Rumours, Consensus and Epidemics on Networks
Lecture Notes, CNI Summer School, July 2025

Ayalvadi Ganesh

July 2025

Supported by a UKIERI/ British Council-funded
Institutional Research and Mobility Partnership

1 Preliminaries: The Poisson Process

A stochastic process N_t, $t \geq 0$ is called a counting process if it is integer-valued, non-decreasing and (by convention) right-continuous. Intuitively, it counts the number of 'points' or 'events' up to any time t.

Definition We say that a counting process $N_t, t \geq 0$ is a Poisson process of rate (or intensity) $\lambda \in \mathbb{R}_+$ if any o fthe following equivalent conditions hold:

1. For all $0 \leq s < t$, he random variable $N_{t+s} - N_t$ is independent of $\{N_u, 0 \leq u \leq t\}$ (the process has independent increments) and $N_{t+s} - N_t$ has a Poisson distribution with mean λs, i.e.,
$$P(N_{t+s} - N_t = k) = \frac{(\lambda s)^k}{k!} e^{-\lambda s}, \quad k = 0, 1, 2, \ldots \tag{1}$$

2. For all $t, h \geq 0$,
$$\begin{aligned} P(N_{t+h} - N_t = 1) &= \lambda h + o(h) \\ P(N_{t+h} - N_t = 0) &= 1 - \lambda h + o(h) \\ P(N_{t+h} - N_t \geq 2) &= o(h), \end{aligned}$$
independent of $\{N_u, 0 \leq u \leq t\}$. Here, for a function f, we write $f(h) = o(h)$ if $f(h)/h \to 0$ as $h \to 0$.

3. Let $0 \leq T_1 \leq T_2 \leq \ldots$ be the increment times of the counting process $(N_t, t \geq 0)$, i.e.,
$$T_n = \inf\{t \geq 0 : N_t \geq n\}.$$
Then, the random variables $T_{n+1} - T_n$ are mutually independent and identically distributed (iid) with an $Exp(\lambda)$ distribution, i.e., $P(T_{n+1} - T_n \geq t) = \exp(-\lambda t)$ for all $t \geq 0$.

We establish the equivalence of [1], [2] and [3] by showing that [1] $\Rightarrow$ [2] $\Rightarrow$ [3] $\Rightarrow$ [1]. 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 because, by the independence of increments, T_1, $T_2 - T_1$, $T_3 - T_2$, ... are iid. Now, let Fdenote the cdf of T_1. We can write

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

where the third equality follows from [1] and the last equality from [2a]. The above equation implies 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 $\text{Exp}(\lambda)$ random variable. Thus, we have shown that T_1 has an $\text{Exp}(\lambda)$ 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 $P(N_s = 0) = e^{-\lambda s}$. 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 $j \leq k$. By conditioning on the time to the first increment, we have

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

Here, f is the density function of T_1, the time to the first increment. The second equality follows from the memoryless property of the exponential distribution, while we have used the exponential density and the induction hypothesis to obtain the third equality. The last equality is obtained by integration.

This completes the proof of the equivalence of the threee definitions. This is important because different definitions might be more convenient at different times.

Remark The reason that this process is called a Poisson process is that the number of points within an interval of duration t has a Poisson distribution, with mean λt. This way of thinking about the process permits an easy generalisation to higher dimensions. We define a point process in $\mathbb{R}^d$ to be a Poisson process with rate or intensity λ if the number of points in a Borel measurable set A has a Poisson distribution with mean $\lambda \text{Leb}(A)$, and the numbers of points in disjoint sets are mutually independent. Here, $\text{Leb}(A)$

denotes the Lebesgue measure of the set A. In fact, one can generalise further to inhomogeneous Poisson processes parametrised by an intensity measure. Let $\boldsymbol{\lambda}$ be a σ-finite Borel measure on $\mathbb{R}^d$. Then, the Poisson process with intensity $\boldsymbol{\lambda}$ is a point process such that the number of points in a Borel measurable set A has a Poisson distribution with mean $\boldsymbol{\lambda}(A)$ (which is well defined for A for which this is finite), and the numbers of points in disjoint sets are mutually independent. Of course, it is far from self-evident that such a stochastic process exists. Kolmogorov's extension theorem can be used to prove its existence. Poisson processes are a vast topic in their own right and we only need the simplest version of them. For further reading, see [7, 8].

We conclude this section with two important properties of Poisson processes, the proofs of which are relegated to the homework problems.

1. Let $\{N_t^i, t \geq 0, i = 1, 2\}$ be independent Poisson processes with rates λ_i, $i = 1, 2$. Let $N_t = N_t^1 + N_t^2$, $t \geq 0$, denote their superposition. Then $\{N_t, t \geq 0\}$ is a Poisson process of rate $\lambda = \lambda_1 + \lambda_2$. By induction, the property extends to the superposition of countably many Poisson processes.

2. Suppose $\{N_t, t \geq 0\}$ is a Poisson process of rate λ and that $\{X_i, i \in \mathbb{N}\}$ is a sequence of iid Bernoulli random variables with parameter p (denoted $Bern(p)$), independent of $\{N_t, t \geq 0\}$. Define $N_t^1 = \sum_{i=1}^{N_t} X_i$ and $N_t^2 = \sum_{i=1}^{N_t}(1 - X_i)$. Then, $\{N_t^1, t \geq 0\}$ and $\{N_t^2, t \geq 0\}$ are known as a Bernoulli splitting of the process $\{N_t, t \geq 0\}$, and are independent Poisson processes with rates λp and $\lambda(1-p)$ respectively. The property extends to splitting into countably many processes.

2 Rumour Spreading on Networks

How does gossip spread? How do technologies diffuse within or across societies? How do fads and fashions take hold? How do people learn about job opportunities or recruit workers? All these are examples of information spread over social networks, networks of friendship, family and acquaintance. They play a role in everything from the smooth functioning of economies to the maintenance of social and cultural groupings. Those are some of the questions that motivate our study of rumour/information spread in networks. Another motivation comes from synthetic networks like the Internet

or peer-to-peer networks, where it may be important to disseminate information rapidly; for example, security updates or routing tables. Rumour spreading mechanisms provide a template for many distributed algorithms on such networks. In these lectures, we shall look at some highly simplified mathematical models of information spread, with the goal of addressing precise questions about the speed of spread.

2.1 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.

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 $T_n = T$. 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,

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

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

$$\begin{aligned} \mathbb{E}[T_{i+1} - T_i] &= \frac{n-1}{n}\Big(\frac{1}{i} + \frac{1}{n-i}\Big), \\ \text{Var}(T_{i+1} - T_i) &= \Big(\frac{n-1}{n}\Big)^2\Big(\frac{1}{i^2} + \frac{1}{(n-i)^2} + \frac{2}{n}\big(\frac{1}{i} + \frac{1}{n-i}\big)\Big). \end{aligned} \tag{3}$$

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 (3) and the linearity of

expectation, we have

$$\begin{aligned}\mathbb{E}[T_n] &= \sum_{i=1}^{n-1} \mathbb{E}[T_{i+1} - T_i] = \frac{n-1}{n} \sum_{i=1}^{n-1} \Big(\frac{1}{i} + \frac{1}{n-i} \Big) \\ &= 2\frac{n-1}{n} \sum_{i=1}^{n-1} \frac{1}{i} \sim 2 \log n. \end{aligned} \tag{4}$$

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 rumour spreading time concentrates closely around this mean value. In order to do so, we 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 (3) that

$$\mathrm{Var}(T_n) = \Big(\frac{n-1}{n} \Big)^2 \sum_{i=1}^{n-1} \Big(\frac{1}{i^2} + \frac{1}{(n-i)^2} + \frac{2}{n} \big(\frac{1}{i} + \frac{1}{n-i} \big) \Big) \sim \frac{\pi^2}{3}. \tag{5}$$

We have used the fact that $\sum_{i=1}^{\infty} 1/i^2 = \pi^2/6$ to obtain the last equivalence. In order to use this variance estimate to show that the random variable T_n concentrates around its mean value, we will Chebyshev's inequality, which is an example of a probability inequality.

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

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

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

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

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

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

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.

Recall that a sequence of random variables X_n is said to converge in probability to a constant c, denoted $X_n \xrightarrow{p} c$ if, for all $\epsilon > 0$, $\mathbb{P}(|X_n - c| > \epsilon)$ tends to zero as n tends to infinity. We now establsih convergence in probability of the rumour spreading times, appropriately rescaled.

Theorem 1. *Let T_n denote the rumour spreading time on K_n, the complete graph on n nodes, in the model described above. Then,*

$$\frac{T_n}{\log n} \xrightarrow{p} 2 \text{ as } n \to \infty.$$

Proof. We apply Chebyshev's inequality to the scaled random variables T_n/n. Using the estimates for the mean and variance of T_n in (4) and (5), we obtain that

$$\mathbb{P}\Big(\Big| \frac{T_n - \mathbb{E}T_n}{\log n} \Big| \geq \frac{\epsilon}{2}\Big) \leq \frac{4\pi^2}{3\epsilon^2 \log^2 n} \to 0.$$

In addition, it holds deterministically that $\mathbb{E}T_n/\log n$ tends to 2 as n tends to infinity, and hence that $|\mathbb{E}T_n/\log n - 2| < \epsilon/2$ for all n sufficiently large. The claim of the theorem now follows from the triangle inequality. □

The above theorem says that the mean rumour spreading time, $\mathbb{E}[T_n]$, is asymptotic to $2\log n$, and that fluctuations around the mean are $o(\log n)$ in size with high probability (i.e., with probability tending to 1 as n tends to infinity). We now give a very precise characterisation of the fluctuations by showing that they are $O(1)$ in size and that the sequence of random variables $T_n - 2\log n$ converge in distribution to a limit. Before doing that, we need to recall the notion of convergence in distribution.

Let $X_n, n \in \mathbb{N}$ be a sequence of random variables, not necessarily defined on the same probability space, and let X be another random variable. Let F_n denote the cumulative distribution function (cdf) of X_n, and F the cdf of X. We say that the sequence of random variables X_n converges weakly to X, or converges in distribution to X, written $X_n \xrightarrow{d} X$, if either of the following equivalent statements hold:

1. At every real number x which is a continuity point of $F(\cdot)$, $F_n(x)$ tends to $F(x)$ as n tends to infinity. In particular, if F is a continuous function, then $F_n(x)$ tends to $F(x)$ for all x.
2. If $g : \mathbb{R} \to \mathbb{R}$ is a bounded, continuous function, then $\mathbb{E}[g(X_n)]$ tends to $\mathbb{E}[g(X)]$ as n tends to infinity.

A proof of their equivalence can be found in standard texts on weak convergence, and is not included here.

We are now ready to state the second main result of this section.

Theorem 2. *Let T_n denote the rumour spreading time on K_n, the complete graph on n nodes, in the model described above. Then,*

$$T_n - 2\log n \xrightarrow{d} \zeta_1 + \zeta_2,$$

where ζ_1 and ζ_2 are independent standard Gumbel random variables, i.e., $\mathbb{P}(\zeta_i \le x) = e^{-e^{-x}}$, $x \in \mathbb{R}$.

The proof is relegated to the problem sheets.

We described and analysed a continuous-time model of rumour spreading above. It has likely occurred to the reader that a discrete-time model is also natural. One could again start with a single node knowing the rumour. Then, in each time step, each informed node picks another node uniformly at random and informs it of the rumour (if it isn't already informed). This model has been studied in [3, 13], where it is shown that all nodes learn the rumour in around $\log_2 n + \log n$ time steps. Despite its apparent simplicity, the discrete-time model is in fact significantly harder to analyse, which is why we have chosen to focus on the continuous-time model, which was studied in [6].

2.2 First passage percolation

Consider the complete graph K_n and suppose that edges have random lengths (or weights) which are iid with an $Exp(1/n)$ distribution. If we fix two nodes u and v, what can we say about the length of the shortest path between u and v (where the length of a path is the sum of the lengths of the edges constituting it) and about the number of hops or edges along this shortest path? Next, if we fix a single node u, what can we say about the lengths of the shortest paths from u to all other nodes? What does the union of these shortest paths look like? It is almost surely a tree because, if there is a cycle, there must be two nodes x and y on this cycle such that the distance from x to y along two different paths is exactly equal. Can we say something about the height of this tree or the sum of edge weights on it? We now address some of these questions. The shortest path problem is also referred to as first passage percolation. Note that, given the edge lengths, there are well-known algorithms for solving the shortest path problem. Our focus here is not on these algorithms, but on the random variables arising from the solution. The model described above has been termed the stochastic mean field model of distance, and various optimisation problems on it have been studied, including the travelling salesman problem, the minimum spanning tree, etc.

We first note that first passage percolation is intimately connected with rumour spreading. We can identify the random edge lengths with the time it would take for a rumour to pass from one end-point of the edge to the other. Indeed, we saw when analysing the spread of a rumour that we can replace the node-based description of the process by an edge-based description whereby messages pass along each edge at the increments of a Poisson

process of rate $1/n$. In other words, messages pass along each edge after $Exp(1/n)$ random times. This is exactly the same as the first passage percolation (FPP) model described above.

One minor difference is that there is a single length associated with each edge in the FPP model, while there is a Poisson process in the rumour spreading model. However, if u communicates a rumour to v, then it is only the first random time sampled by u between when it hears the rumour and when it communicates it to v that is relevant; future communications are irrelevant. Another minor difference is that the rumour spreading model effectively happens on a directed graph (edges (u, v) and (v, u) sample independent exponentials) whereas FPP is on an undirected graph. But in rumour spreading, what matters is whether u first informs v or v first informs u; that the recipient of the message then communicates it back to the sender after a further random time is irrelevant. Thus, these models are in fact identical. We can use this observation to derive results about shortest paths and related objects in the FPP model.

Let $u \neq v$ be two vertices chosen uniformly at random in K_n. (Alternatively, the vertices are chosen first and then the edge lengths are sampled.) We are interested in $D(u, v)$, the random distance between u and v, namely the length of the shortest path between them. We now compute the mean of this random variable. As noted above, $D(u, v)$ has the same distribution as the time for a rumour started at u to reach v. We saw in the previous section that the time for a rumour started at u to reach k other nodes, denoted T_{k+1}, can be written as the sum of independent exponential random variables, $T_{j+1} - T_j \sim Exp(j(n-j)/n)$. Thus,

$$\mathbb{E}[T_{k+1}] = \mathbb{E}\Big[\sum_{j=1}^{k} T_{j+1} - T_j] = \sum_{j=1}^{k} \mathbb{E}[T_{j+1} - T_j] = \sum_{j=1}^{k} \frac{n}{j(n-j)}.$$

In order to compute the mean time until node v learns the rumour, we need to know in what order different nodes learn the rumour. Let $S(t)$ denote the set of nodes that know the rumour at time t, and let $S_k = S(T_k)$ be the first k nodes to know the rumour. Then, $S_1 = S(0) = \{u\}$. Define $K = \inf\{k : v \in S_{k+1}$. In other words, v is the K^{th} node (excluding the source, u) to learn the rumour. Now, it is clear by symmetry that all permutations of the order in which nodes learn the rumour are equally likely, and hence that K is uniform in $\{1, \ldots, n-1\}$. It follows from the tower rule

that the mean time until node v learns the rumour is given by

$$\begin{aligned}\mathbb{E}[T_{K+1}] = \mathbb{E}[\mathbb{E}[T_{K+1}|K]] &= \frac{1}{n-1}\sum_{k=1}^{n-1}\mathbb{E}[\mathbb{E}[T_{K+1}|K=k]] \\ &= \frac{1}{n-1}\sum_{k=1}^{n-1}\sum_{j=1}^{k}\frac{n}{j(n-j)} = \frac{1}{n-1}\sum_{j=1}^{n-1}\sum_{k=j}^{n-1}\frac{n}{j(n-j)} \\ &= \frac{1}{n-1}\sum_{j=1}^{n-1}\frac{n}{j} \sim \log n.\end{aligned}$$

We have thus shown that the expected distance between two typical points in the edge-weighted graph described above scales as $\log n$. In comparison, $\max_{v\in V} D(u,v)$ has the distribution of the time for a rumour started at u to reach all nodes, which we saw was asymptotic to $2\log n$, in expectation and in probability.

Next, we turn to the union of shortest paths from u to all other nodes. We already noted that this union must be a tree, which we will call the shortest path tree with root u, denoted $SPT(u)$. If we orient all its edges away from the root, then the path from u to any node v describes the path along which v first heard the rumour; the immediate predecessor of v along this path, which we call its parent, is the node from which v first heard the rumour, and so on.

We now study the construction and some properties of the SPT. We can construct the tree by following the rumour spreading process. Starting from u, the times at which other nodes hear the rumour from u are independent $Exp(1/n)$ random variables. Thus, the first node to hear the rumour is equally likely to be any of the other nodes. We sample this node, which we denote by u_1, uniformly at random from $V\setminus\{u\}$ and add it to the broadcast tree by drawing an edge between u and u_1. Once u_1 hears the rumour, it also starts spreading it. The residual times until any node w hears the rumour from u or u_1 are independent $Exp(1/n)$ random variables, and likewise for all nodes other than u and u_1. Hence, the next node to hear the rumour, denoted u_2, is drawn uniformly from $V\setminus\{u,u_1\}$, and it is equally likely to have heard the rumour from u or u_1. We add u_2 to the broadcast tree, attaching it with an edge to whichever of u or u_1 it first heard the rumour from. Continuing this process, at each step we add a node to the broadcast, attaching it to an existing node chosen uniformly at random. The resulting tree is known as the **uniform (random) recursive tree** (URT).

We shall compute the height of a node chosen uniformly at random within the *SPT*, as well as the height of the node added at the k^{th} step, for each k. Notice that at each time step, a new node is added to the tree, along with an edge to an existing node chosen uniformly at random. Let $\mathcal{T}_k$ denote the *URT* with k nodes, and H_k the height of a node chosen uniformly at random from $\mathcal{T}_k$; $\mathcal{T}_k$ is a random tree and H_k a random variable. Now consider a node sampled uniformly at random from $\mathcal{T}_{k+1}$. With probability $k/k+1$, it was already present in $\mathcal{T}_{\|}$, in which case its height has the same H_k. With the residual probability $1/k+1$, it is the new node that was added at step $k+1$. whose height is distributed like $H_k + 1$, since it was attached to a node chosen uniformly at random from $\mathcal{T}_k$. Thus, we see that

$$H_{k+1} = H_k + \xi_{k+1}, \text{ where } \xi_{k+1} \sim Bern\Big(\frac{1}{k+1}\Big) \text{ and } \xi_{k+1} \perp\!\!\!\perp H_k. \tag{6}$$

Clearly, $H_1 = 0$. Hence, we obtain from the above recursion that

$$H_n = \sum_{k=1}^{n-1} \xi_{k+1},$$

where the ξ_k are independent Bernoulli random variables, with parameters $1/k$. It follows that

$$\mathbb{E}[H_n] = \sum_{k=2}^{n} \frac{1}{k} \sim \log n, \quad \text{Var}(H_n) = \sum_{k=2}^{n} \Big(\frac{1}{k} - \frac{1}{k^2}\Big) \sim \log n.$$

As with the rumour spreading time, we can also prove a limit theorem for the sequence of random heights, H_n. Using the Lyapunov Central Limit Theorem for independent, but non-identically distributed summands, it is easy to show (as you will in the problem sheet) that:

Theorem 3. *Let H_n denote the height of a node chosen uniformly at random from a uniform random recursive tree with n nodes. Then,*

$$\frac{H_n - \log n}{\sqrt{\log n}} \xrightarrow{d} N(0,1),$$

where $N(0,1)$ denotes the standard normal distribution with zero mean and unit variance.

Next, let $\tilde{H}_k$ denote the random height of the k^{th} node in the tree, so that $\tilde{H}_1 = 0$. Since this node is attached to a pre-existing node chosen uniformly

at random, it is clear that $\tilde{H}_k$ has the same distribution as $H_{k-1}+1$. Thus, the results derived above for H_k carry over to $\tilde{H}_k$ with very minor modification. One can also ask about the height of the *URT*, namely the maximum height of all nodes in $\mathcal{T}_n$. It was shown in [14] that this random variable is asymptotic to $e \log n$. While the *URT* captures the shape of the *SPT*, it does not tell us about the weights or lengths of the edges in this tree. For a detailed exploration of these and other properties, we refer the reader to [18].

2.3 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 do the shape of a graph and the communication rates on different edges affect the spreading time?

Let $G = (V, E)$ be a directed graph on n nodes, and let R be an $n \times n$ non-negative matrix with entires $r_{ij} \geq 0$. A graph is implicitly specified by R if we interpret its non-zero entries as corresponding to directed edges. We consider the following rumour spreading model. There are independent Poisson processes, one associated with each directed edge; the Poisson process on edge (i, j) has rate r_{ij}. At an increment time of the Poisson process on edge (i, j), node i communicates the rumour to node j if i knows the rumour by this time. Otherwise, it does nothing. If G is the complete graph, and R is the matrix with all elements equal to $1/n$, 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 at any time 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 matrix R of contact rates. We shall make use of the following definition in our analysis of the rumour spreading time.

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

$$\Phi(R) = \min_{S \subset V, S \neq \emptyset} \frac{\sum_{i \in S, j \in S^c} r_{ij}}{\frac{1}{n}|S| \cdot |S^c|}. \tag{7}$$

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 r_{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 $\sum_{i \in S_k, j \in S_k^c} r_{ij}$. (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 $\sum_{i \in S_k, j \in S_k^c} r_{ij}$. Consequently,

$$\mathbb{E}[T_{k+1} - T_k | S_k] = \frac{1}{\sum_{i \in S_k, j \in S_k^c} r_{ij}}. \tag{8}$$

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 (7) and (8) that

$$\mathbb{E}[T_{k+1} - T_k | S_k] \leq \frac{1}{\Phi(R)} \frac{n}{k(n-k)} = \frac{1}{\Phi(R)} \Big(\frac{1}{k} + \frac{1}{n-k} \Big). \tag{9}$$

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 (9), to get

$$\mathbb{E}[T_n] \leq \sum_{k=1}^{n-1} \frac{1}{\Phi(R)} \Big(\frac{1}{k} + \frac{1}{n-k} \Big) = \frac{2}{\Phi(R)} \sum_{k=1}^{n-1} \frac{1}{k} \sim \frac{2 \log n}{\Phi(R)}. \tag{10}$$

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

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

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

Example. Let $G = (V, E)$ be the complete graph, and suppose $r_{ij} = 1/n$ 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 $\Phi(R)$ 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

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

and so

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

irrespective of the choice of S. Hence, taking the minimum over S gives us $\Phi(R) = 1$. Substituting this in (10), 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. In your homework problems, you'll see examples both of when the bound is good and when it is not.

3 Averaging and Consenus

Collective decision-making is ubiquitous in the natural world. For example, bees from a hive typically forage together and need to jointly decide between alternative food sources. An ant colony moving to a new nest have to collectively agree on the choice of nest site. These decisions appear to be made collectively rather than by a leader. Similarly, flocks of birds and shoals of fish exhibit complex patterns of movement in which the group retains cohesion while moving, without there being a clear leader or controller. Even bioluminescent bacteria appear to employ some form of quorum sensing to decide whether to luminesce. Biologists have long been interested in understanding how very simple organisms can collectively exhibit quite complex behaviours.

Social influence also plays a major role in human decision-making, for example in the adoption of products or technologies. Examples include competition between different video recording formats (VHS vs. Betamax), or operating systems (Windows vs. Apple Mac, or iPhone vs. Android), or social networks (Facebook vs. Instagram vs. TikTok). In some of these cases, there is a substantial switching cost (buying new hardware), while in others the switching cost may be small (time to set up a new account and upload material). There are often "network externalities" - benefits to adopting one of the competing technologies that don't have to do with its intrinsic qualities, but only to do with who and how many have adopted it. In the examples above, developers are more likely to produce films or software for the more popular platform, where they can expect to sell more copies. In the social network example, you would prefer to be on the network that most of your friends are on, because that would make it easier to communicate with them. These factors render the decision collective rather than individual. A consequence is that the question of which technology gets adopted may depend not just on its quality, but also on the random decisions of early adopters. Hence, understanding these mechanisms is of

interest to economists and social scientists.

Computer scientists are also interested in such problems. For example, many applications involve geographically distributed databases. Updates at one copy need to be propagated to the others and conflicts resolved. The task of maintaining consistency across the copies corresponds to a consensus problem. Computer scientists are particularly interested in solutions to the consensus problem that can achieve Byzantine fault tolerance, i.e., that are resilent to a small number of faulty or malicious agents. A topical instance of such problems arises in blockchains, where forking may occur because an update has not completed propagating across the network before a further modification is made to the chain. This can lead to inconsistent copies, which need to be reconciled by a consensus mechanism. Another topical engineering application is in swarm robotics, where large swarms of simple and low cost robots cooperate in order to perform some task, such as surverying a region, or carrying out search and rescue or providing communications support in disaster zones.

Not all collective decision problems involve consensus. In some cases, such as wireless spectrum access over a shared medium or route choice in networks, it is desirable that agents choose different options in order to reduce congestion and achieve load balancing. While this is also an important class of problems, we will focus on consensus problems as there are significant differences in the mathematical approach to studying these other coordination problems. Finally, we further distinguish between consensus problems in which the choice set is finite, such as the choice of foraging sites by bees, or video recording formats by consumers, and ones where it is a continuum, such as the choice of direction of movement in flocks of birds. We call the latter averaging problems; here, it is possible for consensus to never be attained in finite time, but the values of different agents may converge to a common value. This is the class of problems we study first.

3.1 The de Groot model

The first model we look at involves agents who start out with an initial preference or opinion that is real-valued, and who want to reach a consensus on a real number that is a (possibly weighted) average of these initial values. An example from above is the case of a flock of birds agreeing on a direction. (In this case, the values are in $\mathbb{R}^3$, but the idea is the same.) The de Groot model is very simple, but was one of the first formal models of "social

learning".

The model can be stated precisely as follows. There are n agents, who each have an initial preference or opinion or value in $\mathbb{R}$. We denote these initial values by the vector $\mathbf{x} = (x_1, x_2, \ldots x_n)$. The agents are located on the nodes of a directed graph $G = (V, E)$, and can only communicate directly with their neighbours in this graph. Time is discrete, and the agents update their values synchronously as follows. Agent i updates its value in time step $k + 1$ to

$$x_i(k+1) = \sum_{j \in V:(i,j) \in E} w_{ij} x_j(k).$$

Here, the w_{ij} are fixed weights, describing how much weight i gives to the opinion of j in updating its own opinion. We assume that the w_{ij} are non-negative and sum to 1, so that $x_j(k+1)$ is a weighted average of the opinions of j's neighbours at time k. In other words, W is a stochastic matrix, and we can describe the evolution of opinions by the linear recursion

$$\mathbf{x}(k+1) = W\mathbf{x}(k). \tag{11}$$

The process is deterministic. We are interested in its long-term behaviour. Do the opinions of each of the agents converge in the long run, i.e., does $x_j(t)$ tend to some x_j as t tends to infinity? Morever, do the agents reach (or approach) agreement in that their opinions all converge to the same value x? If so, we say that the agents asymptotically reach consensus. Another question we could ask is how quickly this consensus is reached. We first consider a couple of examples.

Examples

1. $$W_1 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

   In this example, the first two agents assign no weight to the third, and vice versa. So the third agent never changes its value, while the first two agents reach consensus on the average of their values after a single time step. Unless this average happens to be the same as the original value of the third agent, the values of all agents do not converge, i.e., consensus is never reached.

   It is clear that the problem arises because the matrix W is not irreducible (by which we mean that the Markov chain with transition

probability matrix W is not irreducible). Is it the case, then, that irreducibility is necessary for consensus? The answer is no, as the following example illustrates.

2. $$W_2 = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/2 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

   As before, agents 1 and 2 average their values, while paying no attention to agent 3. But now, instead of ignoring agents 1 and 2, agent 3 copies the value of agent 1. It is clear that, in time step 1, agents 1 and 2 will both update their values to $(x_1(0) + x_2(0))/2$, and will keep this value in all subsequent time steps. Agent 3 will also adopt this value in time step 2, and will keep it from then onwards. Thus, the agents reach consensus in time step 2.

   While neither W_1 nor W_2 is irreducible, the key difference is that there is a unique closed communicating class (comprised of states 1 and 2) in W_2, while there are two closed classes in W_1 ($\{1, 2\}$ and $\{3\}$). This is the essential feature. The transient states (i.e., the corresponding agents) have no impact on the value of any agent in the long run. Thus, we may ignore them from the point of view of consensus. If we restrict attention to the recurrent states, and the corresponding submatrix of the weight matrix (which can be seen to be stochastic as well), then we obtain an irreducible matrix from W_2, but not from W_1.

   Henceforth, we will ignore transient states in our discussion of the de Groot model, and restrict attention to the recurrent states. If we do that, then irreducibility is indeed a *necessary* condition to achieve consensus?

   Is it also *sufficient*? The following example shows that it is not.

3. $$W_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

   In each round, agent 1 copies agent 2 and vice versa, so they simply keep exchanging values, without converging to a common value. Thus, even though W_3 is irreducible, there is no consensus.

   The reader will no doubt have noticed that the problem is that the Markov chain with transition probability matrx W_3 is periodic, with

period 2. It is not hard to see that if the weight matrix is periodic, then consensus will not be achieved (except for special initial conditions). Thus, aperiodicity is also *necessary* for consensus, along with irreducibility.

Are irreducibility and aperiodicity together *sufficient* for consensus? The answer is yes, by the ergodic theorem for Markov chains. We provide a proof below.

Theorem 4. *Consider the de Groot model recursion* $\mathbf{x}(k+1) = W\mathbf{x}(k)$, $k = 0, 1, 2, \ldots$, *and suppose that the stochastic weight matrix* W *is irreducible and aperiodic. Let* π *be a row vector denoting the invariant distribution of the Markov chain with transition probablity matrix,* W*. (Note that* π *is unique by the irreducibility of* W*.) Then,*

$$\mathbf{x}(k) \to \mathbf{x}(\infty) = (\pi\mathbf{x}(0))\mathbf{1} \text{ as } k \to \infty,$$

where $\mathbf{1}$ *denotes the all-1 column vector.*

Proof. The solution of the de Groot recursion is given by $\mathbf{x}(k) = W^k\mathbf{x}(0)$. Now, by the ergodic theorem for irreducible, aperiodic Markov chains, W^k_{ij} tends to π_j as k tends to infinity, for any i. In other words, W^k tends to the rank-1 matrix, $\mathbf{1}\pi$. It follows that

$$\mathbf{x}(k) \to \mathbf{1}\pi\mathbf{x}(0),$$

as claimed. □

The theorem says that the values at all agents converge to the common value of $\pi\mathbf{x}(0)$, i.e., consensus is achieved. Observe that the value on which consensus is achieved is a weighted linear combination of the initial values at the different agents, with weights given by the vector π. Thus, π_j can be thought of as the relative *influence* of agent j in determining the consensus value.

Example Let $G = (V, E)$ be a connected, undirected graph with vertex set V and edge set E. We identify the vertices with agents, and the edges with which agents can influence which other ones. Suppose agents update their opinions according to the de Groot mode, assiqning equal weight to each of their neighbours. Then the weight matrix W has elements $w_{ij} = 1/deg(i)$, where $deg(i)$ denotes the degree of node i in the graph G.

In order to find the influence of different nodes, we need to compute the invariant distribution π, which is the probability distribution satisfying the

equation $\pi W = \pi$. It is easy to verify that if we take $\pi_i = deg(i)/2|E|$, where $|E|$ denotes the cardinality of the edge set, then π satisfies the detailed balance equations, $\pi_i w_{ij} = \pi_j w_{ji}$ for all $i, j \in V$. Hence π is the invariant distribution, and the Markov chain with transition matrix W is reversible.

Thus, in this example, the influence of a node is proportional to its degree. If we take G to be the star graph, consisting of a central hub and $n-1$ leaves, each of which is connected to the hub and to no other vertex, then the hub has degree $n-1$ and each leaf has degree 1. It follows that, in the star graph, the hub alone accounts for half the total influence, with each of the leaves having influence $1/2(n-1)$.

Speed of convergence Suppose the weight matrix W is irreducible and aperiodic. How quickly do the opinion vectors $\mathbf{x}(k)$ converge to the consensus vector, $\mathbf{x}(\infty)$? In order to answer this, we need to recall the Perron-Frobenius theorem. We call a matrix non-negative (resp. strictly positive) if all its elements are non-negative (resp. positive).

Theorem 5 (Perron-Frobenius theorem)**.** *Suppose that $A \in \mathbb{R}^{n \times n}$ is a non-negative matrix and that A^k is strictly positive for some $k > 0$. Then, the following hold:*

(i) A has a positive eigenvalue λ such that $|\lambda_i| < \lambda$ for all other eigenvalues λ_i, which could be real or complex.

(ii) The eigenvector corresponding to λ is strictly positive, and it is the only non-negative eigenvector.

Remarks. The positive eigenvalue λ which has the largest absolute value among all eigenvalues of A is referred to as the Perron root or Perron eigenvalue of A. Part (ii) refers to right eigenvectors, but also applies to left eigenvectors, as can be seen by applying the theorem to A^T.

We will not prove this theorem. A proof can be found in any book on non-negative matrices, e.g., [1] or [15].

We now apply the theorem to our stochastic matrix W. The assumptions of irreducibility and aperiodicity imply that there is a $k \geq 1$ such that W^k is strictly positive. Hence, the theorem is applicable.

Since the eigenvector corresponding to the eigenvalue 1 is the all-1 vector, which is non-negative, it follows that 1 must be the Perron eigenvalue of W.

Consequently, all other eigenvalues are strictly smaller than 1 in absolute value. We order them as $1 = \lambda_1 > |\lambda_2| \geq |\lambda_3| \geq \ldots$. Suppose first that the eigenvalues are all distinct. This is not a restrictive assumption as the property holds for generic matrices; the set of matrices W with repeated eigenvalues has zero Lebesgue measure. If the eigenvalues are distinct, then W is diagolisable, i.e., $W = P\Lambda P^{-1}$, where Λ is the diagonal matrix of eigenvalues, and P the matrix whose columns are the corresponding eigenvectors. Consequently,

$$W^k = P \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & \lambda_2^k & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n^k \end{pmatrix} P^{-1} = P \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 \end{pmatrix} P^{-1} + O(|\lambda_2|^k).$$

As we know that W^k converges to $\pi\mathbf{1}$, it must be the case that $W^k = \pi\mathbf{1} + O(|\lambda_2|^k)$, from which it follows that $\|\mathbf{x}(k) - \mathbf{x}(\infty)\| = O(|\lambda_2|^k)$, for any ℓ_p norm, $\|\cdot\|$. In other words, the gap between the opinion vector at time k and the consensus value decays geometrically to zero, with decay rate determined by the second larget eigenvalue, in absolute value, of the weight matrix, W.

The same result holds without assuming that eigenvalues are distinct. The justification is a bit more involved, proceeding through the Jordan decomposition of W, and is left to the interested reader.

An Application: The Wisdom of Crowds

One of the arguments in favour of collective decision-making is the notion of the "wisdom of crowds", which says that by aggregating information across multiple poorly informed agents, we can make better decisions than any single one of them would be likely to make. The idea was formalised at least as far back as 1785, in Condorcet's celebrated Jury Theorem, where he showed that if each member of a jury had a slightly better than 50% chance of reaching the correct verdict, and jurors' decisions were independent, then majority voting could be made to have as high a chance as desired of reaching the correct verdict by choosing a large enough jury.

An analogous question was studied in the context of the de Groot model in [5]. Suppose that a population of agents wishes to learn the true value of some quantity, represented by a real number, x. Suppose that initially each agent i obtains a noisy private signal, $x_i(0)$, of the true value, and that they subsequently update their estimates according to the de Groot model.

Then, for a large enough population of agents, do their individual estimates approach consensus on the true value, x?

Golub and Jackson formalise this question as follows. Consider a sequence of irreducible and aperiodic weight matrices, $W^{(n)}$, indexed by the number of agents, n, which tends to infinity. Let $x_i^{(n)}(t)$ denote the opinion of agent i at time t in the n^{th} system. Suppose that the agents are initialised with values $x_i^{(n)}(0)$ that are independent random variables, with common mean x, and variances $\sigma_i^2(n)$ that are uniformly bounded above and below, i.e., away from infinity and zero. They seek sufficient conditions on the weight matrices $W^{(n)}$ to guarantee that $x_i^{(n)}(\infty)$, the consensus value in the n^{th} system, which does not depend on i, converges in probability to x as n tends to infinity. In other words, as the population size gets large, the consensus value reached by the population converges to the true value, x.

Their main result is that such convergence takes place if and only if the influence of the most influential agent vanishes in the limit, i.e., $\max_{i=1}^n \pi_i^{(n)}$ tends to zero as n tends to infinity. Here, $\pi^{(n)}$ denotes the unique invariant distribution corresponding to the irreducible and aperiodic stochastic matrix, $W^{(n)}$.

This result is pretty straightforward. By our analysis of the de Groot model, we see that the consensus value in the n^{th} system is given by $\pi^{(n)}\mathbf{x}^n(0)$. This is a random variable with mean $x\pi^{(n)}\mathbf{1} = x$, and variance $\sum_{i=1}^n (\pi_i^{(n)})^2\sigma_i^2(n)$. Thus, it converges in probability to x if and only if the variance tends to zero. (The "only if" is obvious, while the "if" follows from Cheybshev's inequality.) Since the $\sigma_i^2(n)$ are bounded uniformly from below, it follows that the sum above cannot tend to zero if $\max_{i=1}^n \pi_i^{(n)}$ does not do so. Thus, it only remains to prove the converse, which we do by bounding the variance. We have that

$$\sum_{i=1}^n (\pi_i^{(n)})^2\sigma_i^2(n) \leq \max_{i=1}^n \pi_i^{(n)} \max_{i=1}^n \sigma_i^2(n) \sum_{i=1}^n \pi_i^{(n)} = \max_{i=1}^n \pi_i^{(n)} \max_{i=1}^n \sigma_i^2(n).$$

As $\max_{i=1}^n \sigma_i^2(n)$ is uniformly bounded, the last term on the RHS above tends to zero if $\max_{i=1}^n \pi_i^{(n)}$ does so. This establishes the claimed result.

While the result precisely characterises the necessary and sufficient conditions for a sequence of weight matrices to asymptotically ensure consensus on the true value, the condition is not easy to check in terms of graph properties. We refer the reader to Golub and Jackson's paper for a detailed

discussion of sufficient conditions for this result on certain specific families of graphs.

3.1.1 The dynamic de Groot model

In many real-world instantiations of opinion dynamics, all agents do not update their opinions synchronously. For example, when an ant colony needs to forage, scouts go out in different directions. Upon identifying a food source, they return to the nest and become recruiters, ferrying individual ants to the source they have located. These ants in turn also become recruiters, repeating the process until some critical mass of ants have been recruited. Thus, the opinion dynamics involve a large number of pairwise interactions, taking place asynchronously and in parallel. A suitable description of this process is given by the recursion

$$\mathbf{x}(k+1) = W(k+1)\mathbf{x}(k), \tag{12}$$

where $W(k), k \in \mathbb{N}$ is a sequence of deterministic or random matrices representing the interactions in successive time steps. For example, if $W(k)$ involves a pairwise interaction between agents i and j, then it would take the general form

$$w_{uv}(k) = \begin{cases} 1, & u = v \notin \{i, j\}, \\ 1 - \alpha_{ij}, & u = v = i, \\ 1 - \alpha_{ji}, & u = v = j, \\ \alpha_{ij}, & u = i, v = j, \\ \alpha_{ji}, & u = j, v = i, \\ 0, & \text{otherwise.} \end{cases}$$

Here, $(\alpha_{uv} \in [0, 1], u \neq v)$ represent the weight assigned by u to v's opinion in a pairwise interaction. We have taken these to be constant over time, but they could also be allowed to vary over time if so desired. The pair of nodes chosen to interact in a given time step could be chosen at random, or according to a specified schedule, or even adaptively based on the outcomes of past interactions. Likewise, the weights could also be adapted.

Another example is one in which agents sequentially broadcast their opinions, according to a fixed or random schedule, and that a directed communication graph $G = (V, E)$ captures who can hear whose broadcast. (This

graph could also be time-varying.) When agent i broadcasts its opinion, each node j which hears the broadcast updates its opinion by computing a convex combination of its current value and the broadcast value. Thus, if node i broadcasts in time step k, then the weight matrix $W(k)$ takes the form

$$w_{uv}(k) = \begin{cases} 1, & u = v \notin \{j : (i,j) \in E\}, \\ \alpha_{ji}, & v = i, u = j : (i,j) \in E, \\ 1 - \alpha_{ji}, & u = v = j : (i,j) \in E, \\ 0, & \text{otherwise.} \end{cases}$$

We would like to know, in scenarios like the two examples above, whether consensus is still attained. Observe that, at any single time step k, the matrix $W(k)$ is not irreducible in these examples. Clearly, some assumptions are needed regarding these matrices in different time steps. We now set out the general framework and assumptions, and the main results, drawn from [17]. We refer the reader to [11] for a comprehensive treatment of this topic.

We focus on the case when $\{W(k), k \in \mathbb{N}\}$ are mutually independent. Note that this includes the case of a determinstic sequence of matrices. We need a few definitions before stating our results.

Definition Let $\{W(k), k \in \mathbb{N}\}$ constitute a sequence of independent $n \times n$ random matrices, also referred to as an independent random chain, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and let $\mathbb{E}$ denote expectation with respect to $\mathbb{P}$. We write $[n]$ to denote the set $\{1, \ldots, n\}$, and S^c to denote the complement of a subset S of $[n]$. We define $W_{SS^c}(k) = \sum_{i \in S, j \in S^c} W_{ij}(k)$.

1. We say that $\{W(k)\}$ is *strongly aperiodic* if there exists $\gamma \in (0, 1]$ such that

$$\mathbb{E}[W_{ii}(k)W_{ij}(k)] \geq \gamma \mathbb{E}[W_{ij}(k)] \text{ for all } i \neq j \in [n], \text{ and all } k \in \mathbb{N}.$$

2. We say that $\{W(k)\}$ is *cut-balanced* if there exists $\alpha > 0$ such that

$$\mathbb{E}[W_{SS^c}(k)] \geq \alpha \mathbb{E}[W_{S^cS}(k)] \text{ for all } S \subset [n] \text{ and all } k \in \mathbb{N}.$$

3. We define the *infinite flow graph* of $\{W(k)\}$ as the random graph $G^\infty = ([n], E^\infty)$ with vertex set $[n]$, and random edge set

$$E^\infty(\omega) = \left\{ (i,j) : \sum_{k=1}^{\infty} (W_{ij}(k,\omega) + W_{ji}(k,\omega)) = \infty \right\}.$$

Remarks

1. If $W_{ii}(k) \geq \gamma$ almost surely (a.s.) for all $i \in [n]$ and all $k \in \mathbb{N}$, then the chain is clearly strongly aperiodic. Conversely, if the chain is strongly aperiodic, then we obtain that

$$\begin{aligned} \mathbb{E}[W_{ii}(k)] &\geq \mathbb{E}[W_{ii}(k)(1 - W_{ii}(k))] = \sum_{j \neq i} \mathbb{E}[W_{ii}(k) W_{ij}(k)] \\ &\geq \gamma \sum_{j \neq i} \mathbb{E}[W_{ij}(k)] = \gamma(1 - \mathbb{E}[W_{ii}(k)]), \end{aligned}$$

   where the first inequality holds because $W_{ii}(k) \geq 0$ and the second inequality follows from the definition of strong aperiodicity, while the two equalities hold because W is a stochastic matrix. It follows from the above equation that $\mathbb{E}[W_{ii}(k)] \geq \tilde{\gamma} = \gamma/(1+\gamma)$ for all $i \in [n]$ and all $k \in \mathbb{N}$.

2. By Kolmogorov's zero-one law, the infinite flow graph of an independent random chain is almost surely equal to a deterministic graph. It has been shown [16] that it is, in fact, equal to the infinite flow graph of the expected chain, $\{\mathbb{E}[W(k)]\}$.

We are now ready to state the main result of this subsection.

Theorem 6. *Let $\{W(k)\}$ be an independent sequence of random stochastic matrices which is cut-balanced and strongly aperiodic. Then, for arbitary $t_0 \geq 0$, the products $W(k : t_0) = W(k) \cdot W(t_0 + 1)$ converge a.s. to a random stochastic matrix $W(\infty : t_0)$ as k tends to infinity. Moreover, for any i and j in the same connected component of the infinite flow graph, the i^{th} and j^{th} rows of the matrix $W(\infty : t_0)$ are almost surely equal.*

In particular, if the infinite flow graph is connected, then $W(\infty : t_0)$ is almost surely a rank one matrix with identical rows, and consensus is achieved starting from any initial condition at time t_0.

3.2 Averaging via gossip

A very different approach to the averaging problem was proposed by Mosk-Aoyama and Shah [9]. More precisely, they present an algorithm for summing values, x_i, $i = 1, \ldots, n$, initially known privately to the n agents. If we

can compute their sum, then we can obtain the average by dividing by n. If n is unknown, then we can separately compute the sum of $y_i \equiv 1$ to estimate the number of nodes. We now present their algorithm for computing the sum.

The algorithm exploits the key insight that, if X_i, $i = 1, \ldots, n$ are independent Exponential random variables with parameters x_i, $i = 1, \ldots, n$, then $X = \min_{i=1}^{n} X_i$ is exponentially distributed with parameter $x = \sum_{i=1}^{n} x_i$, which is the quantity that we wish to calculate. Thus, if the agents independently generate random variables $X_i \sim Exp(x_i)$, and collaboratively compute $X = \min_{i=1}^{n} X_i$, then X provides an estimate of x, since $\mathbb{E}[X] = 1/x$. Admittedly, this is a rather noisy estimate. But it is easy to overcome this drawback. Suppose each agent i generates a vector $\mathbf{X}^i = (X_1^i, \ldots, X_k^i)$, whose components are independent and identically distributed (iid) with an $Exp(x_i)$ distribution. Suppose, too, that the random vectors $\mathbf{X}^i$ are mutually independent. Then, the componentwise minimum of these vectors, which we denote $\mathbf{X}$, has independent components, each with an $Exp(x)$ distribution. Hence, by the weak law of large numbers, $(X_1 + \ldots + X_k)/k$ converges in probability to $1/x$, the mean of the $Exp(x)$ dsitribution. Consequently, $k/(X_1 + \ldots + X_k)$ provides a consistent estimator of x, the quantity of interest. One can also compute confidence intervals for any given k.

We now present an algorithm for computing the minimum that may be implemented either in discrete or continuous time; these are also referred to as synchronous and asynchronous versions. We will analyse the continuous time (asynchronous) version. The algorithm is parametrised by a non-negative matrix W, which we take to be stochastic; this is needed for the synchronous version but can be relaxed in the asynchronous setting. We first present the synchronous version.

Time is discrete. In each time step t, each agent i maintains an estimate $\mathbf{Y}^i(t)$ of the vector $\mathbf{X}$ of component-wise minima, on which the agents are seeking to reach consensus. Agent i initiates its estimate to $\mathbf{Y}^i(0) = \mathbf{X}^i$, the random vector that it generated. Subsequently, in each time step t, each agent i randomly picks an agent with which to communicate, choosing agent j with probability W_{ij}, independent of the past. We assume a Push model, that when i chooses to communicate with j, it sends its current estimate, $\mathbf{Y}^i(t)$ to j; the Pull version is analogous. Suppose that, in time step t, agent j receives estimates from agents $i_1, \ldots, i_k$. Then it updates its estimate by setting

$$\mathbf{Y}^j(t+1) = \wedge_{i_1,\ldots,i_k} \left(\mathbf{Y}^j(t), \mathbf{Y}^{i_1}(t), \ldots, \mathbf{Y}^{i_k}(t) \right),$$

where $\wedge$ denotes the componentwise minimum operation. In words, agent j computes the componentwise minimum of its current estimate as well as all the estimates it received in time step t. If it received no messages, its estimate remains unchanged.

One could consider a variant in which, if an agent receives multiple messages, it picks one of them at random and computes the componentwise minimum with its current estimate. While our analysis can be easily modified to study this variant, the results can be very different. (*Hint.* Take $k = 1$ and analyse both variants on the star graph, in which a central hub node is connected to each of $n-1$ leaves and there are no edges between leaves. Take $W_{ij} = 1/n-1$ if i is the hub and j a leaf and $W_{ij} = 1$ if i is a leaf and j the hub.)

The asynchronous version of the algorithm operates in continuous time, with agent i sending its current estimate to agent j at the increment times of a Poisson process of rate or intensity W_{ij}; the Poisson processes corresponding to distinct ordered pairs of agents are mutually independent. When j receives a message from i, it updates its estimate by computing the coordinatewise minimum of its current estimate and the message. We now present an analysis of this version, obtaining bounds on the time until all agents reach consensus on the minimum vector.

Suppose first that $k = 1$. Let $*$ denote the agent whose inital random variable $X^i(0)$ is smallest, i.e., $* \in \arg\min i : X^i(0) = \min_j X^j(0)$. Since the $X^i(0)$ are continuous random variables, the minimiser is almost surely unique. Initially, only the estimator of $*$ coincides with the minimum value $X^*(0)$. When $*$ performs a Push operation at some time t_1, choosing agent j_1 as the target, then j_1 will update its estimate to $X^*(0)$ as its current estimate cannot be smaller. Thus, at this time, two nodes have the minimum value. When either of them subsequently pushes their value, a third node will learn the minimum (if they did not choose the other as the target) and so on. Notice that after time t_1, the values of $*$ and j_1 cannot change, as they cannot receive a message with a smaller value that $X^*(0)$. Thus, the minimum value spreads through the population like a rumour, and we want to know how long it takes the rumour to reach everyone in the population. We now study this rumour spreading model.

Consider a rumour that is initially known to a single agent in the set $[n]$. After node i learns the rumour, it pushes the rumour at the increment times of a Poisson process (PP) of rate $1 = \sum_j W_{ij}$, choosing j as its target with probability W_{ij}, independent of the past. The PPs corresponding to

distinct agents are assumed to be mutually independent. Equivalently (since Bernoulli splittings of PPs yield independent PPs), i pushes the rumour to j at the jump times of a PP of rate W_{ij}, with all these PPs being mutually independent. We want to know the time by which all agents learn the rumour. In order to analyse the rumour spreading time, it will be useful to define the *conductance* of W, which we define as

$$\Phi(W) = \min_{S \subset [n]: S \neq \emptyset, [n]} \frac{nW_{SS^c}}{|S||S^c|}, \text{ where } W(S, S^c) = \sum_{i \in S, j \in S^c} W_{ij}. \tag{13}$$

Here, S^c denotes the complement of S. The minimum in the definition of $\Phi(W)$ is taken over all non-trivial subsets of $[n]$.

Let s denote the source of the rumour and $S(t)$ the set of nodes that knows the rumour at time t, so that $S(0) = \{s\}$. Let T_k denote the first time that exactly (or at least) k nodes know the rumour, so that $T_1 = 0$. Now, conditional on S_{T_k}, the random time until one more node learns the rumour is given by the time until some node in S_{T_k} pushes the rumour to some node in $S^c_{T_k}$. By the superposition property for independent PPs, we say that, conditional on S_{T_k}, the random variable $T_{k+1} - T_k$ is exponentially distributed, with parameter $W_{S_{T_k}, S^c_{T_k}}$, i.e.,

$$(T_{k+1} - T_k)|S_{T_k} \sim Exp(W_{S_{T_k}, S^c_{T_k}}),$$

where we write $\sim$ to denote equality in distribution. But it follows from (13) that

$$W_{S_{T_k}, S^c_{T_k}} \leq \frac{|S_{T_k}||S^c_{T_k}|}{n} \Phi(W) = \frac{k(n-k)}{n} \Phi(W).$$

Let τ_k, $k = 1, \ldots, n-1$ be independent exponential random variables, with parameters $k(n-k)\Phi(W)/n$. It follows from the above that

$$(T_{k+1} - T_k)|S_{T_k} \preceq \tau_k, \quad k = 1, \ldots, n-1,$$

for arbitrary subsets S_{T_k}. Here, for random variables X, Y, we write $X \preceq Y$ to denote that X is stochastically dominated by Y. Since the above inequality holds for arbitrary S_{T_k}, we obtain by summing it over $k = 1, \ldots, n-1$ that

$$T_n \preceq \tau_1 + \ldots + \tau_{n-1}. \tag{14}$$

Here, T_n is the time until all n nodes know the rumour, i.e., the rumour spreading time. The above equation gives a stochastic bound on this random variable.

A straightforward calculation yields that

$$\mathbb{E}[T_n] \leq \sum_{j=1}^{n-1} \mathbb{E}[\tau_j] = \sum_{j=1}^{n-1} \frac{1}{\Phi(W)} \left(\frac{1}{k} + \frac{1}{n-k} \right) \sim \frac{2 \log n}{\Phi(W)},$$

where, for sequences a_n and b_n, we write $a_n \sim b_n$ to denote that a_n/b_n tends to 1 as n tends to infinity. In fact, by computing the variance and applying Chebyshev's inequality, it can be shown that $(\tau_1 + \ldots + \tau_{n-1})/\log n$ converges in probability to $2/\Phi(W)$. One can also derive tail bounds on this sum of independent exponential random variables. It can be shown that

$$\mathbb{P}\left(T_n > \frac{c \log n}{\Phi(W)} \right) \leq \mathbb{P}\left(\sum_{j=1}^{n-1} \tau_j > \frac{c \log n}{\Phi(W)} \right) \leq n^{-(c-2)}. \tag{15}$$

This completes our analysis of the case $k = 1$, in which the consensus time is the same as the rumour spreading time.

What can we say about larger k? In this case, we have multiple rumours, one started at each agent which initially minimises one of the components of the vectors $\mathbf{X}^i$. We want to bound the time until all rumours reach all agents. As there are k rumours, a naive bound is obtained by pretending the rumours spread sequentially, one after another. This yield a bound of order $k \log n/\Phi(W)$ for the consensus time. However, this is too loose. We can in fact do much better. Let $T(i)$ denote the rumour spreading time starting from the i^{th} (in some numbering) of the k agents (or fewer) which achieve the minimum of some component of the initial vectors $\mathbf{X}^i$. The maximum of $T(i)$, $i = 1, \ldots, k$ is the time by which all agents learn all rumours, and hence their estimated vectors have reached consensus. In other words, the consensus time is $\max_{i=1}^{j} T(i)$. We will bound this from above by $\max_{i=1}^{n} T(i)$, the time by which rumours started at each of the agents have reached all other agents. We now have the following by the union bound:

$$\mathbb{P}\left(\max_{i=1}^{n} T(i) > \frac{c \log n}{\Phi(W)} \right) \leq \sum_{i=1}^{n} \mathbb{P}\left(T(i) > \frac{c \log n}{\Phi(W)} \right) \leq n^{-(c-3)}.$$

It follows that for any $c > 3$, the consensus time is bounded above by $c \log n/\Phi(W)$, with high probability.

3.3 The 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 directed graph, $G = (V, E)$. 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 $|E|$ independent Poisson processes (PPs), one associated with each edge; we denote by q_{uv} the rate of the PP associated with the directed edge (u, v). At every time that there is an increment of the PP associated with edge (u, v), node u copies the preference of node v 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. A special case of this model that we will study in more detail arises when G is the complete directed graph on n nodes, and the rates on all edges are identical; without loss of generality (wlog), we take the rate to be $1/n$, so that the total rate at which a node initiates contacts is 1 (or rather, $(n-1)/n$, but we are interested in large n aysmptotics).

Note that we have described an asynchronous model in continuous time. It is identical, up to a random time change, to the asynchronous discrete time model in which, at each time step, an ordered pair of agents (u, v) is selected with probability $q_{uv}/\sum_{i,j} q_{ij}$, independent of the past, and then u copies the preference of v. It is different from a synchronous discrete time model in which, in each time step, each agents choose another agent (according to a specified probability distribution, and indepndent of the past and of the choices of other agents) and copies their preference. The synchronous model

is harder to analyse due to simultaneous changes of preference at all agents rather than jut one. We will not study the synchronous model in any depth but may mention related work on this model.

We denote by $X_v(t)$ the state of node v at time t, and by $\mathbf{X}(t)$ the vector, $(X_v(t), v \in V)$. It should be clear from the verbal 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. 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. 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:

$$\begin{aligned} q(\mathbf{x}, \mathbf{x} + e_v) &= (1 - x_v) \sum_{w \in V} q_{vw} x_w, \\ q(\mathbf{x}, \mathbf{x} - e_v) &= x_v \sum_{w \in V} q_{vw} (1 - x_w). \end{aligned} \tag{16}$$

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 the total rate at which v contacts some node w in state 1, which is given by the sum on the RHS. Likewise, the second equation gives the rate for node v to move from state 1 to state 0. We shall assume the following throughout the remainder of this section.

Assumption The directed graph $G = (V, E)$ is *strongly connected*, i.e., there is a *directed path* from v to w for every pair of nodes v, w.

Observe that the all-0 and all-1 vectors, which we'll denote $\mathbf{0}$ and $\mathbf{1}$, are absorbing states of the Markov process $\mathbf{X}(t)$; no vertex can change state if the system has reached either of these states. Moreover, the assumption above says that every node can influence every other node as there is a directed path between them, 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 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.

Note that the assumption of strong connectivity is essential for absorption in one of the two states, $\mathbf{0}$ or $\mathbf{1}$, to be guaranteed. To see this, we consider a counterexample. First, consider a graph consisting of 3 nodes: $V = \{u, v, w\}$

and $E = \{(v,u),(v,w)\}$. 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 v can't influence them, and they can't influence each 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, one for each connected component of the graph.

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. We shall answer the first question for general graphs and the second question for the complete graph with symmetric rates.

In order to study consensus probabilities, we will make use of the Optional Stopping Theorem for martingales. We now introduce these for the benefit of the reader.

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

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 roulette 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 only 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 $X_t, t \in \mathbb{N}$ 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 7. *(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. Suppose that T is finite almost surely, i.e., $\mathbb{P}(T < \infty) = 1$. 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 Optional Stopping Theorem can be extended to continuous time but this involves a number of technicalities. In order to avoid them, we will instead work with the asynchronous discrete-time Markov chain described above, obtained by watching the continuous-time chain at jump times of its constituent PPs. We denote this process by $\mathbf{Y}(n)$, $n \in \mathbb{N}$, i.e., we set $\mathbf{Y}(0) = \mathbf{X}(0)$ and $\mathbf{Y}(n) = \mathbf{X}(T_n)$ for $n \geq 1$, where $T_n \in \mathbb{R}_+$ denotes the n^{th} jump time of any of the PPs associated with the various edges. If we define $q^* = \sum_{(v,w)\in E} q_{vw}$ to be the sum of the jump rates of the PPs corresponding to all the edges, then T_n, $n \in \mathbb{N}$ is a PP with intensity q^*. Furthermore, as noted above, $\mathbf{Y}(n)$, $n \in \mathbb{N}$, is a discrete-time Markov chain on the same state space, $\{0,1\}^V$, as $\mathbf{X}(t)$, $t \in \mathbb{R}_+$, with one-step transition probabilities

matrix given by:

$$\begin{aligned} p(\mathbf{x}, \mathbf{x} + e_v) &= \frac{1 - x_v}{q^*} \sum_{w \in V} q_{vw} x_w, \\ p(\mathbf{x}, \mathbf{x} - e_v) &= \frac{x_v}{q^*} \sum_{w \in V} q_{vw}(1 - x_w), \\ p(\mathbf{x}, \mathbf{x}) &= 1 - \sum_{v \in V} (p(\mathbf{x}, \mathbf{x} + e_v) + p(\mathbf{x}, \mathbf{x} - e_v)). \end{aligned} \tag{17}$$

With these preliminaries, we are ready to analyse the probability of reaching consensus on a given value. Notice that the Markov process $\mathbf{X}(t)$, $t \in \mathbb{R}_+$, hits the all-1 state, $\mathbf{1}$, before the all-0 state, $\mathbf{0}$, if and only if the Markov process $\mathbf{Y}(n)$, $n \in \mathbb{N}$, does so as well. Thus, the probability of reaching consensus on a given value, say 1, is the same in both the continuous and asynchronous discrete time models. It will be easier to study consensus probabilities in the discrete-time model.

3.3.1 Consensus probabilities

Let Q be a rate matrix (infinitesimal generator) with off-diagonal elements q_{vw} taken to be the contact rates (at which v copies w) specified above; the diagonal elements are then determined by the requirement that the row sums should all be zero. Since Q is the rate matrix of a finite state Markov process 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 Markov process is irreducible, and the invariant distribution is unique and strictly positive (i.e., π_v is not zero for any $v \in V$). Define $M(t) = \pi \mathbf{Y}(t)$ for $t \in \mathbb{N}$, the product of the row vector π and the column vector $\mathbf{Y}(t)$, for $Y(t)$, $t \in \mathbb{N}$, defined as above. We now claim the following.

Theorem 8. *The stochastic process $M(t) = \pi \mathbf{Y}(t) = \sum_{v \in V} \pi_v Y_v(t)$ is a martingale.*

Proof. As $\mathbf{Y}(t)$, $t \in \mathbb{N}$, is a Markov chain, it suffices to show that $\mathbb{E}[M(t+1) - M(t) | \mathbf{Y}(t) = \mathbf{x}] = 0$ for all states $\mathbf{x} \in \{0,1\}^V$. We shall show this now. We have that

$$\mathbb{E}[M(t+1) - M(t) | \mathbf{Y}(t) = \mathbf{x}] = \sum_{v \in V : x_v = 0} \pi_v p(\mathbf{x}, \mathbf{x} + e_v) - \sum_{v \in V : x_v = 1} \pi_v p(\mathbf{x}, \mathbf{x} - e_v), \tag{18}$$

since at most a single node changes state in any time step and, by the definition of $M(t)$, a change in state of node v from 0 to 1 increases $M(t)$ by π_v, while a change from 1 to 0 decreases it by π_v. Now, substituting (17) in (18), we get

$$\begin{aligned} & q^*(\mathbf{x})\mathbb{E}[M(t+1) - M(t)|\mathbf{Y}(t) = \mathbf{x}] \\ &= \sum_{\substack{v,w\in V:\\ x_v=0,x_w=1}} \pi_v q_{vw} - \sum_{\substack{v,w\in V:\\ x_v=1,x_w=0}} \pi_v q_{vw} \quad = \sum_{\substack{v,w\in V:\\ x_v=0,x_w=1}} (\pi_v q_{vw} - \pi_w q_{wv}). \end{aligned}$$

We now add and subtract

$$\sum_{v,w\in V:x_v=0,x_w=0} \pi_v q_{vw} \equiv \sum_{v,w\in V:x_v=0,x_w=0} \pi_w q_{wv}$$

on the RHS of above expression, to obtain that

$$\begin{aligned} & q^*(\mathbf{x})\mathbb{E}[M(t+1) - M(t)|\mathbf{Y}(t) = \mathbf{x}] = \\ & \sum_{v\in V:x_v=0} \sum_{w\in V} \pi_v q_{vw} - \sum_{v\in V:x_v=0} \sum_{w\in V} \pi_w q_{wv}. \end{aligned} \tag{19}$$

Now, for each $v \in V$, $\sum_{w\in V} \pi_v q_{vw} = \pi_v \sum_{w\in V} q_{vw} = 0$, since the row sums of the rate matrix Q are zero. Moreover, $\sum_{w\in V} \pi_w q_{wv} = 0$, because π is an invariant distribution and hence satisfies the global balance equations $\pi Q = \mathbf{0}$. Therefore, it follows from (19) that

$$\mathbb{E}[M(t+1) - M(t)|(\mathbf{Y}(u), u < t), \mathbf{Y}(t) = \mathbf{x}] = \mathbb{E}[M(t+1) - M(t)|\mathbf{Y}(t) = \mathbf{x}] = 0,$$

which implies that $(M(t), t \in \mathbb{N})$ 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 $\mathbf{Y}(0) = \mathbf{X}(0)$. (We initialise the discrete-time chain in the same state as the continuous-time chain.) Let $M(0) = \pi\mathbf{X}(0)$ denote the value of the martingale corresponding to the given initial condition. Let T denote the absorption time (of the discrete-time chain) in one of the two absorbing states, and note that it is a stopping time. Now, T is finite almost surely because it is the absorption time of a finite state Markov chain. Moreover, $M(t) = \pi\mathbf{Y}(t)$ is bounded between 0 and 1 for all t as $Y_v(t)$ is $\{0, 1\}$-valued for each $v \in V$ and π is a probability vector over V. Hence, by the Optional Stopping Theorem, $\mathbb{E}[M(T)] = M(0) = \pi\mathbf{X}(0)$. But

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

Re-arranging the above, we get

$$\mathbb{P}(\mathbf{Y}(\tau) = \mathbf{1}) = \mathbb{E}[M(T)] = M(0) = \pi\mathbf{X}(0). \tag{20}$$

This is the probability of reaching consensus on $\mathbf{1}$ for the discrete-time chain $\mathbf{Y}(t)$, $t \in \mathbb{N}$, and hence also for the continuous-time chain $\mathbf{X}(t)$, $t \in \mathbb{R}_+$. (We remind the reader that the sequence of distinct states visited is the same for both chains, but the discrete-time chain has a random number of 'lazy' transitions in which the state remains unchanged between successive jumps of the continuous-time chain. These lazy transitions correspond to jump times of the underlying PPs driving the voter model dynamics at which an agent contacts another agent when the same opinion and hence does not change state. Hence, they both reach consensus on the same value.)

Notice the similarity of the final result with that of the de Groot model. There, the value on which consensus was reached was $\pi\mathbf{X}(0)$. Here, consensus is reached on either $\mathbf{0}$ or $\mathbf{1}$, but the expected value is still the same, $\pi\mathbf{X}(0)$. In particular, the invariant distribution π still determines the influence of different nodes or agents in determining the final outcome.

Example Let $G = (V, E)$ be a connected, undirected graph, and suppose $q_{vw} = 1/\deg(v)$ for all $v \in V$. In words, nodes v communicates at the increment times of independent unit rate PPs, and when a node v communicates, it chooses one of its neighbours uniformly at random and copies its opinion. It can be readily verified thaat the Markov process with state space V and generator Q defined as above as reversible with respect to the probability distribution $\pi_v = deg(v)/2|E|$, where $2|E|$ is twice the number of edges and is equal to the sum of vertex degrees. Hence π is the unique invariant distribution of this irreducible (by connectedness of G) Markov process.

By (20), the probability of reaching consensus on 1 starting from an initial state $\mathbf{X}(0)$ is given by $\pi\mathbf{X}(0) = \sum_{v\in V} \deg(v)X_v(0)/2|E|$. Thus, as in the de Groot model, the influence of a node in determining the consensus value is proportional to its degree. In the special case that G is the complete directed graph and $q_{vw} = 1/n$ for all $v \neq w$, the uniform distribution is invariant; the probability of reaching consensus on 1 depends only on the number of initial votes in favour of 1 (and not which nodes hold these votes). In fact, the probability of reaching consensus on 1 is exactly equal to the initial proportion of agents favouring this option.

3.3.2 Consensus 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,

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

Recalling that $T_n = 0$, we obtain that

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

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. An alternative measure of complexity, which does not depend on the somewhat arbitrary choice of rates of the underlying Poisson processes is to count the number of pairwise communications needed to reach consensus. This if of order n^2 for the voter model on the complete graph on n nodes.

3.4 Variants of the voter model

We first consider generalisations of the voter model with more than two possible opinions. We then allow for stubborn agents, which never change opinion. In this case, consensus is never reached, but we can ask for the long-run fraction of time that voters hold each of the possible opinions. These generalisations are straightforward.

One of the drawbacks of the voter model is that it is slow to reach consensus. If one wishes to use it as a mechanism to reach consensus on the majority initial opinion, which is desirable in many applications, then its probability of doing so correctly is small. For example, on the complete graph with symmetric contact rates, we saw that the probability of reaching consensus on the minority value is equal to the proportion of agents initially preferring it. Thus, the voter model is not a robust or fast mechanism for determining

the majority vote. This has led to interest in many variants of the voter model designed to be fast and to achieve small error probabilities, typically decaying in the population size. Interest in other variants have come from various applications in social networks. We briefly study some of these variants including the biased voter model and the majority voter model. Finally, we discuss voter models on evolving networks.

Voter model with multiple opinions Consider the voter model as above, where agents are associated with the vertices of a directed graph $G = (V, E)$, and each agent has an opinion from a finite set $[K] = \{1, \ldots, K\}$. We may also refer to the possible opinions as colours. As before, agent $u \in V$ contacts $v \in V$ at the increment times of a PP of intensity q_{uv} and adopts the opinion of v. The opinion dynamics are described by a Markov process on the state space $[K]^V$. We assume without loss of generality that $K \leq |V|$ as we may restrict attention to the set of colours used in the initial configuration; no new colours can emerge during the process. If G is strongly connected, then the absorbing state of this Markov process are precisely the monochromatic configurations in which all nodes have the same colour. Thus, consensus is eventually reached on a single opinion.

The probability of reaching consensus on a given opinion k can be obtained by considering the two-opinion model studied above, with all opinions other than k treated as a single 'not-k' opinion. The analysis of consensus times using coalescing random walks is still applicable. The time for all random walks to coalesce continues to provide an upper bound on the consensus time. Thus, no new analysis is needed, either for consensus probabilities or consensus times, when generalising to voter models with multiple opinions.

Biased voter model Suppose one of the options, say option 1, is intrinsically better than the others, and this is either known in advance or the quality of an option is measured whenever an agent changes state. An example scenario in which this is a reasonable assumption is that of site choice in social insects, for foraging, nesting or some other purpose. Here, agents may have some information regarding the quality of the site while trying to reach consensus. We incorporate this in the voter model by allowing different contact rates depending on the state, as detailed below.

We will consider the voter model on the complete graph, with symmetric contact rates. However, instead of the contact rate being $1/n$ along each directed edge (v, w), we take it to be q_1/n if node v is in state 1, and q_0/n if it is in state 0. Note that it does not depend on the state of w (which v does

not know until contacting it). In other words, a node in state 1 waits an $Exp(q_1)$ time before contacting another node, chosen uniformly at random (including itself), and copying its opinion; a node in state 0 waits an $Exp(q_0)$ time before doing the same. We take $q_1 < q_0$ to represent 1 being the better option, so that opinion 1 is held for longer on average before reconsidering it.

The symmetry inherent in the problem means that we do not need to keep track of the states of all individual nodes, but only of the numbers in each state. Let $X(t)$ denote the number of nodes with opinion 1 at time t. Then it is easy to see that $(X(t), t \geq 0)$ evolves as a Markov process on the state space $\{0, 1, \ldots, n\}$ with transition rates

$$q_{k,k+1} = q_0 \frac{k(n-k)}{n}, \quad q_{k,k-1} = q_1 \frac{k(n-k)}{n},$$

so that the embedded jump chain is a discrete-time Markov chain with transition probabilities

$$p_{k,k+1} = \frac{q_0}{q_0 + q_1}, \quad p_{k,k-1} = \frac{q_1}{q_0 + q_1}. \tag{21}$$

It is easy to verify that $M(t) = (q_1/q_0)^{X(t)}$ is a martingale. Indeed, considering the jump chain, this corresponds to verifying that

$$\frac{q_0}{q_0 + q_1}\Big(\frac{q_1}{q_0}\Big)^{k+1} + \frac{q_1}{q_0 + q_1}\Big(\frac{q_1}{q_0}\Big)^{k-1} = \Big(\frac{q_1}{q_0}\Big)^{k} \text{ for all } k \in \{1, \ldots, n-1\}.$$

Defining $T \in \mathbb{N}$ to be the first time step that the jump chain hits either 0 or n, it is easy to see that T is a stopping time which is finite almost surely. The martingale $M(t)$ is bounded between 0 and 1. Hence, the Optional Stopping Theorem can be applied, and we obtain using it that

$$\Big(\frac{q_1}{q_0}\Big)^{n} \mathbb{P}(X_T = n) + \mathbb{P}(X_T = 0) = \Big(\frac{q_1}{q_0}\Big)^{X(0)}.$$

Hence, the probability of reaching consensus on the better option, 1, is given by

$$\mathbb{P}(X_T = n) = \frac{1 - (q_1/q_0)^{X(0)}}{1 - (q_1/q_0)^{n}} \approx 1 - \Big(\frac{q_1}{q_0}\Big)^{X(0)} \text{ for large } n.$$

Observe from the above expression that, as long as $X(0) \geq 1$, $\mathbb{P}(X(T) = n)$ is bounded away from zero, uniformly in n. In other words, as long as a single agent initially prefers the better option, there is a strictly positive

probability of reaching consensus on this better option, no matter the size of the population holding the opposite opinion initially! Moreover, the probability of reaching consensus on 0, the worse option, decays geometrically in the number of agents initially preferring the better option. In particular, if we start with some non-zero fraction of the population preferring the better option, then the error probability decays exponentially in the population size.

The above analysis of the biased voter model pertains to the complete graph. Unfortunately, unlike the voter model, the analysis does not extend readily to general networks. It does extend to regular graphs, where the same martingale works. But in general, the function $M(t) = (q_1/q_0)^{X(t)}$ is not a martingale. It is possible to obtain some bounds when the graph is close to regular. We refer the reader to [4] for details.

Majority voter model See [2], [10]

Three-state model See [12]

4 The Contact Process (SIS epidemic) on Networks

The spread of infectious diseases bears some similarities to the spread of rumours. In particular, each individual who becomes infected spreads it anew and, at least in the sense of statistical averages, the behaviours of all agents are very similar, and can be mathematically abstracted as being identical. The study of mathematical models of epidemics has a long history, going back to the work of Daniel Bernoulli (1766), who tried to quantify the benefits, in terms of increased life expectancy, of inoculation against smallpox. Modern mathematical epidemiology begins with the work of Kermack and McKendrick, who proposed differential equation models similar to the law of mass action in chemistry to describe the fractions of susceptible and infected individuals in a population. More recently, there has been interest in stochastic models, as well as in the effects of network structure.

The rumour spreading model studied earlier can be as a special case of an epidemic model. It refers to what is known as the simple or SI epidemic. Here, every node (individual) is in one of two states, susceptible (S) or infected (I). Infected individuals can infect other individuals through an

interaction, which may be physical contact or close proximity, or may be mediated through carriers (e.g. mosquitoes) or shared water sources or other means. Ignoring the details, we may abstract it by a graph, where an edge between two nodes reflects a potential for such interaction between the corresponding individuals. Thus, a complete graph would represent a situation where any two individuals in the population are equally likely to interact, while other graph structures might emphasise family relationships or sexual contacts.

In the SI model, a node, once infected, remains infectious for ever. It contacts its neighbours in the graph according to some specified random process (usually a Poisson process of a given rate), and infects it if it is susceptible. Eventually, all nodes become infected if the graph is connected, just as earlier all nodes eventually learnt the rumour. The question of the interest is the time for the infection to reach all nodes. We have already studied this model in the rumour-spreading context and won't repeat the analysis here.

The SI model is not particularly realistic as an epidemic model and so now turn to two other models. In the SIR model, there are 3 possible states for each node, susceptible, infected or removed (R). A susceptible node may become infected by one of its neighbours and is subsequently capable of infecting other nodes. But after some deterministic or random time, an infectious node is removed, and no longer plays a role in the spread of infection. You can think of removal either as the individual dying and hence no longer affecting the infection spread, or recovering from infection and subsequently being immune to re-infection. While these two outcomes may be quite different for the individual, they are equivalent for modelling purposes. In the SIR model, questions of interest are typically about how many individuals eventually become infected, and hence removed, and possibly about how long it takes for this to happen.

Another commonly used model is the SIS model, known as the contact process in the mathematics literature. Here, individuals may be susceptible or infected, and an infected individual can infect susceptible neighbours as in the SI and SIR models. The difference is that, infected individuals may recover but, unlike in the SIR model, they don't remain in this state forever. Instead, recovered individuals are immediately susceptible to re-infection. This can be thought of as modelling infections in which there is no long term immunity. The SIS model is also called the contact process. In infinite populations, it is possible for nodes to switch forever between susceptible and infected states. In finite populations, it is possible to reach a state in

which no nodes are infected. In that case, infection can never subsequently re-appear in the population. The question of interest in this model is the random time until this occurs, namely until the infection dies out. The SIR and SIS models are both widely used. We shall focus on the SIS model, but the methods introduced can be adapted to study the SIR model.

Let $G = (V, E)$ be an undirected graph with vertex set V and edge set E. We shall 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 two 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? This will be the focus of this section. We begin with an upper bound on T, the recovery time.

4.1 Upper bound on the recovery time

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:

$$\begin{aligned} q(\mathbf{x}, \mathbf{x} + e_v) &= \alpha(1 - x_v) \sum_{w:(w,v)\in E} x_w \\ q(\mathbf{x}, \mathbf{x} - e_v) &= \beta x_v. \end{aligned} \tag{22}$$

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:

$$\begin{aligned} \tilde{q}(\mathbf{y}, \mathbf{y} + e_v) &= \alpha \sum_{w:(w,v)\in E} y_w \\ \tilde{q}(\mathbf{y}, \mathbf{y} - e_v) &= \beta y_v. \end{aligned} \tag{23}$$

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 (of which we shall see more later).

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. Recall that a probability space is a triple $(\Omega, \mathcal{F}, P)$ consisting of a sample space Ω, a collection $\mathcal{F}$ of subsets of Ω containing the null set and being closed under complements and countable unions (such a collection of sets is called a σ-algebra), and a probability measure P defined on the sets in $\mathcal{F}$, i.e., a measure which is non-negative and countably additive, and satisfies the normalisation condition $P(\Omega) = 1$. Recall, too, that a random variable is a function f from Ω to the real numbers. A stochastic process can also be thought of as a function on Ω, but taking values in the space of possible sample paths, rather than the space of real numbers.

If all that sounds rather abstract and complicated, you'll be relieved to

know that we will take a much more informal and operational approach to coupling. We say 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. We recall the example we saw earlier in the context of stochastic ordering. 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, $X \preceq Y$ in the stochastic order, so we may wish to 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 $\mu < \lambda$. Now, for an arbitrary interval $[a, b]$, we know that the number of points of the X_t process in this interval is a Poisson random variable of mean $\lambda(b-a)$, while the number of points of the Y_t process in the same interval is Poisson with mean $\mu(b-a)$. It can be verified that the latter random variable is stochastically dominated by the former. But 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 but X_v doesn't, then there must have been a strictly positive probability for X_v not to decrease. In other wrods, the probability that X_v decreases must have been strictly smaller than 1. But the probability that X_v decreases is $X_v(0)/Y_v(0)$. Hence, either both started at the same value at time 0, in which case both decreased, or Y_v started with a strictly bigger value at time 0, in which case the inequality $X_v(\tau) \leq Y_v(\tau)$ still holds (despite Y_v decreasing while X_v stays the same).

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, possible 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, as is any Markov process we shall encounter in this course, but we will not prove it.)

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 9. *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 (23) 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 (23) 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

$$\mathbb{E}[N_t] \leq \mathbf{1}^T e^{(\alpha A - \beta I)t}\mathbf{X}(0) \quad \forall\, t \geq 0. \tag{24}$$

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 (24) that

$$\mathbb{E}[N_t] \leq \|\mathbf{1}\|e^{-(\beta - \alpha\rho)t}\|\mathbf{X}(0)\| = e^{-(\beta-\alpha\rho)t}\sqrt{n}\sqrt{N_0} \leq ne^{-(\beta-\alpha\rho)t}. \tag{25}$$

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 (25) 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. □

4.2 Lower bound on the recovery time

References

[1] R. B. Bapat and T. E. S. Raghavan. *Non-negative Matrices and Applications*. Cambridge University Press, 1997.

[2] James Cruise and Ayalvadi Ganesh. Probabilistic consensus via polling and majority rules. *Queueing Systems*, 78(2):99–120, 2014.

[3] A.M. Frieze and G.R. Grimmett. The shortest-path problem for graphs with random arc-lengths. *Discrete Applied Mathematics*, 10(1):57–77, 1985.

[4] Ayalvadi Ganesh, Sabine Hauert, and Emma Valla. Consensus in the weighted voter model with noise-free and noisy observations. *Swarm Intelligence*, pages 1–42, 2025.

[5] Benjamin Golub and Matthew O. Jackson. Naïve learning in social networks and the wisdom of crowds. *American Economic Journal: Microeconomics*, 2(1):112–49, February 2010.

[6] Svante Janson. One, two and three times log n/n for paths in a complete graph with random weights. *Combinatorics, Probability and Computing*, 8(4):347–361, 1999.

[7] J. F. C. Kingman. *Poisson processes*, volume 3 of *Oxford Studies in Probability*. The Clarendon Press Oxford University Press, 1993.

[8] Günter Last and Mathew Penrose. *Lectures on the Poisson Process*. Institute of Mathematical Statistics Textbooks. Cambridge University Press, 2017.

[9] Damon Mosk-Aoyama and Devavrat Shah. Computing separable functions via gossip. In *Proceedings of the Twenty-Fifth Annual ACM Symposium on Principles of Distributed Computing*, PODC '06, page 113–122, New York, NY, USA, 2006. Association for Computing Machinery.

[10] Arpan Mukhopadhyay, Ravi R Mazumdar, and Rahul Roy. Voter and majority dynamics with biased and stubborn agents. *Journal of Statistical Physics*, 181(4):1239–1265, 2020.

[11] Angelia Nedich. Convergence rate of distributed averaging dynamics and optimization in networks. *Foundations and Trends® in Systems and Control*, 2(1):1–100, 2015.

[12] E. Perron, D. Vasudevan, and M. Vojnovic. Using three states for binary consensus on complete graphs. In *IEEE INFOCOM*, pages 2527–2535, 2009.

[13] Boris Pittel. On spreading a rumor. *SIAM Journal on Applied Mathematics*, 47(1):213–223, 1987.

[14] Boris Pittel. Note on the heights of random recursive trees and random m-ary search trees. *Random Structures & Algorithms*, 5(2):337–347, 1994.

[15] E. Seneta. *Non-negative Matrices and Markov Chains*. Springer, New York, 2006.

[16] Behrouz Touri and Angelia Nedic. On approximations and ergodicity classes in random chains. *IEEE Transactions on Automatic Control*, 57(11):2718–2730, 2012.

[17] Behrouz Touri and Angelia Nedić. Product of random stochastic matrices. *IEEE Transactions on Automatic Control*, 59(2):437–448, 2013.

[18] Remco van der Hofstad, Gerard Hooghiemstra, and Piet van Mieghem. Size and weight of shortest path trees with exponential link weights. *Combinatorics, Probability and Computing*, 15(6):903–926, 2006.