A NOTE ON THE ASYMPTOTIC UNIFORMITY OF MARKOV CHAINS WITH RANDOM RATES

JACOB CALVERT, FRANK DEN HOLLANDER, AND DANA RANDALL

ABSTRACT. The stationary distribution of a continuous-time Markov chain is generally a complicated function of its transition rates. However, we show that if the transition rates are i.i.d. random variables with a common distribution satisfying certain tail conditions, then the resulting stationary distribution is close in total variation distance to the distribution that is proportional to the inverse of the exit rates of the states. This result, which generalizes and makes a precise prediction in [CBV$^+$21], constitutes the first rigorous validation of an emerging physical theory of order in nonequilibrium systems. The proof entails showing that the stationary distribution of the corresponding "jump chain," i.e., the discrete-time Markov chain with transition probabilities given by the normalized transition rates, is asymptotically uniform as the number of states grows, which settles a question raised in [BCC12] under certain assumptions.

1. INTRODUCTION AND MAIN RESULT

Background. The stationary distribution of an irreducible continuous-time Markov chain generally depends in a complicated way on its transition rates. The Markov chain tree theorem [LR86] makes this dependence explicit: A Markov chain on $n \geq 2$ states $[n] = \{1, \dots, n\}$ with transition rate matrix $Q = \{Q_{i,j}\}_{i \neq j}$ and unique stationary distribution π_Q satisfies

$$\pi_Q(i) \propto \sum_{\mathcal{T}} \prod_{(j,k) \in \mathcal{T}_i} Q_{j,k}, \qquad i \in [n]. \tag{1}$$

The sum in Eq. (1) ranges over spanning trees $\mathcal{T}$ of the underlying adjacency graph in which the vertices represent states, and $\mathcal{T}_i$ is the corresponding tree with its edges "directed toward" vertex i.

An important consequence of Eq. (1) is that there cannot be a function of only the rates incident to a state that determines the relative probability of the state under π_Q for every Q. This is akin to the central difficulty of non-equilibrium statistical physics, namely that there can be no analogue of the Boltzmann–Gibbs distribution for non-equilibrium steady states [Lan75]. In fact, even when Q and π_Q satisfy a detailed balance—the Markov chain analogue of thermal equilibrium—the

INSTITUTE FOR DATA ENGINEERING AND SCIENCE, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GA 30332

MATHEMATICAL INSTITUTE, LEIDEN UNIVERSITY, EINSTEINWEG 55, 2333 CC LEIDEN, THE NETHERLANDS

SCHOOL OF COMPUTER SCIENCE, GEORGIA INSTITUTE OF TECHNOLOGY, ATLANTA, GA 30332

E-mail addresses: calvert@gatech.edu, denholla@math.leidenuniv.nl, randall@cc.gatech.edu.

dependence of π_Q on the rates can still be complicated. Specifically, in this case, Eq. (1) becomes

$$\pi_Q(i) \propto \prod_{(j,k)\in\mathcal{P}_i} Q_{j,k}, \qquad i \in [n], \tag{2}$$

where $\mathcal{P}_i$ is a directed path in the underlying adjacency graph from a fixed reference state to i. For example, if Q is tridiagonal, then the relative probability of i potentially depends on many rates that are "far" from state i.

Although the stationary distribution is generally a complicated function of all the transition rates, recent work suggests that broad classes of Markov chains have stationary distributions that are closely approximated by simple functions of the exit rates $q_i = \sum_{k\neq i} Q_{i,k}$ [CR24]. For physical systems modeled by such Markov chains, short-time dynamical information suffices to predict the long-run fraction of time spent in a state. This is the claim of an emerging physical theory of order in non-equilibrium steady states, called "rattling" [CBV$^+$21]. Rattling theory has been validated through simulations of driven spin glasses [Gol21] and mechanical networks [KPSE23], numerical studies of Lorenz equations [JW21] and coupled Langevin equations with timescale separation [CE18], and experiments with robot swarms [CBV$^+$21] and systems of active microparticles [YBB$^+$22]. Despite mounting empirical evidence, rattling theory has yet to be proved to hold for a broad class of dynamical systems.

Goal. The purpose of this note is to rigorously establish a first instance of the rattling phenomenon in a large class of random Markov chains. Specifically, we prove that if the rates $\{Q_{i,j}\}_{i\neq j}$ are i.i.d. non-negative random variables with a finite mean μ and a cumulative distribution function $F(x) = \mathbf{P}(Q_{1,2} \leq x)$ satisfying certain left and right tail conditions, then π_Q converges in total variation distance to the distribution ν_Q that is inversely proportional to the exit rates:

$$\nu_Q(i) = \frac{q_i^{-1}}{\sum_{k\in[n]} q_k^{-1}}, \qquad i \in [n]. \tag{3}$$

Note that Q is non-reversible, so there is no simple expression for π_Q. Nonetheless, $\pi_Q(i)$ is closely approximated by a simple function of the rates incident to i (Fig. 1).

The convergence of π_Q to ν_Q follows from the fact that both converge to the uniform distribution $u_n = (1/n, \ldots, 1/n)$ on $[n]$. Our main theorem establishes this convergence for π_Q and ν_Q, as well as for the stationary distribution π_P of P, the jump chain of Q. The jump chain of Q is the discrete-time Markov chain with transition probabilities $P_{i,j} = Q_{i,j}/q_i$.

Assumption and main theorem. To state our main theorem, we need the following assumption on the cumulative distribution function of the transition rates $F(x) = \mathbf{P}(Q_{1,2} \leq x)$, $x \geq 0$.

Assumption (H). There exists a bounded sequence $(\epsilon_m)_{m\geq 1}$ of positive real numbers satisfying

$$\lim_{m\to\infty} \frac{\epsilon_m}{r_m} = \infty \quad \text{with} \quad r_m = \sqrt{(\log m)/m},$$

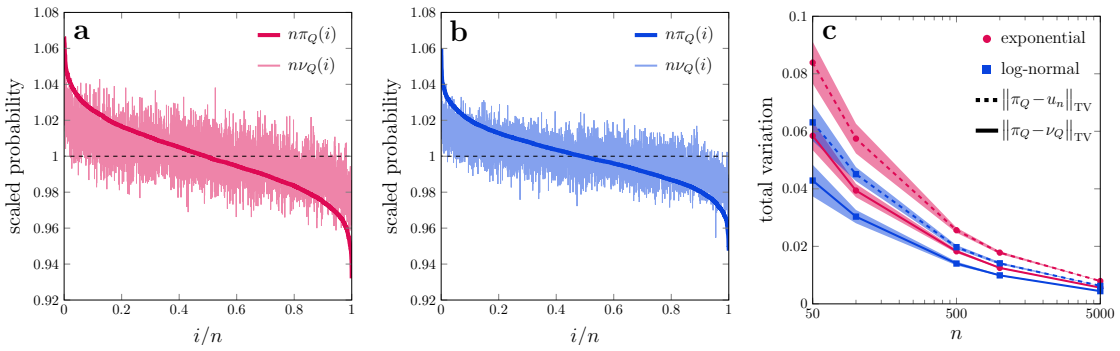

FIGURE 1. Comparison of π_Q, ν_Q, and the uniform distribution u_n for Markov chains with i.i.d. transition rates. (**a**,**b**) Values of $n\pi_Q(i)$ are placed in descending order and plotted with the corresponding values of $n\nu_Q(i)$, for $n = 5000$. The dashed line represents the uniform distribution. The transition rates are either (**a**) standard exponential or (**b**) log-normal with unit variance. (**c**) The distribution π_Q is closer in total variation distance to ν_Q than to uniform in each case, but both distances decrease as n increases from 50 to 5000. The solid and dashed lines represent averages over 10 independent trials, while the shaded regions represent these averages ± 1 standard deviation.

such that

(H1) $$\sum_{m\geq 1} m^2(1 - F(1/\epsilon_m)) < \infty,$$

(H2) $$\sum_{m\geq 1} m^2 F(\epsilon_m) < \infty.$$

Note that if ϵ_m were equal to $1/m$, then the conditions in **(H1)** and **(H2)** would be equivalent to the finiteness of the third moments of $Q_{1,2}$ and $1/Q_{1,2}$, respectively. However, since ϵ_m must grow faster than $\sqrt{(\log m)/m}$, it instead suffices for their $(5+\delta)$th moments to be finite for some $\delta > 0$.

Our main theorem is the following.

Theorem 1.1. *If F satisfies* (**H**)*, then, a.s. as $n \to \infty$,*

$$\|\nu_Q - u_n\|_1 = O(r_n), \qquad \|\pi_Q - u_n\|_1 = O(r_n/\epsilon_n), \quad \pi \in \{\pi_P, \pi_Q\}. \tag{4}$$

The asymptotic uniformity of ν_Q only requires the upper tail condition (**H1**), because it follows from the concentration of exit rates. The intuition behind the asymptotic uniformity of π_P is the fact that, although the transition probability matrix P is random, the corresponding Markov chain jumps to all other states with comparable probabilities. In particular, it rapidly converges to equilibrium. The uniformity of π_P settles a question raised in the context of random matrix theory by Bordenave, Caputo, and Chafaï [BCC12], except that we require stronger assumptions on the tails of the cumulative distribution function F (namely, [BCC12] asked whether π_P is asymptotically uniform when F has finite variance). The uniformity of π_P and ν_Q implies the same of π_Q, because π_Q is proportional to their product.

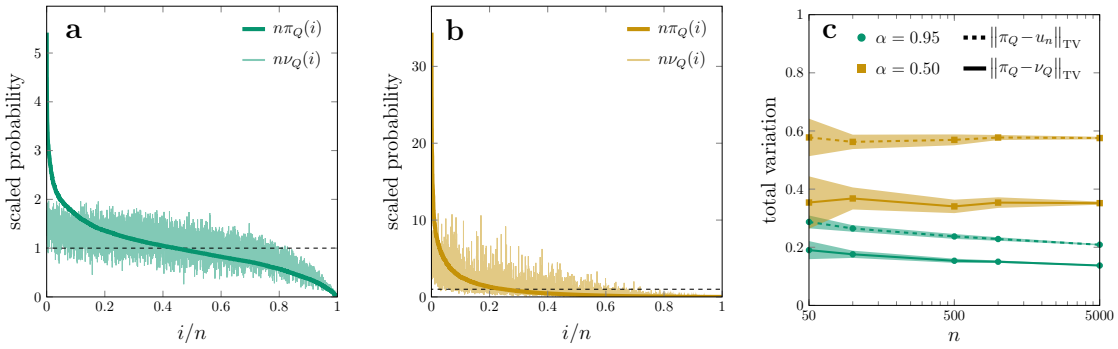

FIGURE 2. Non-uniformity of Markov chains with i.i.d. heavy-tailed transition rates. (**a**–**c**) The simulation parameters are the same as those in Fig. 1, except that the rates are equal in distribution to $X^{-1/\alpha}$, where X is a standard exponential random variable, and either (**a**) $\alpha = 0.95$ or (**b**) $\alpha = 0.5$. Note that if F is the cumulative distribution function of $X^{-1/\alpha}$, then $1 - F(x) \sim x^{-\alpha}$.

Theorem 1.1 and the triangle inequality imply that π_Q converges to ν_Q in total variation distance, which establishes that the main claim of rattling theory holds in a large class of random Markov chains.

Corollary 1.1. *If F satisfies* (**H**)*, then, a.s. $n \to \infty$,*

$$\|\pi_Q - \nu_Q\|_1 = O(r_n/\epsilon_n).$$

Note that if the support of F is bounded away from 0 and ∞, then (**H**) holds with all ϵ_n equal to a constant $\epsilon > 0$. In this case, the error term is simply $O(r_n)$.

Open problems. Although it may be possible to weaken hypotheses (**H1**) and (**H2**), it appears that conditions on the left and right tails of F are necessary. For example, Fig. 2 shows that the total variation distance of π_Q to ν_Q and u_n can be substantial when the rate distribution has a heavy right tail. In this case, it may still be possible for ν_Q to approximate π_Q, but in a different, namely, for a uniformly random state X the quantities $\log \pi_Q(X)$ and $\log \nu_Q(X)$ are highly correlated [CR24]. A related problem is to identify a natural class of random Markov chains for which π_P is close to uniform, but π_Q and ν_Q are not. This is true, for example, of Markov chains that have doubly stochastic jump chains and exit rates that vary over many exponential scales.

Organization. In Section 2, we state several key propositions and prove Theorem 1.1. In Section 3–Section 5 we prove these propositions. Section 3 addresses the concentration of the row and column sums of Q, because it serves as an input to the others. Section 4 shows that P is close to being doubly stochastic. Section 5 shows that P converges to equilibrium rapidly.

2. KEY INGREDIENTS AND PROOF OF THEOREM 1.1

We focus on proving the asymptotic uniformity of π_P, because the analogous result for π_Q follows more easily from the same ingredients and was already proved in [BCC14] by bounding

the spectrum of Q. The latter strategy does not immediately apply to π_P, as it is based on the independence of the entries of Q. Instead, our main tool is the correspondence between the rate of convergence to equilibrium of a Markov chain and the robustness of its stationary distribution to perturbations of its transition probabilities [CM01].

Key ingredients. We use one instance of the above correspondence, namely, one that measures the distance from equilibrium by

$$d(t) = \max_{i\in[n]} \|P^t(i,\cdot) - \pi_P\|_1, \qquad t \in \mathbb{Z}_+,$$

and the size of a perturbation $E \in \mathbb{R}^{n\times n}$ by its maximum absolute row sum

$$\|E\|_\infty = \max_{i\in[n]} \sum_{j\in[n]} |E_{i,j}|.$$

It states that if a Markov chain P converges rapidly and if P is close in $\|\cdot\|_\infty$ to a second Markov chain S, then the stationary distributions of P and S are close in total variation distance.

Proposition 2.1 (Theorem 3.1 of [Mit05])**.** *Let P be uniformly geometrically ergodic, in the sense that there exist positive numbers $\rho < 1$ and $C < \infty$ such that*

$$d(t) \le C\rho^t, \qquad t \in \mathbb{Z}_+. \tag{5}$$

If S is a stochastic matrix with invariant distribution π_S, then

$$\|\pi_P - \pi_S\|_1 \le \left(\hat{t} + \frac{C\rho^{\hat{t}}}{1-\rho}\right) \|P - S\|_\infty, \tag{6}$$

where $\hat{t} = \lceil \log_\rho C^{-1} \rceil$.

In light of Proposition 2.1, to prove that π_P is approximately uniform, it suffices to show that P is close to a doubly stochastic matrix in the ∞-norm and that P converges rapidly. The following propositions address these aspects in turn. We will show that, because the row and column sums of Q are concentrated, P is a small perturbation of a doubly stochastic matrix.

Proposition 2.2. *If F satisfies* (**H**)*, then, a.s. for all sufficiently large n, there exists a matrix $E \in \mathbb{R}^{n\times n}$ such that $P + E$ is doubly stochastic. Additionally, $\|E\|_\infty = O(r_n)$ a.s. as $n \to \infty$.*

Estimating the speed at which P converges to equilibrium is complicated by the fact that P is generally non-reversible and its entries are dependent. However, we will show that, because the row sums of Q are concentrated and because extremely small rates are rare, P converges to equilibrium rapidly in n.

Proposition 2.3. *If F satisfies* (**H**)*, then P a.s. satisfies Eq.* (5) *with $\rho = \sqrt{1 - \epsilon_n/2\mu}$ and $C = 2$, for all sufficiently large n.*

Propositions 2.1 to 2.3 together imply the conclusion of Theorem 1.1 regarding π_P. The conclusions regarding π_Q and ν_Q require a concentration result. To state this, we denote by $\{X_i\}_{i\in[2n]}$ the n row sums and n column sums of Q, excluding the diagonal entries:

$$X_i = \begin{cases} \sum_{j\in[n]\setminus\{i\}} Q_{i,j}, & i \in [n], \\ \sum_{j\in[n]\setminus\{i\}} Q_{j,i}, & i \in [2n] \setminus [n]. \end{cases}$$

Additionally, because the proof involves truncating the summands of X_i by $1/\epsilon_n$, we introduce notation for the mean of the truncated rate distribution. Specifically, we denote

$$\mu_n = \left(\frac{n-1}{n}\right) \mathbf{E}\left(Q_{1,2}\, \mathbb{1}_{\{Q_{1,2}\le 1/\epsilon_n\}}\right).$$

We include the asymptotically negligible factor of $(n-1)/n$ to write $(n-1)$ times the truncated mean of F more compactly as $\mu_n n$. The following remark explains why we can assume that μ_n is uniformly positive despite the truncation.

Remark 2.1. We may assume without loss of generality that if (**H1**) or (**H2**) holds, then (i) $\mu_n \ge 3\mu/4$ and (ii) $\epsilon_n/2\mu \le 1/2$ for all $n \ge 2$. Note that if either hypothesis holds for a sequence $(\epsilon_m)_{m\ge1}$, then it also holds for the sequence $(\epsilon'_m)_{m\ge1}$ with $\epsilon'_m = c\epsilon_m$ for a constant $c \in (0,1)$ because F is non-decreasing. Since $(\epsilon_m)_{m\ge1}$ is bounded by assumption, we can always choose c small enough depending on F such that (i) and (ii) are true. The proofs of several propositions require that $(\mu_n)_{n\ge2}$ is uniformly positive, and further require that $(1-\epsilon_n/2\mu)^{-1} \le 2$. Note that replacing ϵ_m by ϵ'_m does not affect the conclusion of Theorem 1.1.

With the preceding notation and remark in mind, we state the final input to the proof of Theorem 1.1.

Proposition 2.4. *If F satisfies* (**H1**)*, then a.s. there exists a random $N < \infty$ such that the event*

$$\mathcal{E}_N = \bigcap_{n\ge N} \bigcap_{i\in[2n]} \left\{ |X_i - \mu_n n| \le \sqrt{2} n r_n/\epsilon_n \right\}$$

occurs. Moreover, when $\mathcal{E}_N$ occurs, the column sums of P satisfy the following estimate a.s. as $n\to\infty$, uniformly in $j \in [n]$:

$$\sum_{k\in[n]} P_{k,j} = 1 + O(r_n).$$

Proof of main theorem. We now prove Theorem 1.1, assuming the preceding propositions.

Proof of Theorem 1.1. The cumulative distribution function F of the rates satisfies (**H**), so the conclusions of Propositions 2.2 to 2.4 apply. Like the conclusions of these propositions, the following equations should be interpreted to hold a.s. as $n\to\infty$. We consider ν_Q, π_P, and π_Q in turn.

First, by Proposition 2.4, the exit rates satisfy

$$q_i = \sum_{j\ne i} Q_{i,j} = \mu_n n(1 + O(r_n)). \tag{7}$$

This estimate holds uniformly in $i \in [n]$, so ν_Q satisfies

$$\nu_Q(i) = \frac{q_i^{-1}}{\sum_{k\in[n]} q_k^{-1}} = \frac{1}{n} + O(r_n/n),$$

which then implies that

$$\|\nu_Q - u_n\|_1 = \sum_{k\in[n]} \left|\nu_Q(i) - \frac{1}{n}\right| = \sum_{k\in[n]} O(r_n/n) = O(r_n).$$

Second, by Proposition 2.3, the distance from equilibrium $d(t)$ satisfies

$$d(t) \le 2\rho^t, \qquad t \in \mathbb{Z}_+,$$

in terms of $\rho = \sqrt{1 - \epsilon_n/2\mu}$ and the sequence $(\epsilon_n)_{n\ge1}$ from (**H**). In particular, the condition of Proposition 2.1 is met for $C = 2$ and this choice of ρ. Proposition 2.1 consequently implies that every $n \times n$ doubly stochastic matrix S satisfies

$$\|\pi_P - u_n\| \le \left(\hat{t} + \frac{2\rho^{\hat{t}}}{1-\rho}\right) \|P - S\|_\infty, \tag{8}$$

where $\hat{t} = \lceil \log_\rho(1/2) \rceil$, because every such S has a uniform stationary distribution. Some straightforward but tedious algebra, which uses the fact that $\epsilon_n/2\mu \le 1/2$ (Remark 2.1), shows that both $\hat{t}$ and $2\rho^{\hat{t}}/(1-\rho)$ are $O(1/\epsilon_n)$. The asymptotic uniformity of π_P then follows from Eq. (8) and Proposition 2.2, which guarantees the existence of an S such that $\|P - S\|_\infty = O(r_n)$:

$$\|\pi_P - u_n\|_1 = O(r_n/\epsilon_n).$$

Third, the conclusion concerning π_Q is due to the identity $\pi_Q(i) \propto \pi_P(i)/q_i$, the estimate for the exit rates in Eq. (7), and the conclusion for π_P:

$$\begin{aligned}\|\pi_Q - u_n\|_1 &= \sum_{i\in[n]} \left|\frac{\pi_P(i)/q_i}{\sum_{k\in[n]} \pi_P(k)/q_k} - \frac{1}{n}\right| \\ &= \sum_{i\in[n]} \left|\pi_P(i)(1 + O(r_n)) - \frac{1}{n}\right| \le \|\pi_P - u_n\|_1 + O(r_n) = O(r_n/\epsilon_n).\end{aligned}$$

□

3. Estimates of transition rates and probabilities

First, we establish the concentration of row and column sums of Q, as stated in Proposition 2.4. We then prepare bounds on the largest and smallest transition probabilities, which will be useful in the sections that follow.

Proof of Proposition 2.4. We must show that a.s. the event

$$\mathcal{A}_{n,i} = \left\{|X_i - \mu_n n| \le \sqrt{2}\, n r_n/\epsilon_n\right\}$$

occurs for all $i \in [2n]$, for all sufficiently large n, that is,

$$\mathbf{P}\Big(\bigcup_{N\geq 1}\bigcap_{n\geq N}\bigcap_{i\in[2n]}\mathcal{A}_{n,i}\Big) = \mathbf{P}\Big(\bigcup_{N\geq 1}\mathcal{E}_N\Big) = 1.$$

By the Borel–Cantelli lemma, it suffices to show that

$$\sum_{n\geq 1}\mathbf{P}\Big(\bigcup_{i\in[2n]}\mathcal{A}_{n,i}^c\Big) < \infty. \tag{9}$$

Our strategy is to reduce Eq. (9) to a bound on $\mathbf{P}(\mathcal{A}_{n,1}^c)$ that we can prove by using the Hoeffding inequality (see [Hoe63] and [Ver18, Theorem 2.6.2]). The summands of $X_1 = \sum_{j=2}^n Q_{1,j}$ are not necessarily bounded, so we must truncate them first. Consider the sum $Y_1 = \sum_{j=2}^n R_j$ of truncated rates $R_j = Q_{1,j}\mathbb{1}_{\{Q_{1,j}\leq 1/\epsilon_n\}}$, and the events

$$\mathcal{V}_n = \{Q_{1,j} \leq 1/\epsilon_n, \forall\, 2 \leq j \leq n\}, \qquad \mathcal{B}_n = \Big\{|Y_1 - \mu_n n| \leq \sqrt{2}\, n r_n/\epsilon_n\Big\}.$$

Note that $X_1 = Y_1$ when $\mathcal{V}_n$ occurs, and $\mathcal{A}_{n,1}$ occurs when both $\mathcal{V}_n$ and $\mathcal{B}_n$ do. The probability $\mathbf{P}(\mathcal{A}_{n,1}^c)$ consequently satisfies

$$\mathbf{P}(\mathcal{A}_{n,1}^c) \leq \mathbf{P}(\mathcal{A}_{n,1}^c \cap \mathcal{V}_n) + \mathbf{P}(\mathcal{V}_n^c) \leq \mathbf{P}(\mathcal{B}_n^c) + \mathbf{P}(\mathcal{V}_n^c). \tag{10}$$

The Hoeffding inequality states that, because the summands of Y_1 are independent random variables taking values in $[0, 1/\epsilon_n]$, the event $\mathcal{B}_n^c$ has probability

$$\mathbf{P}(\mathcal{B}_n^c) \leq 2\exp\left(-\frac{2(\sqrt{2}\, n r_n/\epsilon_n)^2}{n(1/\epsilon_n)^2}\right) = 2n^{-4}.$$

Concerning $\mathcal{V}_n^c$, the union bound implies that

$$\mathbf{P}(\mathcal{V}_n^c) \leq \sum_{j=2}^n \mathbf{P}(Q_{1,j} > 1/\epsilon_n) \leq n(1 - F(1/\epsilon_n)).$$

We use the preceding bounds to show that the sum in Eq. (9) is finite:

$$\begin{aligned}\sum_{n\geq 1}\mathbf{P}\Big(\bigcup_{i\in[2n]}\mathcal{A}_{n,i}^c\Big) &\leq \sum_{n\geq 1} 2n\,\mathbf{P}(\mathcal{A}_{n,1}^c)\\ &\leq \sum_{n\geq 1} 2n\big(\mathbf{P}(\mathcal{B}_n^c) + \mathbf{P}(\mathcal{V}_n^c)\big)\\ &\leq \sum_{n\geq 1} 2n\big(2n^{-4} + n(1 - F(1/\epsilon_n))\big) < \infty.\end{aligned}$$

The first inequality holds by the union bound and the fact that $(X_i)_{i\in[2n}$ are identically distributed, the second holds by (10), the third holds by the preceding bounds on $\mathbf{P}(\mathcal{B}_n^c)$ and $\mathbf{P}(\mathcal{V}_n^c)$, and the fourth holds by (**H1**).

Next, we estimate $\sum_k P_{k,j}$. Under assumption (**H1**) and when $\mathcal{E}_N$ occurs, each row and column sum of the rate matrix equals $\mu_n n + O(n r_n/\epsilon_n)$. Hence, the column sums of the transition

probability matrix satisfy

$$\sum_k P_{k,j} = \sum_k \frac{Q_{k,j}}{\sum_{\ell \neq k} Q_{k,\ell}} = \frac{\sum_k Q_{k,j}}{\mu_n n + O(n r_n/\epsilon_n)} = \frac{\mu_n n + O(n r_n/\epsilon_n)}{\mu_n n + O(n r_n/\epsilon_n)} = 1 + O(r_n/\epsilon_n). \tag{11}$$

The final equality uses the fact that $(\mu_n)_{n\geq 2}$ is uniformly positive (see Remark 2.1). □

Next, we bound the largest transition probability. While the strength of (**H1**) is necessary for estimating the row and column sums, the bound it implies on the largest transition probability is stronger than we need.

Proposition 3.1. *If F satisfies* (**H1**)*, then, a.s. for all sufficiently large n,*

$$\max_{i\neq j} P_{i,j} \leq 2/\mu n \epsilon_n.$$

Proof. It suffices to show that, a.s. for all sufficiently large n,

$$\max_{i\neq j} Q_{i,j} \leq 1/\epsilon_n,$$

because, by Proposition 2.4, this implies that

$$\max_{i\neq j} P_{i,j} = \max_{i\neq j} \frac{Q_{i,j}}{\sum_{k\neq i} Q_{i,k}} = \frac{\max_{i\neq j} Q_{i,j}}{\mu_n n(1+O(r_n))} \leq 2/\mu n \epsilon_n.$$

Note that the upper bound uses the fact that μ_n is at least $3\mu/4$ (Remark 2.1).

By the Borel–Cantelli lemma, it is enough to show that the probability of the event $\{\max_{i\neq j} Q_{i,j} > 1/\epsilon_n\}$ is summable. This follows from a union bound and the fact that the $Q_{i,j}$ are identically distributed, and the hypothesis (**H1**):

$$\mathbf{P}\Big(\max_{i\neq j} Q_{i,j} > 1/\epsilon_n\Big) \leq n^2\, \mathbf{P}\big(Q_{1,2} > 1/\epsilon_n\big) \leq n^2(1 - F(1/\epsilon_n)).$$

□

Lastly, we bound the smallest transition probability. The proof of the following proposition motivates the form of (**H2**).

Proposition 3.2. *If F satisfies* (**H**)*, then, a.s. for all sufficiently large n,*

$$\min_{i\neq j} P_{i,j} \geq \epsilon_n/2\mu n.$$

Proof. We claim that, a.s. for all sufficiently large n, the smallest off-diagonal rate satisfies

$$\min_{i\neq j} Q_{i,j} \geq \epsilon_n, \tag{12}$$

in terms of the sequence $(\epsilon_n)_{n\geq 1}$ from (**H2**). Indeed, a union bound and the fact that the $(Q_{i,j})$ are identically distributed implies that

$$\mathbf{P}\Big(\min_{i\neq j} Q_{i,j} \leq \epsilon_n\Big) \leq n^2 F(\epsilon_n).$$

Under hypothesis (**H2**), the right-hand side is summable. The Borel–Cantelli lemma therefore implies the claim.

Using Eq. (12), we bound $P_{i,j}$ from below according to

$$\min_{i\neq j} P_{i,j} = \min_{i\neq j} \frac{Q_{i,j}}{\sum_{k\neq i} Q_{i,k}} \geq \frac{\min_{i\neq j} Q_{i,j}}{\mu_n n(1+O(r_n))} \geq \frac{\epsilon_n}{2\mu n}.$$

The first inequality holds for all sufficiently large n under hypothesis (**H1**), due to Proposition 2.4. The second holds by Eq. (12) and the fact that $\mu_n \leq \mu$. □

4. Approximation by a doubly stochastic Markov chain

The goal of this section is to prove Proposition 2.2, which entails finding a doubly stochastic matrix close to P in the ∞-norm. To our knowledge, there is no formula for the doubly stochastic matrix closest to a given stochastic one, nor for the ∞-norm of their difference. However, there is such a formula for the closest matrix with row and column sums equal to 1, but possibly with negative entries, in the distance induced by the Frobenius norm [Kho98]. We use this matrix as our starting point, and then perturb it to make its entries non-negative, resulting in a doubly stochastic matrix that is close enough to obtain a good estimate from Eq. (6).

Construction. First, given a stochastic matrix P, we define $G = G(P)$ to be the matrix with entries

$$G_{i,j} = P_{i,j} + \frac{1}{n}\left(1 - \sum_{k\in[n]} P_{k,j}\right).$$

It is easy to check that all the row and column sums of G are equal to 1. However, G is not necessarily doubly stochastic because some of its entries may be negative. For example, in our case, the diagonal entries of P are equal to 0. Consequently, if the jth column sum of P exceeds 1, then $G_{i,j}$ is negative.

To make G non-negative while preserving its row and column sums, we will add to it the matrix $H = H(\delta)$ with entries

$$H_{i,j} = \begin{cases} \delta & i = j, \\ -\delta/(n-1) & i \neq j, \end{cases}$$

where δ is the largest absolute entry of $G - P$, i.e.,

$$\delta = \max_{j\in[n]} \frac{1}{n}\left|1 - \sum_{k\in[n]} P_{k,j}\right|.$$

Note that all the row and column sums of H are equal to 0. The matrix $G + H$ is doubly stochastic a.s. for all sufficiently large n because the column sums of P tightly concentrate around 1. Consequently, the diagonal entries of G are not too negative, and the off-diagonal entries of G are sufficiently positive to redistribute $\delta/(n-1)$ per entry to the diagonal according to H, without making an off-diagonal entry negative.

Proof of Proposition 2.2. We must show that, a.s. for all sufficiently large n, $G + H$ is doubly stochastic and $E = (G + H) - P$ satisfies the ∞-norm bound in the statement.

First, because the row and column sums of $G + H$ are equal to 1 for any choice of δ, it suffices to show that its entries are non-negative. By our choice of δ, the diagonal entries of $G + H$ are necessarily non-negative, while the off-diagonal entries satisfy

$$\min_{i\neq j}(G+H)_{i,j} \geq \min_{i\neq j} P_{i,j} - \max_{j\in[n]} \frac{1}{n}\left|1 - \sum_{k\in[n]} P_{k,j}\right| - \frac{\delta}{n-1} = \min_{i\neq j} P_{i,j} - \left(1 + \frac{1}{n-1}\right)\delta. \tag{13}$$

Under assumption (**H2**), there exists a sequence $(\epsilon_n)_{n\geq 1}$ such that $\epsilon_n = \omega(r_n)$ and, by Proposition 3.2, for which $P_{i,j} = \Omega(\epsilon_n/n)$ a.s. Under assumption (**H1**), Proposition 2.4 further implies that $\delta = O(r_n/n)$ a.s., so the lower bound in Eq. (13) is non-negative for all sufficiently large n:

$$\min_{i\neq j}(G+H)_{i,j} \geq \min_{i\neq j} P_{i,j} - \left(1 + \frac{1}{n-1}\right)\delta = \Omega(\epsilon_n/n) - O(r_n/n) = \Omega(\epsilon_n/n).$$

Second, we observe that the absolute row sums of E are identical. Hence, by Proposition 2.4, the ∞-norm of E satisfies

$$\|E\|_\infty \leq \sum_{j\in[n]} \frac{1}{n}\left|1 - \sum_{k\in[n]} P_{k,j}\right| + 2\delta = O(r_n).$$

□

5. Convergence rate via global minorization

We use Doeblin's minorization condition to establish Eq. (5) with a value of ρ that depends on n through the sequence $(\epsilon_n)_{n\geq 1}$ in (**H**), and a value of C that does not. The following statement appears in [MT09].

Lemma 5.1. *If there exist $\ell \geq 1$, $\eta > 0$, and a probability measure λ on $[n]$ such that*

$$(P^\ell)_{i,j} \geq \eta\lambda(j), \qquad i, j \in [n], \tag{14}$$

then

$$d(t) \leq (1-\eta)^{\lfloor t/\ell \rfloor}, \qquad t \in \mathbb{Z}_+.$$

Because the typical off-diagonal entry of P is $1/n$, it would seem that Doeblin's condition holds with η of constant order and $\lambda(i)$ identically equal to $1/n$. However, since P has diagonal entries equal to 0, we consider P^2 instead. The conclusion of Lemma 5.1 then implies Proposition 2.3.

Proof of Proposition 2.3. The proof entails uniformly bounding below $(P^2)_{i,j}$. As in the proof of Proposition 2.2, equations that involve random quantities are meant to hold a.s. for all sufficiently large n.

We begin by observing that the entries of P^2 deterministically satisfy

$$(P^2)_{i,j} = \sum_{k\in[n]} P_{i,k}P_{k,j} \geq \Big(\min_{k\neq l} P_{k,l}\Big)\Big(1 - \max_{k\neq l} P_{k,l}\Big), \qquad i,j \in [n].$$

According to Propositions 3.1 and 3.2, which apply under assumptions (**H1**) and (**H2**), the extreme off-diagonal transition probabilities satisfy

$$\min_{k\neq l} P_{k,l} \geq \epsilon_n/2\mu n \quad \text{and} \quad \max_{k\neq l} P_{k,l} \leq 2/\mu n\epsilon_n,$$

where $(\epsilon_n)_{n\geq 1}$ is the sequence in (**H2**). Since $\epsilon_n = \omega(r_n)$ by assumption, the quantity $2/\mu n\epsilon_n$ is $o(1)$. As a consequence, the minorization condition Eq. (14) holds with $\ell = 2$, $\eta = \epsilon_n/2\mu$, and $\lambda = u_n$. The distance $d(t)$ then satisfies the following bound, in terms of $\rho = \sqrt{1-\eta}$:

$$d(t) \leq (1-\eta)^{\lfloor t/2\rfloor} \leq (1-\eta)^{-1}\rho^t.$$

The first inequality holds by Lemma 5.1, while the second is due to the bound $\lfloor t/2\rfloor \geq t/2 - 1$. Lastly, recall from Remark 2.1 that $(1-\eta)^{-1} \leq 2$ for all $n \geq 2$. □

Acknowledgements. This work was supported by the National Science Foundation (NSF) under grant no. DMS-1928930, while the authors were in residence at the Simons Laufer Mathematical Sciences Institute in Berkeley, California, during the Spring 2025 semester. J.C. and D.R. further acknowledge support through the NSF award CCF-2106687 and the US Army Research Office Multidisciplinary University Research Initiative award W911NF-19-1-0233.

REFERENCES

[BCC12] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Circular law theorem for random Markov matrices. *Probability Theory and Related Fields*, 152(3):751–779, 2012.

[BCC14] Charles Bordenave, Pietro Caputo, and Djalil Chafaï. Spectrum of Markov generators on sparse random graphs. *Communications on Pure and Applied Mathematics*, 67(4):621–669, 2014.

[CBV+21] Pavel Chvykov, Thomas A. Berrueta, Akash Vardhan, William Savoie, Alexander Samland, Todd D. Murphey, Kurt Wiesenfeld, Daniel I. Goldman, and Jeremy L. England. Low rattling: A predictive principle for self-organization in active collectives. *Science*, 371(6524):90–95, 2021.

[CE18] Pavel Chvykov and Jeremy England. Least-rattling feedback from strong time-scale separation. *Physical Review E*, 97:032115, 2018.

[CM01] Grace E. Cho and Carl D. Meyer. Comparison of perturbation bounds for the stationary distribution of a Markov chain. *Linear Algebra and its Applications*, 335(1):137–150, 2001.

[CR24] Jacob Calvert and Dana Randall. A local–global principle for nonequilibrium steady states. *Proceedings of the National Academy of Sciences*, 121(42):e2411731121, 2024.

[Gol21] Jacob Gold. *Self-organized fine-tuned response in a driven spin glass*. Ph.D. thesis, Massachusetts Institute of Technology, February 2021.

[Hoe63] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[JW21] Zachary Jackson and Kurt Wiesenfeld. Emergent, linked traits of fluctuation feedback systems. *Physical Review E*, 104:064216, 2021.

[Kho98] Raja Nicolas Khoury. Closest matrices in the space of generalized doubly stochastic matrices. *Journal of Mathematical Analysis and Applications*, 222(2):562–568, 1998.

[KPSE23] Hridesh Kedia, Deng Pan, Jean-Jacques Slotine, and Jeremy L. England. Drive-specific selection in multistable mechanical networks. *The Journal of Chemical Physics*, 159(21):214106, 2023.

[Lan75] Rolf Landauer. Inadequacy of entropy and entropy derivatives in characterizing the steady state. *Physical Review A*, 12(2):636–638, 1975.

[LR86] Frank Leighton and Ronald Rivest. Estimating a probability using finite memory. *IEEE Transactions on Information Theory*, 32(6):733–742, 1986.

[Mit05] Alexander Y. Mitrophanov. Sensitivity and convergence of uniformly ergodic Markov chains. *Journal of Applied Probability*, 42(4):1003–1014, 2005.

[MT09] Sean Meyn and Richard L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2nd edition, 2009.

[Ver18] Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2018.

[YBB+22] Jing Fan Yang, Thomas A. Berrueta, Allan M. Brooks, Albert Tianxiang Liu, Ge Zhang, David Gonzalez-Medrano, Sungyun Yang, Volodymyr B. Koman, Pavel Chvykov, Lexy N. LeMar, Marc Z. Miskin, Todd D. Murphey, and Michael S. Strano. Emergent microrobotic oscillators via asymmetry-induced order. *Nature Communications*, 13(1):5734, 2022.