta I)t}\mathbf{X}(0),$$

since we assume that the $\mathbf{Y}$ process is started in the same initial state as the $\mathbf{X}$ process. Now, since $\mathbf{X}(t) \leq \mathbf{Y}(t)$ for all $t \geq 0$, it follows that $\mathbb{E}[\mathbf{X}(t)] \leq e^{(\alpha A-\beta I)t}\mathbf{X}(0)$ for all $t \geq 0$. If we let N_t denote the total number of infected nodes at time t, then $N_t = \sum_{v\in V} X_v(t) = \mathbf{1}^T\mathbf{X}(t)$, where $\mathbf{1}$ denotes the all-1 vector of length n. Thus, we obtain that

$$\mathbb{E}[N_t] \leq \mathbf{1}^T e^{(\alpha A-\beta I)t}\mathbf{X}(0) \quad \forall\, t \geq 0. \tag{18}$$

Let $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ denote the eigenvalues of A (which can be ordered as they are real, since A is symmetric). It is straightforward to verify that the eigenvalues of $e^{(\alpha A-\beta I)t}$ are given by $e^{(\alpha\lambda_i-\beta)t}$, $i = 1, 2, \ldots, n$. Hence, the spectral radius of this matrix is $e^{(\alpha\rho-\beta)t}$, which decays exponentially in t by the assumption that $\alpha\rho < \beta$ in the statement of the theorem. We shall now use the fact that if a symmetric $n \times n$ matrix B has spectral radius η, then $\|B\mathbf{x}\| \leq \eta\|\mathbf{x}\|$ for arbitrary vectors $\mathbf{x} \in \mathbb{R}^n$, where $\|\cdot\|$ denotes the usual Euclidean norm of a vector. We will say a proof of this fact later in the course. Combining this fact with the Cauchy-Schwarz inequality, which

states that $|\mathbf{x}^T\mathbf{y}| \leq \|\mathbf{x}\|\|\mathbf{y}\|$ for arbitrary vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we obtain from (18) that

$$\mathbb{E}[N_t] \leq \|\mathbf{1}\| e^{-(\beta-\alpha\rho)t} \|\mathbf{X}(0)\| = e^{-(\beta-\alpha\rho)t}\sqrt{n}\sqrt{N_0} \leq n e^{-(\beta-\alpha\rho)t}. \qquad (19)$$

Next, observe that the event $\{N_t \geq 1\}$ is the same as the event $\{T > t\}$, as both say that the epidemic has not died out by time t; there is at least one infected node at time t. Hence, we obtain using Markov's inequality that

$$\mathbb{P}(T > t) = \mathbb{P}(N_t \geq 1) \leq \frac{\mathbb{E}[N(t)]}{1},$$

which, together with (19) yields the first claim of the theorem. To obtain the second claim, we use the fact that for any non-negative random variable X (i.e., a random variable which has zero probability of taking negative values), we can write $\mathbb{E}[X] = \int_0^\infty P(X > x)dx$, which you can verify by integrating by parts. Hence, combining the above bound on $P(T > t)$ with the trivial bound that probabilities are no bigger than 1, we get

$$\begin{aligned}
\mathbb{E}[T] &= \int_0^\infty \mathbb{P}(T > t)dt \leq \int_0^\infty \min\{1, ne^{-(\beta-\alpha\rho)t}\}dt \\
&= \int_0^{\log n/(\beta-\alpha\rho)} 1dt + \int_{\log n/(\beta-\alpha\rho)}^\infty ne^{-(\beta-\alpha\rho)t}dt \\
&= \frac{\log n}{\beta-\alpha\rho} + \int_0^\infty e^{-(\beta-\alpha\rho)s}ds = \frac{\log n + 1}{\beta-\alpha\rho}.
\end{aligned}$$

We made the change of variables $s = t - \frac{\log n}{\beta-\alpha\rho}$ to obtain the penultimate equality. This completes the proof of the theorem. □